Functional characterization and application of exapted transposable element gene family *MUSTANG-A*

By

Yang Shao

Department of Biology

McGill University, Montreal

June 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

"I've seen things you people wouldn't believe.
Attack ships on fire off the shoulder of Orion.
I watched C-beams glitter in the dark near the Tannhäuser Gate.

Blade Runner

Table of Contents

ABSTRACT

Transposable elements (TEs) are DNA sequences that are able to "jump" in the genome. Despite most TE jumping events having neutral or deleterious effects, they also promote genetic innovation. These TEs that confer benefits are called exapted transposable elements (ETEs). The *MUSTANG-A* (*MUG-A*) family, derived from *Mutator*-like elements (MULEs), is the first ETE family identified systematically in *Arabidopsis thaliana*. The mutants of *MUG-A* genes displayed developmental defects and abiotic stress sensitivity. The main goal of my thesis is through exploring the molecular profiles of *MUG-A* genes, combined with genetic and physiological evidence to understand the mechanisms underlying these phenotypes. Moreover, the agricultural potential of ETEs is also explored using *MUG4* as a representative.

In Chapter II, I found all MUG-A proteins have a strict nuclear localization and mediate transcriptional repression in plants. They also form homodimers and heterodimers with each other. Genetic evidence suggests *MUG-A* genes have redundant roles in plant development. In addition, RNA-seq analysis suggests a set of pathways are regulated by *MUG-A* genes. This detailed molecular and genetic evidence suggests MUG-A proteins may work as a transcriptional repressor complex that regulates a set of genes essential for plant development. In Chapter III, I focus on the mechanism and agriculture potential of *MUG4* under salt stress conditions. MUG4 specifically prevents plants from accumulating extra sodium, potassium, and calcium. The expression of abiotic stress-responsive genes was significantly higher in *mug4* mutant plants than wild-type plants under salt stress conditions. Furthermore, overexpression of AtMUG4 increases salt tolerance in both *A. thaliana* and *Camelina sativa*. In Chapter IV, I further dissect the functional roles of three domains in MUG4 using a site-directed mutagenesis approach. The N-

terminal domain is essential for the nucleus subcellular localization of MUG4 and the E254 mutation in the middle domain partially abolished the nuclear targeting. All three domains play critical roles in the interaction of MUG4 with MUG1 and MUG2. In addition, The N-terminal and C-terminal domains are important for plant salt response and the middle domain might be required for optimal salt defence. The ability of MUG4 to defend from high salinity largely correlates with its heterodimer formation with MUG1 and MUG2.

In the end, the work in the present thesis fills the void of a detailed molecular profile of *MUG-A* genes, creates a deeper understanding of the function of MUG4 in salt response, revealed discrete roles of its multiple domains, opens a new door for valuable roles of ETEs in plant adaptation to stress and provides great potential for crop improvement.

RÉSUMÉ

Les éléments transposables (TEs) sont des séquences d'ADN capables de se déplacer d'un endroit à un autre dans le génome. Bien que la plupart des insertions des TE aient des effets neutres ou délétères, elles favorisent également l'innovation génétique, et des insertions des TE bénéfiques ont été signalées à l'occasion. Ces TE qui confèrent des avantages sont appelés éléments transposables exaptés (ETEs). La famille *MUSTANG-A* (*MUG-A*), dérivée de la superfamille des transposons *Mutator*-like elements (MULEs), est la première famille des ETE identifiée systématiquement dans *Arabidopsis thaliana*. Les mutants des gènes de *MUG-A* ont montré des défauts développementaux et la sensibilité abiotique d'effort. L'objectif principal de ma thèse est d'explorer les profils moléculaires des gènes *MUG-A*, combinés avec des preuves génétiques et physiologiques pour comprendre les mécanismes sous-jacents à ces phénotypes. De plus, le potentiel agricole des ETE est également exploré en utilisant *MUG4* comme représentant.

Au chapitre 2, j'ai effectué un ensemble complet d'expériences pour explorer les profils moléculaires des gènes *MUG-A*. Toutes les protéines MUG-A ont une localisation nucléaire stricte et toutes peuvent médier la répression transcriptionnelle chez les plantes. Grâce à l'analyse phénotypique génétique, il a été démontré que les gènes *MUG-A* peuvent former des homodimères et des hétérodimères avec des fonctions qui se chevauchent et qui sont bénéfiques pour le développement des plantes. Ces preuves moléculaires et génétiques détaillées suggèrent que les protéines *MUG-A* fonctionnent comme un complexe répresseur transcriptionnel qui régule un ensemble de gènes essentiels au développement des plantes. Dans le chapitre 3, je me concentre sur le mécanisme moléculaire et l'application de *MUG4* dans le terrain dans des conditions de stress salin. Grâce à une série de tests physiologiques, j'ai découvert que *MUG4*

empêche spécifiquement les plantes d'accumuler du sodium, du potassium et du calcium supplémentaires. L'expression des gènes sensibles au stress abiotique était significativement plus élevée chez les plantes mutantes *mug4* que chez les plantes de type sauvage dans des conditions de stress salin. La surexpression d'AtMUG4 augmente la tolérance au sel à la fois chez *A. thaliana* et *Camelina sativa*. Dans le chapitre 4, je décortique les rôles fonctionnels de trois domaines dans MUG4 en utilisant une approche de mutagenèse dirigée. Le domaine N-terminal est essentiel pour la localisation subcellulaire du noyau de MUG4 et la mutation E254 dans le domaine médian a partiellement aboli le ciblage nucléaire. Les trois domaines jouent un rôle essentiel dans l'interaction de MUG4 avec MUG1 et MUG2. De plus, les domaines N-terminal et C-terminal sont importants pour la réponse au sel de la plante et le domaine médian pourrait être requis pour une défense optimale contre le sel. La capacité à défendre la salinité élevée de MUG4 corrèle en grande partie sa formation d'hétérodimères avec MUG1 et MUG2.

À la fin, le travail de la thèse a comblé le vide du profil moléculaire détaillé des gènes *MUG-A*, créé une compréhension plus approfondie de la fonction de MUG4 sous-jacente au phénotype de ses mutants, révélé les rôles discrets de ses multiples domaines, ouvre une nouvelle porte à de précieux rôles des été dan' l'adaptation des plantes au stress et offre un grand potentiel pour l'amélioration des cultures.

ACKNOWLEDGEMENTS

I sit in the Redpath Library and look at the snow on the campus outside of the window, all sorts of feelings well up in my mind. First, I would like to sincerely thank my supervisor Thomas Bureau. His calm, patient and trust allow me to wander around the unknown world of science, his unique training way protected my enthusiasm for scientific research and cultivated my confidence and independence, his profound and thoughtful insight sharpened my thinking and brought my scope to a higher level. I feel truly honoured to be his student. I would also like to thank my supervisory committee members, Drs. Daniel Schoen, and Tamara Western for organizing a journal club in my closely-related field to build and enhance my scientific sense and also their availability to answer my questions and provide constructive advice. I am grateful for the extra academic support from Dr. Hugo Zheng. Many thanks to the former and current members of the Bureau lab**.** A special thank you to Zoé Joly-Lopez for being so generous to give of your time to help me troubleshoot with research ideas and lab techniques. Thank you, Maia Kaplan, Navei Hernández and John Alexander Aguirre for being ideal lab companions, as well as for reviewing, proofreading my draft. I would also like to acknowledge the financial support provided by China Scholarship Council.

Many thanks to my dear friends, Anna, Tiffany, Maia, Lucie, Sarah for the inspirations and all the efforts to integrate me into the local circle. Thank Zheng Huan, Gongteng, Meteora and their families for always being there no matter up or down time throughout my Ph.D. Finally, I want to thank my lover, Jiaqi Sun. His passion for research, honesty and sincerity personality inspired me to become a better me.

PREFACE AND CONTRIBUTIONS OF AUTHORS

This dissertation of my Ph.D. study is comprised of five chapters, an introduction and literature review (Chapter I), three original research chapters (Chapter II, III and IV) and a conclusion chapter (Chapter V). All the research chapters contain their own abstract, introduction, results, discussion, methods, and reference. All the chapters in the thesis are written by me.

Chapter I is a literature review, giving an introduction on the widespread knowledge of TEs, in the field of TE classification, main TE families, TE silencing and activation. Also, I described the ETE gene families, their function in the host and discussed the potential application of ETEs for crop improvements.

Chapter II is a work aiming to fill the blank of the molecular profile of a particular ETE gene family, *MUG-A* which consists of four members, *MUG1* to *MUG4*. I performed all the experimental work. Transcriptional activation activity in yeast and plant, subcellular localization determination, bimolecular fluorescent complimentary assay, yeast two-hybrid assay were performed with the assist of Jiaqi Sun, bioinformatic analysis of RNA-seq data, *mug1 mug2 mug4* triple mutant building and phenotyping were performed by me. The RNA-seq data were generated by Zoé Joly-Lopez. The plasmids used in subcellular localization determination were built by me. The original plasmids used for yeast two-hybrid assay and bimolecular fluorescent complimentary assay were kindly provided by Drs. Yuhai Cui and Dr. Shoji Mano respectively and modified by me.

Chapter III verified AtMUG4 as an ETE that is essential for plant salt response, explored the pathway MUG4 may be involved in *Arabidopsis thaliana* by using a physiological and

molecular approach. In this project, I also build the transgenic lines in *Arabidopsis thaliana* and *Camelina sativa* that overexpress AtMUG4 and tested that these plants are tolerant to salt stress. I designed and produced all the results of this project with the advice of my supervisor and supervisory committee. Mingyong Cho, Doris Zhang, Alice Liu helped with testing the plant survival rate under LiCl, KCl, $CaCl_2$ and relative growth rate under osmotic stress, Cl^--dominant stress and oxidative stress that are present in Figures 3 and 4. Jiaqi Sun helped with all the phenotypic pictures taken. Emilio Vello performed the relative growth rate analysis in Figure 3.

Chapter IV further dissected the domain function of MUG4. This work described a structural basis of MUG4, revealed discrete roles of its three domains and provided functional support for how MUG4 works for salt tolerance. I designed and performed all the experiments, including the site-directed mutagenesis to determine the domains that are essential for MUG4 nuclear targeting, the interaction of MUG4 with MUG1 and MUG2 and for plant salt response.

ABBREVIATIONS

ABA - Abscisic Acid

ALP - Antagonist of Like heterochromatin Protein

BiFC - Bimolecular Fluorescence Complementation

CaMV - Cauliflower Mosaic Virus

cDNA - complementary DNA

Col-0 - Columbia-*0*

CSB - Cockayne syndrome B

cYFP - C-terminal Yellow Fluorescent Protein

DAPI - 4′,6-diamidino-2-phenylindole

DBD - DNA-Binding Domain

DCL - DICER-like

DEG - Differential Expressed Genes

DDM- Decrease in DNA Methylation

DIRS - *Dictyostelium* Intermediate Repeat Sequence

ds RNA - double-stranded RNA

ETE - Exapted Transposable Element

FAR1 - Far-red impaired response 1

FBS - FHY3/FAR1-binding site

FHL - FHY1-like

FHY3 - Far-red elongated hypocotyl 3

FRS - Far1 Related Sequence

GAL4AD - GAL4 Activation Domain

GAL4BD - GAL4 Binding Domain

GFP - Green Fluorescent Protein

GL2 - GLABRA 2

GM - Genetically modified

GO - Gene Ontology

Gret1 - grapevine retrotransposon 1

hAT - hobo/Ac/Tam

HDP - HARBINGER DERIVED PROTEIN

ITR - Inverted Terminal Repeat

LFY - LEAFY

LINE - Long Interspersed Nuclear Element

LTR - Long Terminal Repeat

MAIL - MAIN-LIKE

MAIN - MAINTENANCE OF MERISTEMS

METHYLTRANSFERASE 1- MET1

MITE - Miniature Inverted Repeat Elements

MUG – MUSTANG

MULEs - *Mutator*-like elements

NGS - Next-Generation Sequencing

NLS - Nuclear Localization Signal

nYFP - N-terminal Yellow Fluorescent Protein

N/C - Nucleus to the Cytosol

PEG - Polyethylene Glycol

phyA - *Phytochrome A*

PGBD3 - PiggyBac Transposable Element Derived 3

piRNAs - PIWI-interacting RNAs

PLE - *Penelope*-like element

PolIVa - RNA polymerase IVa

PRC2 - POLYCOMB REPRESSIVE COMPLEX2

PTGS - Post-Transcriptional Gene Silencing

qRT-PCR - Quantitative RT-PCR

RAG1 - Recombination *Activating 1*

RdDM - RNA-directed DNA methylation

RDR - RNA-dependent RNA polymerase

RISC - RNA-induced silencing complex

RNAi - RNA interference

RNA-seq - RNA-sequencing

RSI - ROSINA

RSL1 - RHD Six-Like 1

RT-PCR - Reverse Transcription PCR

RTE - retrotransposable element

SINE - Short Interspersed Nuclear Element

siRNAs - short interfering RNAs

SPL - SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE

TE - Transposable Element

TF - Transcription Factor

TGS - Transcriptional Gene Silencing

THAP - Thanatos Associated Proteins

TIR - Terminal Inverted Repeat

TSD - Tandem Site target Duplication

V(D)J - Variable, Diversity and Joining

YFP - Yellow Fluorescent Protein

YR - Tyrosine Recombinase

Y2H - Yeast Two-Hybrid

Chapter I

Introduction and Literature Review

Introduction

Transposable elements (TEs) are DNA sequences that could move around in the genome. The first transposable element was found in maize by Barbara McClintock in 1950 (McClintock 1950). TEs can be grouped into two classes based on their transposition mechanism. Class I elements include retrotransposons, and their mobility is mostly via the "copy and paste" mechanism mediated by their RNA; Class II or DNA transposons use "cut-and-paste" mechanism to change their positions. Both classes can be subdivided into autonomous and non-autonomous TEs according to if they carry a transposase-coding region enabling them to transpose by themselves (Joly-Lopez and Bureau 2018).

These jumping sequences had been called "selfish DNA" or "junk DNA" for several years because the previous studies indicated that TEs were only making additional copies of themselves in the genome and they had no obvious benefit for the hosts (Doolittle and Sapienza 1980). In recent decades, the perception of TEs has changed. More evidence showed that the transposon-derived genes can make contributions to the hosts (Cowan, Hoen et al. 2005). In plants, the greatest number of transposon-derived genes are DNA transposons (Volff 2006). The TE-derived genes, which become beneficial to hosts, are named exapted transposable elements (ETEs) (Hoen and Bureau 2015).

In *Arabidopsis thaliana,* two gene families derived from whole transposases were found by forward genetics. One is *FAR1/FHY3* family which origins from *MudrA*, a transposase from *MULE*. FAR1/FHY3 proteins are transcription factors (TFs) involved in the phyA signal transduction (Hudson, Lisch et al. 2003). The other one is the *DAYSLEEPER* gene, derived from

a transposase of the *hAT* superfamily, which is essential for plant development (Bundock and Hooykaas 2005).

MUSTANG is the first ETE family discovered in *Arabidopsis thaliana* by reverse genetics (Cowan, Hoen et al. 2005). *MUG1, MUG2, MUG7* and *MUG8,* the members of *MUSTANG*, were experimentally verified to play fundamentally important roles in development and evolution (Joly-Lopez, Forczek et al. 2012). Later, more ETEs were discovered in Arabidopsis using integrative genomic analysis (Hoen and Bureau 2015) and their T-DNA insertion lines were shown to have abiotic stress-associated phenotypes (Joly-Lopez, Forczek et al. 2017).

The intensity and frequency of some climate and weather extremes have been detected to increase over time (Hoegh-Guldberg, Jacob et al. 2018), concerns about climate-related "collapse" of ecosystems are also growing with it, especially for the marginal, already-stressed agricultural ecosystems (Cheeseman 2016, Canadell and Jackson 2021). Environmental stresses induced by climate change include extreme temperatures, high winds, drought, soil salinity, and flood, which reduce agricultural productivity, decreases economic returns, and add soil erosions (Shrivastava and Kumar 2015). Therefore, it is prudent to respond by developing major improvements in plant breeding programs and agricultural technology.

Genetically modified (GM) technology is applied to overcome the limitations to crop production brought by these adverse conditions for better growth and other traits in need, as GM technology only transfers the specific gene or genes to the receipt plants, which is precise and efficient (Lemaux 2008). TEs make up a large component of plant genomes, and stress conditions drive TE activation. However, as TEs were considered as "Junk DNA", they were excluded from mainstream functional research. ETEs were shown to be widespread in plants, most ETEs are

evolved into TFs, also, they are associated with abiotic stress-related phenotypes. Thus, the ETEs will be a promising reservoir for broadening the diversity of the gene pool selected for crop abiotic stress-associated traits. This review introduces TEs, ETEs, and the application of TEs in crop traits and the potential application of ETEs for crop improvement associated with abiotic stress.

1. Transposable element

1.1. The discovery of transposable elements

In the early decade of the 20th century, genes were considered to be fixed on the chromosome, like beads on a necklace (Morgan 1910). The bead theory was challenged in 1944 when Barbara McClintock worked on the mechanisms of the mosaic colour patterns of maize seed and the unstable inheritance of this mosaicism. She made the surprising discovery that Dissociation causes chromosome break and affects the on and off on its neighbouring genes when the Activator was also present, which included making certain stable mutations unstable. She summarized the data of Ac and Ds in a 1950 PNAS Classic Article, “The origin and behaviour of mutable loci in maize” (McClintock 1950). Although the mobile element concept was accepted relatively soon by maize geneticists, the implications took decades to be widely recognized. In 1963, mobile genetic elements were discovered in bacteriophages (Taylor 1963). Soon, these mobile elements were also found in bacteria (Shapiro 1969), and Drosophila (Engels and Preston 1981). After that, the scientific community gradually accepted that these mobile elements were not just in maize but were widespread across species. These mobile genetic elements are now known as transposable elements (TEs), or transposons.

1.2. The selfish label of transposable elements

Even though the concept of transposable elements was accepted, researchers still consider TEs have subtle effects on genetic diversity until the *P* elements were discovered in Drosophila. In the 1970s, scientists found a special phenomenon named hybrid dysgenesis in Drosophila research. It occurs when certain strains of *D. melanogaster* are crossed with one another, leading to a syndrome of sterility, mutation, chromosome breakage, male recombination, transmission distortion, and nondisjunction (Kidwell, Kidwell et al. 1977). Some years afterwards, this phenomenon was eventually shown to be associated with the *P* element (Bingham, Kidwell et al. 1982). Unlike laboratory-bred females, wild-type females express an inhibitor to P transposase function. This inhibitor reduces the disruption to the genome caused by the movement of *P* elements, allowing fertile progeny. The cross of laboratory females that lack the *P* transposase inhibitor and wild-type males that have *P* elements, leads to lethal to the progeny or rendering them sterile (Bingham, Kidwell et al. 1982). This funding brought TEs back of interest again, not only because they were able to move within the genome, but also because the *P* element was able to spread rapidly to multiple populations of *D. melanogaster* across the world (Silva and Kidwell 2004, Yoshitake, Inomata et al. 2018) and thus TEs play a significant role number of in the fundamental genetic processes. However, the deleterious effects of TEs confirmed the label that TEs are “selfish” and then the selfish DNA was generally accepted for many years (Doolittle and Sapienza 1980, Hickey 1982).

1.3. Types of TEs and main superfamilies

As more and more TEs were found gradually and massively in different species of the eukaryotic kingdom, scientists started to organize them into classifications. All eukaryotic TEs can be classified into two types, Class I RNA transposable element, or retrotransposons and Class II DNA transposable element, or DNA transposons mainly based on their transposition intermediate. Both classes of TEs include two types of TEs based on the insertion mechanism if they carry the transposase-coding gene. One type is autonomous elements, which can sustain and propagate themselves by encoding the genes that produce transposase to promote their own transposition, and the other type is non-autonomous elements, which use the transposition machinery of closely related or even unrelated families to mediate their transposition (Joly-Lopez and Bureau 2018).

After the division of Class I and Class II, orders are also made accoding to the insertion mechanism, the overall organization, and enzymology. Each order is further divided into superfamilies, they share the same replication strategy but are distinguished by large-scale features, such as the structure of a protein, non-coding domains, the short site target duplication (TSD) generated on both flanks upon TE insertion. Superfamilies can be subsequently divided into families, which are defined by DNA sequence conservation.

1.3.1. Class I Retrotransposons

Retrotransposons, also known as Class I RNA transposable elements move from one location to another in the genome by converting RNA back into DNA through the process of reverse transcription using an RNA transposition intermediate. Reverse transcription usually initiates at a

short sequence located immediately downstream of the 5'-LTR and is termed the primer binding site. Specific host tRNAs bind to the primer binding site and act as primers for reverse transcription, which occurs in a complex and multi-step process, ultimately producing a double-stranded cDNA and finally integrated into a new location, creating TSDs and adding a new copy in the host genome. Through the copy and paste mechanism, retrotransposons propagate themselves massively in eukaryotic genomes such as humans (Cordaux and Batzer 2009) and maize (Stitzer, Anderson et al. 2021). Retrotransposons are only present in eukaryotes, however, retroviruses such as HIV were thought to have evolved from retrotransposons by acquiring an envelope protein (Dodonova, Prinz et al. 2019).

There are five orders of retrotransposons: long terminal repeat (LTR), long interspersed nuclear element (LINE), short interspersed nuclear element (SINE), *Dictyostelium* intermediate repeat sequence (DIRS), and *Penelope*-like element (PLE) (Wicker, Sabot et al. 2007).

LTR retrotransposons have direct long terminal repeats that range from 100 bp to 5 kb. *gag* and *pol* are two genes found in all retrotransposons, which are required for their replication. *gag* encodes structural proteins with a capsid and a nucleocapsid domain, that form virus-like particle proteins in the cytoplasm, inside which reverse-transcription occurs. *pol* is located at 3′ of *gag* and encodes three enzymes: a protease, a reverse transcriptase endowed with an RT (reverse-transcriptase) and an RNAse H domains, and an integrase (IN) (Harrison, Tuske et al. 2021). The *gag* and *pol* genes are transcribed in the same mRNA, however, there are three ways to express them. One way is generating a single *gag-pol* ORF. Another is generating a premature stop codon or frameshift between *gag* and *pol*. The other is generating three or more ORFs that separate *gag* and *pol* (Gao, Havecker et al. 2003). LTR retrotransposons consist of five superfamilies: *copia*, gypsy, *BEL-Pao*, *Retroviruses* and endogenous retroviruses (Wicker,

Sabot et al. 2007). The major structural difference between *copia* and *gypsy* superfamilies is the order of the reverse transcriptase (RT) and integrase domains in their *pol* genes (Friesen, Brandes et al. 2001). *copia* and *gypsy* have a high copy number (up to a few million copies per haploid nucleus) in the genome of plants, animals, fungi, and protista. BEL-Pao superfamily contains bona fide LTRs, encode GAG and POL proteins, and create a 4–6 bp TSD upon insertion. It distributes specifically to animals (Filée, Farhat et al. 2021). *Retroviruses* are evolutionarily close to Gypsy LTR retrotransposons but adopted a viral lifestyle through envelope protein (ENV). They are restricted to vertebrates, with some exceptions in *Drosophila melanogaster* (Wicker, Sabot et al. 2007). *Endogenous retroviruses* (*ERV*) is a vertically inherited proviral sequence that through inactivation or deletion of the domains that enable extracellular mobility in retrovirus, which transit through the germline for propagation (Nelson, Hooley et al. 2004).

LINEs are retrotransposons that lack LTRs but can reach several kilobases in length and are widespread in the genome of many eukaryotes. They consist of five major superfamilies: *R2*, *RTE*, *Jockey*, *L1*, and *I*. *R2* elements distribute exclusively in animals. They encode a sequence-specific endonuclease that is responsible for its insertion at a unique site in the 28S rRNA genes and also encode a reverse transcriptase (RT) that uses the cleaved DNA to prime reverse transcription of the R2 transcript (Yang, Malik et al. 1999, Eickbush and Eickbush 2015). *RTE* (*retrotransposable element*) superfamily was first identified in *Caenorhabditis elegans* and later was discovered in vertebrates and arthropods (Malik and Eickbush 1998). The *RTE* elements contain a domain with homology to the apurinic-apyrimidinic (AP) endonucleases in addition to the RT domain. They have extensive truncations, giving rise to SINE elements (Wicker, Sabot et al. 2007). *Jockey* is a superfamily found only in Arthropoda. The element is ~ 5 kb in size and consist of two ORFs, the first is 568 aa residues long and quite diverse; the second is 916 aa

residues long, encodes an apurinic endonuclease and a reverse transcriptase (RT), is well conserved among the elements (Tambones, Haudry et al. 2019). *LINE-1* (*L1*) superfamily appears to be the most active autonomous retrotransposons in most mammals, which is the only retrotransposon that can retrotranspose in somatic cells *in vivo* (Faulkner and Billon 2018). The members of Superfamily *I* are identified by if they contain RNaseH (Wicker, Sabot et al. 2007).

Short interspersed nuclear elements (SINEs) are non-autonomous retrotransposons that amplify themselves throughout eukaryotic genomes. They range from ~ 100 to 700 bp in size. Unlike other retrotransposons that contain RT domains, SINEs don not encode for any protein product. Thus, SINEs have to exploit LINE transposition components to help them reverse-transcribed and inserted back into a new location in the genome, even though the LINE-binding proteins prefer binding to LINE RNA. SINEs usually consist of parts derived from tRNA and LINEs. The tRNA portion contains an RNA polymerase III promoter to make sure the LINE portions of SINEs can be transcribed into RNA. The LINE component of SINEs can be recognized by the LINE-binding proteins to be used for transposition (Kramerov and Vassetzky 2011). SINEs can be classified into three superfamilies: *tRNA*, *7SL* and *5S*. Monomeric *tRNA* SINE superfamilies are present in almost all the eukaryotic genomes and by far the most frequent SINEs. These elements consist of a 5′ tRNA-related region, a central region, and a stretch of homopolymeric adenosine residues. *tRNA* SINEs can also be dimeric such as *Twin* SINEs and *RathE2* SINEs from *Arabidopsis thaliana (Pélissier, Bousquet-Antonelli et al. 2004)*. SINEs with *7SL* RNA-derived heads belong to the *7SL* superfamily. *Alu* is the element of the *7SL* superfamily, which propagates in the human genome successfully, having a copy number well over 1 million copies, contributing almost 11% of the human genome (Deininger 2011). Elements from the *5S* superfamily are derived from 5S rRNA. They are transcribed from the type 1 internal pol III

promoter. The 3′-end of the elements is significantly similar to that of *CR1-like* non-LTR retrotransposons, which are also not flanked by TSD, and their 3′ ends are composed of $(ACATT)_n$ and $(ATT)_n$ (Kapitonov and Jurka 2003).

DIRS retrotransposons are tyrosine recombinase (YR) retrotransposons that can be detected in diverse species, ranging from green algae to animals and fungi (Wicker, Sabot et al. 2007). They contain an internal complementary region that is required to reconstitute a full-length element from incomplete RNA transcripts and is essential to complete retrotransposition (Malicki, Spaller et al. 2020). This order includes three superfamilies: *DIRS*, *Ngaro*, and *VIPER*.

The *PLE* superfamily was first found in *Drosophila virilis*, later was found in diverse eukaryotes, including amoebae, fungi, cnidarians, rotifers, flatworms, roundworms, fish, amphibia, and reptilia. *PLE* encodes an RT and an endonuclease. Members have LTR-like flanking DNA that can be in a direct or an inverse orientation (Evgen'ev and Arkhipova 2005).

1.3.2. Class II DNA transposons

Class II DNA transposons are transposable elements that move via a DNA intermediate, in contrast to class I TEs, retrotransposons, that move via an RNA intermediate. The DNA intermediate could be either single or double-stranded. Compared to the retrotransposons that are often found in eukaryotes, DNA transposons have been found in both prokaryotes and eukaryotes, and make up a significant portion of eukaryotic genomes (Babakhani and Oloomi 2018). In prokaryotes, DNA transposons usually carry additional genes for antibiotic resistance. Transposons can transfer from a plasmid to other plasmids or from a DNA chromosome to a

plasmid and *vice versa* that causing the horizontal transfer of antibiotic resistance genes (Babakhani and Oloomi 2018).

There are three main mechanisms for the movement of DNA transposons: cut and paste for most DNA transposons with terminal inverted repeats (TIR), rolling circle for *Helitrons*, and self-synthesizing for *Polintons*. Based on the move mechanisms, DNA can be further divided into superfamilies. The "cut and paste" transposons are ordered by the size of their terminal inverted repeats (TIRs) and the TSDs. These TEs increase their copy numbers by transposing during chromosome replication from a position that has already been replicated to another that the replication fork has not yet reached. Transposition of these elements requires a transposase enzyme that recognizes the TIRs and cuts both strands at each end of the transposon, and also three DNA sites: two TIRs and one target site. The transposase will bind to the TIR sites and mediate synapsis of the transposon ends, cut the element from flanking DNA, carry the element and insert it into the target site. The addition of the new DNA into the target site causes short gaps on both side of the inserted segment and repaired by host gap repair machinery resulting in the target sequence duplication (TSD) that are characteristic of transposition (Craig 1995). In total nine superfamilies are classified: *Tc1–Mariner*, *hAT, Mutator*, *Merlin*, *Transib*, *P*, *PiggyBac*, *PIF-Harbinger*, *CACTA*. *Tc1–Mariner* is named after its two best-studied members, the *Tc1* transposon of *C. elegans* and the *mariner* transposon of *Drosophila.* This superfamily is found in all animals, protists and bacteria. Recently, a new family of plant-genome restricted *Tc1/mariner* elements were found, which were named *PlantMar* (Dupeyron, Baril et al. 2020). The name of *hAT* superfamily derives from three of its members: the *hobo* element from *Drosophila melanogaster,* the *Ac* element from *Zea mays*, and the *Tam3* element from *Antirrhinum majus*. *hAT* elements are widely distributed across eukaryotic genomes, usually

having a size of 2.5-5 kb with short TIRs and short TSDs generated during transposition (Kempken and Windhofer 2001). *Mutator* elements have a high transposition rate and tend to insert near or close to genes, in consequence, they are the most mutagenic transposons. Members of the *Mutator* DNA transposon superfamily are termed 'MULEs' (*Mutator*-like elements). The structure of MULEs resembles classic "cut-and-paste" DNA TEs, with TIRs at each end enclosing a transposase domain containing a catalytic DDE motif, and often an additional zinc finger DNA-binding motif. MULE TIRs are typically considerably longer than those in other DNA TE superfamilies, commonly being hundreds of base pairs in length. Meanwhile, some families of MULE appear to lack TIRs completely (non-TIR MULEs) (Yu, Wright et al. 2000). *Merlin* transposons were identified in diverse animal and eubacterial genomes, they display a motif similar to the DDE motif found in many transposases and integrases, generating an 8-bp or 9-bp TSD upon insertion. They are closely related to the IS*1016* group of bacterial insertion sequences (Feschotte 2004). *Transib* was identified by *in silico* reconstruction of ancient fossil TE sequences in the *Drosophila melanogaster* and *Anopheles gambiae* genomes. These transposons are 3 ~ 4 kb long and contain 9 ~ 60 bp TIRs flanked by 5-bp GC-rich TSDs (Chen and Li 2008). As mentioned earlier *P* elements were discovered in *Drosophila* as the causative agents of hybrid dysgenesis and also found in other eukaryotes. All *P* elements have a conserved structure containing 31 bp TIR and 11 bp internal inverted repeats located at the (Thanatos Associated Proteins) THAP domain of the transposase (Majumdar and Rio 2015). *PiggyBac* elements were originally isolated from the cabbage looper moth in the 1980s, to date, they are shown to have a broad host spectrum from yeast to mammals. The *PiggyBac* transposase recognizes transposon-specific inverted terminal repeat sequences (ITRs) located on both ends of the transposon and specifically integrates the transposon into TTAA target sites (Yusa 2015).

PIF-Harbinger elements were shown to be widespread in the plant kingdom and composed of two subgroups, *PIF* and *Pong*. Elements representing both subgroups were related to certain miniature inverted repeat elements (MITEs), like *Tourist* in maize (Bureau and Wessler 1992) and *mPING* in rice (Yang, Zhang et al. 2007). *CACTA* elements are named of their highly conserved CACTA TIR motif. A full-length *CACTA* element encodes a transposase and a protein of unknown function. During transposition, it creates a 3-bp TSD. *CACTAs* are found in genomes across algae to vascular plants to animals (Buchmann, Löytynoja et al. 2014).

Crypton elements represent a unique class of DNA transposons using YR to cut and rejoin the recombining DNA molecules. *Cryptons* were originally identified in fungi and later in the sea anemone, sea urchin and insects. They don not have TIRs or long direct repeats. Instead, they have short direct repeats (4 ~ 6 bp) at both ends of the transposon which are considered as substrates for recombination (Kojima and Jurka 2011).

Helitron elements are the eukaryotic rolling-circle transposable elements. They are hypothesized to transpose via a single-stranded DNA intermediate rolling circle replication mechanism and integrate between an AT host dinucleotide. Canonical *Helitrons* typically begin with a 5′ T (C/T) and terminate with CTR. The 3′ end frequently has a short (~ 11 bp) palindromic sequence that forms a hairpin structure. Autonomous *Helitrons* encodes a Y2-type YR like the IS*91* rolling-circle transposons, with a helicase domain and replication initiator activity. During the transposition process, *helitrons* frequently capture and disperse host gene fragments, producing chimeric transcripts (Borgognone, Castanera et al. 2017).

Polintons, also called *Mavericks*, are named after two proteins, a DNA polymerase and a retroviral-like integrase. They are large DNA transposons, around 15 ~ 20 kb in size, examples

have been reported up to 40 kb. *Polintons* encode up to 10 proteins, the key elements being the protein-primed type B DNA polymerase and the retroviral-like integrase. Most *polintons* also encode an adenoviral-like cysteine protease, an FtsK-like ATPase, and viral jelly roll capsids proteins (Krupovic, Bamford et al. 2014).

1.4. TE silencing and activation

1.4.1. TE silencing

To date, transposable elements have been identified in almost all eukaryotic species that have been investigated. TEs make up a significant component of the genomes of both plants and animals. Owing to the high copy numbers and the ability to move around the genome, the transposition of transposons has a profound impact on the host genome. For example, the insertion of a transposon in or near a gene may disrupt the coding sequences of the gene or perturb its expression pattern; the failure to repair DNA breaks associated with transposon excision or insertion may result in chromosome rearrangements; providing sites for non-homologous recombination during meiosis (Klein and O'Neill 2018). Gene expression alteration, gene deletion and insertion, and chromosome rearrangements, most effects from TE activation are deleterious (Hollister and Gaut 2009). Not surprisingly, the host develops several defence mechanisms to suppress the activation of transposons, such as a variety of small RNA, chromatin, DNA modification pathways (Sigman and Slotkin 2016) and sequence-specific repressors such as KRAB zinc-finger proteins (Bourque, Burns et al. 2018). These strategies can be briefly divided into two mechanisms: transcriptional gene silencing (TGS) and post-transcriptional gene

silencing (PTGS). In this review, I will summarize recent trends in understanding TE silencing in eukaryotes, with the main emphasis on plants.

Two decades ago, scientists found in Arabidopsis, TEs are enriched in the highly methylated centromeric region and packed into heterochromatin. They are also concentrated in other heterochromatin regions such as the knob, which are highly methylated. The same observation was also found in maize. Those reports suggest TEs are the primary targets of DNA methylation (Okamoto and Hirochika 2001). Nowadays, it is known silencing TEs at the transcriptional level is via RNA-directed DNA methylation (RdDM). Lisch proposed a "host nuclear immune response system" hypothesis (Lisch 2009). In which, transposons produce aberrant RNAs, which are analogous to antigens. The host has a surveillance system, once detects these aberrant RNAs, the host stimulates a nuclear immune response. The initial trigger for TGS can be tandem duplications or inverted duplications introduced by chromosome rearrangements of transposon sequences, such as *Mu killer*. RNA polymerase IVa (PolIVa) transcripts the aberrant RNAs from transposons. The RNAs are then transported to the Cajal body and duplexed into double-stranded RNA (ds RNA) by the RNA-dependent RNA polymerase (RDR) 2 and subsequently sliced to 24 nucleotides (nt) small interfering (si) RNAs by the DICER-like (DCL) 3. These 24 nt-siRNAs are bound by the ARGONAUTE (AGO) 4 proteins and form into an RNA-induced silencing complex (RISC) and interact with the nascent RNA transcribed by the RNA PolVb. AGO4 then recruits multiple proteins including SU(VAR)3-9 HOMOLOG (SUVH) 4/5/6 and DOMAINS REARRANGED METHYLASE (DRM) 1/2 that mediate repressive histone modification (H3K9me2) and DNA methylation, respectively, thus in turn to promote the production of more aberrant RNA (Cho 2018).

In contrast to TGS, post-transcriptional regulation of TEs is through RNA interference (RNAi) mechanism. A common pathway for post-transcriptionally silencing (PTGS) is mediated by *DCL1*-dependent microRNAs (miRNAs), which trigger 21-nucleotide secondary short interfering RNAs (siRNAs) via RNA-DEPENDENT RNA POLYMERASE 6 (RDR6), DCL4 and ARGONAUTE 1 (Creasey, Zhai et al. 2014). TEs escaped from TGS or newly introduced TEs transcribed into RNAs, which undergo frequent ribosome stalling caused by their unfavourable codon usage. These TEs also give rise to 21- or 22-nt siRNAs, triggering RNAi. Ribosome stalling induces TE transcripts truncation and localization to cytoplasmic siRNA bodies, both of which are essential prerequisites for RDR6 targeting (Kim, Wang et al. 2021). siRNA is then loaded onto a RISC, which includes a member of the Argonaute (AGO) family. The AGO-associated siRNA is then guided to a target TE transcript, which is cleaved and degraded by the RISC complex.

In animals, there is a special pathway for germline DNA methylation called PIWI-interacting RNAs (piRNAs) pathway. The 23–30nt germline-specific piRNAs are loaded into PIWI clade Argonaute proteins and guide sequence-specific TGS and PTGS (Zhang, Yu et al. 2021). In *C. elegans*, *Tc1* elements give rise to siRNAs that degrades the target TE RNAs through PTGS. In addition, siRNAs also interact with piRNAs to degrade TEs (Siomi, Sato et al. 2011), explaining observed *Tc1* activity in *C. elegans* somatic cells, but not in the germ cells. Intriguingly, piRNAs and TE-associated siRNAs also target non-TE mRNAs and long non-coding RNAs (lncRNAs), exerting certain regulatory roles in various biological processes (Ali, Peffers et al. 2021). Furthermore, the Krüppel-associated box zinc finger protein (KRAB-ZFP) family in mammalian genomes, plays an important role in the recognition and transcriptional silencing of TEs, including the active ETn, IAP, and L1 families (Wolf, de Iaco et al. 2020).

Besides the above-mentioned mechanisms, alterations of chromatin packaging and condensation are also a strategy for TE silencing. For example, in *Arabidopsis thaliana*, the chromatin-remodelling protein Decrease in DNA Methylation (DDM1) is essential for the silencing of TEs and the condensation of chromatin (Zemach, Kim et al. 2013). A recent study showed dNxf2 is associated with the Panoramix–dNxf2-dependent TAP/p15 silencing (Pandas) complex that counteracts the canonical RNA exporting machinery and restricts transposons to the nuclear peripheries to drive heterochromatin formation (Zhao, Cheng et al. 2019).

1.4.2. TE activation

Even under the suppression of diverse TE silencing mechanisms in the host, TEs are evolutionarily able to circumvent silencing and survive (Fultz, Choudury et al. 2015). A major question in the field is how can a certain type of transposable element remain active in the same genome with others that are silenced? Multiple mechanisms of TE evasion of silencing have been proposed. One reason could be that at a certain condition, these silencing restrictions are lifted that allow the conventional genes to be expressed in the host. For example, to reset imprinted genes in primordial germ cells, the host needs to reduce the genome-wide of DNA methylation, which provides TEs with a chance to promote their expression (Bourque, Burns et al. 2018). In ciliates, they silence transposons by way of deleting them from their somatic genome using a series of intricate sequences targeting pathways involving siRNAs and transposases, part of which consists of a transnuclear comparison between maternal soma macronucleus and daughter germline micronucleus (Allen and Nowacki 2017). Another example is the alternative splicing of transposon *P*- elements and *Gypsy* in Drosophila germline and soma (Teixeira, Okuniewska et al. 2017). The piRNA pathway introduces TGS and PTGS by

regulating precursor mRNA splicing of transcripts from *P*-transposable elements and *Gypsy in vivo*, leading to the production of the non-transposase-encoding mature mRNA isoform in Drosophila germ cells and cultured ovarian somatic cells.

There is other evidence supporting the impaired DNA methylation increases the activity of transposons. For example, under laboratory conditions, two Arabidopsis lines of a population from a cross between a wild-type and a *met1* homozygous plant carrying a null mutation for *MET1 (METHYLTRANSFERASE 1)* showed aberrant development phenotype due to the movement of a *copia*-type retrotransposon (*Évadé*), which was independently activated and inserted into three unlinked loci: *LEAFY* (*LFY*) on chromosome 5, *BRI1* on chromosome 4, and *VAR2* on chromosome 2 (Mirouze, Reinders et al. 2009); double mutation of CHROMOMETHYLASE 3 (CMT3) and MET1 triggers a high-frequency transposition of *CACTA elements* in Arabidopsis and *Tos17* and *Tos19* in rice (Kato, Miura et al. 2003). Mutations in Arabidopsis chromatin remodelling factor DDM1 also leads to the activation of *MULEs* (Singer, Yordan et al. 2001) and *CAC1*, an endogenous *CACTA* family transposon (Miura, Yonebayashi et al. 2001). The mutation in chromatin remodelling as well as various components of the RdDM pathway reactive the retrotransposons *ATGP3*, *COPIA13*, *COPIA21*, *VANDAL21*, *EVADÉ* and *DODGER* (Benoit, Drost et al. 2019).

From the perspective of TE, an ideal scenario is to be expressed and active in the germline to propagate, but not in the soma, where expression and transposition would gain TEs no benefits but silence, however, organisms such as plants, do not differentiate germ lineage cells early in development. Besides the loosen up of epigenetic silencing from the host, some plant transposons have a special way to be transcriptionally activated, which is biotic and abiotic

stresses (Benoit, Drost et al. 2019). For example, in Arabidopsis wild-type cells with functional TGS and PTGS silencing mechanisms, the *ONSEN* retrotransposons are able to remain active under heat stress. They have acquired a heat-responsive element recognized by plant-derived heat stress defence factor, heat shock factor A2, resulting in transcription and synthesis of extrachromosomal DNA copies under elevated temperatures (Ito, Gaubert et al. 2011). However, the retrotransposition of *ONSEN* was stimulated only in mutants impaired in the biogenesis of siRNAs, suggesting the combination of loosening or impaired siRNA pathway and environmental stress is a prerequisite for *ONSEN* retrotransposition (Ito, Gaubert et al. 2011). Similar activation under heat stress was also found in rice, in which *Go-on*, a *copia* family LTR retrotransposon is activated (Cho, Benoit et al. 2019). Under biotic stress, *Bs1*, a transposable element is shown to move into the maize *Adhi* gene following barley stripe mosaic virus infection (Johns, Mottinger et al. 1985). Another report is that the expression of *Tnt1* was highly induced in protoplasts isolated from tobacco leaf tissue, activated by a series of stress such as protoplast isolation, wounding and plasmolysis (Pouteau, Huttner et al. 1991) and inserted into *nitrate reductase (nia)* gene (Grandbastien 1998). In Sicilian blood orange, *Tcs1* retroelement is inserted to the adjacent loci of *Ruby*, an MYB transcriptional activator gene of anthocyanin production. *Tcs1 is activated by* cold stress and therefore regulates the expression of *Ruby* (Butelli, Licciardello et al. 2012). In tomatoes, the accumulation of the transcription of LTR-type retrotransposon *Rider* and its transposition is triggered by drought stress and relies on abscisic acid (ABA) signalling (Benoit, Drost et al. 2019). Also, another LTR retrotransposon, *MESSI* from Solanum, can be transcribed specifically in the shoot apical meristems of tomato. The author believes *MESSI* escaping from TGS is induced by developmental signals rather than stress (Sanchez, Gaubert et al. 2019).

Most of the above-mentioned mobile elements belong to the *copia* TE family. It is consistent with a recent report to characterize the *Arabidopsis thaliana* mobilome (Quadrana, Silveira et al. 2016). In its analysis, the author shows the composition and activity of the 'mobilome' vary extensively between accessions with climate and genetic factors. While the loci controlling adaptive responses to the environment are the most frequent transposition targets. For example, *ATCOPIA2* and *ATCOPIA78* share the highest number of geo-climatic variables correlated with copy numbers. In the same time, *ATCOPIA78* is shown to have the strongest correlation between temperature annual range and copy number.

Although many mechanisms are there for TE to escape from circumventing silencing, among them, the one that has gained the most attention is TE integration into an active gene to prevent the TE from silencing. In a systematic survey of active *Mutator* TEs in *A. thaliana*, the widespread natural *AtMu1c* transposition was characterized from the analysis of over 200 accessions. *AtMu1c* has the highest expression for inserting within 3′ untranslated regions of protein-coding genes (Kabelitz, Kappel et al. 2014), suggesting insertion sites in genic 3′ untranslated regions are safe havens to avoid silencing. In the above-mentioned example, the position of the *MULE* element and proximity to a gene is critical for the evasion of silencing of the TE. In maize, individual *Mutator* TE copies are also able to transposition despite siRNAs which generated from silenced and fragmented homologous copies exist (Fultz, Choudury et al. 2015). Because different TE families have different insertion site preferences, the ability to escape from silencing may be higher for some DNA TEs (such as *Mutator*) that target genic regions for insertion.

2. Exapted transposable element

The survival TEs that escaped from the host silencing mechanism could be harmful, neutral, or beneficial towards the host genome. The harmful effects have been documented extensively, as mentioned earlier, they integrate into regulatory or coding regions of host genes or induce chromosome recombination. As a result, deleterious TE insertions are removed from population by purifying selection. Sometimes TE insertions are effectively neutral, these TE insertions that do not serve host function, the rate and pattern of sequence evolution of which generally follows that of unconstrained, neutrally evolving DNA, leading to the accumulation of disabled and nonreplicative TE 'vestigiofossil' throughout genomes. Thus, the genome information shows the long history of TE activity during evolution, reflecting a balance between the TEs amplification strategy and the host defence against them. However, occasionally, survival TEs could be beneficial to the host. From another evolutionary perspective, the activation of TEs finely tuned regulatory activities that favoured their propagation and have predisposed them to be co-opted for the regulation of host genes. As mentioned earlier, some TEs become active under stress conditions, and this activation is often taken as evidence for an adaptive role of TEs in the stress response. Stress conditions and other environmental stimuli drive the activity of certain TE families and consequently new TE insertions, which can give rise to altered gene expression patterns and phenotypes. Such connections between TE-mediated increases in diversity and an accelerated rate of genome evolution provide powerful mechanisms for plants to adapt to new environmental conditions more rapidly. The theory behind this is that activation of TEs would lead to an increase in the mutation rate which generates variability and natural selection will eventually choose certain mutations that are good for the fitness of the organism. Additionally,

because some TEs are known to obtain stress-response regulatory sequences, if TEs are activated by stress, they have the capacity to distribute the stress-response elements throughout the genome to help the host genome develop novel stress-responding networks.

Furthermore, the survival of TEs in the host genome provides extra DNA to be repurposed, (or co-opted, domesticated, exapted) for the host genome to serve cellular functions that become beneficial to the host organism. These TE coding sequences that originally encode transposases for the mobilization of TE then are exapted to have novel roles as conventional host genes, are called exapted transposable elements (ETEs) (Joly-Lopez and Bureau 2018). In recent years, a growing number of ETEs in various organisms across bacterial, plant and animal were identified by next-generation sequencing (NGS) approaches. Here we address a few well-studied ETE or ETE families, with the main focus on plants and their roles in host genomes.

2.1. *ROSINA* in *Antirrhinum majus*

ROSINA (RSI) is a part of *TamRSI*, a *CACTA* transposable element in *Antirrhinum majus* (Roccaro, Li et al. 2007). It was isolated as a DNA binding factor that was able to bind to the CArG-box in the promoter of MADS-box gene *DEFICIENS* in *Antirrhinum majus*. The exapted function of *RSI* is modulating petal and stamen development. Sequence analysis shows that the sequence orientation of *RSI* is opposite to the putative transposase of *TamRSI*. Interestingly, *RSI* is still evolving within the TIRs of this CACTA transposon, suggesting it is a newly ETE (Hoen and Bureau 2015).

2.2. The *SLEEPER* family

SLEEPER, short for DAYSLEEPER, this ETE family is derived from a transposase of hAT superfamily. *DAYSLEEPER* was first discovered fortuitously in a yeast one-hybrid screen aimed to identify Arabidopsis proteins bind to the Kubox1 motif in the promoter of *Ku70*, a DNA repair gene (Bundock and Hooykaas 2005). *DAYSLEEPER* encodes a protein containing a C2H2 type BED-zinc finger domain and an hATC dimerization domain. However, the protein lacks the catalytical DDE motif which is typical for hAT transposases. *DAYSLEEPER* has also lost the TIR sequences that are essential for transposition and flanking genomic TSDs. The *daysleeper* knockout mutants in Arabidopsis showed dramatic developmental defects, such as failing to develop normal leaves or floral organs. While the plants overexpressing DAYSLEEPER can grow but still showed slow growth and aberrant flower. The transcriptional expression profiling of the DAYSLEEPER-overexpressing plants revealed a set of upregulated genes, although none of them were found to have a Kubox1 motif in their promoter. Microscopy observation showed DAYSLEEPER is localized into the nucleus, which is consistent with the molecular function of the protein. There is a divergent homolog of *DAYSLEEPER* in *A. thaliana* which lacks the K/R rich conserved array adjacent to the BED-zinc finger, named CYTOSLEEPER as the protein is cytoplasmic. Unlike *DAYSLEEPER* knock-out lines, cytosleeper mutants are indistinguishable from wild-type plants (Knip, Hiemstra et al. 2013). *DAYSLEEPER*-like genes were found distributed across angiosperms, such as Basal angiosperms as well as grasses (Poaceae) and dicotyledonous plants, and most species have not only one copy. However, no *SLEEPER* genes were found in gymnosperms, mosses and algae. In *Oryza sativa*, homozygous *RICESLEEPER*1 and 2 T-DNA insertion mutants are lethal. In addition, the heterozygous line can grow but

showed abnormalities, such as empty panicles, yielding only very few seeds (Knip, de Pater et al. 2012). These results suggest *RICESLEEPER*1 and 2 are also essential for normal development in rice. Altogether, these studies demonstrated the pivotal role of DAYSLEEPER in higher plants.

2.3. *FAR1-related sequence* (*FRS*) family

In plants, the first discovered ETE genes are *FAR-RED IMPAIRED RESPONSE1 (FAR1)* and *FAR-RED ELONGATED HYPOCOTYLS3 (FHY3)*. *FAR1* and *FHY3* are homologs that were initially identified in *A. thaliana*. They are the founding members of the *FAR1*-related sequence (*FRS*) ETE family. *FRS* sequences and domain architecture display significant similarity to Mutator-like element (MULE) transposases such as Jittery and MURA of maize (Wang and Wang 2015), suggesting the close evolutionary relationship that the *FRS* family is expated from the MULE TE superfamily. *FRS* sequences are conserved in both eudicots and monocots, suggesting that exaptation of this ETE family occurred before the monocot–dicot split. Comparative genomic analyses showed the various expansion pattern of *FRS* genes in different species. Taking advantage of the high-quality chromosome-level genome assembly in rose, the variety *Rosa wichuraiana* 'Basye's Thornless' was discovered to contain 91 *RwFAR1/FRS-like* genes. In the meanwhile, in *Rosa chinensis* 'Old Blush', only 52 *RwFAR1/FRS-like* genes were found. The significant expansion of the *FRS* gene family in Basye's Thornless correlates well with the switch to prostrate-to-erect growth of shoots upon flowering in rose, with potential links to molecular processes related to light signalling, shade perception and flowering time (Zhong, Jiang et al. 2021). In comparison to two related species, *Medicago truncatula* and *Medicago sativa*, a large expansion of *FRS* family, in total 291 gene members were also discovered in the wild species *Medicago ruthenica*. *FRS* genes make the largest TF family in *M.*

ruthenica and regulate a set of genes that are involved in responding to abiotic stress (Wang, Ren et al. 2021). Therefore, the expansion of *FRS* genes may contribute to the higher tolerance of *M. ruthenica* to various abiotic stresses, making it a valuable model to study molecular mechanisms underlying tolerance to environmental stress in legume forages and also an alternative as a legume forage crop. The same regulatory role of *FRS* family is also found in the tea plant *Camellia sinensis*. The transcriptome profile of *Camellia sinensis* showed that *FRS* genes respond to salt stress and shading treatment. Quantitative RT-PCR (qRT-PCR) result suggests *FRS* genes in *Camellia* can be regulated by biotic and abiotic stresses, such as high salt, high/low temperature, polyethylene glycol (PEG)-mimicked drought and ABA treatment (Liu, An et al. 2021). These reports are consistent with a study performed in *A. thaliana* showing that mutant plants knocked out for genes belonging to the *FAR1/FHY3* family display significant phenotypical traits when challenged by several abiotic stresses (Joly-Lopez, Forczek et al. 2017).

In model organism *A. thaliana*, *FRS* genes were most thoroughly characterized. FAR1 and FHY3 act cooperatively as transcription activators modulating far-red light-responsive gene expression downstream of the photoreceptor phytochrome A (Hudson, Ringli et al. 1999). The DNA-binding domain of FAR1 and FHY3 recognize specific motifs called FHY3/FAR1-binding site (FBS) that are predominantly located at the transcription start site of promoters. FBS is enriched in the promoters of numerous genes exhibiting diurnal or circadian cycling, for example, *CIRCADIAN CLOCK ASSOCIATED1 (CCA1)*, which is a central gene controlling the circadian clock. Besides the DNA-binding function, FAR1 and FHY3 can form heterodimers and activate multiple genes expression for various biological functions and cellular processes such as chlorophyll biosynthesis, circadian rhythm, shade tolerance, seed germination, flowering, plant immunity and stress responses (Liu, Ma et al. 2020). Interestingly, FHY3 and FAR1 also work as

transcriptional repressors. FHY3/FAR1 complex physically interacts with three SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) TFs, SPL3, SPL4, and SPL5. This complex negatively regulates the expression of a set of genes in which SPL binds to the promoter regions (Xie, Zhou et al. 2020). These regulated genes are involved in flowering regulation, including *FRUITFUL (FUL), LFY* and *APETALA1 (AP1)*. Thus, FHY3 and FAR1 integrate light signals with the miR156-SPL module-mediated ageing pathway to regulate Arabidopsis flowering. Altogether, *FRS* ETE family can be depicted as multifaceted molecular players involved in plant development and response to various environmental stimuli (Wang and Wang 2015).

2.4. *MAINTENANCE OF MERISTEMS* (*MAIN*) and *MAIN-LIKE* (*MAIL*) genes

MAIN was first identified as a cellular factor required for shoot apical meristem and root apical meristem organization maintenance by sustaining genome integrity in stem cells and their descendants cells. Arabidopsis *main* mutants showed short roots, misshaped leaves, reduced fertility and partial fasciation of stems, also the mutant plants exhibit increased DNA damage and are hypersensitive to DNA-damaging agents (Wenig, Meyer et al. 2013). Later, three closely related genes were found in the *A. thaliana* genome, named *MAIN*-related genes *MAIN-LIKE1* (*MAIL1,* At2g25010)*, MAIN-LIKE2* (*MAIL2,* At2g04865) and *MAIN-LIKE3* (*MAIL3,* At1g48120). Arabidopsis *MAIL* genes encode nuclear proteins and show expression in all tissues. Based on their protein sequence similarity, MAIN and MAIL form a sub-family of four proteins containing a conserved aminotransferase-like domain which is similar to a DNA binding motif found in transposases. MAIL1 and its closely related proteins are considered to be encoded by ETEs (Ühlken, Horvath et al. 2014). Later, it was shown that *MAIN* and *MAIL1* defined an

alternative silencing pathway independent of DNA methylation and short interfering RNAs. Mutants for *MAIL1* or *MAIN* exhibit release of silencing and appear to show impaired condensation of pericentromeric heterochromatin (Ikeda, Pélissier et al. 2017). Furthermore, MAIN and MAIL1 physically interact together, as well as with a putative serine/threonine phosphoprotein phosphatase (PPP) called PP7-LIKE (PP7L). Loss-of-function mutants for PP7L phenocopy *main* mutants, *mail1* single mutants and *mail1 pp7l* double mutants, such as photosynthetic defects and strong developmental phenotype associated with misregulation of several genes. (de Luxán-Hernández, Lohmann et al. 2020, Nicolau, Picault et al. 2020).

More recently, the genetic interaction between the DRM2, CMT3 and MAIN is deciphered. RNA sequencing (RNA-seq) experiments show upregulation of several TEs in the DNA methylation-deficient *drm1 drm2 cmt3 (ddc)* triple mutant; a milder degree of TE activation can be found in the *main-3* single mutant, with the significant enrichment of pericentromeric TEs among the upregulates TEs. The *ddc main-3* mutant showed an exacerbation of TE silencing defects, with a large number of pericentromeric TEs being specifically upregulated in this mutant background. These results suggest that DRM2, CMT3 and MAIN are part of different epigenetic pathways that act redundantly or synergistically to repress TEs (Nicolau, Picault et al. 2020).

2.5. The *Harbinger* transposase-like ETEs

2.5.1. ANTAGONIST OF LIKE HETEROCHROMATIN PROTEIN (ALP)

ALP1 protein was first identified in a large forward genetic screen for suppressors and enhancers of the Arabidopsis Polycomb group member, *LIKE HETEROCHROMATIN PROTEIN1* mutant *lhp1*(Hartwig, James et al. 2012). *ALP1* encodes a protein similar to the transposase of

Harbinger, a transposon that belongs to the *PIF* superfamily. In Arabidopsis, in total eight *Harbinger*-like genes. However, phylogenetic analysis shows the *ALP1* clade is clustered with four other plant proteins of unknown function from soybean, poplar, grapevine, and castor bean, whereas the closest Arabidopsis homolog, At3g55350, was in a distinct branch (Hartwig, James et al. 2012). Proteomic studies showed ALP1 is associated with the Polycomb group protein complex POLYCOMB REPRESSIVE COMPLEX2 (PRC2), which is an ancient H3K27me3 histone methyltransferase that was first identified genetically in Drosophila through its role in repression of developmental target genes (Velanis, Perera et al. 2020). In the same screen, ANTAGONIST OF LIKE HETEROCHROMATIN PROTEIN 2 (ALP2) was identified as its mutant was similar to *alp1* that gave a partial suppression of the narrow, curled leaf and early flowering phenotype of the *lhp1* mutant. The alignment of plant ALP2 proteins with selected *Harbinger* DNA binding proteins and the exapted *Harbinger* protein HDP2 from *A. thaliana* showed sequence similarity, suggesting ALP2 is an exapted transposon that highly diverged from *Harbinger* transposase. BLAST searches revealed that ALP2 is well conserved across angiosperms. Phylogenetic analysis confirmed ALP2 sequences form a single, well-supported group distinct from those of HDP2 proteins or *Harbinger* transposase DNA binding domains, consistent with an ancient origin in land plants similar to that of ALP1. Notably, all of the land plant ALP2 proteins lacked the three tryptophan residues that are conserved in *Harbinger* DNA binding transposases and the HDP2 exapted and are required for their DNA binding activity. ALP2 is therefore unlikely to bind DNA. Importantly, ALP2 is required for the interaction between ALP1 and MULTICOPY SUPPRESSOR OF IRA1 (MSI1), a core component of the H3K27me3 histone methyltransferase complex PRC2, and may displace the accessory components EMBRYONIC FLOWER 1 (EMF1) and LIKE

HETEROCHROMATIN PROTEIN 1 (LHP1) to form a variant PRC2 complex to antagonize PRC2-mediated H3K27me3 deposition and gene silencing (Velanis, Perera et al. 2020). The functional studies clearly showed the importance of ALP1 and ALP2, most likely by controlling PRC2 activities during important developmental transitions. However, more experiments and structural approaches of higher resolution are needed to reveal the significance of the various protein in the ALP1-ALP2-MSI1 interaction.

2.5.2. The Harbinger transposon-derived proteins (HDPs)

HARBINGER DERIVED PROTEIN 1 (*HDP1*) and *HARBINGER DERIVED PROTEIN 2* (*HDP2*) were discovered by a genetic forward screen. This screen system is an *Arabidopsis* transgene reporter system, in which expression of the 35S promoter-driven *SUC2* transgene (*35S::SUC2*) causes a DNA demethylation pathway-dependent short-root phenotype on sucrose medium (Lei, Zhang et al. 2015). The null recessive mutants, *hdp1-1*, *hdp1-2* and *hdp2-1* displayed normal root length, though mapping and genetic complementation, HDP1 and HDP2 were suggested to cause the transgene silencing. Both *HDP1* and *HDP2* are derived from *Harbinger* transposons. A BLASTP search using the *Harbinger* transposon database revealed that HDP1 shares about 30% identity with Harbinger transposase sequences. Phylogenetic analysis suggests HDP1 exapted from *Harbinge*r transposase and the derivation of HDP1 had likely happened before the emergence of angiosperms. HDP2 protein contains a SANT/Myb/trihelix DNA-binding motif that is conserved among *Harbinger* proteins from *Arabidopsis lyrata*. The DBD of HDP2 is capable of binding DNA in vitro, and at the genomic level, the protein is significantly enriched at MBD7 locations. However, the other fraction of the HDP2 is divergent from *Harbinger* transposase. Proteomic studies revealed that the HDP1 and HDP2 proteins interact together as

well as with several components of the increased DNA methylation 1 (IDM1) histone acetyltransferase complex, including the methyl-DNA-binding protein MBD7, to prevent DNA hypermethylation and TE silencing (Duan, Wang et al. 2017). Thus, the exapted HDP1 and HDP2 are recruited to histone acetyltransferase complex to promote the expression level of transposons and antagonize DNA methylation-mediated gene silencing.

2.6. The *MUSTANG* (*MUG*) family

Unlike the *FRS* or other ETE families in Arabidopsis that were identified by classical genetic approaches, *MUSTANG* (*MUG*) was the first ETE family discovered in *A. thaliana* and *O. sativa* by systematic bioinformatics approach (Cowan, Hoen et al. 2005). MUG genes are derived from the transposase mudrA of MULEs and are widely distributed across angiosperms. The *A. thaliana* genome encodes eight MUG homologs gathering into two distinct clades: MUGA (MUG1 to MUG4) and MUGB (MUG5 to MUG8) (Joly-Lopez, Forczek et al. 2012). It has been suggested that MUGA and MUGB clades come from two independent exaptation events (Joly-Lopez, Hoen et al. 2016). Although mug1, mug2 and mug3 simple mutants are phenotypically indistinguishable from wild-type plants, high-order mutant combinations such as *mug1 mug2* or *mug1 mug2 mug3* can seriously impact plant growth, flower development and fertility (Joly-Lopez, Hoen et al. 2016). Similarly, the mug7 mug8 double mutant displays severe growth defects (Joly-Lopez, Forczek et al. 2012). Like FAR1 and MURA, MUG proteins carry the typical MULE transposase domain, middle catalytic domain, and a c-terminal SWIM domain. In addition, the MUGB MUG7 and MUG8 carry the Phox and Bem1p (PB1) domain of an unknown function, most likely acquired through a transduplication event (Joly-Lopez, Forczek et al. 2012). It was proposed that MUG proteins act as putative transcriptional regulators of plant

development. Recently, the MUG genes were shown to be associated with abiotic stress phenotype (Joly-Lopez, Forczek et al. 2017). Starting from the mutant phenotype, the *MUG-A* family will be described in detail in the other chapters of this thesis.

2.7. The *Gary* family

Where the aforementioned exapted families were mostly distributed in angiosperms, *GARY* has only been identified in grasses and not in eudicots (Hoen and Bureau 2012). Same with the *MUG* family, *GARY* was detected *in silico* by seeking *hAT* transposase-like sequences. The homologous genes of *GARY* were only found in one or two copies in barley, two diverged copies in rice and two similar copies in hexaploid wheat (Muehlbauer, Bhau et al. 2006). Like other exapted TEs, GARY is not mobile, as shown by the lack of TIR and key residues required for mobility (Muehlbauer et al., 2006). The function of *GARY* is not known, however, its grass-specific distribution and its expression pattern in wheat and barley suggest that it may have a grass-specific reproductive function. In addition, the phylogenetic distributions of *GARY* indicate molecular exaption is an ongoing process in monocots (Hoen and Bureau 2012).

3. The role of TE and ETE in crop improvement

3.1. The effects of TEs for crop traits

In plants, TEs play an important role in agronomic traits and crop domestication. Many TE-induced mutations have been selected during plant domestication. Among them, *Rider*, a high-copy retrotransposon was reported in several cases for its contribution to various phenotypes of agronomical interest. For example, the locus *SUN* in the *Solanaceae* family controls plant

morphology by producing an elongated and oval-shaped fruit. The variation in the locus *SUN* leads to differentiation of fruit shape between some varieties of *Solanum lycopersicum* and its wild parent *Solanum pimpinellifolium* (Wu, Xiao et al. 2011). Through genome structure analysis, it is found in elongated fruits, *Rider* mediated an inter-chromosomal duplicate event, which created the *Sun* locus, and put it in a new genomic context, resulting in increased expression *SUN* in the fruit compared to the ancestral copy and hence altered fruit shape (Xiao, Jiang et al. 2008). Another *Rider*-induced phenotype that has been selected for tomato breeding is jointless pedicel which reduces pedicel abscission and thus facilitates yielding and post-harvest fruit quality. *Rider* inserts into the first intron of the jointless allele *Solyc12g038510* and disrupts the function of MADS-box protein SlMBP21 by providing an alternative transcription start site which results in an early nonsense mutation (Roldan, Périlleux et al. 2017). *Rider* can also change fruit colour. A *Rider* insertion within the gene *PSY1*, which encodes a fruit-specific phytoene synthase involved in carotenoid biosynthesis is strongly associated with yellow tomato compared to the wild-type red tomato (Domínguez, Dugas et al. 2020). Not only *Rider*, but *Gret1* (*grapevine retrotransposon 1*) also generates colour difference. In grapes, red plant pigments called anthocyanins determine the colour of grape skins. *VvmybA1* is a *Myb*-related gene that regulates anthocyanin biosynthesis. The insertion of *Gret1* in *VvmybA1* leads to the white-skinned cultivars Italia and Muscat of Alexandria (Kobayashi, Goto-Yamamoto et al. 2004). Similar insertion of an MYB transcriptional activator of anthocyanin production was found in orange and leads to varieties of oranges. *Citrus sinensis Ruby* gene encodes Myb TF that is involved in the regulation of anthocyanin biosynthesis. In the Navalina variety, the *Ruby* gene is normally functional and shows a rather limited relative expression in the fruit pulp. In the Tarocco variety, the copia-like LTR retrotransposon *Tcs1* inserted upstream of the

coding sequence of *Ruby*, services as a new promoter of *Ruby*, resulting in an increased Ruby expression with consequent accumulation of anthocyanins in the pulp. *Tcs1* is activated by low temperatures. The recombination between the LTRs of *Tcs1* led to the enhanced transcriptional activity of *Ruby*, which gives rise to the Maro variety, with even more anthocyanin in the pulp (Butelli, Licciardello et al. 2012).

TEs not only regulate the expression of nearby genes but also influence gene activity long distance from the coding region. In maize, *Hopscotch* inserted approximately 60 kb upstream from the *TEOSINTE BRANCHED1* (*TB1*) gene, which was a major contributor to the increase in apical dominance as well as a different position of male and female inflorescences during maize domestication. *Hopscotch* acts as an enhancer of *TB1* gene expression and partially explains the increased apical dominance in maize in respect to its progenitor, teosinte (Studer, Zhao et al. 2011). Another trait selected for TE long-distance regulation is flowing time in maize, which leads maize to adapt to local temperate zones, enabling the plant to reach the mature state within a shorter growing season. *Vegetative to generative transition 1* (*Vgt1*), which locates about 70 kb upstream of the APETALA2 (AP2) TF, negatively regulates flowering time. It was demonstrated that the insertion of a *MITE* leads to the formation of *Vgt1* (Ducrocq, Madur et al. 2008).

3.2. The implication of ETEs in agronomic improvement for abiotic stress

Plants are sessile organisms, they have to endure the negative impact of environmental factors such as salinity, drought, extreme temperatures, oxidative stress, metal toxicity and multiple biotic stresses. These adverse conditions greatly limit the distribution of plants, alter their growth and development, and reduce crop productivity. To survive or adapt to environmental stresses,

plants have evolved multiple strategies, such as altering growth and development through specialized metabolism, modifications in morphology. In extreme conditions, they may compress their life cycle to survive the stresses in the form of seeds (Haak, Fukao et al. 2017). A lot of studies from different fields, such as molecular genetics, physiology, biochemistry have uncovered complex regulatory processes that respond to abiotic stress and lead tolerance in plants. The plant responds to the stresses in multiple layers too, such as the cell wall, plasma membrane, cytoplasm, nucleus, chloroplasts, mitochondria, endoplasmic reticulum and peroxisomes (Zhang, Lv et al. 2018). Stress-induced signal transduction often triggers a genome-wide transcriptional reprogramming process and eventually generates protective mechanisms, such as osmotic adjustment, detoxification, stress-induced damage repair and amplifying or weaken the stress signal (Zhang, Zhu et al. 2021). The transcriptional regulation of stress-responsive genes is a crucial part of the plant response to a range of abiotic stresses and the transcriptional reprogramming is through transcription regulators. Functional characterization of key transcriptional regulators such as AP2/ERF, WRKY, MYB, NAC, bZIP that govern multiple signalling processes and directly regulate stress-responsive genes has contributed to dissecting intricate regulatory networks (Yoon, Seo et al. 2020). Transcription regulators are perfect candidates for genetic engineering of agricultural crops resistant to abiotic stress (Zaikina, Rumyantsev et al. 2019). They are also of great interest in molecular marker-assisted selection in plant breeding programmes (Hasan, Choudhary et al. 2021).

ETEs perform various beneficial functions to the host genome, such as centromere binding, chromosome segregation, meiotic recombination, heterochromatin formation, V(D)J recombination, genome stability, and translational regulation (Feschotte and Pritham 2007). Among them, an important role is exapted into transcription regulators (Feschotte 2008). Most

documented ETEs derived from DNA transposons have retained inherent characteristics of the transposase, by encoding for proteins for transcription regulators, where DNA binding, transcriptional activation or suppression ability are required (Ritter, Inigo et al. 2017, Liu, Ma et al. 2020). In plants, all the ETE families that have been characterized in *A. thaliana* has been uncovered to be derived from DNA transposons and function as transcriptional regulators (Hoen and Bureau 2012, Joly-Lopez and Bureau 2018). Recently, novel protein-coding ETEs in *A. thaliana* are associated with abiotic stress phenotypes, suggesting essential roles of them in plant tolerance to abiotic stress (Joly-Lopez, Forczek et al. 2017). Crops, such as maize and rice, have a large amount of ETEs in their genomes. Therefore, ETEs will be a promising reservoir for broadening the pool of the transcriptional regulator selected for crop abiotic stress-associated traits.

Reference

Ahmed, A. U., J. Hoddinott, N. Abedin and N. Hossain (2021). "The impacts of GM foods: results from a randomized controlled trial of Bt eggplant in Bangladesh." American Journal of Agricultural Economics **103**(4): 1186-1206.

Ali, S. A., M. J. Peffers, M. J. Ormseth, I. Jurisica and M. Kapoor (2021). "The non-coding RNA interactome in joint health and disease." Nature Reviews Rheumatology: 1-14.

Allen, S. E. and M. Nowacki (2017). "Necessity is the mother of invention: ciliates, transposons, and transgenerational inheritance." Trends in genetics **33**(3): 197-207.

Babakhani, S. and M. Oloomi (2018). "Transposons: the agents of antibiotic resistance in bacteria." Journal of basic microbiology **58**(11): 905-917.

Benoit, M., H.-G. Drost, M. Catoni, Q. Gouil, S. Lopez-Gomollon, D. Baulcombe and J. Paszkowski (2019). "Environmental and epigenetic regulation of Rider retrotransposons in tomato." PLOS Genetics **15**(9): e1008370.

Bingham, P. M., M. G. Kidwell and G. M. Rubin (1982). "The molecular basis of PM hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family." Cell **29**(3): 995-1004.

Borgognone, A., R. Castanera, E. Muguerza, A. G. Pisabarro and L. Ramírez (2017). "Somatic transposition and meiotically driven elimination of an active helitron family in Pleurotus ostreatus." Dna Research **24**(2): 103-115.

Bourque, G., K. H. Burns, M. Gehring, V. Gorbunova, A. Seluanov, M. Hammell, M. Imbeault, Z. Izsvák, H. L. Levin and T. S. Macfarlan (2018). "Ten things you should know about transposable elements." Genome biology **19**(1): 1-12.

Buchmann, J. P., A. Löytynoja, T. Wicker and A. H. Schulman (2014). "Analysis of CACTA transposases reveals intron loss as major factor influencing their exon/intron structure in monocotyledonous and eudicotyledonous hosts." Mobile DNA **5**(1): 1-15.

Bundock, P. and P. Hooykaas (2005). "An Arabidopsis hAT-like transposase is essential for plant development." Nature **436**(7048): 282-284.

Bureau, T. E. and S. R. Wessler (1992). "Tourist: a large family of small inverted repeat elements frequently associated with maize genes." The Plant Cell **4**(10): 1283-1294.

Butelli, E., C. Licciardello, Y. Zhang, J. Liu, S. Mackay, P. Bailey, G. Reforgiato-Recupero and C. Martin (2012). "Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges." The Plant Cell **24**(3): 1242-1255.

Canadell, J. G. and R. B. Jackson (2021). "Ecosystem collapse and climate change." Ecological Studies **241**.

Cheeseman, J. (2016). Food security in the face of salinity, drought, climate change, and population growth. Halophytes for food security in dry lands, Elsevier**:** 111-123.

Chen, S. and X. Li (2008). "Molecular characterization of the first intact Transib transposon from Helicoverpa zea." Gene **408**(1-2): 51-63.

Cho, J. (2018). "Transposon-derived non-coding RNAs and their function in plants." Frontiers in plant science **9**: 600.

Cho, J., M. Benoit, M. Catoni, H.-G. Drost, A. Brestovitsky, M. Oosterbeek and J. Paszkowski (2019). "Sensitive detection of pre-integration intermediates of long terminal repeat retrotransposons in crop plants." Nature plants **5**(1): 26-33.

Cordaux, R. and M. A. Batzer (2009). "The impact of retrotransposons on human genome evolution." Nature Reviews Genetics **10**(10): 691-703.

Cowan, R. K., D. R. Hoen, D. J. Schoen and T. E. Bureau (2005). "MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms." Molecular Biology and Evolution **22**(10): 2084-2089.

Craig, N. L. (1995). "Unity in transposition reactions." Science **270**(5234): 253-253.

Creasey, K. M., J. Zhai, F. Borges, F. Van Ex, M. Regulski, B. C. Meyers and R. A. Martienssen (2014). "miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis." Nature **508**(7496): 411-415.

de Luxán-Hernández, C., J. Lohmann, W. Hellmeyer, S. Seanpong, K. Wöltje, Z. Magyar, A. Pettkó-Szandtner, T. Pelissier, G. De Jaeger and S. Hoth (2020). "PP7L is essential for MAIL1-mediated transposable element silencing and primary root growth." The Plant Journal **102**(4): 703-717.

Deininger, P. (2011). "Alu elements: know the SINEs." Genome biology **12**(12): 1-12.

Dodonova, S. O., S. Prinz, V. Bilanchone, S. Sandmeyer and J. A. G. Briggs (2019). "Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses." Proc Natl Acad Sci U S A **116**(20): 10048-10057.

Doolittle, W. F. and C. Sapienza (1980). "Selfish Genes, the Phenotype Paradigm and Genome Evolution." Nature **284**(5757): 601-603.

Duan, C.-G., X. Wang, S. Xie, L. Pan, D. Miki, K. Tang, C.-C. Hsu, M. Lei, Y. Zhong and Y.-J. Hou (2017). "A pair of transposon-derived proteins function in a histone acetyltransferase complex for active DNA demethylation." Cell research **27**(2): 226-240.

Dupeyron, M., T. Baril, C. Bass and A. Hayward (2020). "Phylogenetic analysis of the Tc1/mariner superfamily reveals the unexplored diversity of pogo-like elements." Mobile DNA **11**(1): 1-14.

Eickbush, T. H. and D. G. Eickbush (2015). "Integration, regulation, and long-term stability of R2 retrotransposons." Microbiology spectrum **3**(2): 3.2. 02.

Engels, W. R. and C. R. Preston (1981). "Identifying P factors in Drosophila by means of chromosome breakage hotspots." Cell **26**(3): 421-428.

Evgen'ev, M. and I. Arkhipova (2005). "Penelope-like elements–a new class of retroelements: distribution, function and possible evolutionary significance." Cytogenetic and genome research **110**(1-4): 510-521.

Faulkner, G. J. and V. Billon (2018). "L1 retrotransposition in the soma: a field jumping ahead." Mobile DNA **9**(1): 1-18.

Feschotte, C. (2004). "Merlin, a new superfamily of DNA transposons identified in diverse animal genomes and related to bacterial IS 1016 insertion sequences." Molecular Biology and Evolution **21**(9): 1769-1780.

Filée, J., S. Farhat, D. Higuet, L. Teysset, D. Marie, C. Thomas-Bulle, S. Hourdez, D. Jollivet and E. Bonnivard (2021). "Comparative genomic and transcriptomic analyses of transposable elements in polychaetous annelids highlight LTR retrotransposon diversity and evolution." Mobile DNA **12**(1): 1-24.

Friesen, N., A. Brandes and J. S. Heslop-Harrison (2001). "Diversity, Origin, and Distribution of Retrotransposons (gypsy and copia) in Conifers." Molecular Biology and Evolution **18**(7): 1176-1188.

Fultz, D., S. G. Choudury and R. K. Slotkin (2015). "Silencing of active transposable elements in plants." Current opinion in plant biology **27**: 67-76.

Gao, X., E. R. Havecker, P. V. Baranov, J. F. Atkins and D. F. Voytas (2003). "Translational recoding signals between gag and pol in diverse LTR retrotransposons." RNA (New York, N.Y.) **9**(12): 1422-1430.

Grandbastien, M.-A. (1998). "Activation of plant retrotransposons under stress conditions." Trends in plant science **3**(5): 181-187.

Haak, D. C., T. Fukao, R. Grene, Z. Hua, R. Ivanov, G. Perrella and S. Li (2017). "Multilevel regulation of abiotic stress responses in plants." Frontiers in plant science **8**: 1564.

Harrison, J. J. E. K., S. Tuske, K. Das, F. X. Ruiz, J. D. Bauman, P. L. Boyer, J. J. DeStefano, S. H. Hughes and E. Arnold (2021). "Crystal Structure of a Retroviral Polyprotein: Prototype Foamy Virus Protease-Reverse Transcriptase (PR-RT)." Viruses **13**(8): 1495.

Hartwig, B., G. V. James, K. Konrad, K. Schneeberger and F. Turck (2012). "Fast isogenic mapping-by-sequencing of ethyl methanesulfonate-induced mutant bulks." Plant physiology **160**(2): 591-600.

Hickey, D. A. (1982). "Selfish DNA: a sexually-transmitted nuclear parasite." Genetics **101**(3-4): 519-531.

Hoegh-Guldberg, O., D. Jacob, M. Bindi, S. Brown, I. Camilloni, A. Diedhiou, R. Djalante, K. Ebi, F. Engelbrecht and J. Guiot (2018). "Impacts of 1.5 C global warming on natural and human systems." Global warming of 1.5 C. An IPCC Special Report.

Hoen, D. R. and T. E. Bureau (2012). Transposable element exaptation in plants. Plant transposable elements, Springer**:** 219-251.

Hoen, D. R. and T. E. Bureau (2015). "Discovery of Novel Genes Derived from Transposable Elements Using Integrative Genomic Analysis." Molecular Biology and Evolution **32**(6): 1487-1506.

Hollister, J. D. and B. S. Gaut (2009). "Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression." Genome research **19**(8): 1419-1428.

Hudson, M., C. Ringli, M. T. Boylan and P. H. Quail (1999). "The FAR1 locus encodes a novel nuclear protein specific to phytochrome A signaling." Genes Dev **13**(15): 2017-2027.

Hudson, M. E., D. R. Lisch and P. H. Quail (2003). "The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway." Plant Journal **34**(4): 453-471.

Ikeda, Y., T. Pélissier, P. Bourguet, C. Becker, M.-N. Pouch-Pélissier, R. Pogorelcnik, M. Weingartner, D. Weigel, J.-M. Deragon and O. Mathieu (2017). "Arabidopsis proteins with a transposon-related domain act in gene silencing." Nature communications **8**(1): 1-10.

Ito, H., H. Gaubert, E. Bucher, M. Mirouze, I. Vaillant and J. Paszkowski (2011). "An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress." Nature **472**(7341): 115-119.

Johns, M. A., J. Mottinger and M. Freeling (1985). "A low copy number, copia-like transposon in maize." The EMBO journal **4**(5): 1093-1101.

Joly-Lopez, Z. and T. E. Bureau (2018). "Exaptation of transposable element coding sequences." Current opinion in genetics & development **49**: 34-42.

Joly-Lopez, Z., E. Forczek, D. R. Hoen, N. Juretic and T. E. Bureau (2012). "A Gene Family Derived from Transposable Elements during Early Angiosperm Evolution Has Reproductive Fitness Benefits in Arabidopsis thaliana." Plos Genetics **8**(9).

Joly-Lopez, Z., E. Forczek, E. Vello, D. R. Hoen, A. Tomita and T. E. Bureau (2017). "Abiotic Stress Phenotypes Are Associated with Conserved Genes Derived from Transposable Elements." Front Plant Sci **8**: 2027.

Joly-Lopez, Z., D. R. Hoen, M. Blanchette and T. E. Bureau (2016). "Phylogenetic and Genomic Analyses Resolve the Origin of Important Plant Genes Derived from Transposable Elements." Mol Biol Evol **33**(8): 1937-1956.

Kabelitz, T., C. Kappel, K. Henneberger, E. Benke, C. Nöh and I. Bäurle (2014). "eQTL mapping of transposon silencing reveals a position-dependent stable escape from epigenetic silencing and transposition of AtMu1 in the Arabidopsis lineage." The Plant Cell **26**(8): 3261-3271.

Kapitonov, V. V. and J. Jurka (2003). "A novel class of SINE elements derived from 5S rRNA." Molecular biology and evolution **20**(5): 694-702.

Kato, M., A. Miura, J. Bender, S. E. Jacobsen and T. Kakutani (2003). "Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis." Current Biology **13**(5): 421-426.

Kempken, F. and F. Windhofer (2001). "The hAT family: a versatile transposon group common to plants, fungi, animals, and man." Chromosoma **110**(1): 1-9.

Kidwell, M. G., J. F. Kidwell and J. A. Sved (1977). "Hybrid dysgenesis in Drosophila melanogaster: a syndrome of aberrant traits including mutation, sterility and male recombination." Genetics **86**(4): 813-833.

Kim, E. Y., L. Wang, Z. Lei, H. Li, W. Fan and J. Cho (2021). "Ribosome stalling and SGS3 phase separation prime the epigenetic silencing of transposons." Nature Plants **7**(3): 303-309.

Klein, S. J. and R. J. O'Neill (2018). "Transposable elements: genome innovation, chromosome diversity, and centromere conflict." Chromosome Research **26**(1): 5-23.

Knip, M., S. de Pater and P. J. Hooykaas (2012). "The SLEEPER genes: a transposase-derived angiosperm-specific gene family." BMC plant biology **12**(1): 1-15.

Knip, M., S. Hiemstra, A. Sietsma, M. Castelein, S. de Pater and P. Hooykaas (2013). "DAYSLEEPER: a nuclear and vesicular-localized protein that is expressed in proliferating tissues." BMC Plant Biol **13**: 211.

Kojima, K. K. and J. Jurka (2011). "Crypton transposons: identification of new diverse families and ancient domestication events." Mobile DNA **2**(1): 1-17.

Kramerov, D. and N. Vassetzky (2011). "Origin and evolution of SINEs in eukaryotic genomes." Heredity **107**(6): 487-495.

Krupovic, M., D. H. Bamford and E. V. Koonin (2014). "Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses." Biology direct **9**(1): 1-7.

Lei, M., H. Zhang, R. Julian, K. Tang, S. Xie and J.-K. Zhu (2015). "Regulatory link between DNA methylation and active demethylation in Arabidopsis." Proceedings of the National Academy of Sciences **112**(11): 3553-3557.

Lemaux, P. G. (2008). "Genetically Engineered Plants and Foods: A Scientist's Analysis of the Issues (Part I)." Annu Rev Plant Biol **59**: 771-812.

Lisch, D. (2009). "Epigenetic regulation of transposable elements in plants." Annual review of plant biology **60**: 43-66.

Liu, Y., M. Ma, G. Li, L. Yuan, Y. Xie, H. Wei, X. Ma, Q. Li, P. F. Devlin and X. Xu (2020). "Transcription factors FHY3 and FAR1 regulate light-induced CIRCADIAN CLOCK ASSOCIATED1 gene expression in Arabidopsis." The Plant Cell **32**(5): 1464-1478.

Liu, Z., C. An, Y. Zhao, Y. Xiao, L. Bao, C. Gong and Y. Gao (2021). "Genome-Wide Identification and Characterization of the CsFHY3/FAR1 Gene Family and Expression Analysis under Biotic and Abiotic Stresses in Tea Plants (Camellia sinensis)." Plants **10**(3): 570.

Majumdar, S. and D. C. Rio (2015). "P transposable elements in Drosophila and other eukaryotic organisms." Microbiology spectrum **3**(2): 3.2. 01.

Malicki, M., T. Spaller, T. Winckler and C. Hammann (2020). "DIRS retrotransposons amplify via linear, single-stranded cDNA intermediates." Nucleic acids research **48**(8): 4230-4243.

Malik, H. S. and T. H. Eickbush (1998). "The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs." Mol Biol Evol **15**(9): 1123-1134.

McClintock, B. (1950). "The Origin and Behavior of Mutable Loci in Maize." Proceedings of the National Academy of Sciences of the United States of America **36**(6): 344-355.

Mirouze, M., J. Reinders, E. Bucher, T. Nishimura, K. Schneeberger, S. Ossowski, J. Cao, D. Weigel, J. Paszkowski and O. Mathieu (2009). "Selective epigenetic control of retrotransposition in Arabidopsis." Nature **461**(7262): 427-430.

Miura, A., S. Yonebayashi, K. Watanabe, T. Toyama, H. Shimada and T. Kakutani (2001). "Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis." Nature **411**(6834): 212-214.

Morgan, T. H. (1910). "Sex Limited Inheritance in Drosophila." Science **32**(812): 120-122.

Muehlbauer, G. J., B. S. Bhau, N. H. Syed, S. Heinen, S. Cho, D. Marshall, S. Pateyron, N. Buisine, B. Chalhoub and A. J. Flavell (2006). "A hAT superfamily transposase recruited by the cereal grass genome." Molecular Genetics and Genomics **275**(6): 553-563.

Nelson, P. N., P. Hooley, D. Roden, H. Davari Ejtehadi, P. Rylance, P. Warren, J. Martin, P. G. Murray and G. Molecular Immunology Research (2004). "Human endogenous retroviruses: transposable elements with potential?" Clinical and experimental immunology **138**(1): 1-9.

Nicolau, M., N. Picault, J. Descombin, Y. Jami-Alahmadi, S. Feng, E. Bucher, S. E. Jacobsen, J.-M. Deragon, J. Wohlschlegel and G. Moissiard (2020). "The plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in Arabidopsis thaliana." PLoS genetics **16**(4): e1008324.

Okamoto, H. and H. Hirochika (2001). "Silencing of transposable elements in plants." Trends in plant science **6**(11): 527-534.

Pélissier, T., C. Bousquet-Antonelli, L. Lavie and J.-M. Deragon (2004). "Synthesis and processing of tRNA-related SINE transcripts in Arabidopsis thaliana." Nucleic acids research **32**(13): 3957-3966.

Pouteau, S., E. Huttner, M. A. Grandbastien and M. Caboche (1991). "Specific expression of the tobacco Tnt1 retrotransposon in protoplasts." The EMBO journal **10**(7): 1911-1918.

Quadrana, L., A. B. Silveira, G. F. Mayhew, C. LeBlanc, R. A. Martienssen, J. A. Jeddeloh and V. Colot (2016). "The Arabidopsis thaliana mobilome and its impact at the species level." elife **5**: e15716.

Roccaro, M., Y. Li, H. Sommer and H. Saedler (2007). "ROSINA (RSI) is part of a CACTA transposable element, TamRSI, and links flower development to transposon activity." Molecular Genetics and Genomics **278**(3): 243-254.

Sanchez, D. H., H. Gaubert and W. Yang (2019). "Evidence of developmental escape from transcriptional gene silencing in MESSI retrotransposons." New Phytologist **223**(2): 950-964.

Shapiro, J. (1969). "Mutations caused by the insertion of genetic material into the galactose operon of Escherichia coli." Journal of molecular biology **40**(1): 93-105.

Shrivastava, P. and R. Kumar (2015). "Soil salinity: A serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation." Saudi journal of biological sciences **22**(2): 123-131.

Sigman, M. J. and R. K. Slotkin (2016). "The first rule of plant transposable element silencing: location, location, location." The Plant Cell **28**(2): 304-313.

Silva, J. C. and M. G. Kidwell (2004). "Evolution of P elements in natural populations of Drosophila willistoni and D. sturtevanti." Genetics **168**(3): 1323-1335.

Singer, T., C. Yordan and R. A. Martienssen (2001). "Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1)." Genes & development **15**(5): 591-602.

Siomi, M. C., K. Sato, D. Pezic and A. A. Aravin (2011). "PIWI-interacting small RNAs: the vanguard of genome defence." Nature reviews Molecular cell biology **12**(4): 246-258.

Stitzer, M. C., S. N. Anderson, N. M. Springer and J. Ross-Ibarra (2021). "The genomic ecosystem of transposable elements in maize." PLoS Genet **17**(10): e1009768.

Tambones, I. L., A. Haudry, M. C. Simão and C. Carareto (2019). "High frequency of horizontal transfer in Jockey families (LINE order) of drosophilids." Mobile DNA **10**(1): 1-15.

Taylor, A. L. (1963). "Bacteriophage-induced mutation in Escherichia coli." Proceedings of the National Academy of Sciences of the United States of America **50**(6): 1043.

Teixeira, F. K., M. Okuniewska, C. D. Malone, R.-X. Coux, D. C. Rio and R. Lehmann (2017). "piRNA-mediated regulation of transposon alternative splicing in the soma and germ line." Nature **552**(7684): 268-272.

Ühlken, C., B. Horvath, R. Stadler, N. Sauer and M. Weingartner (2014). "MAIN-LIKE 1 is a crucial factor for correct cell division and differentiation in A rabidopsis thaliana." The Plant Journal **78**(1): 107-120.

Velanis, C. N., P. Perera, B. Thomson, E. de Leau, S. C. Liang, B. Hartwig, A. Förderer, H. Thornton, P. Arede and J. Chen (2020). "The domesticated transposase ALP2 mediates formation of a novel Polycomb protein complex by direct interaction with MSI1, a core subunit of Polycomb Repressive Complex 2 (PRC2)." PLoS genetics **16**(5): e1008681.

Volff, J. N. (2006). "Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes." Bioessays **28**(9): 913-922.

Wang, H. and H. Wang (2015). "Multifaceted roles of FHY3 and FAR1 in light signaling and beyond." Trends Plant Sci **20**(7): 453-461.

Wang, T., L. Ren, C. Li, D. Zhang, X. Zhang, G. Zhou, D. Gao, R. Chen, Y. Chen and Z. Wang (2021). "The genome of a wild Medicago species provides insights into the tolerant mechanisms of legume forage to environmental stress." BMC biology **19**(1): 1-17.

Wenig, U., S. Meyer, R. Stadler, S. Fischer, D. Werner, A. Lauter, M. Melzer, S. Hoth, M. Weingartner and N. Sauer (2013). "Identification of MAIN, a factor involved in genome stability in the meristems of Arabidopsis thaliana." The plant journal **75**(3): 469-483.

Wessler, S. R. (2006). "Transposable elements and the evolution of eukaryotic genomes." Proceedings of the National Academy of Sciences of the United States of America **103**(47): 17600-17601.

Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel and A. H. Schulman (2007). "A unified classification system for eukaryotic transposable elements." Nature Reviews Genetics **8**(12): 973-982.

Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel and A. H. Schulman (2007). "A unified classification system for eukaryotic transposable elements." Nat Rev Genet **8**(12): 973-982.

Wolf, G., A. de Iaco, M.-A. Sun, M. Bruno, M. Tinkham, D. Hoang, A. Mitra, S. Ralls, D. Trono and T. S. Macfarlan (2020). "KRAB-zinc finger protein gene expansion in response to active retrotransposons in the murine lineage." Elife **9**: e56337.

Xie, Y., Q. Zhou, Y. Zhao, Q. Li, Y. Liu, M. Ma, B. Wang, R. Shen, Z. Zheng and H. Wang (2020). "FHY3 and FAR1 integrate light signals with the miR156-SPL module-mediated aging pathway to regulate Arabidopsis flowering." Molecular Plant **13**(3): 483-498.

Yang, G., F. Zhang, C. N. Hancock and S. R. Wessler (2007). "Transposition of the rice miniature inverted repeat transposable element mPing in Arabidopsis thaliana." Proceedings of the National Academy of Sciences **104**(26): 10962-10967.

Yang, J., H. S. Malik and T. H. Eickbush (1999). "Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements." Proceedings of the National Academy of Sciences **96**(14): 7847-7852.

Yoshitake, Y., N. Inomata, M. Sano, Y. Kato and M. Itoh (2018). "The P element invaded rapidly and caused hybrid dysgenesis in natural populations of Drosophila simulans in Japan." Ecology and evolution **8**(19): 9590-9599.

Yu, Z. H., S. I. Wright and T. E. Bureau (2000). "Mutator-like elements in Arabidopsis thaliana: Structure, diversity and evolution." Genetics **156**(4): 2019-2031.

Yusa, K. (2015). "piggyBac Transposon." Microbiol Spectr **3**(2): Mdna3-0028-2014.

Zemach, A., M. Y. Kim, P.-H. Hsieh, D. Coleman-Derr, L. Eshed-Williams, K. Thao, S. L. Harmer and D. Zilberman (2013). "The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin." Cell **153**(1): 193-205.

Zhang, G., T. Yu, S. S. Parhad, S. Ho, Z. Weng and W. E. Theurkauf (2021). "piRNA-independent transposon silencing by the Drosophila THO complex." Developmental Cell **56**(18): 2623-2635. e2625.

Zhang, H. M., J. H. Zhu, Z. Z. Gong and J. K. Zhu (2021). "Abiotic stress responses in plants." Nature Reviews Genetics.

Zhang, Y., Y. Lv, N. Jahan, G. Chen, D. Ren and L. Guo (2018). "Sensing of abiotic stress and ionic stress responses in plants." International journal of molecular sciences **19**(11): 3298.

Zhao, K., S. Cheng, N. Miao, P. Xu, X. Lu, Y. Zhang, M. Wang, X. Ouyang, X. Yuan and W. Liu (2019). "A Pandas complex adapted for piRNA-guided transcriptional silencing and heterochromatin formation." Nature cell biology **21**(10): 1261-1272.

Zhong, M.-C., X.-D. Jiang, W.-H. Cui and J.-Y. Hu (2021). "Expansion and expression diversity of FAR1/FRS-like genes provides insights into flowering time regulation in roses." Plant diversity **43**(2): 173-179.

Link between Chapter I and Chapter II

In the first chapter, I wrote a literature review as my thesis start to introduce the current concept of TEs, ETEs and discussed the potential application of ETEs. TEs are DNA sequences that can jump around the genome. They were considered to only propagate themselves in the genome and make no contribution towards the hosts. Thus, they were called "selfish DNA" or "junk DNA" for a while. In recent decades, more and more genomes were revealed thanks to the dramatic advancement of DNA sequencing technology, whereby some functional genes identified by forward genetics were discovered that derived from TEs. These TE-derived genes, which become beneficial to hosts, are ETEs. Later, several ETE families derived from the main TE superfamilies were mined by a systematical bioinformatic approach in Arabidopsis. Among those ETEs, *MUSTANG* (*MUG*) was the first ETE family discovered *in silico*. Previously, Drs. Zoé Joly-Lopez and Ewa Forczek showed *mug1 mug2* and *mug7 mug8* double mutants have pleiotropic phenotype in Arabidopsis. However, no molecular profile of *MUG* gene family has been published when I started my graduate study. To fill the blank, in Chapter II, I explored the molecular profile of *MUG-A*, a subfamily of *MUG*, consisting of four gene members, *MUG1* to *MUG4*. A comprehensive set of experiments were performed to detect the subcellular localization, protein-protein interaction, the transcriptional activation or repression capacity of *MUG-A* genes. Phenotyping analyses were also applied to determine the genetic relationship of MUG-A genes.

Chapter II

The molecular characterization of *MUSTANG-A* gene family

Abstract

Exapted transposable elements (ETEs) are transposable elements (TEs) which are repurposed into genes with novel functions that make significant contributions to the host. *MUSTANG-A (MUG-A)* gene family, derived from *Mutator*-Like Element (MULE) transposase gene *MudrA*. It is the first ETE family found in plants using a systematic bioinformatics approach. Previously, *MUG-A genes* were reported to be critical for plant development in *Arabidopsis thaliana* due to the pleiotropic phenotype in *mug1 mug2* double mutants. However, no detailed molecular profile of any *MUG-A* member has been published. To fill this gap and obtain a deeper understanding of the function of the *MUG-A* genes, I performed a comprehensive set of experiments to characterize their molecular profiles. Fluorescent microscopy analysis demonstrates that all MUG-A proteins have a strict nuclear localization, and all of them can mediate transcriptional repression in plants. RNA-Seq data showed evidence that the loss of MUG4 leads to a set of genes change their expression levels and these genes are enriched into photosytheis, response to stress pathways. Yeast two-hybrid (Y2H) and bimolecular fluorescence complementation (BiFC) assay revealed that the MUG-A proteins can form homodimers and heterodimers with other MUG-A proteins. The genetic phenotypic analysis confirmed and bolstered the initial observations, which shows that the *MUG-A* proteins have overlapping essential functions that are beneficial to plant fitness. These detailed molecular and genetic evidence suggest MUG-A proteins work as a transcriptional repressor complex that regulates a set of genes that are essential for plant development. This work provides a key step toward understanding the mechanism of how *MUG-A* ETEs benefit the host.

Introduction

TEs are diverse genomic elements that do something that most genes cannot: they move within the genome. The first TEs were found in maize by Barbara McClintock in 1950 (McClintock 1950). TEs can be grouped into two classes based on their transposition mechanism. Class I or retrotransposons, move via a "copy and paste" mechanism mediated by their RNA; Class II, or DNA transposons, use a "cut and paste" mechanism to change their positions. Both classes can be divided into autonomous and non-autonomous TEs depending on if they carry a functional transposase (Wessler 2006).

These jumping sequences had been called "selfish DNA" or "junk DNA" for several years since they are deleterious towards the host as their mobility can disrupt the host genes, such as integrating into regulatory or coding regions of conventional genes, inducing ectopic or nonallelic recombination events (Jangam, Feschotte et al. 2017). In recent decades, more evidence showed that transposon-derived genes make contributions to the hosts (Cowan, Hoen et al. 2005). In plants, the greatest number of transposon-derived genes are DNA transposons (Volff 2006). The TE-derived genes, which become beneficial to hosts, are named ETEs (Hoen and Bureau 2015). They become exapted mainly through three mechanisms (Sinzelle, Izsvak et al. 2009). The first is that a complete transposase gene becomes exapted. An example is the *RAG1* gene, derived from a DNA transposase, involved in V(D)J recombination in the immune system of jawed vertebrates. The second mechanism is creating a chimeric protein by fusing a whole transposase gene with a host gene. For instance, *PiggyBac Transposable Element Derived 3* (*PGBD3*) transposon inserted into the *Cockayne syndrome B* (*CSB*) gene 43 million years ago and created a fusion protein CSB-PGBD3, which regulates the expression of nearby genes in

humans (Gray, Fong et al. 2012). The third mechanism is the fusion of a portion of a transposase with a host gene. For instance, the THAP proteins, with the DNA-binding domain derived from *P* transposase, have important roles in cell proliferation, cell cycle control, and apoptosis (Sabogal, Lyubimov et al. 2010).

The above examples were found through forward genetics. To date, in *A. thaliana,* two gene families derived from whole transposases that have been discovered and studied, were also found by forward genetics. One is FAR1/FHY3 family which originated from *MudrA*, a transposase gene from *MULE*. FAR1/FHY3 proteins are TFs involved in phyA signal transduction (Hudson, Lisch et al. 2003). The other is the *DAYSLEEPER* gene, derived from a transposase gene of the *hAT* TE superfamily, which is also essential for plant development (Bundock and Hooykaas 2005).

Compared with the large number of TEs in the genome, the number of mined ETEs is limited and there are likely more undiscovered ETEs in the genome. Bioinformatics approaches were thus established to search for more genes that have the same pattern as the evolved transposes-like genes. *MUSTANG* (*MUG*) is such a gene family which was found through reverse genetics (Cowan, Hoen et al. 2005). The *MUG* gene family can be divided into *MUSTANGA (MUG-A)* and *MUSTANG-B (MUG-B)*. *MUG-A* genes share same gene architecture with *mudrA*, while *MUG-B* genes have an additional PB1 domain encoding sequence in the N-terminal. Previously, *MUG1* and *MUG2* in the *MUG-A* family were reported to be involved in diverse processes impacting plant fitness, including flowering, growth and reproduction (Joly-Lopez, Forczek et al. 2012). However, no detailed molecular characterization of a *MUG-A* family member has yet to

be described. Therefore, this study aims to fill the gap and uncover the biological function of the *MUG-A* gene family.

Results

All MUG-A proteins are localized to the nucleus.

Finding out the protein subcellular localization is important for understanding the function of proteins. The ancient transposase MURA and the ETE FAR1 were experimentally verified that they are both located in the nucleus (Hudson, Ringli et al. 1999, Ono, Kim et al. 2002). Are MUG-A proteins still localized in the nucleus or in a different subcellular structure? First, I used the program LOCALIZER (http://localizer.csiro.au/) to predict their subcellular localization. All MUG-A proteins are predicted to have a nuclear localization signal (NLS) (Figure 2.1a). I then experimentally verified the prediction. MUG1-mCherry, MUG2-mCherry, MUG3-mCherry and MUG4-mCherry fusion proteins were expressed under the control of the 35S promoter and infiltrated into the leaves of *Nicotiana benthamiana*. DAPI staining was used to visualize nucleus localization. As expected, all the MUG-A proteins showed exclusive and clear nucleus signal overlapped with DAPI staining (Figure 2.1b). Besides transient expression in tobacco, I also made stable transgenic *Arabidopsis* lines expressing MUG4-GFP driven by 35S promoter or native promoter and GFP alone to confirm the localization. With the GFP construct alone, the GFP signal was found constitutively expressed both in the nucleus and in the cytoplasm. Consistent with the transient expression result, both the 35S promoter-driven and native protomer-driven stable transgenic lines expressing MUG4-GFP showed a clear nucleus signal overlapped with DAPI staining, while free GFP does not overlap with DAPI (Figure 2.1c). These results demonstrate that all MUG-A proteins are localized to the nucleus of the cells.

Domain analysis of MUG-A proteins.

Domains are the building blocks of protein structures, typically associated with a specific function. To investigate the molecular function of *MUG-A* gene family members, I first searched the domain architecture of MUG-A proteins. InterPro protein signature database searches (https://www.ebi.ac.uk/interpro/) showed MUG-A proteins share a conserved domain architecture: An N-terminal C2H2-type zinc-chelating domain (IPR004332), a central putative core transposase domain (IPR018289), and a C-terminal SWIM domain (IPR007527), which are the same as the ancestral transposase (Figure 2.2a).

Given that the domain architecture is conserved, I calculated the identity among the domains using protein blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Figure 2.2b). The N-terminal domain of MUG-A proteins share an identity score of 32.69% ~ 38.30% to MURA, however, the sequence identity is not significant to FAR1 according to blastp results. I then calculated the identity of the N-terminal domain between MURA and FAR1, which is not significant, either. However, while not significant, both MURA and FAR1 were reported to bind to DNA sequences: MURA binds to the TIR sequence of MULEs (Benito and Walbot 1997), and FAR1 binds to the promoters of conventional genes such as FHY1 and FHL (Lin, Ding et al. 2007). Therefore, MUG-A proteins may bind to DNA of conventional genes in a manner that evolved directly from the way in which MURA transposase binds to terminal inverted repeats (TIRs) sequences (Cowan, Hoen et al. 2005).

The middle domain of MUG1 and MUG4 share 37.89% and 30.53% to MURA, 24.21% and 29.85% to FAR1, respectively. As to MUG2 and MUG3, they share 29.07% and 32.61% to MURA, not significant to FAR1. In MURA, this domain has not been investigated

experimentally, however, the acidic amino acid triad, DDE or DDD in the domain that is conserved among the MULE transposase are proved to be essential for catalyzing the "cut and paste" transposition reaction (Yuan and Wessler 2011), similar with the catalytic motif of the retroviral integrase (Makarova, Aravind et al. 2002). Interestingly, the middle domain of FAR1 and FHY3, the homologous of FAR1, shares 27.27% and 20.45% identity scores to MURA. Instead of catalyzing the transposition process, the middle domain of FHY3 is essential for transcriptional regulatory activity and its homodimerization or heterodimerization with FAR1 (Lin, Teng et al. 2008). As MUG-A proteins are also ETEs, they likely evolved new functions from MURA, like FAR1 and FHY3.

The SWIM domain was designated after SWI2/SNF2 and MuDR, the autonomous DNA transposon that harbours the transposase MURA. MUG1, MUG2, MUG3 and MUG4 share 32.26%, 33.33%, 40.00% and 38.71% identity with MURA, respectively, while 41.67%, 43.48%, 34.78%, 34.78% with FAR1. This domain appears to be versatile with DNA-binding or protein-protein interaction (Makarova, Aravind et al. 2002). As the N-terminal domain of MURA performs the TIR-DNA binding function, the C-terminal SWIM domain is likely to perform the protein-protein interaction function to form a dimer when a DNA transposase binds and cuts target DNA. In addition, previous experimental evidence shows that the C-terminal domain of FHY3 is essential for its transcriptional regulatory activity and its homodimerization or heterodimerization with FAR1 (Lin, Teng et al. 2008). With the high identity score, the SMIM domain of MUG-A proteins could be used for protein-protein interaction.

MUG-A proteins do not have transcriptional activation activity in yeast.

FAR1 has an experimentally verified transcriptional activation activity in both yeast and plants (Wang and Deng 2002, Hudson, Lisch et al. 2003). As MUG-A proteins share a high identity score with FAR1 in the C-terminal SWIM domain, and this domain is used for transcriptional activation activity of FHY3, the homologous partner of FAR1, I tested if MUG-A proteins also have transcriptional activation activity using a modified yeast one-hybrid assay (Lu, Tang et al. 2010). The full-length MUG1, MUG2, MUG3, MUG4, and FAR1 were linked with the GAL4 binding domain (GAL4BD), then transformed into yeast Y2H Gold reporter strain. All the cells can grow in the control medium. The colour of the yeast expressing FAR1 and the positive control GAL4BD linked with GAL4 activation domain (GAL4AD) are white, while the one expressing MUG1, MUG2, MUG3, MUG4 and negative control, GAL4BD alone are red. The yeast containing MUG4, and the negative control plasmid were unable to grow in the selective medium, whereas the cells containing FAR1 and GAL4BD-AD grew well. The cells containing MUG1, MUG2 and MUG3, can grow in the selective medium but slowly and weakly (Figure 2.3a). However, MUG4 does not. These results suggest that FAR1 has transcriptional activation activity in yeast, which is consistent with the previous finding. The red colour of MUG1, MUG2 and MUG3 in the control media suggests the accumulation of the substrate of ADE2 protein, thus, the weak growth of MUG1, MUG2 and MUG3 in the selective media may be a false positive phenotype.

MUG-A proteins have transcriptional suppression ability in plants.

Since MUG-A genes were only found in angiosperms (Joly-Lopez, Hoen et al. 2016), not in yeast, I tested whether MUG-A proteins have intrinsic transcriptional regulatory ability in plants. The full-length *MUG-A* and *FAR1* cDNA sequences were fused with the GLABRA 2 (GL2)

DNA binding domain (GL2DBD) which can bind to the promoter region of RSL1 (pRSL1) (Lin, Ohashi et al. 2015). The control, which was transformed with pRSL1-GFP only, serves as the basal expression level driven by RLS1 promoter in tobacco leaves. As expected, the tobacco leaves that were co-transformed with pRSL1-GFP and GL2DBD fused with FAR1 showed a strong GFP signal compared to the background (Figure 2.3b), with a 3-fold intensity relative to the control (Figure 2.3c). The tobacco leaves that were co-transformed with pRSL1-GFP and GL2DBD fused with MUG4 almost lost all the GFP signal (Figure 2.3b), which is consistent with the yeast one-hybrid result, that MUG4 does not have transcriptional activation activity. Furthermore, the signal of MUG4 is even weaker than the background, less than 0.5-fold relative to the control (Figure 2.3c), which suggests MUG4 suppressed the expression of the pRSL1 in plant. At the same time, the signal of MUG1, MUG2 and MUG3 are also weaker than control (Figure 2.3b), whereas the GFP intensity is less than half of the control (Figure 2.3c), indicating that MUG1, MUG2 and MUG3 has transcriptional suppression ability in plants.

MUG-A proteins can form homodimer with themselves and heterodimer with other MUG-A proteins *in planta*.

Homodimerization is a common feature for DNA transposases (Benito and Walbot 1997, Bhasin, Goryshin et al. 2000). The SWIM domain of the ETE FAR1 is essential for its homodimer formation (Hudson, Lisch et al. 2003) and heterodimer formation with FHY3 (Wang and Deng 2002). As MUG-A proteins share a highly conserved SWIM domain with FAR1, I tested if MUG-A proteins also have this characteristic using bimolecular fluorescence complementation (BiFC) in planta. In the assay, Yellow Fluorescent Protein (YFP) is split into N-terminal (nYFP) and C-terminal (cYFP) parts. MUG1, MUG2, MUG3 and MUG4 were fused to the nYFP or

cYFP upstream, respectively. As shown in Figure 2.4a left panel, strong YFP fluorescence was observed of the MUG1, MUG2, MUG3 and MUG4 self-interacts. Additionally, FAR1 was used as a negative control with the corresponding BiFC constructs, as expected, no fluorescence signal was detected (Figure 2.4a, right panel). These results suggest that all the MUG-A proteins can form homodimers.

I then checked if MUG-A proteins can form heterodimers like FAR1 and FHY3. The MUG1, MUG2, MUG3 and MUG4 constructions described above were introduced into tobacco cells in pairwise combinations. The reconstitution of fluorescent signal mediated by all the MUG-A protein-protein pairwise interaction was detected in the nuclei of the tobacco epidermal cells (Figure 2.4b). This result suggests that MUG1, MUG2, MUG3 and MUG4 can form heterodimers.

To validate the BiFC results, I applied the Y2H assay. MUG4 was fused with GAL4AD, meanwhile, MUG1, MUG2, MUG3, MUG4 and FAR1 were fused with GAL4BD. Even though MUG1-GAL4BD and MUG2-GAL4BD co-transformed with GAl4AD yeast strains could grow slowly on SD Trp-Leu-His, MUG1-GAL4BD, MUG2-GALBD co-transformed with MUG4-Gal4AD yeast strains grow faster (Figure 2.4c). The results suggest that MUG4 can form heterodimers with MUG1 and MUG2.

MUG4 regulates different pathways during development.

To further elucidate the mechanisms of how *MUG-A* proteins work as transcriptional repressors, using mug4 as a representative, I analyzed the transcriptome of *mug4-2* and wild-type obtained

by RNA-Seq. The transcriptome of 10-day-old *mug4-2* mutants seedlings was compared to that in 10-day-old wild-type seedlings.

In total 700 differential expression genes, among them, 431 genes were upregulated, and 269 genes were downregulated in the *mug4-2* line (adjust P value<0.05; log2 fold change>0.5 and <-0.5). To obtain the functional information of these genes, I searched for gene ontology (GO) terms using agriGO web-based tool and database. Eighty-seven significant GO terms are found (Statistical test method: Fisher; Multi-test adjustment method: Yekutieli (FDR under dependency); Significance Level: 0.05; Minimum number of mapping entries: 5). The most significant GO terms are photosynthesis, response to stimulus and response to stress (Supplemental Figure 2.1).

mug1 mug2 mug4 mutants showed more severe defects compared to _mug1 mug2_ double mutants.

Previously, it was shown that the *mug1 mug2* double mutants exhibit strong phenotypes for traits associated with plant fitness (Joly-Lopez, Forczek et al. 2012). To understand the functional significance of all MUG-A proteins to plant development, I compared the phenotypic traits of the *MUG-A* gene single mutants, double and triple mutant combinations at different developmental stages, such as rosette diameter, flowering time, and primary stem height, according to (Joly-Lopez, Hoen et al. 2016). Consistent with the previous observation, all the single mutants of MUG-A genes did not show a significant difference compared to Col-0. *mug1 mug2* shows the most severe defects compared to other double mutants. Interestingly, the *mug1 mug2 mug4* triple mutants grew similarly to the *mug1 mug2 mug3* reported before (Joly-Lopez, Hoen et al. 2016) (Figure 2.5b). I observed that the *mug1 mug2 mug4* plants have even more severer defects than

mug1 mug2 double mutants: chlorotic seedlings, smaller rosette (Figure 2.5a), delayed first leave emergence and flowering time (Figure 2.5c), and extremely short primary stem (Figure 2.5c).

Discussion

MUG-A proteins are active transcriptional repressors.

ETEs evolved novel functions from ancient transposase genes. *MUG-A* is such an ETE family, the strong phenotype of *mug1 mug2* double mutants led me to explore the underlying molecular mechanism of how *MUG-A* ETEs benefit plant fitness. The MUG-A protein domain architecture is still conserved among the ancient transposase MURA and ETE FAR1 (Figure 2.2). Thus, I investigated the molecular function of the MUG-A proteins to examine whether they have a conserved function or develop novel functions. Conserved as the ancient transposase MURA and reported ETE FAR1, the MUG-A proteins are still located in the nucleus (Figure 2.1). However, as MUG-A proteins lost their TIRs, they are not functional TEs. FAR1/FHY3 ETE family are reported as transcriptional activators regulating phyA signalling network. Unlike ETE FAR1, MUG-A proteins do not have transcriptional activation ability in yeast or plants (Figure 2.3). Surprisingly, the MUG-A proteins suppress the expression of the reporter in a tobacco transient transfection system (Figure 2.3), suggesting they are active repressors.

While they share the same domain architecture, why do MUG-A proteins work as transcriptional repressors instead of transcriptional activity activators? One possibility is that MUG-A proteins did not gain or lose the active site that is essential for transcriptional activation activity. The previous report demonstrates that the mutagenesis of D283, D288, G305, C579, and H591 aborted the transcriptional activation activity of FHY3 in yeast cells (Lin, Teng et al. 2008). I

aligned the middle and C-terminal domains of FHY3 and MUG-A proteins. The alignment demonstrates that the corresponding sites of D283 in FHY3 are P375, P373, P354, P214 in MUG1, MUG2, MUG3, and MUG4, respectively. The G305 of FHY3 remains conservative in MUG1, MUG2 and MUG3, but turns into Alanine in MUG4. The C579 and H591 remain conserved among MUG-A proteins (Figure 2.6). This result suggests that the different amino acids corresponding to the position D283 in FHY3, the amino acid proline in MUG-A proteins disabled the transcriptional activation activity. Another possibility is that MUG-A proteins interact with other proteins that suppressed their transcriptional activation activity. A few examples support this assumption, for example, CCA1, LHY, and PIF1 interact with FHY3 and suppress the transcriptional activation of FHY3 on ELF4 and HEMB1; HY5 interacts with FHY3 and suppresses the transcriptional activation of FHY3 on FHY1 and PHR1 (Ma and Li 2018). Thus, like the way how FHY3 disables its transcriptional activation activity, some proteins may interact with MUG-A proteins that suppress their function.

MUG-A proteins may work as a transcriptional repressor complex.

As described above, MUG-A proteins could form homodimers as well as heterodimers with each other (Figure 2.4). This result leads me to ask whether the homodimer or heterodimer are endogenous physically existing or because the domain structure of the MUG-A proteins remains the same, it allows them to interact with each other. Comparing the expression profile of *MUG-A* genes in eFP browser (http://bar.utoronto.ca/efp2/Arabidopsis/Arabidopsis_eFPBrowser2.html), I found that MUG-A proteins share a similar expression pattern (Supplemental Figure 2.2). Furthermore, at the cell level, they all locate to the nucleus. In addition, all of them have transcriptional repressor ability. Commonly, transcriptional repressors form complexes, such as

FRS7 and FRS12, which are the members of FAR1 RELATED SEQUENCE (FRS) gene family. Unlike FAR1, these two proteins form a transcriptional repressor complex regulating flowering time and growth in *Arabidopsis* (Ritter, Inigo et al. 2017). Another example is ASR3, which is a transcriptional repressor, it forms a homodimer complex and negatively regulates plant innate immunity and immune gene expression (Li, Jiang et al. 2015). Additional evidence is that in the Y2H assay, failed to detect MUG4 interactions with MUG3 and MUG4 (Figure 2.4), suggesting that the complex of MUG3 and MUG4 is an active repressor, which is strong enough to suppress the expression of the reporter genes. Taken together, the MUG-A proteins probably work as a complex to repress the expression of target genes.

MUG-A proteins have an overlapping function.

A few TFs have redundant functions in regulating plant growth and development. The above-mentioned FAR1 and FHY3 have redundant roles in controlling adult plant growth as suggested by the observation that the *fhy3-4* and *far1-2* single mutant plants did not differ much from the Col.0 wild-type but *fhy3 far1* double mutant did (Wang, Tang et al. 2016). Also, in the Aux/IAA protein family, 13 different aux/iaa loss-of-function mutants do not show any obvious morphological phenotypes, even more, the *iaa5 iaa6 iaa19* triple mutant is still very similar to wild-type, indicating the functional redundancy among members of Aux/IAA family (Wang, Chang et al. 2007). In the *MUG-A* gene family, all the MUG-A single mutants do not show any obvious phenotype during plant development stage. However, the double mutant of *mug1 mug2* shows pleiotropic traits. Furthermore, I observed an additive effect on the triple mutants of *mug1 mug2 mug3* and *mug1 mug2 mug4,* which show a smaller rosette, delayed flowering time, shorter stem height compared to *mug1 mug2* double mutant (Figure 2.5). This result suggests MUG-A

proteins may have an overlapping function. Such functional redundancy among *MUG-A* genes may be required to achieve maximal regulatory ability of the transcriptional repression complex.

FIGURES AND TABLES

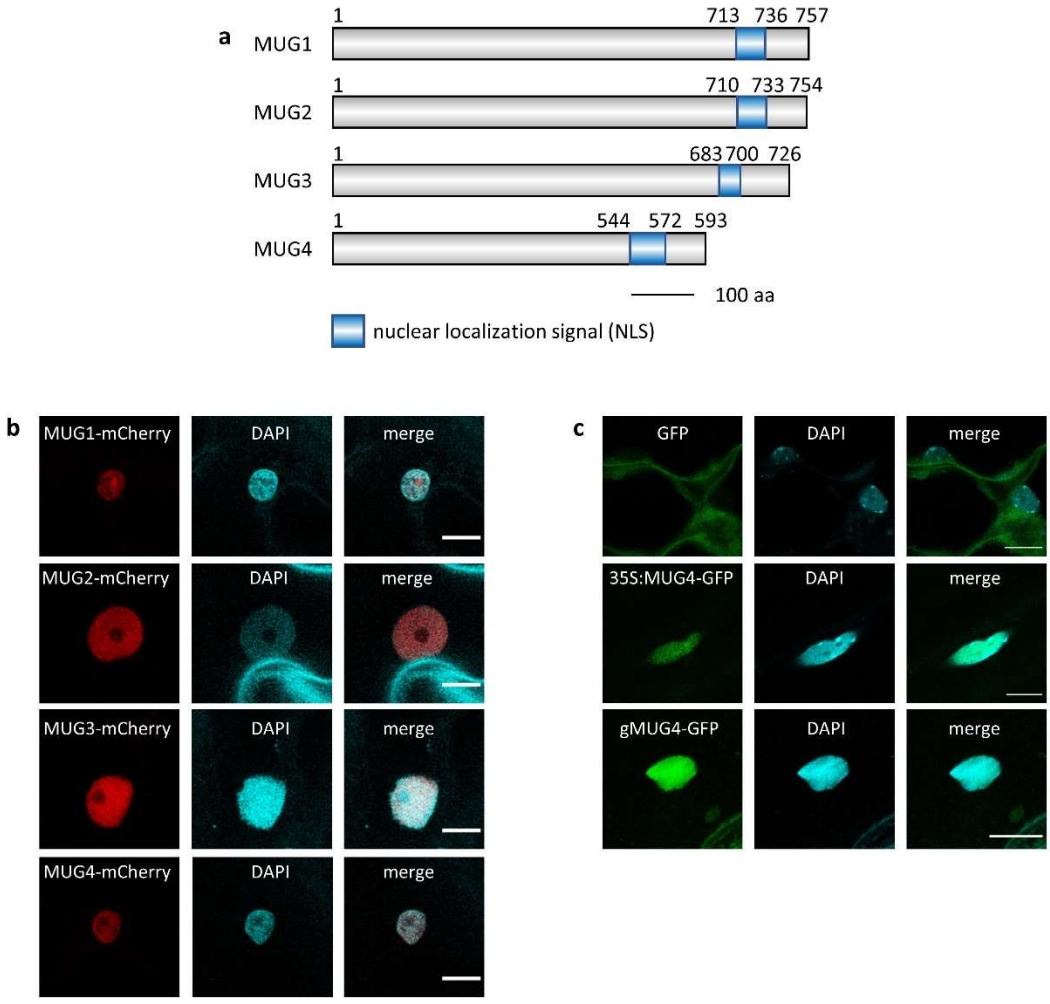
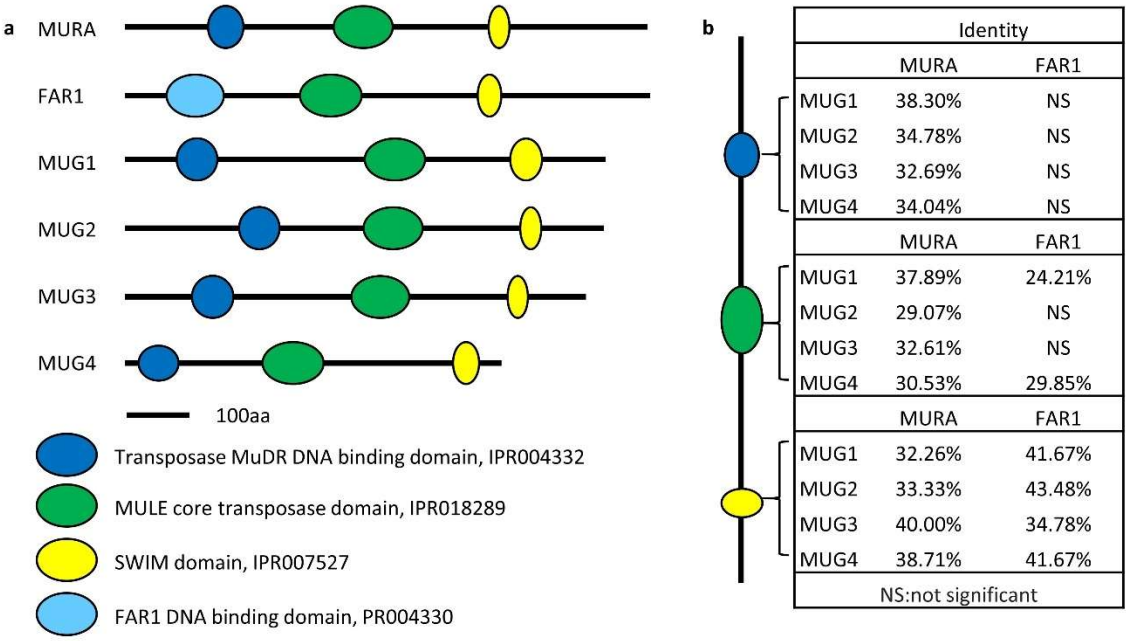


Figure 2.1. MUGA proteins localized to the plant nucleus **a** MUG-A proteins are all predicted to have a signal peptide by the LOCALIZER program. MUG1: 713–736 aa; MUG2: 710–733 aa; MUG3: 683–700 aa; MUG4: 544–572 aa. **b** Leaf tissues of *N. benthamiana* transiently expressing the MUG1: mCherry, MUG2: mCherry, MUG3: mCherry and MUG4: mCherry were stained with DAPI for nuclei location reference and examined by confocal microscopy. Bar = 10 μm. **c** Root cells of *A. thaliana* stable expressing GFP alone (first row) and T3 individual lines

expressing 35s: MUG4: GFP (second row) and pMUG4: MUG4: GFP (third row) were stained with DAPI for nuclei location reference and examined by confocal microscopy. Bar = 10 μm.



Identity		
	MURA	FAR1
MUG1	38.30%	NS
MUG2	34.78%	NS
MUG3	32.69%	NS
MUG4	34.04%	NS
	MURA	FAR1
MUG1	37.89%	24.21%
MUG2	29.07%	NS
MUG3	32.61%	NS
MUG4	30.53%	29.85%
	MURA	FAR1
MUG1	32.26%	41.67%
MUG2	33.33%	43.48%
MUG3	40.00%	34.78%
MUG4	38.71%	41.67%
NS:not significant		

Figure 2.2. Domain analysis of MUG-A proteins comparing with MURA and FAR1. **a** Graphical representation with domain predictions, the MURA, MUG1, MUG2, MUG3, and MUG4 are all predicted to share three conserved domains: An N-terminal C2H2-type zinc-chelating transposase MuDR domain (IPR004332), a central MULE putative core transposase domain (IPR018289), and a C-terminal SWIM domain (IPR007527). For FAR1, while the middle and C-terminal domain remain conserved, the N-terminal domain evolved into its own FAR1 DNA binding domain, PR004330. **b** domain identity calculation of MUG1, MUG2, MUG3, and MUG4 against MURA and FAR1 using blastp.

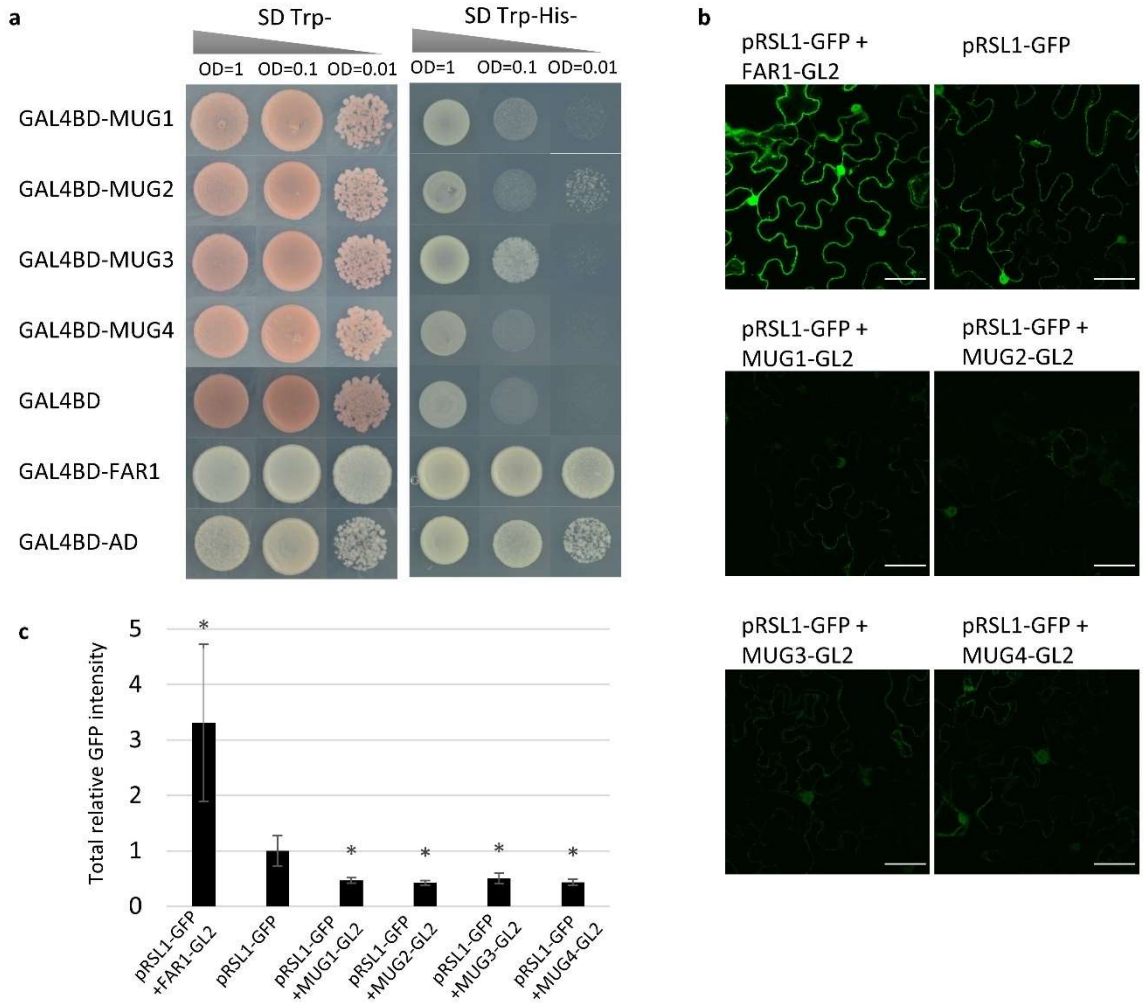
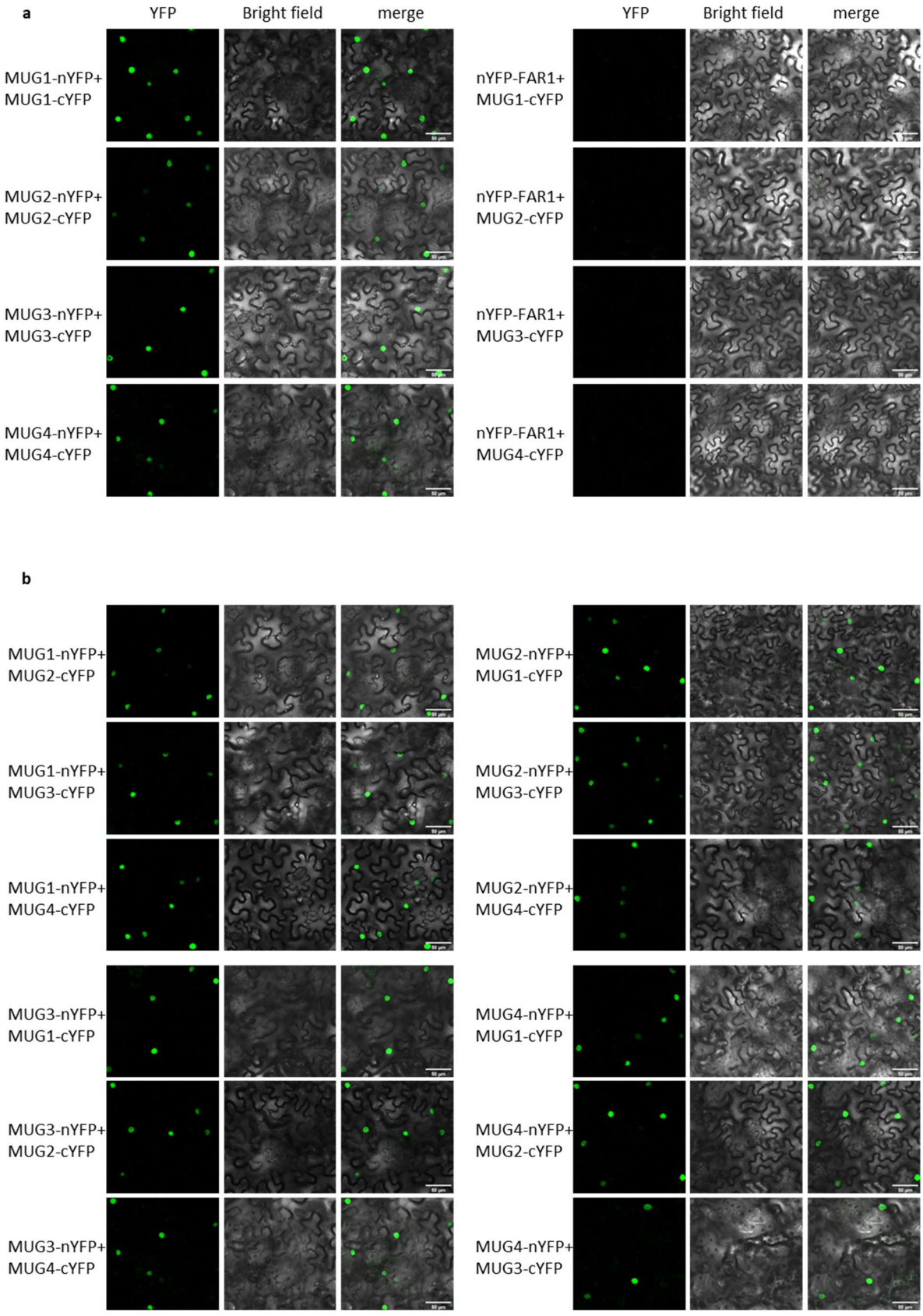


Figure 2.3. Transcriptional regulatory ability assay among MUG-A proteins. **a** Transcriptional activation assay in yeast**.** On the control plates SD Trp-, all the yeast strain can grow. The yeast strain expressing GAL4BD linked with MUG1 (GAL4BD-MUG1), GAL4BD linked with MUG2 (GAL4BD-MUG2), GAL4BD linked with MUG3 (GAL4BD-MUG3), GAL4BD linked with MUG4 (GAL4BD-MUG4) and GAL4BD alone turned red color while the ones expressing GAL4BD linked with FAR1 (GAL4BD-FAR1) and GAL4BD linked with GAL4AD (GAL4BD-AD) stay white color on the control plate. The yeast cells transformed with GALBD only and

GAL4BD-MUG4 cannot grow on SD Trp- His-. The yeast strains expressing GAL4BD-FAR1 and GAL4BD-AD can grow on SD Trp- His-. **b** Transcriptional regulatory assay in plant. *N. benthamiana* leaf epidermis transfected with the RSL1 promoter-driven GFP gene (pRSL1-GFP) work as the basal control. GL2 DNA binding domain can bind to pRSL1. The leaves co-transformed with pRSL1-GFP and 35Spro-FAR1-GL2 (pRSL1-GFP + FAR1-GL2) is set to the positive control. The leaves co-transformed with pRSL1-GFP and 35Spro-MUG1- GL2 (pRSL1-GFP + MUG1-GL2), 35Spro-MUG2-GL2 (pRSL1-GFP + MUG1-GL2), 35Spro-MUG3-GL2 (pRSL1-GFP + MUG3-GL2), 35Spro-MUG4-GL2 (pRSL1-GFP + MUG4-GL2) were tested. Representative fluorescence images of *N. benthamiana* leaf epidermis described above are shown. The structures of the genes are illustrated on the top of each image. Bars = 50 μm. **c** Quantification of the intensities of the GFP fluorescence signal. Relative intensity of the signal was calculated with the mean value for pRSL1-GFP intensity set as 1 (mean ± sd, n = 10). Asterisks indicate that the values are significantly different from pRSL1-GFP values (*P < 0.001, Student's t test).


a
YFP
Bright field
merge
MUG1-nYFP+
MUG1-cYFP
MUG2-nYFP+
MUG2-cYFP
MUG3-nYFP+
MUG3-cYFP
MUG4-nYFP+
MUG4-cYFP
YFP
Bright field
merge
nYFP-FAR1+
MUG1-cYFP
nYFP-FAR1+
MUG2-cYFP
nYFP-FAR1+
MUG3-cYFP
nYFP-FAR1+
MUG4-cYFP
b
MUG1-nYFP+
MUG2-cYFP
MUG1-nYFP+
MUG3-cYFP
MUG1-nYFP+
MUG4-cYFP
MUG2-nYFP+
MUG1-cYFP
MUG2-nYFP+
MUG3-cYFP
MUG2-nYFP+
MUG4-cYFP
MUG3-nYFP+
MUG1-cYFP
MUG3-nYFP+
MUG2-cYFP
MUG3-nYFP+
MUG4-cYFP
MUG4-nYFP+
MUG1-cYFP
MUG4-nYFP+
MUG2-cYFP
MUG4-nYFP+
MUG3-cYFP

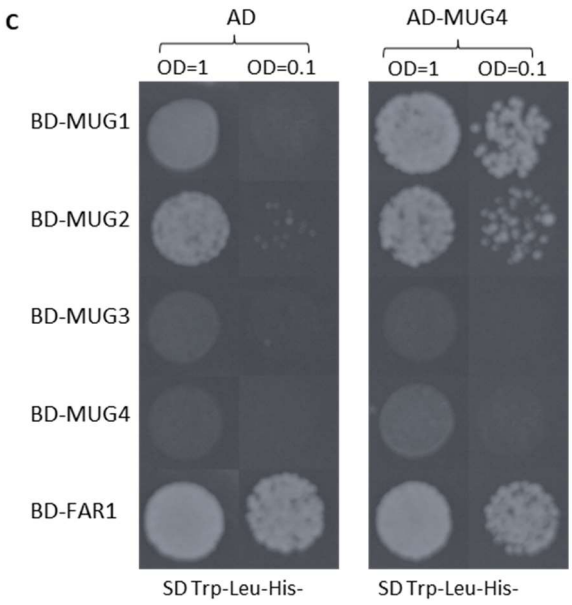


Figure 2.4. Determination of the homodimer or heterodimer formation among MUG-A proteins. **a** Confocal image of MUG1, MUG2, MUG3 and MUG4 self-interacts using BiFC, left panel. Negative control in the BiFC system with the nYFP-FAR1 construction. No BiFC signals were observed between FAR1 and MUG1, MUG2, MUG3 or MUG4 (right panel). **b** Pairwise BiFC experiments between MUG1, MUG2, MUG3 and MUG4. Protein partners were fused to an N-terminal fragment or C-terminal fragment of YFP, respectively, and co-infiltrated into tobacco leaves. BiFC signals between MUG1, MUG2, MUG3 and MUG4 were observed in nuclear (green). Bars = 50 μm**. c** Y2H confirmation of MUG4 self-interaction and its interaction with MUG1 and MUG2. MUG1-GAL4BD and MUG2-GAL4BD co-transformed with GAl4AD yeast strains could also grow a little bit on SD Trp-Leu-His (left panel). MUG1-GAL4BD, MUG2-GALBD co-transformed with MUG4-GAL4 AD yeast strains grow more (right panel). The results show MUG1 and MUG4 forms heterodimer. The combination of MUG3-GAL4BD and MUG4-GAL4AD, MUG4-GAL4BD and MUG4-GAL4AD failed to active the reporter genes and the corresponding yeast strains cannot grow on SD Trp-Leu-His.

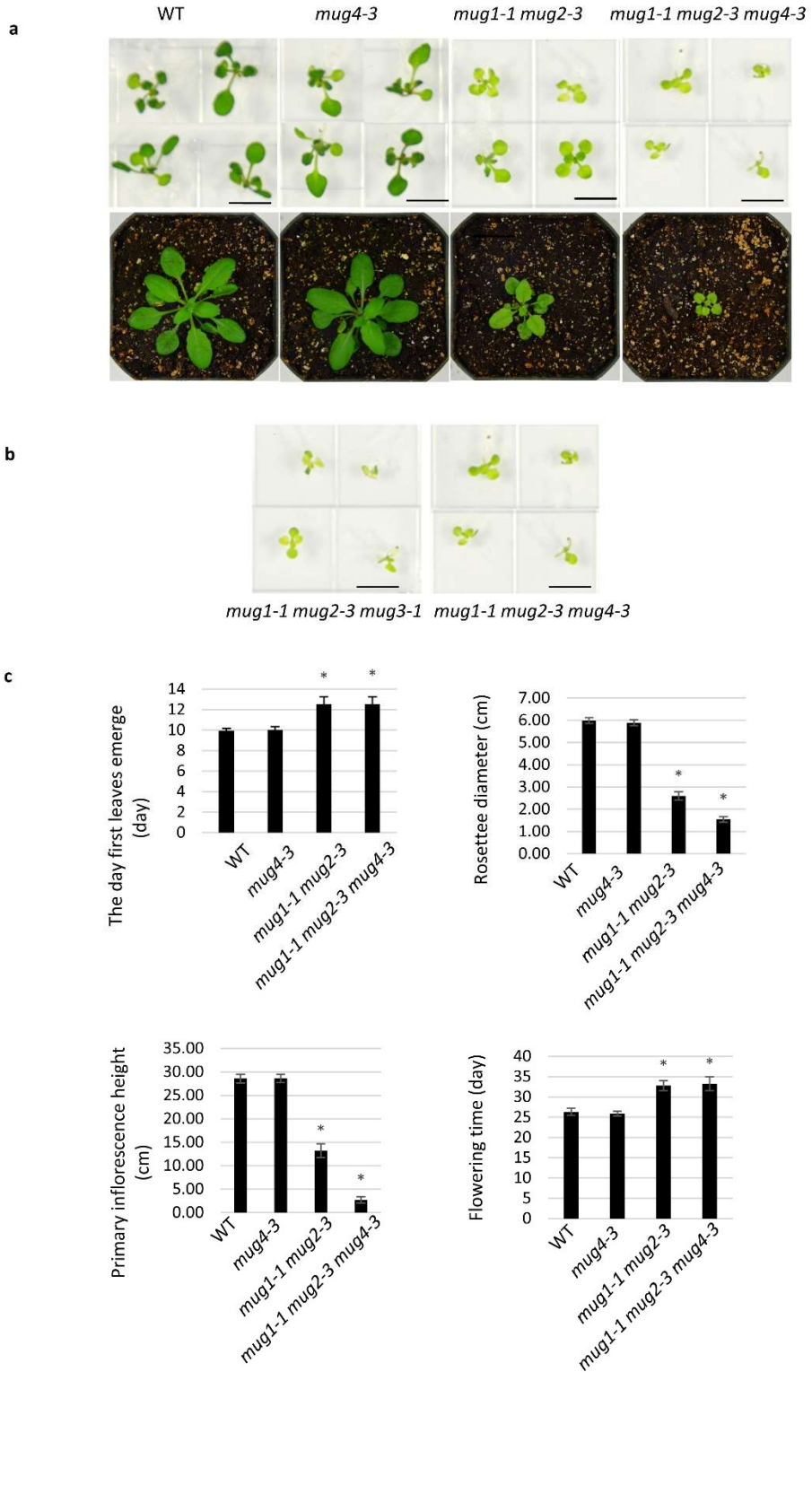

a
WT
mug4-3
mug1-1 mug2-3
mug1-1 mug2-3 mug4-3
b
mug1-1 mug2-3 mug3-1
mug1-1 mug2-3 mug4-3
c
The day first leaves emerge (day)
Rosettee diameter (cm)
Primary inflorescence height (cm)
Flowering time (day)
WT
mug4-3
mug1-1 mug2-3
mug1-1 mug2-3 mug4-3

Figure 2.5. *mug1 mug2 mug4* has additive phenotype to *mug1 mug2* double mutant. **a** the 15-day seedlings and 23-day plants in pots of Col-0*, mug4-3, mug1-1 mug2-3, and mug1-1 mug2-3 mug4-3*. **b** the 15-day seedlings of *mug1-1 mug2-3 mug3-1* and *mug1-1 mug2-3 mug4-3*, scale bar=0.5 cm. **c** the quantification of the day first leaves emerges, the rosette diameter, the primary inflorescence height and flowering time of Col-0*, mug4-3, mug1-1 mug2-3, and mug1-1 mug2-3 mug4-3* (n = 16). Error bars represent ±SE, P-values: P < 0.001. * indicates significant difference from the Col-0.

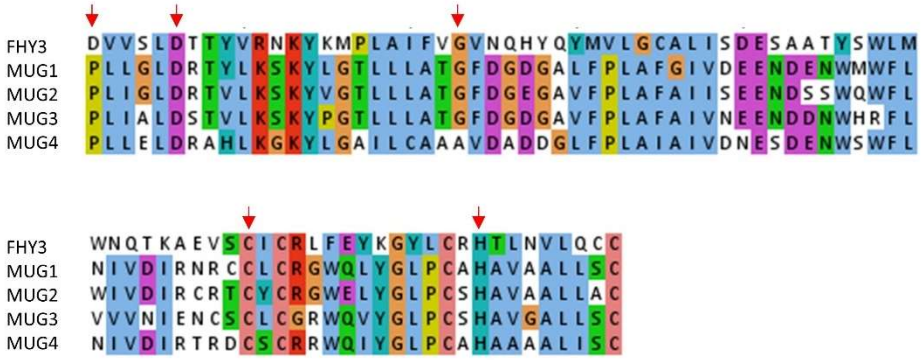
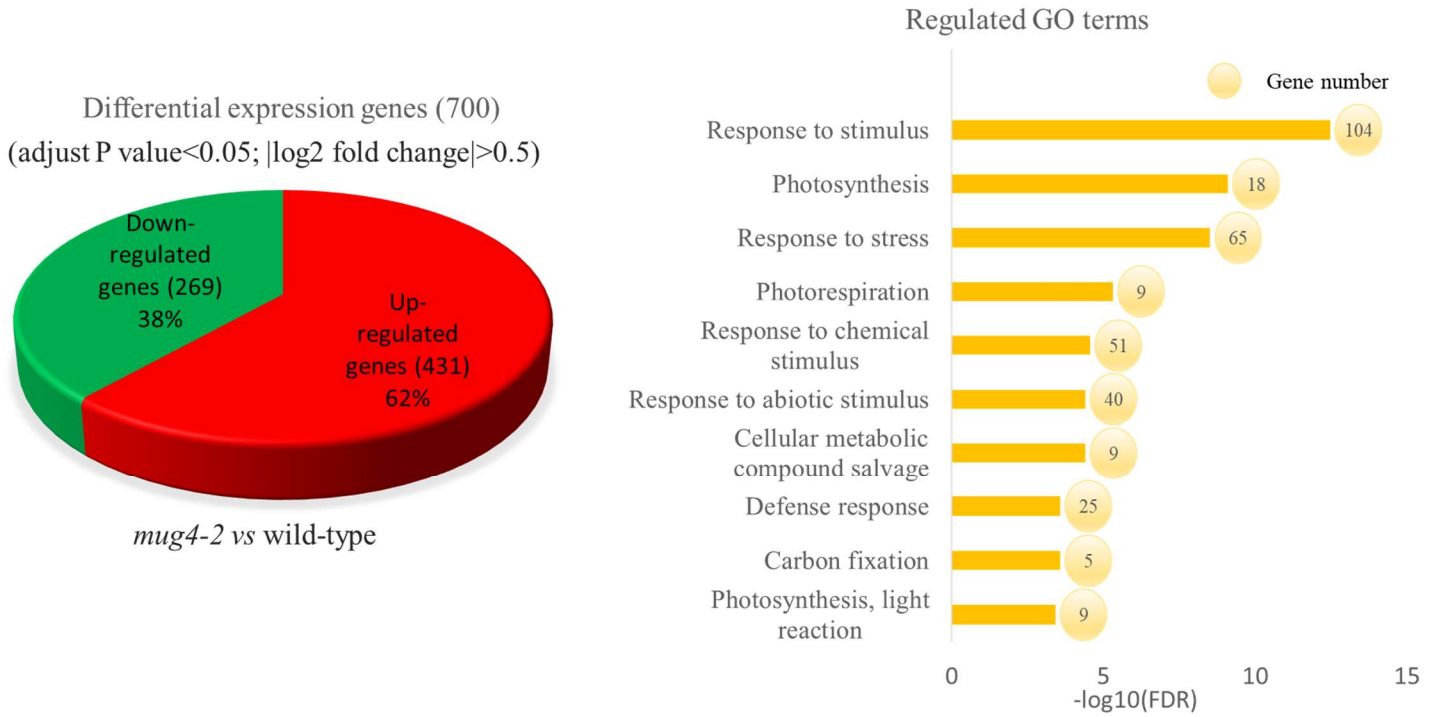


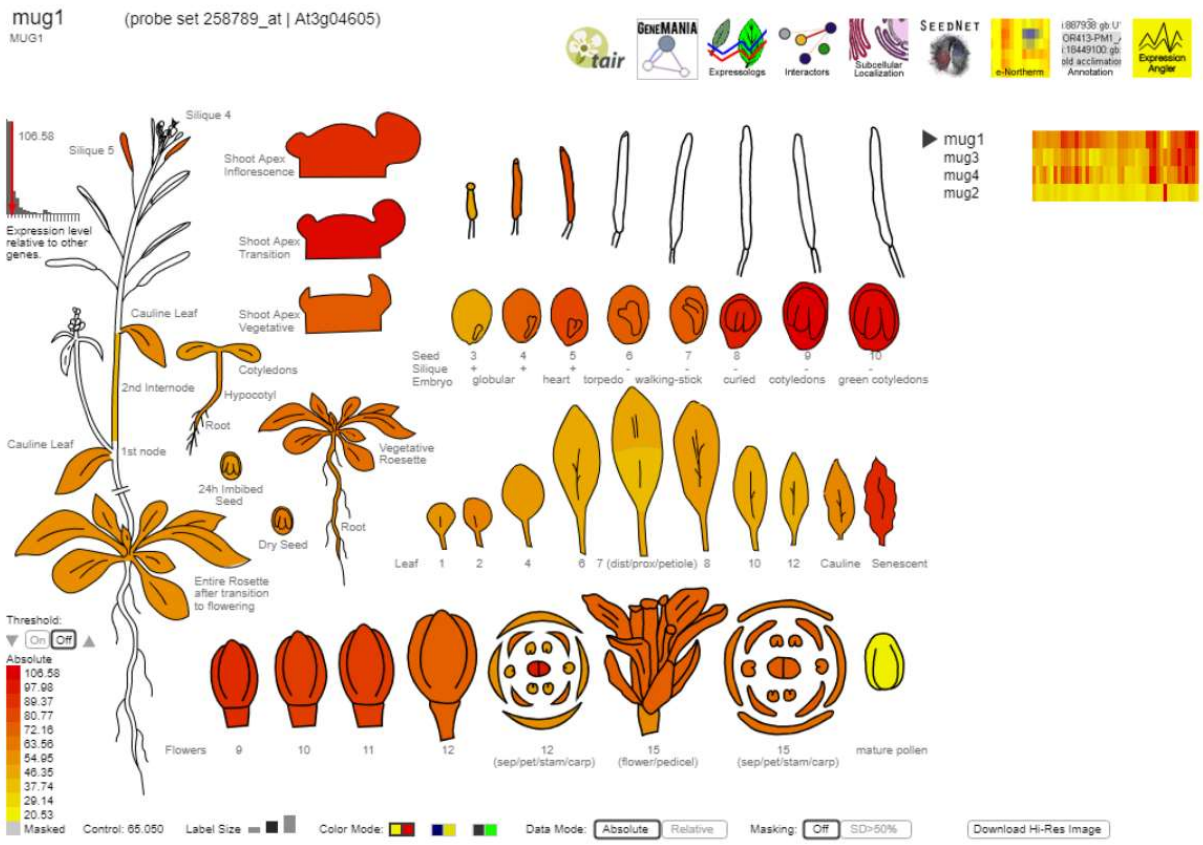
Figure 2.6. The alignment among FHY3 and MUG-A proteins of the middle MULE domain and C-terminal Swim domain. Up panel, MULE domain. Down panel, SWIM domain. Red arrows indicate the sites that are essential for the transcriptional activation activity of FHY3 (D283, D288, G305, C579, and H591) and corresponding amino acid in MUG-A proteins.

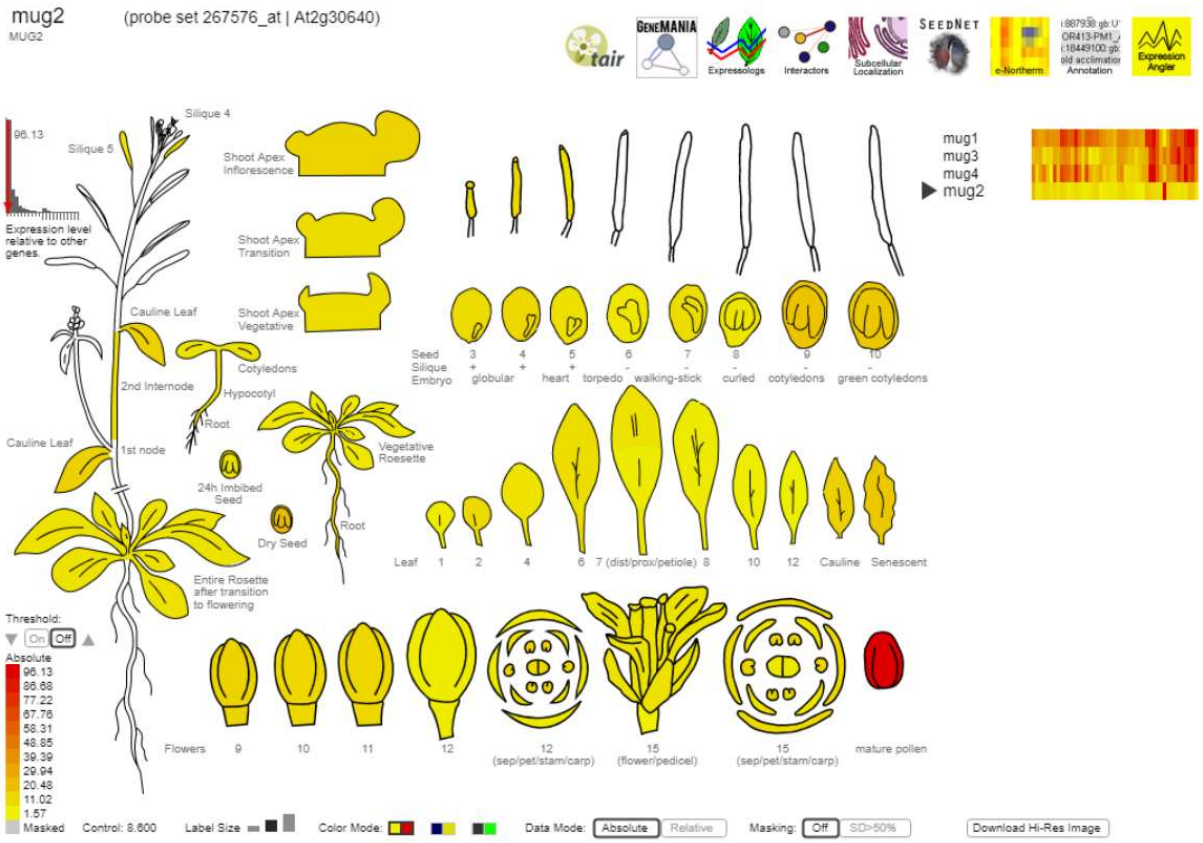
Primer name	Sequence	Usage
mug1-1 LP	TGA CGA TGC ACA AGA TTA AG	genotyping
mug1-1 RP	TGA CCC AAG TGA TTT ATT CC	genotyping
mug2-3 LP	TGA GAC AAT GGA CAC ACA AGC	genotyping
mug2-3 RP	GTG GGT CAT TGC ATG AAT AGG	genotyping
mug4-3 LP	ATC CCG TGT CCG TAT ATC CA	genotyping
mug4-3 RP	CGG TTG TTT ATT GCG TAT CG	genotyping
lbb3.3	ATTTTGCCGATTTCGGAAC	genotyping
XbaI-gMUG4-FP	AATCTAGAAGCAAAAGGAGAGAGAAAGAAAA	mug4pro:MUG4 entry vector construction
XbaI-MUG4-FP	AAAtctagaATGGCTGATGGAGCTCTAATTACTTTAG	35s:MUG4 entry vector construction
BamHI-MUG4-RP	AAggatccTCAAATGGGCTGTGTGCATTTCTTC	MUG4 entry vector construction
knp1-mCherry	Aaaggtaccttacttgtacagctcgtccatgc	pDOE-04-mCherry vector construction
mcherryrev	Atggtgagcaagggcgaggag	pDOE-04-mCherry vector construction
S-B-VP16-FP	AAAgtcgacggatccATGgacgcgctagacgatttcgatc	pDOE-04-mCherry-VP16-GL2-V2 vector construction
GL2-R	TCAGCAATCTTCGATTTGTAGACTTCTC	pDOE-04-mCherry-VP16-GL2-V2 vector construction
BamHI-MUG4	AAAggatccATGGCTGATGGAGCTCTAATTACTTTA	pDOE-04-mCherry-MUG4-GL2-V2 vector construction
c-MUG4-RP	ggAATGGGCTGTGTGCATTTCTTCTG	pDOE-04-mCherry-MUG4-GL2-V2 vector construction
BamHI-FAR1	AAAggatccATGGATTTGCAAGAGAATCTGGTTAG	pDOE-04-mCherry-FAR1-GL2-V2 vector construction
c-FAR1-RP	ggTAGCTGCCTTGATGAACTACCAGG	pDOE-04-mCherry-FAR1-GL2-V2 vector construction
V2-M1-FP	gttcatttcatttggagaggacacgctcgacgATGGCCAATCGCGATTTGATGC	pDOE-04-mCherry-MUG1-GL2-V2 vector construction
V2-M1-RP	cggagtcgtgggggtaaatcccggacccggAATAGGTGCAGCACAAGTGGTTCTG	pDOE-04-mCherry-MUG1-GL2-V2 vector construction
V2-M2-FP	gttcatttcatttggagaggacacgctcgacgATGACAAATAACGAGCTGATTATGTCG	pDOE-04-mCherry-MUG2-GL2-V2 vector construction
V2-M2-RP	gcggagtcgtgggggtaaatcccggacccggTATAGGAGCTGTGCAAGTTGTTCTG	pDOE-04-mCherry-MUG2-GL2-V2 vector construction
V2-M3-FP	gttcatttcatttggagaggacacgctcgacgATGGCAAACGACGAGCTTGTG	pDOE-04-mCherry-MUG3-GL2-V2 vector construction
V2-M3-RP	gcggagtcgtgggggtaaatcccggacccggCATAGGAGCAGTGCAAGTTGTTCTG	pDOE-04-mCherry-MUG3-GL2-V2 vector construction

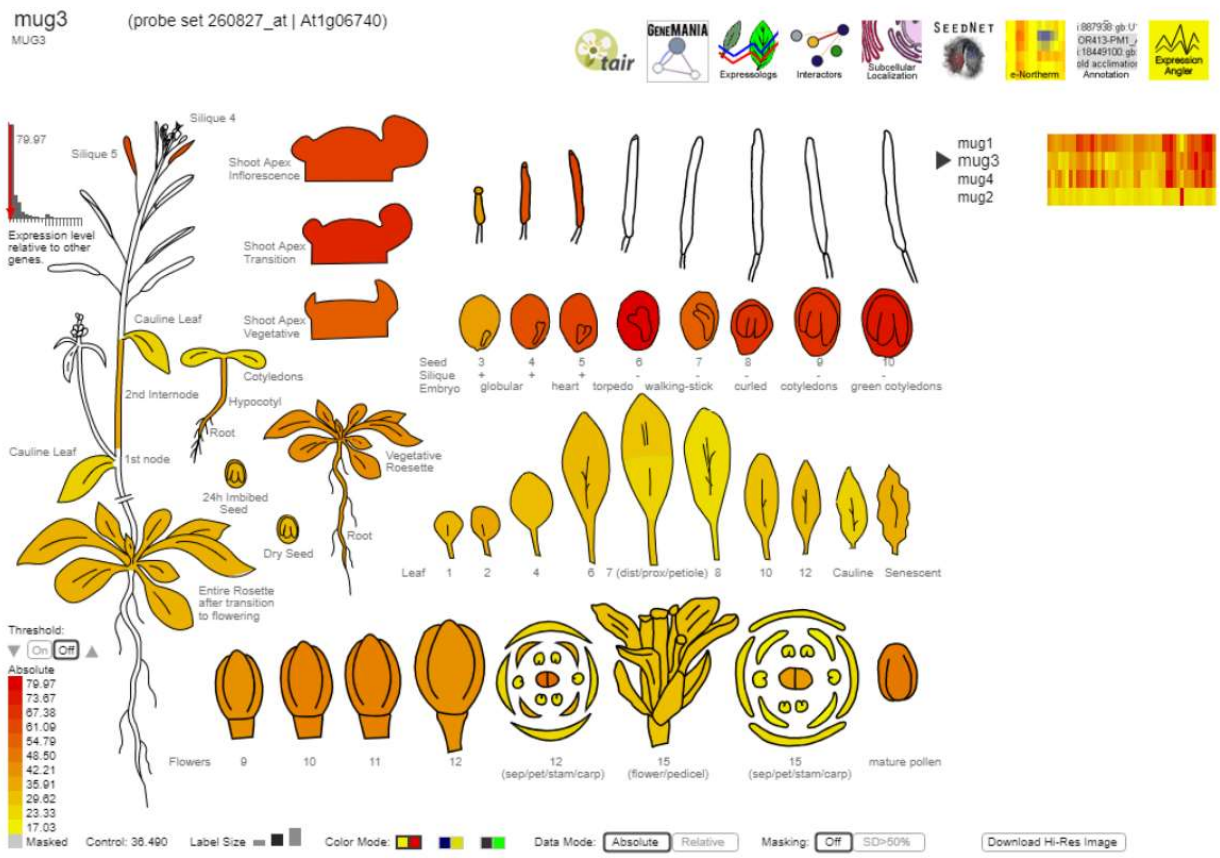
Supplemental Table 2.1. The primer information used in this paper.

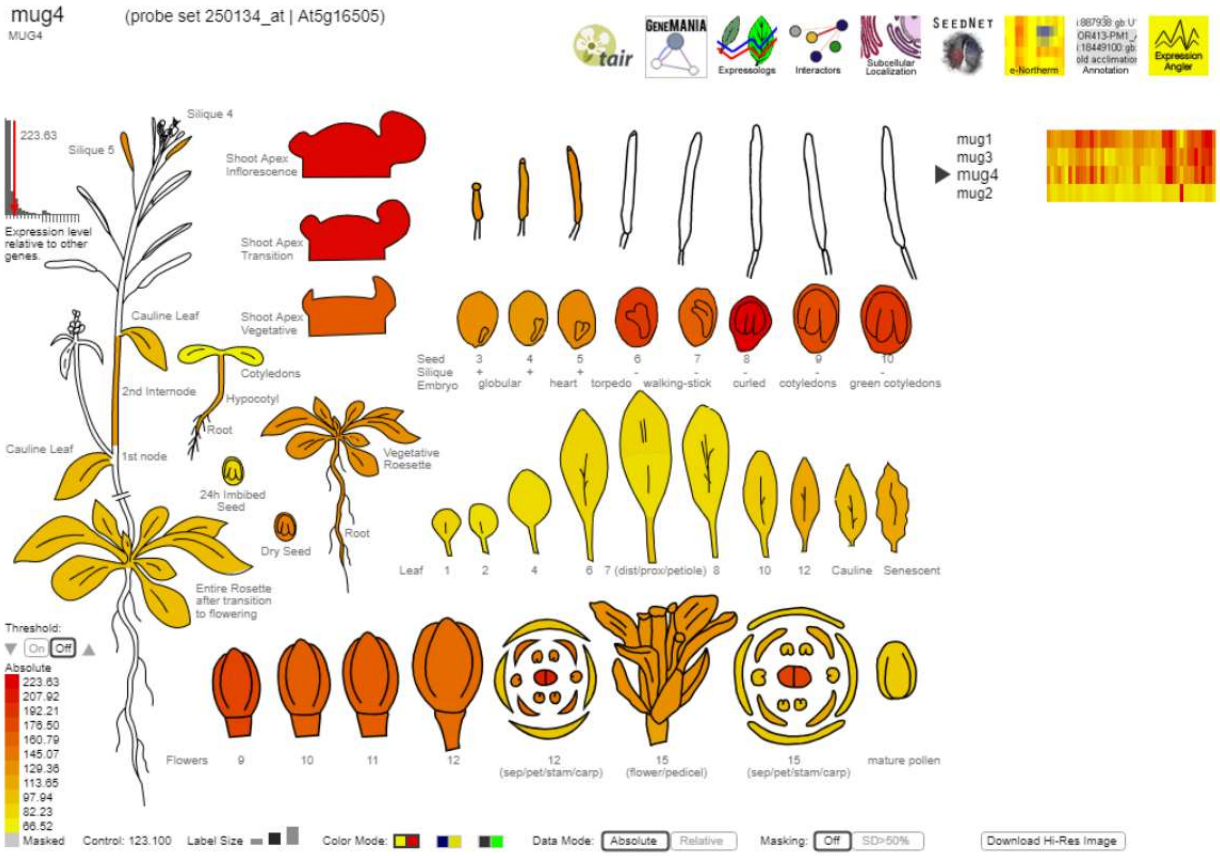


Supplemental Figure 2.1. The regulated genes distribution and GO enrichment of regulated genes in *mug4-2*.

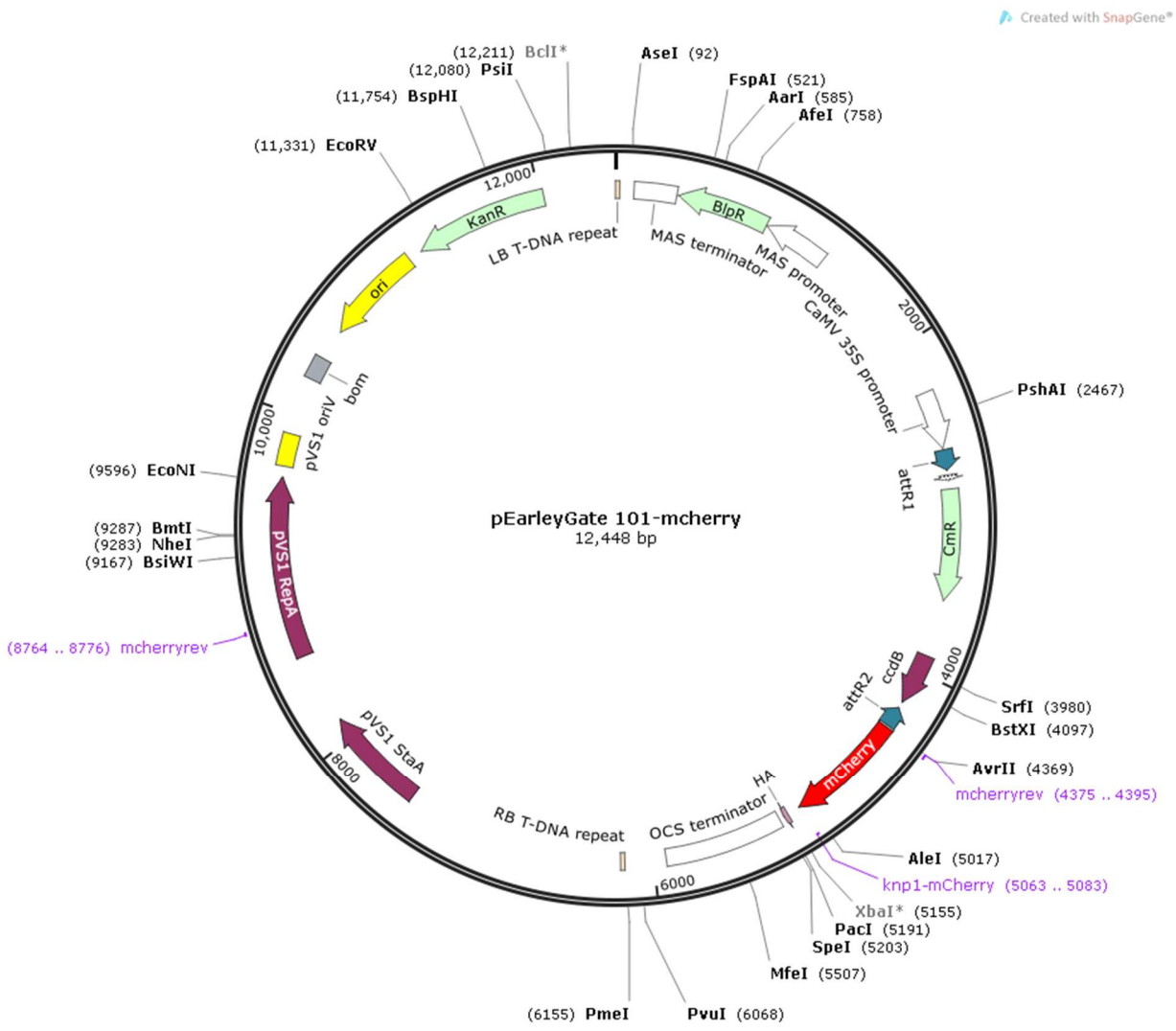

mug1
MUG1
(probe set 258789_at | At3g04605)
tair
GeneMANIA
Expressologs
Interactors
Subcellular Localization
SeedNet
e-Northern
Expression Angler
106.58
Silique 5
Silique 4
Expression level relative to other genes.
Shoot Apex Inflorescence
Shoot Apex Transition
Shoot Apex Vegetative
Cauline Leaf
Cotyledons
2nd Internode
Hypocotyl
Root
Cauline Leaf
1st node
24h Imbibed Seed
Dry Seed
Vegetative Roesette
Root
Entire Rosette after transition to flowering
mug1
mug3
mug4
mug2
Seed
Silique
Embryo
3 + globular
4 + 5 + heart
6 - torpedo
7 - walking-stick
8 - curled
9 - cotyledons
10 - green cotyledons
Leaf 1 2 4 6 7 (dist/prox/petiole) 8 10 12 Cauline Senescent
Flowers 9 10 11 12
12 (sep/pet/stam/carp)
15 (flower/pedicel)
15 (sep/pet/stam/carp)
mature pollen
Threshold:
On Off
Absolute
106.58
97.98
89.37
80.77
72.16
63.56
54.95
46.35
37.74
29.14
20.53
Masked
Control: 65.050
Label Size
Color Mode:
Data Mode: Absolute Relative
Masking: Off SD>50%
Download Hi-Res Image

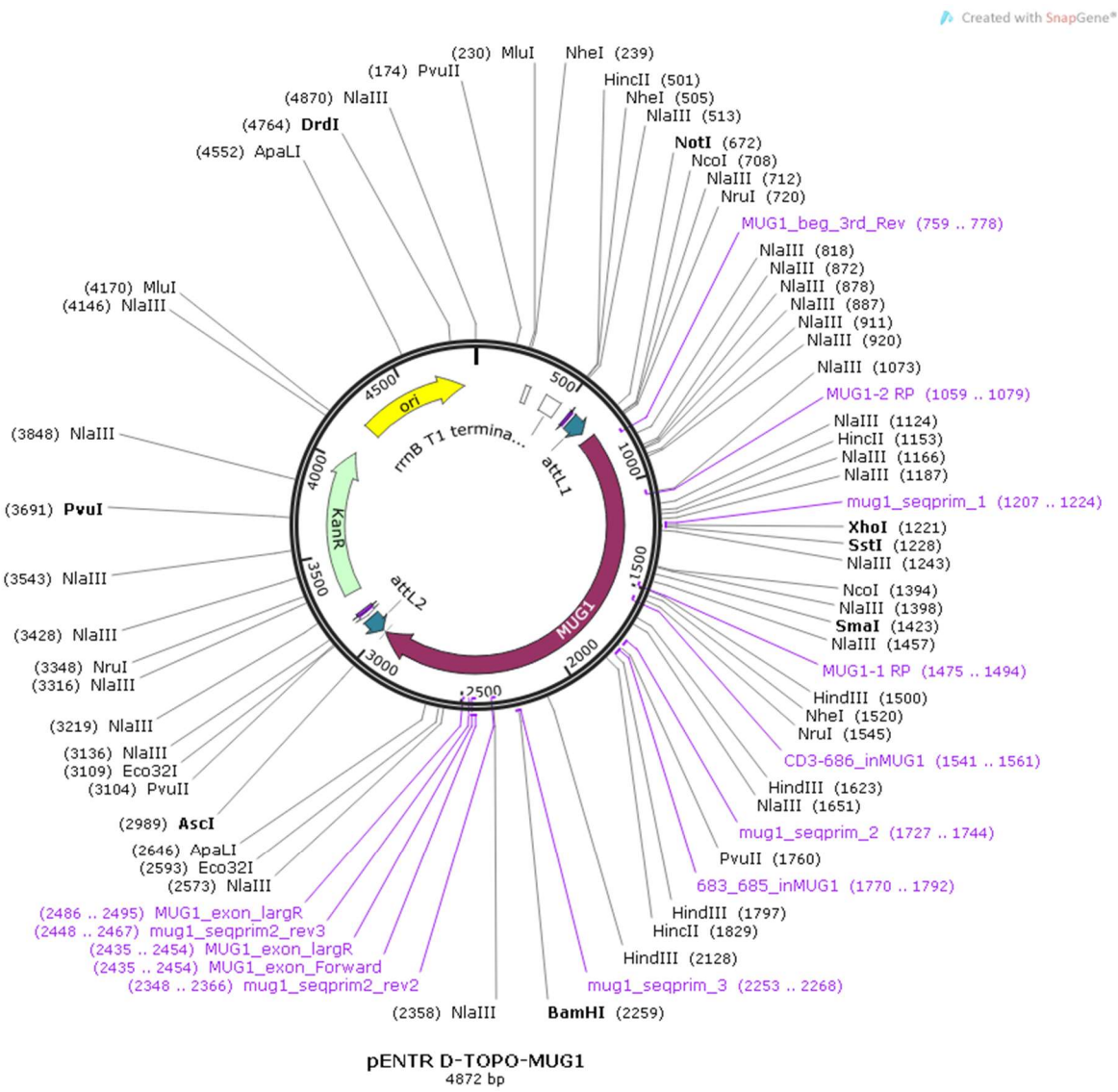

mug2
MUG2
(probe set 267576_at | At2g30640)
tair
GeneMANIA
Expressologs
Interactors
Subcellular Localization
SeedNet
e-Northern
Expression Angler
96.13
Expression level relative to other genes.
Silique 4
Silique 5
Shoot Apex Inflorescence
Shoot Apex Transition
Shoot Apex Vegetative
Cauline Leaf
2nd Internode
Cotyledons
Hypocotyl
Root
Cauline Leaf
1st node
24h Imbibed Seed
Dry Seed
Vegetative Roesette
Root
Entire Rosette after transition to flowering
mug1
mug3
mug4
mug2
Seed
Silique
Embryo
3 globular
4 heart
5 torpedo
6 walking-stick
7 curled
8 cotyledons
9 green cotyledons
10
Leaf 1 2 4 6 7 (dist/prox/petiole) 8 10 12 Cauline Senescent
Flowers 9 10 11 12 12 (sep/pet/stam/carp) 15 (flower/pedicel) 15 (sep/pet/stam/carp) mature pollen
Threshold:
On Off
Absolute
96.13
86.68
77.22
67.76
58.31
48.85
39.39
29.94
20.48
11.02
1.57
Masked
Control: 8.600
Label Size
Color Mode:
Data Mode: Absolute Relative
Masking: Off SD>50%
Download Hi-Res Image

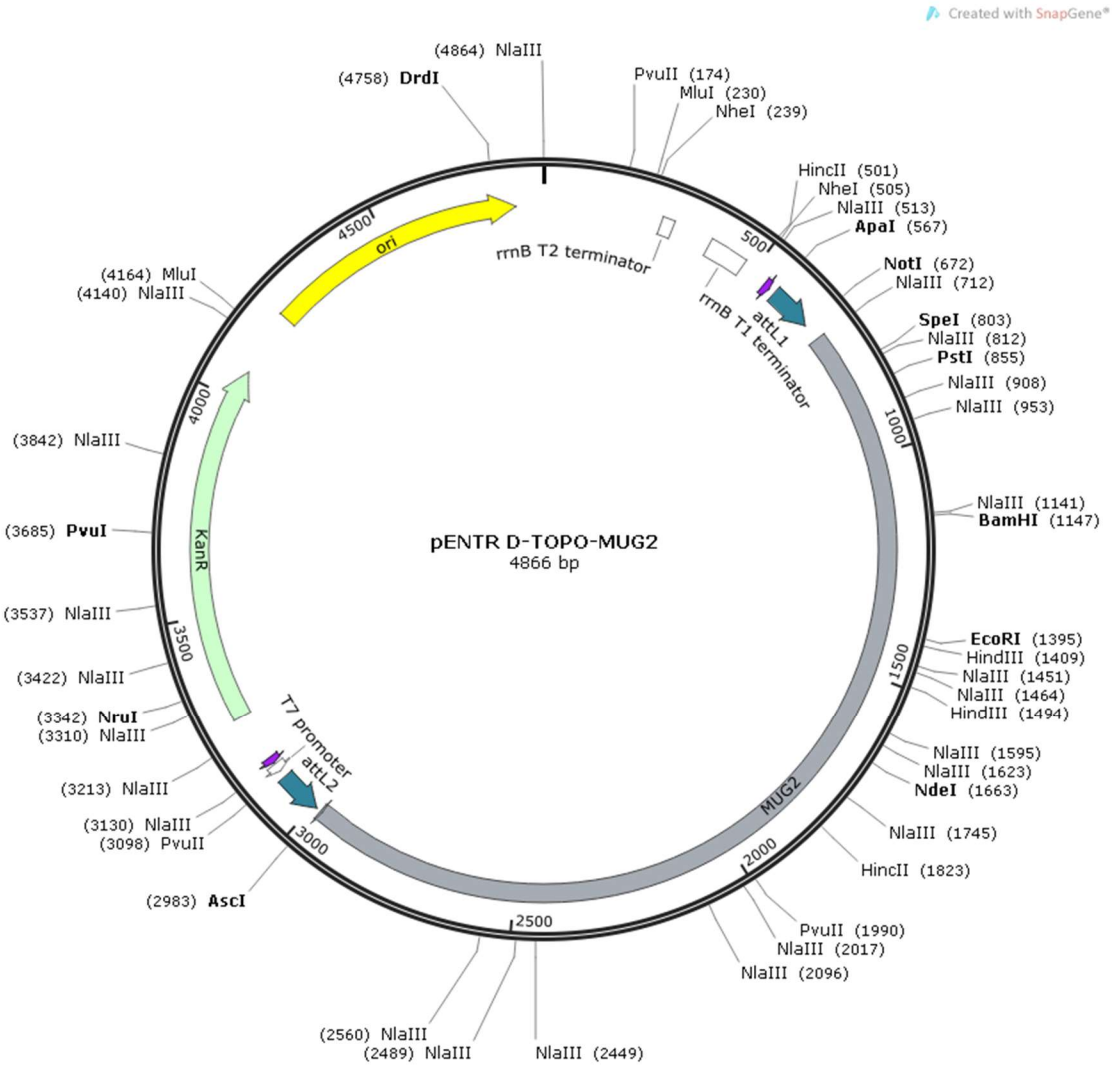

mug3
MUG3
(probe set 260827_at | At1g06740)
tair
GeneMANIA
Expressologs
Interactors
Subcellular Localization
SeedNet
e-Northern
Expression Angler
Silique 4
79.97
Silique 5
Expression level relative to other genes.
Shoot Apex Inflorescence
Shoot Apex Transition
Shoot Apex Vegetative
Cauline Leaf
Cotyledons
2nd Internode
Hypocotyl
Root
Cauline Leaf
1st node
24h Imbibed Seed
Dry Seed
Vegetative Roesette
Root
Entire Rosette after transition to flowering
mug1
mug3
mug4
mug2
Seed
Silique
Embryo
3 + globular
4 + heart
5 + torpedo
6 - walking-stick
7 - curled
8 - cotyledons
9 - green cotyledons
10 -
Leaf 1 2 4 6 7 (dist/prox/petiole) 8 10 12 Cauline Senescent
Flowers 9 10 11 12
12 (sep/pet/stam/carp)
15 (flower/pedicel)
15 (sep/pet/stam/carp)
mature pollen
Threshold:
On Off
Absolute
79.97
73.67
67.38
61.09
54.79
48.50
42.21
35.91
29.62
23.33
17.03
Masked
Control: 36.490
Label Size
Color Mode:
Data Mode: Absolute Relative
Masking: Off SD>50%
Download Hi-Res Image

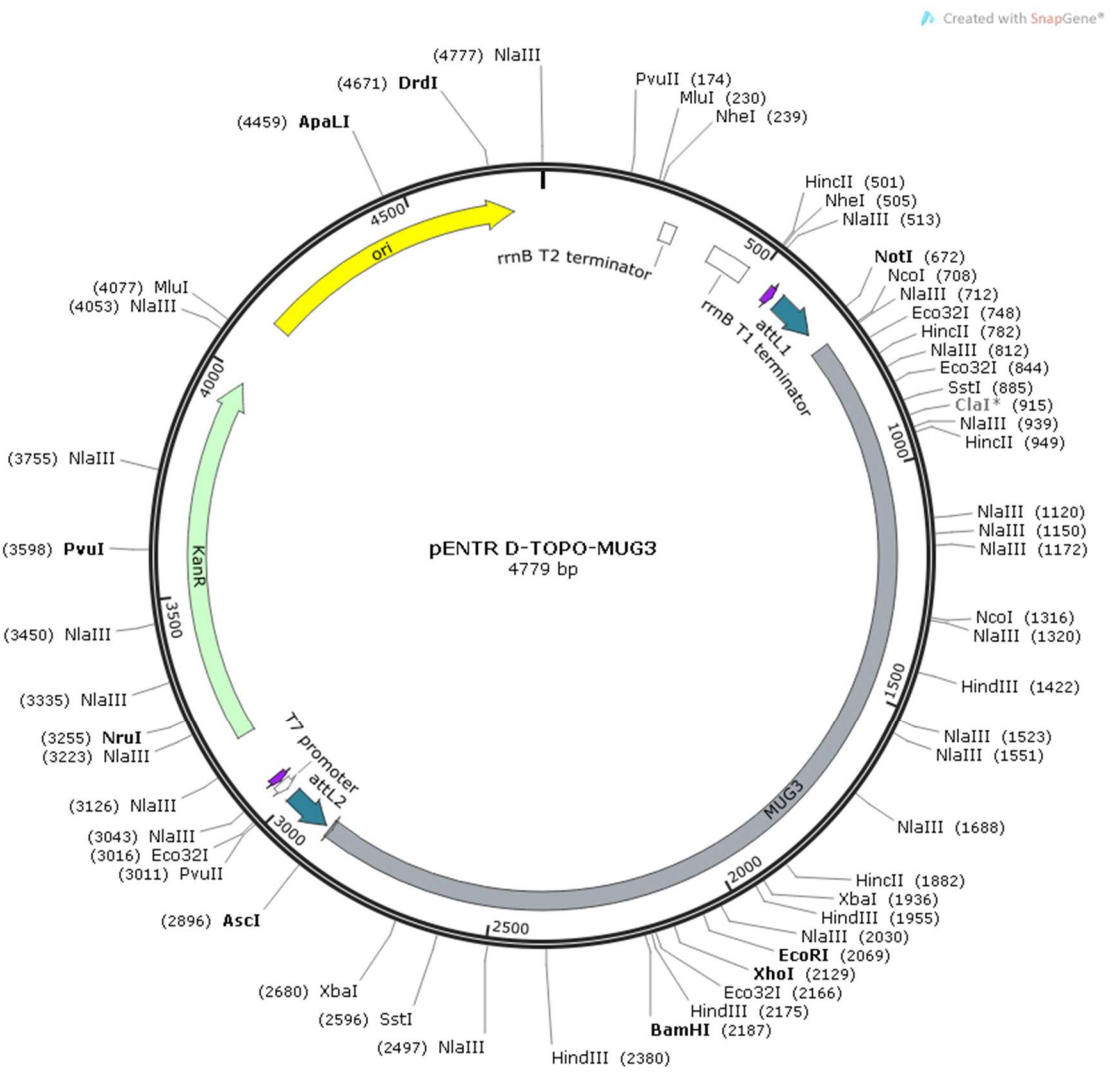


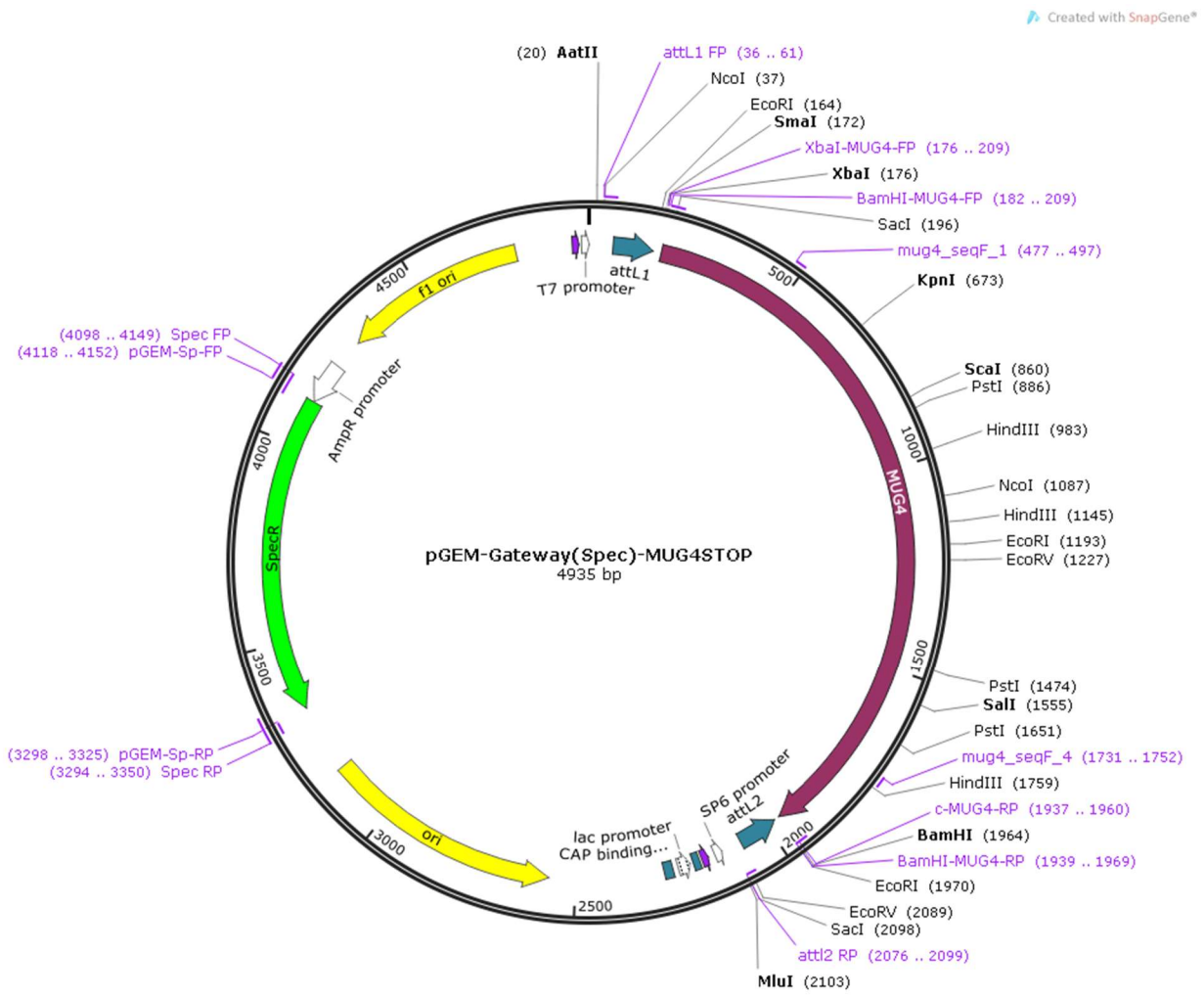
Supplemental Figure 2.2 The expression pattern of *MUG-A* genes collected from eFP Browser.

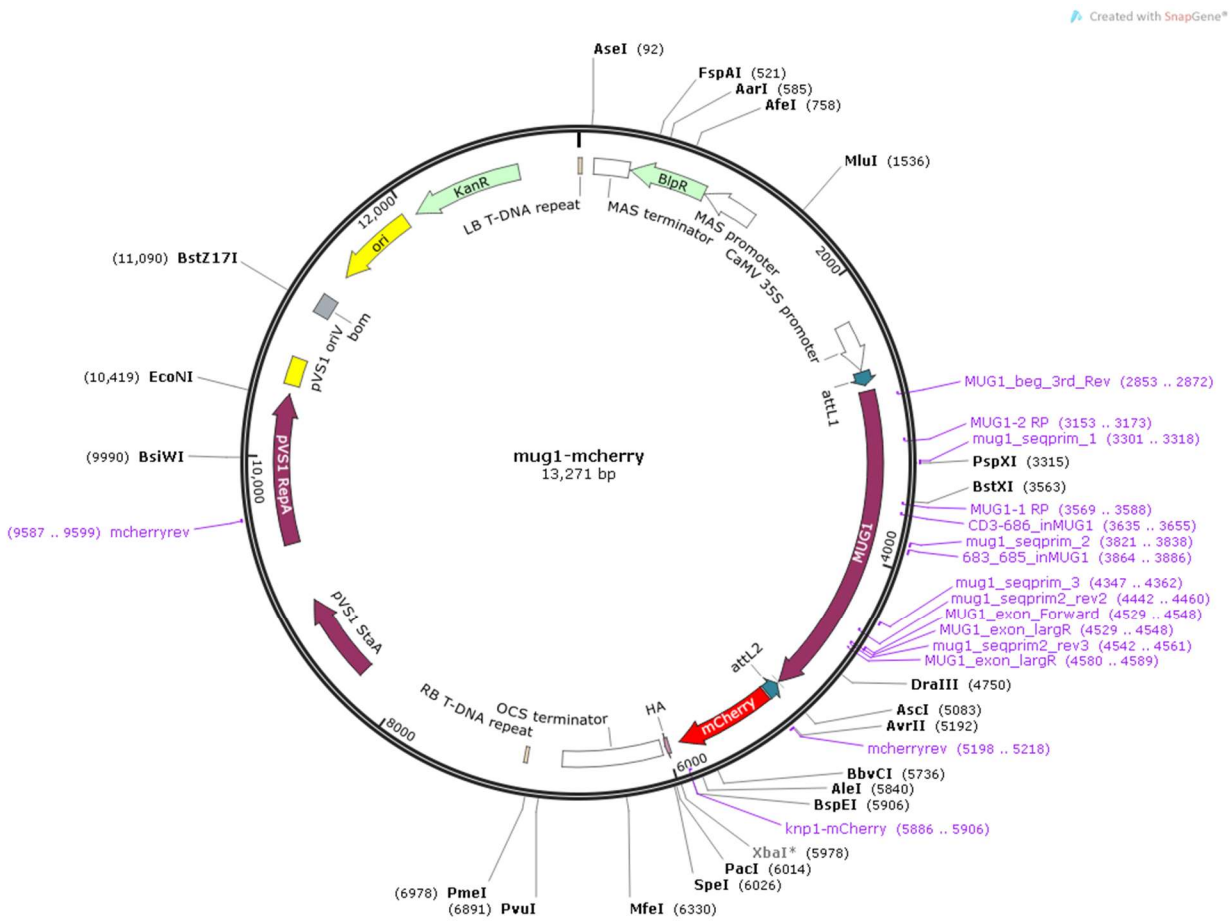

Created with SnapGene®
pEarleyGate 101-mcherry
12,448 bp
(12,211) BclI*
(12,080) PsiI
(11,754) BspHI
(11,331) EcoRV
AseI (92)
FspAI (521)
AarI (585)
AfeI (758)
12,000
KanR
LB T-DNA repeat
BlpR
MAS terminator
MAS promoter
CaMV 35S promoter
2000
PshAI (2467)
attR1
CmR
ori
bom
pVS1 oriV
10,000
(9596) EcoNI
(9287) BmtI
(9283) NheI
(9167) BsiWI
pVS1 RepA
(8764 .. 8776) mcherryrev
attR2
ccdB
4000
SrfI (3980)
BstXI (4097)
AvrII (4369)
mcherryrev (4375 .. 4395)
mCherry
HA
OCS terminator
RB T-DNA repeat
pVS1 StaA
8000
6000
AleI (5017)
knp1-mCherry (5063 .. 5083)
XbaI* (5155)
PacI (5191)
SpeI (5203)
MfeI (5507)
(6155) PmeI
PvuI (6068)

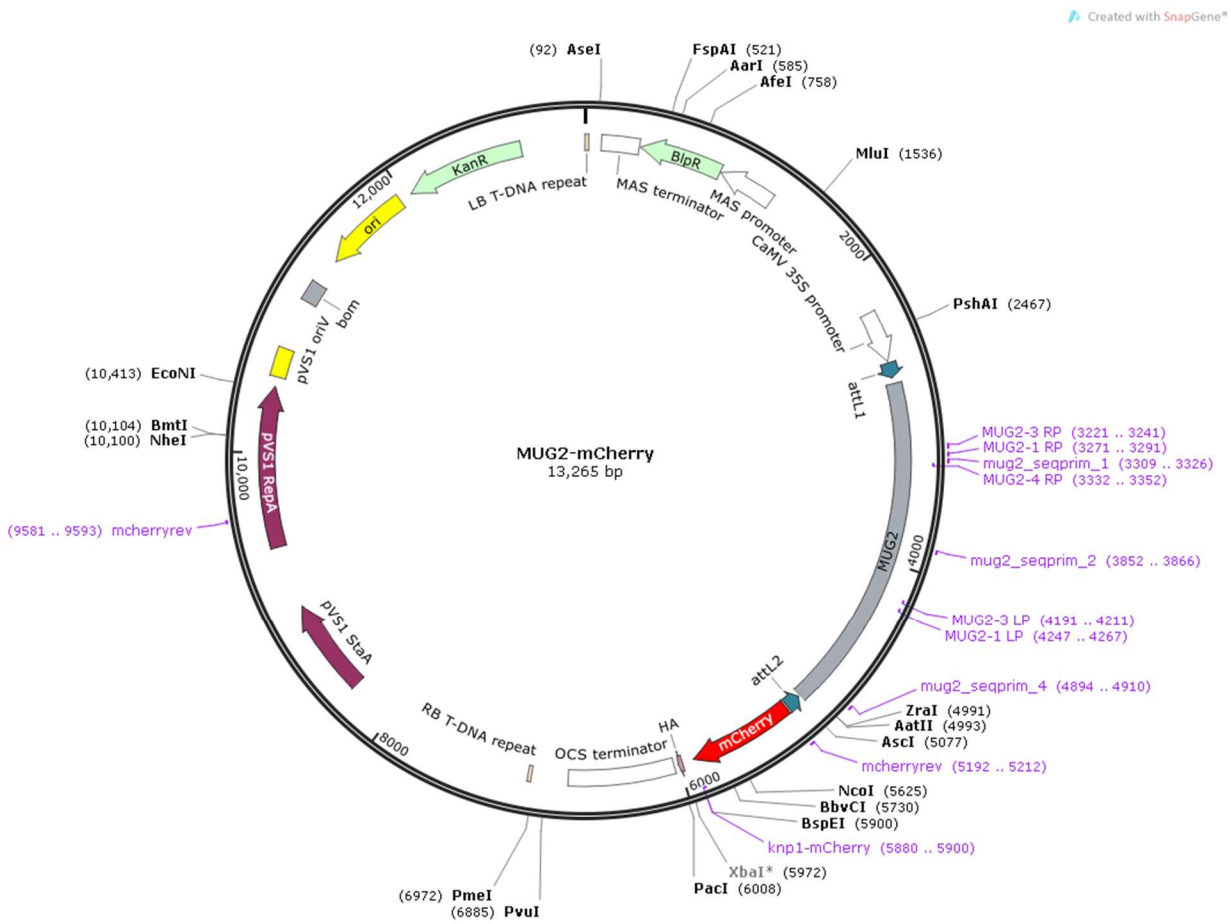

Created with SnapGene®
(230) MluI
NheI (239)
(174) PvuII
HincII (501)
NheI (505)
(4870) NlaIII
NlaIII (513)
(4764) DrdI
NotI (672)
(4552) ApaLI
NcoI (708)
NlaIII (712)
NruI (720)
MUG1_beg_3rd_Rev (759 .. 778)
NlaIII (818)
NlaIII (872)
(4170) MluI
NlaIII (878)
(4146) NlaIII
NlaIII (887)
NlaIII (911)
NlaIII (920)
NlaIII (1073)
MUG1-2 RP (1059 .. 1079)
(3848) NlaIII
NlaIII (1124)
HincII (1153)
NlaIII (1166)
NlaIII (1187)
(3691) PvuI
mug1_seqprim_1 (1207 .. 1224)
XhoI (1221)
SstI (1228)
NlaIII (1243)
(3543) NlaIII
NcoI (1394)
NlaIII (1398)
(3428) NlaIII
SmaI (1423)
NlaIII (1457)
(3348) NruI
MUG1-1 RP (1475 .. 1494)
(3316) NlaIII
HindIII (1500)
NheI (1520)
(3219) NlaIII
NruI (1545)
(3136) NlaIII
CD3-686_inMUG1 (1541 .. 1561)
(3109) Eco32I
HindIII (1623)
(3104) PvuII
NlaIII (1651)
(2989) AscI
mug1_seqprim_2 (1727 .. 1744)
(2646) ApaLI
PvuII (1760)
(2593) Eco32I
683_685_inMUG1 (1770 .. 1792)
(2573) NlaIII
HindIII (1797)
(2486 .. 2495) MUG1_exon_largR
HincII (1829)
(2448 .. 2467) mug1_seqprim2_rev3
HindIII (2128)
(2435 .. 2454) MUG1_exon_largR
(2435 .. 2454) MUG1_exon_Forward
mug1_seqprim_3 (2253 .. 2268)
(2348 .. 2366) mug1_seqprim2_rev2
(2358) NlaIII
BamHI (2259)
ori
rrnB T1 termina...
attL1
attL2
KanR
MUG1
500
1000
1500
2000
2500
3000
3500
4000
4500
pENTR D-TOPO-MUG1
4872 bp

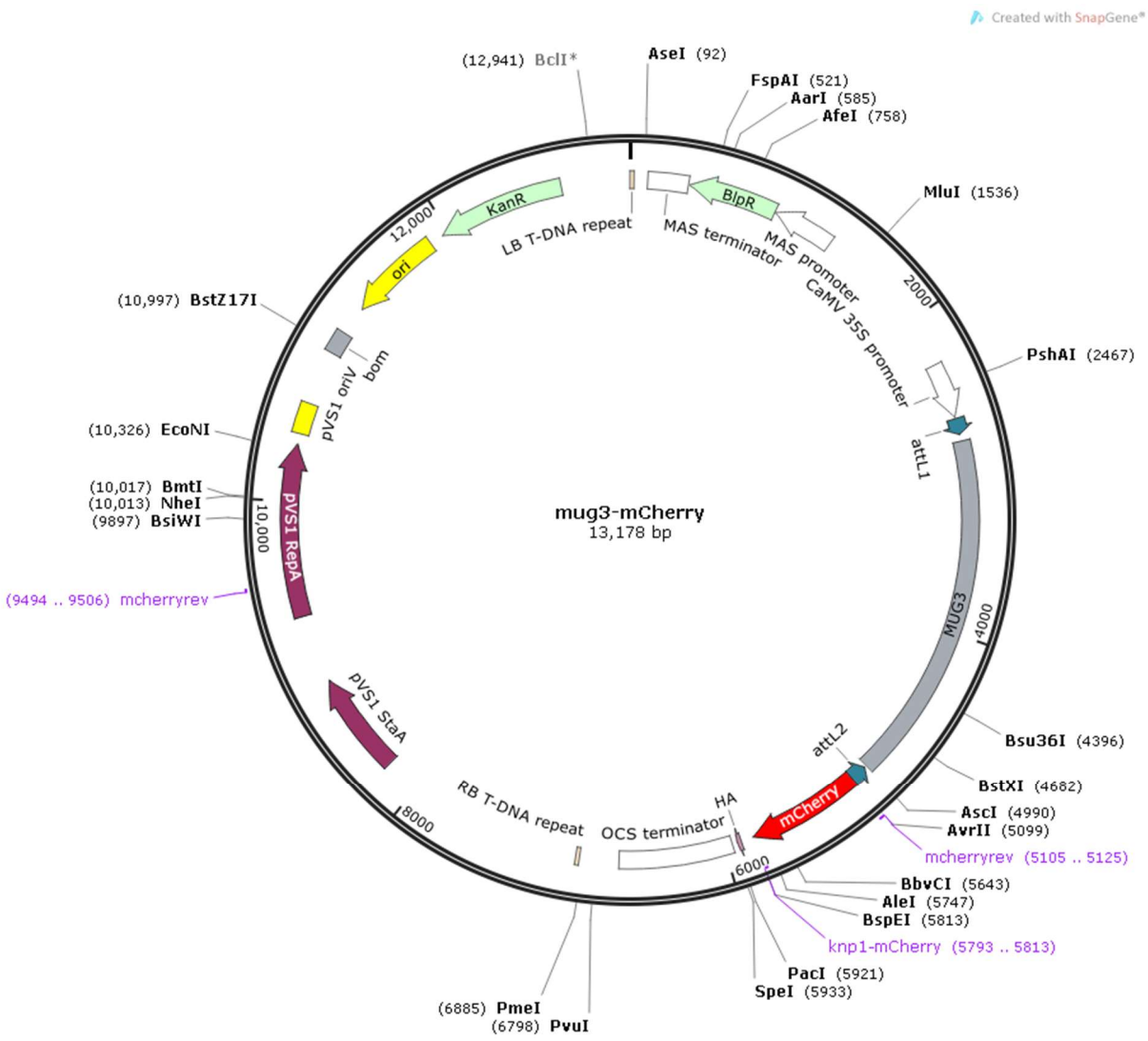

Created with SnapGene®
pENTR D-TOPO-MUG2
4866 bp
(4864) NlaIII
(4758) DrdI
PvuII (174)
MluI (230)
NheI (239)
HincII (501)
NheI (505)
NlaIII (513)
ApaI (567)
NotI (672)
NlaIII (712)
SpeI (803)
NlaIII (812)
PstI (855)
NlaIII (908)
NlaIII (953)
NlaIII (1141)
BamHI (1147)
EcoRI (1395)
HindIII (1409)
NlaIII (1451)
NlaIII (1464)
HindIII (1494)
NlaIII (1595)
NlaIII (1623)
NdeI (1663)
NlaIII (1745)
HincII (1823)
PvuII (1990)
NlaIII (2017)
NlaIII (2096)
NlaIII (2449)
(2489) NlaIII
(2560) NlaIII
(2983) AscI
(3098) PvuII
(3130) NlaIII
(3213) NlaIII
(3310) NlaIII
(3342) NruI
(3422) NlaIII
(3537) NlaIII
(3685) PvuI
(3842) NlaIII
(4140) NlaIII
(4164) MluI
ori
rrnB T2 terminator
rrnB T1 terminator
attL1
attL2
T7 promoter
KanR
MUG2
500
1000
1500
2000
2500
3000
3500
4000
4500

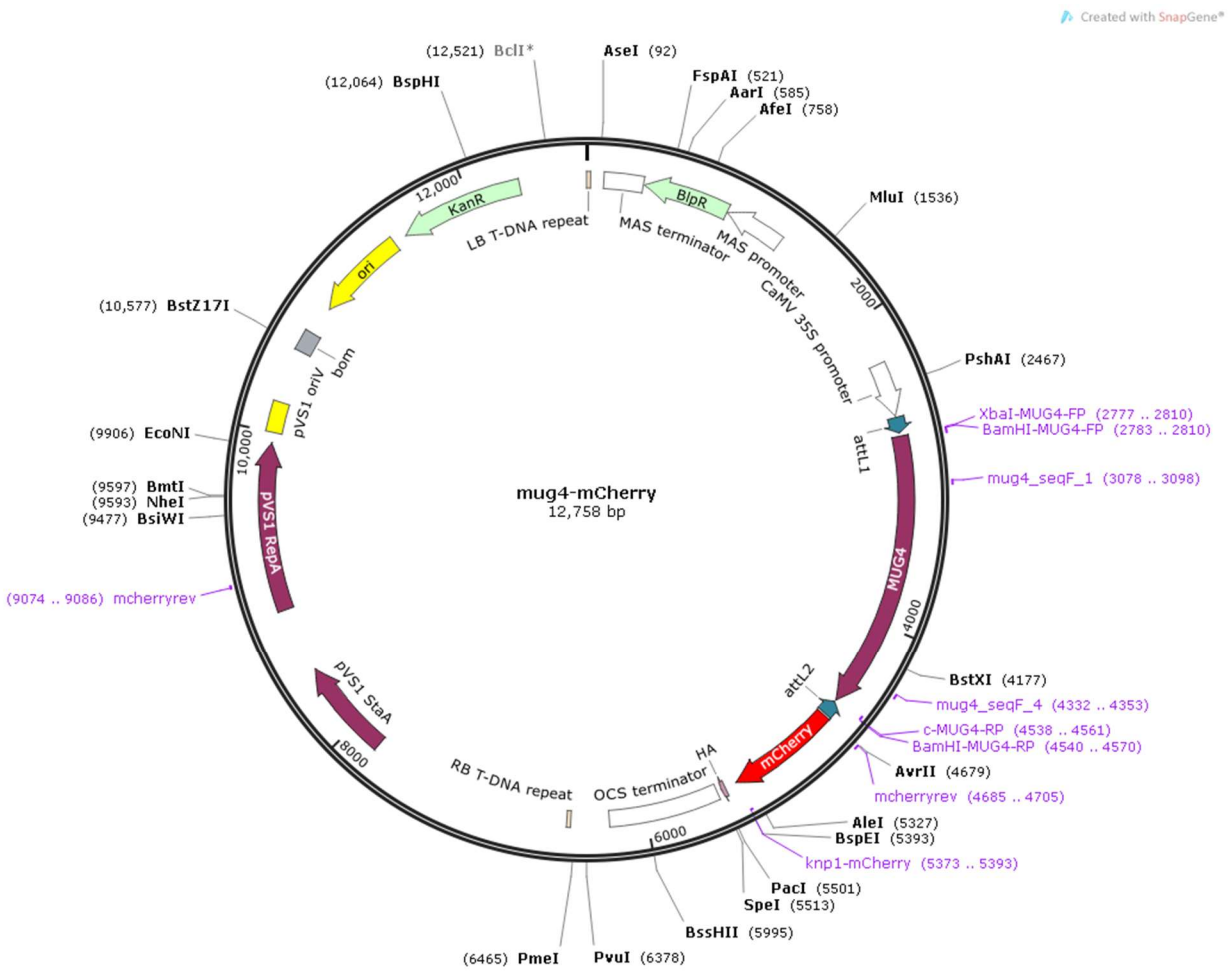

Created with SnapGene®
pENTR D-TOPO-MUG3
4779 bp
(4777) NlaIII
(4671) DrdI
(4459) ApaLI
PvuII (174)
MluI (230)
NheI (239)
HincII (501)
NheI (505)
NlaIII (513)
NotI (672)
NcoI (708)
NlaIII (712)
Eco32I (748)
HincII (782)
NlaIII (812)
Eco32I (844)
SstI (885)
ClaI* (915)
NlaIII (939)
HincII (949)
NlaIII (1120)
NlaIII (1150)
NlaIII (1172)
NcoI (1316)
NlaIII (1320)
HindIII (1422)
NlaIII (1523)
NlaIII (1551)
NlaIII (1688)
HincII (1882)
XbaI (1936)
HindIII (1955)
NlaIII (2030)
EcoRI (2069)
XhoI (2129)
Eco32I (2166)
HindIII (2175)
BamHI (2187)
HindIII (2380)
(2497) NlaIII
(2596) SstI
(2680) XbaI
(2896) AscI
(3011) PvuII
(3016) Eco32I
(3043) NlaIII
(3126) NlaIII
(3223) NlaIII
(3255) NruI
(3335) NlaIII
(3450) NlaIII
(3598) PvuI
(3755) NlaIII
(4053) NlaIII
(4077) MluI
rrnB T2 terminator
rrnB T1 terminator
attL1
attL2
T7 promoter
ori
KanR
MUG3
500
1000
1500
2000
2500
3000
3500
4000
4500

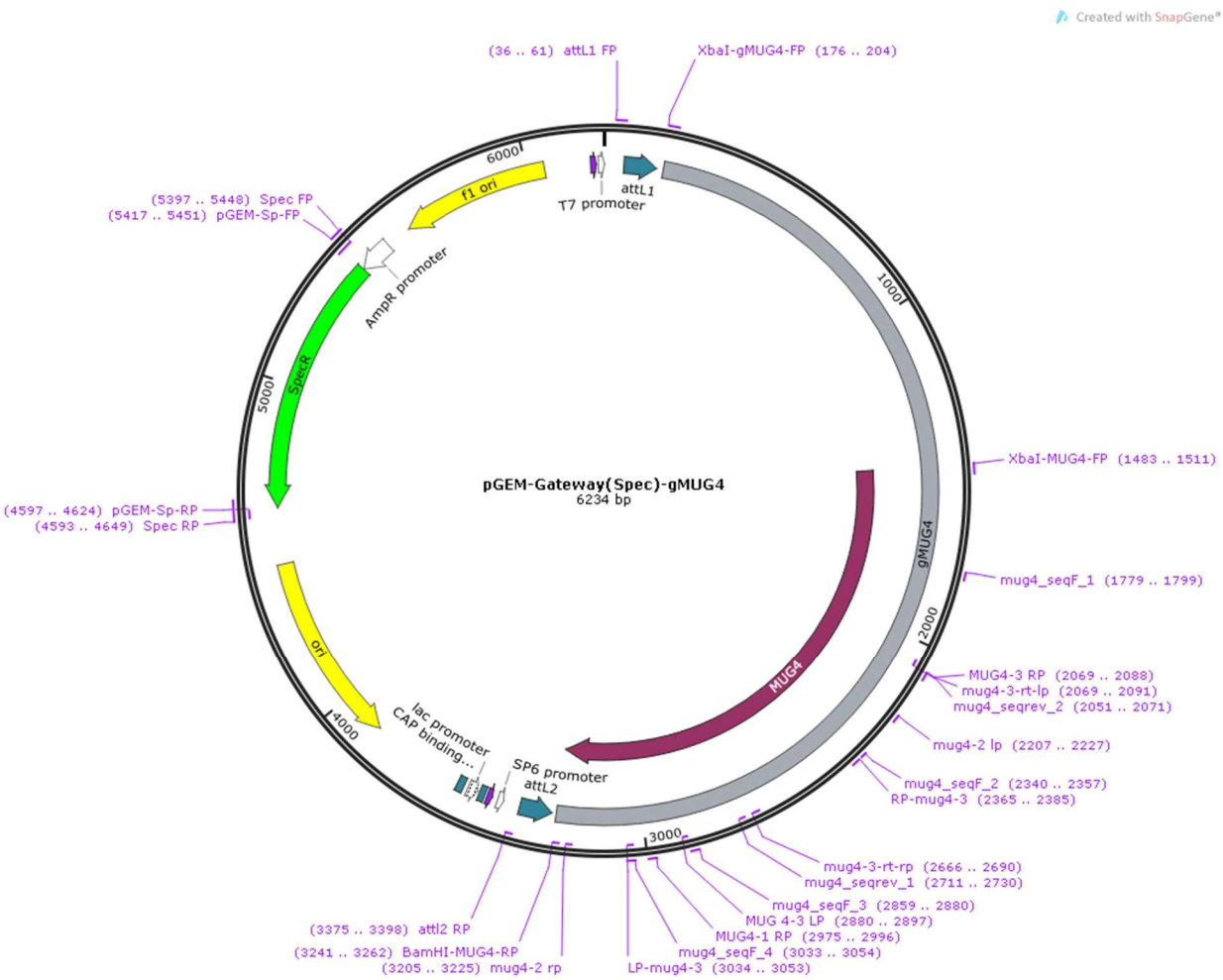

Created with SnapGene®
(20) AatII
attL1 FP (36 .. 61)
NcoI (37)
EcoRI (164)
SmaI (172)
XbaI-MUG4-FP (176 .. 209)
XbaI (176)
BamHI-MUG4-FP (182 .. 209)
SacI (196)
mug4_seqF_1 (477 .. 497)
KpnI (673)
ScaI (860)
PstI (886)
HindIII (983)
NcoI (1087)
HindIII (1145)
EcoRI (1193)
EcoRV (1227)
PstI (1474)
SalI (1555)
PstI (1651)
mug4_seqF_4 (1731 .. 1752)
HindIII (1759)
c-MUG4-RP (1937 .. 1960)
BamHI (1964)
BamHI-MUG4-RP (1939 .. 1969)
EcoRI (1970)
EcoRV (2089)
SacI (2098)
attl2 RP (2076 .. 2099)
MluI (2103)
attL1
T7 promoter
f1 ori
(4098 .. 4149) Spec FP
(4118 .. 4152) pGEM-Sp-FP
AmpR promoter
SpecR
(3298 .. 3325) pGEM-Sp-RP
(3294 .. 3350) Spec RP
ori
lac promoter
CAP binding...
SP6 promoter
attL2
MUG4
pGEM-Gateway(Spec)-MUG4STOP
4935 bp
500
1000
1500
2000
2500
3000
3500
4000
4500

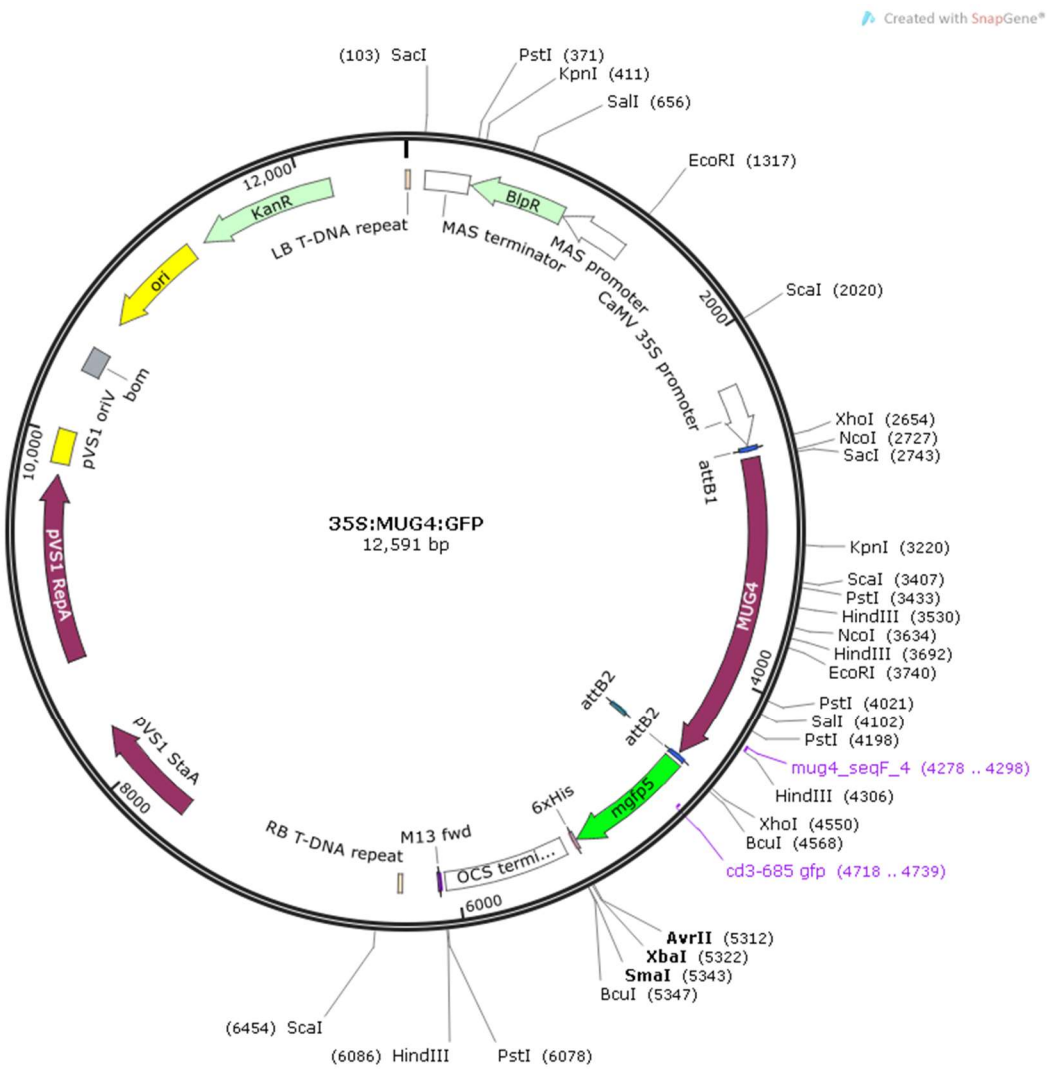

Created with SnapGene®
mug1-mcherry
13,271 bp
AseI (92)
FspAI (521)
AarI (585)
AfeI (758)
MluI (1536)
2000
BlpR
MAS promoter
CaMV 35S promoter
attL1
MUG1
4000
attL2
mCherry
HA
6000
OCS terminator
RB T-DNA repeat
8000
pVS1 StbA
pVS1 RepA
10,000
LB T-DNA repeat
MAS terminator
KanR
12,000
ori
pVS1 oriV
bom
(11,090) BstZ17I
(10,419) EcoNI
(9990) BsiWI
(9587 .. 9599) mcherryrev
MUG1_beg_3rd_Rev (2853 .. 2872)
MUG1-2 RP (3153 .. 3173)
mug1_seqprim_1 (3301 .. 3318)
PspXI (3315)
BstXI (3563)
MUG1-1 RP (3569 .. 3588)
CD3-686_inMUG1 (3635 .. 3655)
mug1_seqprim_2 (3821 .. 3838)
683_685_inMUG1 (3864 .. 3886)
mug1_seqprim_3 (4347 .. 4362)
mug1_seqprim2_rev2 (4442 .. 4460)
MUG1_exon_Forward (4529 .. 4548)
MUG1_exon_largR (4529 .. 4548)
mug1_seqprim2_rev3 (4542 .. 4561)
MUG1_exon_largR (4580 .. 4589)
DraIII (4750)
AscI (5083)
AvrII (5192)
mcherryrev (5198 .. 5218)
BbvCI (5736)
AleI (5840)
BspEI (5906)
knp1-mCherry (5886 .. 5906)
XbaI* (5978)
PacI (6014)
SpeI (6026)
MfeI (6330)
(6978) PmeI
(6891) PvuI

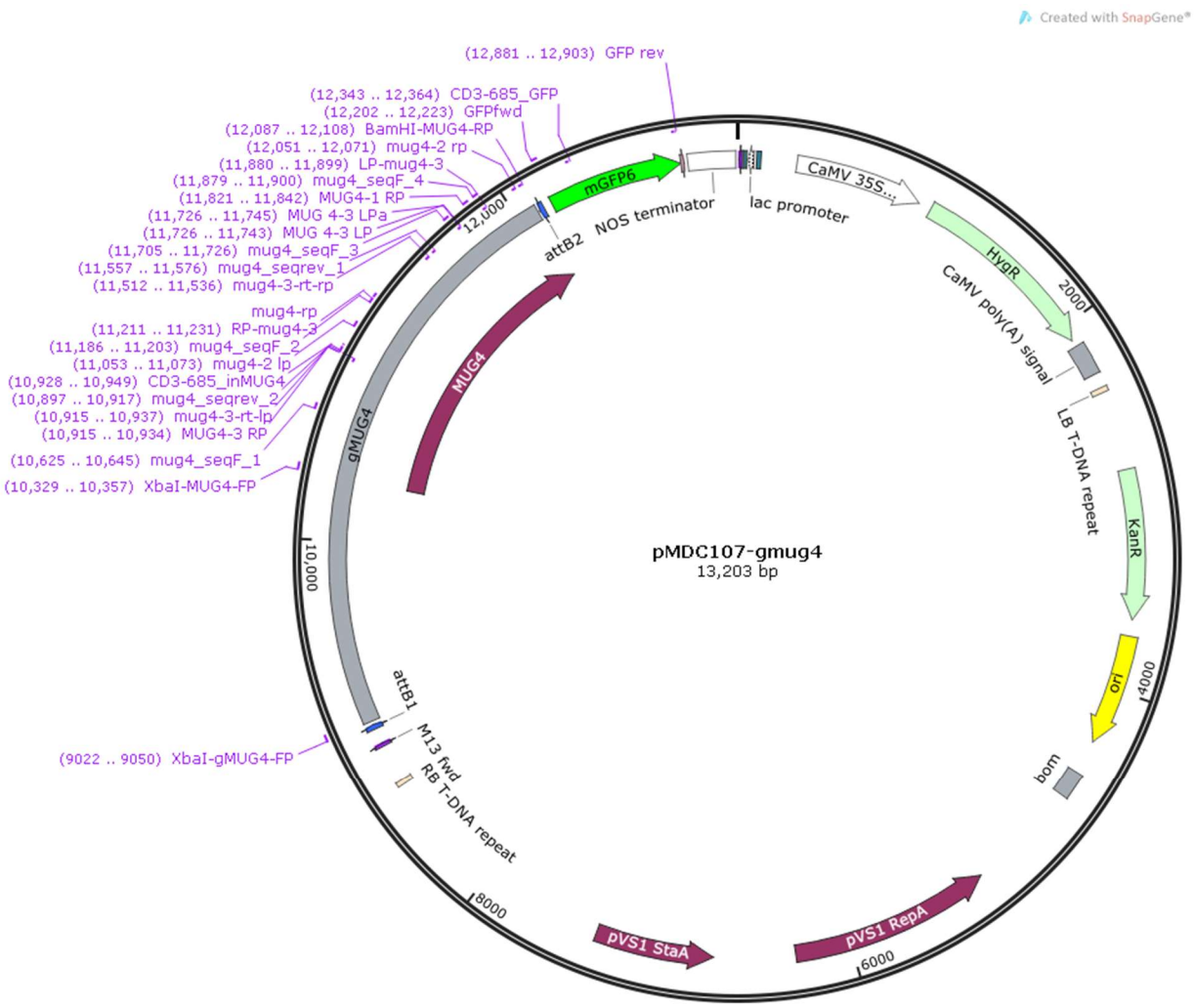

Created with SnapGene®
MUG2-mCherry
13,265 bp
(92) AseI
FspAI (521)
AarI (585)
AfeI (758)
MluI (1536)
PshAI (2467)
(10,413) EcoNI
(10,104) BmtI
(10,100) NheI
(9581 .. 9593) mcherryrev
KanR
LB T-DNA repeat
MAS terminator
BlpR
MAS promoter
CaMV 35S promoter
attL1
MUG2
attL2
mCherry
HA
OCS terminator
RB T-DNA repeat
pVS1 StaA
pVS1 RepA
pVS1 oriV
bom
ori
MUG2-3 RP (3221 .. 3241)
MUG2-1 RP (3271 .. 3291)
mug2_seqprim_1 (3309 .. 3326)
MUG2-4 RP (3332 .. 3352)
mug2_seqprim_2 (3852 .. 3866)
MUG2-3 LP (4191 .. 4211)
MUG2-1 LP (4247 .. 4267)
mug2_seqprim_4 (4894 .. 4910)
ZraI (4991)
AatII (4993)
AscI (5077)
mcherryrev (5192 .. 5212)
NcoI (5625)
BbvCI (5730)
BspEI (5900)
knp1-mCherry (5880 .. 5900)
XbaI* (5972)
PacI (6008)
(6972) PmeI
(6885) PvuI
2,000
4,000
6,000
8,000
10,000
12,000

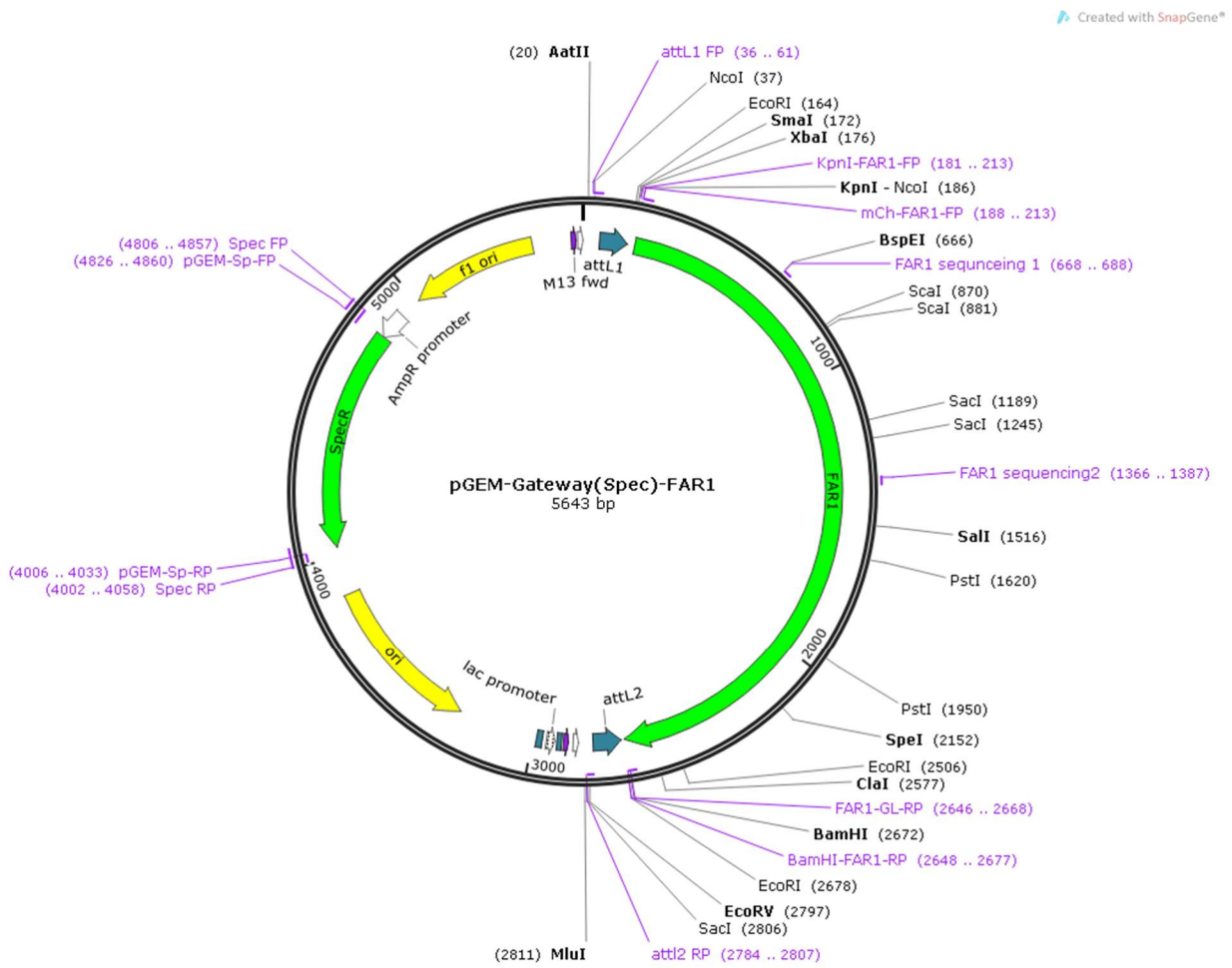

Created with SnapGene*
mug3-mCherry
13,178 bp
(12,941) BclI*
AseI (92)
FspAI (521)
AarI (585)
AfeI (758)
MluI (1536)
PshAI (2467)
Bsu36I (4396)
BstXI (4682)
AscI (4990)
AvrII (5099)
mcherryrev (5105 .. 5125)
BbvCI (5643)
AleI (5747)
BspEI (5813)
knp1-mCherry (5793 .. 5813)
PacI (5921)
SpeI (5933)
(6885) PmeI
(6798) PvuI
(10,997) BstZ17I
(10,326) EcoNI
(10,017) BmtI
(10,013) NheI
(9897) BsiWI
(9494 .. 9506) mcherryrev
KanR
BlpR
LB T-DNA repeat
MAS terminator
MAS promoter
CaMV 35S promoter
attL1
MUG3
attL2
mCherry
HA
OCS terminator
RB T-DNA repeat
pVS1 StaA
pVS1 RepA
pVS1 oriV
bom
ori
2000
4000
6000
8000
10,000
12,000

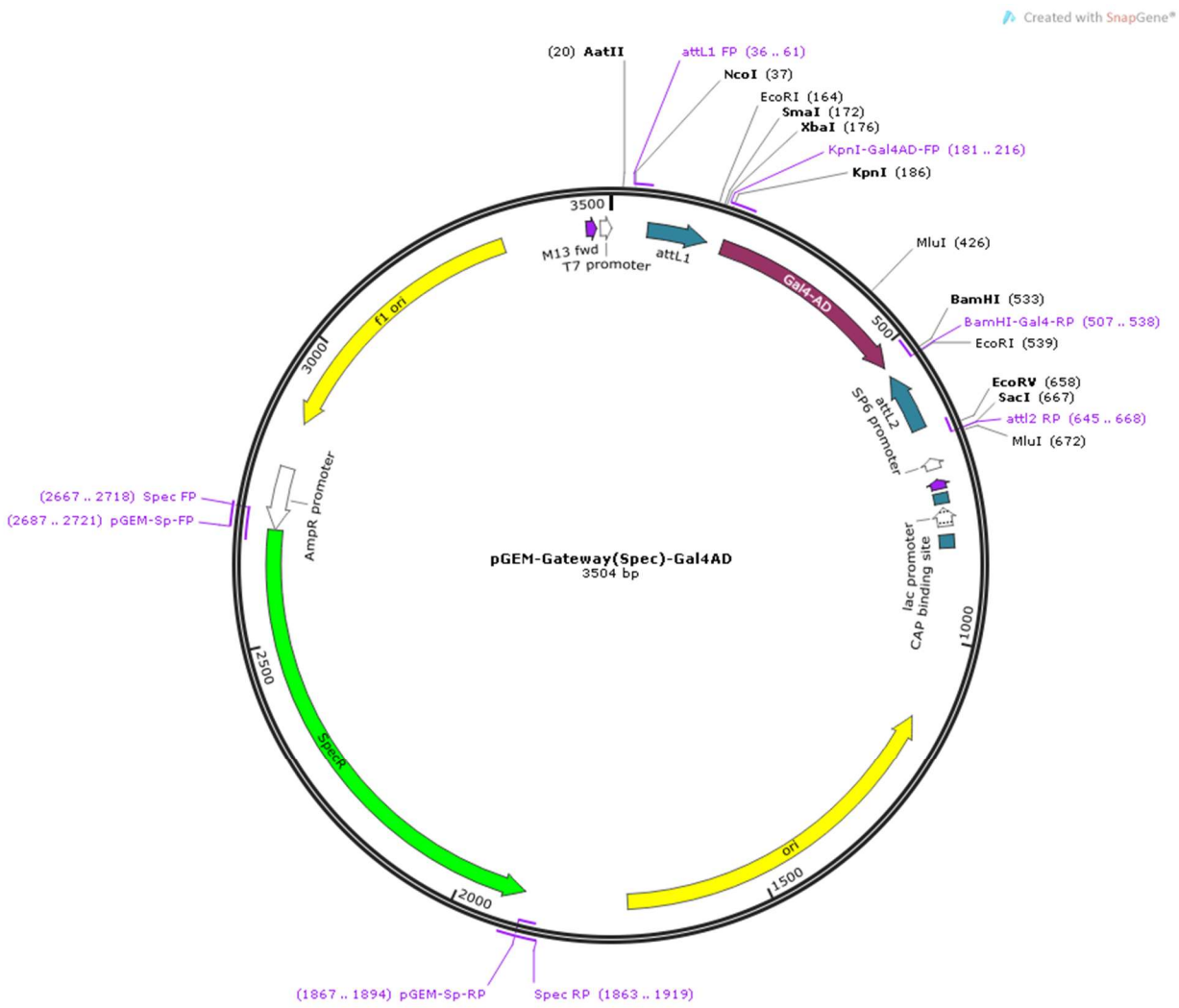

Created with SnapGene®
mug4-mCherry
12,758 bp
(12,521) BclI*
AseI (92)
(12,064) BspHI
FspAI (521)
AarI (585)
AfeI (758)
MluI (1536)
2000
PshAI (2467)
XbaI-MUG4-FP (2777 .. 2810)
BamHI-MUG4-FP (2783 .. 2810)
mug4_seqF_1 (3078 .. 3098)
4000
BstXI (4177)
mug4_seqF_4 (4332 .. 4353)
c-MUG4-RP (4538 .. 4561)
BamHI-MUG4-RP (4540 .. 4570)
AvrII (4679)
mcherryrev (4685 .. 4705)
AleI (5327)
BspEI (5393)
knp1-mCherry (5373 .. 5393)
PacI (5501)
SpeI (5513)
BssHII (5995)
6000
PvuI (6378)
(6465) PmeI
8000
RB T-DNA repeat
OCS terminator
HA
mCherry
attL2
MUG4
attL1
CaMV 35S promoter
MAS promoter
MAS terminator
BlpR
LB T-DNA repeat
KanR
12,000
ori
bom
pVS1 oriV
(10,577) BstZ17I
10,000
(9906) EcoNI
(9597) BmtI
(9593) NheI
(9477) BsiWI
pVS1 RepA
(9074 .. 9086) mcherryrev
pVS1 StaA

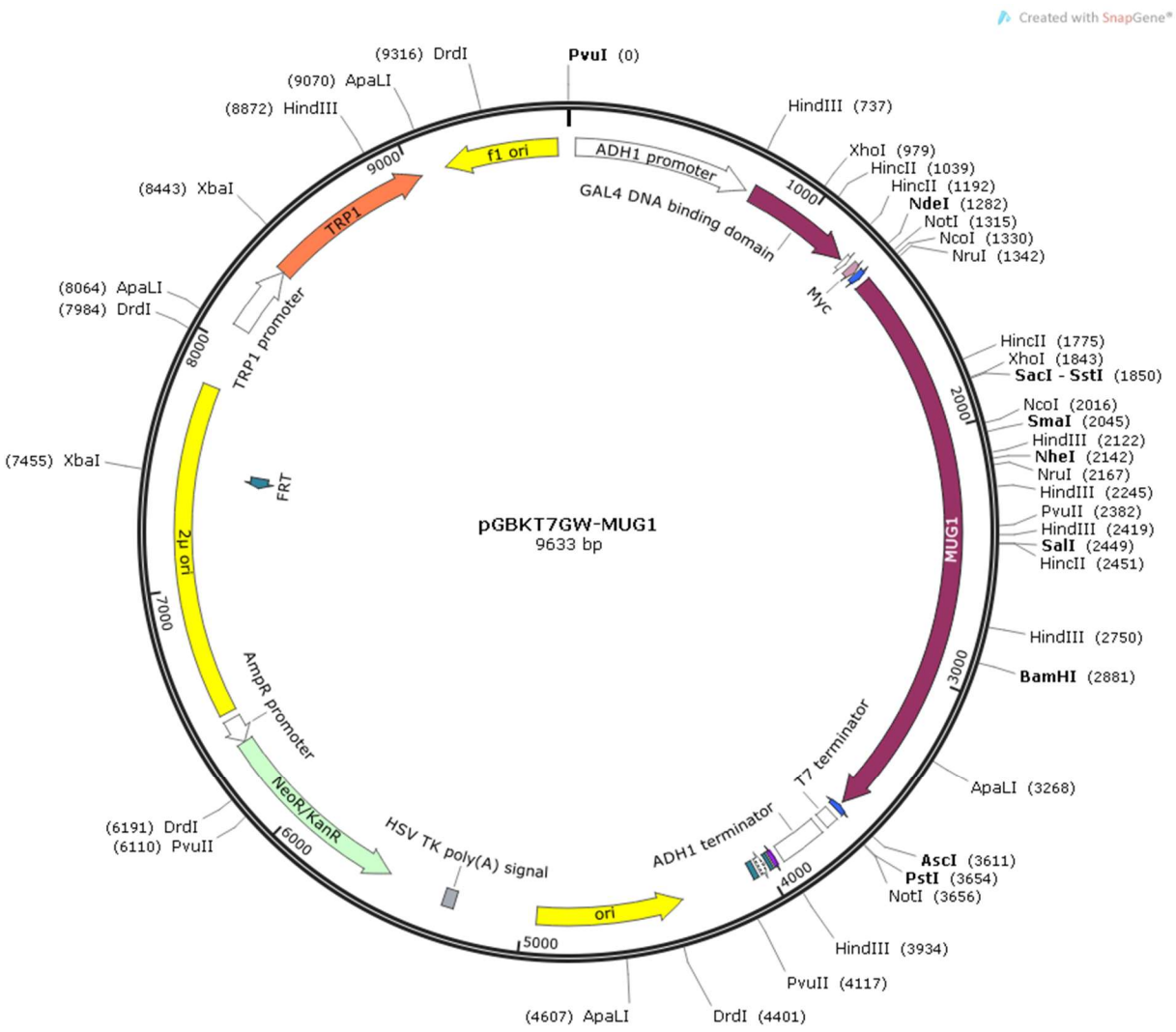

Created with SnapGene®
(36 .. 61) attL1 FP
XbaI-gMUG4-FP (176 .. 204)
6000
f1 ori
attL1
T7 promoter
1000
(5397 .. 5448) Spec FP
(5417 .. 5451) pGEM-Sp-FP
AmpR promoter
5000
SpecR
pGEM-Gateway(Spec)-gMUG4
6234 bp
XbaI-MUG4-FP (1483 .. 1511)
gMUG4
(4597 .. 4624) pGEM-Sp-RP
(4593 .. 4649) Spec RP
mug4_seqF_1 (1779 .. 1799)
2000
MUG4
ori
MUG4-3 RP (2069 .. 2088)
mug4-3-rt-lp (2069 .. 2091)
mug4_seqrev_2 (2051 .. 2071)
4000
lac promoter
CAP binding...
SP6 promoter
attL2
mug4-2 lp (2207 .. 2227)
mug4_seqF_2 (2340 .. 2357)
RP-mug4-3 (2365 .. 2385)
3000
mug4-3-rt-rp (2666 .. 2690)
mug4_seqrev_1 (2711 .. 2730)
mug4_seqF_3 (2859 .. 2880)
MUG 4-3 LP (2880 .. 2897)
MUG4-1 RP (2975 .. 2996)
mug4_seqF_4 (3033 .. 3054)
LP-mug4-3 (3034 .. 3053)
(3375 .. 3398) attl2 RP
(3241 .. 3262) BamHI-MUG4-RP
(3205 .. 3225) mug4-2 rp

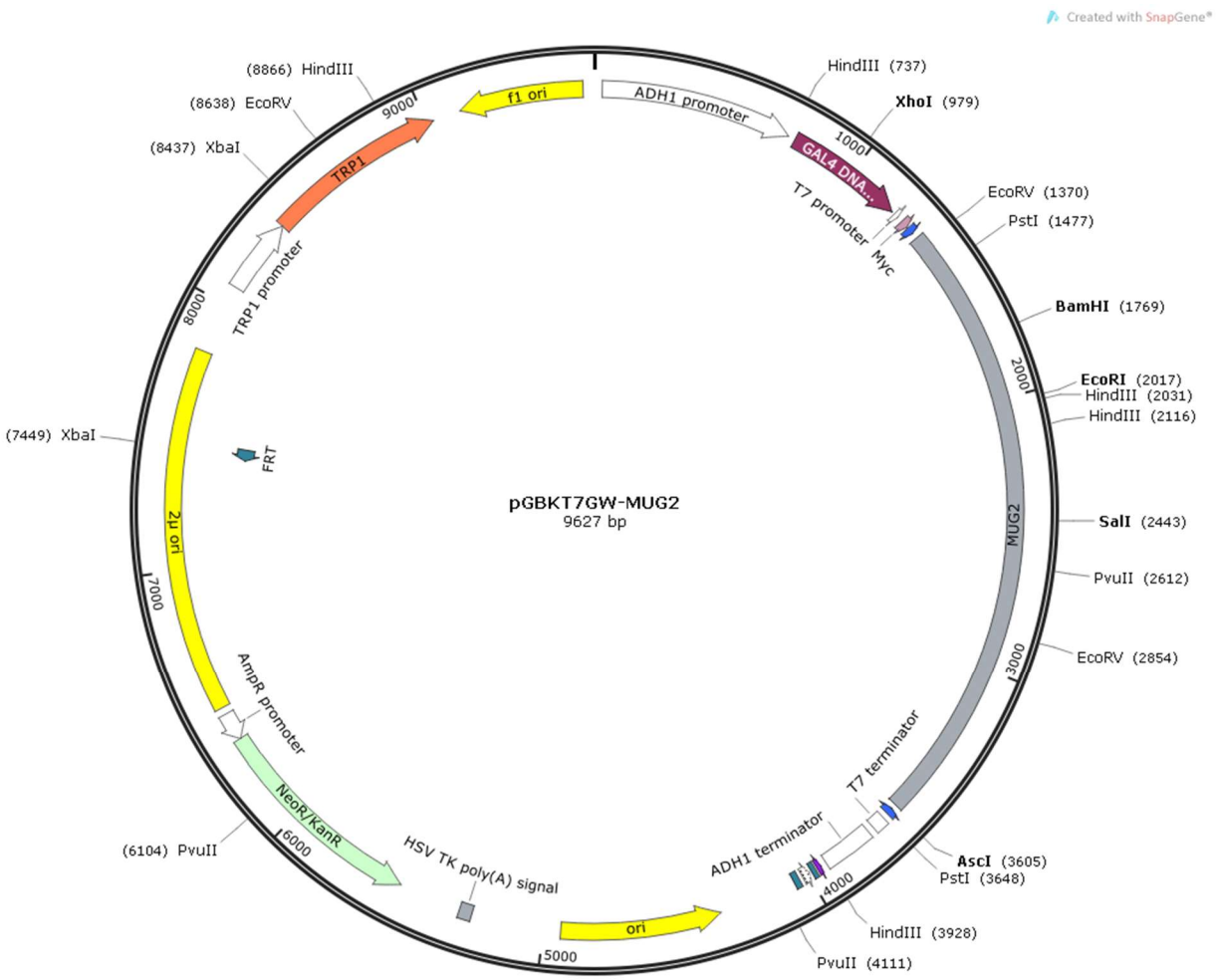

Created with SnapGene®
(103) SacI
PstI (371)
KpnI (411)
SalI (656)
EcoRI (1317)
ScaI (2020)
XhoI (2654)
NcoI (2727)
SacI (2743)
KpnI (3220)
ScaI (3407)
PstI (3433)
HindIII (3530)
NcoI (3634)
HindIII (3692)
EcoRI (3740)
PstI (4021)
SalI (4102)
PstI (4198)
mug4_seqF_4 (4278 .. 4298)
HindIII (4306)
XhoI (4550)
BcuI (4568)
cd3-685 gfp (4718 .. 4739)
AvrII (5312)
XbaI (5322)
SmaI (5343)
BcuI (5347)
PstI (6078)
HindIII (6086)
ScaI (6454)
12,000
2000
4000
6000
8000
10,000
KanR
LB T-DNA repeat
MAS terminator
BlpR
MAS promoter
CaMV 35S promoter
ori
bom
pVS1 oriV
pVS1 RepA
pVS1 StaA
attB1
attB2
attB2
MUG4
mgfp5
6xHis
M13 fwd
OCS terml...
RB T-DNA repeat
35S:MUG4:GFP
12,591 bp

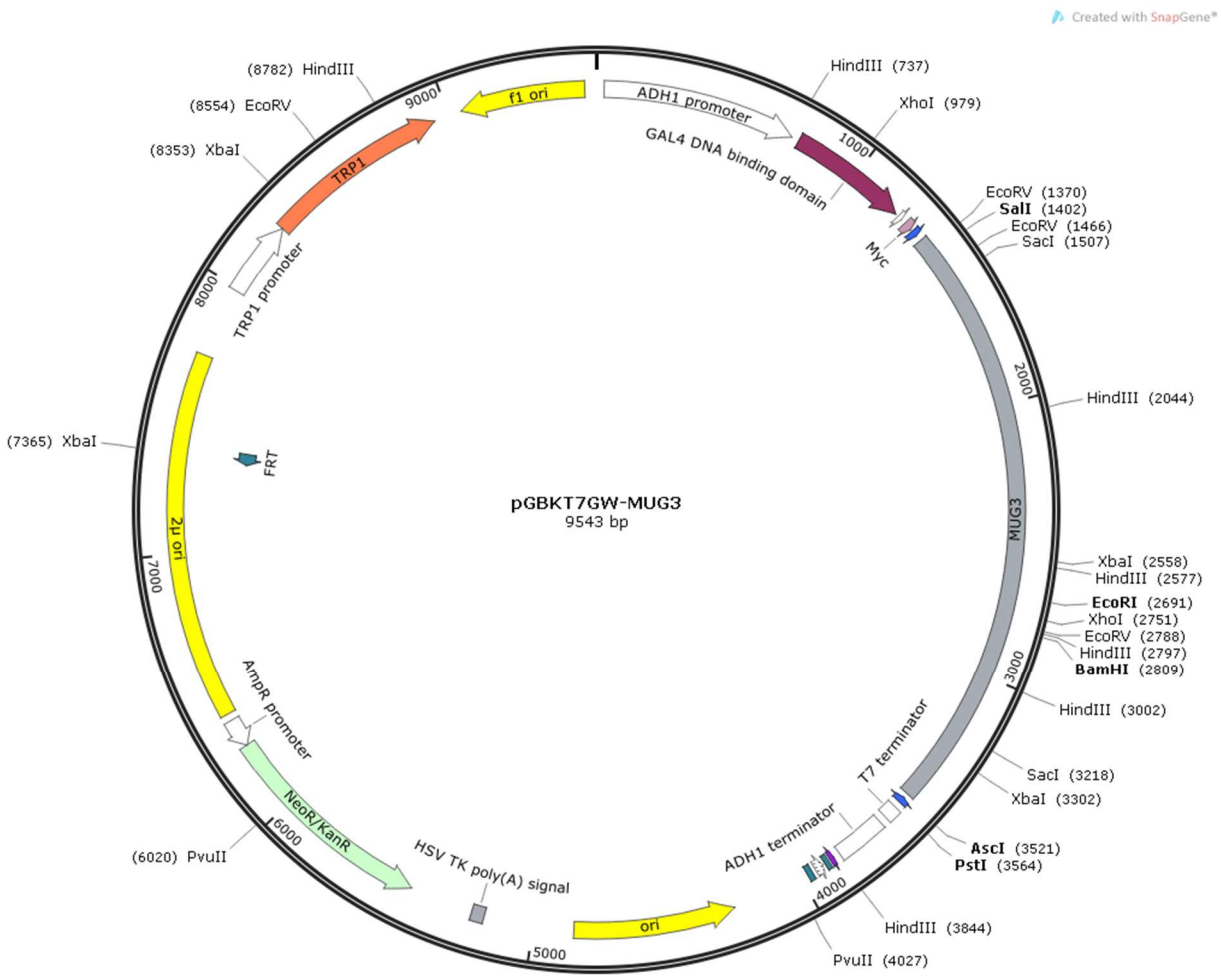

Created with SnapGene®
(12,881 .. 12,903) GFP rev
(12,343 .. 12,364) CD3-685_GFP
(12,202 .. 12,223) GFPfwd
(12,087 .. 12,108) BamHI-MUG4-RP
(12,051 .. 12,071) mug4-2 rp
(11,880 .. 11,899) LP-mug4-3
(11,879 .. 11,900) mug4_seqF_4
(11,821 .. 11,842) MUG4-1 RP
(11,726 .. 11,745) MUG 4-3 LPa
(11,726 .. 11,743) MUG 4-3 LP
(11,705 .. 11,726) mug4_seqF_3
(11,557 .. 11,576) mug4_seqrev_1
(11,512 .. 11,536) mug4-3-rt-rp
mug4-rp
(11,211 .. 11,231) RP-mug4-3
(11,186 .. 11,203) mug4_seqF_2
(11,053 .. 11,073) mug4-2 lp
(10,928 .. 10,949) CD3-685_inMUG4
(10,897 .. 10,917) mug4_seqrev_2
(10,915 .. 10,937) mug4-3-rt-lp
(10,915 .. 10,934) MUG4-3 RP
(10,625 .. 10,645) mug4_seqF_1
(10,329 .. 10,357) XbaI-MUG4-FP
(9022 .. 9050) XbaI-gMUG4-FP
mGFP6
attB2
NOS terminator
lac promoter
CaMV 35S...
HygR
CaMV poly(A) signal
LB T-DNA repeat
KanR
ori
bom
attB1
M13 fwd
RB T-DNA repeat
pVS1 StaA
pVS1 RepA
MUG4
gMUG4
12,000
2000
4000
6000
8000
10,000
pMDC107-gmug4
13,203 bp

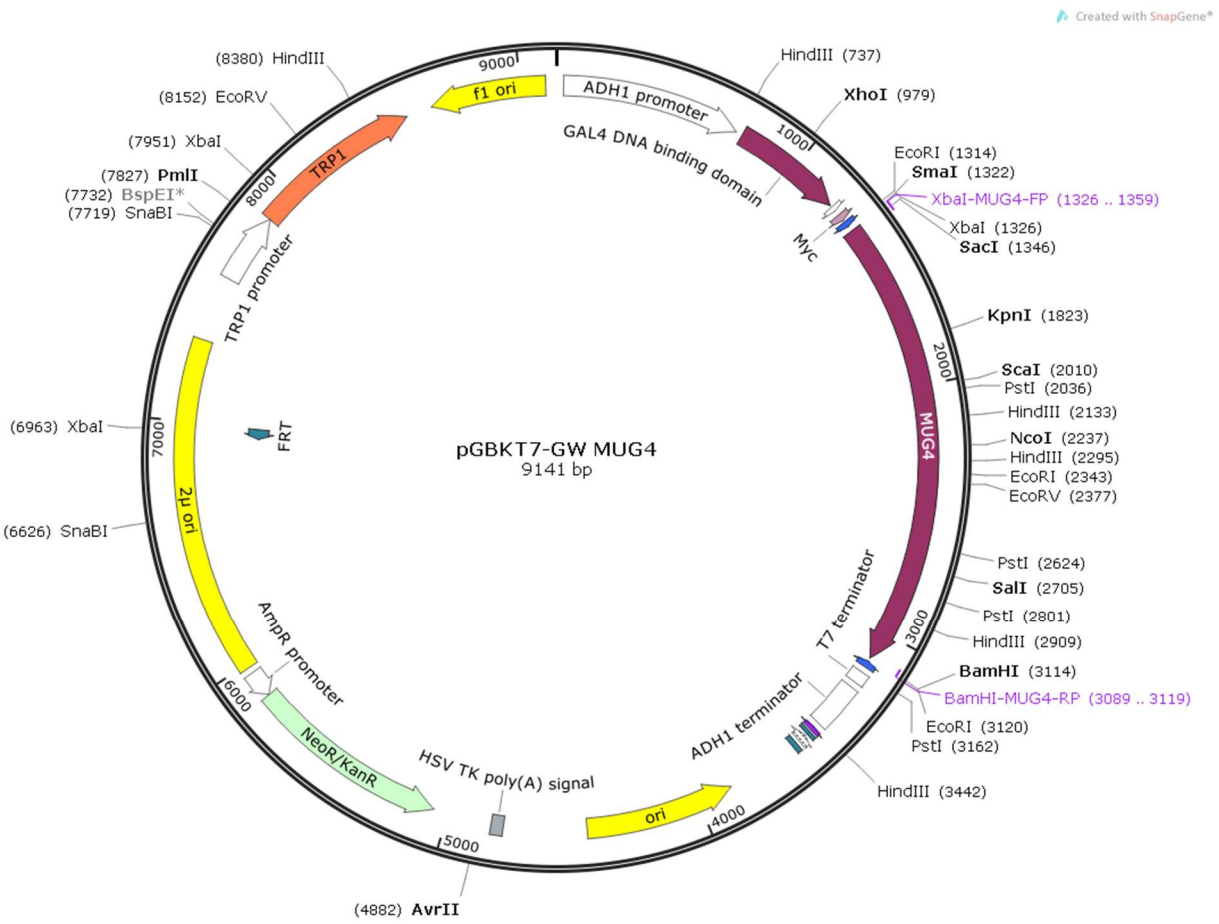

Created with SnapGene®
(20) AatII
attL1 FP (36 .. 61)
NcoI (37)
EcoRI (164)
SmaI (172)
XbaI (176)
KpnI-FAR1-FP (181 .. 213)
KpnI - NcoI (186)
mCh-FAR1-FP (188 .. 213)
BspEI (666)
FAR1 sequnceing 1 (668 .. 688)
ScaI (870)
ScaI (881)
SacI (1189)
SacI (1245)
FAR1 sequencing2 (1366 .. 1387)
SalI (1516)
PstI (1620)
PstI (1950)
SpeI (2152)
EcoRI (2506)
ClaI (2577)
FAR1-GL-RP (2646 .. 2668)
BamHI (2672)
BamHI-FAR1-RP (2648 .. 2677)
EcoRI (2678)
EcoRV (2797)
SacI (2806)
attl2 RP (2784 .. 2807)
(2811) MluI
attL2
lac promoter
ori
(4006 .. 4033) pGEM-Sp-RP
(4002 .. 4058) Spec RP
SpecR
AmpR promoter
f1 ori
(4806 .. 4857) Spec FP
(4826 .. 4860) pGEM-Sp-FP
attL1
M13 fwd
FAR1
pGEM-Gateway(Spec)-FAR1
5643 bp
1000
2000
3000
4000
5000

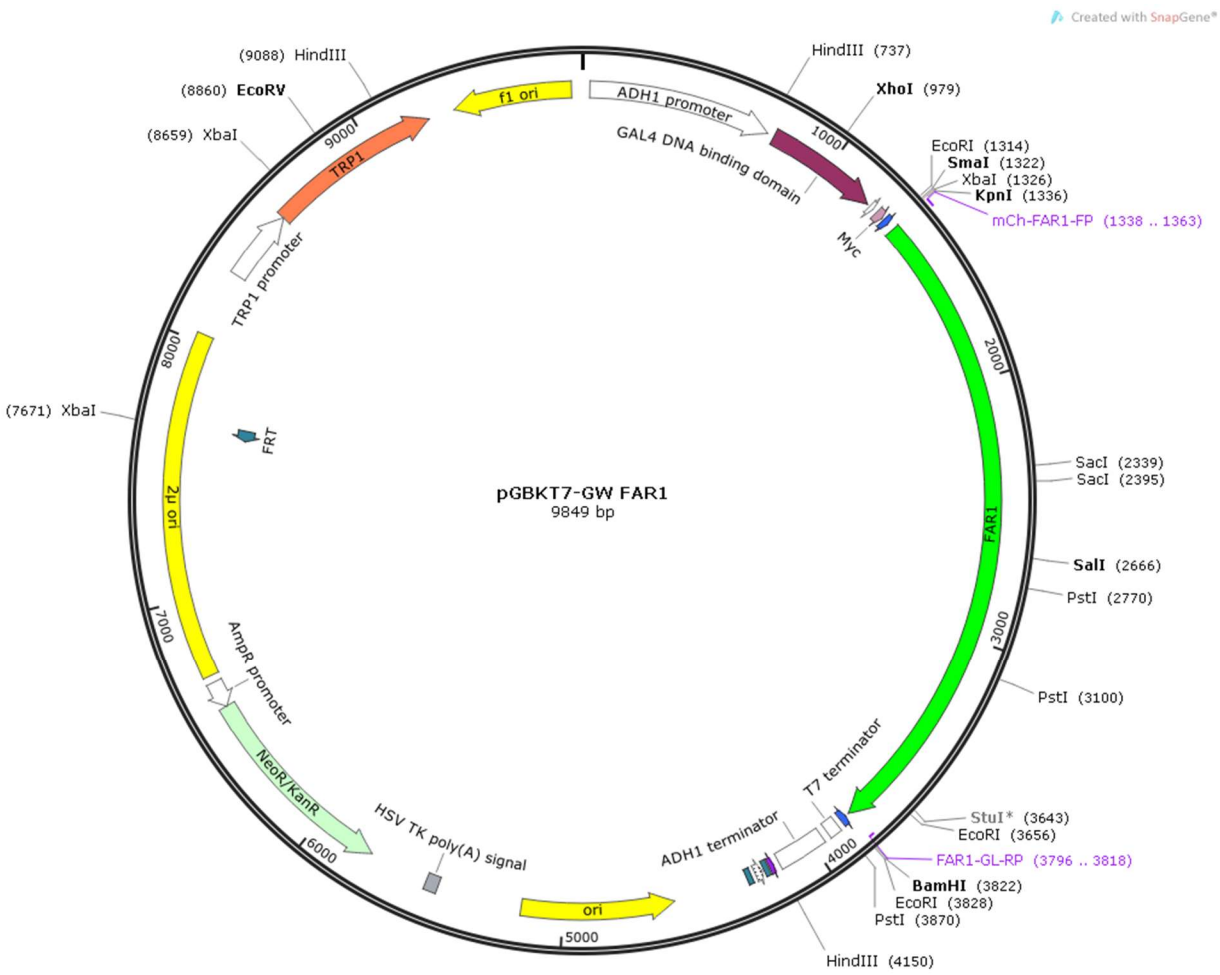

Created with SnapGene®
(20) AatII
attL1 FP (36 .. 61)
NcoI (37)
EcoRI (164)
SmaI (172)
XbaI (176)
KpnI-Gal4AD-FP (181 .. 216)
KpnI (186)
MluI (426)
BamHI (533)
BamHI-Gal4-RP (507 .. 538)
EcoRI (539)
EcoRV (658)
SacI (667)
attl2 RP (645 .. 668)
MluI (672)
(2667 .. 2718) Spec FP
(2687 .. 2721) pGEM-Sp-FP
(1867 .. 1894) pGEM-Sp-RP
Spec RP (1863 .. 1919)
3500
500
1000
1500
2000
2500
3000
M13 fwd
T7 promoter
attL1
Gal4-AD
attL2
SP6 promoter
lac promoter
CAP binding site
AmpR promoter
SpecR
ori
f1 ori
pGEM-Gateway(Spec)-Gal4AD
3504 bp

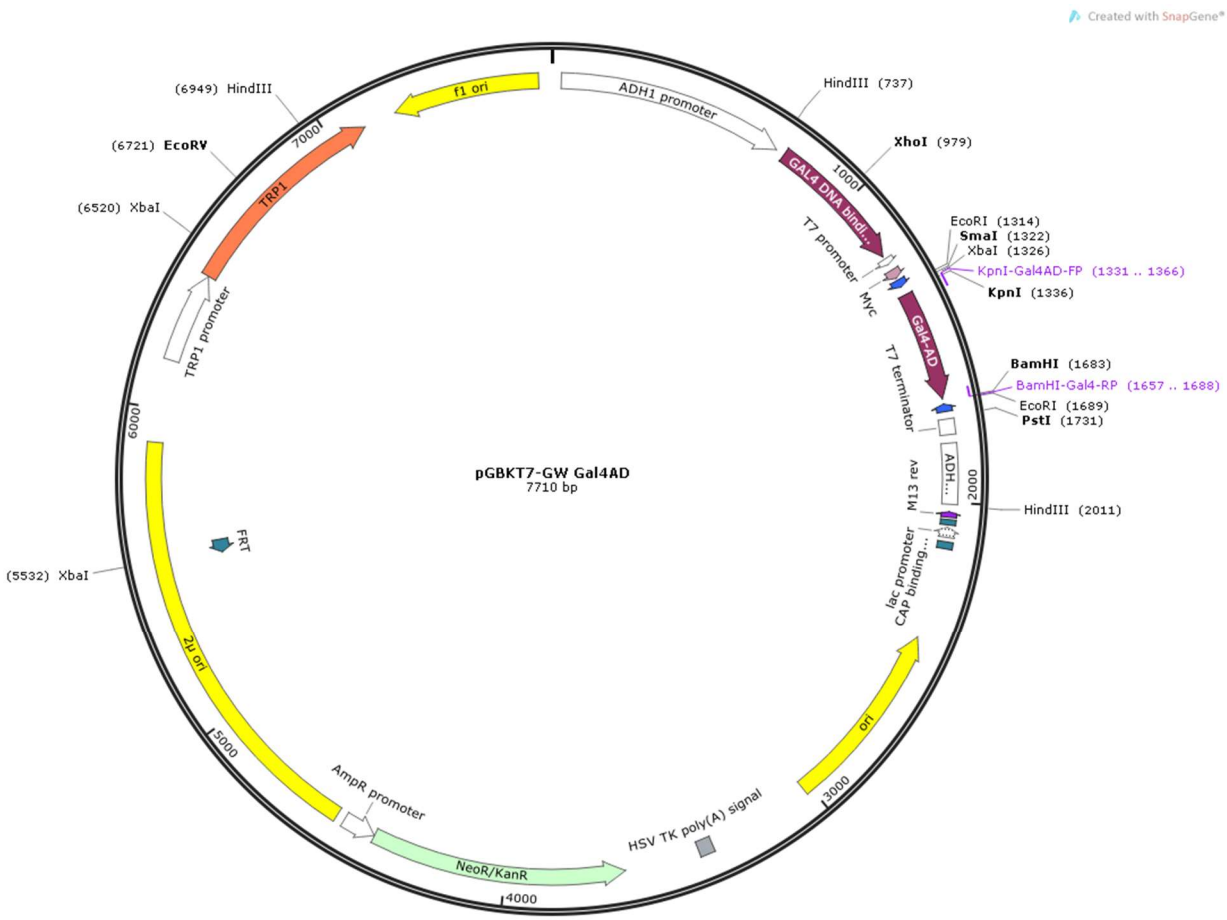

Created with SnapGene®
pGBKT7GW-MUG1
9633 bp
PvuI (0)
HindIII (737)
XhoI (979)
HincII (1039)
HincII (1192)
NdeI (1282)
NotI (1315)
NcoI (1330)
NruI (1342)
HincII (1775)
XhoI (1843)
SacI - SstI (1850)
NcoI (2016)
SmaI (2045)
HindIII (2122)
NheI (2142)
NruI (2167)
HindIII (2245)
PvuII (2382)
HindIII (2419)
SalI (2449)
HincII (2451)
HindIII (2750)
BamHI (2881)
ApaLI (3268)
AscI (3611)
PstI (3654)
NotI (3656)
HindIII (3934)
PvuII (4117)
DrdI (4401)
(4607) ApaLI
5000
HSV TK poly(A) signal
ori
ADH1 terminator
T7 terminator
4000
3000
2000
1000
MUG1
Myc
GAL4 DNA binding domain
ADH1 promoter
f1 ori
9000
TRP1
TRP1 promoter
8000
(8872) HindIII
(9070) ApaLI
(9316) DrdI
(8443) XbaI
(8064) ApaLI
(7984) DrdI
FRT
(7455) XbaI
2µ ori
7000
AmpR promoter
NeoR/KanR
6000
(6191) DrdI
(6110) PvuII

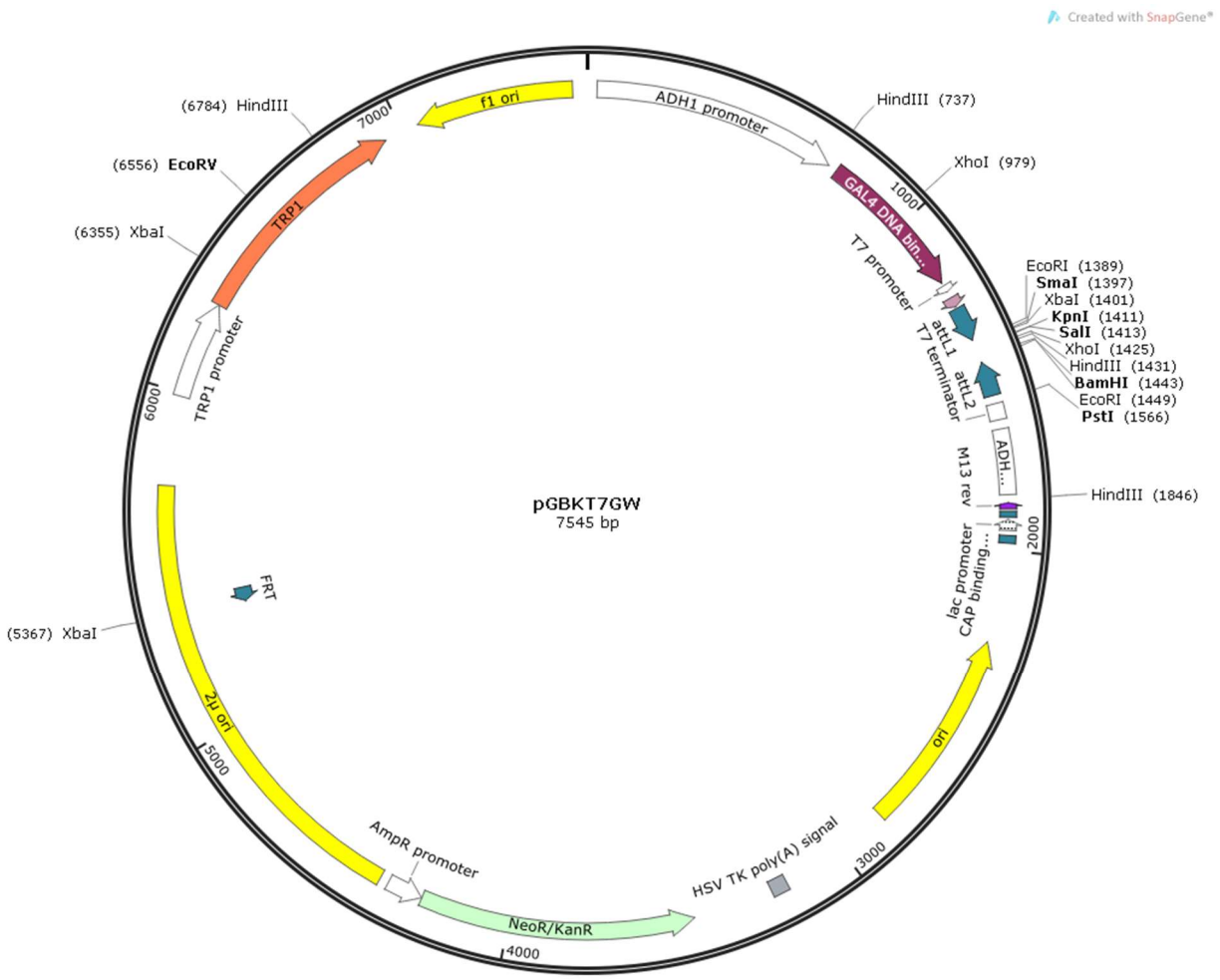

Created with SnapGene®
pGBKT7GW-MUG2
9627 bp
f1 ori
ADH1 promoter
HindIII (737)
XhoI (979)
GAL4 DNA...
T7 promoter
Myc
EcoRV (1370)
PstI (1477)
BamHI (1769)
EcoRI (2017)
HindIII (2031)
HindIII (2116)
SalI (2443)
PvuII (2612)
EcoRV (2854)
MUG2
T7 terminator
ADH1 terminator
AscI (3605)
PstI (3648)
HindIII (3928)
PvuII (4111)
ori
HSV TK poly(A) signal
NeoR/KanR
AmpR promoter
(6104) PvuII
2µ ori
(7449) XbaI
FRT
TRP1 promoter
TRP1
(8437) XbaI
(8638) EcoRV
(8866) HindIII
1000
2000
3000
4000
5000
6000
7000
8000
9000


Created with SnapGene®
pGBKT7GW-MUG3
9543 bp
f1 ori
ADH1 promoter
GAL4 DNA binding domain
Myc
MUG3
T7 terminator
ADH1 terminator
ori
HSV TK poly(A) signal
NeoR/KanR
AmpR promoter
2µ ori
TRP1 promoter
TRP1
FRT
HindIII (737)
XhoI (979)
EcoRV (1370)
SalI (1402)
EcoRV (1466)
SacI (1507)
HindIII (2044)
XbaI (2558)
HindIII (2577)
EcoRI (2691)
XhoI (2751)
EcoRV (2788)
HindIII (2797)
BamHI (2809)
HindIII (3002)
SacI (3218)
XbaI (3302)
AscI (3521)
PstI (3564)
HindIII (3844)
PvuII (4027)
(6020) PvuII
(7365) XbaI
(8353) XbaI
(8554) EcoRV
(8782) HindIII
1000
2000
3000
4000
5000
6000
7000
8000
9000

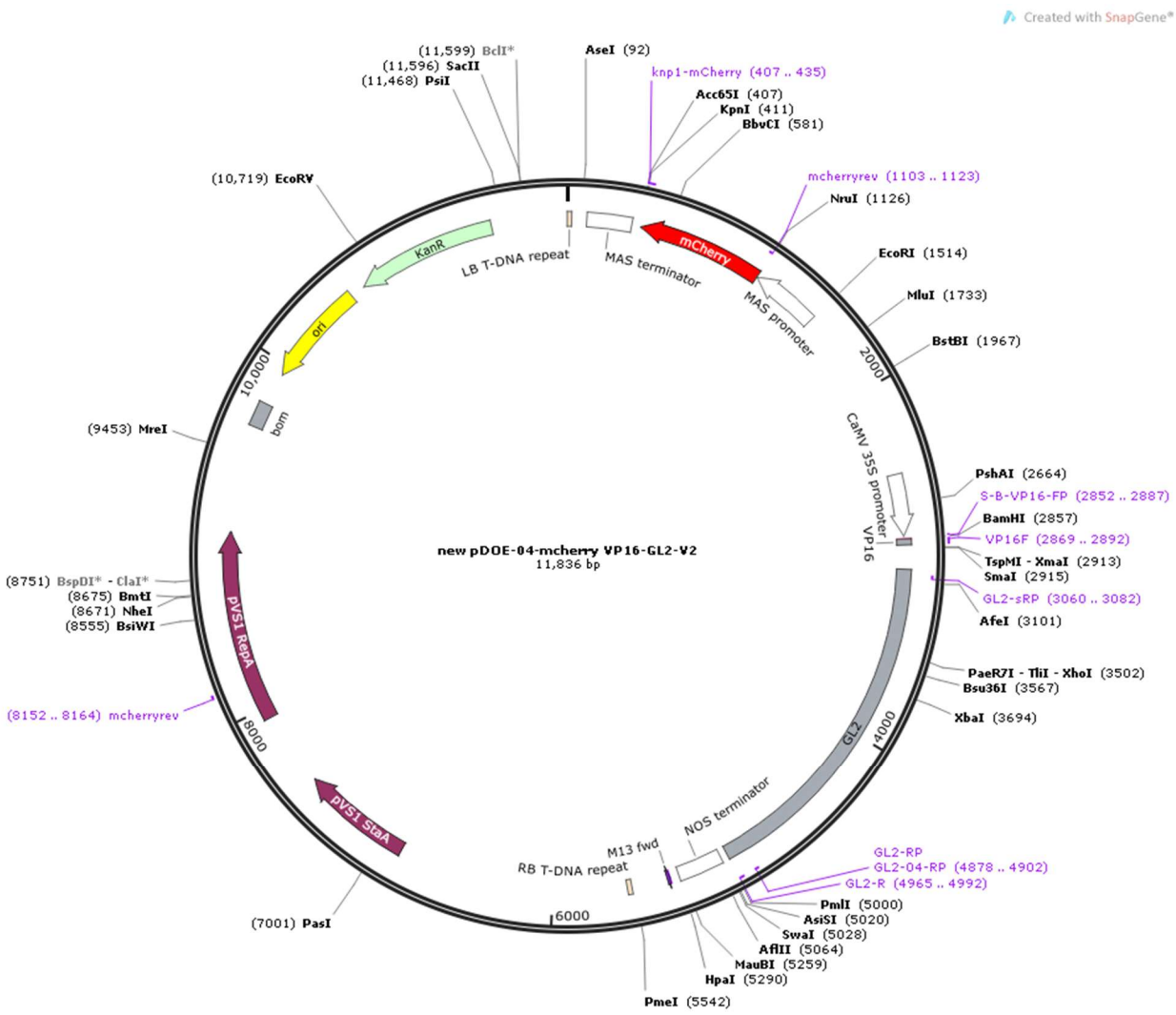

Created with SnapGene®
pGBKT7-GW MUG4
9141 bp
9000
f1 ori
ADH1 promoter
GAL4 DNA binding domain
1000
HindIII (737)
XhoI (979)
EcoRI (1314)
SmaI (1322)
XbaI-MUG4-FP (1326 .. 1359)
XbaI (1326)
SacI (1346)
Myc
MUG4
2000
KpnI (1823)
ScaI (2010)
PstI (2036)
HindIII (2133)
NcoI (2237)
HindIII (2295)
EcoRI (2343)
EcoRV (2377)
3000
PstI (2624)
SalI (2705)
PstI (2801)
HindIII (2909)
BamHI (3114)
BamHI-MUG4-RP (3089 .. 3119)
EcoRI (3120)
PstI (3162)
HindIII (3442)
T7 terminator
ADH1 terminator
4000
ori
5000
HSV TK poly(A) signal
(4882) AvrII
NeoR/KanR
6000
AmpR promoter
2µ ori
7000
(6626) SnaBI
(6963) XbaI
FRT
TRP1 promoter
8000
TRP1
(7719) SnaBI
(7732) BspEI*
(7827) PmlI
(7951) XbaI
(8152) EcoRV
(8380) HindIII

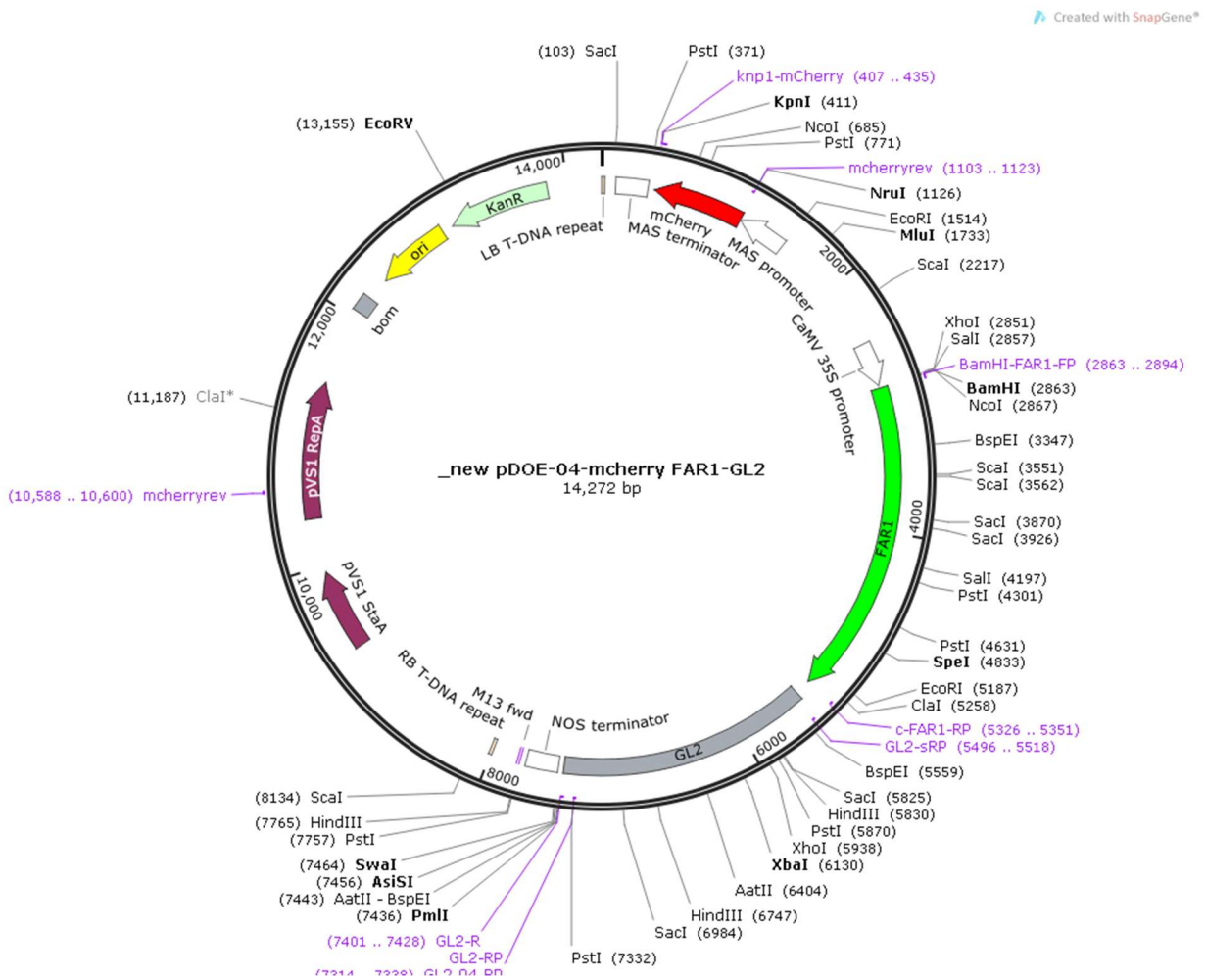

Created with SnapGene®
pGBKT7-GW FAR1
9849 bp
f1 ori
ADH1 promoter
GAL4 DNA binding domain
Myc
FAR1
T7 terminator
ADH1 terminator
ori
HSV TK poly(A) signal
NeoR/KanR
AmpR promoter
2µ ori
TRP1 promoter
TRP1
FRT
HindIII (737)
XhoI (979)
EcoRI (1314)
SmaI (1322)
XbaI (1326)
KpnI (1336)
mCh-FAR1-FP (1338 .. 1363)
SacI (2339)
SacI (2395)
SalI (2666)
PstI (2770)
PstI (3100)
StuI* (3643)
EcoRI (3656)
FAR1-GL-RP (3796 .. 3818)
BamHI (3822)
EcoRI (3828)
PstI (3870)
HindIII (4150)
(9088) HindIII
(8860) EcoRV
(8659) XbaI
(7671) XbaI
1000
2000
3000
4000
5000
6000
7000
8000
9000

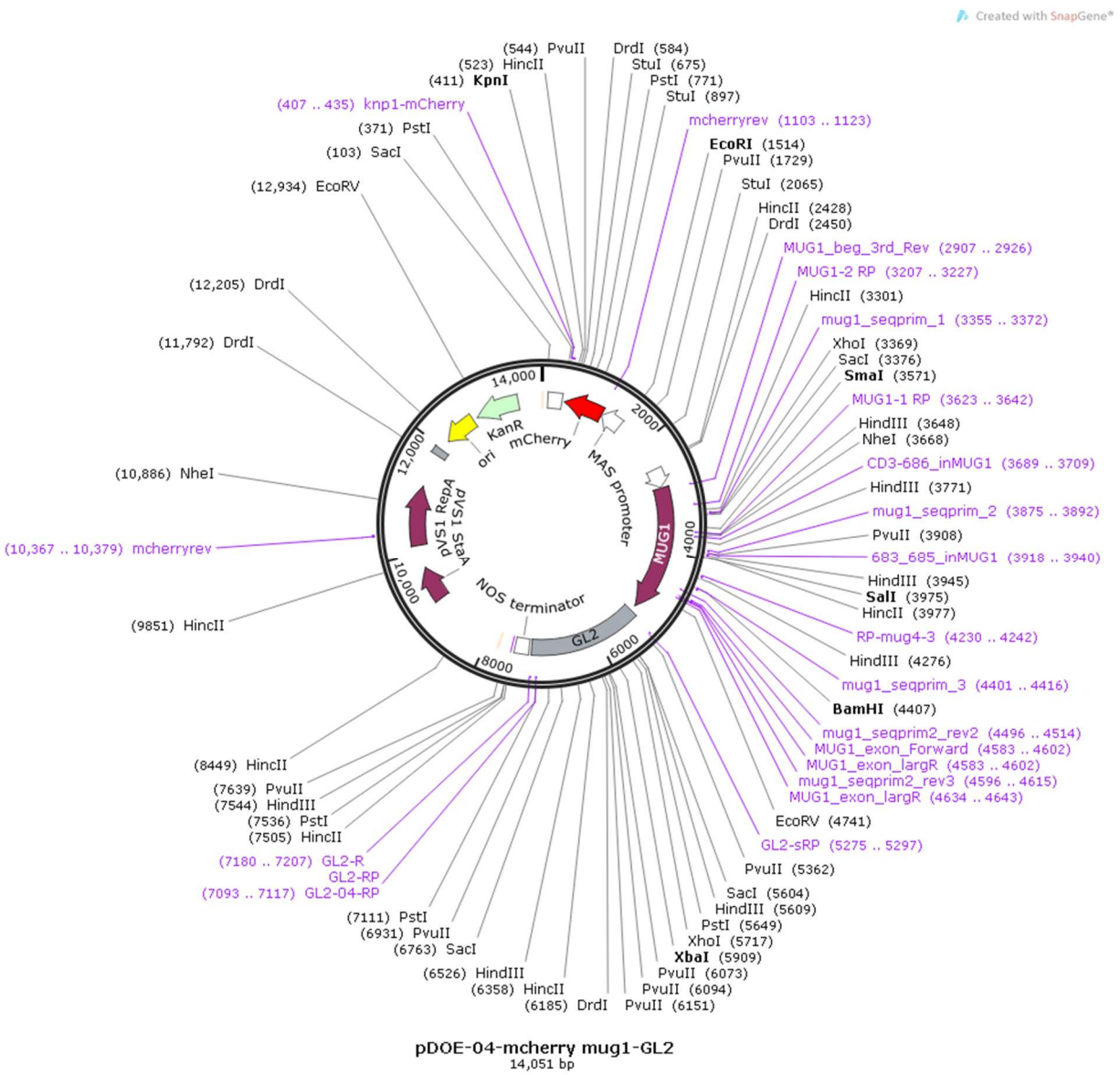

Created with SnapGene®
pGBKT7-GW Gal4AD
7710 bp
f1 ori
ADH1 promoter
HindIII (737)
XhoI (979)
1000
GAL4 DNA bindi...
T7 promoter
Myc
Gal4-AD
T7 terminator
ADH...
M13 rev
lac promoter
CAP binding...
ori
3000
HSV TK poly(A) signal
NeoR/KanR
4000
AmpR promoter
5000
2µ ori
FRT
(5532) XbaI
6000
TRP1 promoter
TRP1
7000
(6949) HindIII
(6721) EcoRV
(6520) XbaI
EcoRI (1314)
SmaI (1322)
XbaI (1326)
KpnI-Gal4AD-FP (1331 .. 1366)
KpnI (1336)
BamHI (1683)
BamHI-Gal4-RP (1657 .. 1688)
EcoRI (1689)
PstI (1731)
2000
HindIII (2011)


Created with SnapGene®
pGBKT7GW
7545 bp
f1 ori
ADH1 promoter
HindIII (737)
XhoI (979)
GAL4 DNA bin...
T7 promoter
attL1
T7 terminator
attL2
EcoRI (1389)
SmaI (1397)
XbaI (1401)
KpnI (1411)
SalI (1413)
XhoI (1425)
HindIII (1431)
BamHI (1443)
EcoRI (1449)
PstI (1566)
ADH...
M13 rev
HindIII (1846)
lac promoter
CAP binding...
ori
HSV TK poly(A) signal
AmpR promoter
NeoR/KanR
2µ ori
FRT
(5367) XbaI
TRP1 promoter
TRP1
(6784) HindIII
(6556) EcoRV
(6355) XbaI
1000
2000
3000
4000
5000
6000
7000

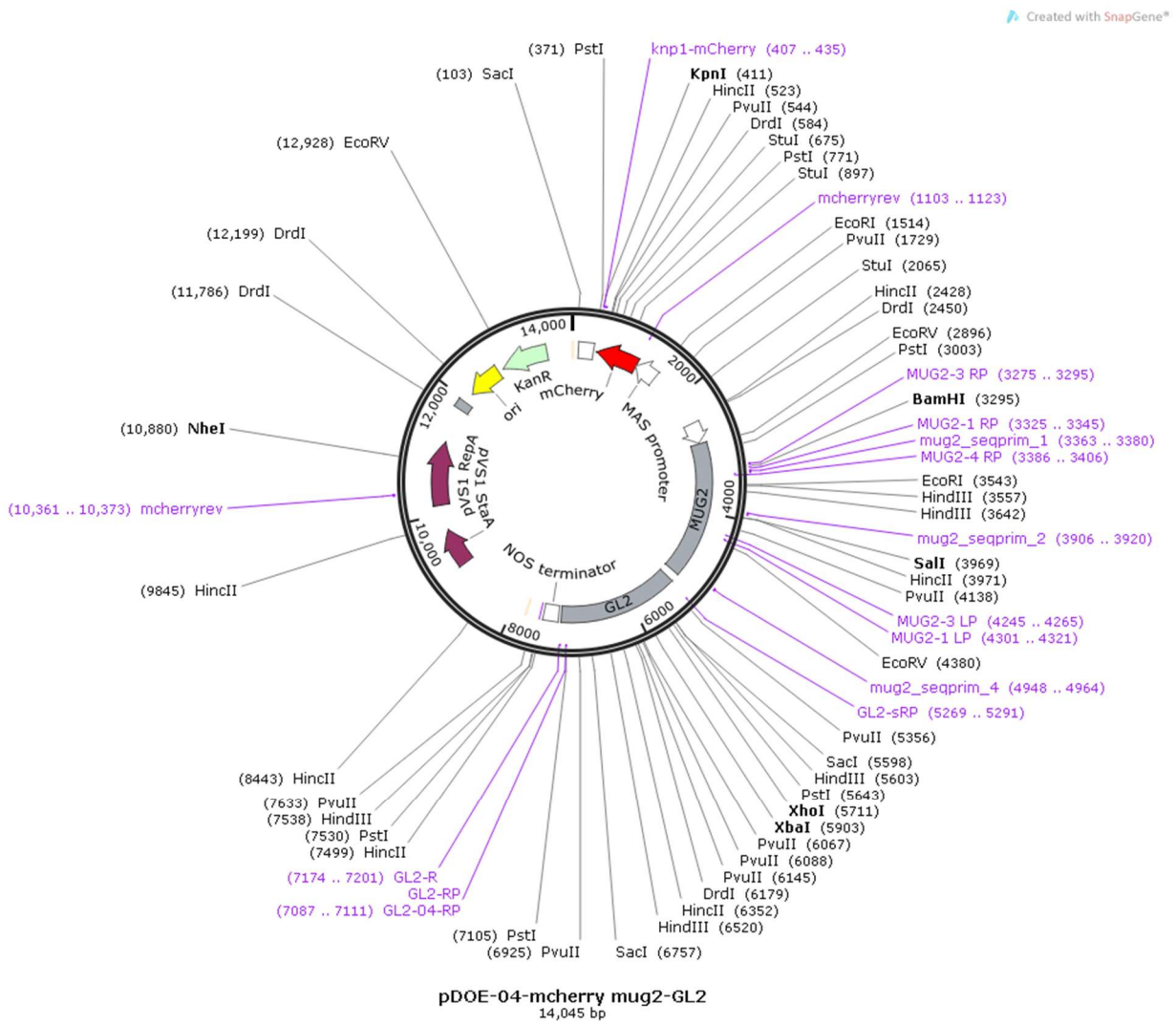

Created with SnapGene®
pDOE-04-mCherry
10,374 bp
10,000
KanR
LB T-DNA repeat
MAS terminator
MAS promoter
2000
CaMV 35S promoter
mCherry
Strep-Tag II
M13 fwd
NOS...
4000
RB T-DNA repeat
pVS1 StaA
6000
pVS1 RepA
8000
bom
ori


Created with SnapGene®
new pDOE-04-mcherry VP16-GL2-V2
11,836 bp
(11,599) BclI*
(11,596) SacII
(11,468) PsiI
AseI (92)
knp1-mCherry (407 .. 435)
Acc65I (407)
KpnI (411)
BbvCI (581)
mcherryrev (1103 .. 1123)
NruI (1126)
EcoRI (1514)
MluI (1733)
BstBI (1967)
2000
CaMV 35S promoter
VP16
PshAI (2664)
S-B-VP16-FP (2852 .. 2887)
BamHI (2857)
VP16F (2869 .. 2892)
TspMI - XmaI (2913)
SmaI (2915)
GL2-sRP (3060 .. 3082)
AfeI (3101)
PaeR7I - TliI - XhoI (3502)
Bsu36I (3567)
XbaI (3694)
4000
GL2
NOS terminator
M13 fwd
RB T-DNA repeat
GL2-RP
GL2-04-RP (4878 .. 4902)
GL2-R (4965 .. 4992)
PmlI (5000)
AsiSI (5020)
SwaI (5028)
AflII (5064)
MauBI (5259)
HpaI (5290)
PmeI (5542)
6000
(7001) PasI
pVS1 StbA
8000
(8152 .. 8164) mcherryrev
pVS1 RepA
(8751) BspDI* - ClaI*
(8675) BmtI
(8671) NheI
(8555) BsiWI
(9453) MreI
bom
10,000
ori
KanR
(10,719) EcoRV
LB T-DNA repeat
MAS terminator
mCherry
MAS promoter

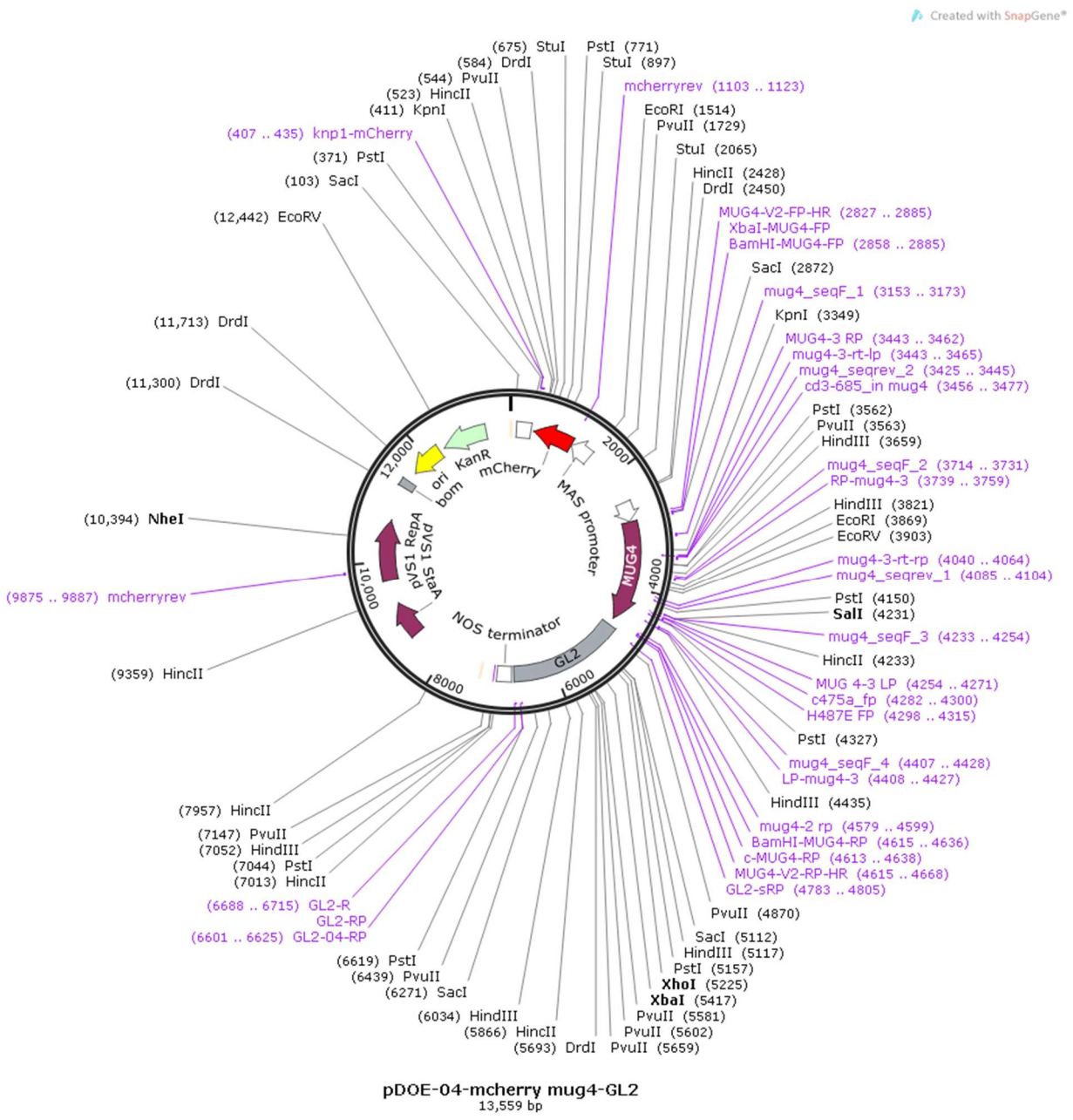

Created with SnapGene®
_new pDOE-04-mcherry FAR1-GL2
14,272 bp
(103) SacI
PstI (371)
knp1-mCherry (407 .. 435)
KpnI (411)
NcoI (685)
PstI (771)
mcherryrev (1103 .. 1123)
NruI (1126)
EcoRI (1514)
MluI (1733)
ScaI (2217)
XhoI (2851)
SalI (2857)
BamHI-FAR1-FP (2863 .. 2894)
BamHI (2863)
NcoI (2867)
BspEI (3347)
ScaI (3551)
ScaI (3562)
SacI (3870)
SacI (3926)
SalI (4197)
PstI (4301)
PstI (4631)
SpeI (4833)
EcoRI (5187)
ClaI (5258)
c-FAR1-RP (5326 .. 5351)
GL2-sRP (5496 .. 5518)
BspEI (5559)
SacI (5825)
HindIII (5830)
PstI (5870)
XhoI (5938)
XbaI (6130)
AatII (6404)
HindIII (6747)
SacI (6984)
PstI (7332)
(7401 .. 7428) GL2-R
GL2-RP
(7436) PmlI
(7443) AatII - BspEI
(7456) AsiSI
(7464) SwaI
(7757) PstI
(7765) HindIII
(8134) ScaI
(10,588 .. 10,600) mcherryrev
(11,187) ClaI*
(13,155) EcoRV
14,000
2000
4000
6000
8000
10,000
12,000
KanR
ori
LB T-DNA repeat
mCherry
MAS terminator
MAS promoter
CaMV 35S promoter
FAR1
GL2
NOS terminator
M13 fwd
RB T-DNA repeat
pVS1 StaA
pVS1 RepA
bom



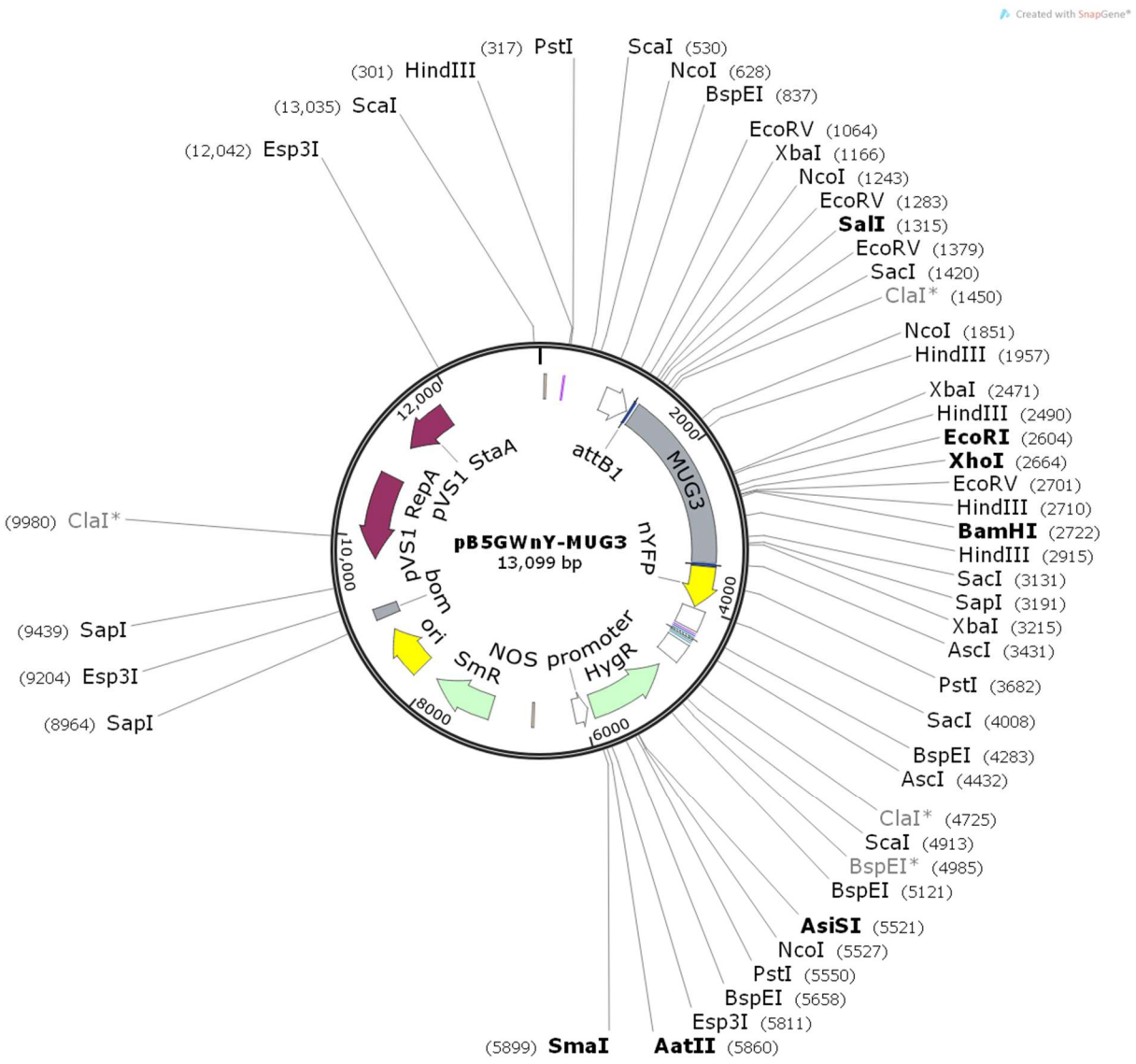
pDOE-04-mcherry mug1-GL2
14,051 bp

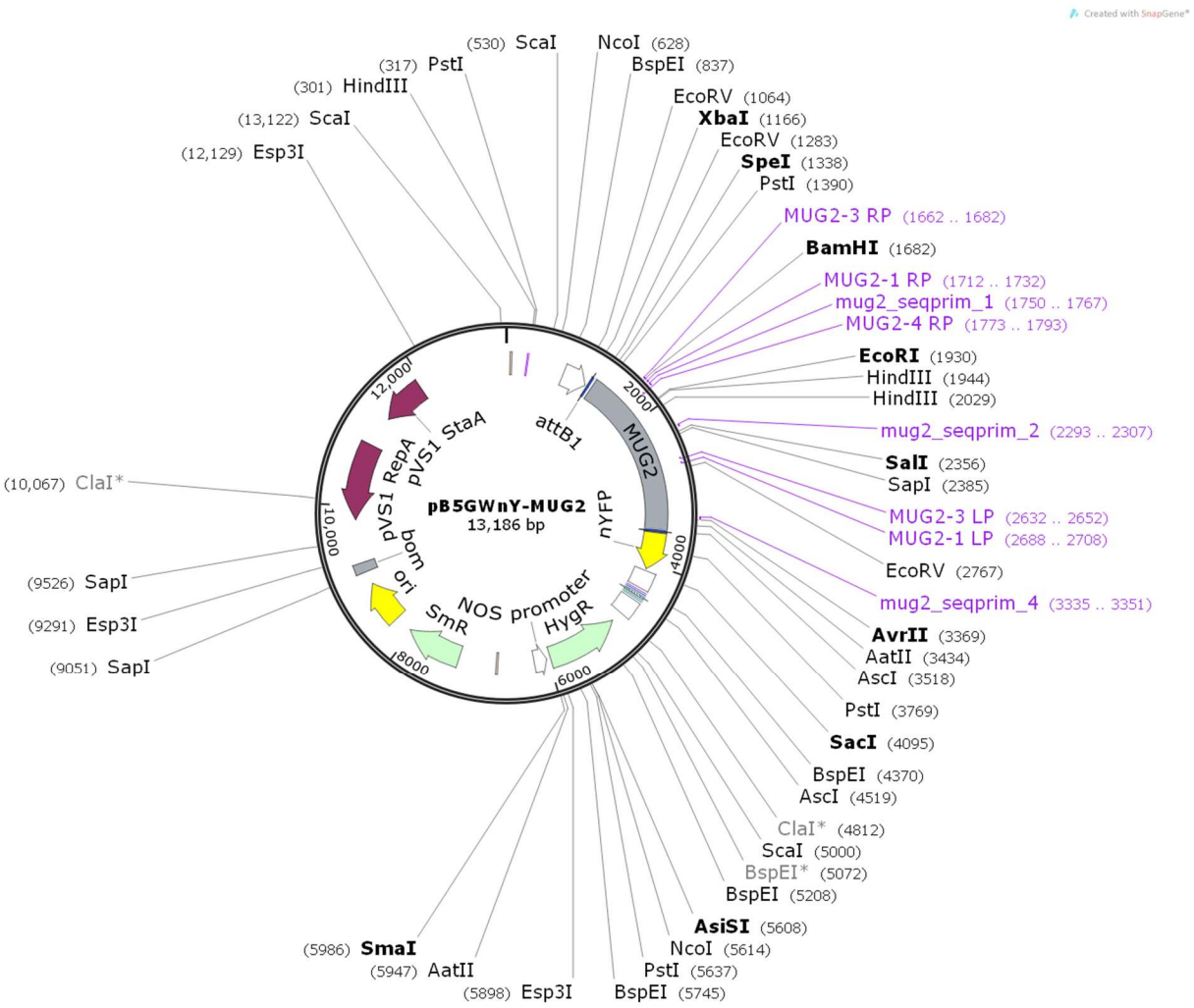


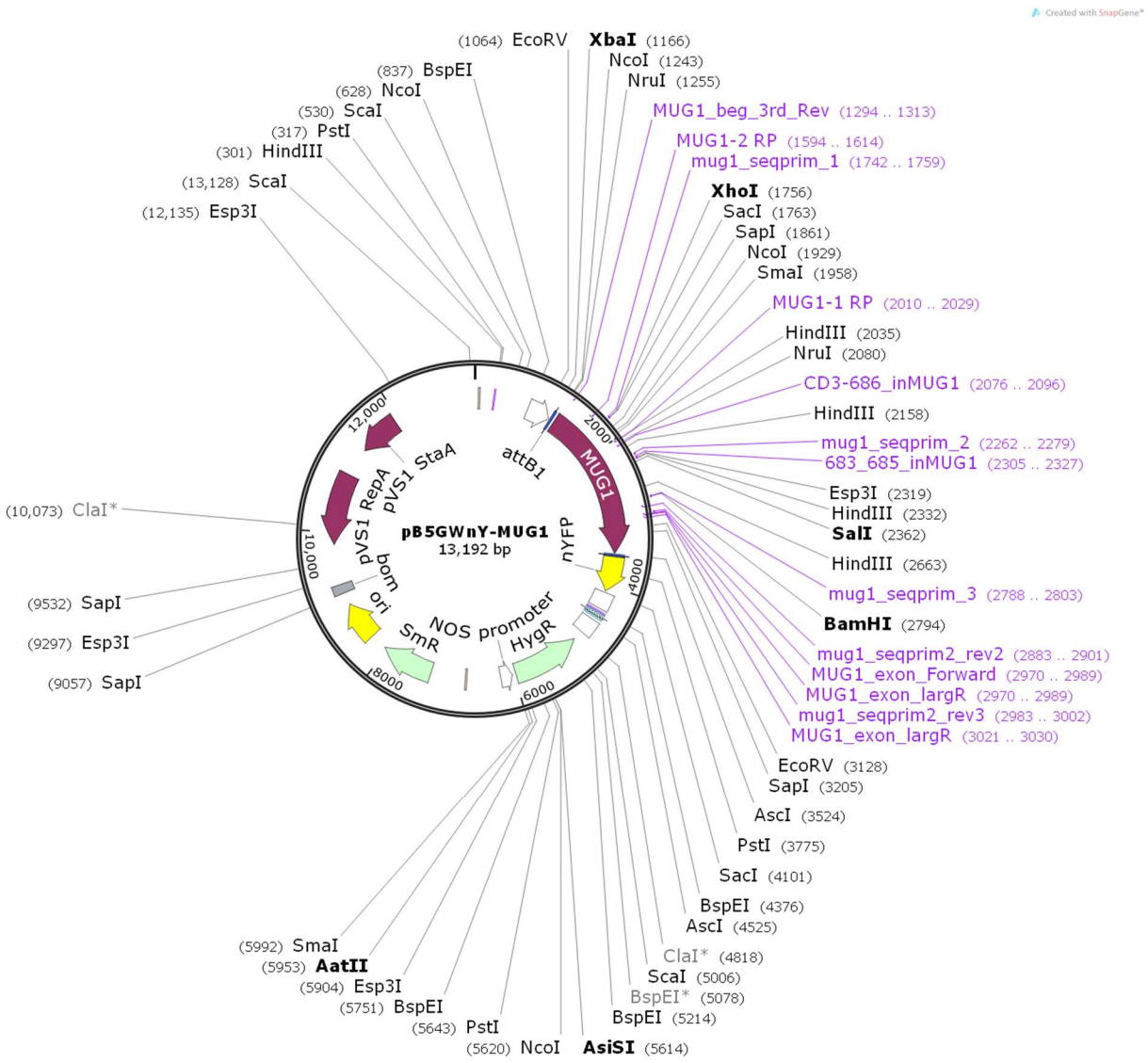
pDOE-04-mcherry mug2-GL2
14,045 bp

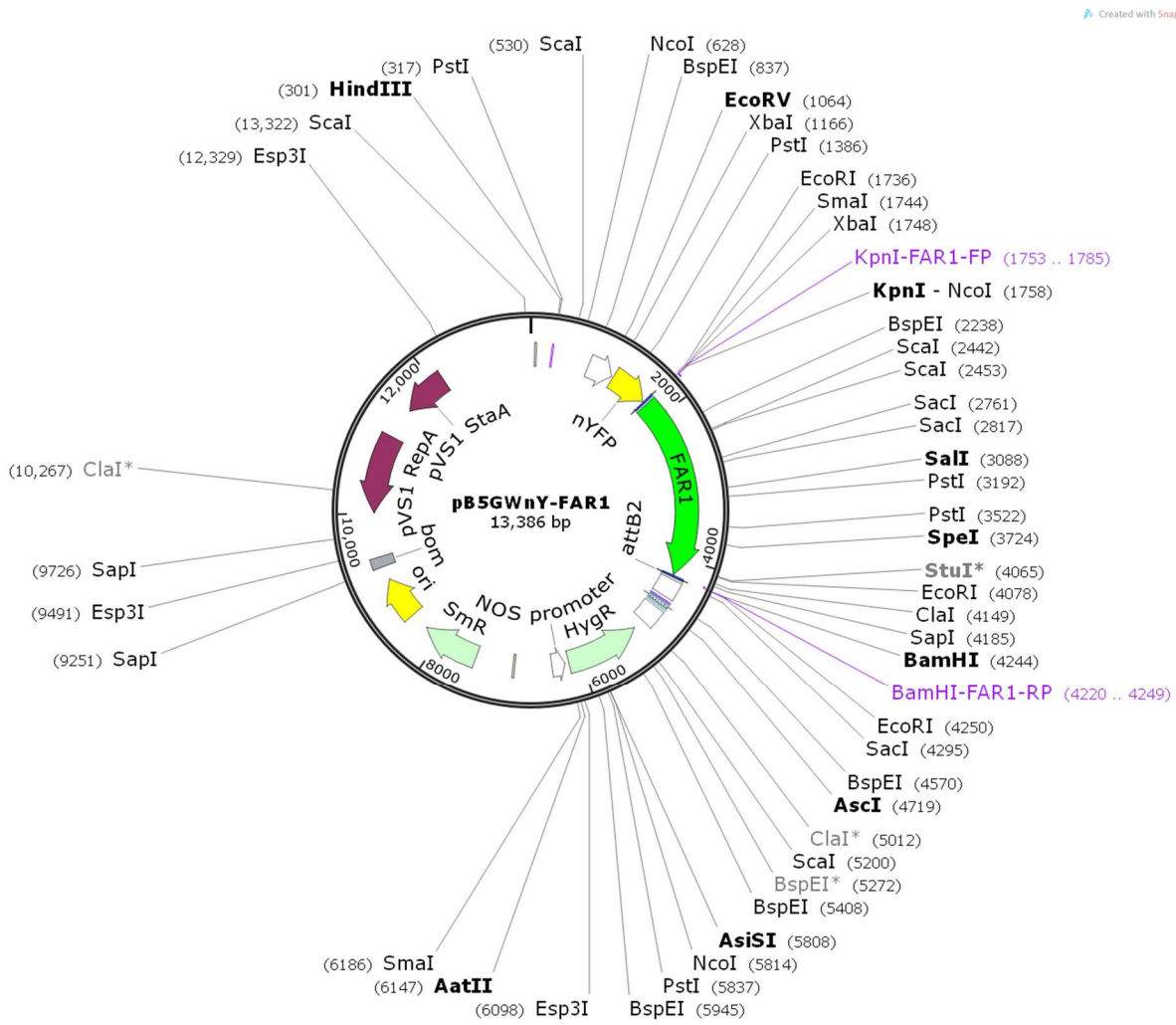


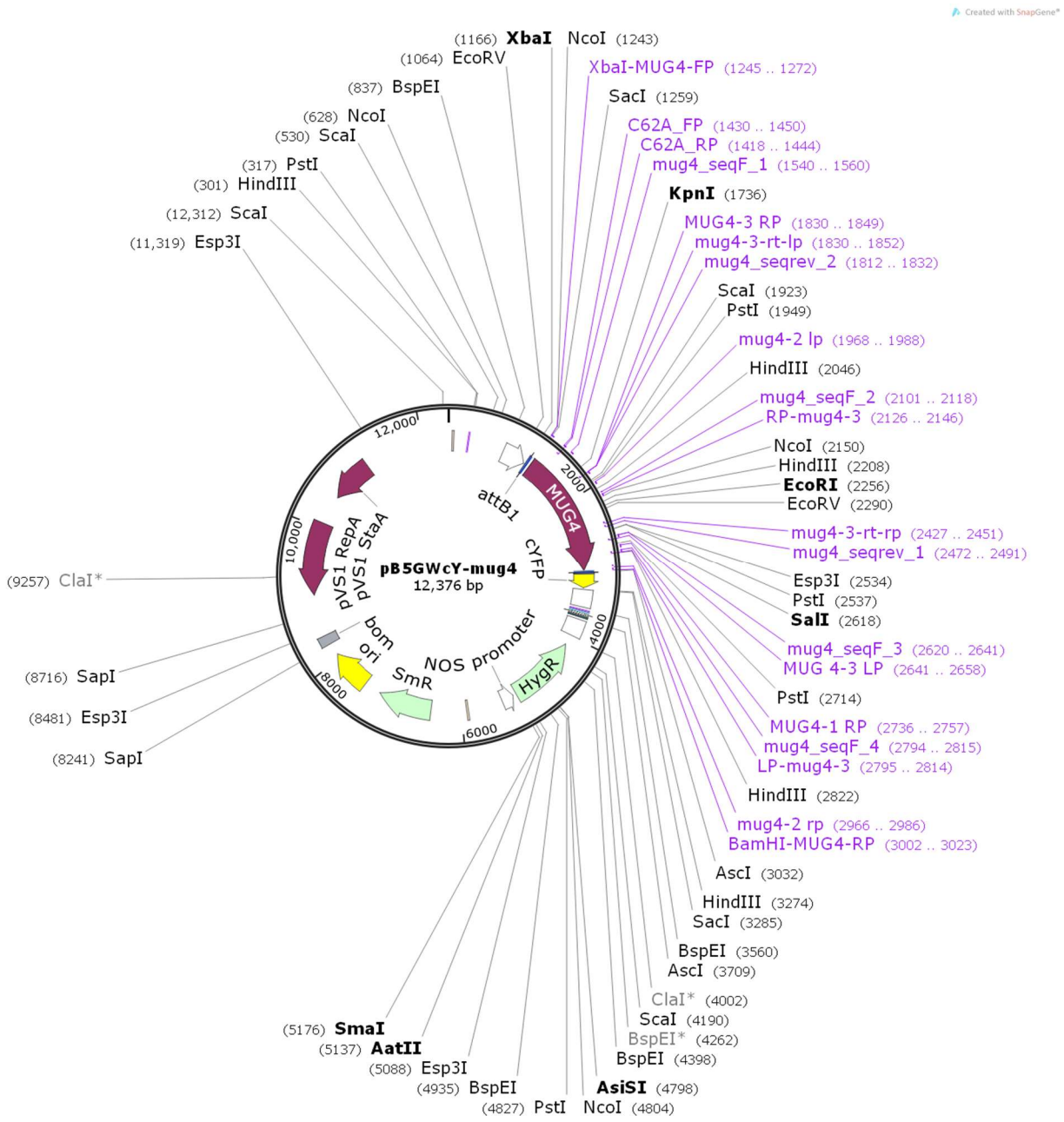
pDOE-04-mcherry mug4-GL2
13,559 bp

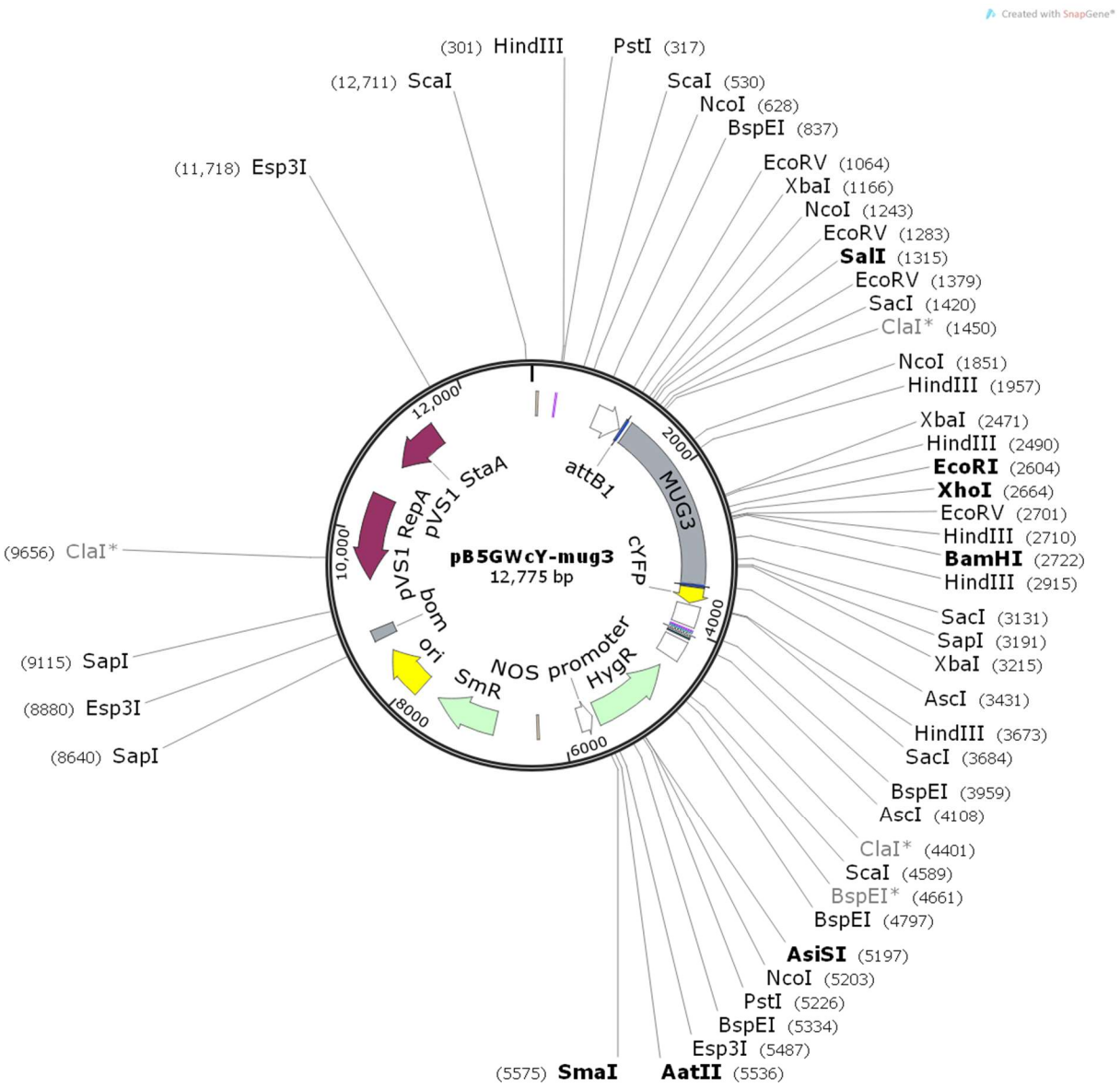

Created with SnapGene®
pB5GWnY-MUG3
13,099 bp
attB1
MUG3
nYFP
NOS promoter
HygR
SmR
ori
bom
pVS1 RepA
pVS1 StaA
2000
4000
6000
8000
10,000
12,000
(317) PstI
(301) HindIII
(13,035) ScaI
(12,042) Esp3I
(9980) ClaI*
(9439) SapI
(9204) Esp3I
(8964) SapI
(5899) SmaI
ScaI (530)
NcoI (628)
BspEI (837)
EcoRV (1064)
XbaI (1166)
NcoI (1243)
EcoRV (1283)
SalI (1315)
EcoRV (1379)
SacI (1420)
ClaI* (1450)
NcoI (1851)
HindIII (1957)
XbaI (2471)
HindIII (2490)
EcoRI (2604)
XhoI (2664)
EcoRV (2701)
HindIII (2710)
BamHI (2722)
HindIII (2915)
SacI (3131)
SapI (3191)
XbaI (3215)
AscI (3431)
PstI (3682)
SacI (4008)
BspEI (4283)
AscI (4432)
ClaI* (4725)
ScaI (4913)
BspEI* (4985)
BspEI (5121)
AsiSI (5521)
NcoI (5527)
PstI (5550)
BspEI (5658)
Esp3I (5811)
AatII (5860)

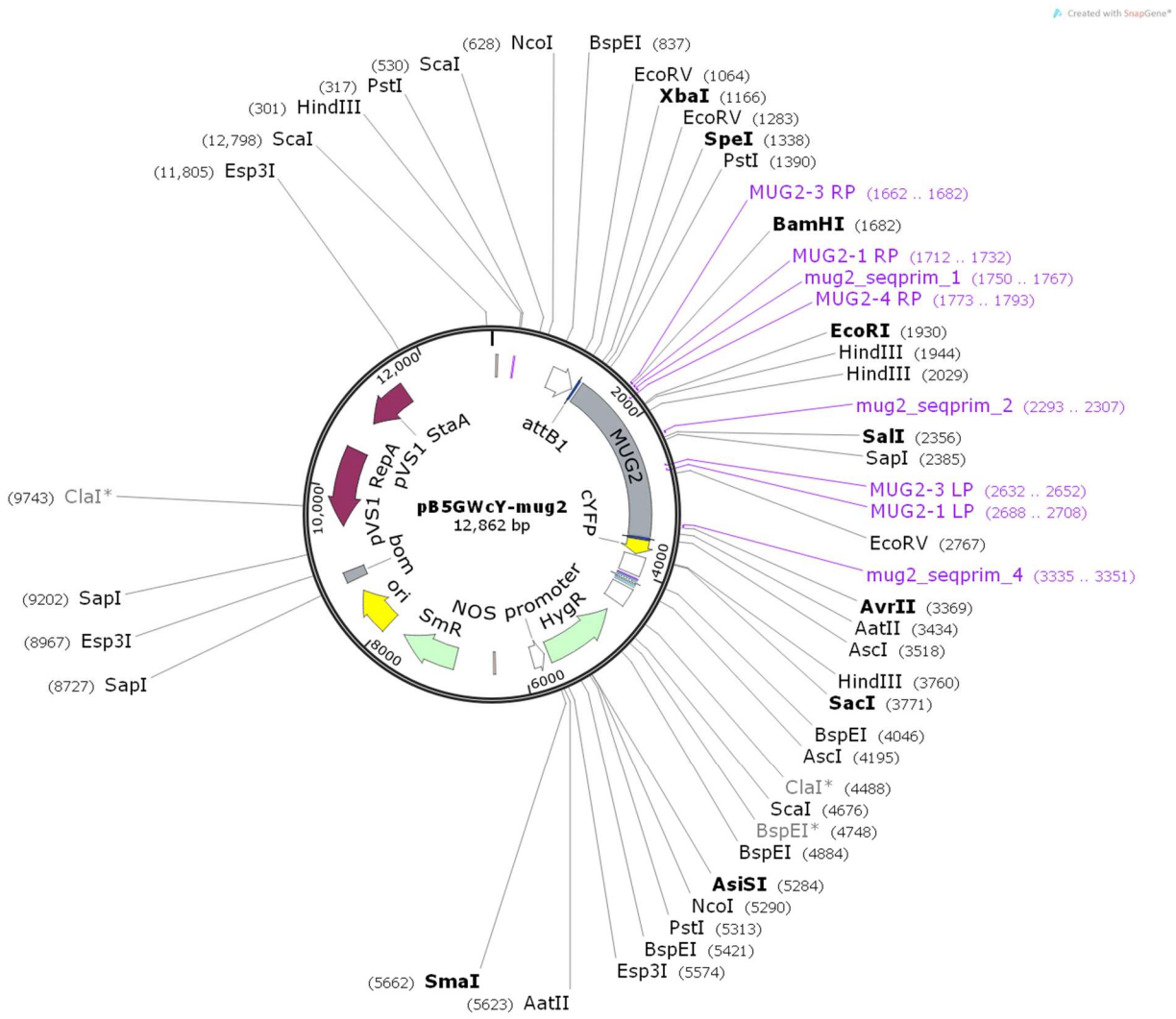

Created with SnapGene®
pB5GWnY-MUG2
13,186 bp
(530) ScaI
NcoI (628)
BspEI (837)
(317) PstI
(301) HindIII
EcoRV (1064)
XbaI (1166)
(13,122) ScaI
EcoRV (1283)
SpeI (1338)
(12,129) Esp3I
PstI (1390)
MUG2-3 RP (1662 .. 1682)
BamHI (1682)
MUG2-1 RP (1712 .. 1732)
mug2_seqprim_1 (1750 .. 1767)
MUG2-4 RP (1773 .. 1793)
EcoRI (1930)
HindIII (1944)
HindIII (2029)
mug2_seqprim_2 (2293 .. 2307)
SalI (2356)
SapI (2385)
MUG2-3 LP (2632 .. 2652)
MUG2-1 LP (2688 .. 2708)
EcoRV (2767)
mug2_seqprim_4 (3335 .. 3351)
AvrII (3369)
AatII (3434)
AscI (3518)
PstI (3769)
SacI (4095)
BspEI (4370)
AscI (4519)
ClaI* (4812)
ScaI (5000)
BspEI* (5072)
BspEI (5208)
AsiSI (5608)
NcoI (5614)
PstI (5637)
BspEI (5745)
(5898) Esp3I
(5947) AatII
(5986) SmaI
(9051) SapI
(9291) Esp3I
(9526) SapI
(10,067) ClaI*
12,000
2000
4000
6000
8000
10,000
attB1
MUG2
nYFP
NOS promoter
HygR
SmR
ori
bom
pVS1 RepA
pVS1 StaA

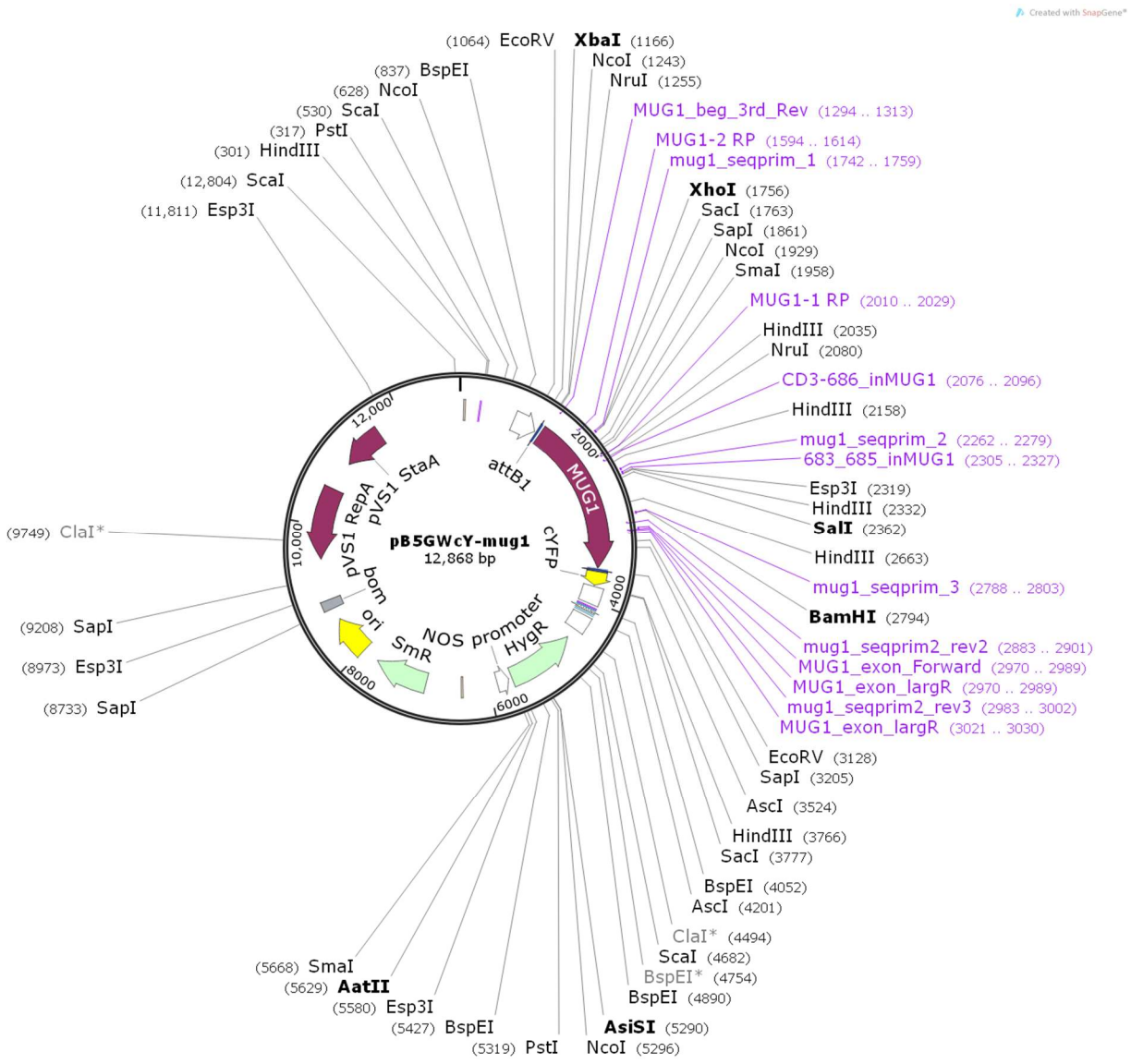

Created with SnapGene®
pB5GWnY-MUG1
13,192 bp
12,000
2000
4000
6000
8000
10,000
attB1
MUG1
nYFP
NOS promoter
HygR
SmR
ori
bom
pVS1 RepA
pVS1 StaA
(1064) EcoRV
XbaI (1166)
NcoI (1243)
NruI (1255)
MUG1_beg_3rd_Rev (1294 .. 1313)
MUG1-2 RP (1594 .. 1614)
mug1_seqprim_1 (1742 .. 1759)
XhoI (1756)
SacI (1763)
SapI (1861)
NcoI (1929)
SmaI (1958)
MUG1-1 RP (2010 .. 2029)
HindIII (2035)
NruI (2080)
CD3-686_inMUG1 (2076 .. 2096)
HindIII (2158)
mug1_seqprim_2 (2262 .. 2279)
683_685_inMUG1 (2305 .. 2327)
Esp3I (2319)
HindIII (2332)
SalI (2362)
HindIII (2663)
mug1_seqprim_3 (2788 .. 2803)
BamHI (2794)
mug1_seqprim2_rev2 (2883 .. 2901)
MUG1_exon_Forward (2970 .. 2989)
MUG1_exon_largR (2970 .. 2989)
mug1_seqprim2_rev3 (2983 .. 3002)
MUG1_exon_largR (3021 .. 3030)
EcoRV (3128)
SapI (3205)
AscI (3524)
PstI (3775)
SacI (4101)
BspEI (4376)
AscI (4525)
ClaI* (4818)
ScaI (5006)
BspEI* (5078)
BspEI (5214)
AsiSI (5614)
(5620) NcoI
(5643) PstI
(5751) BspEI
(5904) Esp3I
(5953) AatII
(5992) SmaI
(9057) SapI
(9297) Esp3I
(9532) SapI
(10,073) ClaI*
(12,135) Esp3I
(13,128) ScaI
(301) HindIII
(317) PstI
(530) ScaI
(628) NcoI
(837) BspEI

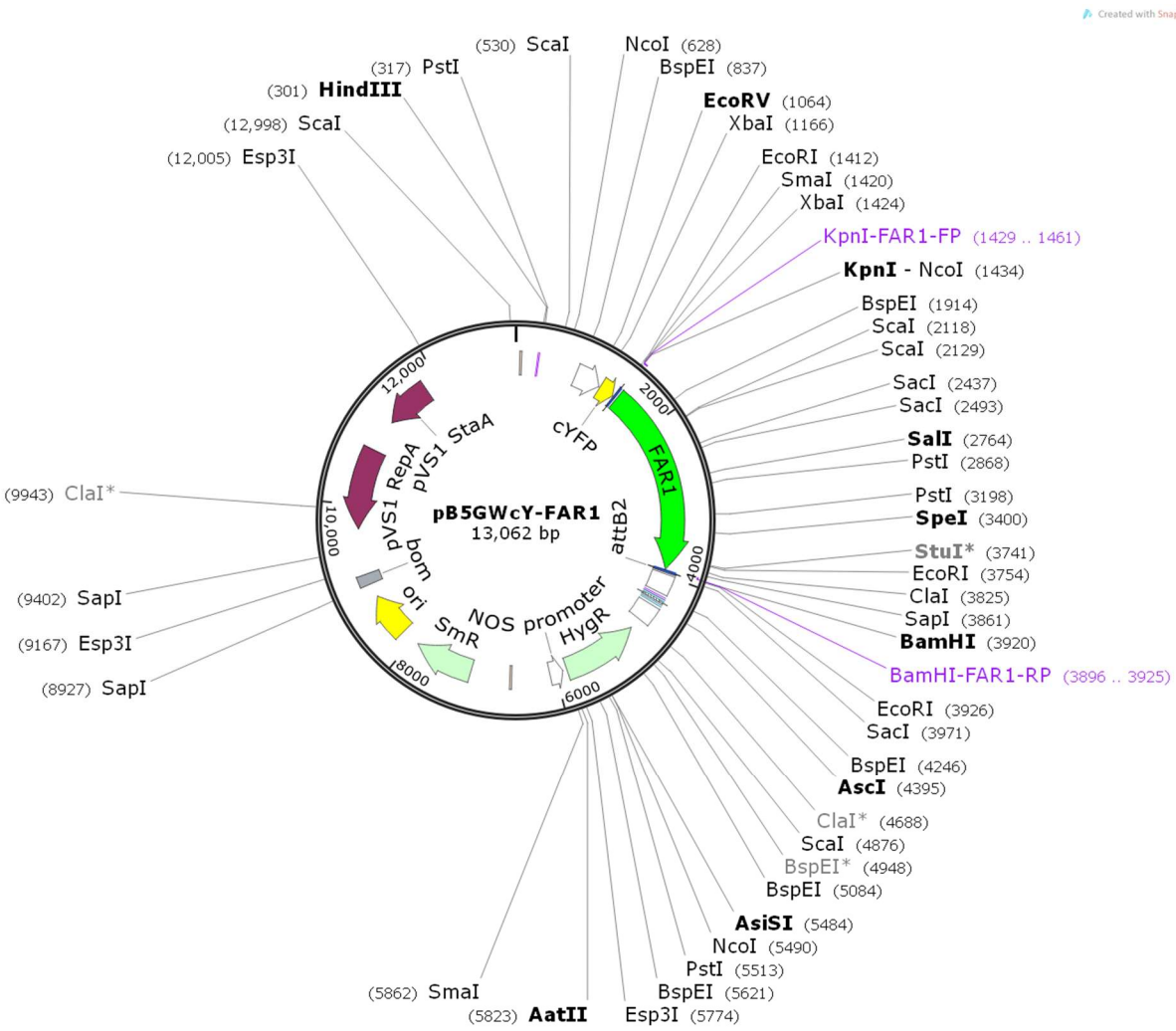

Created with SnapGene®
pB5GWnY-FAR1
13,386 bp
(530) ScaI
NcoI (628)
BspEI (837)
(317) PstI
(301) HindIII
EcoRV (1064)
XbaI (1166)
(13,322) ScaI
PstI (1386)
(12,329) Esp3I
EcoRI (1736)
SmaI (1744)
XbaI (1748)
KpnI-FAR1-FP (1753 .. 1785)
KpnI - NcoI (1758)
BspEI (2238)
ScaI (2442)
ScaI (2453)
SacI (2761)
SacI (2817)
SalI (3088)
PstI (3192)
PstI (3522)
SpeI (3724)
StuI* (4065)
EcoRI (4078)
ClaI (4149)
SapI (4185)
BamHI (4244)
BamHI-FAR1-RP (4220 .. 4249)
EcoRI (4250)
SacI (4295)
BspEI (4570)
AscI (4719)
ClaI* (5012)
ScaI (5200)
BspEI* (5272)
BspEI (5408)
AsiSI (5808)
NcoI (5814)
PstI (5837)
BspEI (5945)
(6186) SmaI
(6147) AatII
(6098) Esp3I
(10,267) ClaI*
(9726) SapI
(9491) Esp3I
(9251) SapI
12,000
2000
4000
6000
8000
10000
pVS1 StaA
pVS1 RepA
nYFP
FAR1
attB2
NOS promoter
HygR
SmR
bom
tio

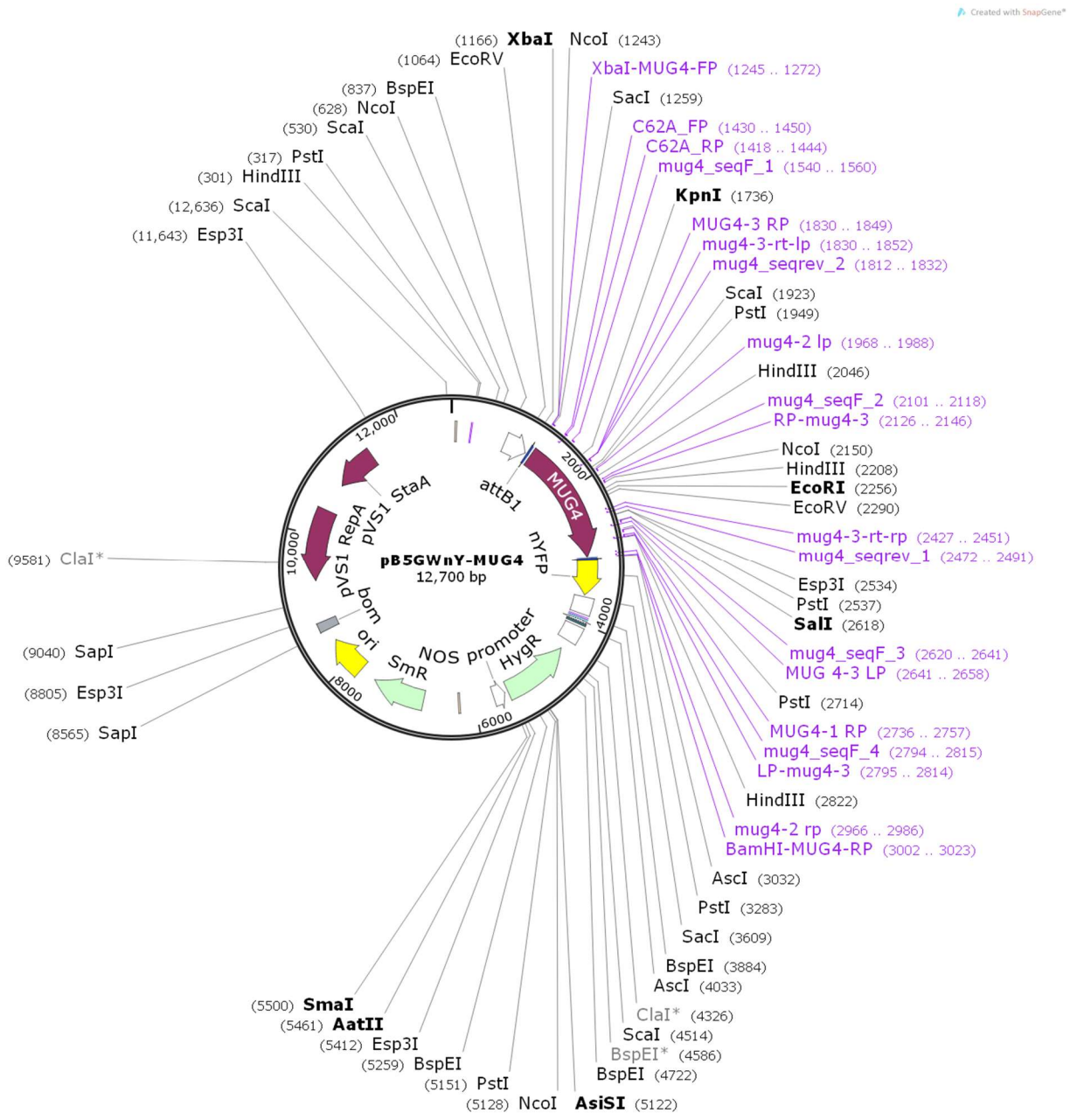

Created with SnapGene®
pB5GWcY-mug4
12,376 bp
12,000
2000
4000
6000
8000
10,000
attB1
MUG4
cYFP
NOS promoter
HygR
SmR
bom
ori
pVS1 RepA
pVS1 StaA
(1166) XbaI
NcoI (1243)
(1064) EcoRV
XbaI-MUG4-FP (1245 .. 1272)
(837) BspEI
SacI (1259)
(628) NcoI
(530) ScaI
C62A_FP (1430 .. 1450)
C62A_RP (1418 .. 1444)
(317) PstI
(301) HindIII
mug4_seqF_1 (1540 .. 1560)
KpnI (1736)
(12,312) ScaI
MUG4-3 RP (1830 .. 1849)
(11,319) Esp3I
mug4-3-rt-lp (1830 .. 1852)
mug4_seqrev_2 (1812 .. 1832)
ScaI (1923)
PstI (1949)
mug4-2 lp (1968 .. 1988)
HindIII (2046)
mug4_seqF_2 (2101 .. 2118)
RP-mug4-3 (2126 .. 2146)
NcoI (2150)
HindIII (2208)
EcoRI (2256)
EcoRV (2290)
mug4-3-rt-rp (2427 .. 2451)
mug4_seqrev_1 (2472 .. 2491)
(9257) ClaI*
Esp3I (2534)
PstI (2537)
SalI (2618)
mug4_seqF_3 (2620 .. 2641)
MUG 4-3 LP (2641 .. 2658)
(8716) SapI
PstI (2714)
(8481) Esp3I
MUG4-1 RP (2736 .. 2757)
mug4_seqF_4 (2794 .. 2815)
(8241) SapI
LP-mug4-3 (2795 .. 2814)
HindIII (2822)
mug4-2 rp (2966 .. 2986)
BamHI-MUG4-RP (3002 .. 3023)
AscI (3032)
HindIII (3274)
SacI (3285)
BspEI (3560)
AscI (3709)
ClaI* (4002)
ScaI (4190)
BspEI* (4262)
BspEI (4398)
(5176) SmaI
(5137) AatII
(5088) Esp3I
(4935) BspEI
(4827) PstI
AsiSI (4798)
NcoI (4804)


Created with SnapGene®
(301) HindIII
PstI (317)
(12,711) ScaI
ScaI (530)
NcoI (628)
BspEI (837)
(11,718) Esp3I
EcoRV (1064)
XbaI (1166)
NcoI (1243)
EcoRV (1283)
SalI (1315)
EcoRV (1379)
SacI (1420)
ClaI* (1450)
NcoI (1851)
HindIII (1957)
XbaI (2471)
HindIII (2490)
EcoRI (2604)
XhoI (2664)
EcoRV (2701)
HindIII (2710)
BamHI (2722)
HindIII (2915)
SacI (3131)
SapI (3191)
XbaI (3215)
AscI (3431)
HindIII (3673)
SacI (3684)
BspEI (3959)
AscI (4108)
ClaI* (4401)
ScaI (4589)
BspEI* (4661)
BspEI (4797)
AsiSI (5197)
NcoI (5203)
PstI (5226)
BspEI (5334)
Esp3I (5487)
(5575) SmaI
AatII (5536)
(9656) ClaI*
(9115) SapI
(8880) Esp3I
(8640) SapI
pB5GWcY-mug3
12,775 bp
12,000
2000
4000
6000
8000
10,000
pVS1 StaA
attB1
MUG3
cYFP
pVS1 RepA
bom ori
NOS promoter
HygR
SmR


Created with SnapGene®
pB5GWcY-mug2
12,862 bp
12,000
2000
4000
6000
8000
10,000
attB1
MUG2
cYFP
NOS promoter
HygR
SmR
ori
bom
pVS1 RepA
pVS1 StaA
(628) NcoI
BspEI (837)
(530) ScaI
EcoRV (1064)
(317) PstI
XbaI (1166)
(301) HindIII
EcoRV (1283)
(12,798) ScaI
SpeI (1338)
(11,805) Esp3I
PstI (1390)
MUG2-3 RP (1662 .. 1682)
BamHI (1682)
MUG2-1 RP (1712 .. 1732)
mug2_seqprim_1 (1750 .. 1767)
MUG2-4 RP (1773 .. 1793)
EcoRI (1930)
HindIII (1944)
HindIII (2029)
mug2_seqprim_2 (2293 .. 2307)
SalI (2356)
SapI (2385)
MUG2-3 LP (2632 .. 2652)
MUG2-1 LP (2688 .. 2708)
EcoRV (2767)
mug2_seqprim_4 (3335 .. 3351)
AvrII (3369)
AatII (3434)
AscI (3518)
HindIII (3760)
SacI (3771)
BspEI (4046)
AscI (4195)
ClaI* (4488)
ScaI (4676)
BspEI* (4748)
BspEI (4884)
AsiSI (5284)
NcoI (5290)
PstI (5313)
BspEI (5421)
Esp3I (5574)
(5662) SmaI
(5623) AatII
(8727) SapI
(8967) Esp3I
(9202) SapI
(9743) ClaI*


Created with SnapGene®
pB5GWcY-mug1
12,868 bp
MUG1
attB1
cYFP
NOS promoter
HygR
SmR
ori
bom
pVS1 RepA
pVS1 StaA
2000
4000
6000
8000
10,000
12,000
(1064) EcoRV
XbaI (1166)
NcoI (1243)
NruI (1255)
MUG1_beg_3rd_Rev (1294 .. 1313)
MUG1-2 RP (1594 .. 1614)
mug1_seqprim_1 (1742 .. 1759)
XhoI (1756)
SacI (1763)
SapI (1861)
NcoI (1929)
SmaI (1958)
MUG1-1 RP (2010 .. 2029)
HindIII (2035)
NruI (2080)
CD3-686_inMUG1 (2076 .. 2096)
HindIII (2158)
mug1_seqprim_2 (2262 .. 2279)
683_685_inMUG1 (2305 .. 2327)
Esp3I (2319)
HindIII (2332)
SalI (2362)
HindIII (2663)
mug1_seqprim_3 (2788 .. 2803)
BamHI (2794)
mug1_seqprim2_rev2 (2883 .. 2901)
MUG1_exon_Forward (2970 .. 2989)
MUG1_exon_largR (2970 .. 2989)
mug1_seqprim2_rev3 (2983 .. 3002)
MUG1_exon_largR (3021 .. 3030)
EcoRV (3128)
SapI (3205)
AscI (3524)
HindIII (3766)
SacI (3777)
BspEI (4052)
AscI (4201)
ClaI* (4494)
ScaI (4682)
BspEI* (4754)
BspEI (4890)
AsiSI (5290)
NcoI (5296)
PstI (5319)
BspEI (5427)
Esp3I (5580)
AatII (5629)
SmaI (5668)
(8733) SapI
(8973) Esp3I
(9208) SapI
(9749) ClaI*
(11,811) Esp3I
(12,804) ScaI
(301) HindIII
(317) PstI
(530) ScaI
(628) NcoI
(837) BspEI


Created with SnapGene®
pB5GWcY-FAR1
13,062 bp
pVS1 StaA
pVS1 RepA
bom
ori
SmR
NOS promoter
HygR
cYFP
FAR1
attB2
2000
4000
6000
8000
10,000
12,000
(530) ScaI
NcoI (628)
BspEI (837)
(317) PstI
(301) HindIII
EcoRV (1064)
XbaI (1166)
(12,998) ScaI
EcoRI (1412)
(12,005) Esp3I
SmaI (1420)
XbaI (1424)
KpnI-FAR1-FP (1429 .. 1461)
KpnI - NcoI (1434)
BspEI (1914)
ScaI (2118)
ScaI (2129)
SacI (2437)
SacI (2493)
SalI (2764)
PstI (2868)
PstI (3198)
SpeI (3400)
StuI* (3741)
EcoRI (3754)
(9943) ClaI*
ClaI (3825)
SapI (3861)
BamHI (3920)
BamHI-FAR1-RP (3896 .. 3925)
(9402) SapI
(9167) Esp3I
(8927) SapI
EcoRI (3926)
SacI (3971)
BspEI (4246)
AscI (4395)
ClaI* (4688)
ScaI (4876)
BspEI* (4948)
BspEI (5084)
AsiSI (5484)
NcoI (5490)
PstI (5513)
BspEI (5621)
(5862) SmaI
(5823) AatII
Esp3I (5774)



Supplemental Figure 2.3 Plasmid maps

Methods

Plant and growth condition

The *mug4-2* (SALK_036408C) and *mug4-3* (SALK_036244C) T-DNA insertion mutants are in the Columbia-0 (Col-0) ecotype background and therefore compared with the Col-0 as wild-type in all analyses presented. The wild-type, T-DNA insertion lines were obtained from the Arabidopsis Biological Resource Center (Ohio State University). The location of the T-DNA insertion in *mug4-2* and *mug4-3* were confirmed by sequencing of the allele-specific PCR products. The *mug1-1 mug2-3* were described and obtained from (Joly-Lopez, Forczek et al. 2012). The *mug1 mug2 mug4* triple mutant was constructed by crossing *mug4-3* and *mug1-1 mug2-3*. Homozygous double mutants and triple mutants were confirmed by genotyping. All the primer information is listed in Supplemental Table 2.1.

Plated seeds were stratified in the dark for 3 days at 4°C and kept on plates for 2 weeks in a growth chamber (Conviron model E15) at 22°C under a 16 h light/8 h dark photoperiod, ~100 mmol quanta/m^2/s light intensity, 60% relative humidity. Seedlings were transplanted to soil with a composition of PRO-MIX (Premier Tech Horticulture, Quebec, Canada): vermiculite: perlite of 2:1:1 in 2 ½ inch square pots and returned to the growth chamber.

Seeds of *Nicotiana benthamiana* were sown on PRO-MIX soil and grown for 3-4 weeks in controlled greenhouse conditions under long-day photoperiod cycles (16 h light/8 h dark) at 22°C.

Subcellular localization detection

Genomic DNA and total RNA were isolated from the leaves of Col-0 using DNeasy Plant Kit and RNeasy Plus Mini Kit, Qiagen, according to the manufacturer's instructions. cDNAs were synthesized from total RNA using SuperScript II reverse transcriptase (Invitrogen) with oligo(dT) primer. The synthesized cDNA was then used as template for subsequent PCR to get the full-length cDNA of *MUG1, MUG2, MUG3* and *MUG4*. The cDNA fragments of *MUG1, MUG2, MUG3* were cloned into pENTR™ /D-TOPO® Vector through Directional TOPO® Cloning according to the manufacturer instructions. The cDNA fragment of *MUG4* was cloned into pGEM-Gate entry vector at the XbaI and BamHI sites (Sun, Movahed et al. 2020). The YFP fragment in the binary Gateway compatible plant transformation vector pEarleyGate 101 was replaced with mCherry fragment by AQUA cloning (Beyer, Gonschorek et al. 2015) to create pEarleyGate 101(mCherry). The *MUG1* to *MUG4* entry vectors and the destination vector pEarleyGate 101(mCherry) were later mixed by gateway LR reaction to generate MUG1-mCherry, MUG2-mCherry, MUG3-mCherry, MUG4-mCherry.

The final constructs were transformed into *Agrobacterium tumefaciens* (strain GV3101) by electroporation (Weigel and Glazebrook 2006) and transfected into *Nicotiana benthamiana* tobacco leaves according to nature protocol (Sparkes, Runions et al. 2006).

To generate the CaMV 35S promoter-driven *35s:MUG4:GFP* construct, the pGEM-MUG4 was cloned into the binary Gateway compatible plant transformation vector pEarleyGate103 by LR reaction. To generate the *MUG4* native promoter-driven *mug4pro:MUG4: GFP* construct, a DNA fragment that starts from 1.3kb upstream region of *MUG4* gene from wild-type genomic

DNA were amplified using PrimeSTAR® Max DNA Polymerase and cloned into pGEM-Gateway entry vector and introduced into the binary vector pMDC107 (provided by Hugo Zheng lab) through LR reaction. All the primers information is listed in Supplemental Table 2.1.

The final *35S:MUG4:GFP* and *mug4pro:MUG4:GFP* constructs were transformed into *A. tumefaciens* and transfected into *A. thaliana mug4-2* by floral dip (Zhang, Henriques et al. 2006). Stable *A. thaliana 35S:MUG4:GFP* and *mug4pro:MUG4:GFP* transgenic plants were confirmed by BASTA and hygromycin B-resistant selection and RT-PCR.

The reagent 4′,6-diamidino-2-phenylindole (DAPI) was used for staining nuclei. Infiltrated tobacco leaves or *A. thaliana* 10-day-old transgenic seedlings were incubated in distilled water supplemented with 5 μg/ml DAPI for 10 min. Afterwards, the plant tissues were mounted on a microscope slide and covered with distilled water for observation. Images were captured using the Leica SP8 confocal system, with a × 40/1.2 water immersion objective. Excitation/emission wavelengths were 587 /610 nm for mCherry, 488/507 nm for GFP and 345/455 nm for DAPI. Post-processing of images was performed with ImageJ. Merged images were generated for reference with a 10 μm scale bar.

Transcription activity assays in yeast

To generate GAL4BD-MUG1, GAL4BD-MUG2, GAL4BD-MUG3, GAL4BD-MUG4, GAL4BD-GAL4AD and GAL4BD-FAR1, the entry plasmids of *MUG1* to *MUG4* were gateway cloned into pGBKT7-GW vectors (Lu, Tang et al. 2010) and generated pGBKT7-GW-MUG1 (BD-MUG1), pGBKT7-GW-MUG2 (BD-MUG2), pGBKT7-GW-MUG3 (BD-MUG3),

pGBKT7-GW-MUG4 (BD-MUG4), pGBKT7-GW-GAL4AD (BD-AD), pGBKT7-GW-FAR1 (BD-FAR1). The above constructs were transformed into Y2HGold reporter strain yeast cells by electroporation. The transformed yeast cells were grown and detected on SD/–Trp medium to evaluate whether the transformation was successful based on growth ability. The positive colonies are mixed and dilute to OD=1, 0.1 in sterilized water. The yeast solutions were then dropped the in the plates of SD/–Trp, SD/–Trp/ His, and SD/–Trp/ His/X-a-Gal to detect the transcriptional activity ability of each protein. The yeast stains with BD-AD, and BD-FAR1 were used as positive controls. The stain with empty pGBKT7-GW vector was used as the negative control.

Transcription activity assays in plant

Based on the fact that GL2 binds to the promoter of RSL1(pRSL1) (Lin, Ohashi et al. 2015), this assay is to test that when MUG1 to MUG4 proteins link with GL2, whether the MUG-A proteins activate GFP that fused with pRSL1.

Backbone construction: pCambia-mCherry (Movahed, Patarroyo et al. 2017) (provided by Hugo Zheng lab) was used as the template, using primer set knp1-mCherry and mcherryrev to amplify the mCherry fragment. The PCR products were treated with DpnI to get rid of the plasmid template, then digested with KpnI restriction enzyme. pDOE-04 plasmid (Gookin and Assmann 2014) (provided by Hugo Zheng lab) was digested with KpnI and NruI. The PCR products and the digested plasmid were ligated and the pDOE-04-mCherry plasmid was generated. VP16:GL2ΔN sequence was amplified using 35s:VP16:GL2ΔN plasmid (provided by Lijia Qu

lab) as the template with primers S-B-VP16-FP and GL2-R, treated with DpnI and then send to digest with SalI. pDOE-04-mCherry was digested with XhoI and SnaBI. The digested VP16:GL2ΔN PCR products were ligated with the digested pDOE-04-mCherry plasmid to generate the pDOE-04-mCherry-VP16-GL2-V2 plasmid.

pDOE-04-mCherry-MUG1/2/3-GL2-V2 construction: The DNA fragments of MUG1 to MUG3 were amplified using their respective entry vector described in the early content using primer sets V2-M1/2/3-FP, V2-M1/2/3-RP. The PCR fragments were treated with DpnI and then cloned into the pDOE-04-mCherry-VP16-GL2-V2 plasmids digested with BamHI and SmaI through homologous recombination.

pDOE-04-mCherry-MUG4/FAR1-GL2-V2 construction: The DNA fragments of MUG4 and FAR1 were amplified using their respective entry vector described in the early content using primer sets BamHI-MUG4/FAR1, c-MUG4/FAR1-RP, respectively. The PCR fragments were treated with DpnI and cloned into the BamHI and SmaI site of the pDOE-04-mCherry-VP16-GL2-V2 plasmid to generate the resultant plasmids.

Nicotiana benthamiana leaves were infiltrated with *Agrobacterium* strain GV301 containing above-described plasmids. Infiltrated leaves were observed using the Leica SP8 point-scanning confocal system on a Leica DMI6000B inverted microscope equipped with spectral fluorescent light detectors (three PMT, one HyD high sensitivity detector). A 63x/1.4 oil objective was used for all the imaging. The laser excitation wavelength was 488 nm and the spectral detection was set 507nm for GFP. Samples are scanned at 400 Hz and standard 16× line averaged for image

recording. Post-processing of images was performed with ImageJ version 1.48 software. Merged images were generated for reference with a 50 μm scale bar.

Bimolecular fluorescence complementation (BiFC) of MUG-A proteins

MUG1 to MUG4 entry vectors were recombined with pB5GWnY and pB5GWcY respectively (Kamigaki, Nito et al. 2016). Each plasmid has either the N-terminal (nYFP) or C-terminal (cYFP) fragment of YFP driven by the 35S plant expression promoter. After sequence was confirmed, each plasmid was transformed separately into *Agrobacterium tumefaciens* strain GV301 and pair-wise infiltrated into tobacco leaves.

BiFC images were captured using the above-mentioned Leica SP8 confocal system, with a × 40/1.2 water immersion objective. Excitation/emission wavelengths were 488/505–530 nm for YFP.Each image captured two separate filter scans, including YFP and bright field. Post-processing of images was performed with ImageJ. A merged image was generated for reference with a 50 μm scale bar.

Yeast two hybrid assay

The entry vector pGEM-MUG4 and destination vector pGADT7-GW were recombined through LR reaction to get the pGADT7-GW-MUG4 (GAL4AD-MUG4). The yeast strain AH109 was first transformed with empty pGADT7-GW, then co-transformed with GAL4BD-MUG1, GAL4BD-MUG2, GAL4BD-MUG3, GAL4BD-MUG4 as controls. For the testing of the interaction among MUG4 and MUG1 to MUG4 proteins, the yeast strain AH109 first

transformed with GAL4AD-MUG4, then co-transformed with GAL4BD-MUG1, GAL4BD-MUG2, GAL4BD-MUG3, GAL4BD-MUG4. Drop the yeast solution in the plates SD/–Trp -Leu as control plate, and SD/–Trp/-Leu/- His, SD/–Trp/-Leu/-His/X-a-Gal plates to detect the protein-protein interaction.

All the construction map are listed in Supplemental Figure 2.3.

RNA-seq analysis

Total RNA was extracted from 10-days wild-type, *mug4-2* seedlings grown on MS media using the RNeasy Plant Mini Kit (Qiagen) and DNase I treated (Promega). Two biological replicates were harvested for each line. RNA samples were processed by first preparing a Trueseq RNA-Seq library (Illumina) and then sequenced at 50 bp single read using Illumina HiSeq 2000 technology at Genome Quebec, Canada. Sequences were filtered and trimmed, respectively, using the FASTQ Quality Trimmer tools (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) with default settings. Reads were subsequently mapped to the TAIR10 version of the Arabidopsis genome using Hisat2 (Kim, Paggi et al. 2019). The mapped reads were then counted using featureCounts for differential gene expression analysis (Liao, Smyth et al. 2014). Differential expressed genes were analyzed in DEseq2 with adjust P value<0.05; log2 fold change>0.5 and <-0.5. GO-enrichment analysis was carried with agriGO web-based tool and database. Parameters: Statistical test method: Fisher; Multi-test adjustment method: Yekutieli (FDR under dependency); Significance Level: 0.05; Minimum number of mapping entries: 5.

Reference

Arkhipova, I. and M. Meselson (2000). "Transposable elements in sexual and ancient asexual taxa." Proceedings of the National Academy of Sciences **97**(26): 14473-14477.

Babu, M. M., L. M. Iyer, S. Balaji and L. Aravind (2006). "The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons." Nucleic Acids Research **34**(22): 6505-6520.

Benito, M. I. and V. Walbot (1997). "Characterization of the maize Mutator transposable element MURA transposase as a DNA-binding protein." Molecular and Cellular Biology **17**(9): 5165-5175.

Beyer, H. M., P. Gonschorek, S. L. Samodelov, M. Meier, W. Weber and M. D. Zurbriggen (2015). "AQUA Cloning: A Versatile and Simple Enzyme-Free Cloning Approach." PloS one **10**(9): e0137652-e0137652.

Bhasin, A., I. Y. Goryshin, M. Steiniger-White, D. York and W. S. Reznikoff (2000). "Characterization of a Tn5 pre-cleavage synaptic complex." J Mol Biol **302**(1): 49-63.

Boeke, J. D., D. J. Garfinkel, C. A. Styles and G. R. Fink (1985). "Ty elements transpose through an RNA intermediate." Cell **40**(3): 491-500.

Boissinot, S., A. Entezam and A. V. Furano (2001). "Selection against deleterious LINE-1-containing loci in the human lineage." Molecular biology and evolution **18**(6): 926-935.

Bundock, P. and P. Hooykaas (2005). "An Arabidopsis hAT-like transposase is essential for plant development." Nature **436**(7048): 282-284.

Chuong, E. B., N. C. Elde and C. Feschotte (2017). "Regulatory activities of transposable elements: from conflicts to benefits." Nature Reviews Genetics **18**(2): 71.

Claeys, H., S. Van Landeghem, M. Dubois, K. Maleux and D. Inzé (2014). "What is stress? Dose-response effects in commonly used in vitro stress assays." Plant physiology **165**(2): 519-527.

Cobb, B. S., S. Morales-Alcelay, G. Kleiger, K. E. Brown, A. G. Fisher and S. T. Smale (2000). "Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding." Genes Dev **14**(17): 2146-2160.

Cowan, R. K., D. R. Hoen, D. J. Schoen and T. E. Bureau (2005). "MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms." Molecular Biology and Evolution **22**(10): 2084-2089.

Demidchik, V. and F. J. M. Maathuis (2007). "Physiological roles of nonselective cation channels in plants: from salt stress to signalling and development." New Phytologist **175**(3): 387-404.

Di Tommaso, P., S. Moretti, I. Xenarios, M. Orobitg, A. Montanyola, J. M. Chang, J. F. Taly and C. Notredame (2011). "T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension." Nucleic Acids Res **39**(Web Server issue): W13-17.

Doolittle, W. F. and C. Sapienza (1980). "Selfish genes, the phenotype paradigm and genome evolution." Nature **284**(5757): 601-603.

Feschotte, C., N. Jiang and S. R. Wessler (2002). "Plant transposable elements: where genetics meets genomics." Nat Rev Genet **3**(5): 329-341.

Flavell, A. J., S. R. Pearce and A. Kumar (1994). "Plant transposable elements and the genome." Current opinion in genetics & development **4**(6): 838-844.

Gagné, F. (2014). Chapter 2 - Tissue Preparation and Subcellular Fractionation Techniques. Biochemical Ecotoxicology. F. Gagné. Oxford, Academic Press**:** 21-31.

Gookin, T. E. and S. M. Assmann (2014). "Significant reduction of BiFC non-specific assembly facilitates in planta assessment of heterotrimeric G-protein interactors." Plant Journal **80**(3): 553-567.

Gray, L. T., K. K. Fong, T. Pavelitz and A. M. Weiner (2012). "Tethering of the Conserved piggyBac Transposase Fusion Protein CSB-PGBD3 to Chromosomal AP-1 Proteins Regulates Expression of Nearby Genes in Humans." Plos Genetics **8**(9).

Greenblatt, I. M. and R. A. Brink (1963). "Transpositions of Modulator in maize into divided and undivided chromosome segments." Nature **197**(4865): 412-413.

Guan, Q. M., J. M. Wu, X. L. Yue, Y. Y. Zhang and J. H. Zhu (2013). "A Nuclear Calcium-Sensing Pathway Is Critical for Gene Regulation and Salt Stress Tolerance in Arabidopsis." Plos Genetics **9**(8).

Hoen, D. R. and T. E. Bureau (2015). "Discovery of Novel Genes Derived from Transposable Elements Using Integrative Genomic Analysis." Molecular Biology and Evolution **32**(6): 1487-1506.

Hudson, M., C. Ringli, M. T. Boylan and P. H. Quail (1999). "The FAR1 locus encodes a novel nuclear protein specific to phytochrome A signaling." Genes Dev **13**(15): 2017-2027.

Hudson, M. E., D. R. Lisch and P. H. Quail (2003). "The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway." Plant Journal **34**(4): 453-471.

Isayenkov, S. V., S. A. Dabravolski, T. Pan and S. Shabala (2020). "Phylogenetic Diversity and Physiological Roles of Plant Monovalent Cation/H(+) Antiporters." Front Plant Sci **11**: 573564.

Isayenkov, S. V. and F. J. M. Maathuis (2019). "Plant Salinity Stress: Many Unanswered Questions Remain." Front Plant Sci **10**: 80.

Jiao, Y., P. Peluso, J. Shi, T. Liang, M. C. Stitzer, B. Wang, M. S. Campbell, J. C. Stein, X. Wei, C. S. Chin, K. Guill, M. Regulski, S. Kumari, A. Olson, J. Gent, K. L. Schneider, T. K. Wolfgruber, M. R. May, N. M. Springer, E. Antoniou, W. R. McCombie, G. G. Presting, M. McMullen, J. Ross-Ibarra, R. K. Dawe, A. Hastie, D. R. Rank and D. Ware (2017). "Improved maize reference genome with single-molecule technologies." Nature **546**(7659): 524-527.

Joly-Lopez, Z. and T. E. Bureau (2018). "Exaptation of transposable element coding sequences." Current opinion in genetics & development **49**: 34-42.

Joly-Lopez, Z., E. Forczek, D. R. Hoen, N. Juretic and T. E. Bureau (2012). "A Gene Family Derived from Transposable Elements during Early Angiosperm Evolution Has Reproductive Fitness Benefits in Arabidopsis thaliana." Plos Genetics **8**(9).

Joly-Lopez, Z., E. Forczek, D. R. Hoen, N. Juretic and T. E. Bureau (2012). "A gene family derived from transposable elements during early angiosperm evolution has reproductive fitness benefits in Arabidopsis thaliana."

Joly-Lopez, Z., E. Forczek, E. Vello, D. R. Hoen, A. Tomita and T. E. Bureau (2017). "Abiotic Stress Phenotypes Are Associated with Conserved Genes Derived from Transposable Elements." Front Plant Sci **8**: 2027.

Joly-Lopez, Z., D. R. Hoen, M. Blanchette and T. E. Bureau (2016). "Phylogenetic and Genomic Analyses Resolve the Origin of Important Plant Genes Derived from Transposable Elements." Mol Biol Evol **33**(8): 1937-1956.

Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek and A. Potapenko (2021). "Highly accurate protein structure prediction with AlphaFold." Nature: 1-11.

Kagale, S., C. Koh, J. Nixon, V. Bollina, W. E. Clarke, R. Tuteja, C. Spillane, S. J. Robinson, M. G. Links, C. Clarke, E. E. Higgins, T. Huebert, A. G. Sharpe and I. A. P. Parkin (2014). "The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure." Nature Communications **5**(1): 3706.

Kagale, S., S. J. Robinson, J. Nixon, R. Xiao, T. Huebert, J. Condie, D. Kessler, W. E. Clarke, P. P. Edger, M. G. Links, A. G. Sharpe and I. A. P. Parkin (2014). "Polyploid Evolution of the Brassicaceae during the Cenozoic Era " The Plant Cell **26**(7): 2777-2791.

Kamigaki, A., K. Nito, K. Hikino, S. Goto-Yamada, M. Nishimura, T. Nakagawa and S. Mano (2016). "Gateway vectors for simultaneous detection of multiple protein− protein interactions in plant cells using bimolecular fluorescence complementation." PloS one **11**(8): e0160717.

Kidwell, M. G. and D. R. Lisch (2001). "Perspective: transposable elements, parasitic DNA, and genome evolution." Evolution **55**(1): 1-24.

Kim, J., S. H. Moon, Y. C. Shin, J. H. Jeon, K. J. Park, K. P. Lee and I. So (2016). "Intracellular spermine blocks TRPC4 channel via electrostatic interaction with C-terminal negative amino acids." Pflugers Arch **468**(4): 551-561.

Kosek, D., A. B. Hickman, R. Ghirlando, S. He and F. Dyda (2021). "Structures of ISC th4 transpososomes reveal the role of asymmetry in copy-out/paste-in DNA transposition." The EMBO journal **40**(1): e105666.

Leung, J., M. Bouvier-Durand, P. C. Morris, D. Guerrier, F. Chefdor and J. Giraudat (1994). "Arabidopsis ABA response gene ABI1: features of a calcium-modulated protein phosphatase." Science **264**(5164): 1448-1452.

Li, B., S. Jiang, X. Yu, C. Cheng, S. Chen, Y. Cheng, J. S. Yuan, D. Jiang, P. He and L. Shan (2015). "Phosphorylation of trihelix transcriptional repressor ASR3 by MAP KINASE4 negatively regulates Arabidopsis immunity." Plant Cell **27**(3): 839-856.

Lin, Q., Y. Ohashi, M. Kato, T. Tsuge, H. Gu, L. J. Qu and T. Aoyama (2015). "GLABRA2 Directly Suppresses Basic Helix-Loop-Helix Transcription Factor Genes with Diverse Functions in Root Hair Development." Plant Cell **27**(10): 2894-2906.

Lin, R., L. Ding, C. Casola, D. R. Ripoll, C. Feschotte and H. Wang (2007). "Transposase-derived transcription factors regulate light signaling in Arabidopsis." Science **318**(5854): 1302-1305.

Lin, R., Y. Teng, H. J. Park, L. Ding, C. Black, P. Fang and H. Wang (2008). "Discrete and essential roles of the multiple domains of Arabidopsis FHY3 in mediating phytochrome A signal transduction." Plant Physiol **148**(2): 981-992.

Liu, K. and S. R. Wessler (2017). "Transposition of Mutator-like transposable elements (MULEs) resembles hAT and Transib elements and V(D)J recombination." Nucleic Acids Research **45**(11): 6644-6655.

Lu, Q., X. Tang, G. Tian, F. Wang, K. Liu, V. Nguyen, S. E. Kohalmi, W. A. Keller, E. W. Tsang, J. J. Harada, S. J. Rothstein and Y. Cui (2010). "Arabidopsis homolog of the yeast TREX-2 mRNA export complex: components and anchoring nucleoporin." Plant J **61**(2): 259-270.

Ma, L. and G. Li (2018). "FAR1-RELATED SEQUENCE (FRS) and FRS-RELATED FACTOR (FRF) Family Proteins in Arabidopsis Growth and Development." Front Plant Sci **9**: 692.

Makarevitch, I., A. J. Waters, P. T. West, M. Stitzer, C. N. Hirsch, J. Ross-Ibarra and N. M. Springer (2015). "Correction: Transposable Elements Contribute to Activation of Maize Genes in Response to Abiotic Stress." PLoS Genet **11**(10): e1005566.

Makarova, K. S., L. Aravind and E. V. Koonin (2002). "SWIM, a novel Zn-chelating domain present in bacteria, archaea and eukaryotes." Trends Biochem Sci **27**(8): 384-386.

McClintock, B. (1950). "The Origin and Behavior of Mutable Loci in Maize." Proceedings of the National Academy of Sciences of the United States of America **36**(6): 344-355.

McGrath, K. C., B. Dombrecht, J. M. Manners, P. M. Schenk, C. I. Edgar, D. J. Maclean, W.-R. d. Scheible, M. K. Udvardi and K. Kazan (2005). "Repressor- and Activator-Type Ethylene Response Factors Functioning in Jasmonate Signaling and Disease Resistance Identified via a Genome-Wide Screen of Arabidopsis Transcription Factor Gene Expression." Plant Physiology **139**(2): 949-959.

Michel, K. and P. W. Atkinson (2003). "Nuclear localization of the Hermes transposase depends on basic amino acid residues at the N-terminus of the protein." J Cell Biochem **89**(4): 778-790.

Mills, R. E., E. A. Bennett, R. C. Iskow and S. E. Devine (2007). "Which transposable elements are active in the human genome?" Trends in Genetics **23**(4): 183-191.

Munns, R. and M. Tester (2008). "Mechanisms of salinity tolerance." Annu. Rev. Plant Biol. **59**: 651-681.

Negi, P., A. N. Rai and P. Suprasanna (2016). "Moving through the Stressed Genome: Emerging Regulatory Roles for Transposons in Plant Stress Response." Frontiers in Plant Science **7**.

Negrao, S., S. M. Schmockel and M. Tester (2017). "Evaluating physiological responses of plants to salinity stress." Annals of Botany **119**(1): 1-11.

Ono, A., S. H. Kim and V. Walbot (2002). "Subcellular localization of MURA and MURB proteins encoded by the maize MuDR transposon." Plant Mol Biol **50**(4-5): 599-611.

Orgel, L. E. and F. H. Crick (1980). "Selfish DNA: the ultimate parasite." Nature **284**(5757): 604-607.

Parida, A. K. and A. B. Das (2005). "Salt tolerance and salinity effects on plants: a review." Ecotoxicol Environ Saf **60**(3): 324-349.

Pasyukova, E., S. Nuzhdin, T. Morozova and T. Mackay (2004). "Accumulation of transposable elements in the genome of Drosophila melanogaster is associated with a decrease in fitness." Journal of Heredity **95**(4): 284-290.

Platt, A., H. C. Ross, S. Hankin and R. J. Reece (2000). "The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase." Proceedings of the National Academy of Sciences **97**(7): 3154-3159.

Ritter, A., S. Inigo, P. Fernandez-Calvo, K. S. Heyndrickx, S. Dhondt, H. Shi, L. De Milde, R. Vanden Bossche, R. De Clercq, D. Eeckhout, M. Ron, D. E. Somers, D. Inze, K. Gevaert, G. De Jaeger, K. Vandepoele, L. Pauwels and A. Goossens (2017). "The transcriptional repressor complex FRS7-FRS12 regulates flowering time and growth in Arabidopsis." Nat Commun **8**: 15235.

Sabogal, A., A. Y. Lyubimov, J. E. Corn, J. M. Berger and D. C. Rio (2010). "THAP proteins target specific DNA sites through bipartite recognition of adjacent major and minor grooves." Nature Structural & Molecular Biology **17**(1): 117-U145.

Samanta, A., T. E. T. Hughes and V. Y. Moiseenkova-Bell (2018). "Transient Receptor Potential (TRP) Channels." Subcell Biochem **87**: 141-165.

Shavrukov, Y. (2013). "Salt stress or salt shock: which genes are we studying?" Journal of Experimental Botany **64**(1): 119-127.

Shi, H., L. Xiong, B. Stevenson, T. Lu and J. K. Zhu (2002). "The Arabidopsis salt overly sensitive 4 mutants uncover a critical role for vitamin B6 in plant salt tolerance." Plant Cell **14**(3): 575-588.

Silver, P. A. (1991). "How proteins enter the nucleus." Cell **64**(3): 489-497.

Sinzelle, L., Z. Izsvak and Z. Ivics (2009). "Molecular domestication of transposable elements: From detrimental parasites to useful host genes." Cellular and Molecular Life Sciences **66**(6): 1073-1093.

Sun, J., N. Movahed and H. Zheng (2020). "LUNAPARK Is an E3 Ligase That Mediates Degradation of ROOT HAIR DEFECTIVE3 to Maintain a Tubular ER Network in Arabidopsis." The Plant Cell **32**(9): 2964-2978.

Tavakkoli, E., P. Rengasamy and G. K. McDonald (2010). "High concentrations of Na+ and Cl–ions in soil solution have simultaneous detrimental effects on growth of faba bean under salinity stress." Journal of Experimental Botany **61**(15): 4449-4459.

Thomas, J., M. Sepahi, B. Arendall and H. Bohnert (1995). "Enhancement of seed germination in high salinity by engineering mannitol expression in Arabidopsis thaliana." Plant, Cell & Environment **18**(7): 801-806.

Volff, J. N. (2006). "Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes." Bioessays **28**(9): 913-922.

Wang, H. and X. W. Deng (2002). "Arabidopsis FHY3 defines a key phytochrome A signaling component directly interacting with its homologous partner FAR1." EMBO J **21**(6): 1339-1349.

Wang, S., Y. Chang, J. Guo and J.-G. Chen (2007). "Arabidopsis Ovate Family Protein 1 is a transcriptional repressor that suppresses cell elongation." The Plant Journal **50**(5): 858-872.

Wang, W., W. Tang, T. Ma, D. Niu, J. B. Jin, H. Wang and R. Lin (2016). "A pair of light signaling factors FHY3 and FAR1 regulates plant immunity by modulating chlorophyll biosynthesis." J Integr Plant Biol **58**(1): 91-103.

Wang, X., Y. He, H. Wei and L. Wang "A clock regulatory module is required for salt tolerance and control of heading date in rice." Plant, Cell & Environment **n/a**(n/a).

Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A P. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede (2018). "SWISS-MODEL: homology modelling of protein structures and complexes." Nucleic Acids Research **46**(W1): W296-W303.

Weiss, S., I. Gottfried, I. Mayrose, S. L. Khare, M. Xiang, S. J. Dawson and K. B. Avraham (2003). "The DFNA15 deafness mutation affects POU4F3 protein stability, localization, and transcriptional activity." Mol Cell Biol **23**(22): 7957-7964.

Wessler, S. R. (2006). "Transposable elements and the evolution of eukaryotic genomes." Proceedings of the National Academy of Sciences of the United States of America **103**(47): 17600-17601.

Wu, S. J., L. Ding and J. K. Zhu (1996). "SOS1, a genetic locus essential for salt tolerance and potassium acquisition." Plant Cell **8**(4): 617-627.

Xu, J., M. Mcpartlon and J. Li (2021). "Improved protein structure prediction by deep learning irrespective of co-evolution information." Nature Machine Intelligence: 1-9.

Xu, X., C. Chen, B. Fan and Z. Chen (2006). "Physical and Functional Interactions between Pathogen-Induced Arabidopsis WRKY18, WRKY40, and WRKY60 Transcription Factors." The Plant Cell **18**(5): 1310-1326.

Yamasaki, K., T. Kigawa, M. Inoue, M. Tateno, T. Yamasaki, T. Yabuki, M. Aoki, E. Seki, T. Matsuda, Y. Tomo, N. Hayami, T. Terada, M. Shirouzu, A. Tanaka, M. Seki, K. Shinozaki and S. Yokoyama (2005). "Solution Structure of an Arabidopsis WRKY DNA Binding Domain." The Plant Cell **17**(3): 944-956.

Yamasaki, K., T. Kigawa, M. Seki, K. Shinozaki and S. Yokoyama (2013). "DNA-binding domains of plant-specific transcription factors: structure, function, and evolution." Trends Plant Sci **18**(5): 267-276.

Yang, J., I. Anishchenko, H. Park, Z. Peng, S. Ovchinnikov and D. Baker (2020). "Improved protein structure prediction using predicted interresidue orientations." Proceedings of the National Academy of Sciences **117**(3): 1496-1503.

Yuan, Y. W. and S. R. Wessler (2011). "The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies." Proceedings of the National Academy of Sciences of the United States of America **108**(19): 7884-7889.

Link between Chapter II and Chapter III

In Chapter II, I performed a comprehensive set of experiments using molecular and genetic technologies to explore the function of *MUG-A* genes. All MUG-A proteins are located in the nucleus. All of them have transcriptional repression capacity in plants. *MUG-A* genes are shown to form homodimers and heterodimers with overlapping functions that are beneficial to plant development. While I began my graduate study, previous lab mates did a phenotypic screening, in which the Arabidopsis T-DNA insertion lines of *MUG4*, a member of *MUG-A* gene family showed a sensitive phenotype to salinity. Within more characteristics are discovered in chapter II, in Chapter III, I focus on the function and application of *MUG4* in the field of salt stress. First, I dissected the NaCl stress into osmatic stress and ion toxic stress and determined which stress MUG4 specifically responds to. Later, physiological analyses were applied to test if *mug4* mutants tend to accumulate more cations compared to wildtype under salt stress. Finally, I generated transgenic lines overexpressing AtMUG4 in Arabidopsis and Camelina and test whether they are more tolerant to salt.

Chapter III

MUG4, an exapted transposable element is essential to salinity tolerance in *Arabidopsis thaliana* and *Camelina sativa*.

Abstract

Transposable elements (TEs) are now known to contribute directly or indirectly to the host, rather than being “selfish DNA”. One specific way is TE exaptation, when a part or a whole transposase is exapted into functional host proteins. *MUSTANG4* (*MUG4*) is one of the exapted transposable elements (ETEs), which is derived from the whole transposase gene *mudrA*. MUG4 has been demonstrated to play key roles in regulating plant development. Recently, it was found that the *Arabidopsis thaliana mug4* mutant seedlings showed a salt-sensitive phenotype. Here, I report that the *A. thaliana mug4* mutants are specifically shot sensitive to exogenous sodium, potassium, and calcium cations. The mutant lines accumulated more sodium, potassium and calcium ions compared to that in wild-type under NaCl, KCl and $CaCl_2$ stresses, respectively. In addition, the expression of abiotic stress-responsive genes was significantly higher in *mug4* mutant plants than wild-type plants under salt stress conditions. Furthermore, overexpression of AtMUG4 increases salt tolerance both in *A. thaliana* as well as *Camelina sativa*. To this end, our study opens a new door for valuable roles of ETEs in plant adaptation to stress and provides great potential for crop improvement.

Introduction

Transposable elements (TEs), also known as "jumping genes”, are DNA sequences with the ability to make copies of themselves or change their position within a genome. They make up a large fraction of eukaryotic genome. There are two distinct types of TEs: Class I or retrotransposons, move via a “copy and paste” mechanism mediated by reverse transcription of

their RNA (Boeke, Garfinkel et al. 1985), and Class II or DNA transposons, use a “cut and paste” mechanism driven by the catalytic activity of the enzyme transposase to change their positions (Greenblatt and Brink 1963).

In the early 1980s, two papers were published back-to-back in Nature (Doolittle and Sapienza 1980, Orgel and Crick 1980) that emphasized the selfish concept of TEs that they spread in a population regardless of their effect on organismal fitness once they have a transmission advantage. This reputation was bolstered by numerous documents that recorded the deleterious mutations generated by TEs (Arkhipova and Meselson 2000, Boissinot, Entezam et al. 2001, Pasyukova, Nuzhdin et al. 2004).

Nowadays, the lines between whether TEs are egocentric or altruistic become blurred. Contrary to the initial belief that TEs were “selfish DNA”, advances in genomics have facilitated the identification of TEs that serve important roles in genome function and evolution (Kidwell and Lisch 2001, Chuong, Elde et al. 2017). The process of turning TEs from“parasites” into useful host genes is termed exaptation. The TEs whose coding regions contribute directly to phenotypic functions of the host are referred to exapted TEs (ETEs) (Joly-Lopez and Bureau 2018).

In plants, a few ETEs has been discovered. For example, *FAR1/FHY3* family originated from *MudrA*, which are TFs involved in phyA signal transduction (Hudson, Lisch et al. 2003). *MUSTANG* (*MUG*) and *DAYSLEEPER*, derived from *MudrA* and a hAT transposase gene, respectively, are also essential for plant development (Bundock and Hooykaas 2005, Joly-Lopez, Forczek et al. 2012). Either discovered by forward genetics or reverse genetics, these reported

ETEs were characterized as limited in plant development. Only one paper published recently demonstrated that ETEs can be repurposed to functions in abiotic stress response with a screening of a pool of ETEs mutants (Joly-Lopez, Forczek et al. 2017). Within this screening, the *Arabidopsis thaliana mug4* mutants were found to have a severe salt sensitive phenotype.

High salinity is a major abiotic stress, which induces both osmotic and ionic stresses on crops, and thereby imposes a negative effect on their growth, development, and ultimately yield (Munns and Tester 2008). *Camelina sativa* is an emerging oilseed crop, its advantages such as relatively short growing period (85–100 days to maturity), low fertilizer requirement, and resistance to cold and drought lead it to be re-embraced as an industrial oil platform crop. However, recent genome sequencing revealed its highly undifferentiated homogenous polyploid genome (Kagale, Koh et al. 2014), posing a challenge for traditional breeding and genetic manipulation for salt tolerance improvement.

In this paper, how MUG4 affects plant salt stress response at physical level and molecular level were carefully studied, whether overexpressing AtMUG4 improves *A. thaliana* salt tolerance is determined. In addition, the homologs of MUG4 among *Brassicaceae* have been discovered and whether the function of MUG4 is conserved has been experimentally verified. Finally, I showed overexpressing AtMUG4 in *Camelina sativa* leads to a significant salt tolerance trait, which sheds light on the great potential of ETEs in agricultural applications.

Results

Genotyping and molecular complementation.

As described in the earlier chapter, MUG4 has an N-terminal MuDR DNA-binding domain, a middle MULE transposase domain and a C-terminal SWIM domain. The T-DNA insertions are predicted in the CDS region (https://www.arabidopsis.org/) and confirmed by sequencing. To confirm the mutants are loss of function lines, RNA was extracted from ten-day-old mutant seedlings and RT-PCR was performed. The two mutants expressed truncated mRNA at a low level up to the T-DNA insertion and no transcripts that continue across the T-DNA insertion were found (Figure 3.1). Thus, the T-DNA insertions effectively knock out MUG4 by disrupting the mRNA.

A subsequent complementation test was performed to confirm the functionality of MUG4. Wild-type (Col-0 background), *mug4-2*, *mug4-3* and two stable rescue lines generated in Chapter 2 were grown in Murashige and Skoog (MS) medium, after 4 days they were transferred to the MS medium supplied with 150 mM NaCl. After 6 days, the survival rate is calculated. Consistent with previous report, the *mug4* mutant seedlings showed an increased salt sensitivity compared with wild-type seedlings. Meanwhile, the two rescue lines led to complementation of the salt sensitive phenotype in *mug4* mutant plants (Figure 3.2). These data demonstrates that *MUG4* is the responsible gene for the salt sensitive phenotype.

MUG4 is specifically sensitive to Na^+, K^+, and Ca^{2+}.

Salinity affects plants in complex pathways. In general, salinity induces direct and indirect stresses on the plant. The direct stresses include the early-occurring stress, osmotic stress, and the late-stage induced ionic stress, caused by the accumulation of Na^+ and Cl^- (Shavrukov 2013).

The indirect stresses are salt-induced secondary stresses, such as oxidative stress caused by the overproduction of reactive oxygen species (ROS) (Negrao, Schmockel et al. 2017). To determine which stress MUG4 is involved in, I tested the response of *mug4* mutants under osmotic stress stimulated by iso-osmotic mannitol (Thomas, Sepahi et al. 1995), ionic stress stimulated by salt mixture of Na^+ and Cl^- dominant equivalent to 150 mM NaCl (Tavakkoli, Rengasamy et al. 2010) and oxidative stress stimulated by 6 mM H_2O_2 (Claeys, Van Landeghem et al. 2014). As shown in Figure 3.3, *mug4* mutants have the same relative growth rate as wild-type under normal condition (Figure 3.3a) as well as osmotic stress (Figure 3.3b), Cl^- stress (Figure 3.3c), and oxidative stress (Figure 3.3d). Meanwhile, *mug4* mutants showed a higher mortality rate than wild-type after the treatment of Na^+ dominant salt mixture (Figure 3.4a), which is comparable with the 150 mM NaCl treatment. These results suggest that the increased sensitivity of *mug4* mutants to salt stress is due to Na^+ toxicity. I then tested if MUG4 specific responds to Na^+ or also other cations. Na, K, and Li are in the same column of the periodic table of the elements, their similar chemical property suggests that they may have similar function in the organisms (Guan, Wu et al. 2013). Divalent cation such as Mg^{2+} and Ca^{2+} were also tested. As shown in Figure 3.4b-4e, the *mug4* mutant lines have a higher mortality rate than WT within extra KCl and $CaCl_2$, while they kept the same growth pattern under LiCl and $MgCl_2$ stresses. Above all, MUG4 is specifically sensitive to Na^+, K^+, and Ca^{2+}.

***mug4-2* mutants show salt sensitive phenotype in shoot during seedling stage and rosette growth stage.**

To determine which stages MUG4 works for plant salt tolerance, *A. thaliana*, Col-0, *mug4-2* plants were treated with salt-stress at gemination stage, seedling stage, and rosette growth stage. When plants directly grow on MS medium supplied with 50 mM NaCl, 75 mM NaCl and 100 mM NaCl, both the *mug4* mutants and wild-type can germinate in the medium (data not shown). After 10 days of growth, the *mug4* mutants showed a smaller shoot size compared to wild-type (data not shown). To determine whether the root growth has a similar pattern as the shoot, the root bending assay is applied (Shi, Xiong et al. 2002). Plants were grown vertically on MS medium for 4 days then transferred to 50 mM NaCl, 75 mM NaCl and 100 mM NaCl MS medium. Interestingly, unlike the Arabidopsis salt overly sensitive (SOS) mutant *sos2−1*, where root growth stopped in salt medium, *mug4* mutants root showed a similar growth pattern to wild-type (Figure 3.5). I then tested the plants response to salt during rosette growth stage. Plants were grown on soil for 15 days then watered with 150 mM NaCl solution and repeat watering after 4 days. Seven days later plants were photographed. *mug4* mutants had a significant decrease in rosette area in stressed plants (Figure 3.6). Fresh weight was also significantly lower in plants exposed to salinity (data not shown).

***mug4-2* accumulates extra cations under cation stresses.**

The specific sensitivity of *mug4-2* to Na^+, K^+, and Ca^{2+} suggests an imbalance of sodium, potassium, and calcium ion exchange process in the plant. Briefly, ions are imported into cells and exported out of cells to keep the net flux accommodating cellular requirements, thus creating ionic homeostasis. When plant cells are exposed to salinity, mediated by high NaCl concentrations, kinetic steady states of ion transport are disturbed (Niu, Bressan et al. 1995).

Plants take several strategies to rebalance the ion homeostasis such as restricting ion influx into the root, reducing ion translocation, and compartmentation of toxic ions into vacuoles (Isayenkov and Maathuis 2019). To determine which mechanism MUG4 is involved in, I measured the Na^+, K^+, Ca^{2+} content in *mug4-2* and wild-type plants under normal state and salt stresses. As shown in Figure 3.7, there was no significant difference between *mug4-2* and wild-type plants under the normal condition. However, after NaCl, KCl and $CaCl_2$ treatment, *mug4-2* accumulated more Na^+, K^+, Ca^{2+} than wild-type. The accumulation of Na^+, K^+, and Ca^{2+} content in *mug4-2* under stresses, combined with the *mug4* mutants shoot specific salt sensitive phenotype, suggests a role of MUG4 in Na^+, K^+, and Ca^{2+} up taking into plant leaves.

***mug4* mutant plants have different transcriptome profiles from wild-type under salt stress.**

In the earlier chapter, I demonstrated MUG4 has a putative DNA-binding domain, localized in the nucleus and have transcriptional suppressor activity. Thus, MUG4 may work as a transcriptional suppressor in response to salt stress. RNA-Seq analysis was performed to reveal the role of MUG4 in salt stress. Compared to the normal conditions, the overall transcriptome profiles of *mug4* mutant lines and wild-type under salt treatment have significant divergence (Figure 3.8). After salt treatment, 1920 genes were up-regulated ($\log2 \geq 1$, FDR ≤ 0.05), and 1265 genes were down-regulated ($\log2 \leq -1$, FDR ≤ 0.05) in the wild-type. In the *mug4-2* mutant, 2556 genes were up-regulated ($\log2 \geq 1$, FDR ≤ 0.05), and 2076 genes were down-regulated ($\log2 \leq -1$, FDR ≤ 0.05) after salt treatment (Figure 3.9a). In *mug4-3*, 2584 genes were up-regulated ($\log2 \geq 1$, FDR ≤ 0.05), 1794 genes were down-regulated ($\log2 \leq -1$, FDR ≤ 0.05) after salt treatment. Since the overall transcriptome profile of *mug4-2* can be

separated from the wild-type under normal condition, and its salt-sensitive phenotype is stronger than *mug4-3*, I then use mug4-2 and wild-type for further GO analysis. Most of the differential expressed genes of *mug4-2* are enriched into terms " Response to stimulus", "Response to stress" (Figure 3.9b).

Overexpression of AtMUG4 leads to tolerance in *A. thaliana* and *C. sativa*

Arabidopsis *mug4* mutants are salt sensitive, which drives me to investigate whether overexpressing MUG4 will make plant more tolerant to salt stress. I first tested overexpressing AtMUG4 in *A. thaliana*. Two Arabidopsis transgenic plants that overexpressed AtMUG4 under the control of the CaMV35S promoter were generated. Under normal condition, transgenic plants grow the same as wild-type plants. However, as shown in Figure 3.2 and Figure 3.4a, the transgenic plants show a higher survival rate than wild-type when exposed to 150 mM NaCl and 150 mM Na^+- dominant salt mix. These results suggest that overexpression of AtMUG4 improved the tolerance to Na^+ stresses in Arabidopsis.

Knowing that overexpressing AtMUG4 improves *A. thaliana* salt tolerance performance, I questioned whether MUG4 has a role to adapt to a high salinity environment in *Camelina sativa.* I transformed AtMUG4 in *C. sativa* and tested the transgenic plants response. A DNA cassette consisting of AtMUG4 cDNA driven by 35S promoter is built in *C. sativa*. The T3 homozygote plants were used for salt tolerance test. As shown in Figure 3.10, under normal conditions, both the two independent transgenic lines and wild-type Camelina grow well. However, after being treated with 200 mM NaCl, wild-type Camelina stopped vegetative growth and started flowering,

while the AtMUG4 overexpression lines continued the vegetative growth and stayed green, developing a similar size without salt treatment. These results suggest overexpressing AtMUG4 prompts more tolerance to salt in *C. sativa*.

Discussion and conclusion

MUG4 improves host adaption capacity to salt.

TEs are normally not actively transposing and remain being silenced by epigenetic suppression from the host. They are activated in response to challenges to the genome (McClintock 1984). This activation is now considered as evidence for adaptive roles of TEs in the stress response (Slotkin and Martienssen 2007). There are several mechanisms that TEs influence the host adaptation capacity, one is the activation of TEs leads to an increase in the mutation rate which generates the variability of the host genome (Casacuberta and Gonzalez 2013). The other is the TEs regulate response elements throughout the genome that helps reprogram stress gene networks (Chuong, Elde et al. 2017). Another mechanism is that a whole transposon gets exapted and works as a conventional gene (Joly-Lopez and Bureau 2018). In this study, I showed AtMUG4, a bona fide ETE, plays an essential role in plant adaptation to salt stress by exapting the whole *MudrA* transposase gene.

MUG4 may be a transcription regulator involved in nonselective cation channels in the shoot.

mug4 mutants salt-sensitive traits were only found in the shoot, not the root, which suggests the salt response function of MUG4 is specific in the shoot. The result *mug4* accumulates more Na^+, K^+, Ca^2 than wild-type when supply exogenous Na^+, K^+, Ca^{2+} suggests MUG4 has a role in restricting Na^+, K^+, Ca^{2+} influx or porting the ions out of the cell in the shoot. In plants, nonselective cation channels (NSCCs) are macromolecular pores in the cell membrane that form aqueous pathways that mediate Na^+, K^+, Ca^{2+} influx (Demidchik and Maathuis 2007). The monovalent cation: proton antiporter-2 (CPA2) family are genes that code transporters that catalyze monovalent cations such as Na^+:H^+ exchange (Isayenkov, Dabravolski et al. 2020). Both types of transporters work in the plasma membrane. MUG4 localizes in the nucleus, not the cell membrane, also it has a putative DNA binding domain, and the early chapter suggests it has the transcriptional repressor function. These results indicate that MUG4 may be involved in salt response by regulating the transporters directly or indirectly. In the RNA-seq analysis, *CNGC19* (*Cyclic Nucleotide Gated Channel 19*) is a up-regulated gene in *mug4*-2 background, which belongs to cyclic nucleotide gated channel family, one of the non-selective cation channels (Kaplan, Sherman et al. 2007) . Salt induced its expression in both wild-type and *mug4-2*, however, the expression level in *mug4-2* is more than 30% increasing than that in wild-type (Figure 3.9c), suggesting CNGC19 maybe a potential target of MUG4.

ETEs in plants have great potential in agriculture applications.

Previously, the *MUG* gene family was found in angiosperm with significant protein similarity, same protein domain architecture and conserved synteny blocks (Joly-Lopez, Forczek et al. 2012). However, having a high possibility closely related proteins do not always share the same

function. For example, the yeast Gal1 and Gal3 proteins are paralogs (73% identity and 92% similarity) that have evolved very different functions: Gal1 turned into a galactokinase by a two amino acid insertion event from its homologs gene Gal3, a transcriptional inducer (Platt, Ross et al. 2000). Thus, experiments should be performed to demonstrate that these homologous sequences are functionally conserved. https://en.wikipedia.org/wiki/Protein_function_prediction - cite_note-Platt2000-7 Here, I showed MUG4 are functionally conserved between *A. thaliana* and *C. sativa* by confirming *C. sativa* lines transformed with AtMUG4 are salt tolerant.

The difficult transformation protocols, complicated genetic systems, and tedious laboratory culture prevents deep gene functional research in crops. Thus, an improved understanding of Arabidopsis would help in the breeding of commercial crops. In the year 2015, Douglas R. Hoen discovered a large number of ETEs in *A. thaliana* (Hoen and Bureau 2015). In the year 2017, Zoé Joly-Lopez showed these ETEs are associated with abiotic stress phenotypes (Joly-Lopez, Forczek et al. 2017). In this paper, I applied MUG4 as a detailed example, showed its vital role to improve superior agronomic traits in salt tolerance of an oil crop, which opens a door for the great agriculture potential of ETEs.

Figures and tables

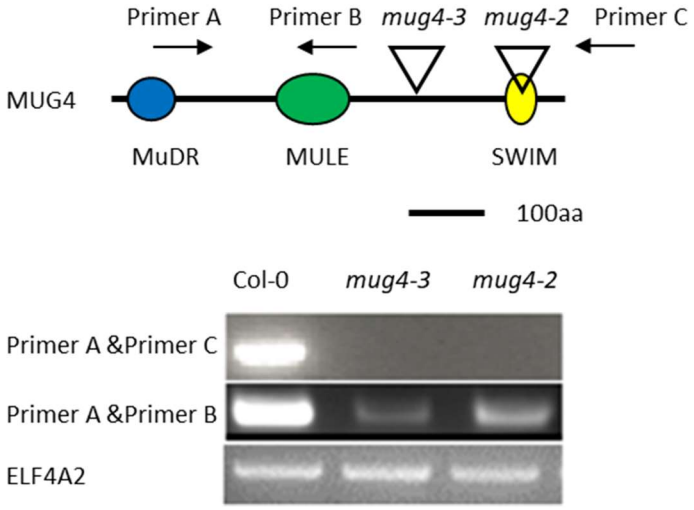


Figure 3.1. The absence of MUG4 expression in T-DNA mutant. Gene structure of *MUG4* and the T-DNA insertion information. The ovals indicate the domains of MUG4. T-DNA insertions are shown in triangles. SALK_036244C (*mug4-3*) is located between the MULE domain and SWIM domain, SALK_034608C (*mug4-2*) is located at the SWIM domain. RT-PCR shows both two T-DNA insertions lead to truncated MUG4 mRNA expression.

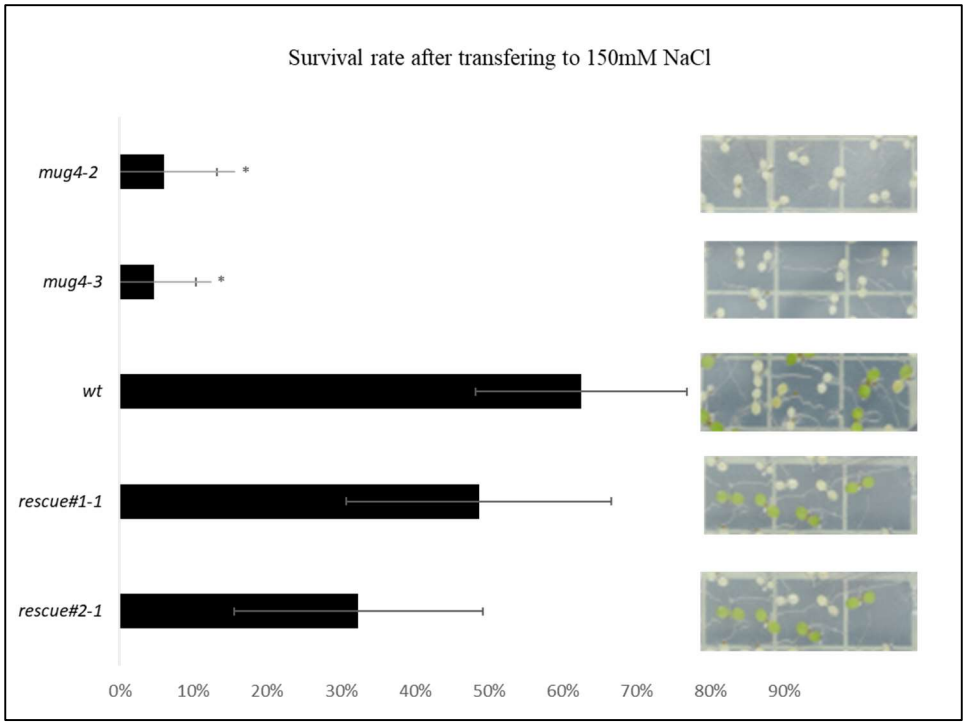


Figure 3.2 The survival rate of several lines under the treatment of 150 mM NaCl. Two *mug4* mutant lines, rescue lines, and overexpression lines, as well as wild-type, were grown on the 1/2 MS media. After 4 days, the seedlings were transferred into 1/2 regular MS and salt medium. The survival rate is calculated after 6 days treated with salt. The survival trait is indicated by cotyledon bleach. n= 36 seedlings. 3 replicates with consistent results were performed. * Indicates $P<0.05$ compared to wild-type.

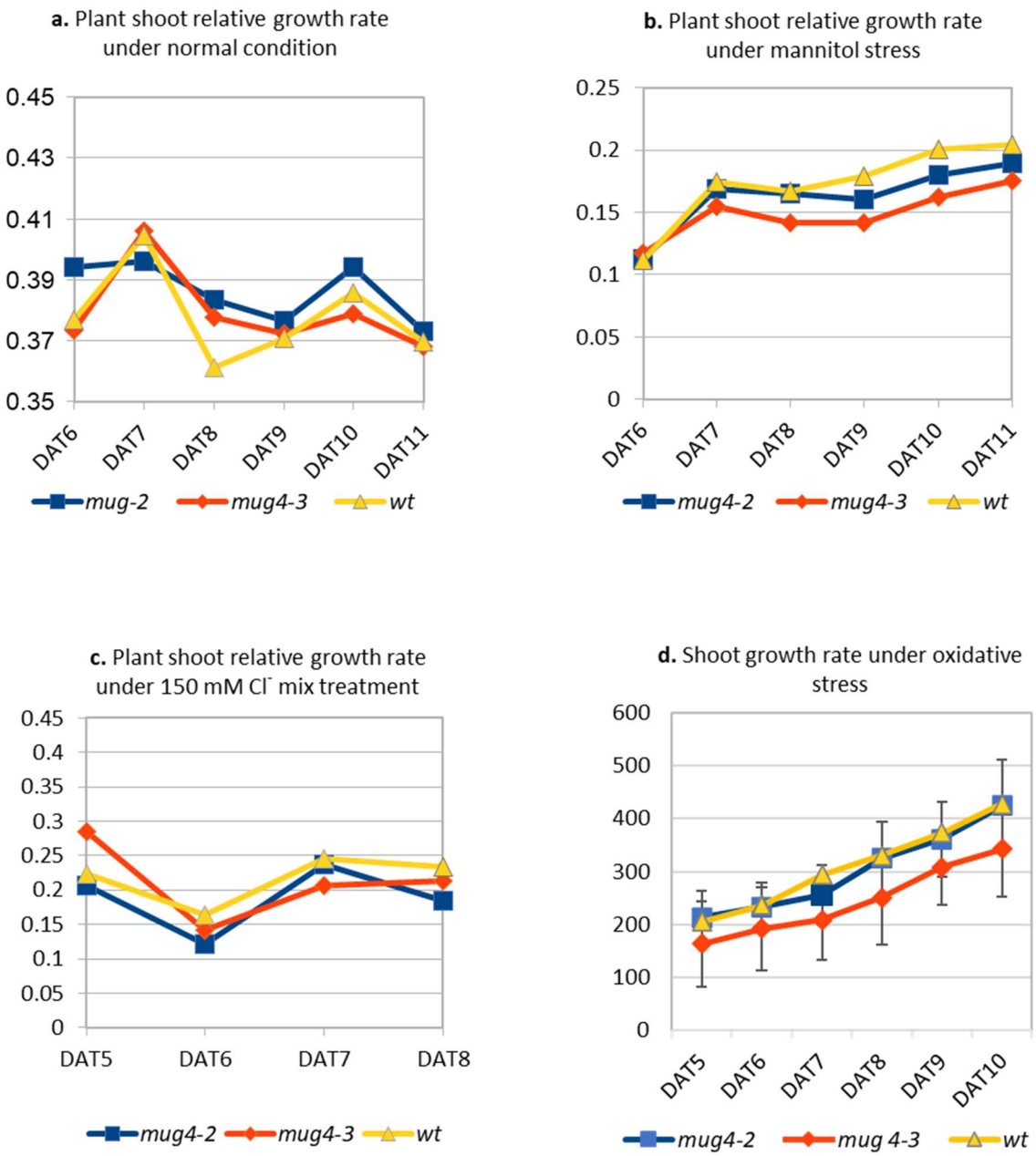


Figure 3.3 *mug4* mutants have the same relative growth rate as wild-type under osmotic stress, Cl^--dominant stress and oxidative stress. Plants were grown on the 1/2 MS media for 4 days then transferred into 1/2 MS and stress medium. The leaf areas were recorded every day after

transferring (DAT). The relative growth rate (RGR) is calculated with the formula: RGR = (ln W2 – ln W1) / (t2-t1).

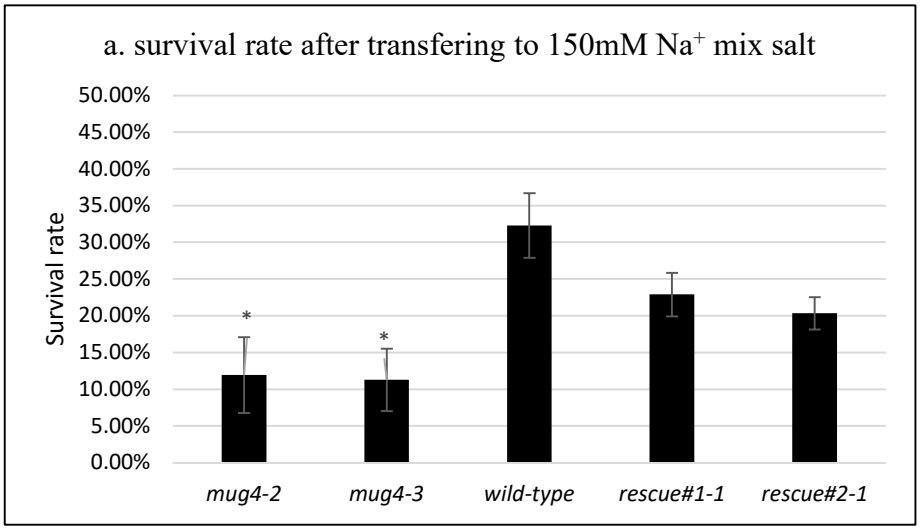


Figure 3.4a The survival rate of plants under the treatment of 150 mM Na^+ - dominant salt mix. Plants were grown on the 1/2 MS media for 4 days then transferred into 1/2 regular MS and salt medium. The survival rate is calculated after 6 days treated with salt. The molarity is indicated by cotyledon bleach. n= 36 seedlings. 3 replicates with consistent results were performed. * indicates $P<0.05$ compared to wild-type.

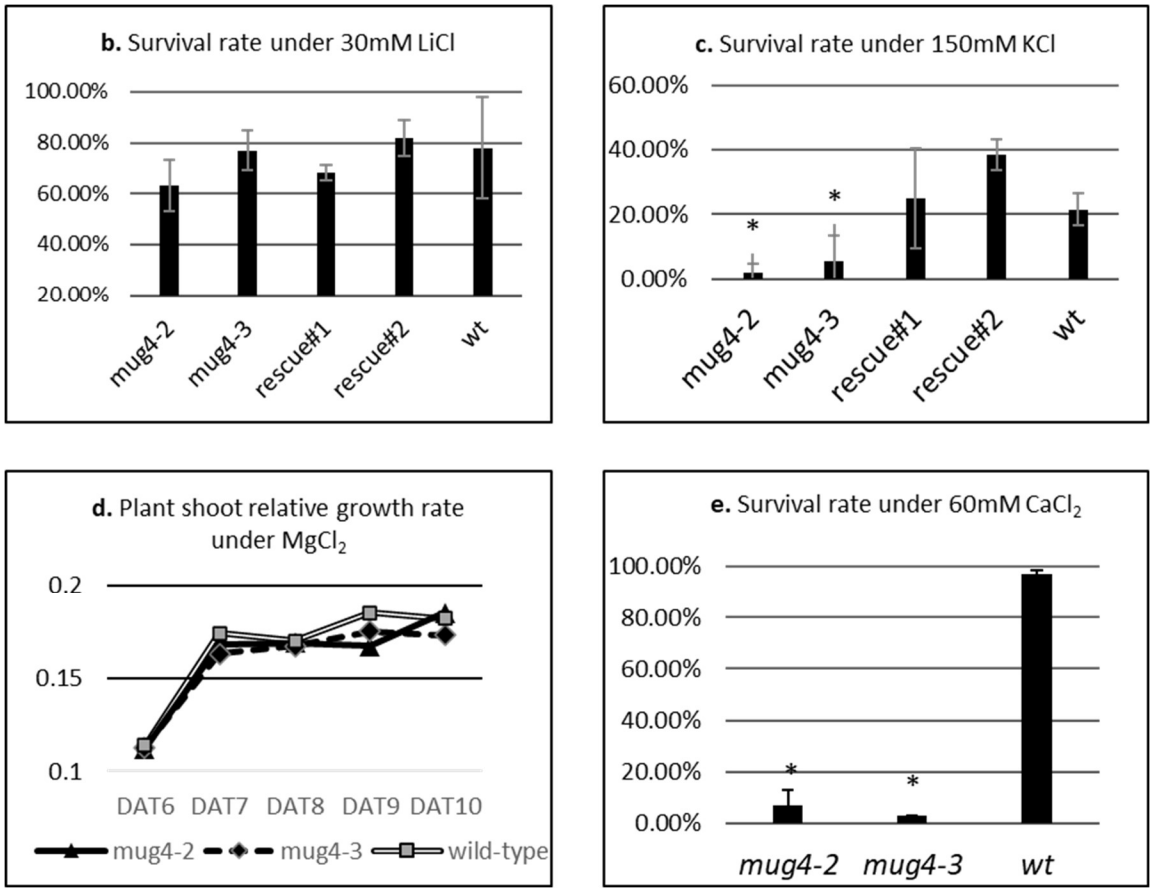


Figure 3.4b-4e Plant survival rate under LiCl, KCl, $CaCl_2$ and relative growth rate under $MgCl_2$.

* indicates $P<0.05$ compared to wild-type.

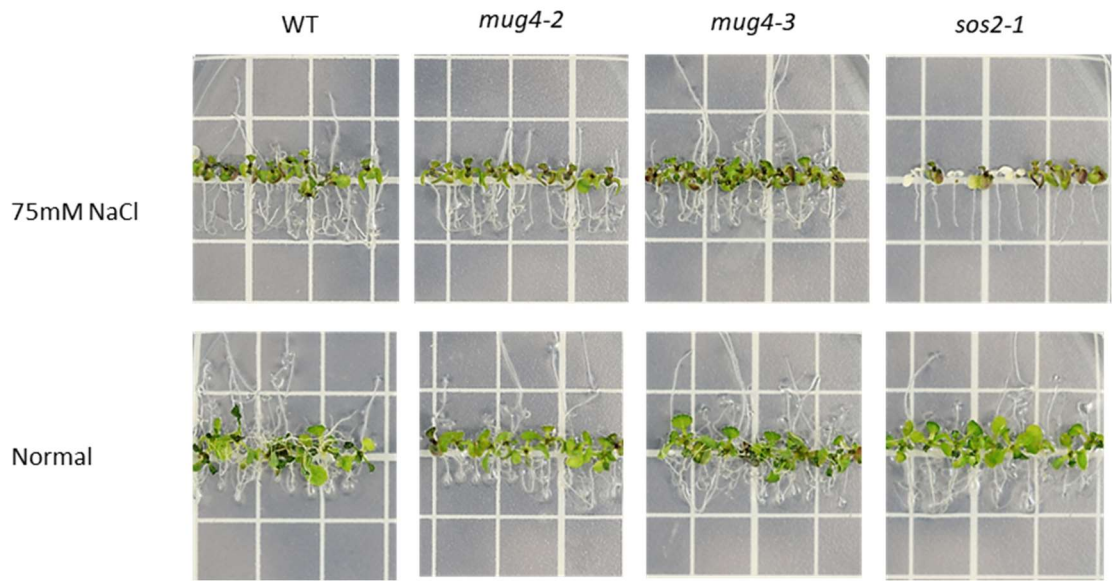


Figure 3.5 Root bending assay. Wild-type, *mug4-2*, *mug4-3*, and *sos2-1* seeds were grown on normal medium vertically, 4 days later the seedlings were transferred to 50, 75, 100mM NaCl medium and the plates were put upside down from the normal medium direction. After 6 days the pictures are taken. This figure shows the representative picture in 75mM NaCl and normal condition.

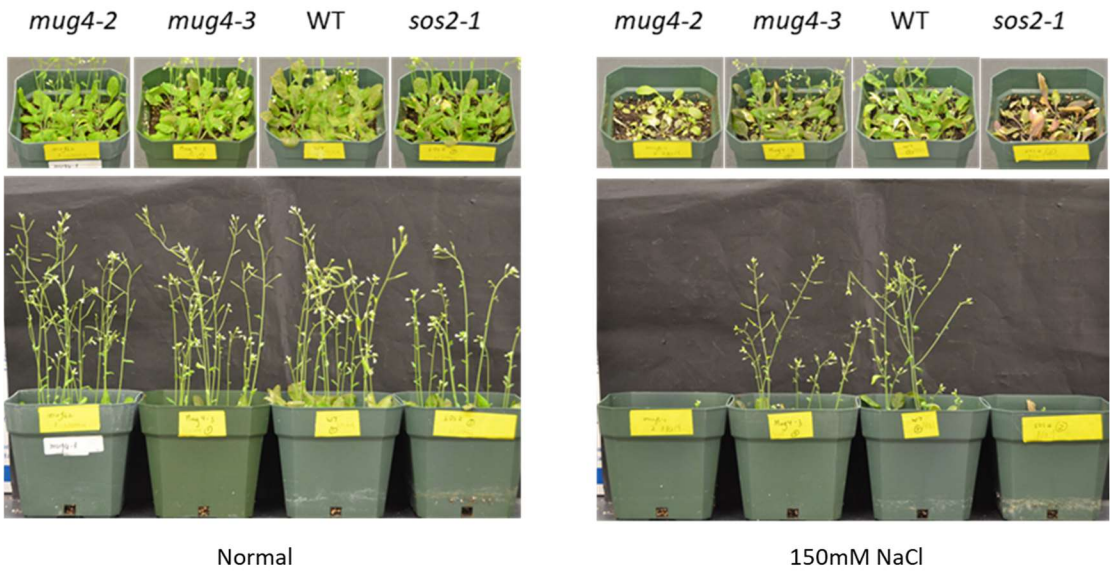


Figure 3.6 Salt assay during rosette growth stage. Wild-type, *mug4-2*, *mug4-3*, and *sos2-1* were grown on soil for 15 days then watered with 150 mM NaCl solution and repeat watering after 4 days. 7 days later plants were photographed.

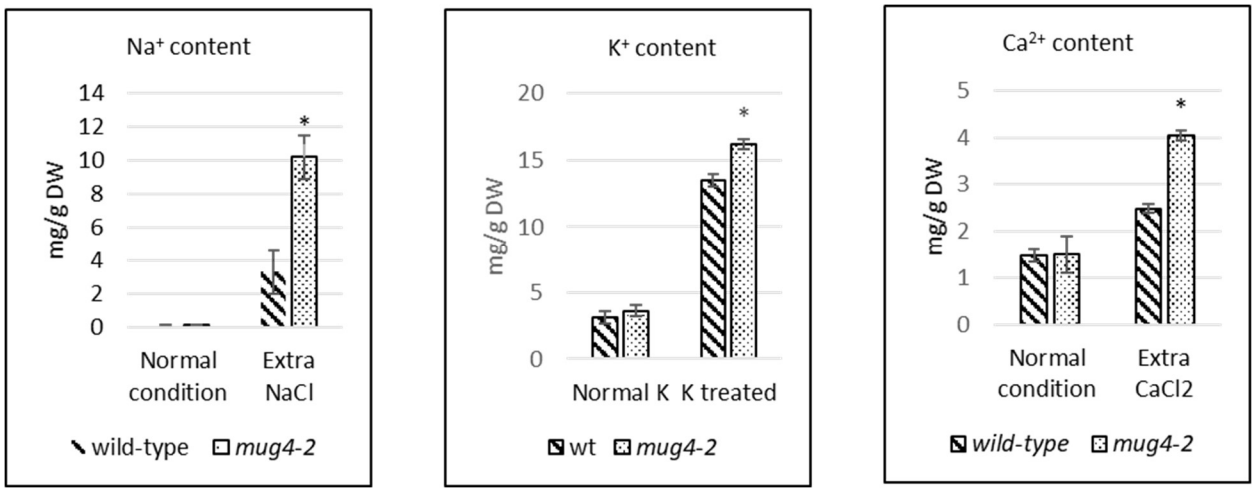


Figure 3.7 The Na, K, Ca content in *mug4-2* mutant and wild-type under normal conditions and 300 mM NaCl, 300 mM KCl, and 150 mM $CaCl_2$. DW, dry weight. Error bars represent SD (n =5). * indicates $P<0.05$ using wild-type in each treatment as reference.

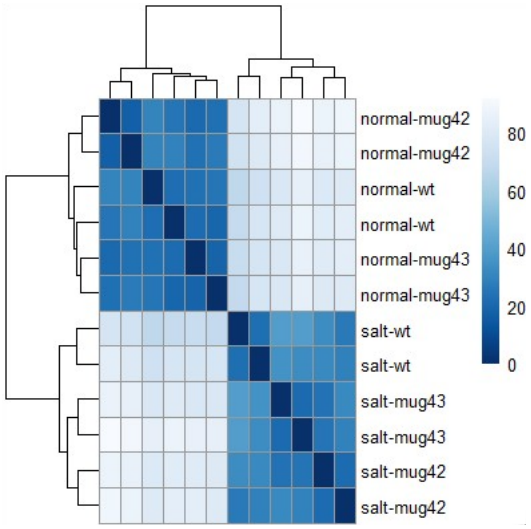


Figure 3.8. Sample distance cluster of the normalized data set. All the 12 samples clustered based on the sample distance. X-axis and Y-axis are the sample cluster method. The unit is the numeric distance among the samples. *mug4-2*, *mug4-3* samples cluster together and are separated from wild-type under salt conditions, while *mug4-3* samples mingle with wild-type in the normal condition.

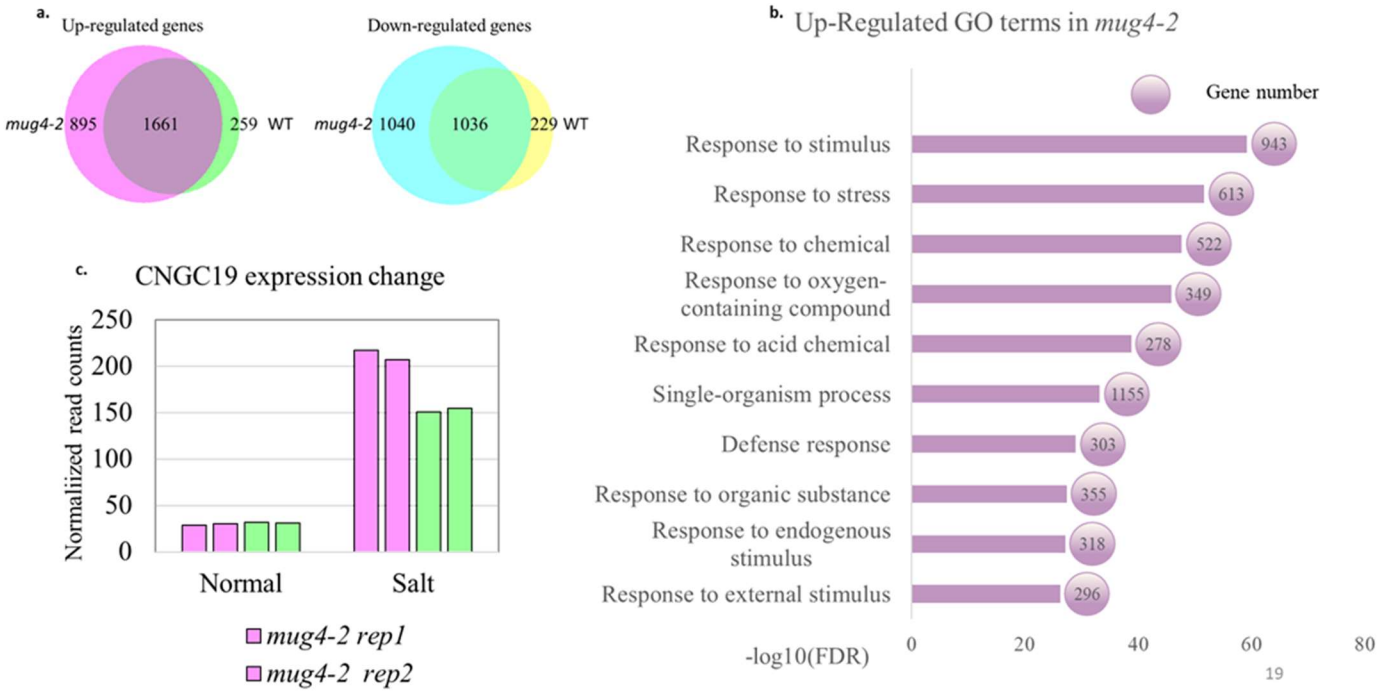


Figure 3.9. RNA-Seq analysis of the differentially expressed genes in the *mug4-2* mutant under salt treatment a. Comparison of genes that were up-regulated and down-regulated in the mug4-2 mutant and wild-type under salt treatment. b. The GO term enrichment analysis of genes that were diffenritentlly expressed in mug4-2.c. The normalized expressing level of CNGC19 in wild-type and mutants under normal and salt conditions, respectively.

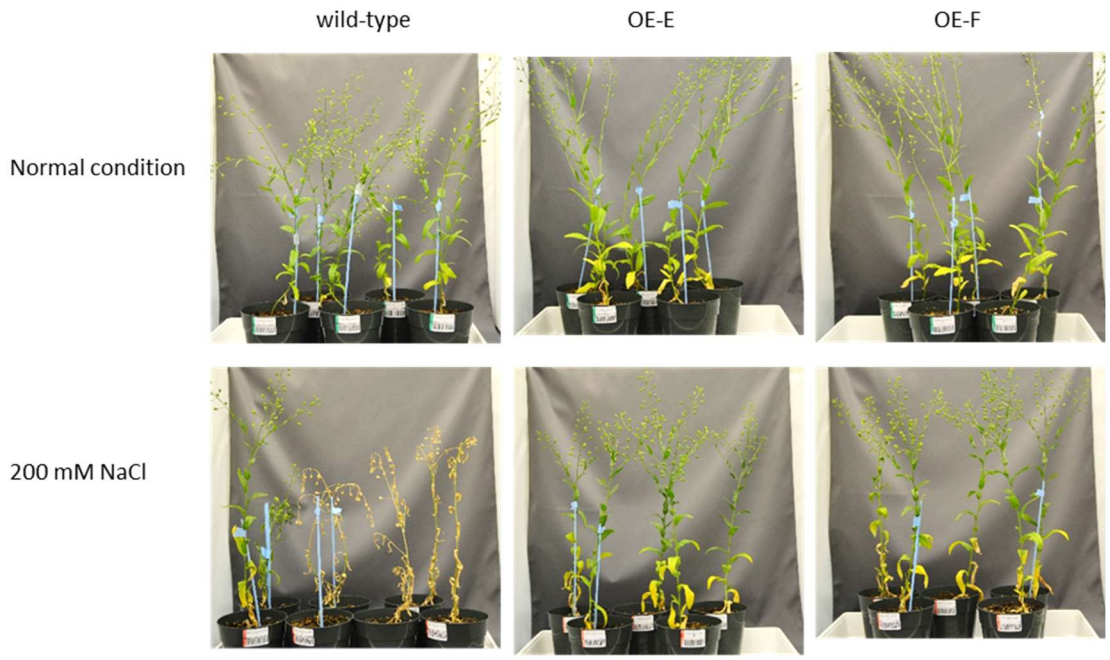


Figure 3.10. Representative photograph of the *C. sativa* transgenic lines and wild-type grown in normal soil and under 200 mM NaCl treatment. The AtMUG4 overexpression lines (OE-E and OE-F) and wild-type both grow well in the normal pots (upper panel). While wild-type has an inhibited growth after salt treatment, the transgenic lines OE-E and OE-F) are more tolerant to NaCl (lower panel).

Methods

Plant materials and growth facilities

The *mug4-2* (SALK_034608C), and *mug4-3* (SALK_036244C) mutant *A. thaliana* plants were obtained from Arabidopsis Biological Resource Center (ARBC). The primer for sequencing the border of T-DNA insertions is LBb1.3, ATTTTGCCGATTTCGGAAC. The RT-PCR primer A: MUG4-3 RP, CGGTTGTTTATTGCGTATCG; Primer B: mug4-3-rt-rp, CCATCATCTGTATGATTGGGAGTTC; Primer C: Mug4-lp, GAATCTCTCCAATCATCTGC.

Plants were grown in growth chamber either in the lab (Conviron model A1000, Canada) or Phytotron located at McGill University, Department of Biology (Montreal, Quebec, Canada; http://www.biology.mcgill.ca/Phytotron).

Construction of transgenic plants

Arabidopsis thaliana transgenic line generation:

Arabidopsis thaliana mug4-2 and Col-0 were grown in soil under greenhouse conditions in McGill greenhouse. The full size of Arabidopsis MUG4 (AtMUG4) cDNA was cloned into the pEarleyGate 103, which is a Gateway ® -compatible plant transformation vector with GFP and 6xHis C-terminal tags. It allows the selection using kanamycin in bacteria, and BASTA in plants. The *35S:MUG4: GFP* construct was transformed into *Agrobacterium tumefaciens* (strain GV3101) by electroporation and the *Agrobacterium* strain that harbours *MUG4* was transferred into *mug4-2* and *A. thaliana* ecotype Col-0 by floral dip. Resistance to BASTA was used to

screen T1 plants. Two stable transgenic lines per background were chosen and confirmed by genotyping and progeny tests.

Camelina sativa transgenic lines generation:

Camelina sativa variety Celine plants were grown in soil under greenhouse conditions in McGill greenhouse. Agrobacterium tumefaciens strain GV301 carrying the above-mentioned binary vector was used to infect *Camelina sativa*. Plants with inflorescences were selected for floral dip transformation. 1 weeks later, repeat the floral dip to increase the transformation rate. In total ~8000 T0 seeds were collected. 1000 seeds were used for screening the transgenic lines. The seeds have 100% gemination rate on the normal MS media. By screening on BASTA, the transformation rate is determined, which is ~ 0.8%. 5 independent T1 transgenic lines were successfully screened on BASTA medium. Since the first-generation lines are often heterozygous, each T1 line were chosen 50 seeds on BASTA medium to run the progeny test. 10 BASTA-positive plants of each line were then selected for T2 seeds collection. Later, the same process was repeated until the stable T4 or T5 seeds were collected in which segregation can't be detected on BASTA media. The stable lines were also tested by genotyping, RT-PCR, and GFP signal detection to confirm the insertion and expression of the AtMUG4 in Camelina. Finally, two randomly picked lines were chosen for the salt test.

Stress assay in *A. thaliana*

The multiple stress assays were adapted from (Joly-Lopez, Forczek et al. 2017). Seeds were sown on square Petri plates with grid (Simport, Canada) containing ½ MS medium buffered at

pH 5.6 with 2.5mM 2-(N-morpholino) ethane sulfonic acid (MES), and 0.8% agar. After 3 days stratification at 4℃ in darkness, the plates were placed in growth chamber programmed for a 16 h light / 8 h dark photocycle (22℃, 70% humidity). To induce stress, 4-day old seedlings were transferred to ½ MS medium supplemented with 300 mM mannitol, an osmolarity same to 150mM NaCl (Gagné 2014); 150 mM Na^+- dominant, made by 22.5 mM Na_2SO_4, 22.5 mM Na_2HPO_4, and 60 mM $NaNO_3$; Cl^- - dominant salt, made by 22.5 mM $CaCl_2$, 22.5 mM $MgCl_2$, and 60 mM KCl according to (Tavakkoli, Rengasamy et al. 2010); 6mM H_2O_2; 150mM KCl, 30 mM LiCl, 60 mM $CaCl_2$ and 75 mM $MgCl_2$.

Ion content measurement

Ion content measurement was adopted from (Shi, Xiong et al. 2002). Plants were grown for 4 weeks in promix-bx soil. Salt treatments were performed by immersing the pots in 300 mM NaCl, KCl and 150 mM $CaCl_2$ solution for 1-4 days. The plants were then harvested for ion content measurement. Total plants were collected and dried at 80 ℃ for at least 2 days and weighed. The samples were digested with HNO3, and the Na, K, Ca concentrations were assayed by atomic emission spectrophotometry in McGill Nutrient and Trace Element Analysis Laboratories.

Stress assay in *C. sativa*

Seeds were grown on ½ MS media for 7 days and then transferred to pots in the greenhouse. Each line grows 10 pots. After 28 days, 5 pots were treated with 200 mM NaCl, and 5 pots were treated with the same amount of water. One week later, another treatment was repeated.

RNA-seq analysis

4-day-old *mug4-2*, *mug4-3*, and wild-type seedlings were grown on MS medium were transferred to MS medium with or without the supply of 150 mM NaCl. RNA of two replicates of *mug4-2*, *mug4-3*, and wild-type samples under the normal condition and salt conditions were extracted from the seedlings in the whole plates after 4 days and transferred to salt plates or normal control plates using the protocols described in chapter 2. In total, 12 separate libraries were generated. The reads generated by the Illumina Hiseq2000 were initially processed to remove adapter sequences and low-quality bases. The reads were then mapped to the *Arabidopsis thaliana* TAIR 10 reference genome using HISAT2. Approximately 55 Gib with each read 50 nucleotides long were obtained. On average, 98.94% of the total reads were aligned to the reference genome. The aligned sequences were then sent for gene read counting using featureCounts with a reference annotation in TAIR (https://www.arabidopsis.org). Gene expression normalization among samples, sample distance cluster and differential expressed genes analysis was performed by using DESeq2. The GO enrichment was performed by AgriGO (Tian, Liu et al. 2017).

References

Arkhipova, I. and M. Meselson (2000). "Transposable elements in sexual and ancient asexual taxa." Proceedings of the National Academy of Sciences **97**(26): 14473-14477.

Boeke, J. D., D. J. Garfinkel, C. A. Styles and G. R. Fink (1985). "Ty elements transpose through an RNA intermediate." Cell **40**(3): 491-500.

Boissinot, S., A. Entezam and A. V. Furano (2001). "Selection against deleterious LINE-1-containing loci in the human lineage." Molecular biology and evolution **18**(6): 926-935.

Bundock, P. and P. Hooykaas (2005). "An Arabidopsis hAT-like transposase is essential for plant development." Nature **436**(7048): 282-284.

Chuong, E. B., N. C. Elde and C. Feschotte (2017). "Regulatory activities of transposable elements: from conflicts to benefits." Nature Reviews Genetics **18**(2): 71.

Claeys, H., S. Van Landeghem, M. Dubois, K. Maleux and D. Inzé (2014). "What is stress? Dose-response effects in commonly used in vitro stress assays." Plant Physiology **165**(2): 519-527.

Demidchik, V. and F. J. M. Maathuis (2007). "Physiological roles of nonselective cation channels in plants: from salt stress to signalling and development." New Phytologist **175**(3): 387-404.

Doolittle, W. F. and C. Sapienza (1980). "Selfish genes, the phenotype paradigm and genome evolution." Nature **284**(5757): 601-603.

Gagné, F. (2014). Chapter 2 - Tissue Preparation and Subcellular Fractionation Techniques. Biochemical Ecotoxicology. F. Gagné. Oxford, Academic Press**:** 21-31.

Greenblatt, I. M. and R. A. Brink (1963). "Transpositions of Modulator in maize into divided and undivided chromosome segments." Nature **197**(4865): 412-413.

Guan, Q. M., J. M. Wu, X. L. Yue, Y. Y. Zhang and J. H. Zhu (2013). "A Nuclear Calcium-Sensing Pathway Is Critical for Gene Regulation and Salt Stress Tolerance in Arabidopsis." Plos Genetics **9**(8).

Hoen, D. R. and T. E. Bureau (2015). "Discovery of Novel Genes Derived from Transposable Elements Using Integrative Genomic Analysis." Molecular Biology and Evolution **32**(6): 1487-1506.

Hudson, M. E., D. R. Lisch and P. H. Quail (2003). "The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway." Plant Journal **34**(4): 453-471.

Isayenkov, S. V., S. A. Dabravolski, T. Pan and S. Shabala (2020). "Phylogenetic Diversity and Physiological Roles of Plant Monovalent Cation/H(+) Antiporters." Front Plant Sci **11**: 573564.

Isayenkov, S. V. and F. J. M. Maathuis (2019). "Plant Salinity Stress: Many Unanswered Questions Remain." Front Plant Sci **10**: 80.

Joly-Lopez, Z. and T. E. Bureau (2018). "Exaptation of transposable element coding sequences." Current opinion in genetics & development **49**: 34-42.

Joly-Lopez, Z., E. Forczek, D. R. Hoen, N. Juretic and T. E. Bureau (2012). "A Gene Family Derived from Transposable Elements during Early Angiosperm Evolution Has Reproductive Fitness Benefits in Arabidopsis thaliana." Plos Genetics **8**(9).

Joly-Lopez, Z., E. Forczek, E. Vello, D. R. Hoen, A. Tomita and T. E. Bureau (2017). "Abiotic Stress Phenotypes Are Associated with Conserved Genes Derived from Transposable Elements." Front Plant Sci **8**: 2027.

Kagale, S., C. Koh, J. Nixon, V. Bollina, W. E. Clarke, R. Tuteja, C. Spillane, S. J. Robinson, M. G. Links, C. Clarke, E. E. Higgins, T. Huebert, A. G. Sharpe and I. A. P. Parkin (2014). "The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure." Nature Communications **5**(1): 3706.

Kagale, S., S. J. Robinson, J. Nixon, R. Xiao, T. Huebert, J. Condie, D. Kessler, W. E. Clarke, P. P. Edger, M. G. Links, A. G. Sharpe and I. A. P. Parkin (2014). "Polyploid Evolution of the Brassicaceae during the Cenozoic Era " The Plant Cell **26**(7): 2777-2791.

Kidwell, M. G. and D. R. Lisch (2001). "Perspective: transposable elements, parasitic DNA, and genome evolution." Evolution **55**(1): 1-24.

Kim, J., S. H. Moon, Y. C. Shin, J. H. Jeon, K. J. Park, K. P. Lee and I. So (2016). "Intracellular spermine blocks TRPC4 channel via electrostatic interaction with C-terminal negative amino acids." Pflugers Arch **468**(4): 551-561.

Munns, R. and M. Tester (2008). "Mechanisms of salinity tolerance." Annu. Rev. Plant Biol. **59**: 651-681.

Negrao, S., S. M. Schmockel and M. Tester (2017). "Evaluating physiological responses of plants to salinity stress." Annals of Botany **119**(1): 1-11.

Orgel, L. E. and F. H. Crick (1980). "Selfish DNA: the ultimate parasite." Nature **284**(5757): 604-607.

Pasyukova, E., S. Nuzhdin, T. Morozova and T. Mackay (2004). "Accumulation of transposable elements in the genome of Drosophila melanogaster is associated with a decrease in fitness." Journal of Heredity **95**(4): 284-290.

Platt, A., H. C. Ross, S. Hankin and R. J. Reece (2000). "The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase." Proceedings of the National Academy of Sciences **97**(7): 3154-3159.

Samanta, A., T. E. T. Hughes and V. Y. Moiseenkova-Bell (2018). "Transient Receptor Potential (TRP) Channels." Subcell Biochem **87**: 141-165.

Shavrukov, Y. (2013). "Salt stress or salt shock: which genes are we studying?" Journal of Experimental Botany **64**(1): 119-127.

Shi, H., L. Xiong, B. Stevenson, T. Lu and J. K. Zhu (2002). "The Arabidopsis salt overly sensitive 4 mutants uncover a critical role for vitamin B6 in plant salt tolerance." Plant Cell **14**(3): 575-588.

Tavakkoli, E., P. Rengasamy and G. K. McDonald (2010). "High concentrations of Na+ and Cl– ions in soil solution have simultaneous detrimental effects on growth of faba bean under salinity stress." Journal of Experimental Botany **61**(15): 4449-4459.

Thomas, J., M. Sepahi, B. Arendall and H. Bohnert (1995). "Enhancement of seed germination in high salinity by engineering mannitol expression in Arabidopsis thaliana." Plant, Cell & Environment **18**(7): 801-806.

Link between Chapter III and Chapter IV

In Chapter III, beginning with salt-sensitive phenotype of AtMUG4 T-DNA insertion mutants, I performed a set of physiological assays, revealed MUG4 specifically prevents plants from accumulating extra sodium, potassium, and calcium. The expression of abiotic stress-responsive genes was also significantly higher in *mug4* mutant plants than wild-type plants under salt stress conditions. Overexpressing AtMUG4 in *A. thaliana* and *Camelina sativa* lead both the two species more tolerant to salt. *MUG4* is an exapted transposable element, having three main structural modules: N-terminal transposase MURA-like domain, followed by MULE catalytic-like domain and C-terminal SWIM-type zinc finger domain. In Chapter IV, I did a deeper investigation based on the result of Chapter III to understand how each domain of MUG4 works in salt stress. Using site-directed mutagenesis approach, I dissect the functional roles of these three MUG4 domains, such as subcellular localization, heterodimer formation characterized in Chapter II. Also, I generated transgenic plants with the targeted mutation to determine the importance of each domain in Arabidopsis salt defence. In the end, this work combined Chapter II and Chapter III, revealed discrete roles of the multiple domains of MUG4 and provide functional support for how MUG4 works in salt tolerance as an ETE.

Chapter IV

Functional dissection of Arabidopsis MUG4 reveals essential roles of its three domains

Abstract

Salinity is a major abiotic stress threatening the growth of plants. I recently showed that *Arabidopsis thaliana MUSTANG4 (MUG4)* is essential for plant salinity defence, and its protein products are required for preventing extra toxic cation accumulation. Unlike the ancestor *mudrA* gene that encodes the transposase MURA that is required for the mobility of *Mutator*-like elements (MULEs), *MUG4* is an exapted transposable element (ETE). It has three main structural modules: N-terminal transposase MURA-like domain, followed by MULE catalytic-like domain and C-terminal SWIM-type zinc finger domain. A systematic analysis using a site-directed mutagenesis approach was performed to dissect the functional roles of these three MUG4 domains, such as subcellular localization, and heterodimer formation. Transgenic plants with the targeted mutation were generated to determine the importance of each domain in Arabidopsis salt defence. The results show MUG4 shares the conserved C2HC zinc finger pocket at DNA sequence level as well as the protein structure level. Also, MUG4 has structural similarity in the N-terminal domain to the AP2/ERF gene family. MUG4 middle domain lacks the original DDE motif that most DNA transposase used for catalyzing the "cut and paste" transposition reaction. The N-terminal C2HC zinc finger domain is essential for the nucleus subcellular localization of MUG4 and the E254 mutation in the middle domain partially abolished the nuclear targeting. All three domains play critical roles in the interaction of MUG4 with MUG1 and MUG2. In addition, the N-terminal and C-terminal domains are important for plant salt response and the middle domain might be required for optimal salt defence. Furthermore, the ability to defend the high salinity of MUG4 largely correlates its heterodimer

formation with MUG1 and MUG2. Together, this work described a structural basis of MUG4, revealed discrete roles of the multiple domains of MUG4 and provide functional support for how MUG4 works in salt tolerance as an ETE.

Introduction

As sessile organisms, plants have to recruit their troops to cope with the constantly changing environment, especially adverse conditions for the growth and development, such as soil salinity, drought, high temperature, etc. Among them, salt is a major environmental factor that threatens plant survival in nature, and limits plant productivity in agriculture. In recent years, the adverse effects of salt stress are exacerbated by increased salt accumulation that is caused by inadequate irrigation and seawater indwelling (Wang, He et al. 2021). How plants use strategies that apply a wide spectrum of programs to adapt to the salt stress is a fundamental biological question. Furthermore, such insight is critical for enhancing plant salt tolerance as well as agricultural productivity.

A few strategies on how plants naturally cope with the salt perturbations have been found, including selectively accumulating or excluding the ions, controlling ion up taking and transporting the ions into leaves, ion compartmentalization to vacuoles, compatible solutes generation, photosynthetic pathway alternation, membrane structure change, antioxidative enzymes and plant hormones induction (Parida and Das 2005). In the early years, forward genetics was applied to discover certain genetic basis in the strategies that plants respond to salt through a phenotype-based approach, such as the finding of Arabidopsis abscisic acid insensitive

1(*ABI1*) gene (Leung, Bouvier-Durand et al. 1994), salt overly sensitive 1 (*SOS1*) gene (Wu, Ding et al. 1996).

In the recent two decades, sequencing technologies have improved and reduced in cost, providing a wealth of genetic code information. The largest component of the genetic code in most eukaryotes are transposable elements (TEs) (Feschotte, Jiang et al. 2002). For example, transposable elements make up 44% of the human genome (Mills, Bennett et al. 2007). In maize, 85% of the genome is annotated as TEs (Jiao, Peluso et al. 2017). Although some transposable elements appear to be non-functional, acting as 'selfish' sequence elements, studies reported at times, these sequences also create novel functions. Especially under environmental stress conditions, the plant transposons tend to be more active for transposition in response to specific environmental stimuli (Flavell, Pearce et al. 1994). Some of them work as regulatory elements, reprogramming host genes circuits and ultimately fine-tuning the host response to environmental challenges (Makarevitch, Waters et al. 2015). Moreover, the stress-induced changes in epigenetic status of TE activity allows TEs to propagate their stress-responsive elements to host genes (Negi, Rai et al. 2016), later on, those transposable elements may settle down and start their new identity as conventional genes. Those transposable elements which exapted into conventional gene functions are named exapted transposable elements (ETEs) (Hoen and Bureau 2015).

In plants, a few ETEs such as *FHY3*, *MUSTANG*, *DAYSLEEPER*, etc. were described to be essential for light signalling (Lin, Ding et al. 2007) and plant development (Bundock and Hooykaas 2005, Joly-Lopez, Forczek et al. 2012). However, no ETEs were well-documented that are critical to abiotic stress, except one study that shows the T-DNA insertion lines of ETEs

are associated with the abiotic stress phenotype (Joly-Lopez, Forczek et al. 2017). Following this, in Chapter III, I further explored the deeper mechanism of how MUG4, an ETE derived from *MUTATOR*, responds to salt in plants and showed that overexpressing MUG4 leads to more salt tolerance in both Arabidopsis and Camelina. In this study, I performed a systematic analysis using a site-directed mutagenesis approach to gain a better understanding of the function of the domain of MUG4 in response to salt and the molecular characteristics of MUG4. The results suggest that the multiple domains of MUG4 perform discrete but essential functions in ensuring its subcellular localization, heterodimer formation and salt response. Additionally, the results suggest that the localization, heterodimer formation and salt response of MUG4 are tightly correlated. Given their predominating presence in the plant genomes, ETEs hold unexplored potential for salt, even abiotic stress response improvement programs. This study applies the structural-functional approach, uses MUG4 as an example to illustrate the importance of ETEs in response to salt stress and highlights the potential value of the ETEs in agriculture.

Result

Determine the essential residues in each domain

As described in chapter II, Arabidopsis MUG4 is predicted to have three regions that are conserved to known a protein family in the NCBI conserved domain database: the N-terminal domain was predicted to be conserved with DBD_Tnp_Mut (DNA binding domain_Transposase_Mutator) superfamily, the central domain was conserved with DDE_Tnp_ISL3 (DDE motif for transpostion_Transposase_ISL3) superfamily and the C-

terminal domain was conserved with ZnF_PMZ (Zinc finger_ Plant Mutator transposase). DBD_Tnp_Mut domain was reported to have the putative DNA-binding function for MULE transposases (Liu and Wessler 2017), where it requires the C2HC zinc-chelating finger motif typified by the presence of a H×C in place of the WRKY H×H motif of the Zn-chelating metal ligands (Babu, Iyer et al. 2006). Previous domain analysis in chapter II suggested that MUG4 has a 34.04% identity to maize MURA in the N-terminal domain (Chapter 2, Figure 2.2). Thus, I aligned the amino acid sequences among MUG4, MURA and the C2HC type of WRKY–GCM1 domain family published in (Babu, Iyer et al. 2006). According to the alignment (Figure 4.1a), MUG4 shares the essential C2HC motif with MURA as well as the HxC family of WRKY-GCM1 domains.

Then I further explored whether the three-dimensional structure of MUG4 is conserved with the WRKY-GCM1 domains. SWISS-MODEL was applied to find a proper template for the full-length MUG4. No template in the protein database (PDB) has an identity of over 30% using the full-length MUG4 as the target, suggesting close homologs of MUG4 are absent. Among the WRKY-related templates that SWISS-MODEL provided, WRKY transcription factor 4 (PDB: 1wj2) has the highest identity, 25.64% with the N-terminal of MUG4. While the WRKY4 C-terminal (WRKY4-C) has been revealed a four-stranded β-sheet structure and a following zinc binding pocket (Yamasaki, Kigawa et al. 2005), the SWISS-MODEL result shows that MUG4 can only match three-stranded β-sheet structure with the WRKY4 template (Figure 4.2a). However, when looking deeper into the overlay of the two structures, the zinc finger binding pocket can be covered. Like the C2H2 residue of WRKY4 that forms a zinc finger pocket at the

ends of the 4th β-sheet, the CCHC residues of MUG4 also formed a zinc-binding pocket (Figure 4.2a). This result suggests that MUG4 might keep the function of zinc chelating. Thus, I generated two mutants: the conserved cysteine (C) residues (C62, C93), which are presumably involved in chelating the zinc atom, were changed into Ala (A) for further analysis. The threading/fold recognition-based tool RaptorX (Xu, Mcpartlon et al. 2021), as well as the *ab initio* structure prediction-based tool trRosetta (Yang, Anishchenko et al. 2020), were then applied to model the MUG4 structure to complement the result from homology modelling-based tool SWISS-MODEL. Both methods predicted an α-helix and a following four-stranded antiparallel β-sheet structure in the N-terminal of MUG4 (Figure 4.2b). In the previous study, I have shown MUG4 has the transcriptional repressor activity, thus, the structure of MUG4 was compared with other well-documented plant transcriptional regulators, including AP2/ERF, NAC, WRKY, B3 and SBP (Yamasaki, Kigawa et al. 2013). Among them, the AP2/ERF family genes, which were reported to have transcriptional repressor activity as well (McGrath, Dombrecht et al. 2005), have a three-stranded antiparallel β-sheet packed along with an α-helix structure (Yamasaki, Kigawa et al. 2013), which is similar to the N-terminal domain of MUG4 predicted by RaptorX and trRosetta (Figure 4.2b). This result supported the hypothesis that MUG4 is a transcriptional repressor at the structural level.

For the central domain, I aligned MUG4 as well as MUG1 to MUG3 with a subset of the reported DDE domains of Mutator (Yuan and Wessler 2011). As shown in Figure 4.1b, all the MUG-A members keep the conservation of the first Aspartic Acid (D) and the third Glutamic Acid (E) in the DDE motif with the Mutator transposase. Interestingly, while the second D has

been evolved into other amino acids, a conserved E residue among MUG-A members was found in the alignment (see Figure 4.1b). The 3D structure of the MUG4 middle domain was also predicted by the above-stated methods. As shown in Figure 4.2c, this middle domain contains a seven-stranded α+β core. The corresponding residues that aligned to the catalytic triad DDE of the ancient transposase (D219, E285 and E384 for MUG4) are shown using the stick model. Both trRosetta and RaptorX predicted a similar structure. Thus, I selected the first Aspartic Acid (D219) in the DDE motif for further analysis. Also, the conserved E (E254 for MUG4) among MUG-A members was selected for mutagenesis studies.

For the C-terminal sequences of MUG-A members, they were aligned with the well-documented ZnF_PMZ sequences (Makarova, Aravind et al. 2002). The essential CxCxnCxH motif could be aligned with confidence (Figure 4.1c). Since no SWIM-type zinc-finger (ZSWIM) structure is available in any protein structure database, it is not able to compare the MUG4 C-terminal structure with the reported SWIM-type zinc finger. As the CCHC motif is still conserved in sequence alignment, I chose the essential cysteine, C475 of MUG4 for further analysis. In summary, to further investigate the domain-function relationship of MUG4, mutagenesis studies were performed at the above described essential or conserved residues in each domain: C62, C93, D219, E254, C475.

The N-terminal domain and middle domain are critical for MUG4 nuclear subcellular location

Previous fluorescence microscopy analysis showed Arabidopsis MUG4 tagged with GFP localizes in the nucleus. To determine which domain affects the subcellular localization, Alanine-scanning mutagenesis was applied to the above-described residues. In addition, as the amino acid triad DDE is an acidic motif, so does E254, residue D219 and E254 were also mutated into the basic amino acid Lysine (K). Transient transformations of 35sp:: MUG4-GFP, 35sp:: C62A-GFP, 35sp:: C93A-GFP, 35sp:: D219A-GFP, 35sp:: D219K-GFP, 35sp:: E254A-GFP, 35sp:: E254K-GFP, 35sp:: C475A-GFP and 35sp:: GFP were performed in tobacco leaves and the respective GFP fluorescence signals were detected. As shown in Figure 4.3, the wild-type MUG4-GFP signal can only be detected in nucleus, so does C475A-GFP. While 35sp:: GFP free GFP, as well as the mutation of C62A, C93A in the DNA-binding domain, leads the GFP signal can be detected in both cytosol and nucleus. This result indicates that the putative DNA-binding domain is essential for MUG4 to localize in the nucleus. Interestingly, in the middle domain, D219 of the DDE motif, as well as E254 of MUG4 which share conserved E in the MUG-A family, both the Alanine and Lysine mutation alternate the GFP signal, showing the middle domain of MUG4 affects the nuclear localization stability.

Mutagenesis in each domain reduced the heterodimer formation of MUG4 with MUG1 and MUG2

The finding that MUG4 interacts with other MUG-A proteins (See chapter 2) drives me to understand which amino acids are essential for the interactions. *FHY3* is also an ETE derived from *mudrA*, as FHY3 related proteins and all the MUG-A proteins can form heterodimers, it is possible that the conserved residues, not their unique amino acids that endow the heterodimer

formation. Lin et al applied a series of mutations and found the D288A, C579A, H591A, D283N, and G305R of FHY3 are essential for the dimerization with FAR1 (Lin et al., 2008). Thus, I aligned the amino acid sequences among MUG-A proteins and FHY3. The C93A, D219, E254, C475, H487 of MUG4 are conserved with the C157, D288, E323, C579, H591 of FHY3, also the corresponding sites are conserved among other MUG-A proteins (See Figure 4.4a). According to the alignment, I performed a Y2H assay using the mutation of MUG4 N-terminal domain C93, middle domain D219, E254, and C-terminal SWIF domain C475 and H487, plus C62 since in MUG-A proteins, the corresponding site is C while in FHY3 is R125. Consistent with the previous results in chapter 2, both MUG1-BD and MUG2-BD interact with wild-type MUG4-AD that the corresponding transgenic yeast cells can grow well on the selective medium, develop blue colour with the supplement of X-α-Gal and activate the reporter genes. The yeast cells that expressed MUG1-BD and MUG2-BD co-transformed with MUG4 single amino acid mutations C62A, C93A, D219A, D219K, E254A, E254K, C475A, H487E that fused with Gal4-AD have limited growth on the selective medium, and dramatically reduced their activity by developing a very faint blue (Figure 4.4b). This result suggests all the three domains of AtMUG4 are essential for its dimerization with MUG1 and MUG2 in yeast cells.

All three domains of MUG4 are essential for plant salt stress

I then further tested the *in vivo* functional significance of MUG4 three domain mutations. Transgenic plants expressing these MUG4 mutant forms driven by MUG4 endogenous promoter in the *mug4-2* mutant background were generated and each mutation version was chosen in one or two stable lines to test. All the plants were grown under the normal condition as well as the

salt condition to test the potential phenotype. Both the wild-type as well as the mutation forms grow well under the normal condition. Consistent with the results in chapter 2, the wild-type mug4p::MUG4-GFP transgenic line successfully rescued the *mug4-2* salt-sensitive phenotype (Figure 4.5). However, all transgenes of the mutant forms failed to complement the *mug4-2* salt-sensitive phenotype (Figure 4.5), although their transcription levels were higher than that of the endogenous *MUG4* gene in a wild-type background (data not shown). Among the mutation forms, the N-terminal DNA-binding domain mutations have the highest mortality rate, followed by the C-terminal domain mutations. The middle domain mutation forms can partially rescue the salt-sensitive phenotype but are not as strong as the endogenous promoter-driven rescue lines. Together, these results suggest that all the three domains of MUG4 are essential for the salt defence function in planta.

Discussion

Essential roles of the N-terminal domain of MUG4 in subcellular localization, heterodimer formation and biological function

Eukaryotic protein translation occurs in the cytoplasm, except for a few proteins made in mitochondria and chloroplasts. To enter into the nucleus, one strategy for the nuclear protein is using the active nuclear localization sequence or nuclear localization signals (NLS) (Silver 1991). I previously showed that MUG4 has the predicted NLS (KKKVVRVENLKRPKRI, RIRPPKTRRPPGRPKKKVVRVENLKRPKR) at the C-terminal (Chapter 2, Figure 2.1a), which may be required for transporting MUG4 into the nucleus. In this work, I find that MUG4

mutations at the N-terminal domain also alter the localization of MUG4. As shown in the structure analysis, the N-terminal domain structure has the conserved C2HC zinc finger pocket. Also, MUG4 is derived from transposase MudrA, which originally has the DNA-binding ability. Both structure analysis and evolutionary history suggests that the MUG4 N-terminal domain is a DNA binding domain. Since C62 and C93 are essential residues for zinc chelating, these mutations probably reduce the strength of the interaction of MUG4 with the target DNA. As a result, the interaction between MUG4 and DNA becomes less stable, probably leading to a severe defect in protein trafficking that results in partial mislocalization in the cytoplasm. Consistent with this, transcriptional regulators localization is often altered by disrupting the DNA-binding function. The mutation of the POU domain in POU4F3, which disrupt its DNA binding function, altered the localization of POU4F3 from nucleus to both nucleus and cytosol (Weiss, Gottfried et al. 2003). So does Ikaros, from which the C2H2 type zinc finger DNA-binding domain is essential for Ikaros nucleus targeting (Cobb, Morales-Alcelay et al. 2000). Also, a report demonstrated that for Hermes transposase, the DNA-binding region is necessary for nuclear localization (Michel and Atkinson 2003). Another interesting finding is that point mutations in the middle domain and C-terminal of ETE FHY3, do not affect its nucleus localization (Lin, Teng et al. 2008), which indirectly hints that the mutation in the DNA-binding domain may disrupt its localization.

The Y2H results show that the mutations in the DNA binding domain reduced the interaction between MUG4 with MUG1 and MUG2. As it is found that the DNA binding domain mutation leads to a distribution of MUG4 in both nucleus and cytosol, which results in less amount of

MUG4 in the nucleus, as such, the available MUG4 to form heterodimers is reduced. Which could be the reason for a reduced interaction of MUG4 DNA-binding domain mutation form with MUG1 and MUG2 observation in the Y2H assay. Another possibility is that the mutations in the DNA-binding domain result in a change in MUG4 protein conformation, which leads to an unstable protein complex. A recent report for TE transposition, shows that ISCth4 transposase recruits an asymmetric dimer complex, which requires its DNA binding domain to recognize the terminal inverted repeat sequences (TIRs) of TEs then the transposase conformation changes, that the dimerization domain forms a highly asymmetric dimer that bond to the TIRs (Kosek, Hickman et al. 2021). Also, Arabidopsis WRKY18, WRKY40, and WRKY60 showed a correlation between the altered DNA binding activities and the formation of heterocomplexes of two WRKY proteins that result from cooperative binding of the proteins or a conformational change of one WRKY protein by the other interacting WRKY protein (Xu, Chen et al. 2006). Both examples suggest that the DNA-binding function is essential for heterodimer formation.

The disruption of the MUG4 DNA binding domain results in the failure of complementation for the salt sensitivity of the *mug4* mutant. As MUG4 is a transcriptional repressor, it is possible that MUG4 directly regulates salt-response genes. Moreover, because the DNA-binding function was disrupted, the regulation ability subsequently gets disrupted.

MUG4 middle domain is optimal for MUG4 subcellular localization and essential for heterodimer formation and biological function

As to the result that the mutations in the middle domain also alter the subcellular localization, as mentioned in the earlier paragraph, because the protein complex is not stable, MUG4 either cannot only partially form a functional protein complex or can not form the protein complex at all. Thus, more MUG4 proteins are dissociated from the complex and are free in the cytosol. I quantified each mutation line, the wild-type MUG4 as well as the free GFP nuclear to cytoplasmic (N/C) GFP signal ratio, then calculated relative N/C using MUG4-GFP signal as the reference. As shown in Figure 4.6, in the DNA-binding domain, both C62A and C93A mutation, the relative N/C GFP signal is dramatically decreased to the level of free GFP. In the middle domain, site D219 does not affect MUG4 subcellular localization significantly, however, site E254 partially altered the nuclear localization. The weaker effects on the middle domain in the DNA-binding domain on MUG4 nucleus localization suggest that the disruption of middle domains may be a secondary effect.

The ancestor transposase whose middle domain is used for catalyzing the mobilization of TE; the middle domain of FHY3 is used for transcriptional activation. Compared to them, the middle domain of MUG4 is keeping the heterodimer formation capacity, however, this domain does not have the catalyzation nor the transcriptional activation function. In this paper, we found that the mutation in the middle domain reduced the heterodimer formation, and partially, not fully complement the salt-sensitive phenotype of the MUG4 mutant.

In chapter 2, I have shown that MUG4 has the transcriptional suppression activity, and in the previous report, it is shown that the transcriptional activation activity of FHY3 is essential for the morphological phenotype of its mutants, thus, the observation of the reduction of heterodimer

formation of MUG4 with MUG1 and MUG2 and the partially rescued phenotype could work as the way that FHY3 does, which is, the transcriptional suppressor is essential for MUG4 mutation salt-sensitive phenotype, and the mutation in MUG4 middle domain leads to a reduced functional transcription suppression complex, which results in a partially rescued phenotype, even though some experiments need to be performed to verify.

MUG4 C-terminal domain may keep a conserved function mechanism with FHY3

As to the mutations in the C-terminal domain, they did not affect the subcellular localization of MUG4, which is consistent with FHY3, that the C-terminal SWIM motif amino acid substitutions on FHY3 do not alter the nuclear targeting (Lin, Teng et al. 2008). Also, the mutations in MUG4 are essential for the heterodimer formation of MUG4 with MUG1 and MUG2, which is the same with FHY3, that the corresponding domain in which is essential for FHY3 homodimerization or its heterodimerization with FAR1 (Lin, Ding et al. 2007, Lin, Teng et al. 2008). For the biological meaning of the SWIM type zinc finger motif, in MUG4, it is also essential for rescuing the salt-sensitive phenotype of the MUG4 mutant. In FHY3, it is also essential for FHY3 biological function. Even though the phenotype between MUG4 and FAR1 is different, the SWIM type zinc finger motif in the two proteins seems to keep a conserved functional mechanism.

Figures and tables

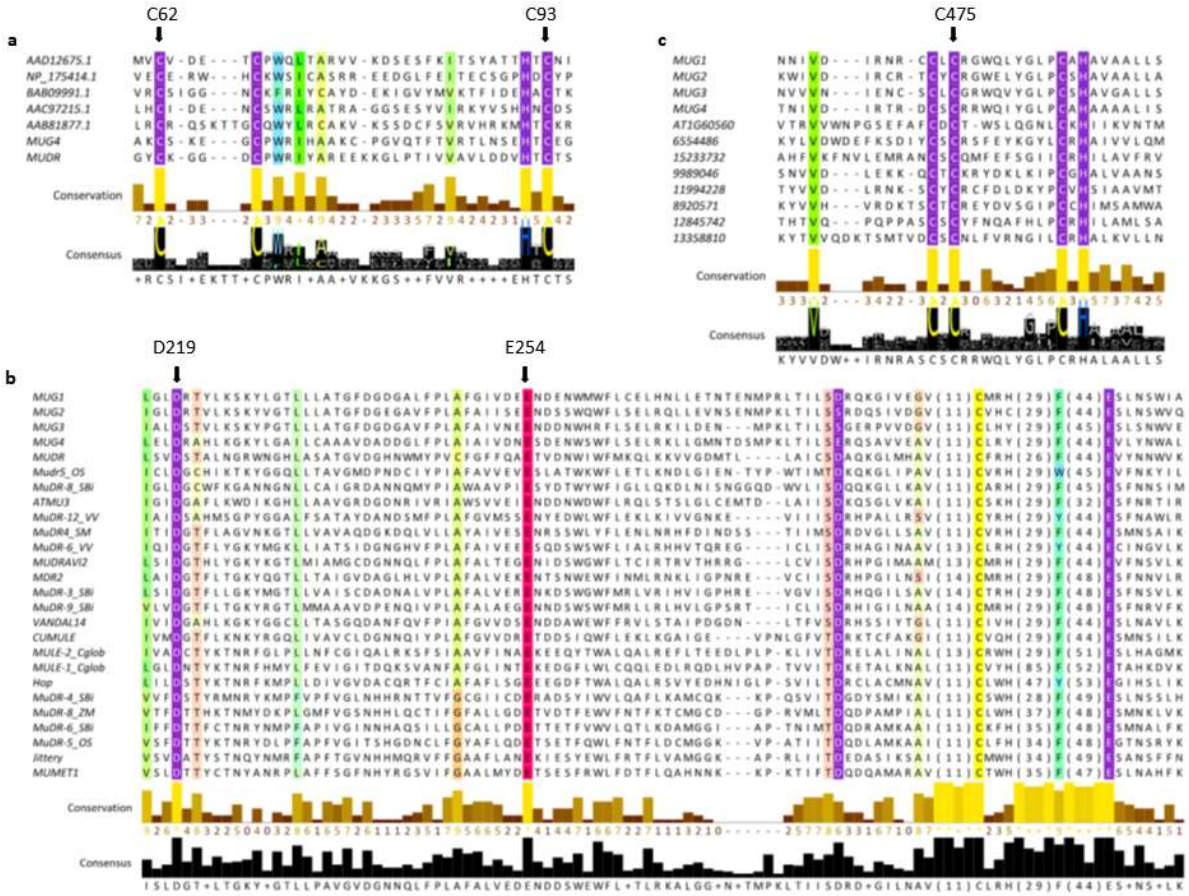


Figure 4.1. Multiple sequence alignment of **a**. the HxC family of WRKY–GCM1 domain family protein (Babu, Iyer et al. 2006), MUG4 and MUDR, sequences names are denoted by the NCBI accession number; **b**. MUG-A proteins and DDE domains of Mutator, sequences name followed (Yuan and Wessler 2011) **c**. MUG-A proteins and ZnF_PMZ sequences are described in (Makarova, Aravind et al. 2002). The colouring reflects the conservation profile at 50% consensus. Colouring scheme: Taylor. The sites shading in purple indicates the essential residues for each reported domain, a: CCHC motif, b: DDE motif, c: CCCH motif. The black arrow above the alignment marks the site position of MUG4 selected for mutagenesis studies.

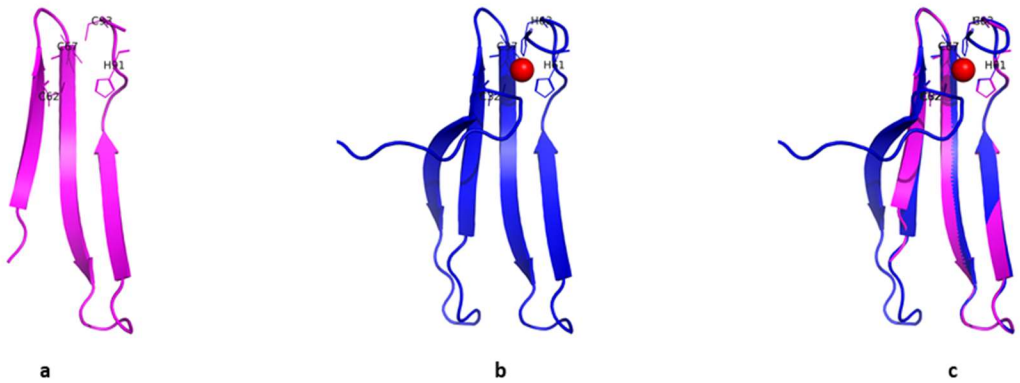


Figure 4.2a. The tertiary structure of the N-terminal domain of MUG4 predicted by SWISS-MODEL. a. the N-terminal domain structure of MUG4 predicted by SWISS-MODEL to the b. NMR solution structure of WRKY4 DNA-binding domain (PDB code 1wj2) and c. their structure overlay. The core structure of MUG4 possesses a three-stranded anti-parallel β-sheet and a zinc finger pocket. The zinc finger pocket of MUG4 is formed by two cysteines, one located at the end of strand 1 and the other located at the beginning of strand 2, a histidine and a cysteine located at the end of strand 3. The zinc finger pocket of WRKY4 is stabilized by a Zn atom (red ball) chelated by two cysteines, one located at the end of strand 2 and the other located at the beginning of strand 3, and two conserved histidines located at the end of strand 4. The atoms that form the zinc finger pocket are shown using the stick model and labelled with positions: C62, C67, H91 and H93 for MUG4.

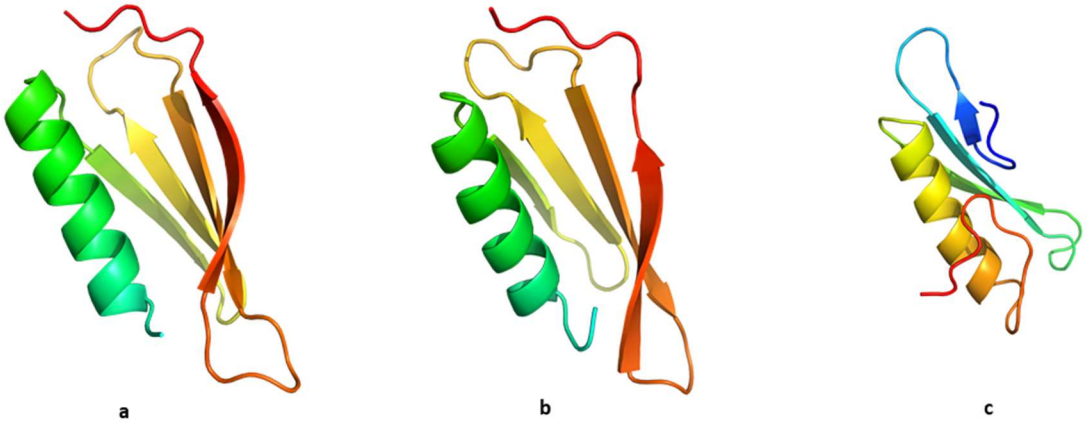


Figure 4.2b. The tertiary structure of the N-terminal domain of MUG4 predicted by RaptorX and trRosetta. a. the prediction of threading/fold recognition-based tool RaptorX; b. the prediction of *ab initio* structure modelling tool trRosetta; and c. the NMR structure of AP2/ERF domain (PDB 2gcc as the representative). Both two methods suggest MUG4 possesses an α-helix and following four-stranded anti-parallel β-sheet. For clarity, a variable half α-helix and a variable half β-strand insertion are not shown in the generated model.

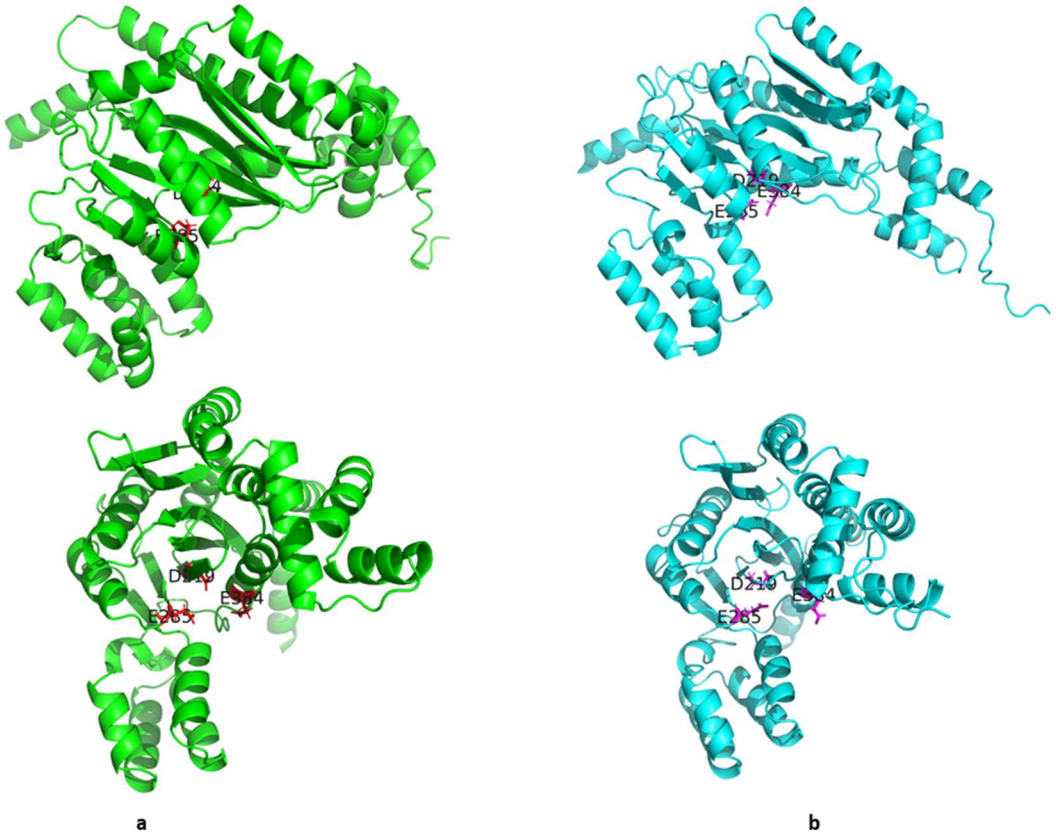


Figure 4.2c. The tertiary structure of the middle domain of MUG4 predicted by a. RaptorX and b. trRosetta. Both methods suggest a seven-stranded α+β core of the middle domain of MUG4. The corresponding residues that aligned to the catalytic triad DDE of the ancient transposase (D219, E285 and E384 for MUG4) are shown using the stick model. The up panel and down panel are two angels to see the triad DEE for MUG4.

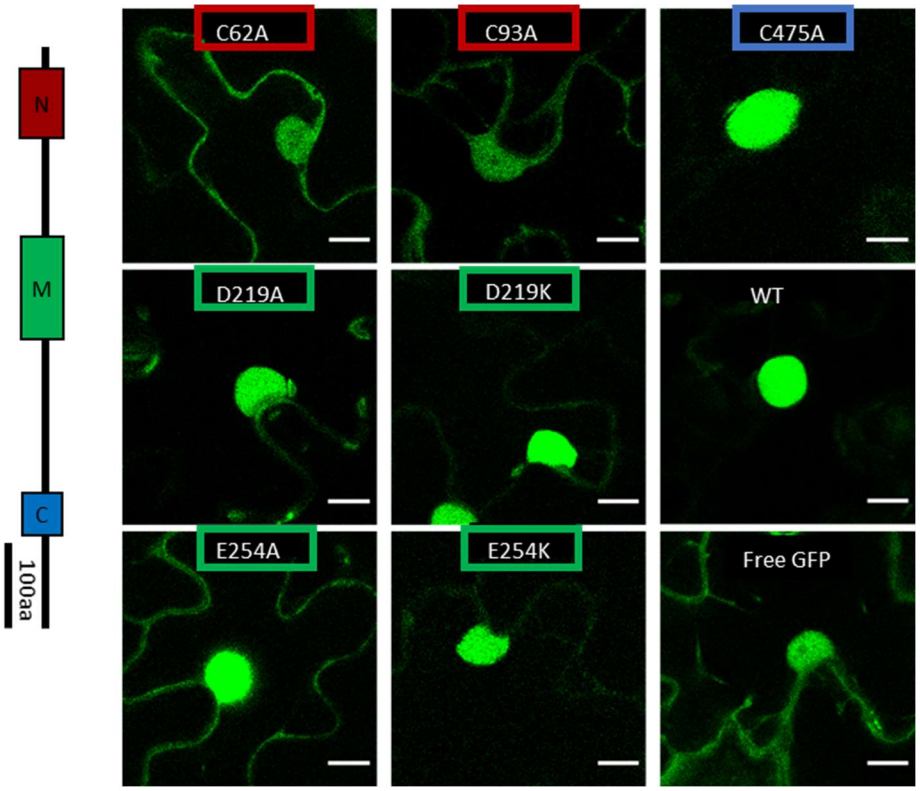


Figure 4.3. The subcellular localization of the site-directed mutagenesis of AtMUG4 in tobacco leaves. Confocal images of transgenic plants expressing 35sp:: MUG4-GFP, 35sp:: C62A-GFP, 35sp:: C93A-GFP, 35sp:: D219A-GFP, 35sp:: D219K-GFP, 35sp:: E254A-GFP, 35sp:: E254K-GFP, 35sp:: C475A-GFP. The wild-type MUG4 and C475A mutant protein show GFP signals are exclusively localized in the nucleus. While the C62A, C93A, D219A, D219K, E254A, E254K mutant protein, which are the mutagenesis in the N terminal as well as the middle domain of MUG4, alternate the GFP signals in both nuclear and cytosol. Free GFP works as the control of cell-wild distribution. Bar=10 μm.

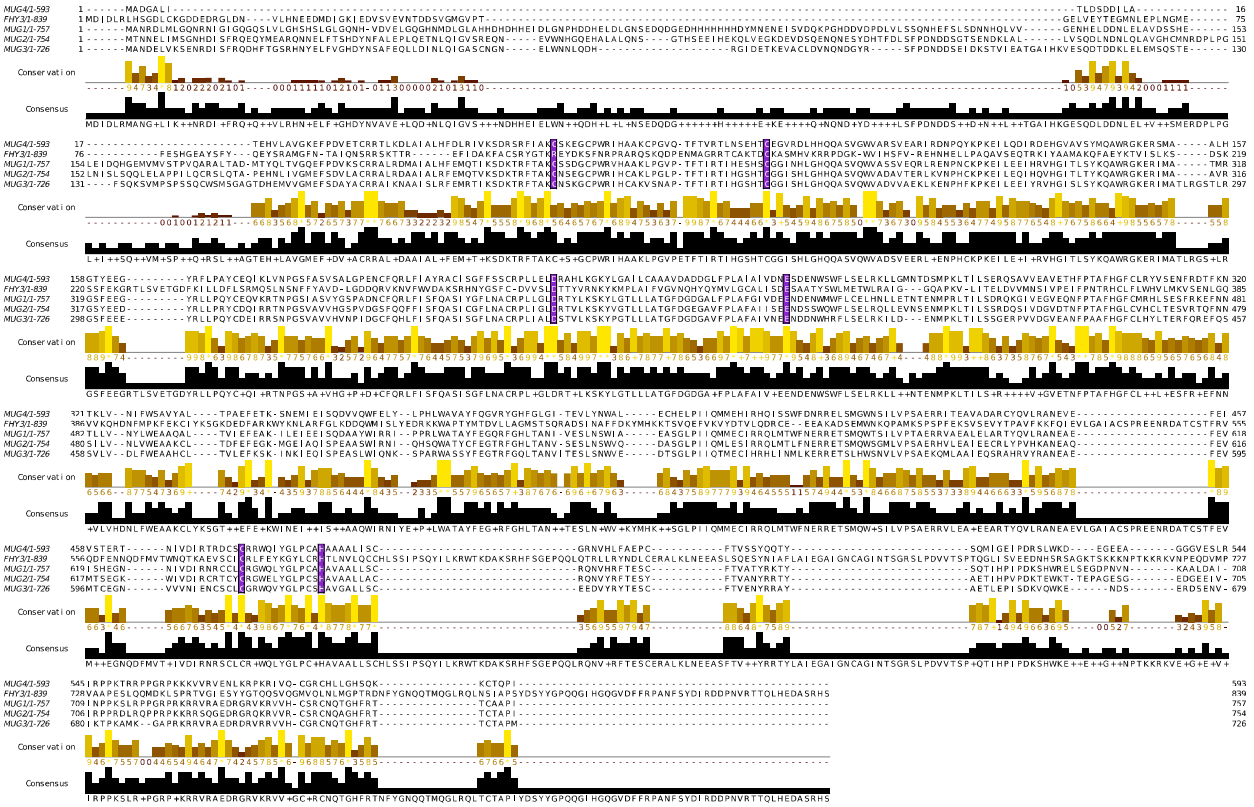


Figure 4.4a. Multiple sequence alignment of MUG1 to MUG4 and FHY3. The positions shading in purple mark the sites of MUG4 that were selected for mutagenesis studies in yeast two-hybrid assay.

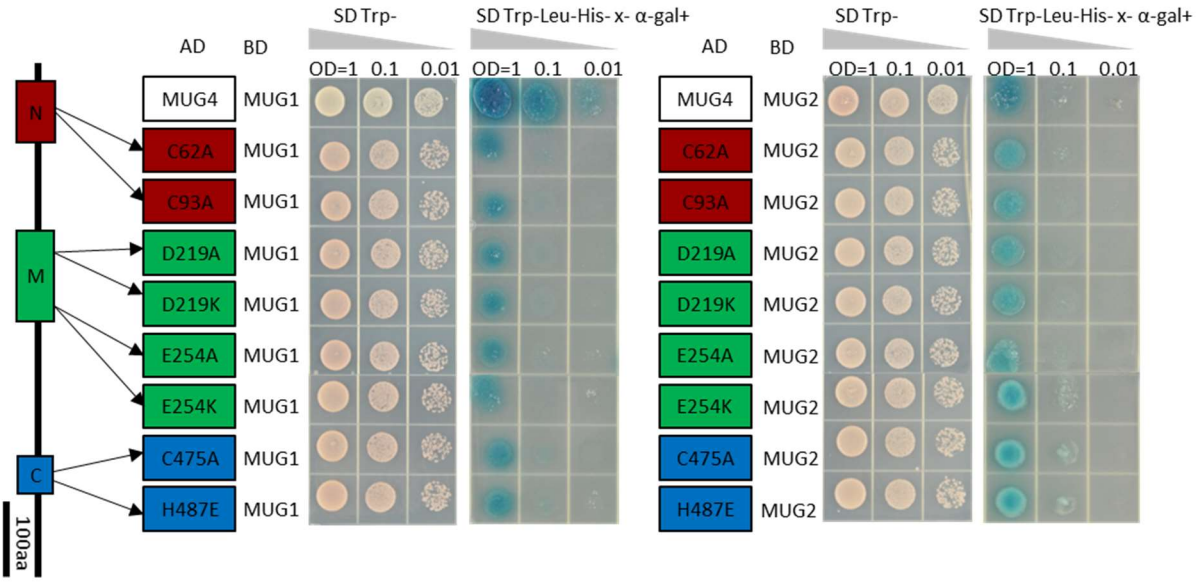


Figure 4.4b. All the three domains of AtMUG4 are essential for its dimerization with MUG1 and MUG2 in yeast cells. Yeast two-hybrid assay of site-directed mutagenesis of MUG4 and their interaction with MUG1 and MUG2. MUG1-GAL4BD co-transformed with MUG4-GAL4AD and MUG2-GAL4BD co-transformed with MUG4-GAl4AD yeast strains could grow on SD Trp-Leu-His and develop blue colour with the supplement of x- α-gal. While the mutation of C62A, C93A, D219A, D219K, E254A, E254K, C475A, H487E in MUG4 that fused with Gal4-AD that co-transformed with MUG1 and MUG2 grow limited on the selective medium and develop faint blue.

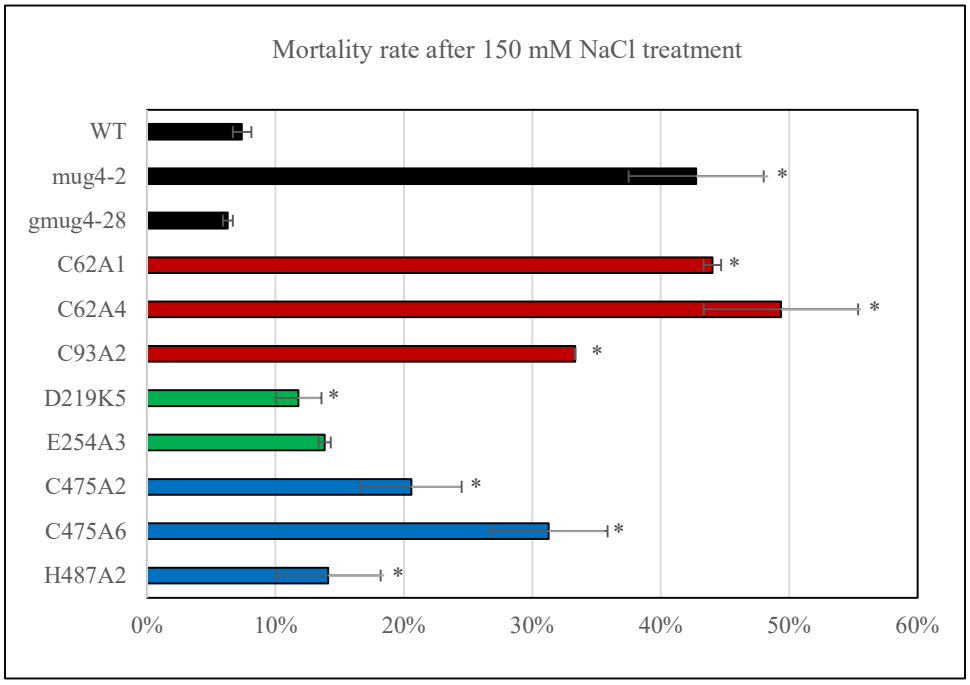


Figure 4.5. The mortality rate of several lines under the treatment of 150 mM NaCl. Various mutant forms of MUG4, as well as wild-type, were grown on the 1/2 MS media. After 4 days, the seedlings were transferred into 1/2 regular MS and salt medium. The molarity is indicated by cotyledon bleach after 3 days treatment. 3 replicates with consistent results were performed. * indicates $P<0.05$ compared to wild-type.

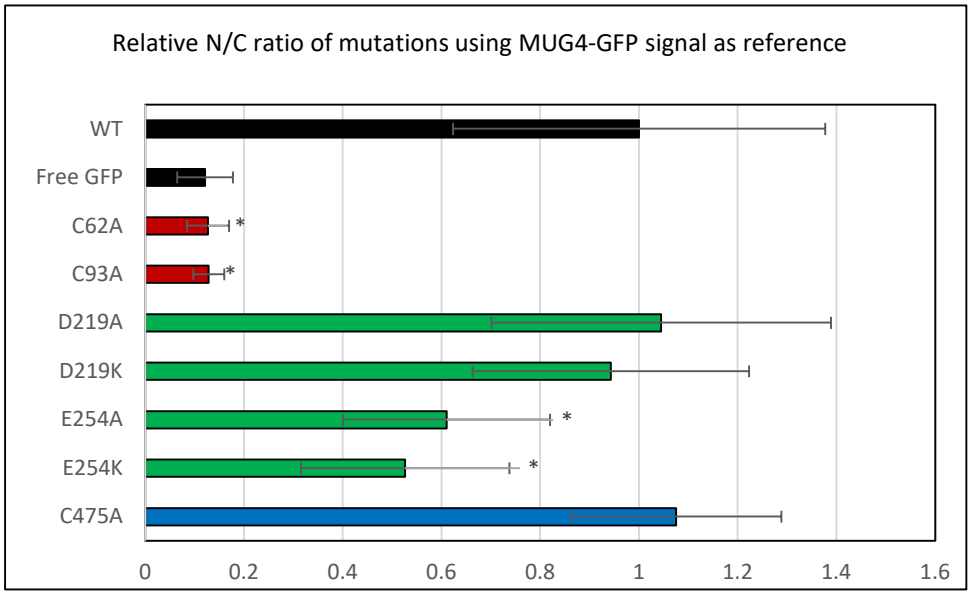


Figure 4.6. Bar plot displaying the quantitative data of the relative Nuclear/Cytosol GFP signal ratio in the mutation lines. The N-terminal domain, middle domain and C-terminal domain are represented by red, green, and blue bars, respectively. The basal level of the GFP signal is MUG4-GFP. Each line uses at least 11 cells for calculating, * indicates $P<0.001$.

Methods

Multiple sequence alignment

Multiple sequence alignment was performed by M-Coffee (Di Tommaso, Moretti et al. 2011) with default setting and were manually adjusted. The result was sent to Jalview for visualization.

Structure modelling of MUG4

Homology modelling-based tool: SWISS-MODEL (Waterhouse, Bertoni et al. 2018) was applied to predict the MUG4 structure. To provide sufficient modelling templates, I checked MUG4 closely related protein in the newly released AlphaFold Protein Structure Database (Jumper, Evans et al. 2021) and found Mutator transposase MURA protein (Gene ID: At5g15685) structure is recorded. This structure was then included as the supplemental modelling template to determine if it has a higher identity than other templates provided in the SWISS-MODEL template database. After comparing the identity of each template, the WRKY4 DNA-binding domain (PDB code 1wj2) was chosen as the modelling template. The remaining sequences other than N-terminal of MUG4 don not fit any template in the SWISS-MODEL.

The threading/fold recognition-based tool RaptorX (Xu, Mcpartlon et al. 2021), as well as the *ab initio* structure prediction-based tool trRosetta (Yang, Anishchenko et al. 2020), were applied with default settings to modelling MUG4 structure. Later, PyMOL (https://pymol.org/) was used for structural analysis and to create figures.

Site-directed Mutagenesis and Plasmid Construction

To generate various mutations for subcellular location detection and Y2H assay, the plasmid pGEM-MUG4 build in chapter 2 was used as the template for site-directed mutagenesis using the AQUA (advanced quick assembly) cloning approach (Beyer, Gonschorek et al. 2015). The primers used for building site-directed mutagenesis entry plasmids were as follows: C62A_FP, 5'-CAA AGC CTC TAA AGA AGG TTG TCC TTG-3'; C62A_RP, 5'-CAA CCT TCT TTA GAG GCT TTG GCT ATG-3'; C93A_FP, 5'-CTG CTG AAG GTG TCC GTG ATC TC-3'; C93A_RP, 5'-ATC ACG GAC ACC TTC AGC AGT ATG-3'; D219A_FP, 5'-GGC TAG AGC ACA CCT TAA AGG GAA G-3'; D219A_RP 5'-CTT TAA GGT GTG CTC TAG CCA ATT CAA G-3'; D219K_FP, 5'-CGG CCT CTT CTT GAA TTG AAG AGA GC-3'; D219K_RP, 5'-CTT CAA TTC AAG AAG AGG CCG ACA AG-3'; E254A_FP, 5'-GCT ATT GTC GAT AAT GCA AGC GAT G-3'; E254A_RP, 5'-CTT GCA TTA TCG ACA ATA GCG ATA GC-3'; E254K_FP, 5'-CGC TAT TGT CGA TAA TAA AAG CGA TG-3'; E254K_RP, 5'-GCT TTT ATT ATC GAC AAT AGC GAT AGC-3'; C475A_FP, 5'-CAG CCC GTC GTT GGC AGA TTT AC-3'; C475A_RP, 5'-AAT CTG CCA ACG ACG GGC TGA G-3'; H487E_FP, 5'-GCT TCA GCA CAC GGC AAT CCG TA; H487E_RP 5'-GGA TTG CCG TGT GCT GAA GCG G. The above-mentioned mutations were then cloned into pEarleyGate103 based backbone resulting plasmids using Gateway™ LR Cloning system.

To generate site-directed mutagenesis of MUG4 transgenic lines in Arabidopsis, the *MUG4* native promoter-driven mutation constructs were built for transformation. The entry vector pGEM-*mug4pro:MUG4* build in chapter 2 was used as the entry vector and introduced into the

binary vector pMDC107 through LR reaction. All constructs were validated by diagnostic restriction digest and sequencing.

Yeast two-hybrid assay

The above-mentioned pEarleyGate103 based backbone resulting plasmids were transformed into yeast strains. The Y2H assays were performed essentially as described in chapter 2, seeing Methods, Yeast two-hybrid assay.

Plant Transformation

The final pEarleyGate103 based backbone mutation vectors were transfected into *Nicotiana benthamiana* tobacco leaves for the subcellular localization detection.

The pMDC107 based mutation binary vectors were electroporated into the Agrobacterium tumefaciens strain GV3101 and transformed into the *mug4-2* mutant via the floral dip method. Ten kanamycin-resistant T1 transgenic lines for each transgene were selected and continued for selecting stable homozygous plants. Phenotypic analyses were performed in the T3 stable generation.

Plant Materials and Growth Conditions

Plated seeds were stratified in the dark for 3 days at 4℃ and kept on plates for 2 weeks in a growth chamber (Conviron model E15) at 22°C under a 16 h light/8 h dark photoperiod, ~100 mmol quanta/m2/s light intensity, 60% relative humidity. Seedlings were transplanted to soil

with a composition of PRO-MIX (Premier Tech Horticulture, Quebec, Canada): vermiculite: perlite of 2:1:1 in 2 ½ inch square pots and returned to the growth chamber.

Confocal Microscopy Observation

Subcellular images were captured using the Leica SP8 confocal system, with a × 40/1.2 water immersion objective. Excitation/emission wavelengths were 488/505–530 nm for GFP. At least ten images containing one or more protoplasts were taken for each of the plasmid construct combinations.

N/C GFP signal quantification

LASX(Leica) was used to quantify the maximal fluorescent intensity of the nucleus and cytosol. The ratio of nucleus to the cytosol (N/C) of different MUG4 mutants was quantified relative to the WT MUG4 N/C value which was set as 1. At least 20 cells were used for the fluorescence quantification of each MUG4 mutant.

Reference

Ali, S. A., M. J. Peffers, M. J. Ormseth, I. Jurisica and M. Kapoor (2021). "The non-coding RNA interactome in joint health and disease." Nature Reviews Rheumatology: 1-14.

Allen, S. E. and M. Nowacki (2017). "Necessity is the mother of invention: ciliates, transposons, and transgenerational inheritance." Trends in Genetics **33**(3): 197-207.

Arkhipova, I. and M. Meselson (2000). "Transposable elements in sexual and ancient asexual taxa." Proceedings of the National Academy of Sciences **97**(26): 14473-14477.

Babakhani, S. and M. Oloomi (2018). "Transposons: the agents of antibiotic resistance in bacteria." Journal of basic microbiology **58**(11): 905-917.

Babu, M. M., L. M. Iyer, S. Balaji and L. Aravind (2006). "The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons." Nucleic Acids Research **34**(22): 6505-6520.

Benito, M. I. and V. Walbot (1997). "Characterization of the maize Mutator transposable element MURA transposase as a DNA-binding protein." Molecular and Cellular Biology **17**(9): 5165-5175.

Benoit, M., H.-G. Drost, M. Catoni, Q. Gouil, S. Lopez-Gomollon, D. Baulcombe and J. Paszkowski (2019). "Environmental and epigenetic regulation of Rider retrotransposons in tomato." PLoS genetics **15**(9): e1008370.

Beyer, H. M., P. Gonschorek, S. L. Samodelov, M. Meier, W. Weber and M. D. Zurbriggen (2015). "AQUA Cloning: A Versatile and Simple Enzyme-Free Cloning Approach." PloS one **10**(9): e0137652-e0137652.

Bhasin, A., I. Y. Goryshin, M. Steiniger-White, D. York and W. S. Reznikoff (2000). "Characterization of a Tn5 pre-cleavage synaptic complex." J Mol Biol **302**(1): 49-63.

Bingham, P. M., M. G. Kidwell and G. M. Rubin (1982). "The molecular basis of PM hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family." Cell **29**(3): 995-1004.

Boeke, J. D., D. J. Garfinkel, C. A. Styles and G. R. Fink (1985). "Ty elements transpose through an RNA intermediate." Cell **40**(3): 491-500.

Boissinot, S., A. Entezam and A. V. Furano (2001). "Selection against deleterious LINE-1-containing loci in the human lineage." Molecular biology and evolution **18**(6): 926-935.

Borgognone, A., R. Castanera, E. Muguerza, A. G. Pisabarro and L. Ramírez (2017). "Somatic transposition and meiotically driven elimination of an active helitron family in Pleurotus ostreatus." Dna Research **24**(2): 103-115.

Bourque, G., K. H. Burns, M. Gehring, V. Gorbunova, A. Seluanov, M. Hammell, M. Imbeault, Z. Izsvák, H. L. Levin and T. S. Macfarlan (2018). "Ten things you should know about transposable elements." Genome biology **19**(1): 1-12.

Buchmann, J. P., A. Löytynoja, T. Wicker and A. H. Schulman (2014). "Analysis of CACTA transposases reveals intron loss as major factor influencing their exon/intron structure in monocotyledonous and eudicotyledonous hosts." Mobile DNA **5**(1): 1-15.

Bundock, P. and P. Hooykaas (2005). "An Arabidopsis hAT-like transposase is essential for plant development." Nature **436**(7048): 282-284.

Bureau, T. E. and S. R. Wessler (1992). "Tourist: a large family of small inverted repeat elements frequently associated with maize genes." The Plant Cell **4**(10): 1283-1294.

Butelli, E., C. Licciardello, Y. Zhang, J. Liu, S. Mackay, P. Bailey, G. Reforgiato-Recupero and C. Martin (2012). "Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges." The Plant Cell **24**(3): 1242-1255.

Canadell, J. G. and R. B. Jackson (2021). "Ecosystem collapse and climate change." Ecological Studies **241**.

Cheeseman, J. (2016). Food security in the face of salinity, drought, climate change, and population growth. Halophytes for food security in dry lands, Elsevier**:** 111-123.

Chen, S. and X. Li (2008). "Molecular characterization of the first intact Transib transposon from Helicoverpa zea." Gene **408**(1-2): 51-63.

Cho, J. (2018). "Transposon-derived non-coding RNAs and their function in plants." Frontiers in plant science **9**: 600.

Cho, J., M. Benoit, M. Catoni, H.-G. Drost, A. Brestovitsky, M. Oosterbeek and J. Paszkowski (2019). "Sensitive detection of pre-integration intermediates of long terminal repeat retrotransposons in crop plants." Nature plants **5**(1): 26-33.

Chuong, E. B., N. C. Elde and C. Feschotte (2017). "Regulatory activities of transposable elements: from conflicts to benefits." Nature Reviews Genetics **18**(2): 71.

Claeys, H., S. Van Landeghem, M. Dubois, K. Maleux and D. Inzé (2014). "What is stress? Dose-response effects in commonly used in vitro stress assays." Plant physiology **165**(2): 519-527.

Cobb, B. S., S. Morales-Alcelay, G. Kleiger, K. E. Brown, A. G. Fisher and S. T. Smale (2000). "Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding." Genes Dev **14**(17): 2146-2160.

Cordaux, R. and M. A. Batzer (2009). "The impact of retrotransposons on human genome evolution." Nature Reviews Genetics **10**(10): 691-703.

Cowan, R. K., D. R. Hoen, D. J. Schoen and T. E. Bureau (2005). "MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms." Molecular Biology and Evolution **22**(10): 2084-2089.

Craig, N. L. (1995). "Unity in transposition reactions." Science **270**(5234): 253-253.

Creasey, K. M., J. Zhai, F. Borges, F. Van Ex, M. Regulski, B. C. Meyers and R. A. Martienssen (2014). "miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis." Nature **508**(7496): 411-415.

de Luxán-Hernández, C., J. Lohmann, W. Hellmeyer, S. Seanpong, K. Wöltje, Z. Magyar, A. Pettkó-Szandtner, T. Pelissier, G. De Jaeger and S. Hoth (2020). "PP7L is essential for MAIL1-mediated transposable element silencing and primary root growth." The Plant Journal **102**(4): 703-717.

Deininger, P. (2011). "Alu elements: know the SINEs." Genome biology **12**(12): 1-12.

Demidchik, V. and F. J. M. Maathuis (2007). "Physiological roles of nonselective cation channels in plants: from salt stress to signalling and development." New Phytologist **175**(3): 387-404.

Di Tommaso, P., S. Moretti, I. Xenarios, M. Orobitg, A. Montanyola, J. M. Chang, J. F. Taly and C. Notredame (2011). "T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension." Nucleic Acids Res **39**(Web Server issue): W13-17.

Dodonova, S. O., S. Prinz, V. Bilanchone, S. Sandmeyer and J. A. G. Briggs (2019). "Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses." Proc Natl Acad Sci U S A **116**(20): 10048-10057.

Domínguez, M., E. Dugas, M. Benchouaia, B. Leduque, J. M. Jiménez-Gómez, V. Colot and L. Quadrana (2020). "The impact of transposable elements on tomato diversity." Nature communications **11**(1): 1-11.

Doolittle, W. F. and C. Sapienza (1980). "Selfish genes, the phenotype paradigm and genome evolution." Nature **284**(5757): 601-603.

Duan, C.-G., X. Wang, S. Xie, L. Pan, D. Miki, K. Tang, C.-C. Hsu, M. Lei, Y. Zhong and Y.-J. Hou (2017). "A pair of transposon-derived proteins function in a histone acetyltransferase complex for active DNA demethylation." Cell research **27**(2): 226-240.

Ducrocq, S., D. Madur, J.-B. Veyrieras, L. Camus-Kulandaivelu, M. Kloiber-Maitz, T. Presterl, M. Ouzunova, D. Manicacci and A. Charcosset (2008). "Key impact of Vgt1 on flowering time adaptation in maize: evidence from association mapping and ecogeographical information." Genetics **178**(4): 2433-2437.

Dupeyron, M., T. Baril, C. Bass and A. Hayward (2020). "Phylogenetic analysis of the Tc1/mariner superfamily reveals the unexplored diversity of pogo-like elements." Mobile DNA **11**(1): 1-14.

Eickbush, T. H. and D. G. Eickbush (2015). "Integration, regulation, and long-term stability of R2 retrotransposons." Microbiology spectrum **3**(2): 3.2. 02.

Engels, W. R. and C. R. Preston (1981). "Identifying P factors in Drosophila by means of chromosome breakage hotspots." Cell **26**(3): 421-428.

Evgen'ev, M. and I. Arkhipova (2005). "Penelope-like elements–a new class of retroelements: distribution, function and possible evolutionary significance." Cytogenetic and genome research **110**(1-4): 510-521.

Faulkner, G. J. and V. Billon (2018). "L1 retrotransposition in the soma: a field jumping ahead." Mobile DNA **9**(1): 1-18.

Feschotte, C. (2004). "Merlin, a new superfamily of DNA transposons identified in diverse animal genomes and related to bacterial IS 1016 insertion sequences." Molecular Biology and Evolution **21**(9): 1769-1780.

Feschotte, C. (2008). "The contribution of transposable elements to the evolution of regulatory networks." Nature reviews. Genetics **9**(5): 397-405.

Feschotte, C., N. Jiang and S. R. Wessler (2002). "Plant transposable elements: where genetics meets genomics." Nat Rev Genet **3**(5): 329-341.

Feschotte, C. and E. J. Pritham (2007). "DNA transposons and the evolution of eukaryotic genomes." Annu. Rev. Genet. **41**: 331-368.

Filée, J., S. Farhat, D. Higuet, L. Teysset, D. Marie, C. Thomas-Bulle, S. Hourdez, D. Jollivet and E. Bonnivard (2021). "Comparative genomic and transcriptomic analyses of transposable elements in polychaetous annelids highlight LTR retrotransposon diversity and evolution." Mobile DNA **12**(1): 1-24.

Flavell, A. J., S. R. Pearce and A. Kumar (1994). "Plant transposable elements and the genome." Current opinion in genetics & development **4**(6): 838-844.

Friesen, N., A. Brandes and J. S. Heslop-Harrison (2001). "Diversity, Origin, and Distribution of Retrotransposons (gypsy and copia) in Conifers." Molecular Biology and Evolution **18**(7): 1176-1188.

Fultz, D., S. G. Choudury and R. K. Slotkin (2015). "Silencing of active transposable elements in plants." Current opinion in plant biology **27**: 67-76.

Gagné, F. (2014). Chapter 2 - Tissue Preparation and Subcellular Fractionation Techniques. Biochemical Ecotoxicology. F. Gagné. Oxford, Academic Press**:** 21-31.

Gao, X., E. R. Havecker, P. V. Baranov, J. F. Atkins and D. F. Voytas (2003). "Translational recoding signals between gag and pol in diverse LTR retrotransposons." RNA (New York, N.Y.) **9**(12): 1422-1430.

Gookin, T. E. and S. M. Assmann (2014). "Significant reduction of BiFC non-specific assembly facilitates in planta assessment of heterotrimeric G-protein interactors." Plant Journal **80**(3): 553-567.

Grandbastien, M.-A. (1998). "Activation of plant retrotransposons under stress conditions." Trends in plant science **3**(5): 181-187.

Gray, L. T., K. K. Fong, T. Pavelitz and A. M. Weiner (2012). "Tethering of the Conserved piggyBac Transposase Fusion Protein CSB-PGBD3 to Chromosomal AP-1 Proteins Regulates Expression of Nearby Genes in Humans." Plos Genetics **8**(9).

Greenblatt, I. M. and R. A. Brink (1963). "Transpositions of Modulator in maize into divided and undivided chromosome segments." Nature **197**(4865): 412-413.

Guan, Q. M., J. M. Wu, X. L. Yue, Y. Y. Zhang and J. H. Zhu (2013). "A Nuclear Calcium-Sensing Pathway Is Critical for Gene Regulation and Salt Stress Tolerance in Arabidopsis." Plos Genetics **9**(8).

Haak, D. C., T. Fukao, R. Grene, Z. Hua, R. Ivanov, G. Perrella and S. Li (2017). "Multilevel regulation of abiotic stress responses in plants." Frontiers in plant science **8**: 1564.

Harrison, J. J. E. K., S. Tuske, K. Das, F. X. Ruiz, J. D. Bauman, P. L. Boyer, J. J. DeStefano, S. H. Hughes and E. Arnold (2021). "Crystal Structure of a Retroviral Polyprotein: Prototype Foamy Virus Protease-Reverse Transcriptase (PR-RT)." Viruses **13**(8): 1495.

Hartwig, B., G. V. James, K. Konrad, K. Schneeberger and F. Turck (2012). "Fast isogenic mapping-by-sequencing of ethyl methanesulfonate-induced mutant bulks." Plant physiology **160**(2): 591-600.

Hasan, N., S. Choudhary, N. Naaz, N. Sharma and R. A. Laskar (2021). "Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes." Journal of Genetic Engineering and Biotechnology **19**(1): 1-26.

Hickey, D. A. (1982). "Selfish DNA: a sexually-transmitted nuclear parasite." Genetics **101**(3-4): 519-531.

Hoegh-Guldberg, O., D. Jacob, M. Bindi, S. Brown, I. Camilloni, A. Diedhiou, R. Djalante, K. Ebi, F. Engelbrecht and J. Guiot (2018). "Impacts of 1.5 C global warming on natural and human systems." Global warming of 1.5 C. An IPCC Special Report.

Hoen, D. R. and T. E. Bureau (2012). Transposable element exaptation in plants. Plant transposable elements, Springer**:** 219-251.

Hoen, D. R. and T. E. Bureau (2015). "Discovery of Novel Genes Derived from Transposable Elements Using Integrative Genomic Analysis." Molecular Biology and Evolution **32**(6): 1487-1506.

Hollister, J. D. and B. S. Gaut (2009). "Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression." Genome research **19**(8): 1419-1428.

Hudson, M., C. Ringli, M. T. Boylan and P. H. Quail (1999). "The FAR1 locus encodes a novel nuclear protein specific to phytochrome A signaling." Genes Dev **13**(15): 2017-2027.

Hudson, M. E., D. R. Lisch and P. H. Quail (2003). "The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway." Plant Journal **34**(4): 453-471.

Ikeda, Y., T. Pélissier, P. Bourguet, C. Becker, M.-N. Pouch-Pélissier, R. Pogorelcnik, M. Weingartner, D. Weigel, J.-M. Deragon and O. Mathieu (2017). "Arabidopsis proteins with a transposon-related domain act in gene silencing." Nature Communications **8**(1): 1-10.

Isayenkov, S. V., S. A. Dabravolski, T. Pan and S. Shabala (2020). "Phylogenetic Diversity and Physiological Roles of Plant Monovalent Cation/H(+) Antiporters." Front Plant Sci **11**: 573564.

Isayenkov, S. V. and F. J. M. Maathuis (2019). "Plant Salinity Stress: Many Unanswered Questions Remain." Front Plant Sci **10**: 80.

Ito, H., H. Gaubert, E. Bucher, M. Mirouze, I. Vaillant and J. Paszkowski (2011). "An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress." Nature **472**(7341): 115-119.

Jiao, Y., P. Peluso, J. Shi, T. Liang, M. C. Stitzer, B. Wang, M. S. Campbell, J. C. Stein, X. Wei, C. S. Chin, K. Guill, M. Regulski, S. Kumari, A. Olson, J. Gent, K. L. Schneider, T. K. Wolfgruber, M. R. May, N. M. Springer, E. Antoniou, W. R. McCombie, G. G. Presting, M. McMullen, J. Ross-Ibarra, R. K. Dawe, A. Hastie, D. R. Rank and D. Ware (2017). "Improved maize reference genome with single-molecule technologies." Nature **546**(7659): 524-527.

Johns, M. A., J. Mottinger and M. Freeling (1985). "A low copy number, copia-like transposon in maize." The EMBO journal **4**(5): 1093-1101.

Joly-Lopez, Z. and T. E. Bureau (2018). "Exaptation of transposable element coding sequences." Current Opinion in Genetics & Development **49**: 34-42.

Joly-Lopez, Z., E. Forczek, D. R. Hoen, N. Juretic and T. E. Bureau (2012). "A Gene Family Derived from Transposable Elements during Early Angiosperm Evolution Has Reproductive Fitness Benefits in Arabidopsis thaliana." Plos Genetics **8**(9).

Joly-Lopez, Z., E. Forczek, D. R. Hoen, N. Juretic and T. E. Bureau (2012). "A gene family derived from transposable elements during early angiosperm evolution has reproductive fitness benefits in Arabidopsis thaliana."

Joly-Lopez, Z., E. Forczek, E. Vello, D. R. Hoen, A. Tomita and T. E. Bureau (2017). "Abiotic Stress Phenotypes Are Associated with Conserved Genes Derived from Transposable Elements." Front Plant Sci **8**: 2027.

Joly-Lopez, Z., D. R. Hoen, M. Blanchette and T. E. Bureau (2016). "Phylogenetic and Genomic Analyses Resolve the Origin of Important Plant Genes Derived from Transposable Elements." Mol Biol Evol **33**(8): 1937-1956.

Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek and A. Potapenko (2021). "Highly accurate protein structure prediction with AlphaFold." Nature: 1-11.

Kabelitz, T., C. Kappel, K. Henneberger, E. Benke, C. Nöh and I. Bäurle (2014). "eQTL mapping of transposon silencing reveals a position-dependent stable escape from epigenetic silencing and transposition of AtMu1 in the Arabidopsis lineage." The Plant Cell **26**(8): 3261-3271.

Kagale, S., C. Koh, J. Nixon, V. Bollina, W. E. Clarke, R. Tuteja, C. Spillane, S. J. Robinson, M. G. Links, C. Clarke, E. E. Higgins, T. Huebert, A. G. Sharpe and I. A. P. Parkin (2014). "The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure." Nature Communications **5**(1): 3706.

Kamigaki, A., K. Nito, K. Hikino, S. Goto-Yamada, M. Nishimura, T. Nakagawa and S. Mano (2016). "Gateway vectors for simultaneous detection of multiple protein− protein interactions in plant cells using bimolecular fluorescence complementation." PloS one **11**(8): e0160717.

Kapitonov, V. V. and J. Jurka (2003). "A novel class of SINE elements derived from 5S rRNA." Molecular biology and evolution **20**(5): 694-702.

Kaplan, B., T. Sherman and H. Fromm (2007). "Cyclic nucleotide-gated channels in plants." FEBS letters **581**(12): 2237-2246.

Kato, M., A. Miura, J. Bender, S. E. Jacobsen and T. Kakutani (2003). "Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis." Current Biology **13**(5): 421-426.

Kempken, F. and F. Windhofer (2001). "The hAT family: a versatile transposon group common to plants, fungi, animals, and man." Chromosoma **110**(1): 1-9.

Kidwell, M. G., J. F. Kidwell and J. A. Sved (1977). "Hybrid dysgenesis in Drosophila melanogaster: a syndrome of aberrant traits including mutation, sterility and male recombination." Genetics **86**(4): 813-833.

Kidwell, M. G. and D. R. Lisch (2001). "Perspective: transposable elements, parasitic DNA, and genome evolution." Evolution **55**(1): 1-24.

Kim, D., J. M. Paggi, C. Park, C. Bennett and S. L. Salzberg (2019). "Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype." Nature Biotechnology **37**(8): 907-915.

Kim, E. Y., L. Wang, Z. Lei, H. Li, W. Fan and J. Cho (2021). "Ribosome stalling and SGS3 phase separation prime the epigenetic silencing of transposons." Nature Plants **7**(3): 303-309.

Klein, S. J. and R. J. O'Neill (2018). "Transposable elements: genome innovation, chromosome diversity, and centromere conflict." Chromosome Research **26**(1): 5-23.

Knip, M., S. de Pater and P. J. Hooykaas (2012). "The SLEEPER genes: a transposase-derived angiosperm-specific gene family." BMC plant biology **12**(1): 1-15.

Knip, M., S. Hiemstra, A. Sietsma, M. Castelein, S. de Pater and P. Hooykaas (2013). "DAYSLEEPER: a nuclear and vesicular-localized protein that is expressed in proliferating tissues." BMC Plant Biol **13**: 211.

Kobayashi, S., N. Goto-Yamamoto and H. Hirochika (2004). "Retrotransposon-induced mutations in grape skin color." Science **304**(5673): 982-982.

Kojima, K. K. and J. Jurka (2011). "Crypton transposons: identification of new diverse families and ancient domestication events." Mobile DNA **2**(1): 1-17.

Kosek, D., A. B. Hickman, R. Ghirlando, S. He and F. Dyda (2021). "Structures of ISC th4 transpososomes reveal the role of asymmetry in copy-out/paste-in DNA transposition." The EMBO journal **40**(1): e105666.

Kramerov, D. and N. Vassetzky (2011). "Origin and evolution of SINEs in eukaryotic genomes." Heredity **107**(6): 487-495.

Krupovic, M., D. H. Bamford and E. V. Koonin (2014). "Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses." Biology direct **9**(1): 1-7.

Lei, M., H. Zhang, R. Julian, K. Tang, S. Xie and J.-K. Zhu (2015). "Regulatory link between DNA methylation and active demethylation in Arabidopsis." Proceedings of the National Academy of Sciences **112**(11): 3553-3557.

Lemaux, P. G. (2008). "Genetically Engineered Plants and Foods: A Scientist's Analysis of the Issues (Part I)." Annu Rev Plant Biol **59**: 771-812.

Leung, J., M. Bouvier-Durand, P. C. Morris, D. Guerrier, F. Chefdor and J. Giraudat (1994). "Arabidopsis ABA response gene ABI1: features of a calcium-modulated protein phosphatase." Science **264**(5164): 1448-1452.

Li, B., S. Jiang, X. Yu, C. Cheng, S. Chen, Y. Cheng, J. S. Yuan, D. Jiang, P. He and L. Shan (2015). "Phosphorylation of trihelix transcriptional repressor ASR3 by MAP KINASE4 negatively regulates Arabidopsis immunity." Plant Cell **27**(3): 839-856.

Liao, Y., G. K. Smyth and W. Shi (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." Bioinformatics **30**(7): 923-930.

Lin, Q., Y. Ohashi, M. Kato, T. Tsuge, H. Gu, L. J. Qu and T. Aoyama (2015). "GLABRA2 Directly Suppresses Basic Helix-Loop-Helix Transcription Factor Genes with Diverse Functions in Root Hair Development." Plant Cell **27**(10): 2894-2906.

Lin, R., L. Ding, C. Casola, D. R. Ripoll, C. Feschotte and H. Wang (2007). "Transposase-derived transcription factors regulate light signaling in Arabidopsis." Science **318**(5854): 1302-1305.

Lin, R., Y. Teng, H. J. Park, L. Ding, C. Black, P. Fang and H. Wang (2008). "Discrete and essential roles of the multiple domains of Arabidopsis FHY3 in mediating phytochrome A signal transduction." Plant Physiol **148**(2): 981-992.

Lisch, D. (2009). "Epigenetic regulation of transposable elements in plants." Annual review of plant biology **60**: 43-66.

Liu, K. and S. R. Wessler (2017). "Transposition of Mutator-like transposable elements (MULEs) resembles hAT and Transib elements and V(D)J recombination." Nucleic Acids Research **45**(11): 6644-6655.

Liu, Y., M. Ma, G. Li, L. Yuan, Y. Xie, H. Wei, X. Ma, Q. Li, P. F. Devlin and X. Xu (2020). "Transcription factors FHY3 and FAR1 regulate light-induced CIRCADIAN CLOCK ASSOCIATED1 gene expression in Arabidopsis." The Plant Cell **32**(5): 1464-1478.

Liu, Z., C. An, Y. Zhao, Y. Xiao, L. Bao, C. Gong and Y. Gao (2021). "Genome-Wide Identification and Characterization of the CsFHY3/FAR1 Gene Family and Expression Analysis under Biotic and Abiotic Stresses in Tea Plants (Camellia sinensis)." Plants **10**(3): 570.

Lu, Q., X. Tang, G. Tian, F. Wang, K. Liu, V. Nguyen, S. E. Kohalmi, W. A. Keller, E. W. Tsang, J. J. Harada, S. J. Rothstein and Y. Cui (2010). "Arabidopsis homolog of the yeast TREX-2 mRNA export complex: components and anchoring nucleoporin." Plant J **61**(2): 259-270.

Ma, L. and G. Li (2018). "FAR1-RELATED SEQUENCE (FRS) and FRS-RELATED FACTOR (FRF) Family Proteins in Arabidopsis Growth and Development." Front Plant Sci **9**: 692.

Majumdar, S. and D. C. Rio (2015). "P transposable elements in Drosophila and other eukaryotic organisms." Microbiology spectrum **3**(2): 3.2. 01.

Makarevitch, I., A. J. Waters, P. T. West, M. Stitzer, C. N. Hirsch, J. Ross-Ibarra and N. M. Springer (2015). "Correction: Transposable Elements Contribute to Activation of Maize Genes in Response to Abiotic Stress." PLoS Genet **11**(10): e1005566.

Makarova, K. S., L. Aravind and E. V. Koonin (2002). "SWIM, a novel Zn-chelating domain present in bacteria, archaea and eukaryotes." Trends Biochem Sci **27**(8): 384-386.

Malicki, M., T. Spaller, T. Winckler and C. Hammann (2020). "DIRS retrotransposons amplify via linear, single-stranded cDNA intermediates." Nucleic acids research **48**(8): 4230-4243.

Malik, H. S. and T. H. Eickbush (1998). "The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs." Mol Biol Evol **15**(9): 1123-1134.

McClintock, B. (1950). "The Origin and Behavior of Mutable Loci in Maize." Proceedings of the National Academy of Sciences of the United States of America **36**(6): 344-355.

McGrath, K. C., B. Dombrecht, J. M. Manners, P. M. Schenk, C. I. Edgar, D. J. Maclean, W.-R. d. Scheible, M. K. Udvardi and K. Kazan (2005). "Repressor- and Activator-Type Ethylene Response Factors Functioning in Jasmonate Signaling and Disease Resistance Identified via a Genome-Wide Screen of Arabidopsis Transcription Factor Gene Expression." Plant Physiology **139**(2): 949-959.

Michel, K. and P. W. Atkinson (2003). "Nuclear localization of the Hermes transposase depends on basic amino acid residues at the N-terminus of the protein." J Cell Biochem **89**(4): 778-790.

Mills, R. E., E. A. Bennett, R. C. Iskow and S. E. Devine (2007). "Which transposable elements are active in the human genome?" Trends in Genetics **23**(4): 183-191.

Mirouze, M., J. Reinders, E. Bucher, T. Nishimura, K. Schneeberger, S. Ossowski, J. Cao, D. Weigel, J. Paszkowski and O. Mathieu (2009). "Selective epigenetic control of retrotransposition in Arabidopsis." Nature **461**(7262): 427-430.

Miura, A., S. Yonebayashi, K. Watanabe, T. Toyama, H. Shimada and T. Kakutani (2001). "Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis." Nature **411**(6834): 212-214.

Morgan, T. H. (1910). "Sex Limited Inheritance in Drosophila." Science **32**(812): 120-122.

Movahed, N., C. Patarroyo, J. Sun, H. Vali, J.-F. Laliberté and H. Zheng (2017). "Cylindrical Inclusion Protein of Turnip Mosaic Virus Serves as a Docking Point for the Intercellular Movement of Viral Replication Vesicles." Plant Physiology **175**(4): 1732-1744.

Muehlbauer, G. J., B. S. Bhau, N. H. Syed, S. Heinen, S. Cho, D. Marshall, S. Pateyron, N. Buisine, B. Chalhoub and A. J. Flavell (2006). "A hAT superfamily transposase recruited by the cereal grass genome." Molecular Genetics and Genomics **275**(6): 553-563.

Munns, R. and M. Tester (2008). "Mechanisms of salinity tolerance." Annu. Rev. Plant Biol. **59**: 651-681.

Negi, P., A. N. Rai and P. Suprasanna (2016). "Moving through the Stressed Genome: Emerging Regulatory Roles for Transposons in Plant Stress Response." Frontiers in Plant Science **7**.

Negrao, S., S. M. Schmockel and M. Tester (2017). "Evaluating physiological responses of plants to salinity stress." Annals of Botany **119**(1): 1-11.

Nelson, P. N., P. Hooley, D. Roden, H. Davari Ejtehadi, P. Rylance, P. Warren, J. Martin, P. G. Murray and G. Molecular Immunology Research (2004). "Human endogenous retroviruses: transposable elements with potential?" Clinical and experimental immunology **138**(1): 1-9.

Nicolau, M., N. Picault, J. Descombin, Y. Jami-Alahmadi, S. Feng, E. Bucher, S. E. Jacobsen, J.-M. Deragon, J. Wohlschlegel and G. Moissiard (2020). "The plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in Arabidopsis thaliana." PLoS genetics **16**(4): e1008324.

Okamoto, H. and H. Hirochika (2001). "Silencing of transposable elements in plants." Trends in plant science **6**(11): 527-534.

Ono, A., S. H. Kim and V. Walbot (2002). "Subcellular localization of MURA and MURB proteins encoded by the maize MuDR transposon." Plant Mol Biol **50**(4-5): 599-611.

Orgel, L. E. and F. H. Crick (1980). "Selfish DNA: the ultimate parasite." Nature **284**(5757): 604-607.

Parida, A. K. and A. B. Das (2005). "Salt tolerance and salinity effects on plants: a review." Ecotoxicol Environ Saf **60**(3): 324-349.

Pasyukova, E., S. Nuzhdin, T. Morozova and T. Mackay (2004). "Accumulation of transposable elements in the genome of Drosophila melanogaster is associated with a decrease in fitness." Journal of Heredity **95**(4): 284-290.

Pélissier, T., C. Bousquet-Antonelli, L. Lavie and J.-M. Deragon (2004). "Synthesis and processing of tRNA-related SINE transcripts in Arabidopsis thaliana." Nucleic acids research **32**(13): 3957-3966.

Platt, A., H. C. Ross, S. Hankin and R. J. Reece (2000). "The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase." Proceedings of the National Academy of Sciences **97**(7): 3154-3159.

Pouteau, S., E. Huttner, M. A. Grandbastien and M. Caboche (1991). "Specific expression of the tobacco Tnt1 retrotransposon in protoplasts." The EMBO journal **10**(7): 1911-1918.

Quadrana, L., A. B. Silveira, G. F. Mayhew, C. LeBlanc, R. A. Martienssen, J. A. Jeddeloh and V. Colot (2016). "The Arabidopsis thaliana mobilome and its impact at the species level." elife **5**: e15716.

Ritter, A., S. Inigo, P. Fernandez-Calvo, K. S. Heyndrickx, S. Dhondt, H. Shi, L. De Milde, R. Vanden Bossche, R. De Clercq, D. Eeckhout, M. Ron, D. E. Somers, D. Inze, K. Gevaert, G. De Jaeger, K. Vandepoele, L. Pauwels and A. Goossens (2017). "The transcriptional repressor complex FRS7-FRS12 regulates flowering time and growth in Arabidopsis." Nat Commun **8**: 15235.

Roccaro, M., Y. Li, H. Sommer and H. Saedler (2007). "ROSINA (RSI) is part of a CACTA transposable element, TamRSI, and links flower development to transposon activity." Molecular Genetics and Genomics **278**(3): 243-254.

Roldan, M. V. G., C. Périlleux, H. Morin, S. Huerga-Fernandez, D. Latrasse, M. Benhamed and A. Bendahmane (2017). "Natural and induced loss of function mutations in SlMBP21 MADS-box gene led to jointless-2 phenotype in tomato." Scientific reports **7**(1): 1-10.

Sabogal, A., A. Y. Lyubimov, J. E. Corn, J. M. Berger and D. C. Rio (2010). "THAP proteins target specific DNA sites through bipartite recognition of adjacent major and minor grooves." Nature Structural & Molecular Biology **17**(1): 117-U145.

Sanchez, D. H., H. Gaubert and W. Yang (2019). "Evidence of developmental escape from transcriptional gene silencing in MESSI retrotransposons." New Phytologist **223**(2): 950-964.

Shapiro, J. (1969). "Mutations caused by the insertion of genetic material into the galactose operon of Escherichia coli." Journal of molecular biology **40**(1): 93-105.

Shavrukov, Y. (2013). "Salt stress or salt shock: which genes are we studying?" Journal of Experimental Botany **64**(1): 119-127.

Shi, H., L. Xiong, B. Stevenson, T. Lu and J. K. Zhu (2002). "The Arabidopsis salt overly sensitive 4 mutants uncover a critical role for vitamin B6 in plant salt tolerance." Plant Cell **14**(3): 575-588.

Shrivastava, P. and R. Kumar (2015). "Soil salinity: A serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation." Saudi journal of biological sciences **22**(2): 123-131.

Sigman, M. J. and R. K. Slotkin (2016). "The first rule of plant transposable element silencing: location, location, location." The Plant Cell **28**(2): 304-313.

Silva, J. C. and M. G. Kidwell (2004). "Evolution of P elements in natural populations of Drosophila willistoni and D. sturtevanti." Genetics **168**(3): 1323-1335.

Silver, P. A. (1991). "How proteins enter the nucleus." Cell **64**(3): 489-497.

Singer, T., C. Yordan and R. A. Martienssen (2001). "Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1)." Genes & development **15**(5): 591-602.

Sinzelle, L., Z. Izsvak and Z. Ivics (2009). "Molecular domestication of transposable elements: From detrimental parasites to useful host genes." Cellular and Molecular Life Sciences **66**(6): 1073-1093.

Siomi, M. C., K. Sato, D. Pezic and A. A. Aravin (2011). "PIWI-interacting small RNAs: the vanguard of genome defence." Nature reviews Molecular cell biology **12**(4): 246-258.

Sparkes, I. A., J. Runions, A. Kearns and C. Hawes (2006). "Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants." Nature protocols **1**(4): 2019-2025.

Stitzer, M. C., S. N. Anderson, N. M. Springer and J. Ross-Ibarra (2021). "The genomic ecosystem of transposable elements in maize." PLoS Genet **17**(10): e1009768.

Studer, A., Q. Zhao, J. Ross-Ibarra and J. Doebley (2011). "Identification of a functional transposon insertion in the maize domestication gene tb1." Nature genetics **43**(11): 1160-1163.

Sun, J., N. Movahed and H. Zheng (2020). "LUNAPARK Is an E3 Ligase That Mediates Degradation of ROOT HAIR DEFECTIVE3 to Maintain a Tubular ER Network in Arabidopsis." The Plant Cell **32**(9): 2964-2978.

Tambones, I. L., A. Haudry, M. C. Simão and C. Carareto (2019). "High frequency of horizontal transfer in Jockey families (LINE order) of drosophilids." Mobile DNA **10**(1): 1-15.

Tavakkoli, E., P. Rengasamy and G. K. McDonald (2010). "High concentrations of Na+ and Cl–ions in soil solution have simultaneous detrimental effects on growth of faba bean under salinity stress." Journal of Experimental Botany **61**(15): 4449-4459.

Taylor, A. L. (1963). "Bacteriophage-induced mutation in Escherichia coli." Proceedings of the National Academy of Sciences of the United States of America **50**(6): 1043.

Teixeira, F. K., M. Okuniewska, C. D. Malone, R.-X. Coux, D. C. Rio and R. Lehmann (2017). "piRNA-mediated regulation of transposon alternative splicing in the soma and germ line." Nature **552**(7684): 268-272.

Thomas, J., M. Sepahi, B. Arendall and H. Bohnert (1995). "Enhancement of seed germination in high salinity by engineering mannitol expression in Arabidopsis thaliana." Plant, Cell & Environment **18**(7): 801-806.

Tian, T., Y. Liu, H. Yan, Q. You, X. Yi, Z. Du, W. Xu and Z. Su (2017). "agriGO v2. 0: a GO analysis toolkit for the agricultural community, 2017 update." Nucleic acids research **45**(W1): W122-W129.

Ühlken, C., B. Horvath, R. Stadler, N. Sauer and M. Weingartner (2014). "MAIN‐LIKE 1 is a crucial factor for correct cell division and differentiation in A rabidopsis thaliana." The Plant Journal **78**(1): 107-120.

Velanis, C. N., P. Perera, B. Thomson, E. de Leau, S. C. Liang, B. Hartwig, A. Förderer, H. Thornton, P. Arede and J. Chen (2020). "The domesticated transposase ALP2 mediates formation of a novel Polycomb protein complex by direct interaction with MSI1, a core subunit of Polycomb Repressive Complex 2 (PRC2)." PLoS genetics **16**(5): e1008681.

Volff, J. N. (2006). "Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes." Bioessays **28**(9): 913-922.

Wang, H. and X. W. Deng (2002). "Arabidopsis FHY3 defines a key phytochrome A signaling component directly interacting with its homologous partner FAR1." EMBO J **21**(6): 1339-1349.

Wang, H. and H. Wang (2015). "Multifaceted roles of FHY3 and FAR1 in light signaling and beyond." Trends Plant Sci **20**(7): 453-461.

Wang, S., Y. Chang, J. Guo and J.-G. Chen (2007). "Arabidopsis Ovate Family Protein 1 is a transcriptional repressor that suppresses cell elongation." The Plant Journal **50**(5): 858-872.

Wang, T., L. Ren, C. Li, D. Zhang, X. Zhang, G. Zhou, D. Gao, R. Chen, Y. Chen and Z. Wang (2021). "The genome of a wild Medicago species provides insights into the tolerant mechanisms of legume forage to environmental stress." BMC biology **19**(1): 1-17.

Wang, W., W. Tang, T. Ma, D. Niu, J. B. Jin, H. Wang and R. Lin (2016). "A pair of light signaling factors FHY3 and FAR1 regulates plant immunity by modulating chlorophyll biosynthesis." J Integr Plant Biol **58**(1): 91-103.

Wang, X. L., Y. Q. He, H. Wei and L. Wang (2021). "A clock regulatory module is required for salt tolerance and control of heading date in rice." Plant Cell and Environment **44**(10): 3283-3301.

Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A P. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede (2018). "SWISS-MODEL: homology modelling of protein structures and complexes." Nucleic Acids Research **46**(W1): W296-W303.

Weigel, D. and J. Glazebrook (2006). "Transformation of agrobacterium using electroporation." CSH protocols **2006**(7).

Weiss, S., I. Gottfried, I. Mayrose, S. L. Khare, M. Xiang, S. J. Dawson and K. B. Avraham (2003). "The DFNA15 deafness mutation affects POU4F3 protein stability, localization, and transcriptional activity." Mol Cell Biol **23**(22): 7957-7964.

Wenig, U., S. Meyer, R. Stadler, S. Fischer, D. Werner, A. Lauter, M. Melzer, S. Hoth, M. Weingartner and N. Sauer (2013). "Identification of MAIN, a factor involved in genome stability in the meristems of Arabidopsis thaliana." The plant journal **75**(3): 469-483.

Wessler, S. R. (2006). "Transposable elements and the evolution of eukaryotic genomes." Proceedings of the National Academy of Sciences of the United States of America **103**(47): 17600-17601.

Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel and A. H. Schulman (2007). "A unified classification system for eukaryotic transposable elements." Nature Reviews Genetics **8**(12): 973-982.

Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel and A. H. Schulman (2007). "A unified classification system for eukaryotic transposable elements." Nat Rev Genet **8**(12): 973-982.

Wolf, G., A. de Iaco, M.-A. Sun, M. Bruno, M. Tinkham, D. Hoang, A. Mitra, S. Ralls, D. Trono and T. S. Macfarlan (2020). "KRAB-zinc finger protein gene expansion in response to active retrotransposons in the murine lineage." Elife **9**: e56337.

Wu, S., H. Xiao, A. Cabrera, T. Meulia and E. van der Knaap (2011). "SUN Regulates Vegetative and Reproductive Organ Shape by Changing Cell Division Patterns1 [C][W][OA]."

Wu, S. J., L. Ding and J. K. Zhu (1996). "SOS1, a genetic locus essential for salt tolerance and potassium acquisition." Plant Cell **8**(4): 617-627.

Xiao, H., N. Jiang, E. Schaffner, E. J. Stockinger and E. Van Der Knaap (2008). "A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit." science **319**(5869): 1527-1530.

Xie, Y., Q. Zhou, Y. Zhao, Q. Li, Y. Liu, M. Ma, B. Wang, R. Shen, Z. Zheng and H. Wang (2020). "FHY3 and FAR1 integrate light signals with the miR156-SPL module-mediated aging pathway to regulate Arabidopsis flowering." Molecular plant **13**(3): 483-498.

Xu, J., M. Mcpartlon and J. Li (2021). "Improved protein structure prediction by deep learning irrespective of co-evolution information." Nature Machine Intelligence: 1-9.

Xu, X., C. Chen, B. Fan and Z. Chen (2006). "Physical and Functional Interactions between Pathogen-Induced Arabidopsis WRKY18, WRKY40, and WRKY60 Transcription Factors." The Plant Cell **18**(5): 1310-1326.

Yamasaki, K., T. Kigawa, M. Inoue, M. Tateno, T. Yamasaki, T. Yabuki, M. Aoki, E. Seki, T. Matsuda, Y. Tomo, N. Hayami, T. Terada, M. Shirouzu, A. Tanaka, M. Seki, K. Shinozaki and S. Yokoyama (2005). "Solution Structure of an Arabidopsis WRKY DNA Binding Domain." The Plant Cell **17**(3): 944-956.

Yamasaki, K., T. Kigawa, M. Seki, K. Shinozaki and S. Yokoyama (2013). "DNA-binding domains of plant-specific transcription factors: structure, function, and evolution." Trends Plant Sci **18**(5): 267-276.

Yang, G., F. Zhang, C. N. Hancock and S. R. Wessler (2007). "Transposition of the rice miniature inverted repeat transposable element mPing in Arabidopsis thaliana." Proceedings of the National Academy of Sciences **104**(26): 10962-10967.

Yang, J., I. Anishchenko, H. Park, Z. Peng, S. Ovchinnikov and D. Baker (2020). "Improved protein structure prediction using predicted interresidue orientations." Proceedings of the National Academy of Sciences **117**(3): 1496-1503.

Yang, J., H. S. Malik and T. H. Eickbush (1999). "Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements." Proceedings of the National Academy of Sciences **96**(14): 7847-7852.

Yoon, Y., D. H. Seo, H. Shin, H. J. Kim, C. M. Kim and G. Jang (2020). "The role of stress-responsive transcription factors in modulating abiotic stress tolerance in plants." Agronomy **10**(6): 788.

Yoshitake, Y., N. Inomata, M. Sano, Y. Kato and M. Itoh (2018). "The P element invaded rapidly and caused hybrid dysgenesis in natural populations of Drosophila simulans in Japan." Ecology and evolution **8**(19): 9590-9599.

Yu, Z. H., S. I. Wright and T. E. Bureau (2000). "Mutator-like elements in Arabidopsis thaliana: Structure, diversity and evolution." Genetics **156**(4): 2019-2031.

Yuan, Y. W. and S. R. Wessler (2011). "The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies." Proceedings of the National Academy of Sciences of the United States of America **108**(19): 7884-7889.

Yusa, K. (2015). "piggyBac Transposon." Microbiol Spectr **3**(2): Mdna3-0028-2014.

Zaikina, E. A., S. D. Rumyantsev, E. R. Sarvarova and B. R. Kuluev (2019). "Transcription factor genes involved in plant response to abiotic stress factors." Ecological genetics **17**(3): 47-58.

Zemach, A., M. Y. Kim, P.-H. Hsieh, D. Coleman-Derr, L. Eshed-Williams, K. Thao, S. L. Harmer and D. Zilberman (2013). "The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin." Cell **153**(1): 193-205.

Zhang, G., T. Yu, S. S. Parhad, S. Ho, Z. Weng and W. E. Theurkauf (2021). "piRNA-independent transposon silencing by the Drosophila THO complex." Developmental Cell **56**(18): 2623-2635. e2625.

Zhang, H. M., J. H. Zhu, Z. Z. Gong and J. K. Zhu (2021). "Abiotic stress responses in plants." Nature Reviews Genetics.

Zhang, X., R. Henriques, S.-S. Lin, Q.-W. Niu and N.-H. Chua (2006). "Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method." Nature protocols **1**(2): 641-646.

Zhang, Y., Y. Lv, N. Jahan, G. Chen, D. Ren and L. Guo (2018). "Sensing of abiotic stress and ionic stress responses in plants." International journal of molecular sciences **19**(11): 3298.

Zhao, K., S. Cheng, N. Miao, P. Xu, X. Lu, Y. Zhang, M. Wang, X. Ouyang, X. Yuan and W. Liu (2019). "A Pandas complex adapted for piRNA-guided transcriptional silencing and heterochromatin formation." Nature cell biology **21**(10): 1261-1272.

Zhong, M.-C., X.-D. Jiang, W.-H. Cui and J.-Y. Hu (2021). "Expansion and expression diversity of FAR1/FRS-like genes provides insights into flowering time regulation in roses." Plant diversity **43**(2): 173-179.

Concluding Remarks

In conclusion, this thesis aims to understand the profile of an expated transposable element family *MUSTANG-A*, the molecular characterization of its members, *MUG1-MUG4*, and the roles of *MUG4* under abiotic stress and the potential application of ETEs in crop improvement.

I first present an overview introduction of TE, such as the types of TE, the activation of TE, TE silencing, and the effects of TE. Then I introduced the main families of exapted transposable elements and their functional characterization. Finally, I reviewed the effects of TEs and ETEs on abiotic stress and agriculture improvement.

In Chapter II, I showed a detailed molecular profile of *MUG-A* family. All MUG-A proteins have a strict nuclear localization and are capable to mediate transcriptional repression in plants. Through genetic and phenotypic analysis, the *MUG-A* genes are shown to form homodimers and heterodimers with overlapping functions that are beneficial to plant development.

In Chapter III, I explored the deeper mechanism of how MUG4 responds to salt in plants. Through a set of physiological assays, I found that MUG4 specifically prevents plants from accumulating extra sodium, potassium, and calcium. The expression of abiotic stress-responsive genes was significantly higher in *mug4* mutant plants than wild-type plants under salt stress conditions. Furthermore, overexpression of AtMUG4 increases salt tolerance in both *A. thaliana* and *Camelina sativa*.

In Chapter IV, I performed a systematic analysis using a site-directed mutagenesis approach to understand the function of each domain of MUG4 in response to salt. The three domains of MUG4 perform discrete but essential functions in ensuring its subcellular localization,

heterodimer formation and salt response. Additionally, the results suggest that the localization, heterodimer formation and salt response of MUG4 are tightly correlated.

Taken together, the work presented in this thesis addresses the function of ETEs from various angles, from molecular functions to applications in crops. Given their predominating presence in the plant genomes, ETEs hold unexplored potential for salt and even abiotic stress response improvement programs.