VIETNAM NATIONAL UNIVERSITY OF FORESTRY
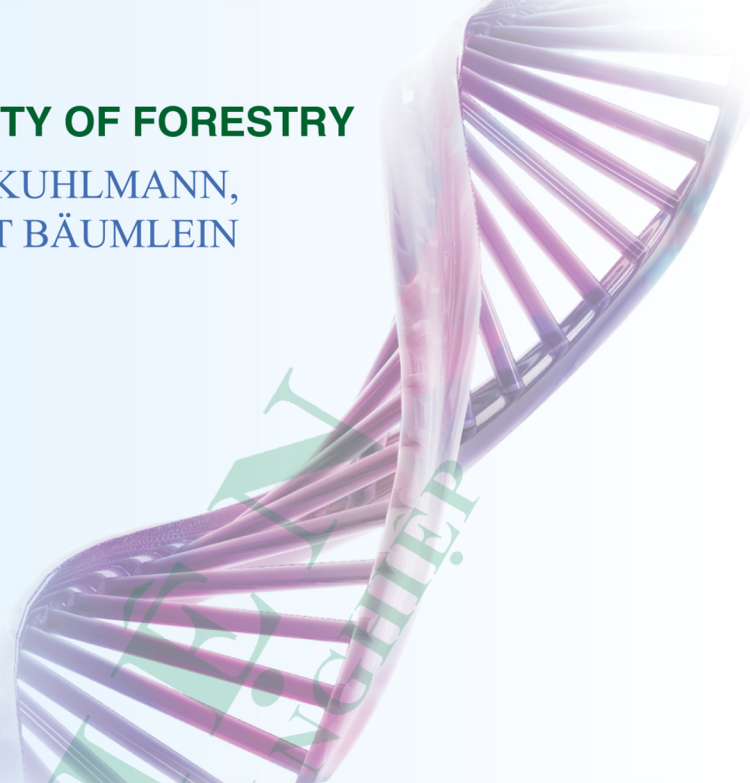
BUI THI MAI HUONG, MARKUS KUHLMANN,
LOTHAR ALTSCHMIED, HELMUT BÄUMLEIN

EFFECTOR OF TRANSCRIPTION (ET)

Novel epigenetic regulators associated with genomic DNA methylation in Arabidopsis

(Monographs)

BUI THI MAI HUONG (CHIEF AUTHOR), MARKUS KUHLMANN,
LOTHAR ALTSCHMIED, HELMUT BÄUMLEIN

EFFECTOR OF TRANSCRIPTION (ET)

Novel epigenetic regulators associated with genomic DNA methylation in Arabidopsis
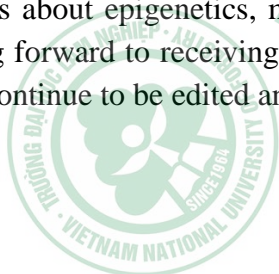
(Monographs)

ABOUT THIS BOOK

Plant development depends on complex regulatory interactions which includes the coordination of numerous transcriptional networks. And interactions of transcription factors with DNA are essential for regulating gene expression, these are often modified through epigenetic mechanisms such as DNA methylation and histone modifications.

Epigenetics caused alterations in the transcriptional potential of a cell which are not caused by changes in the DNA sequence including DNA methylation and histone modifications leading to gene silencing.

Plant-specific effectors of transcription (ET) are characterised by the highly conserved ET repeats, which are involved in zinc and DNA binding. Moreover, ETs share a GIY-YIG domain, which involved in DNA nicking activity. Therefore, it was hypothesised that ETs might act as epigenetic regulators.

This book We describe a more detailed functional analysis of the ET gene family using a genetic approach. T-DNA insertion mutants of Arabidopsis have been isolated and genotyped. A thorough phenotypic description of single and double mutants reveals pleiotropic developmental effects. This includes for instance the failed fusion of the two polar nuclei as prerequisite for double fertilisation and endosperm development. The *et* mutants exhibit a conspicuous homoeotic transformation of flower organs with anthers transformed into carpels containing rather well developed stigmata and ovules with nearly normal embryo sac formation. The endosperm nuclei of the mutants exhibit unusually large nucleoli probably indicating a high synthetic activity. Finally, the mutants germinate precociously when still attached to the mother plant with the cotyledons and not the root tip penetrating the seed coat first. Searching for putative target genes a comparative molecular analysis including deep RNA sequencing and genome-wide methylation studies have been performed. Together, the results provide strong evidence for the conclusion that ETs are novel epigenetic regulators of reproductive processes and act *via* the regulation of the DNA methylation status of plant genomes.

We hope that this book will provide the students, researchers and lecturers with the knowledges about epigenetics, molecular genetics in research as well as in teaching. We are looking forward to receiving your sharing and suggestions to this book, in order to this book can continue to be edited and added more valuable data.

TABLE OF CONTENTS

Chapter 1
THE CHARACTERIZATION AND PHENOTYPIC ANALYSIS OF SINGLE MUTANT ALLELES *ET1, ET2, ET3*, DOUBLE MUTANTS ALLELES *ET1/ET2* OF THE ET GENE FAMILY

Chapter 2
THE IDENTIFICATION OF DIFFERENTIALLY METHYLATED TARGET REGIONS USING A WHOLE GENOME METHYLATION APPROACH

Chapter 3

THE IDENTIFICATION OF PUTATIVE TARGET GENES BY TRANSCRIPTOME ANALYSIS USING DEEP SEQUENCING TECHNOLOGY

6

LIST OF ABBREVIATIONS

FM	Functional megaspore
PMC	Pollen mother cells
TAG	Triacylglycerols
ABA	Abscisic acid
GA	Gibberellic acid
AP	APETALA
SEP	SEPALLATA
siRNA	Small interfering RNA
Col 0	Columbia-0
Ws	Wassilewskija-2
bHLH	Basic Helix Loop Helix
DAPI	4, 6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
HRT	HORDEUM REPRESSOR OF TRANSCRIPTION
mRNA	Messenger RNA
MS	Murashige and Skoog
RNA	Ribonucleic acid
UV	Ultraviolet
EDTA	Ethylenediamine tetraacetic acid
SDS	Sodium dodecyl sulfate
DMSO	Dimethyl sulfoxide
cDNA	Complementary DNA
PCR	Polymerase chain reaction
RT-PCR	Reverse transcription PCR
SDS	Sodium dodecyl sulfate
T-DNA	Transferred DNA
MYB	Myeloblastosis
MET1	Methyltransferease1
CMT3	Chromomethylase3

HDA6	Histone deacetylase
SUVH	SUPPRESSOR OF VARIEGATION 3-9 HOMOLOGUE
RDR2	RNA-dependent-RNA polymerase 2
DCL3	Dicer-like protein
HEN1	HUA ENHANCER 1
AGO4	ARGONAUTE4 protein
POL IV	Polymerase IV
DRM2	DNA methyltransferase
ET	EFFECTOR OF TRANSCRIPTION
DML	DEMETER-LIKE
DME	DEMETER
ROS1	REPRESSOR OF SILENCING1
BER	Base excision repair
MEA	MEDEA
X-gal	5-bromo-4-chloro-3-indolyl-ß-D-galactoside
IPTG	Isopropyl-ß-D-thiogalactoside
EGTA	Ethyleneglycol tetraacetic acid
MALDI-TOF	Matrix assisted laser desorption ionization-time-of-flight
TAIR	The arabidopsis information resources
KNAT	Knottedlike homeobox of arabidopsis thaliana
AtET1,2,3	Indicates the gene (Capital, italic)
AtET1,2,3	Indicates the protein (Capital)
et1,2	Indicates the mutant allele (Lowercase, italic)

Chapter 1

THE CHARACTERIZATION AND PHENOTYPIC ANALYSIS OF SINGLE MUTANT ALLELES *ET1*, *ET2*, *ET3*, DOUBLE MUTANTS ALLELES *ET1/ET2* OF THE ET GENE FAMILY

Bui Thi Mai Huong[1], Helmut Bäumlein[2]

1: Vietnam National University of Forestry

2: Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

1.1. ABSTRACT

EFFECTOR OF TRANSCRIPTION (ET) are plant-specific regulatory proteins, which consists of GIY-YIG domain and conserved ET repeats domain. In which DNA single strand cutting GIY-YIG domain are structural similarity with bacterial UVRC proteins, ET repeats domain are involved in zinc and DNA binding. It was hypothesised that ETs might act as epigenetic regulators. The screening single mutant alleles *et1*, *et2*, *et3* and generating double mutant alleles *et1/et2* have been proposed to identifying and confirming the function of ET family. Identifying a mutation in the gene of interest by PCR screening of pools of insertion lines, using one primer corresponding to interest genes and one primer corresponding to the end of the insertion element. Two double mutants have been generated by crossing the homozygous mutant *et1-1* with homozygous *et2-1* and *et2-3* mutants. Homozygous double mutants *et1-1 et2-1* and *et1-1 et2-3* have been selected and characterized in the F2 generation. And with the single mutant alleles *et1*, *et2* and double mutant alleles *et1/et2* have been analysed phenotype by histological methods. The phenotypic *et* mutants exhibited several highly interesting observation, including effects on flower organ identity, gametophyte development, endosperm development, immature seed germination, pollen development and seed development.

1.2. BACKGROUND

1.2.1. The EFFECTOR OF TRANSCRIPTION (ET) gene family

EFFECTOR OF TRANSCRIPTION (ET) genes were originally isolated by using South Western screens of transcription factors important for embryonic gene regulation (Ellerström et al., 2005; Ivanov et al., 2008). Previous work found a plant specific class of gene regulators of barley, broad bean, rape seed and Arabidopsis (Raventós et al., 1998; Ellerström et al., 2005; Ivanov et al., 2008) designated as HORDEUM REPRESSOR OF TRANSCRIPTION (HRT) in monocots and EFFECTOR OF TRANSCRIPTION (ET) in dicots. The characteristic feature of the ET factors is a variable number of highly

conserved (C, R, H, K), zinc and DNA binding ET repeats as shown for *Brassica (rape), Arabidopsis, Vicia (bean), Hordeum (barley)* and *Physcomitrella (moss)*. Strictly **plant specific**, **zinc- and DNA-binding** protein family characterised by highly conserved cystein-containing **repeat domains** ($\underline{\mathbf{C}}$-X_{8-9} –$\underline{\mathbf{C}}$-X_{10} -$\underline{\mathbf{C}}$-X_2-$\underline{\mathbf{H}}$) (Derbyshire et al., 1997, Ellerström et al., 2005; Ivanov et al., 2005, 2008, 2012) (Fig 1.1). ET proteins share variable numbers of highly conserved cysteine-histidine containing, zinc- and DNA binding repeats also found in lower plants such as the moss *Physcomitrella patens* demonstrating their evolutionary conservation (Ellerström et al., 2005; Ivanov et al., 2008).

Fig.1.1. Structure of ET repeat domains in Plants

Besides these DNA binding ET-repeats, ET factors share a characteristic DNA single strand cutting domain (GIY-YIG) with structural similarity to that of bacterial UVRC proteins and so called homing nucleases, They all contain a highly conserved arginine residue known as to be part of the active center (Derbyshire et al., 1997; Aravind et al., 1999; Verhoeven et al., 2000; Stoddard, 2005). The bacterial UVRC protein is essential for DNA excision repair to UV-induced DNA lesions like thymidine-dimers and introduces two single strand cuts 8 bp 5` and 4 bp 3` of the lesion (Friedberg et al., 1995; Moolenaar et al., 1998). A C-terminal domain consist of an Endonuclease V (EndoV) and Helix-hairpin-Helix (HhH) domain which is required for the 5`cut, whereas the N-terminal GIY-YIG domain inserts the 3`nick (Lin and Sancar, 1992; Derbyshire et al., 1997; Aravind et al., 1999; Kowalski et al., 1999; Verhoeven et al., 2000; Van Roey et al., 2002). The sequence similarity between plant ET factors and UVRC is only restricted to the single strand cutting GIY-YIG domain. This suggests that an ancestral bacterial GIY-YIG domain

12

has been recruited by ET proteins and attached to the DNA-binding ET repeats to create a novel plant specific regulatory protein (Fig. 1.2).
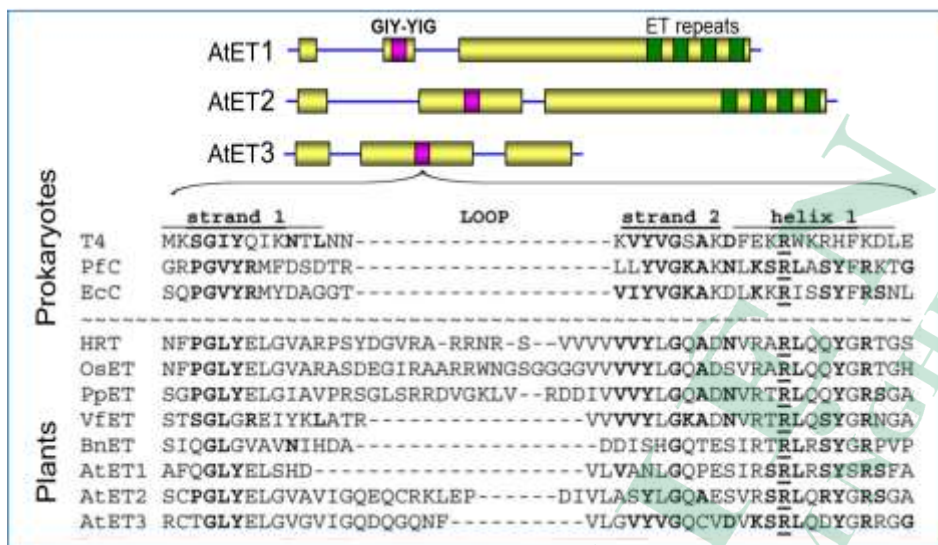
Fig.1.2. Structure of GIY-YIG domains in Prokaryotes and Plants

In Arabidopsis the ET family included three genes: *AtET1* (AT4G26170) is located on the fourth chromosome, and while the other genes, *AtET2* (AT5G56780) and *AtET3* (AT5G56770) are located on the fifth chromosome (Ellerström et al., 2005; Ivanov, 2005; Ivanov et al., 2008). The *AtET1* and *AtET2* genes are intact coding sequence, whereas *AtET3* is a truncated version of *AtET2* due to the lack of the ET repeat domain. In three ET genes the GIY-YIG domain is encoded by the second exon which is consistent with a corresponding domain shuffling event during protein evolution. The functionality of the ET genes was demonstrated by substituting the AtET2 GIY-YIG domain for the corresponding domain of the *E. coli* UVRC protein (Ivanov et al., 2008). We proposed a hypothesis is that the nicking activity of the plant ET factor GIY-YIG domain may be involved in the catalysis of changes in higher order DNA structures, such as, nucleosome sliding or the relaxation of supercoiled chromatin domains as a requirement for regulated gene expression (Choi et al., 2002; Xiao et al., 2003; Haince et al., 2006; Ju et al., 2006). Alternatively, the domain could be involved in active de-methylation processes as described for the plant regulators DEMETER (DME) and REPRESSOR OF SILENCING1 (ROS1) (Choi et al., 2002; Gong et al., 2002; Xiao et al., 2003; Choi et al., 2004; Morales-Ruiz et al., 2006). DME can induce single strand nicks in the MEDEA (MEA) promoter as part of a DNA de-methylation pathway involved in the epigenetic imprinting of the MEA gene. ROS1 represses homology-dependent transcriptional silencing by de-methylation the target promoter DNA (Gong et al., 2002). Thus, a possible evolutionary scenario is that plant-specific ET factors have recruited a single GIY-YIG domain from prokaryotic repair-related proteins by a domain shuffling process, joining this domain to the DNA-binding ET repeat. The resulting plant specific protein is involved in repair processes but acts as a gene

regulator (Fig. 1.3). The regulatory mechanism in part analogous to the function of DME and ROS1 might include the insertion of nicks, with an impact on higher order structures of chromatin packed DNA or on the genomic DNA methylation pattern required for differentiation processes for instance during seed development.
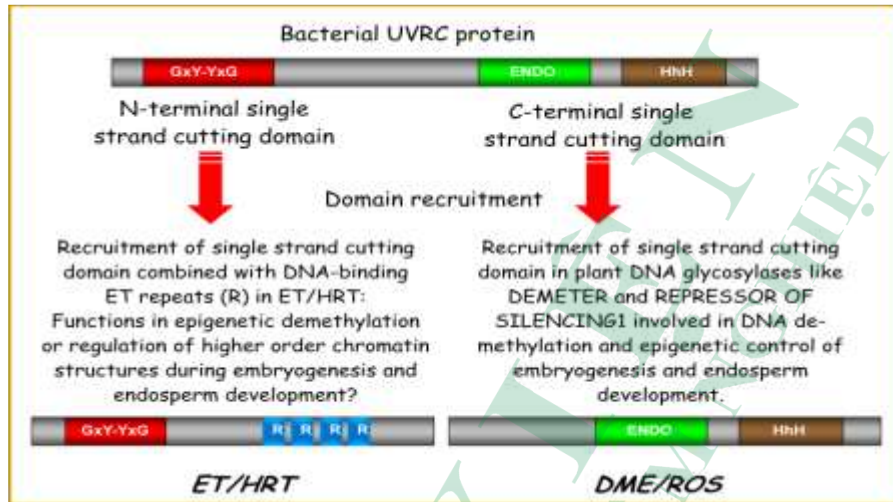
Fig.1.3. Putative protein evolution scenario including the recruitment of two different DNA-single strand cutting domains of bacterial UVRC proteins by the plant regulatory proteins HRT/ET and DME/ROS

The HRT/ET factors have adopted the N-terminal GxY-YxG domain whereas DME and ROS exploit the C-terminal single strand cutting domain of UVRC.
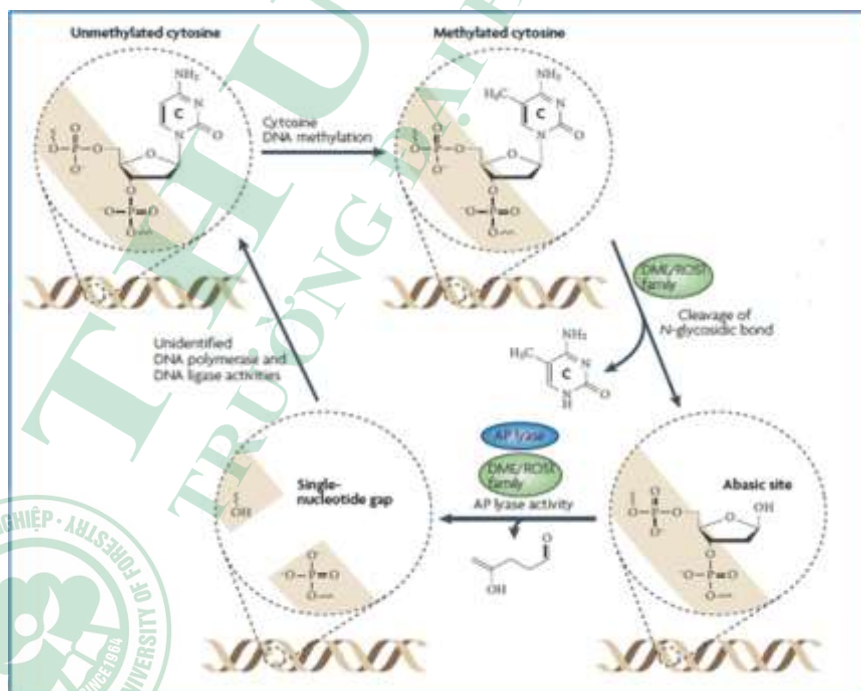
Fig.1.4. Active DNA demethylation through DNA glycosylase activity and base excision repair

In *Arabidopsis thaliana* (green proteins), methylated (CH$_3$) cytosine (bold C) bases are removed by the DEMETER (DME)/REPRESSOR OF SILENCING 1 (ROS1) family of bifunctional 5-methylcytosine glycosylases. First, the methylated cytosine base is released by cleavage of the *N*-glycosidic bond, generating an abasic site. Next, the phosphodiester linkage is broken both 3′ and 5′ of the abasic site through apyrimidic (AP) lysase activity, generating a single-nucleotide gap in the DNA. The DNA is then proposed to be repaired by unknown DNA polymerase and DNA ligase activities, resulting in a net loss of cytosine methylation (He et al.,2011).

1.2.2. *Plant reproductive processes*

Plant reproductive processes included major developmental pathways: female and male sporogenesis, female and male gametogenesis, double fertilisation, embryogenesis, endosperm formation, seed maturation including synthesis of storage compounds, acquisition of desiccation tolerance and dormancy as well as germination (Fig. 1.5).
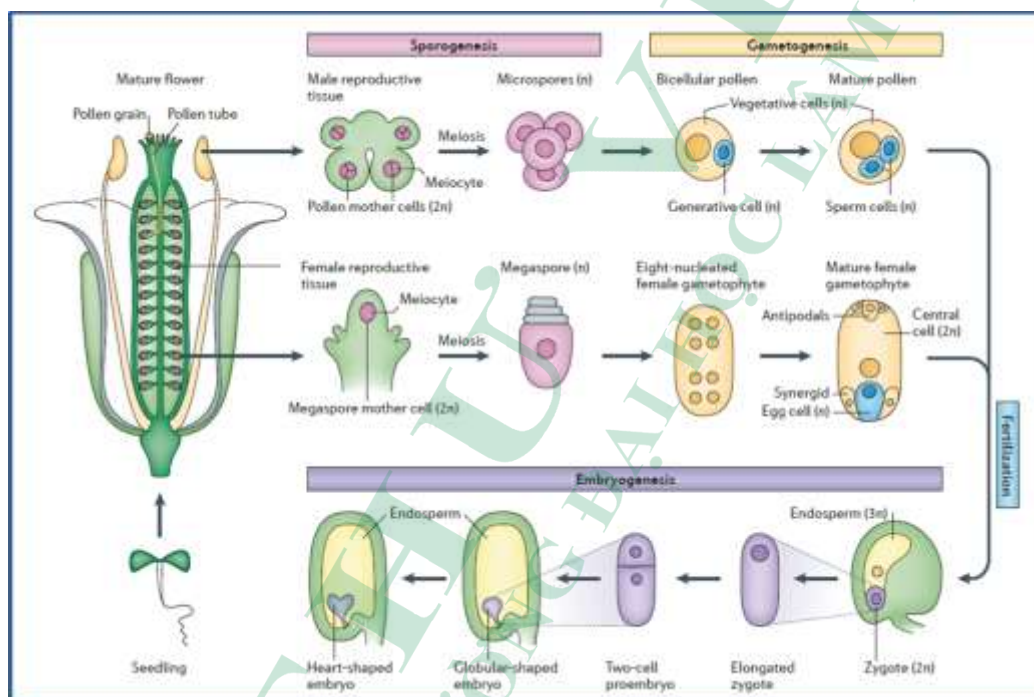
Fig.1.5. Scheme of sexual plant reproduction

(Kawashima and Berger, 2014).

Sporogenesis is initiated both in pollen mother cells (PMC) and megaspore mother cell (MMC) both undergoing meiosis. Whereas all four meiotic products survive in the male pathway, but only the functional megaspore (FM) survives in the female pathway. The plant gametophyte development generate the male gametes which are the sperm cells and the female gametes which are the egg cell and the homodiploid central cell. In the plant specific double fertilization process two sperm cells fuse with the egg cell and central cell to produce the embryo proper and the endosperm, respectively.

1.2.2.1. Female gametophyte of Arabidopsis

The female gametophyte was grown in the ovules on the carpel′s ovary. The female gametophyte of Arabidopsis (Fig.1.6) is similarity the monosporic polygonum type shared by more than 70% of flowering plants (Maheshwari, 1950; Wilemse and van Went, 1984; Haig, 1990; Huang and Russell, 1992; Yadegari and Drews, 2004). A special subepidermal cell, the megaspore mother cell, undergoes meiosis. Three of the meiotic products degenerate by apoptosis , only the chalazal spore survived as functional megaspore (FM). The FM also undergoes three mitosis to generate an eight-nucleate embryo sac. Two polar nuclei fuse to the homodiploid nucleus called the central cell. Two synergids form and the egg cell form the egg apparatus at the micropylar end and three antipodal which were positioned at the chalazal end and degenerate in the mature embryo sac which then ready for fertilisation.
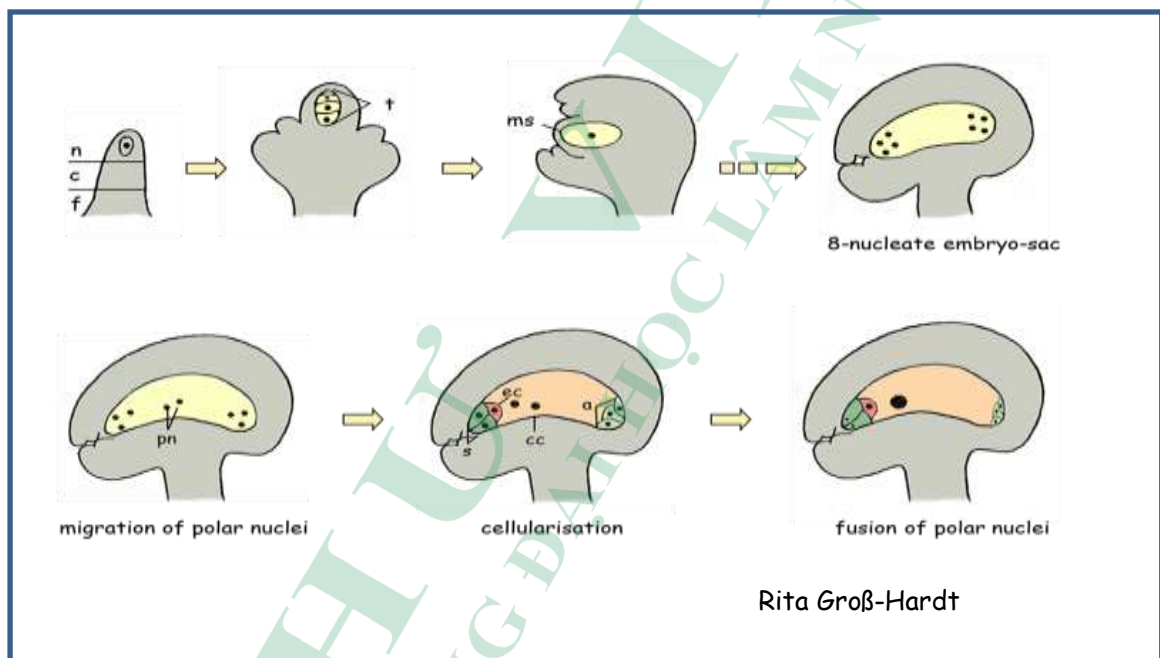
Fig.1.6. Sporogenesis and gametophyte development of Arabidopsis thaliana

1.2.2.2. Male gametophyte of Arabidopsis

In plants, pollen is produced within the anthers of the flower. The reproductive cells give rise to the microspores and the non-reproductive cells form discrete anther tissues layers, including the epidermal, cortical and tapetal cell layers surrounding the sporogenous cells. During microsporogenesis the pollen mother cell (PMC) undergoes meiosis generating four surviving haploid microspores. And then, these unicellular microspores undergo a first mitosis to form two unequal cells, including a large vegetative cell and a small generative cell. The generative cell divides undergo the second mitosis to form the two sperm cells (Fig.1.7).
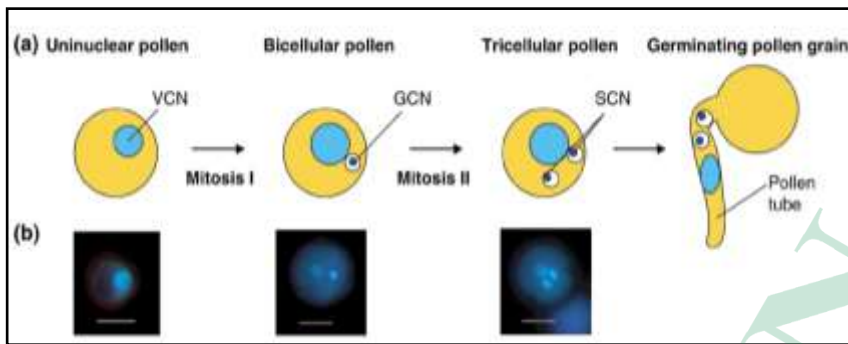
Fig.1.7. Male gametophyte development of Arabidopsis

(Park et al., 1998, Honys and Twell, 2004).

The pollen mother cell undergoes meiosis to form four microspores which subsequently undergo two haploid mitotic steps leading to the mature microgametophyte which consists of a vegetative cell and two sperm cells. a) schematic drawing, b) fluorescence microscopy.

1.2.2.3. Embryogenesis

Embryogenesis of Arabidopsis is initiated by double fertilization. The vegetative cell of the male gametophyte grow to form the pollen tube, interacts with the synergids of the female gametophyte and delivers the two sperm cells. One sperm cell fuses with the egg cell to form the zygote which is initial for the development of the embryo proper. The second sperm cell fuses with the second homodiploid central cell to form the triploid endosperm which is a storage tissue for embryo nutrition (Brown et al., 1999).

The higher plant embryogenesis can be divided into two distinct phases (Fig.1.8). The early phase is characterized by cell proliferation and morphogenesis, including postfertilization and globular-heart transition with the basic body plan of shoot-root polarity being established (West and Harada, 1993; Goldberg et al., 1994; Laux and Jurgens, 1997). A later phase of maturation is characterized by storage compound synthesis mainly in the cotyledons, desiccation tolerance and dormancy (West and Harada, 1993; Goldberg et al., 1994; Lotan et al., 1998; Harada, 2001; Raz et al., 2001).
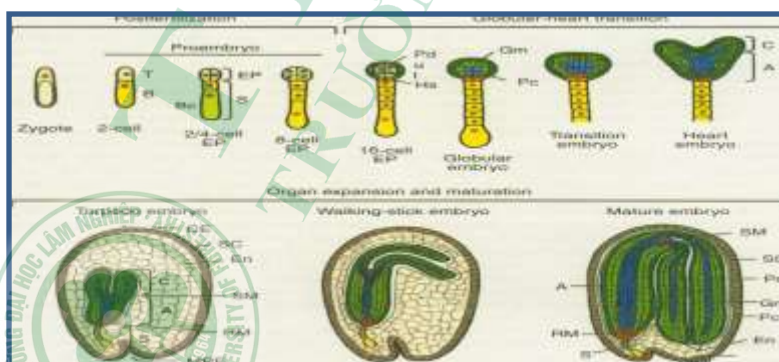
Fig.1.8. Scheme of Arabidopsis embryogenesis

(Goldberg et al., 1994).

17

A generalized overview of plant embryogenesis. Characteristic stages of embryo development comprise the globular-, heart-, torpedo- and walking stick-stages with shoot- (SAM) and root- (RAM) apical meristems indicated. Abbreviations: T, terminal cell; B, basal cell; EP, embryo proper; S, suspensor; Bc, suspensor basal cell; Pd, protoderm; u, upper tier; I, lower tier; Hs, hypophysis; Pc, procambium; Gm, ground meristem; C, cotyledon; A, axis; MPE, micropylar end; CE, chalazal end; SC, seed coat; En, endosperm; SM, shoot meristem; and RM, root meristem.

The storage products of Arabidopsis consists of lipids, proteins and carbohydrates. Seed lipids are stored as triacylglycerols (TAG) in oil bodies (Murphy, 1993; Herman, 1995). TAGs are synthesised from the late heart stage and continues through the torpedo stage until the embryo desiccates. The TAG core is surrounded by a phospholipid monolayer and oleosins, which are special proteins involved in the preservation of the oleosome structure during seed desiccation (Huang et al., 1994; Mansfield and Briarty, 1992). Lipids are analysed by lipases to provide the main energy source of the growing seedling. Seed storage proteins are the primary source of carbon and nitrogen for the growing seedling. In developing Arabidopsis seeds, there are two types of seed storage proteins, the 12S globulins (cruciferins) and the 2S albumins (napins). They are synthesized at the rough endoplasmic reticulum and sorted into the protein storage vacuoles (Müntz, 1998). Starch is found in the plastids of embryo cells and seed coat cells (Focks and Benning, 1998). Later starch is mainly detected in the outer and inner cell layers of the outer integument, but not in the mature embryo (Western et al., 2000; Kim et al., 2005).

1.2.2.4. Germination

Seed germination is mainly regulated by the ratio of the two phytohormones abscisic acid (ABA) and gibberellic acid (GA) (Koornneef et al., 1998; White and Rivin, 2000). ABA concentration increases during late embryogenesis, reaches a peak in the maturation phase and decreases in mature seeds. In this phase ABA prevents germination. Therefore, the mutants affected to ABA synthesis which fail to express maturation specific messengers and are intolerant to desiccation (Black, 1991). Gibberellins (GA) play an important role in the regulation of cell division and expansion as well as in seed germination (Olszewski et al., 2002). External GA application causes premature seed germination (Debaujon and Koornneef, 2000; White and Rivin, 2000).

1.2.2.5. Flowering control

Floral meristems form the four different floral organs: sepals, petals, stamens and carpels (Coen and Carpenter, 1993) including concentric rings, called the whorls, around the flanks of the meristem. In Arabidopsis the four whorls are arranged as follows: the first outermost whorl is 4 green sepals; the second whorl consists of 4 petals with white color at maturity; the third whorl is composed of six stamens, two of which are shorter than the other four and the fourth whorl is the gynoecium or pistil, which consists of an ovary with

two fused carpels, each containing numerous ovules and a short style capped with a stigma (Fig.1.9).
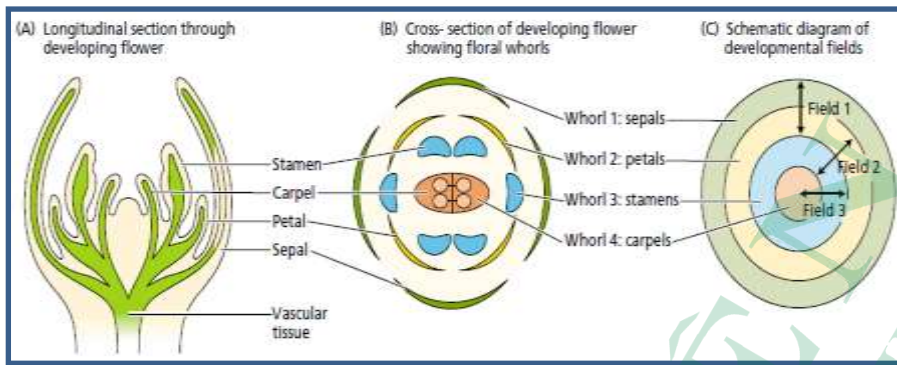
Fig.1.9. Scheme of the Arabidopsis flower

(Bewley et al. 2000).

The floral organs are produced as successive whorls (concentric circles), starting with the sepals, petals, stamens and carpels. According to the ABC model, the functions of each whorl are determined by the expression patterns of specific floral organ identity genes.

Homoeotic mutants with changed organ identity define such as several transcription factors of the MADS box class. At least five MADS box genes are known to specify floral organ identity in Arabidopsis: APETALA1 (AP1), APETALA2 (AP2), APETALA3 (AP3), PISTILATA (PI) and AGAMOUS (AG) (Bowman et al., 1989; Weigel and Meyerowitz, 1994) which are summarized as A (AP1, AP2), B (AP3, PI) and C (AG) function with A expressed in the first and second whorl, B identify active in the second and third whorl and C is activated in the third and the fourth whorl. The type E activity is encoded by SEPALLATA (SEP),1, 2, 3 and 4 (Pelaz et al., 2000) and might be required for the combinatorial function of A, B and C (Soltis et al., 2007).
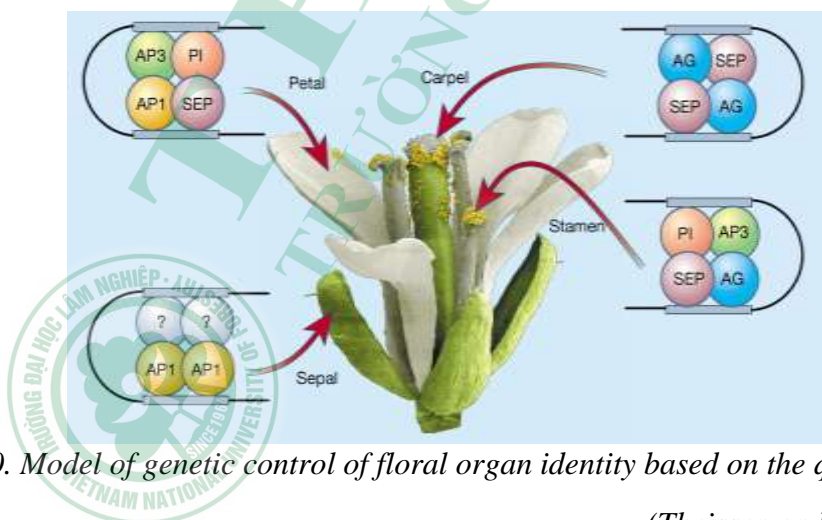
Fig.1.10. Model of genetic control of floral organ identity based on the quartet model

(Theissen and Saedler, 2001).

1.3. MATERIALS AND METHODS

1.3.1. Materials

1.3.1.1. Plant materials

Arabidopsis thaliana ecotypes Columbia-0 (Col) and *Wassilewskija-2* (Ws) were obtained from Gene Regulation Group (IPK, Gatersleben, Germany) and used throughout this study as wild type control experiments. T-DNA insertion lines have been received from Nottingham Arabidopsis stock center. From the genetic and molecular analysis of several SALK lines the following stable mutant lines have been obtained: *et1-1; et1-5; et2-3; et3-2; et3-3*. The line *et2-1* has been isolated from the Arabidopsis Knock-out Facility (AKF) at the University of Wisconsin Biotechnology center.

1.3.1.2. Bacterial strains

Several bacterial strains were used for different purposes such as DNA cloning, plasmid DNA amplification, sequencing etc..

Bacterial strains	Genotype/phenotype and reference
Escherichia coli XL1-Blue:	*rec*A1, *end*A1, *gyr*A96, *thi*-1, *hsd*R17, *supE44* lac [F *pro*AB, *lac*IqZΔM15, Tn10(tet^R)],*rel*A1; (Stratagene, La Jolla, CA).
Escherichia coli DH5α:	F⁻, φ80d/*lac*ZΔM15, *rec*A1, *end*A1, *gyr*A96, t*hi-1*, *hsd*R17(rK-, mK+), *sup*E44, *rel*A1, *deoR*, Δ(*lac*ZYA*arg*F) U169; (Grant et al, 1990)
Agrobacterium tumefaciens:	GV2260 (Deblare et al., 1985)

1.3.1.3. Enzymes, markers, antibiotics

Enzymes

EcoRI, T4 DNA ligase, *pfu* DNA polymerase, Dream Taq DNA polymerase, Shrimp Alkaline Phosphatase (SAP), Platinum Taq polymerase, RNase inhibitor, Reverse transcriptase, 50X advantage ®2 DNA polymerase mix (Fermentas, Vilnius, Lithuania); DNase I, RNase I (Roche, Germany),

Markers

DNA Smart Ladder (Eurogentec, Seraing, Belgium); GeneRuler™ 1kb DNA Ladder Plus, PageRuler™ Prestained Protein Ladder (Fermentas, Vilnius, Lithuania).

Antibiotics

Ampicillin, Kanamycin, Rifampicin, Spectinomycin (Duchefa, Netherlands).

Other chemicals

X-gal (5-bromo-4-chloro-3-indolyl-ß-D-galactoside) and IPTG (Roche, Germany); Murashige-Skoog (MS) medium basal salt mixture including vitamins and microelements (Duchefa, The Netherlands); sucrose, glucose, malachite green, fuchsin acid, orange G, chloral hydrate, glacial acetic acid, nonidet P-40, DMSO, PIPES, EGTA, DAPI, formalin, sodium chloride, magnesium sulphate heptahydrate, potassium chloride, tris-base, yeast extract, glycerol, glycine (Carl Roth, Germany).

1.3.1.4. Commercial kits

GeneJET plasmid miniprep kit, GeneJET gel extraction kit, RevertAid first strand cDNA synthesis kit, DNA labelling kit (Fermentas, Vilnius, Lithuania); RNeasy kit, DNeasy plant mini kit, Epitect bisulfite kit, QIAquick PCR purification kit, QIAquick gel extraction kit, Qiagen plasmid purification kit mini (Qiagen, Hilden, Germany); TA cloning® kit, Zero Blunt® TOPO Cloning kit, BD SMART RACE cDNA Amplification kit (Takara, Japan).

1.3.1.5. Vectors

Various vectors were used for DNA amplification, cloning genes into plants and other purposes.

Vector	Features	Reference or source
pCR®II	Ampicillin^r, Kanamycin^r	Invitrogen, Carlsbad, CA
pCR®4Blunt-TOPO	Ampicillin^r, Kanamycin^r	Invitrogen, Carlsbad, CA
pDONR^TM/Zeo	Kanamycin^r, Zeocin^r	Invitrogen, Carlsbad, CA
pGV2260	Rifampicin, Spectrinomycin	Invitrogen

1.3.1.6. Primers and oligonucleotides

Primers for PCR and sequencing

Primer name	Sequence 5'-3'	Tm(ºC)
ET1-1HuF	AAG AGA GAC GAC TAC ATT CGA ACT AAT C	68
ET1-1HuR	AGT ACC ATC TTC TAG TAA GAC TCC ACA AG	66
ET1-5F	CATCGC CTA TCA AGT ATC AGC TTC CC	68
ET1-RACE1	AGG AGT AGT CCG CAA AAG TCT TGC GA	68
ET1-RACE2	GGG TTT ACG CAG AAA CAT AGA TCG GGC	72
ET2-3HuF	AAT ACC CGA TGA ACA GAT TTA CAT ATT	63
ET2-3HuR	GAG GTA AGT TCT GGA CTC TGT ATC TAC C	69
XR2	TGG GAA AAC CTG GCG TTA CCC AAC TTA AT	69
ET3F	GGA ATG AGA ATC ACC TAA CCT CTG C	66

Primer name	Sequence 5'-3'	Tm(ºC)
ET3R	CTA CAC ATT GTC CGA CAT ATA CAC C	64
LBa1	TGGTTCACGTAGTGGGCCATCG	66
XR2-LB	CAT TTT ATA ATA ACG CTG CGG ACA TCT AC	66
LBb1	GCG TGG ACC GCT TGC TGC AAC T	68
Rba3	CGG CTT GTC CCG CGT CAT C	64
8409 – LB	ATA TTG ACC ATC ATA CTC ATT GC	57

Primers for RT - PCR

Primer name	Sequence 5'-3'	Tm(ºC)
ET1–RACE 1	AGG AGT AGT CCG CAA AAG TCT TGC GA	68
ET1–RACE 2	GGG TTT ACG CAG AAA CAT AGA TCG GGC	72
ET2–RT–ACF	ATG GAA TTC GGC GAC GGC GTT TCC TTC G	73
ET2–RT–ACR	CTC GGA CTT TGG CGG TGT CTG TTT TTC G	72
AP3 F	CTA ACA CCA CAA CGA AGG AGA TC	63
AP3 R	GAA GGT AAT GAT GTC AGA GGC AG	63
ACT2-F	TCG GTG GTT CCA TTC TTG CT	57
ACT2-R	GCT TTT TAA GCC TTT GAT CTT GAG AG	55

1.3.1.7. Solutions and buffers

10 x TAE buffer	
Tris-base	242 g
H3BO3	57,1 ml
EDTA	100 ml
Distilled water	up to 1000 ml
Extraction buffer for plant genomic DNA	
Tris-HCl pH 7.5	0.20 M
NaCl	0.25 M
EDTA pH 8.0	25 mM
SDS	1%
Alexander staining solution	
Ethanol 95%	10 ml
Malachite green solution (1% in 95% ethanol)	1 ml
Fuchsin acid (1% in water)	5 ml
Orange G (1% in water)	0.5 ml

22

Phenol	5 g
Chloral hydrate	5 g
Glycerol	25 ml
Glacial acetic acid	2 ml
Distilled water	50 ml
DAPI staining solution	
Nonidet P-40	0.01%
DMSO	10%
PIPES	50 mM
EGTA	5 mM
DAPI	1 mg/ml
Clearing solution	
Chloral hydrate	40g
Water	10ml
Glycerol	10ml
Formalin	5ml
Luria-Bertani-Medium (LB	
Trypton	10g
Yeast extract	5g
Sodium chloride	5,8g
Mg sulphate heptahydrate	2,46g
Agar	15g
Distilled water	up to 1 l
SOC- Medium	
Trypton	1g
Yeast extract	0,5g
5M NaCl	200µl
KCl 1M	250µl
Distilled water	up to 1 l
Rich medium for Arabidopsis	
MS salt mixture	4,3g
Sucrose	10g
Vitamin solution	10ml
Agar (0,8%; for plates)	8g
Distilled water	up to 1 l, pH 5,8

1.3.2. Methods

1.3.2.1. Extraction of genomic DNA

Genomic DNA extraction from plants was performed according to Edwards *et al*., 1991. Leaf tissue (~100 mg) was ground in liquid nitrogen into fine powder and suspended in 400 µl of extraction buffer. The suspension was centrifuged for 10 min at 13.000 rpm in a microcentrifuge. The supernatant was collected into a new tube and the DNA was precipitated with an equal volume of isopropanol. DNA was collected by centrifugation for 10 minutes, washed in 70% ethanol, dried and resuspended in 50 µl TE buffer. DNA concentration was determined by Nanodrop® ND-1000 spectrophotometer (NanoDrop Technologies Inc., USA).

1.3.2.2. Screening and verifying T-DNA insertion mutants

T-DNA insertion lines *et1-1; et1-5; et2-3; et3-3; et3-2* in Columbia (Col) background were isolated from the Salk Institute collection of T-DNA lines transformed with pROK2 (http://signal.salk.edu/cgi-bin/tdnaexpress). The T-DNA specific primer LBa1 (O'Malley *et al.*, 2007) was used in combination with either forward or reversed gene specific primers. The line *et2-1* was isolated from the collection of the *Arabidopsis* Knock-out Facility (AKF) at the University of Wisconsin Biotechnology center, following a pool screening for insertion in *AtET2* gene in the *Wassilewskija* (Ws) background. The population lines were transformed with the T-DNA vector pD991-AP3 (Krysan *et al.*, 1999). The presence of T-DNA was verified by PCR using T-DNA right border XR2 primer (Zhao *et al.*, 2002; Ivanov *et al.*, 2008) in combination with a gene specific primer.

The primer combinations were as follows: Wild type ET1: ET1-1HUF/ET1-1HUR; T-DNA *et1-1*: ET1-1HuF/LBa1; Wild type ET1: ET1-5F/ET1-RACE1; T-DNA *et1-5*: ET1-RACE1/LBa1; Wild type ET2: ET2_RT_ACF/ET2_RT_ACR; T-DNA *et2-1*: ET2_RT_ACR/XR2; Wild type ET2: ET2-3HUF/ET2-3HUR; T-DNA *et2-3*: ET2-3HUR/LBa1; Wild type ET3: ET3F/ET3R; T-DNA *et3-2*: ET3R/8409-LB; Wild type ET3: ET3F/ET3R; T-DNA *et3-3*: ET3R/8409-LB.

PCR conditions: 95°C in 5min; (95 °C: 30s, 60°C: 30s, 72°C: 40s) repeat in 40 cycles, 72 °C in 5min.

1.3.2.3. Cloning methods and sequencing

Basic molecular methods such as enzymatic digestion, DNA ligation, DNA gel electrophoreses were performed according to standard protocols (Sambrook and Russell, 2001). DNA fragments were isolated and purified from agarose gel by QIAquick gel extraction kit (Qiagen, Hilden, Germany) and GeneJET gel extraction kit (Fermentas, Vilnius, Lithuania). DNA sequences were determined at the Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK, Gatersleben, Germany) or commercially by MWG Biotech Company (Ebersberg, Germany). Plasmid extractions and purifications were done using Qiagen Plasmid kit and Fermentas GeneJET plasmid miniprep kit according to the protocol recommended by the manufactures.

1.3.2.4. cDNA synthesis and RT-PCR

First strand cDNA was synthesized by reverse transcription from total RNA using Revert Aid H Minus First strand cDNA synthesis kit (Fermentas, Vilnius, Lithuania). 1 µg of total RNA and 1 µl oligo (dT) primer were added to each tube to obtain a total volume of 11 µl. Priming was carried out at 70 ºC for 5 minutes, then 1 µl of ribonuclease inhibitor (20 units/µl), 2 µl of 10 mM dNTP mix and 4 µl of 5X RT buffer were added to each reaction tube. The reaction mixture was incubated at 37 ºC for 5 minutes and 1 µl of Reverse transcriptase (200units/µl) was added. The reaction mix was incubated for 1 h at 42 ºC for an hour, heated for 10 minutes at 70 ºC and stored at -20 °C for further uses.

Reverse-transcription PCR

RT-PCR reaction to measure transcript amounts was performed using the primers ET1-1HUF/ET1-1HUR for ET1 transcript, ET2_RT_ACF/ET_RT_ACF for ET2 transcript.

RT-PCR conditions*:* ET1, ET2*:* 95 °C in 5min, (95°C: 30s, 60°C: 30s, 72°C: 1min), 72°C in 5 min.

1.3.2.5. Generating and finding double mutants alleles et1-1/et2-1 and et1-1/et2-3

Cross the homozygous mutants *et1-1* to *et2-1* and cross the homozygous *et1-1* and *et2-3.* Allow the F1 to self, and plant the F2 seed. Using molecular markers to find the double homozygote.

1.3.2.6. Seed germination and premature seed germination.

Seeds were collected from desiccated siliques and kept for one month in a dark and dry place. Seed were surface sterilized and spread on petri dishes with MS-agar. Germination rate was determined after 1 day. Premature seeds were collected from green siliques and grown on plates containing Murashige and Scoog medium (MS, Duchefa). Germination rates were determined for up to 12 days.

1.3.2.7. Generation of transgenic lines with central cell-specific marker

attB PCR conditions

Initial denaturation: 94 °C, 2 min

35 cycles
----
94 °C, 30 s
61 °C, 30 s
72 °C, 1 min

Final extension: 72 °C, 5 min

The fragments were cut and purified by Qiaquick kit and used for BP reaction.

The BP recombination reaction was performed as follows: 3µl attB-PCR product, 1 µl donor vector (pDONR/Zeo), 2 µl BP clonase II enzyme, 4 µl TE buffer, pH 8. The reaction was kept at room temperature overnight and then transformed into DH5α. Plasmid DNA was purified by Qiaquick kit and and resequenced. The LR recombination reaction was performed to transfer the gene of interest into an attR-containing destination vector to create an attB-containing expression clone. LR reaction conditions: 1.5 µl entry clone, 1.5 µl destination vector (pBGW), 4 µl 5X LR clonase reaction buffer, 13 µl TE buffer pH 8; incubation at 25 ºC for 1 h, addition of 2 µl of 2 µg/µl proteinase K, incubation at 37 ºC for 10 minutes and transformation of *E. coli*. Selected plasmid clones were purified and resequenced. Finally, the destination vector was transformed into the *Agrobacterium* strain GV2260. The culture was grown overnight at 28ºC for 2 days in YEB medium containing rifampicin, spectinomycin and carbenicillin. A stock culture was kept with glycerin 60% in -80ºC before transformation into Arabidopsis.

1.3.2.8. Pollen staining by Alexander

Inflorescences were collected from adult plants and fixed for 1-3 hours at 4 °C in acetic acid:ethanol (1:3). Anthers of mature flowers were isolated, transferred to a slide with a drop of Alexander solution (Alexander, 1969). Stained pollens were visualized under a Zeiss Axioplan 2 microscope to check pollen viability.

1.3.2.9. Pollen staining by DAPI

Analysis of mature pollen with DAPI was performed as previously described (Park et al., 1998). 5-10 flowers were incubated in 200 µl DAPI solution overnight at 4 °C and examined by UV epi-illumination using a Zeiss Axioplan 2 microscope (Zeiss, Jena, Germany).

1.3.2.10. Clearing

Various plant tissues were collected and fixed in acetic acid:ethanol (1:3) at 4 °C overnight, dehydrated in an ethanol series (90%; 80%; 70%; 30 min per step) and stored overnight in 70% ethanol at 4 ºC. Ethanol was replaced with clearing solution. After 3 day at 4 °C tissue was observed using a Zeiss Axioplan 2 microscope (Zeiss, Jena, Germany).

1.3.2.11. Raster electron microscopy (REM)

Preparation and analysis of samples using REM were performed in cooperation with Dr. T. Rutten (Structural Cell Biology, IPK). Isolated flowers fixed overnight with 4% formaldehyde in 50 mM phosphate buffer pH 7,0. After washing with buffer and dehydration in an ethanol series, samples were critical point dried in a Bal-Tec critical point dryer (Bal-Tec AG, Balzers, Switzerland). Dried specimens were attached onto aluminium sample blocks and gold coated in an Edwards S150B sputter coater (Edwards

High Vacuum Inc., Crowley, West Sussex, UK). Spikes were examined in a Hitachi S4100 SEM (Hisco Europe, Ratingen, Germany) at 5 kV acceleration voltage. Digital recordings were made and saved as Tif-files.

1.3.2.12. Confocal laser-scanning microscopy

Flowers of plants were emasculated, and whole-mount preparations of ovules were analyzed by microscopy 48h after emasculation. CFP fluorescence signal was studied with a Zeiss LSM 510 META or LSM780 confocal laser-scanning microscope (Zeiss, Jena, Germany). Fluorophore was detected with a 458 nm laser line in combination with a 480-520 nm band-pass (CFP). Identity of fluorophores was confirmed by photo spectrometric analysis with the help of the META-detector. This work has been performed in collaboration with Dr. T. Rutten, IPK Gatersleben.

1.4. RESULTS

1.4.1. Characterisation of et mutants

1.4.1.1. Mutants in AtET1

The mutant lines in *AtET1* were obtained from the SIGNAL T-DNA collection (http://signal.salk.edu/cgi-bin/tdnaexpress) and called *et1-1* (SALK_000422) and *et1-5* (SALK_006710). And then, these plants for each mutant line were verified in the first generation to determine the line was heterozygous or homozygous. Genomic DNA was isolated and used for PCR with gene specific primers in combination with the T-DNA-specific primer LBa1. The sizes of PCR products were determined to be 655 bp and 1047 bp for *et1-1* and *et1-5,* respectively.
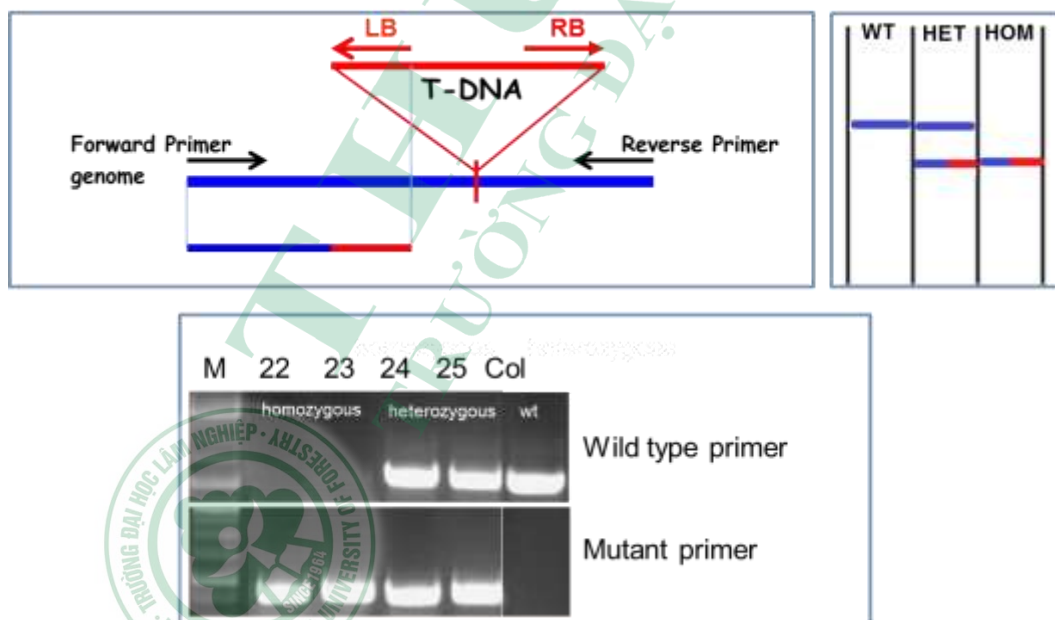

Fig.1.11. Detection of the T-DNA insertion in et1-1.

Upper panel: Principal strategy for T-DNA detection using a gene specific primer for the left T-DNA border (LB) and two gene specific primers to detect the wild type allele. *Lower panel:* PCR analysis of two homozygous *et1-1* lines (22, 23), two heterozygous *et1-1* lines (24, 25) and wild type (Col) using ET1-1HUF and ET1-1HUR as gene specific primers and ET1-1HUF and LBa1 to detect the T-DNA insertion.

By using SMART-RACE technique to amplify and sequence the 5´-terminal part of the transcript ET1 we suggest that the database predicted gene model of *AtET1* needs to be corrected. The results of the sequence of the RACE amplified fragment show that the gene start ET1 needs to be shifted as shown in Fig.1.3. Based on this new gene model, the T-DNA insertion in *et1-5* allele is now located far up in the 5`flanking region, 699 bp in front of the translation start and in the *et1-1* allele the T-DNA insertion is positioned in the second exon (Fig. 1.3).
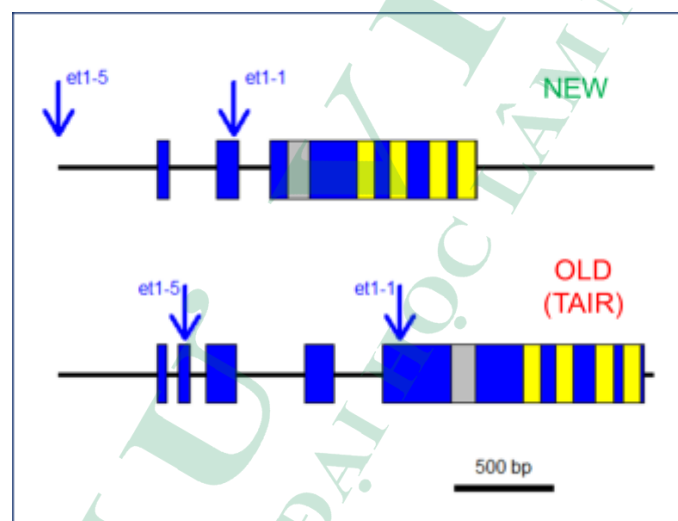
Fig.1.12. Comparision of the TAIR data base predicted gene model and the new experimentally confirmed gene model

(Bui Thi Mai Huong et al., 2021).

1.4.1.2. Mutants in AtET2

The *et2-1* mutant line was sequenced and determined by PCR analysis using gene specific primers (ET2_RT_ACF and ET2_RT_ACR) as well as the T-DNA right border primer XR2 in combination with ET2_RT_ACR. We have selected two lines the homozygous *et2-1* as shown Fig 1.13. The expected fragment length is 851 bp (Fig.1.13). The homozygous *et2-3* lines (line 1, line 2, line 3) also were identified by PCR analysis using the specific primers (ET2_RT_ACF and ET2_RT_ACR) and T-DNA primer (ET2_RT_ACF and LBa1). The expected fragment length was 1284 bp and 1369 bp for the mutant and wild type allele, respectively.
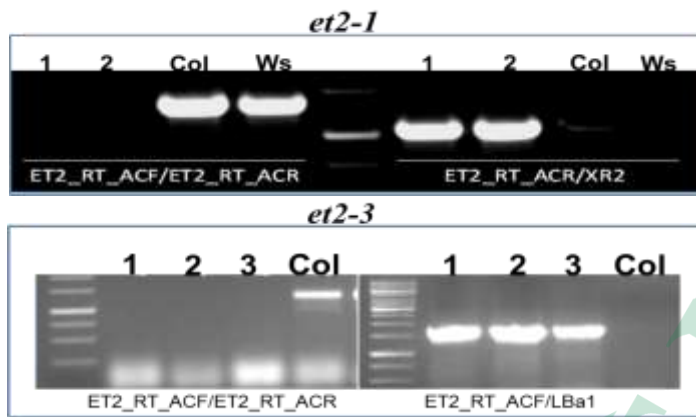
Fig.1.13. Detection of T-DNA insertions in mutants AtET2.

Upper panel: *et2-1* mutant using wild type primers (RT_ACF/RT_ACR) and mutant primers (ET2_RT_ACR/XR2) 1, 2, mutant lines; Col, Ws ecotypes Columbia and *Wassilewskija*; GM, size marker. Lower panel: *et2-3* mutant using gene specific primers ET2_RT_ACF/ET2_RT_ACR and mutant primers ET2_RT_ACF/LBa1.

1.4.1.3. Mutants in AtET3

Two *et3-3* and *et3-2* mutant lines have been identified and characterized for *AtET3*. Both mutants were shown to be homozygous. The insertions in both lines are located close to each other within the 5`-flanking gene region (Fig.1.6).

In total there are 6 mutant lines as summarise in Fig.1.14. The single mutant *et1-1* was combined with the single mutants *et2-3* and *et2-1* to generate homozygous double mutants (see below).
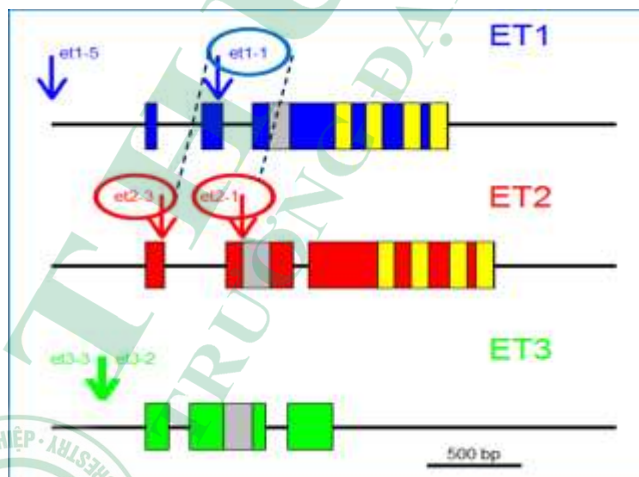
Fig.1.14. Gene model and positions of T-DNA-insertions of the ET gene family.

The positions of T-DNA insertions are indicated by arrows. ET repeats and the GIY-YIG single strand cutting domain are given in yellow and grey, respectively. The dashed lines indicate the alleles which have been combined as homozygous double mutants.

29

Tab.1.1 Nucleotide positions of 6 T-DNA insertion mutants relative to the ATG start.

Gene	Mutant	Position	Insert relative to ATG
ET1	*et1-5*	promoter	-699
	et1-1	Exon2	383
ET2	*et2-3*	Exon1	85
	et2-1	Exon2	518
ET3	*et3-3*	promoter	-239
	et3-2	promoter	-216

1.4.1.4. Loss of transcripts in single mutants et1-1, et2-1 and et2-3

RT-PCR on total RNA from the homozygous mutant plants was performed to verify the homozygous status of single mutants (*et1-1, et2-3, et2-1*) are the absence of the wild type *AtET1* and *AtET2* mRNA, respectively. And actin primers were used to control RNA quality and quantity. Whereas, The *AtET1* and *AtET2* products also be amplified from the wild type plant which demonstrates the intact of *AtET1, AtET2* mRNA (Fig.1.15). The results demonstrate that the *et1-1, et2-3* and *et2-1* can be used as suitable tools for functional studies of the ET gene family in Arabidopsis.
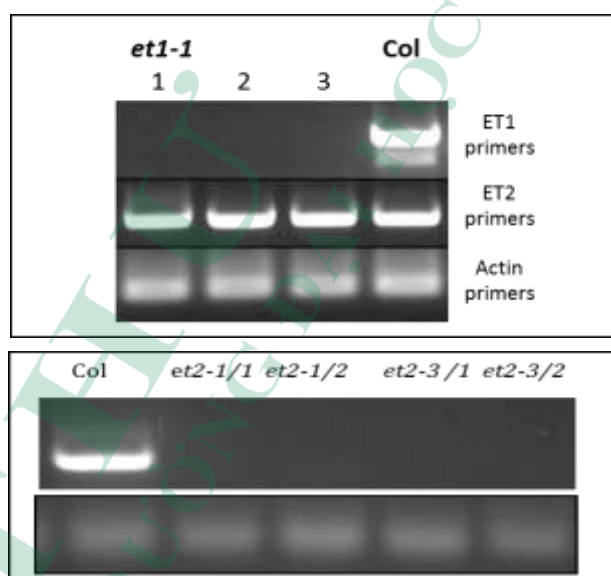
*Fig.1.15. RT-PCR to demonstrate the knock out character of
single mutants et1-1, et2-1 and et2-3.*

Upper panel: et1-1 mutant and wild type probed with *ET1* specific primers which span the insertion site show the loss of the *ET1* transcript. Priming with *ET2* specific primers demonstrate that the *ET2* transcript is not affected. *Lower panel:ET2* specific primers which span the insertion site were used to show the loss of the *ET2* transcript. Actin gene was used in both cases as loading control.

30

1.4.1.5. Gene family evolution

There are three genes ET1, ET2 and ET3 in the ET gene family in *A. thaliana*. ET1 and ET2 encode all characteristic ET sequence motifs, including the ET repeats and the GIY-YIG domain, but ET3 is lacking the C-terminal ET repeats. Therefore it is considered a nonfunctional pseudogene. Here, we only focused analysis on ET1 and ET2, because they contain ET-domain. ET genes are exclusively found in plants, therefore we suggested that their involvement in plant-specific processes. When ET genes were aligned to the common ancestor of mosses and seed plants, we identified the ET2-type gene as ancestral (Fig. 1.16). ET2 consists of three exons of which the second encodes the GIY-YIG domain and the third contains the characteristic ET repeats. ET1 probably resulted from deletion of the second intron of ET2 and an insertion of a complete exon into the first intron of ET2 (Fig. 1.17). The characteristics of ET1 gene is only found in species of the family Brassicaceae (Fig. 1.17). The evolutionary origin of ET1 might be the α-whole genome duplication event in this family (Hohmann et al., 2015).
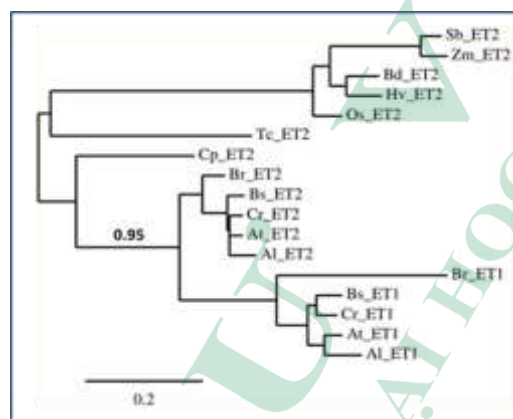
Fig.1.16. Phylogenetic tree of ET proteins and ET1-specific second exon in Brassicaceae.

Protein sequences were identified via Blastp in the Phytozome databases (phytozome.jgi.doe.gov) for *Arabidopsis lyrata* (Al), *A. thaliana* TAIR9 (At), *Boechera stricta* (Bs), *Brachypodium distachyon* (Bd), *Brassica rapa* (Br), *Capsella rubella* (Cr), *Carica papaya* (Cp), *Oryza sativa* (Os), *Sorghum bicolor* (Sb), *Theobroma cacao* (Tc) and *Zea mays* (Zm). The HRT gene (*Hordeum vulgare*; GenBank accession CAA04677), which is an ET2 type gene, was added. ET1 type genes could only be identified in Brassicaceae genomes (Al, At, Bs, Br, Cr), which evolved after the a-whole genome duplication event c. 47 million years ago (Hohmann et al., 2015), while all genomes encode ET2 types. A phylogenetic tree was calculated using the web service at www.phylogeny.fr ('one click' method with Gblocks for curation of the MUSCLE alignment; Dereeper et al., 2008). The phylogenetic tree clearly shows that ET1 types and ET2 types are sister groups within the Brassicaceae. The bootstrap value is given for the node separating Brassicaceae from other plant species and for splitting of ET1 and ET2 types of the Brassicaceae species (Francesca Tedeschi, Bui Thi Mai Huong., 2019)

Non-Brassicaceae

```
Phypat     HDRCFS-EW----------------------------------TILVGPS-DWKEQAT
Horvul     HDSLFS-PW----------------------------------KVLVGPS-DWEDHSA
Orysat     HDSLFS-PW----------------------------------KVLVGPS-DWEDHAA
Soltub     HDSAFS-DW----------------------------------KILVGPN-DWEDYLL
Fraves     HDSHFS-NW----------------------------------KVLVGPS-DWEDYWL
Poptri     HDSSFS-KW----------------------------------QLLIGPS-DWQDYFL
Vicfab     HDSSFS-EW----------------------------------KILIGPS-DWEDYSK
Cansat     HDRHFS-RW----------------------------------EILVGPS-DWEDHSL
Cucmel     HDSSFS-KW----------------------------------ETLVASS-DWEDYSL
Gosrai     HDSYFS-EW----------------------------------KLLICPH-GWKN---
Thecac     HDSQFS-KW----------------------------------KVLIGPD-DWVDYSI
Carpap     HDTCFS-KW----------------------------------KVLVGAN-DWEDYCK
```

Brassicaceae

```
Brarap_ET1 PGGVFS-RWQGFARSMLLPKPFSETAELRRTVADYSLISRGLA--PKILREAKGNREDLRV
Thepar_ET1 HDAVSS-KWQGFARSMLLRKPFSEAAELKRTLADNSLISRDLG--PKILIGAK-DREDFRK
Caprub_ET1 HDAAAISKWQGFARSMLLPISETAELRRTYADYSLISRDLG--PKIRVGAS-DKENFRK
Aratha_ET1 HDPFFS-KWQGFARSMFLRKPISETAELRKTFADYSLISRDLGPKPKILIGAN-DKENFRE
Aralyr_ET1 HDAVSS-KWQGFARSMLLRKPVSETAELRKIFADYSLISRDLG--PKILIGAS-DKENFRK
Boestr_ET1 HDAVSSKKWQGFARSMLLRKPVSETAELRKIFADYSLISRDLG--PKILIGAS-DKENFRK
Brarap_ET2 HDAVFS-KWK----------------------------------ILIGSN-DWEDFKN
Thepar_ET2 HDAVFS-KWK----------------------------------VLIGSN-DWEDFKN
Aralyr_ET2 HDAVFS-KWQ----------------------------------VLIGSN-DWEDFQT
Aratha_ET2 HDTVFS-KWQ----------------------------------VLIGSN-DWEDFKN
Caprub_ET2 HDSVFS-NWQ----------------------------------VLIGSN-DWEDFKN
Boestr_ET2 HDVVFS-KWQ----------------------------------VLIGSN-DWEDFKN
```

Fig.1.17. Amino acid sequence alignment using the MUSCLE program of various Brassicaeceae and non-Brassicaceae species

Phypat, *Physcomitrella patens*; Horvul, *H. vulgare*; Orysat, *O. sativa*; Soltub, *Solanum tuberosum*; Fraves, *Fragaria vesca*; Poptri, *Populus trichocarpa*; Vicfab, *Vicia faba*; Cansat, *Cannabis sativa*; Cucmel, *Cucumber melon*; Gosrai, *Gossypium raimondii*; Thecac, *T. cacao*; Carpap, *C. papaya*; Brarap, *B. rapa*; Thepar, *Thellungiella parvula*; Caprub, *C. rubella*; Aratha, *A. thaliana*; Aralyr, *A. lyrata*; Boestr, *B. stricta*

(Francesca Tedeschi, Bui Thi Mai Huong.,2019).

1.4.1.6. Generation of double et1/et2 mutants

AtET1 and *AtET2* are two closely related proteins. They share an overall amino acid identity of 40%, especially in the ET repeats (58%). To two double mutants have been generated by crossing the homozygous mutant *et1-1* both with homozygous *et2-1* and *et2-3* mutants. Homozygous double mutants *et1-1/et2-1* and *et1-1/et2-3* have been selected and characterized in the F2 generation (Fig.1.18).
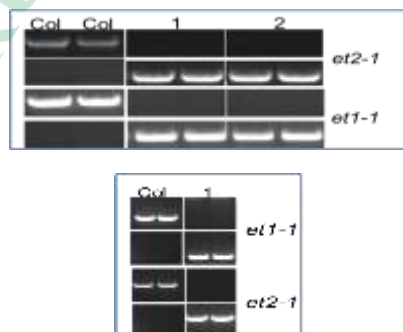

Fig.1.18. Genotyping of et1/et2 double mutants.

Upper panel: The homozyogous double mutant *et1-1/et2-1* of lines 1 and 2. *Lower panel:* The homozyogous double mutant *et1-1/et2-3* of line 1.

1.4.1.7. Phenotypic characterisation of single mutants (et1, et2) and double mutants (et1-1/et2-1; et1-1/et2-3).

The phenotype of single mutants and double mutant revealed several highly interesting observations. They include flower organ identity, gametophyte development, endosperm development, immature seed germination, pollen development and seed development.

a. Homeotic transformation of flower organs in single and double mutants.

The Phenotypic flower organs of *et* mutants reveals unusual numbers of flower organs such as sepals, petals and stamens compare to wild type (Fig. 1.19).
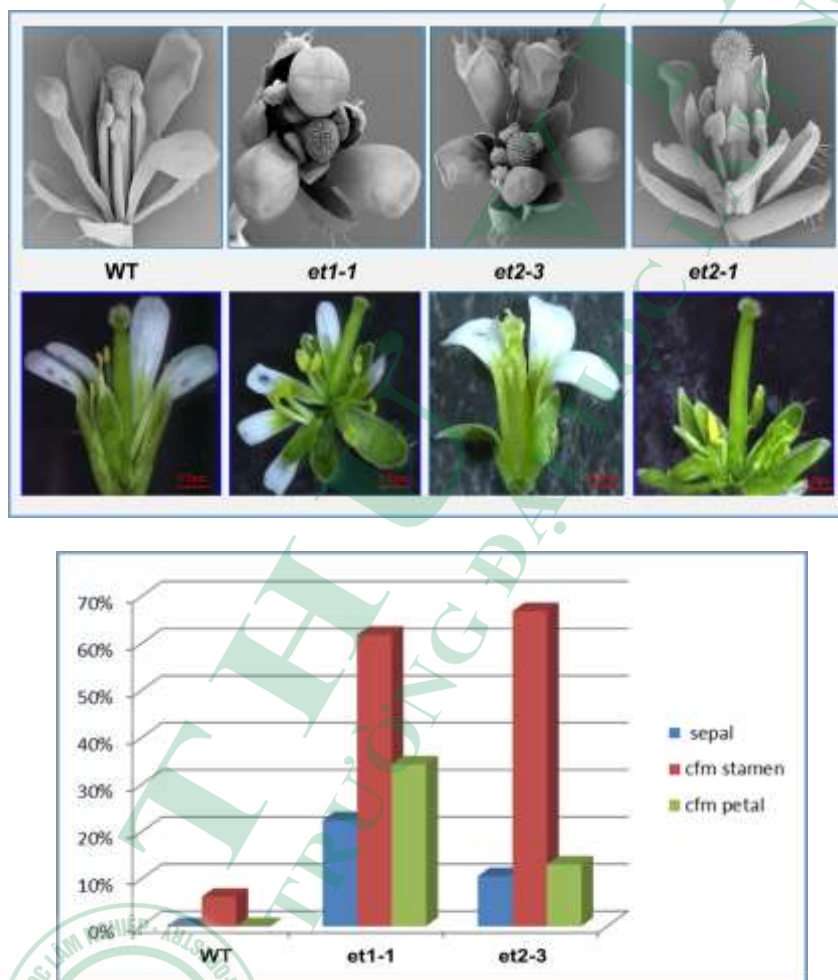

Fig.1.19. Changed numbers of flower organs in et mutants.

Upper panel: The pictures of wild type and selected mutant flowers with two petals and two sepals in *et1-1,* with three petals in *et2-3* and five petals in *et2-3*. *Lower panel:* The effect has been quantified in 160 flowers each.

Remarkably, *et2-1* mutant exhibit homoeotic transformations of anthers into carpel-like structures including the occurrence of stigma-like structures as well as ectopic ovules as shown in Fig. 1.20.
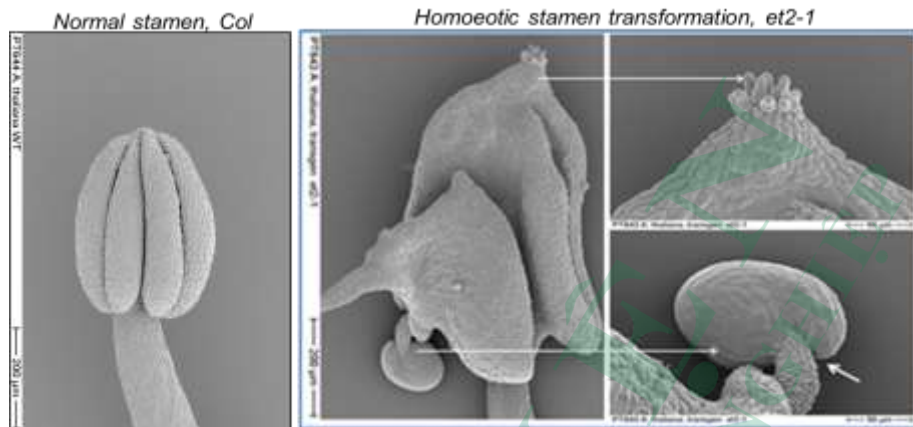

Fig.1.20. Homoeotic transformation of anthers into carpel-like structures in et2-1 mutants.

REM pictures of wild type anthers and various homoeotic transformations of anthers into carpel-like structures including stigma and ovule formation (arrows).

To further characterize the stamen-derived ovules in more detail the tissue was cleared and analyzed by DIC microscopy. As shown in Fig. 1.21 the ectopic ovules contain a nearly normal gametophyte with fully developed egg cell, synergids and fused polar nucleus of the central cell. However, the normal polarity with synergids next to the micropyle followed by egg cell and central is distorted in these ovules.
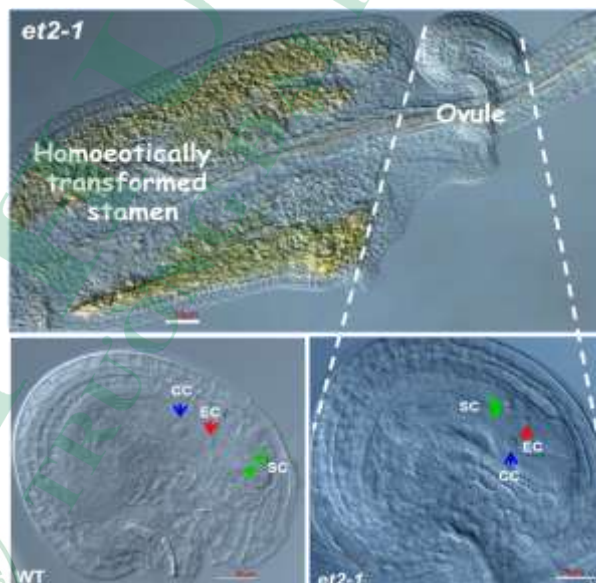

Fig.1.21. Stamen-derived ovules of the et2-1 mutant contain a fully developed gametophyte with egg cell (red arrow), two synergids (green arrows) and a homodiploid central cell nucleus (blue arrow). However, the normal polarity of the gametophytic cell types is partially distorted

The severity of the homoeotic transformation is quantitatively and qualitatively further increased in the *et1-1 et2-1* double mutant as shown in Fig. 1.22. The double mutant exhibits multiple ovule- and stigma-like structures.
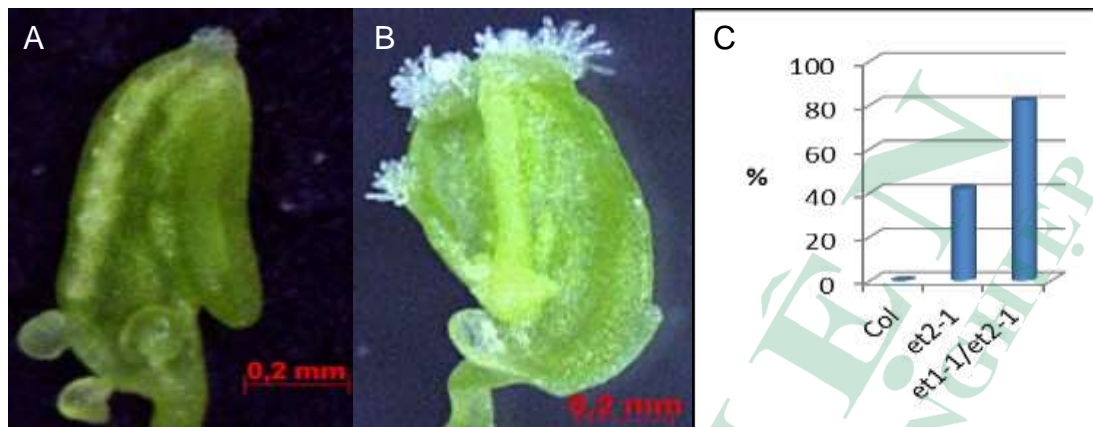

Fig.1.22. The homoeotic transformation of stamen into carpel-like structures in the et2-1 mutant (see above) is further enhanced in the et1-1 et2-1 double mutant with multiple ovule and stigma formation (A, B). The effect has been quantified in 180 flowers each (C).

b. Distortions of gametophyte development.

The Arabidopsis female gametophyte, the embryo sac, develops within the ovule consists of two synergids, one egg cell, one central cell and three antipodal with the latter degenerating at the mature stage. In wild type the homodiploid nucleus of the central cell results from the fusion of the two polar nuclei. As shown in Fig. 1.23 all *et* mutants exhibit non fused central cell nuclei. In general *et* mutant embryo sacs show up to 20% distortions such as multiple non-fused polar nuclei or missing egg cells or synergid cells.
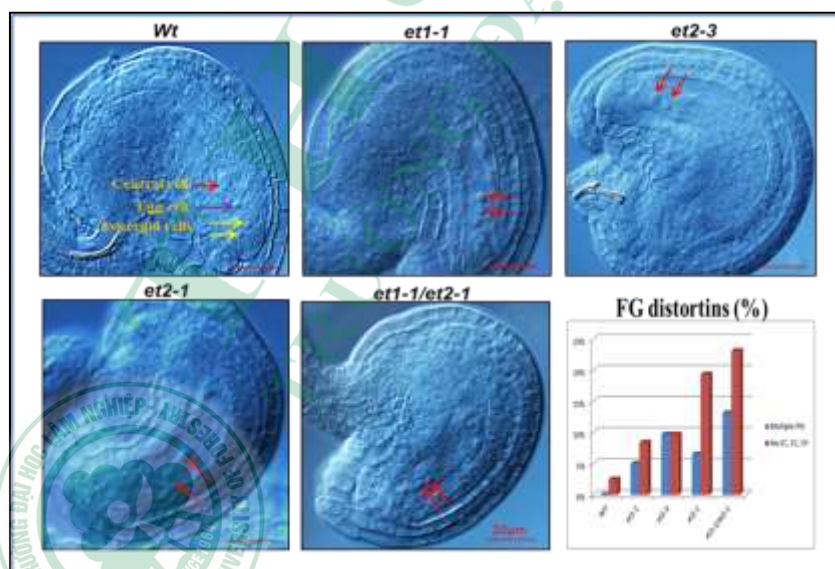

Fig.1.23. Distorted embryo sac development in et mutants. In the mature wild type embryo sac the two polar nuclei fuse.

In all et mutants this does not occur and the two polar nuclei remain non-fused (blue arrows). The gametophytic distortions have been quantified. PN, polar cell; EC, egg cell; CC, central cell, SY; synergids, Scale bar: 20 µm.

To further characterize the polar nuclei fusion, a central cell specific reporter construct has been used. The construct consists of the central cell-specific DD65 promoter controlling the AmCyan fluorescent protein gene. As shown in Fig. 1.24 the construct specifically labels the central cell in the wild type, whereas the marker signal is missing in the *et2-1* mutant.
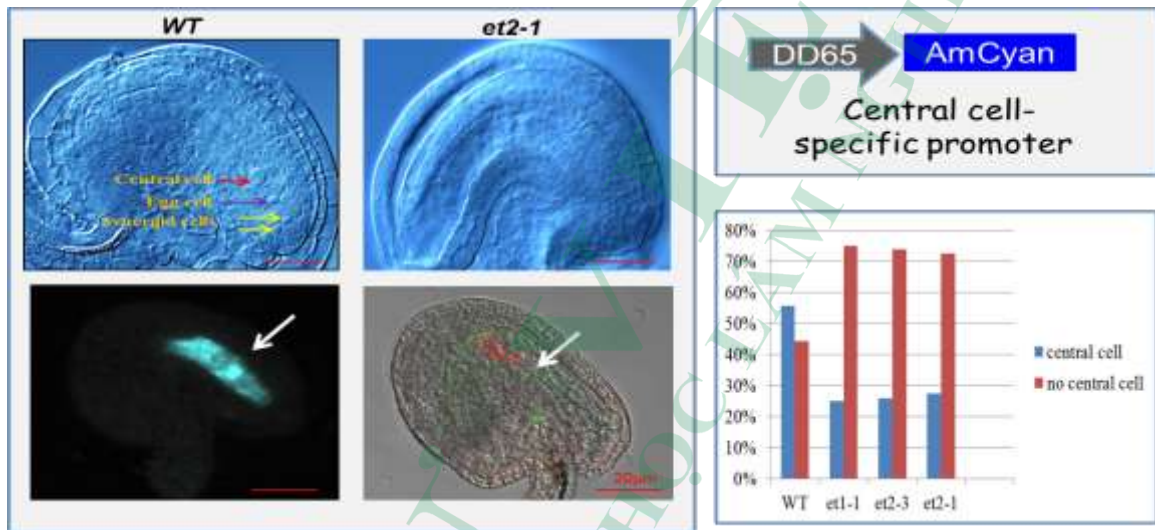
Fig.1.24. Distorted central cell differentiation in the et2-1 mutant.

Right panel: The central cell-specific promoter DD65 and AmCyan fluorescent protein gene. The gametophytic distortions have been quantified. *Left panel:* The marker line is controlled by the central cell-specific promoter DD65 and specifically labels the central cell in wild type. The signal is missing in the collapsed embryo sac of the *et2-1* mutant.

c. Distortions of endosperm development.

Another aspect of the *et* mutant phenotype concerns the endosperm differentiation. The endosperm nuclei exhibit a characteristically changed morphology with greatly enlarged nuclei in all *et* mutants, possibly indicating an enhanced synthetic activity (Fig. 1.25).

36

Fig.1.25. Affected endosperm differentiation in et mutants.

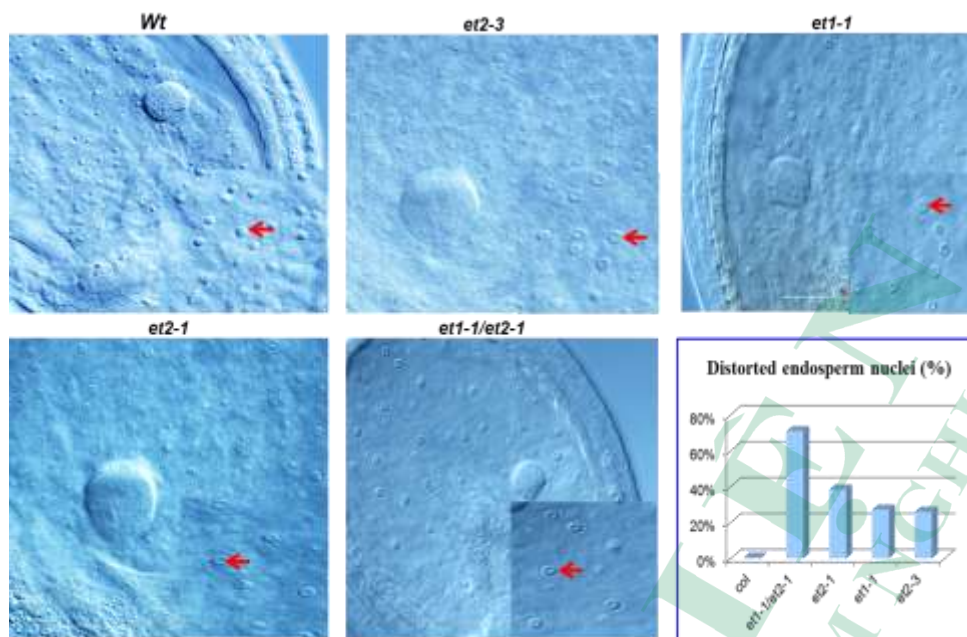The nucleoli of mutant endosperm nuclei are greatly enlarged in the *et single mutants and* double mutant in comparison to wild type. Enlarged nucleoli are considered to reflect an increased synthetic activity. The morphological effect has been quantified in the mutants (right panel), Scale bar: 20 µm.

d. Precocious germination of et mutants

Germination is initiated by the penetration of the radical through the surrounding seed coat. However, *et* mutants exhibit a strong phenotype of precocious germination. Immature seeds start to germinate already within the siliques. Remarkably, germination does not occur as in mature wild type seeds with the root tip in front, but in the mutant the seedling permeates the seed coat at the side with the cotyledons appearing first (Fig. 1.26). A similar behavior was observed when immature seeds were germinated *in vitro* (Fig. 1.27).

Fig.1.26. Precocious germination of et mutants

(Francesca Tedeschi, Bui Thi Mai Huong, 2019).

The immature seed of the *et2-3* mutant start to germinate already within the silique. The seedling permeates the seed coat at the side of the seed with the cotyledon appearing first.
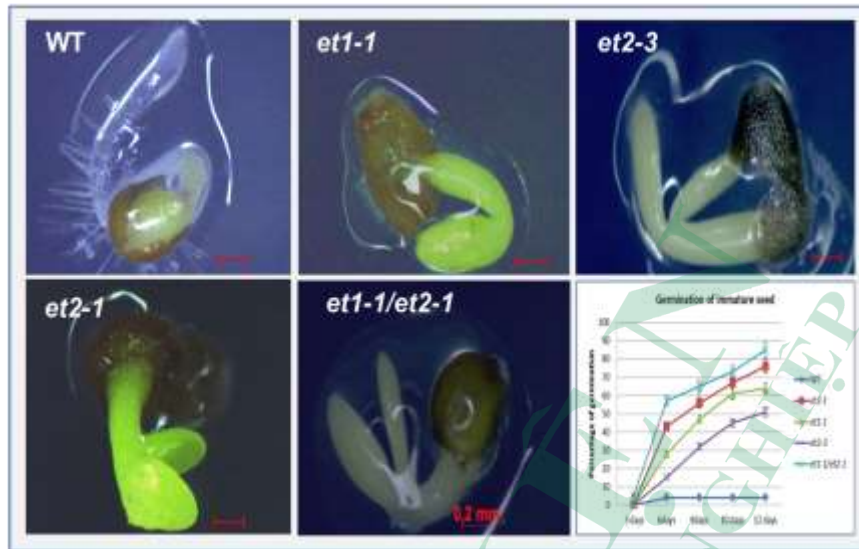
Fig.1.27. Precocious germination of et mutants in vitro.

Germinating wild type and et mutants seedlings: In wild type the radicle appears first, whereas the cotyledons show up first in the et mutants. The phenotype has been quantified (right panel) with 200 seeds each. Scale bar: 0.2 mm.

e. Distortions of pollen development.

The male gametophyte, the pollen grain or microgametophyte, develops within the anther and consists of two sperm cells encased within a vegetative cell. Pollen of wild type and *et* mutants was analyzed using DAPI staining. In wild type the vegetative nucleus and the two generative nuclei were clearly distinguishable, whereas significant numbers of abnormal and collapsed pollen grains were detected in *et* mutants (Fig. 1.28, Fig. 1.29).
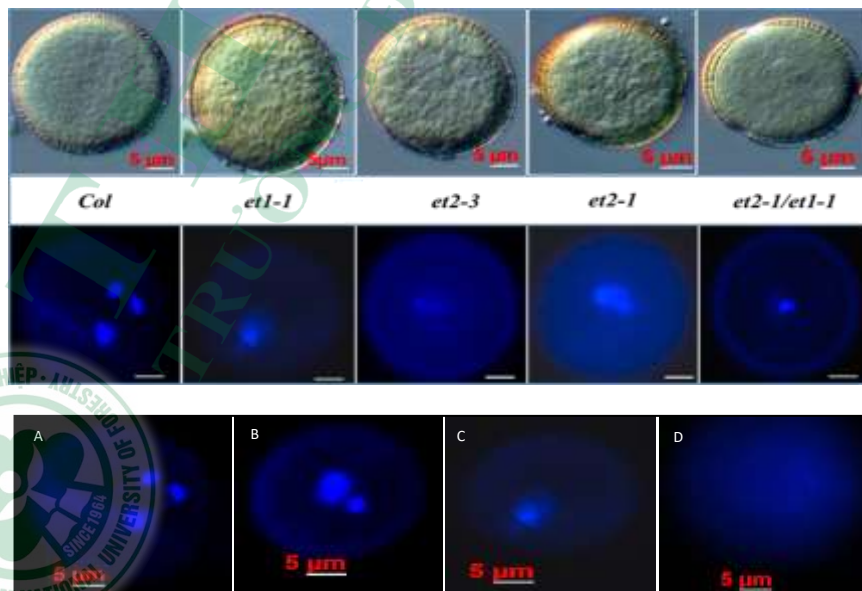
Fig.1.28. DAPI stained pollen nuclei.

Upper panel: Various distortions of pollen with one most likely vegetative nucleus in *et* single mutants and double mutant. *Lower panel:* Wild type pollen show the larger vegetative nucleus and the two generative nuclei (A). Various distortions of pollen differentiation including only one generative nucleus (B), one most likely vegetative nucleus (C) and completely collapsed pollen (D) are shown for the *et1-1* mutant.
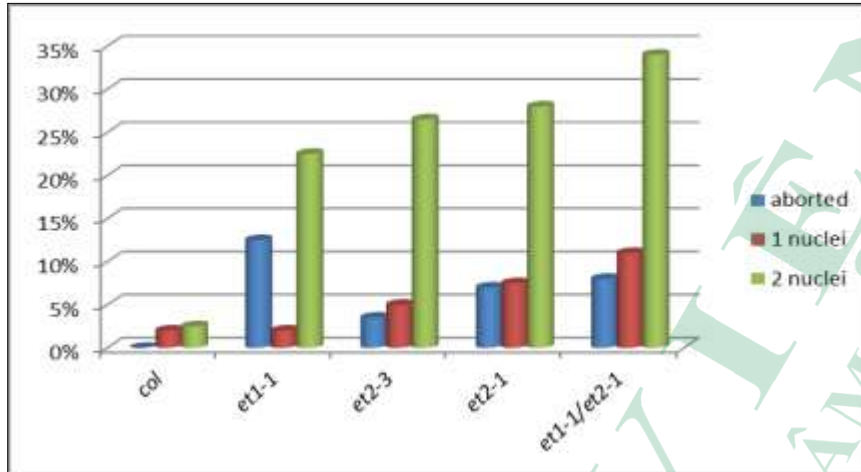
Fig.1.29. Quantification of pollen nuclei distortions in et mutants.
200 pollen each have been analyzed.

f. ET influence on seed set

As a likely consequence of the distorted embryo sac and pollen development a high degree of seed sterility is detected in *et* mutants. Approximately 80% of the *et1-1 et2-1* double mutant seeds are sterile with less than 3% sterility in wild type. The effect on seed set was less pronounced for the two single mutants *et1-1* and *et2-3* (Fig. 1.30).
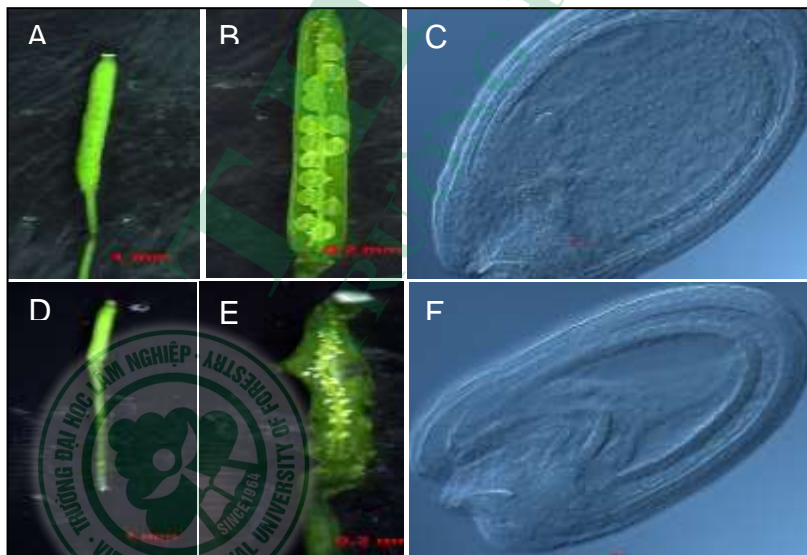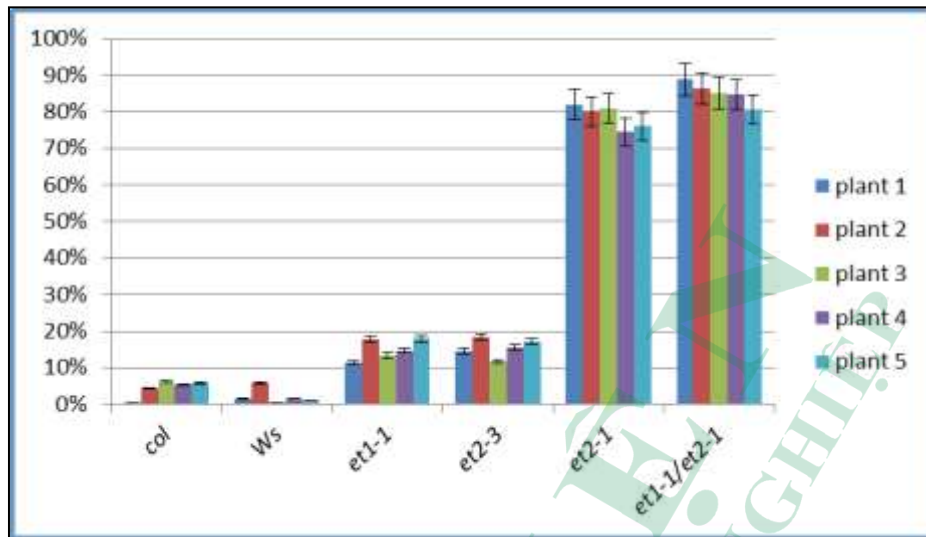
Fig.1.30. Seed set in wild type and et mutants.

Upper panel: Fertile silique of *et2-1*mutant (A, B) and early embryogenesis in cleared seeds (C); sterile silique of *et2-1*(D, E) and collapsed embryogenesis in cleared seeds (F). *Lower panel:* Quantification of sterility in wild type and *et* mutants.

1.5. DISCUSSION

EFFECTOR OF TRANSCRIPTION (ET) factors have been originally isolated as DNA-binding proteins using seed specific gene promoter sequence motifs. However, extensive interaction studies failed to identify a specific sequence motif. Ectopic expression of ET factor genes resulted in severe growth distortions including dwarf growth, late flowering, reduced germination rates, strong anthocyanin accumulation, reduced lignification etc.. Together, these observations indicated a putative function as repressors of transcription of GA-regulated genes including KNAT genes involved in cell differentiation. This is also supported by studies on a barley homologue, HORDEUM REPRESSOR OF TRANSCRIPTION (HRT), which was shown to repress the expression of amylase genes in barley aleuron cells. Remarkably, ET/HRT factors are exclusively found in plant genomes including phylogenetically old species such as the moss *Physcomitrella*. These findings strongly suggest that ETs are involved in the regulation of plant specific processes. All ET factors share a variable number of characteristic, highly conserved cystein-histidine containing ET repeats involved in zinc and DNA binding. In addition to these repeats ETs have a so called GIY-YIG domain in common. These domains are found in bacterial repair proteins such as UVRC and are known to function in the insertion of single strand cuts in DNA. Remarkably, functionally analogous domains have been recruited by plant specific regulators of DNA methylation such as DME and ROS1 leading to the speculation that ET might be also involved in the regulation of the methylation status of genomic DNA. But, despite the much more of these observations

(Wohlfarth, 1996; Raventós et al., 1998; Ellerström et al., 2005, Ivanov et al. 2008), the molecular mode of action and the principal functional importance of ET factors in plant differentiation and development remain obscure. For a more detailed functional analysis of the ET gene family, a genetic approach has been applied. Thus, T-DNA insertion mutants of Arabidopsis have been genotyped followed by detailed phenotypic and molecular analysis including deep RNA sequencing and whole genome methylation studies. Together, these data provide strong indications that ETs are novel epigenetic regulators of the DNA methylation status in plant genomes leading to pleiotropic developmental effects which include gametophytic distortions, homoeotic transformations of flower organs, affected endosperm differentiation as well as precocious ectopic germination.

1.5.1. Gametophytic cell differentiation

The typical mature gametophyte of Arabidopsis consists of two synergids, the egg cell and the homodiploid central cell with the two nuclei fused together. The most obvious gametophytic phenotype observed in *et* mutants concerns the distorted fusion of the two polar nuclei. A relatively large numbers of mutations described previously share this phenotype. Thus, defects in genes encoding BiP1 and BiP2 -molecular chaperons in the ER- exhibit non-fused polar nuclei (Maruyama et al. 2010). Polar nuclei fusion also fails in mutants with miss-specification of gametophytic cells fate, such as *clotho* and *lachesis* (Gross-Hardt et al., 2007). Remarkably, non-fused polar nuclei are also observed in mutants affecting mitochondrial genes such as *nuclear fusion defective (nfd1)* (Portereiko et al., 2006), *gametophytic factor 2 (gfa2)* (Christensen et al., 2002) and *syco1* (Kägi et al., 2010). Finally, there are two MADS-domain proteins, AGAMOUS-LIKE80 (AGL80) and AGAMOUS-LIKE61 (AGL61) which are expressed in the central cell and exhibit similar defects (Bemer et al., 2008, Steffen et al., 2008). Failure in polar nuclei fusion might be caused by mis-differentiation of gametophytic cell types, such as synergids, egg cell and central cell. The gametophytic cell differentiation is thought to be triggered by an auxin gradient, although this view was currently challenged (see below). Nevertheless, genes involved in auxin synthesis and signalling (YUCCA10 is a paternally expressed gene (PEG) and AUXIN RESPONSE FACTOR 17 is a maternally expressed gene (MEG) have been identified as targets of imprinting processes (Köhler et al., 2012). This invites the speculation that ETs mediate changes in the methylation status of such genes and could contribute to the phenotype of non-fused polar nuclei. The failure of polar nuclei fusion in *et* mutants also resembles the phenotype of mutants with defects in the gene encoding the glucose 6-phosphate translocator (GPT1). Thus, *gpt1* mutation also affects the fusion of the polar nuclei during embryo sac development (Niewiadomski et al., 2005). Interestingly, a *GPT* gene was preliminarily found to be down-regulated in *et* mutants (not shown). Currently it is not known whether the above mentioned genes anyhow interact with ET-factors – all questions which obviously need further investigations.

1.5.2. Homoeotic transformation of flower organs

A rather remarkable phenotype of *et* mutants includes the homoeotic transformation of anthers into carpel-like structures including the occurrence of ectopic ovules and stigmata. Following the predictions of the ABC model (Theißen et al., 2001) described in the introduction, this phenotypic transition of stamens into carpels is best explained by the assumption that ET-factors would either inhibit the activity of the B-function AP3/PI or enhance the C-function AG (see Fig. 2.31).
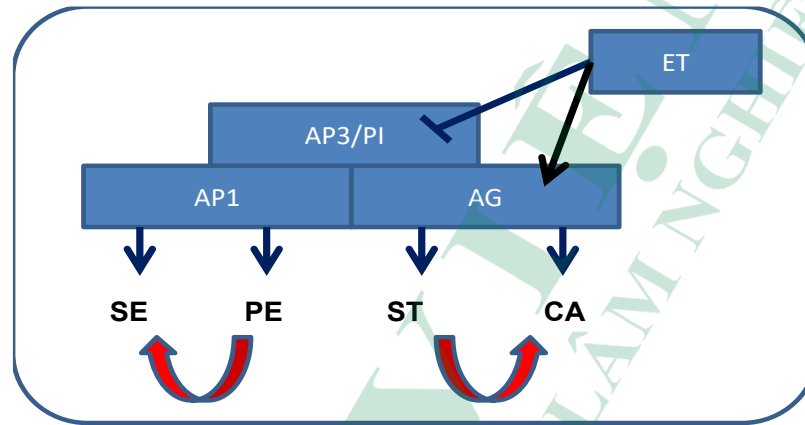

Fig.1.31. Schematic ABC model with proposed ET function to either inhibit the B-function (AP3/PI) or to enhance the C-function (AG).

Preliminary measurements of AP3/PI-transcript levels in *et* mutants indeed show a reduced level of the transcript. Further support for this view comes from the observation that the phenotype of the *ap3-3* mutant closely resembles that of *et* mutants (Krizek and Meyerowitz, 1996). Finally, the observed phenotypic differences in flower organs in *et* mutants are also similar to flower phenotypes observed in homozygous mutants (*dme-1*) of the *DEMETER* gene (Choi et al., 2002).

It was surprising to observe that the ectopic ovules on the stamen derived carpel-like structures contain a rather well developed embryo sac including two synergids, an egg cell and a fused homodiploid polar nuclei. This shows that the ovule differentiation is triggered a mainly autonomous developmental pathway including the sporophyte-gametophyte transition and gametophyte formation. However, a closer inspection reveals that the normal polarity along the embryo sac is partially distorted with the usual order of synergids, egg cell and polar nucleus being scrambled (Fig. 1.21). The differentiation of gametophytic cell fates was proposed to be determined by an auxin gradient within the embryo sac. High auxin levels would control the differentiation of synergids, a lower auxin concentration would then trigger the egg cell differentiation followed by even lower auxin levels and the occurrence of the polar nuclei (Pagnussat et al., 2009). But this view has recently been challenged (Lituiev et al. 2013) and re-challenged (Panoli et al., 2015). Although the

42

existence and relevance of an autonomous gametophytic auxin gradient is currently not clear, most authors agree on the importance of a phyothormone-mediated control of gametophytic cell fate. Since genes involved in auxin synthesis and signaling are targets of imprinting (Köhler et al., 2012), one might assume that ET-mediated methylation processes indirectly influence the gametophytic cell fate and function (see also above).

1.5.3. Endosperm differentiation

Another aspect of the *et* mutant phenotype concerns the endosperm differentiation (Fig. 2.11). The primary endosperm nucleus divides and forms a syncytium of free nuclei. At early heart stage the endosperm starts to become cellular before it degenerates during the cotyledon stages with the aleuron layer being a remnant of the endosperm in the mature seed. At a globular stage the nuclei of the early endosperm of *et* mutants contain unusually large nucleoli, known to be the site of ribosomal RNA synthesis. Enlarged nucleoli are usually connected to an increased synthetic cellular activity (Shaw and Doonan, 2005; Baker, 2013). The observation invites the speculation that in *et* mutants the early endosperm cells are precociously activated, perhaps for cell proliferation or a precocious initiation of early storage compounds synthesis or even for their early mobilisation.

1.5.4. Precocious seed germination

Finally, *et* mutants show a precocious germination phenotype both *in vivo* and *in vitro* (Fig. 2.14, Fig. 2.15). The germination starts already within the silique, a phenotype which resembles the phenomenon of pre-harvest sprouting in cereals (Gao and Ayele, 2014). Usually germination is initiated when the seedling penetrates the seed coat with the radicula ahead. In contrast, the precocious germination of *et* mutants occurs with the still green cotyledons first. The cotyledon-first-phenotype is also retained when isolated seeds are allowed to germinate *in vitro*. Similar phenotypes of germination with the cotyledons ahead have been described for ABA-immunomodulated tobacco seeds (Phillips et al., 1997). Seeds of plants which express anti-ABA scFV antibodies underwent a form of precocious germination when removed from the capsules and germinate by the emergence of the cotyledons first. Further, species of the genus *Aethionema* exhibit a seed dimorphism possibly controlled by an epigenetic pathway. One type germinates normally with the radicula coming out first and another type germinates-similar to the here described Arabidopsis *et* mutants- with the cotyledons ahead (G. Leubner, University of London, personal communication). Again one might speculate that ETs are involved in epigenetic pathway involved in the regulation of the ratio between the phytohormones GA and ABA. This view fits with the well established knowledge that the ratio of both hormones is crucial for the maintenance of dormancy *versus* initiation of seed germination, with ABA favouring dormancy and GA triggering germination. Assuming that ET acts as an epigenetic repressor of GA activity (see above), indeed one would predict an early germination phenotype for *et* mutants as observed.

1.6. CONCLUSSION

There are six single mutant lines have been genotyped in ET gene family, including *et1-1*; *et1-5*; *et2-1*; *et2-3*; *et3-2* and *et3-3*. However, we only focus on *et1-1*, *et2-1* and *et2-3* because T-DNA insertions are positioned in the exon of ET1 and ET2. Moreover, the two double mutant lines (*et1-1/et2-1*; *et1-1/et2-3*) have been generated by crossing *et1-1* with *et2-1* and *et2-3*. All the *et* single mutant lines and the double mutant lines have been analysed phenotype and we have received the pleiotropic development effects of the phenotypic *et* mutants such as gametophytic distortions, homoeotic transformations of flower organs, endosperm differentiation and precocious germination.

REFERENCES

1. Alexander, M. P. (1969). Differential staining of aborted and nonaborted pollen. *Stain Technol.* 44: 117-122.

2. Aravind, L., D. R. Walker and E. V. Koonin (1999). Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res.* 27: 1223-1242.

3. Baker, N. E. (2013). Developmental regulation of the nucleolus size during *Drosophila* eye differentiation. *PLoS One* 8, e58266.

4. Bemer, M., M. Wolters-Arts, U. Grossniklaus and G. C. Angenent (2008). The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in Arabidopsis ovules. *Plant Cell* 20: 2088-2101.

5. Berger, F. and A. Chaudhury (2009). Parental memories shape seeds. *Trends in Plant Sci.* 14: 550-556.

6. Berger, F., P. E. Grini and A. Schnittger (2006). Endosperm: An integrator of seed growth and development. *Curr. Opin. Plant Biol.* 9: 664-670.

7. Bewley, J. D., F. D. Hempel, S. McCormick and P. Zambryski (2000). Reproductive Development. In: Biochemistry and Molecular Biology of Plants. *American Society of Plant Biologists*, pp. 988–1034.

8. Black, M. (1991). Involvement of ABA in the physiology of the developing and mature seeds. In: Abscisic acid physiology and biochemistry. *Oxford Bios. Scientific Publishers*. 99-124.

9. Bowman, J. L., D. R. Smyth and E. M. Meyerowitz (1989). Genes directing flower development in Arabidopsis. *Plant Cell* 1: 37–52.

10. Brown, R. C., B. E. Lemmon, H. Nguyen and O. A. Olsen (1999). Development of endosperm in A*rabidopsis thaliana*. *Sexual Plant Reproduction* 12: 32–42.

11. Choi, Y., M. Gehring, L. Johnson, M. Hannon, J. J. Harada, R. B. Goldberg, S. E. Jacobsen and R. L. Fischer (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. *Cell* 110: 33-42.

12. Choi, Y., J. J. Harada, R. B. Goldberg and R. L. Fischer (2004). An invariant aspartic acid in the DNA glycosylase domain of DEMETER is necessary for transcriptional activation of the imprinted MEDEA gene. *Proc. Natl. Acad. Sci.* 101: 7481-7486.

13. Christensen, C. A., S. W. Gorsich, R. H. Brown, L. G. Jones, J. Brown, J. M. Shaw and G. N. Drews (2002). Mitochondrial GFA2 is required for synergid cell death in Arabidopsis. *Plant Cell* 14: 2215-2232.

14. Coen, E. S. and R. Carpenter (1993). The metamorphosis of flowers. *Plant Cell* 5: 1175–1181.

15. Debeaujon, I. and M. Koornneef (2000). Gibberellin requirement for Arabidopsis seed germination is determined both by testa characteristics and embryonic abscisic acid. *Plant Physiol.* 122: 415-424.

16. Derbyshire, V., J. C. Kowalski, J. T. Dansereau, C. R. Hauer and M. Belfort (1997). Two-domain structure of the *td* intron-encoded endonuclease I-*Te*vI correlates with the two-domain configuration of the homing site. *J. Mol. Biol.*265: 494-506.

17. Ellerström, M., W. Reidt, R. Ivanov, J. Tiedemann, M. Melzer, A. Tewes, T. Moritz, H. P. Mock, F. Sitbon, L. Rask and H. Bäumlein (2005). Ectopic expression of EFFECTOR OF TRANSCRIPTION perturbs gibberellin-mediated plant developmental processes. *Plant Mol. Biol.*59: 663-681.

18. Focks, N. and C. Benning (1998). Wrinkled1: A novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant Physiol.* 118: 91-101.

19. Friedberg, E. C., G. C. Walker and W. Siede (1995). Nucleotide excision repair in prokaryotes. DNA Repair and Mutagenesis. ASM Press, Washington, DC.

20. Gao, F. and B. T. Ayele (2014). Functional genomics of seed dormancy in wheat: Advances and prospects. *Frontiers in Plant Science* 5: 1-11.

21. Goldberg, R. B., G. Paiva and R. Yadegari (1994). Plant embryogenesis: Zygote to seed. *Science* 266: 605–614.

22. Gong, Z., T. Morales-Ruiz, R. R. Ariza, T. Roldán-Arjona, L. David and J. K. Zhu (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell* 111: 803-814.

23. Gross-Hardt, R., C. Kägi, N. Baumann, J. M. Moore, R. Baskar, W. B. Gagliano, G. Jürgens and U. Grossniklaus (2007). LACHESIS restricts gametic cell fate in the female gametophyte of Arabidopsis. *PLoS Biol.* 5: e47.

24. Grossniklaus, U. and K. Schneitz (1998). Genetic and molecular control of ovule development and megagametogenesis. *Seminars in Cell & Devl. Biol.* 9: 227-238.

25. Grossniklaus, U., C. Spillane, D. R. Page and C. Köhler (2001). Genomic imprinting and seed development: Endosperm formation with and without sex. *Curr. Opin. Plant Biol.*4: 21-27.

26. Haig, D. (1990). New perspectives on the angiosperm female gametophyte. *Bot. Rev.* 56: 236–274.

27. Haince, J. F., M. Rouleau and G. G. Poirier (2006). Gene expression needs a break to unwind before carrying on. *Science* 312: 1752-1753.

46

28. Harada, J. J. (2001). Role of Arabidopsis LEAFY COTYLEDON genes in seed development. *Plant Physiol.*158: 405-409.

29. Heard, E. and R. A. Martienssen (2014).Transgenerational epigenetic inheritance: Myths and mechanisms. *Cell* 157: 95-109

30. Herman, E. M. (1995). Cell and molecular biology of seed oil bodies. In: Seed development and germination. *New York*. Pp: 195-214.

31. Hofmeister, W. (1851) Entfaltung und Fruchtbildung höherer Kryptogamen (Moose, Farn, Equisetaceen, Rhizocarpeen und Lycopodiaceen) und die Samenbildung der Coniferen. Vergleichende Untersuchungen der Keimung. Leipzig.

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

33. Honys, D. and D. Twell(2004).Transcriptome analysis of haploid male gametophyte developmentin Arabidopsis.*Genome Biology* 5:R85.1-R85.13.

34. Huang, K. H., C. C. Akoh and M. C. Erickson (1994). Enzymatic modification of melon seed oil: Incorporation of eicosapentaenic acid. *Journal of Agricultural and Food Chemistry* 42: 2646-2648.

35. Huang, B.Q. and S. D. Russell (1992). Female germ unit: Organization, isolation, and function. *Int. Rev. Cytol.* 140: 233–292.

36. Ivanov, R., J. Tiedemann, A. Czihal, A. Schallau, L. H. Diep, H. P. Mock, C. Bernhard, A. Tewes and H. Bäumlein (2008). EFFECTOR OF TRANSCRIPTION2 is involved in xylem differentiation and includes a functional DNA single strand cutting domain. *Developmental Biology* 313: 93-106.

37. Ivanov, R. (2005). Molecular characterization of the EFFECTOR OF TRANSCRIPITION (ET) gene family in Arabidopsisand its role in plant development. PhD thesis, Halle, Germany, Martin-Luther-Universität Halle-Wittenberg.

38. Ju, B. G., V. V. Lunyak, V. Perissi, I. Garcia-Bassets, D. W. Rose, C. K. Glass and M. G. Rosenfeld (2006). A topoisomerase IIß-mediated dsDNA break required for regulated transcription. *Science* 312: 1798-1802.

39. Junker, A., A. Hartmann, F. Schreiber and H. Bäumlein (2010). An engineer´s view on seed development. *Trends in Plant Sci.* 15: 303-307.

40. Kawashima, T. and F.Berger (2014). Epigenetic reprogramming during male gametogenesis. *Nature Reviews Genetics* 15: 613–624.

41. Kägi, C., N. Baumann, N. Nielsen, Y. D. Stierhof and R. Groß-Hardt (2010). The gametic central cell of Arabidopsis determines the lifespan of adjacent accessory cells. *Proc.Natl. Acad. Sci.* 107: 22350-22355.

42. Kim, Y. C., M. Nakajima, A. Nakayama and I. Yamaguchi (2005). Contribution of gibberellins to the formation of *Arabidopsis* seed coat through starch degradation. *Plant Cell Physiol.* 46: 1317-1325.

43. Koornneef, M., K. Leon-Kloosterziel, S. Schwartz and J. Zeevaart (1998).The genetic and molecular dissection of abscisic acid biosynthesis and signal transduction in Arabidopsis. *Plant Physiol. Biochem.*36:83-89.

44. Köhler, C., P. Wolff and C. Spillane (2012). Epigenetic mechanisms underlying genomic imprinting in plants. *Annu. Rev. Plant Biol.* 63: 331-352.

45. Krizek, B. A. and E. M. Meyerowitz (1996). Mapping the protein regions responsible for the functional specificities of the Arabidopsis MADS domain organ-identity proteins. *Proc. Natl. Acad. Sci.* 93: 4063-4070.

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

47. Latchman, D. S. (2008). Eukaryotic transcription factors. Academic Press, Fifth Edition.

48. Laux, T. and G. Jurgens (1997). Embryogenesis: A New Start in Life. *Plant Cell* 9: 989–1000.

49. Lituiev, D. S., G. K. Krohn, B. Müller, D. Jackson, B. Hellriegel, T. Dresselhaus and U. Grossniklaus (2013). Theoretical and experimental evidence indicates that there is no detectable auxin gradient in the angiosperm female gametophyte. *Development* 140: 4544-4553.

50. Lotan, T., M. Ohto, K. M. Yee, M. A. L. West, R. Lo, R. W. Kwong, K. Yamagishi, R. L. Fischer, R. B. Goldberg and J. J. Harada (1998). Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93: 1195-1205.

51. Maheshwari, P. (1950). An Introduction to the Embryology of Angiosperms. (New York: McGraw-Hill).

52. Magnani, E., K. Sjölander and S. Hake (2004). From Endonucleases to transcription factors: Evolution of the AP2 DNA binding domain in plants. *Plant Cell* 16: 2265–2277.

53. Mansfield, S. G. and L. G. Briarty (1992). Cotyledon cell development in *Arabidopsis thaliana* during reserve deposition. *Canadian Journal of Botany-Revue Canadienne De Botanique* 70: 151-164.

54. Maruyama, D., T. Endo and S. Nishikawa (2010). BiP-mediated polar nuclei fusion is essential for the regulation of endosperm nuclei proliferation in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* 107: 1684-1689.

55. Moolenaar, G. F., R. S. Uiterkamp, D. A. Zwijnenburg and N. Goosen (1998). The C-terminal region of the *Escherichia coli* UvrC protein, which is homologous to the C-terminal region of the human ERCC1 protein, is involved in DNA binding and 5´-incision. *Nucleic Acids Res.*26: 462-468.

56. Morales-Ruiz, T., A. P. Ortega-Galisteo, M. I. Ponferrada-Marín, M. I. Martínez-Macías, R. R. Ariza and T. Roldán-Arjona (2006). DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proc. Natl. Acad. Sci.* 103: 6853-6858.

57. Murphy, D. J. (1993). Structure, function and biogenesis of storage lipid bodies and oleosins in plants. *Prog Lipid Res.* 32: 247-280.

58. Müntz, K. (1998). Deposition of storage proteins. *Plant Molecular Biology* 38: 77-99.

59. Niewiadomski, P., S. Knappe, S. Geimer, K. Fischer, B. Schulz, U. S. Unte, M. G. Rosso, P. Ache, U. Flügge and A. Schneider (2005). The Arabidopsis plastidic glucose 6-phosphate/phosphate translocator GPT1 is essential for pollen maturation and embryo sac development. *Plant Cell* 17: 760–775.

60. Olszewski, N., T. P. Sun and F.Gubler (2002). Gibberellin signaling: biosynthesis, catabolism, and response pathways. *Plant Cell* 14: S61–S80.

61. Park, S. K., R. Howden and D.Twell (1998). The *Arabidopsis thaliana* gametophytic mutation *gemini pollen1* disrupts microspore polarity, division asymmetry and pollen cell fate. *Development* 125: 3789-3799.

62. Pagnussat, G. C., M. Alandete-Saez,J. L. Bowman and V. Sundaresan (2009).Auxin-dependent patterning and gamete specification in the Arabidopsis female gametophyte. *Science* 324: 1684-1689.

63. Panoli, A., M. V. Martin, M. Alandete-Saez, M. Simon, C. Neff, R. Swarup, A. Bellido, L. Yuan, P. C. Pagnussat and V. Sundaresan (2015).Auxin import and local auxin biosynthesis are required for mitotic divisions, cell expansion and cell specification during female gametophyte development in *Arabidopsis thaliana*. *PLoS One 10*: 1-23.

64. Pelaz, S., G. S. Ditta, E. Baumann, E. Wisman and M. F. Yanofsky (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* 405: 200–203.

65. Phillips, J., O. Artsaenko, U. Fiedler, C. Horstmann, H. P. Mock, K. Müntz and U. Conrad (1997). Seed-specific immunomodulation of abscisic acid activity induces a developmental switch. *EMBO J.* 16: 4489-4496.

66. Portereiko, M. F., L. Sandaklie-Nikolova,A.Lloyd,C. A. Dever, D. Otsuga and G. N. Drews (2006). NUCLEAR FUSION DEFFECTIVE1 encodes the Arabidopsis RPL21 M protein and is required for karyogamy during female gametophyte development and fertilization. *Plant Phys.* 141: 957-965.

67. Raz, V., J. H. Bergervoet and M.Koornneef (2001). Sequential steps for developmental arrest in Arabidopsis seeds. *Development* 128: 243-252.

68. Raventós, D., K. Skriver, M. Schlein, K. Karnahl, S. W. Rogers, J. C. Rogers and J. Mundy (1998): HRT, a novel zinc finger, transcriptional repressor from barley. *J. Biol. Chem.* 273: 23313-23320.

69. Riechmann, J. L. and O. J. Ratcliffe (2000). A genomic perspective on plant transcription factors. *Curr. Opinion Plant Biol.* 3: 423-434.

70. Sabelli, P. A. and B. A. Larkins (2009). The contribution of cell cycle regulation to endosperm development. *Sex. Plant Reprod.* 22: 207-219.

71. Schmitz, R. J., M. D. Schultz, M. A. Urich, J. R. Nery, M. Pelizzola, O. Libiger, A. Alix, R. B. McCosh, H. Chen, N. J. Schorkand J. R. Ecker (2013). Patterns of population epigenomic diversity. *Nature* 495: 193–198.

72. Shaw, P. and J. Doonan (2005). The nucleolus. Playing by different rules?.*Cell cycle* 4:102-105.

73. Soltis, D. E., A. S. Chanderbali, S. Kim,M. Buzgo and P. S. Soltis (2007). The ABC Model and its applicability to basal angiosperms. *Annals of Botany* 100: 155–163.

74. Steffen, J. G., I. H. Kang, M. F. Portereiko , A. Lloyd and G. N. Drews (2008). AGL61 interacts with AGL80 and is required for central cell development in Arabidopsis. *Plant Physiol.* 148: 259-268.

75. Stoddard, B. L. (2005). Homing endonuclease structure and function. *Q. Rev. Biophys.* 38: 49-95.

76. Sussman, M. R., R. M. Amasino, J. C. Young, P. J. Krysan and S.Austin-Phillips (2000). The ArabidopsisKnockout Facility at the University of Wisconsin–Madison. *Plant Physiol.* 124: 1465-1467.

77. Theissen, G. and H. Saedler (2001). Floral quartets. *Nature* 409: 469–471.

78. Theißen, G., T. Münster, L. U. Wingen, W. Faigl, S. Werth and H. Saedler (2001).Characterization of three GLOBOSA-like MADS-box genes from maize: Evidence for ancient paralogy in one class of floral homeotic B-function genes of grasses. *Gene* 262: 1-13.

79. Weber, H., L. Borisjuk and U. Wobus (2005). Molecular physiology of legume seed development. *Ann. Rev. Plant Biol.* 56: 253-279.

50

80. Weigel, D. and E. M.Meyerowitz (1994). The ABCs of floral homeotic genes. *Cell* 78: 203–209.

81. Weigel, D. andV. Colot (2012). Epialleles in plant evolution. *Genome Biology* 13: 1-6.

82. Verhoeven, E. E., M. V. Kesteren, G. F. Moolenaar, R. Visse and N. Goosen (2000). Catalytic sites for 3' and 5' incision of *Escherichia coli* nucleotide excision repair are both located in UvrC. *The Journal of Biological Chemistry* 275: 5120–5123.

83. West, M. and J. J.Harada (1993). Embryogenesis in higher plants: An overview. *Plant Cell* 5: 1361–1369.

84. Western, T. L., D. J. Skinner and G. W. Haughn (2000). Differentiation of mucilage secretory cells of the Arabidopsis seed coat. *Plant Physiol.* 122: 345-356.

85. White, C. N. and C. J. Rivin (2000). Gibberellins and seed development in maize. II. Gibberellin synthesis inhibition enhances abscisic acid signaling in cultured embryos. *Plant Physiol.*122: 1089-1097.

86. Willemse, M. T. M. and J. L. van Went (1984). The female gametophyte. *In: Embryology of Angiosperms*. (Berlin: Springer-Verlag), pp. 159–196.

87. Wohlfarth, T. (1996). Regulation samenspezifischer Gene der Ackerbohne *Vicia faba*: Analyse von cis-Elementen sowie Isolierung und Charakterisierung von Transkriptionsfaktoren. Dissertation, Martin-Luther-Universität Halle-Wittenberg.

88. Yadegari, R. and G. N. Drews (2004). Female gametophyte development. *Plant Cell* 16: S133-S141.

89. Xiao, W., M. Gehring, Y. Choi, L. Margossian, H. Pu, J. J. Harada, R. B. Goldberg, R. I. Pennell and R. L. Fischer (2003). Imprinting of the MEA polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Developmental Cell* 5: 891-901.

90. Zhao, Y., A. K. Hull, N. R. Gupta, K. A. Goss, J. Alonso, J. R. Ecker, J. Normanly, J. Chory and J. L. Celenza (2002). Trp-dependent auxin biosynthesis in Arabidopsis: Involvement of cytochrome P450s CYP79B2 and CYP79B3. *Genes Dev.* 16: 3100-3112.

Chapter 2

THE IDENTIFICATION OF DIFFERENTIALLY METHYLATED TARGET REGIONS USING A WHOLE GENOME METHYLATION APPROACH

Bui Thi Mai Huong[1], Helmut Bäumlein[2], Markus Kuhlmann[2]

1: Vietnam National University of Forestry

2: Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

2.1. ABSTRACT

The chapter 1 mentioned working hypothesis concerning the molecular function of ET factors suggests that they might act as novel regulators of epigenetic methylation patterns. The whole genome of *et* mutants have been analysed by bisulphite sequencing techniques comparing to wild type as control. These experiments showed high similarity between three replicates. Comparing to Col-0 for *et1-1*, *et2-3* and *et1-1/et2-3* lines detected 352, 373 and 275 highly differentially methylated regions, respectively. DNA methylation analyses revealed the ET specific differentially methylated regions such as some regions hDMR for RdDM (RNA-directed DNA methylation) pathway and the second region for the NERD (Needed for RDR2-independent DNA methylation) pathway. It is concluded that ET factors represent novel plant specific epigenetic regulators of reproductive tissue development acting on DNA-methylation.

2.2. BACKGROUND

2.2.1. DNA methylation and epigenetics

Epigenetics refers to processes causing dynamic alterations in the transcriptional potential of a cell which are not caused by changes in the DNA sequence including for instance DNA methylation and histone modifications for instance leading to gene silencing. In principle, there are two different mechanisms of gene silencing known: the RNA-directed DNA methylation pathway and an RNA interference pathway for transcriptional gene silencing (He et al., 2011). DNA-methylation occurs by the addition of a methyl group to the cytosine bases of DNA to form 5-methylcytosine (He et al., 2011). Typically, DNA methylation is removed during zygote formation and re-established through successive cell divisions during development (Jaenisch and Bird, 2003). In animals, methylation occurs exclusively in the symmetric context CG and CHG, whereas in plants, methylation occures also in the assymmetic CHH context, where H stands for A, T or C (Lister et al., 2009). In Arabidopsis about 7% of the whole genome is methylated-

among this 24% CG, 6.7% CHG and 1.7% CHH (Cokus et al., 2008). DNA-methylation is considered to act as a protective mechanism to prevent the activation of retrotransposons but is also required for differential transcription regulation during differentiation and development (Köhler et al., 2012).

Three pathways for DNA methylation regulation have been described: a) *de novo* methylation, b) maintenance methylation and c) de-methylation (Fig.2.1).
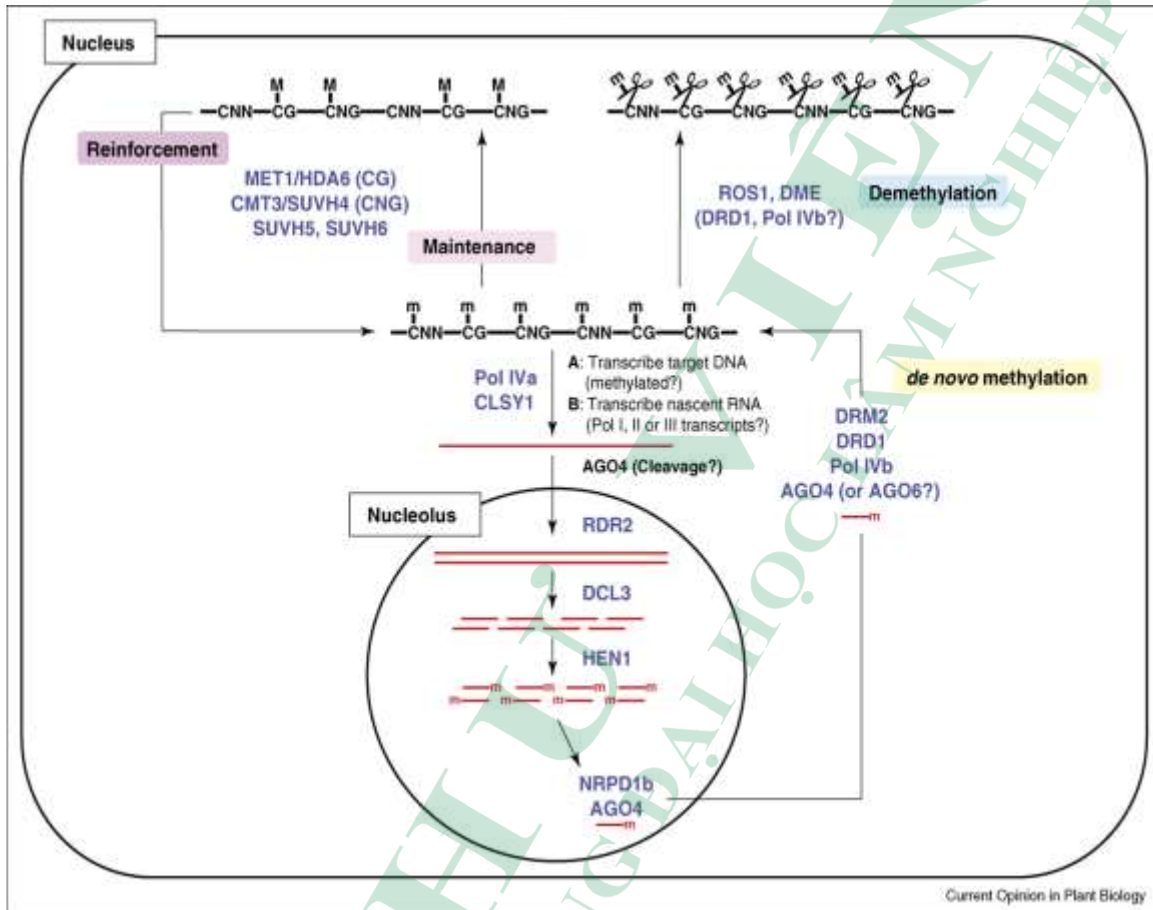

Fig.2.1. schematic model for the dynamic regulation of methylation

(Matzke and Mosher, 2007):

*Middle:*The *de novo* methylation pathway: Pol IVa together with protein CLSY1 transcribes the target locus, which might already be lightly methylated (´m´) or associated with specific histone modifications (A). Alternatively, Pol IVa might transcribe a nascent RNA produced at the target locus by Pol I, II or III (B).

Top, left: The maintenance pathway: CG and CNG methylation can be maintained during DNA replication by MET1 and CMT3, respectively. Locus-specific histone modifications that are catalyzed by HDA6, SUVH4, SUVH5 and/or SUVH6 help to maintain cytosine methylation (´M´) and reinforce the silent state.

Top, right: The demethylation pathway: DNA methylation can be lost in nondividing cells by a base excision repair-type mechanism that involves DNA glycosylase/lyase proteins such as ROS1 and DME (Choi et al., 2002; Gehring et al., 2009).

2.2.2. De novo methylation

RNA-directed DNA methylation (RdDM) was first discovered in 1994 in viroid-infected tobacco plants (Viswanatha and Tian-Kang, 2009). A single-stranded transcript of polymerase IV (POL IV) is transferred from nucleus to the nucleolus by an unspecified mechanism, where it is copied into double stranded RNA by RNA-dependent-RNA polymerase 2 (RDR2). The double-stranded RNA is cleaved into 24-nt primary siRNAs by a Dicer-like protein DCL3 (Matzke and Mosher, 2007). The siRNAs are methylated at their ends by HUA ENHANCER 1 (HEN1) and then the siRNA is loaded on a RISC complex (RNAi-induced silencing complex). This complex contains the ARGONAUTE4 protein (AGO4), which interacts with the C-terminal domain of NRPD1b (largest subunit of POLIV). This complex moves out of the nucleolus into the nucleoplasm, where NRPD2a subunit is added to form functional POL IVb complex. In addition, nascent non-coding RNA transcripts produced by POL V have been suggested to serve as scaffolds for recruiting the AGO4-containing RdDM effector complex by base-pairing with guide siRNAs (Wierzbicki et al.,2008). The functional RdDM effector complex directs the *de novo* DNA methyltransferase DRM2 (Domains Rearranged Methyltransferase 2) to specific chromatin regions to catalyze new DNA methylation (Matzke and Mosher, 2007; He et al., 2011).

2.2.3. Maintenance methylation

In Arabidopsis about one-third of genes have CG methylation in their coding region, which can be maintained by methyltransferease 1 (MET1) (Matzke et al., 2007). CHG methylation can be maintained by chromomethylase 3 (CTM3) or SUPPRESSOR OF VARIEGATION 3-9 HOMOLOGUE 4 (SUVH4 also known as KYP) and SUVH5 and SUVH6 (Matzke and Mosher, 2007). Finally, the CHH methylation is maintained by CMT3 and DRM2 (Law and Jacobsen, 2010).

2.2.4. DNA demethylation

Active demethylation occurs in plants by DNA glycosylase activity, probably in combination with the base excision repair (BER) pathway. DNA glycosylases include DEMETER (DME) and REPRESSOR OF SILENCING 1 (ROS1) as well as DEMETER-LIKE2 (DML2) and DEMETER-LIKE3 (DML3) (Choi et al., 2002; Gehring et al., 2009).

2.3. MATERIALS AND METHODS

2.3.1. Materials

Arabidopsis thaliana ecotypes Columbia-0 (Col) and *Wassilewskija-2* (Ws) were obtained from Gene Regulation Group (IPK, Gatersleben, Germany) and used throughout

this study as wild type control experiments. T-DNA insertion lines *et1-1; et2-3; et1-1/et2-1; et1-1/et2-3* were generated as showed in chapter 1.

2.3.2. Bacterial strains

Several bacterial strains were used for different purposes such as DNA cloning, plasmid DNA amplification, sequencing etc… as showed in chapter 1.

2.3.3. Enzymes, markers, antibiotics

Enzymes, markers, antibiotics, other chemicals were used for different experiments as showed in chapter 1.

2.3.4. Commercial kits

GeneJET plasmid miniprep kit, GeneJET gel extraction kit, RevertAid first strand cDNA synthesis kit, DNA labelling kit (Fermentas, Vilnius, Lithuania); DNeasy plant mini kit, Epitect bisulfite kit, QIAquick PCR purification kit, QIAquick gel extraction kit, Qiagen plasmid purification kit mini (Qiagen, Hilden, Germany).

2.3.5. Vectors

Various vectors were used for DNA amplification, cloning genes into plants and other purposes as showed in chapter 1.

2.3.6. Methods

Methylation studies

Ten days old Arabidopsis seedlings were harvested and immediately frozen in liquid nitrogen. Total DNA was isolated from 100 mg ground plant material using DNeasy Plant Mini kit (Quiagen). DNA concentration was quantified at a Nanodrop ND-1000 spectrophotometer (NanoDrop technologies Inc., USA). Total RNA was extracted from rosette leaves by using the Qiagen Plant RNeasy kit (Qiagen GmbH, Hilden, Germany). About 1 µg genomic DNA was split to 300 bp average size with a Covaris S2 instrument using the following settings for 120 s in frequency sweeping mode: intensity 5, duty cycle 10%, 200 cycles per burst. Then the DNA was purified by Qiaquick PCR purification columns. Libraries were generated by using the NEBNext DNA Sample Prep Reagent Set 1 (New England Biolabs) according to the Illumina Genomic Sample Prep Guide. After size selection, the non-methylated cytosine residues were converted to uracil by using the EpiTect Plus DNA Bisulfite kit (Qiagen) according to the manufacturer's protocol. Sequencing was done by an Illumina GAII instrument. Processing of genomic reads was performed by using the SHORE pipeline v.0.9.0 to trim and quality filter the reads (Ossowski, S. et al., 2008). The high quality sequences were aligned to Col-0 reference genome with Genome Mapper that supports the alignment of bisulphite converted reads (Schneeberger et al., 2009). The data processing was performed as described before (Becker et al., 2011).

2.4. RESULTS

The single-strand cutting function on DNA were performed by the GIY-YIG domain in the N-terminal and the ET factors (Ivanov et al., 2008) may be associated to the changes in higher order DNA structures, such as nucleosome sliding or the relaxation of supercoiled chromatin domains as a prerequisite for regulated gene expression (Choi et al., 2002; Xiao et al., 2003; Haince et al., 2006; Ju et al., 2006). In addition, these domains could be involved in active demethylation processes as described demethylases DME DEMETER (DME) and REPRESSOR OF SILENCING1 (ROS1) (Choi et al., 2002; Gong et al., 2002; Xiao et al., 2003; Choi et al., 2004; Morales-Ruiz et al., 2006). To test this, the methylation of the whole genome of *et* mutants have been analysed with wild type as control. Therefore, whole genome bisulphite sequencing (WGBS) of Col-0, *et1-1* and *et2-3* single mutants and the *et1-1/et2-3* double mutant were performed. As ET factors showed their maximum of expression in reproductive tissues (Ivanov et al., 2008), therefore, the analysis was focused on flower buds (Smyth et al., 1990). In total 527 differentially methylated regions (DMR) could be detected by using the iPlant visualization tool. The positions of these DMRs have been mapped on the Arabidopsis genome (Fig. 2.2).
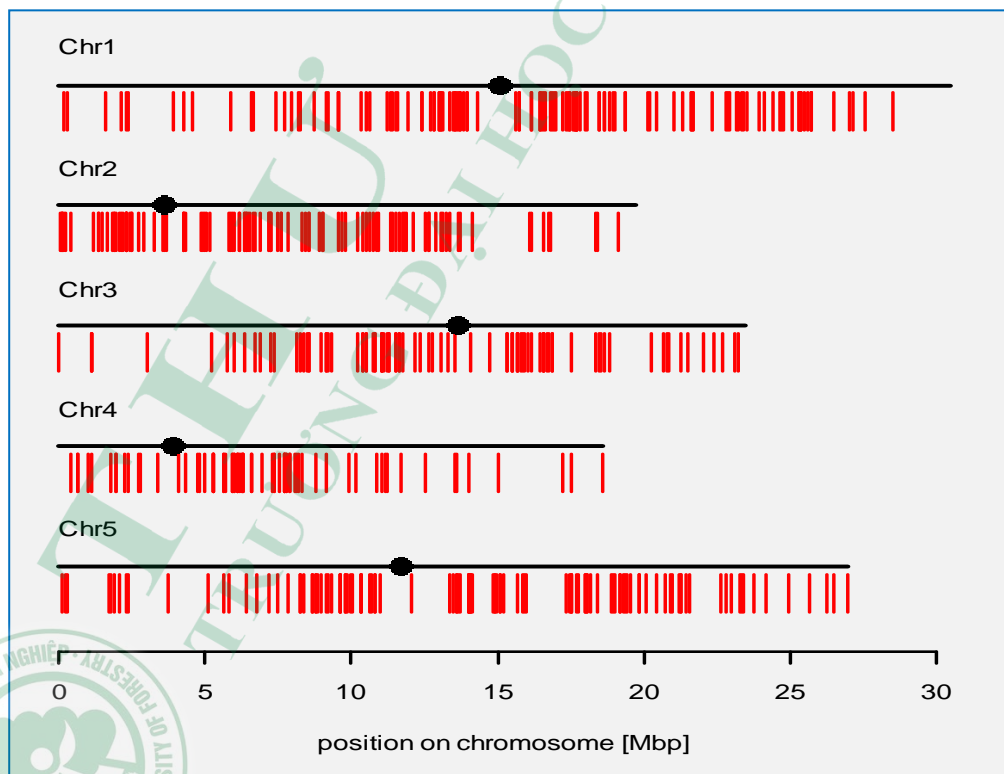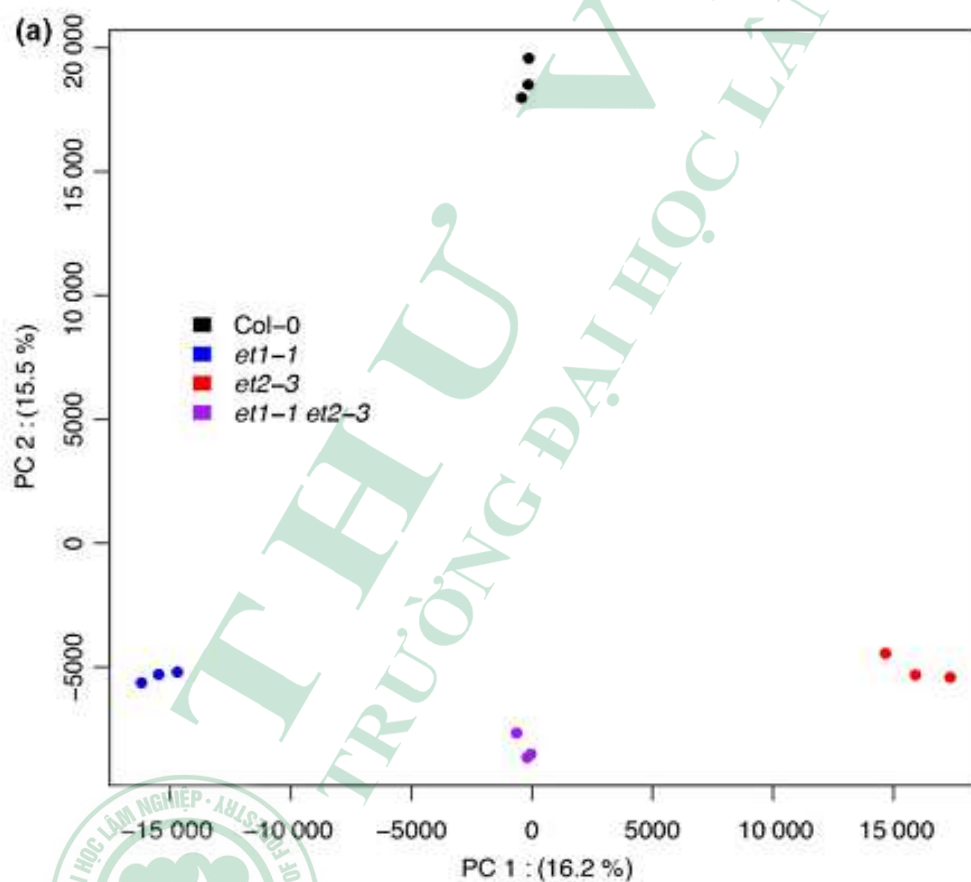

Fig. 2.2. Positions of differentially methylated regions in et mutants compared to wild type.

56

Principal component analysis (PCA) of differential DNA methylation showed clear high similarity between three replications (Fig.2.3a). Methylated regions (MRs) were identified in every sample using a previously published algorithm (Hagmann et al., 2015). The mutant lines detected 352 highly differentially methylated regions (hDMRs) for *et1-1*, 373 for *et2-3* and *275* for the double mutant (*et1-1/et2-3*) comparing to Col-0 (see Table S2 for a list of hDMRs in New Phytologist (2019) 221: 261–278. doi: 10.1111/nph.15439). The analysis of the hDMRs revealed the most of loss methylation in the *et* mutants compared to Col-0 (Fig. 2.3b), mainly in the symmetric CG context (see Fig. 2.4a). Genomic regions detected having hDMRs coincided mainly with transposable elements (TEs), and hDMRs were overrepresented 2 kb upstream and 2 kb downstream of protein-coding sequences (Fig. 2.3c). Methylated regions that were classified as non-DMRs showed minor variation in methylation, confirming the specificity of our algorithm (see Fig. 2.4d).
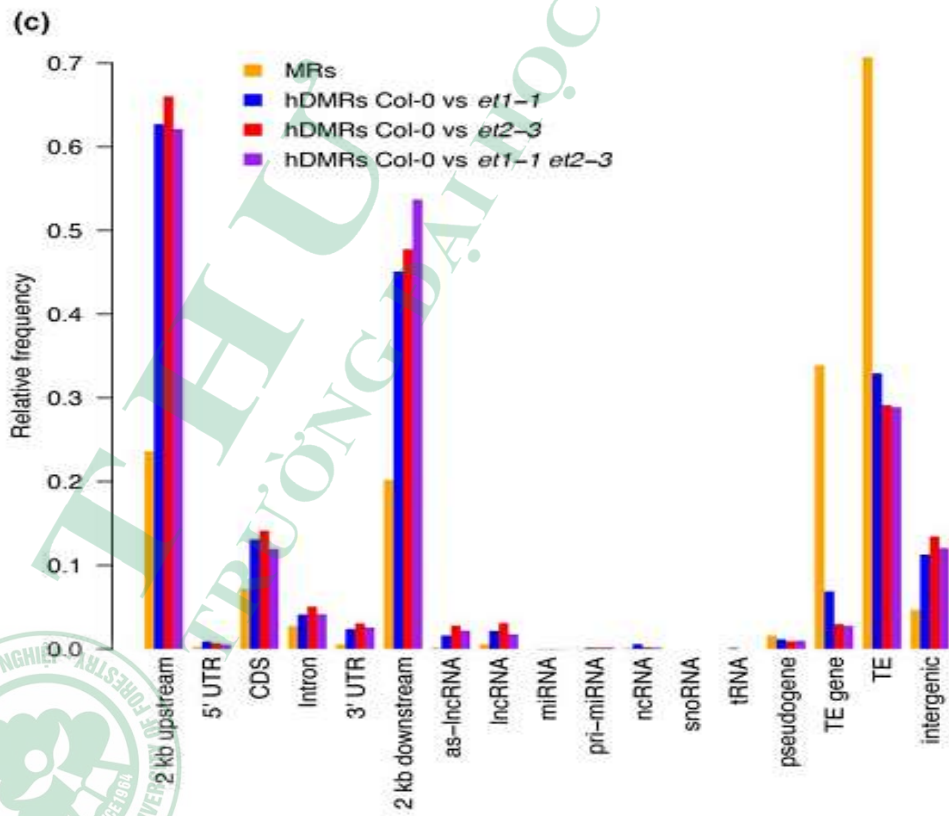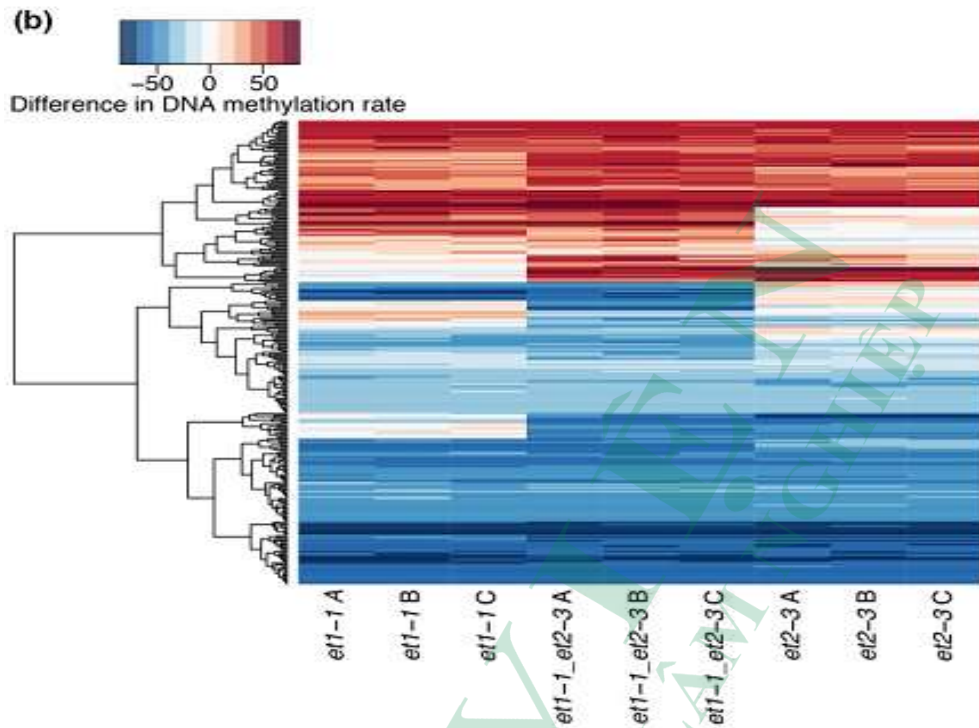
Fig. 2.3. Genome-wide methylation analysis of et mutants in flower.

(a) Principal component (PC) analysis of methylation rates within highly differentially methylated regions (hDMRs). For each hDMR, the average methylation rate was calculated per sample from the methylation rates of all cytosines contained within the region. Percentages indicate the amount of variance explained by the respective PC. (b) Gains and losses of methylation in *et1-1*, *et1-1/ et2-3* and *et2-3* hDMRs of all contexts (CG, CHG, CHH). Each line in the heat map represents an hDMR. Gains and losses are expressed as difference of the methylation rate in the mutant to the average of the three Col-0 replicates. (c) Annotation of cytosines in methylated regions (MRs) and hDMRs (Francesca Tedeschi, Bui Thi Mai Huong, 2019).
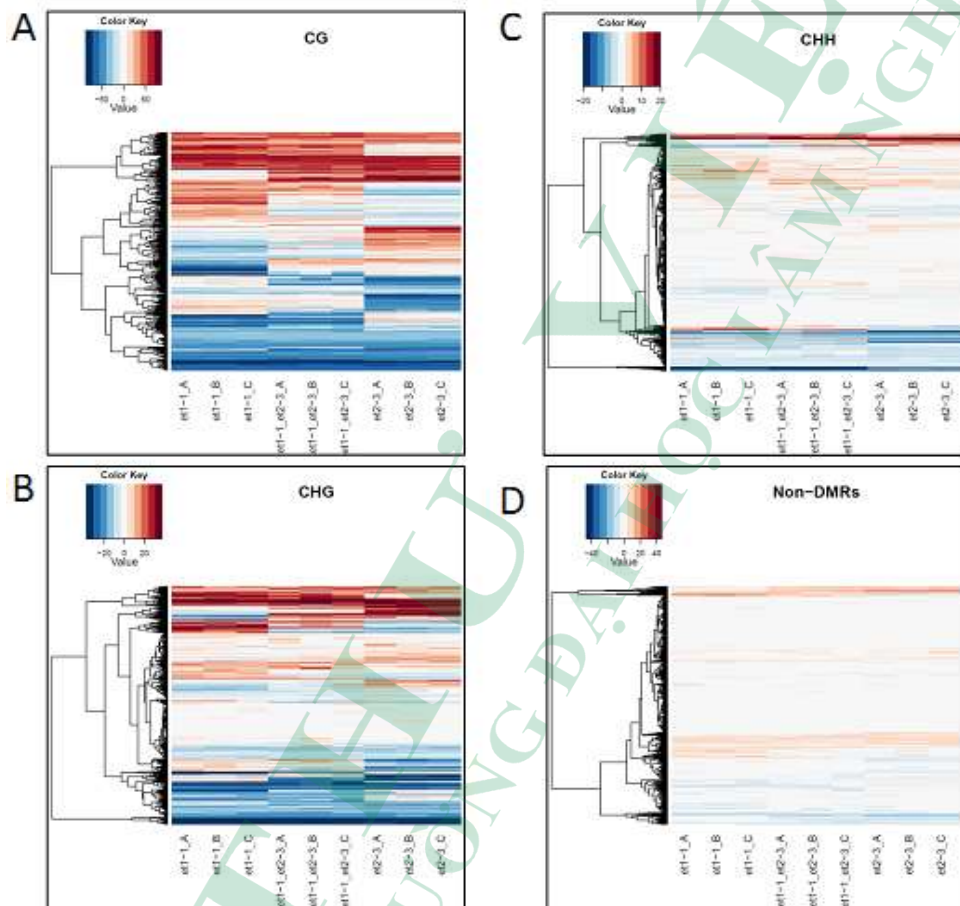

Fig. 2.4. Cluster analysis of et mutant hDMRs relative to Col-0 methylation.

Displayed is the difference in methylation rates of identified hDMRs by WGBS of *et1-1*, *et2-3* and *et1-1/et2-3* double mutant flowers.

Red color indicates increase in methylation, Blue color indicates decrease of methylation relative to Col-0.

A difference of methylation in CG context,

B. Difference in CHG context.

C. Difference in CHH context. H = ATG.

D. Analysis of nonDMRs.

Analysis was performed in triplicate. (A, B and C)

Among identified hDMRs, MPF (hDMR686) is showed *et1-1* specific hypermethylation. Whereas, MPF (Methylated region near Flowering locus C, AT5G10140) was identified as a marker region for loss of demethylation function (Penterman et al., 2007a; Zhai et al., 2008). To confirm the identified hDMRs the available et-T-DNA insertion mutant lines *et1-1*, *et2-1* and *et2-3* were tested by clonal bisulphite sequencing analysis (Fig. 2.5). AT1G26400 (FAD-Berberine-binding protein), AT1G34245 (EPF2, Epidermal Protein Factor2), AT5G20240 (Pi-Pistillata), AT2G36490 (ROS1-Repressor of silencing 1) hypermethylated in ros1 and dme mutants (Fig. 2.6; F2.7), and AtSN1 as a reference region for RdDM pathaway (Kuhlmann & Mette, 2012) were tested. At AT1G26400 and AT1G34245, AT5G20240, AT2G36490 have showed a significant increase of cytosine methylation which was detectable for both alleles of the et2 mutant (*et2-1* and *et2-3*). These increase were preferentially caused by an increase of methylation in the symmetric CG context (Fig. 2.5).
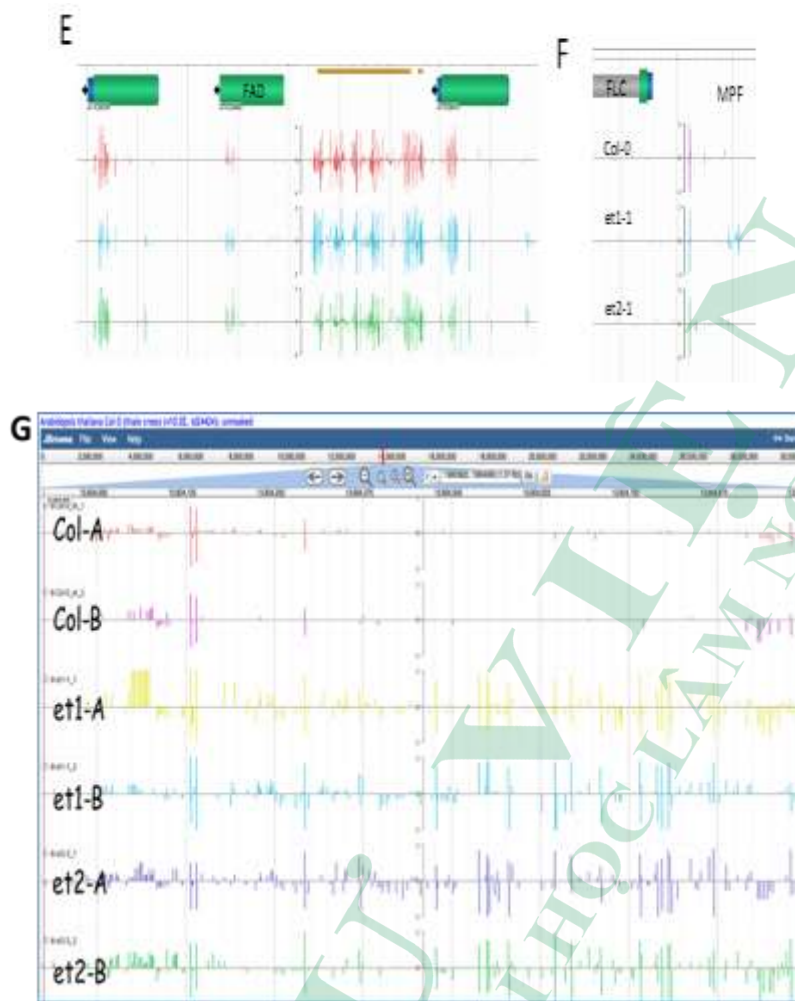

60
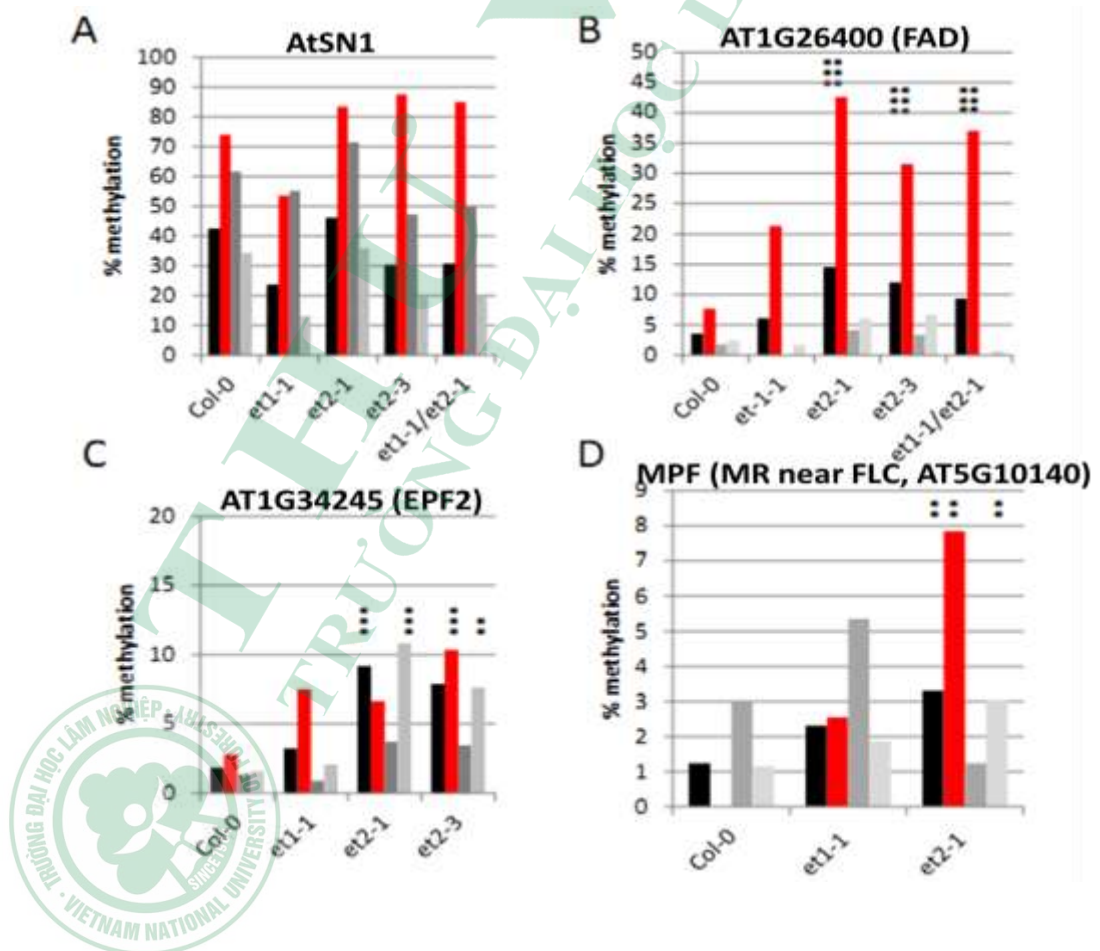
*Fig. 2.5. DNA methylation of selected region analysed in detail
by bisulfite sequencing [%].*

N indicates the number of clones sequenced per target and genotype. Stars indicate significant difference to Col-0: ****<0,005, **<0,01 by Chi square test. Cumulative methylation levels at all cytosines in the* analysed region (black columns), cytosines in CG context (red columns), CHG context (grey columns; H stands for A, C or T) and CHH context (bright grey columns) are given in %. **A**. Facultative heterochromatic AtSN1: N=22 (Col-0), 7(*et1-1*), 6(*et2-1*), 10(*et2-3*), 10(*et1-1/et2-1*). **B**. AT1G26400 (FAD-Berberine-binding protein): N=19 (Col-0), 14 (*et1-1*), 18(*et2-1*), 15 (*et2-3*), 11 (*et1-1/et2-1*); **C**. AT1G34245 (EPF2): N= 36(Col-0), 56 (*et1-1*), 27 (*et2-1*), 29 (*et2-3*). **D**. MPF (methylated region near FLC, AT5G10140), N= 25 (Col-0), 28 (*et1-1*), 20 (*et2-1*). **E**. Browser view for corresponding regions AT1G26400. **F** . MPF, methylated region near FLC, AT5G10140. And **G.** Methylation pattern at a selected position on chromosome 1. The chosen position is 1.07 kb long between position 13603922 and 13604993. Two independent biological replicates have been analyzed (A, B) and demonstrate the high reliability of the technique.
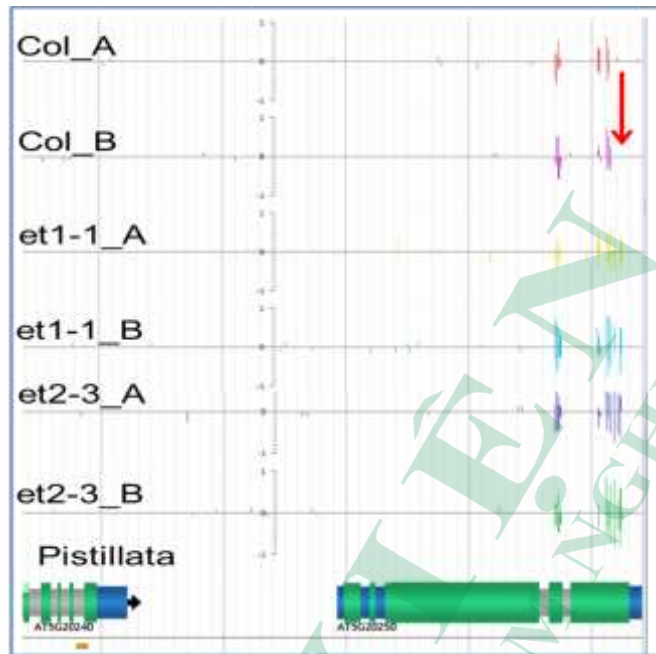
61

Fig. 2.6. Differential methylation pattern at Pistillata in et1 and et2 mutants.

The gene product is known to be involved in flower development where it is required to regulated the B function. Pistillata were found to be hypermethylated in the et mutants. Note the high reproducibility of the two independent biological replicates A and B. The basic figure is derived from the iPlant Visualisation Tool.

Fig. 2.7. Differential methylation pattern at ROS1 in et1 and et2 mutants.

The gene product is known to be involved in demethylation of DNA, therefore, ROS1 were found to be hypermethylated in all sequence contexts in the et mutants. Note the high reproducibility of the two independent biological replicates A and B. The basic figure is derived from the iPlant Visualisation Tool.
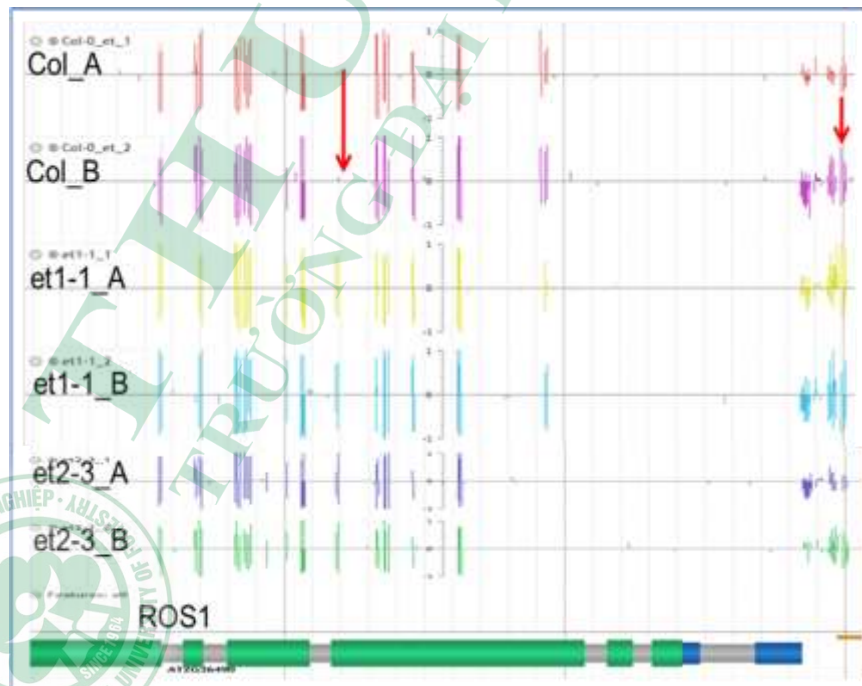
The largest hDMR (hDMR180: 1159 bp) was located on chromosome 1 : 28515015, that is a HELITRON1 element (AT1TE93275) (Fig.2.8). This region is methylated in wild type but unmethylated in three mutant lines (*et1-1*,*et2-3* and *et2-1*). The position of the differential methylation pattern in wild type and mutants precisely coincides with the position of the transposable element Helitron 1 in the promoter of the gene AT1G75950. This gene encodes for an E3-SCF protein also known as ASK1 (Arabidopsis SKP1 homologue 1) and has been identified to be involved in flower development. In cooperation with UFO and LEAFY gene, they regulates the B function required for flower development in Arabidopsis (Zhao et al., 2001). In addition, Demethylation of this region was characteristic for nerd mutant plants (Pontier et al., 2012). Loss of this GYF-and zinc-finger (CCCH-type) domain-containing protein function led to definition of a plant-specific chromatin-based RNA silencing pathway depending on RDR1/6. The second region defining the NERD pathway was psORF (AT5G35935). This region was also detected as hDMR750 in the *et* mutants.
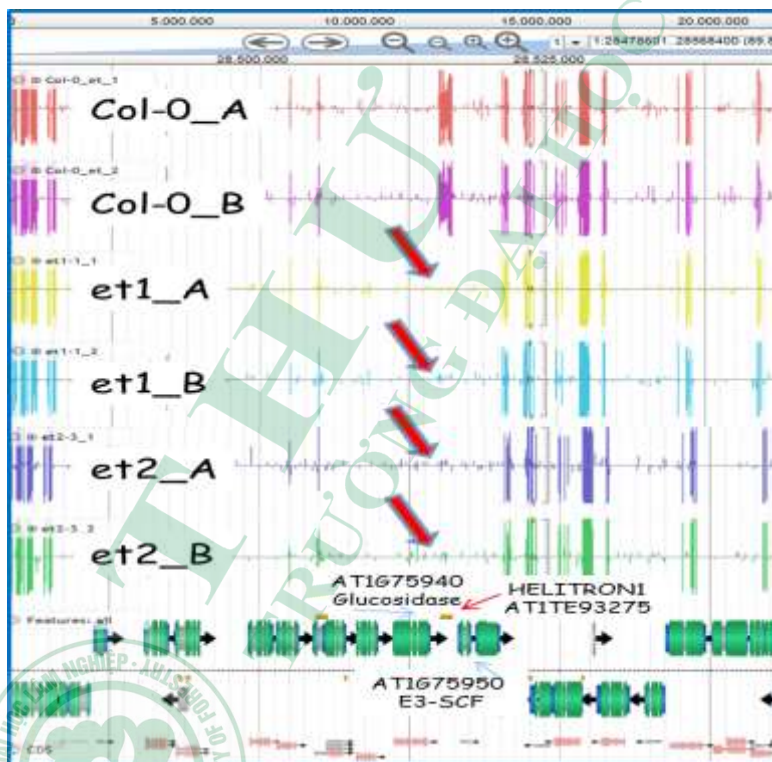

Fig. 2.8. Differential methylation pattern at the DMR diffM133.

The transposable element Helitron1 (AT1TE93275) positioned in the putative promoter region of the gene AT1G75950 is unmethylated in both mutants. The gene

product is known to be involved in flower development. Note the high reproducibility of the two independent biological replicates A and B. The basic figure is derived from the iPlant Visualisation Tool.

Another selected example shows the methylation pattern in the DMR diffM105 . This region exhibits different methylation with the gene AT1G61810 encoding β - glucosidase (Fig 2.9). The gene is methylated in wild type and the *et2-3* mutant but not in *et1-1* mutant. This preliminarily indicates the existence of ET1 and ET2 specific targets. An adjacent gene (AT1G61800) encodes a glucose 6-phosphate translocator and exhibits a differentially expression with a log2R of -3.8.
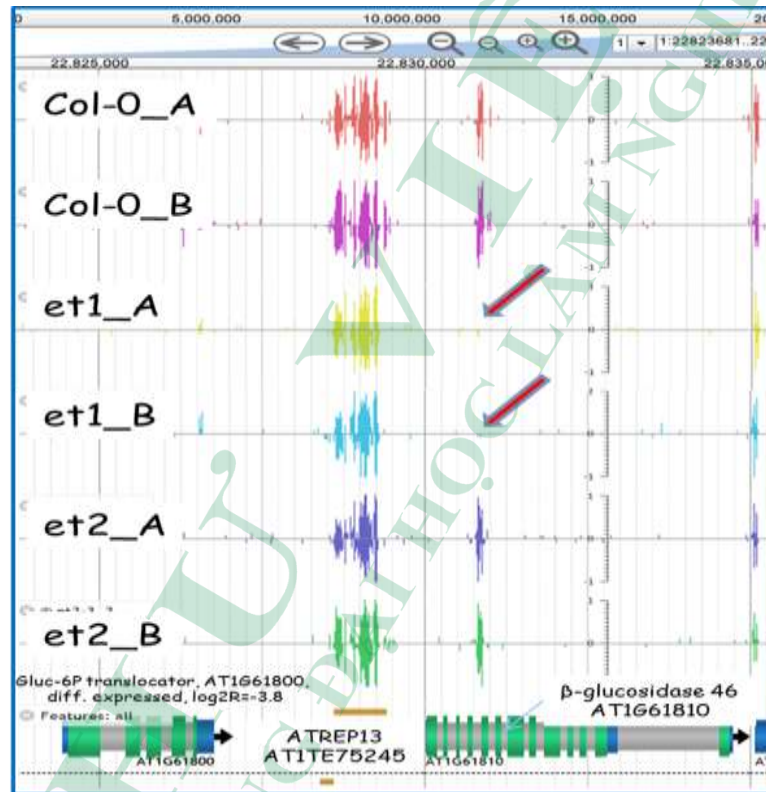

Fig. 2.9. Differential methylation pattern at the DMR diffM105.

The methylation is missing in *et1-1* but occurs in *et2-3* and wild type. The adjacent gene for glucose 6-phosphate translocator has been found to be differentially expressed.

The analysis of various other DMR demonstrates that all combinatorial possibilities of methylation between wild type and the three mutants do occur. The identified hDMRs in the *et1* and *et2* single mutants overlapped substantially (Fig. 2.10), therefore, to propose similar regulatory function of ET1 and ET2 at these shared loci. Differential DNA methylation with respect to Col-0 was similar in both mutants for a large fraction of hDMRs (Fig. 2.3c). In total, 70% of hDMRs showed the same directional methylation change. However, For example, the diffM024 region is methylated in wild type and *et2-3*, but not in *et1-1*. At diffM048 methylation is detected in wild type and *et1-1* with

64

methylation missing in *et2-3*. Finally, in diffM001 no methylation is found in wild type and *et2-3* whereas the region is methylated in *et1-1*, which suggests antagonistic roles of ET1 and ET2 for methylation of these loci (Fig. 2.11).
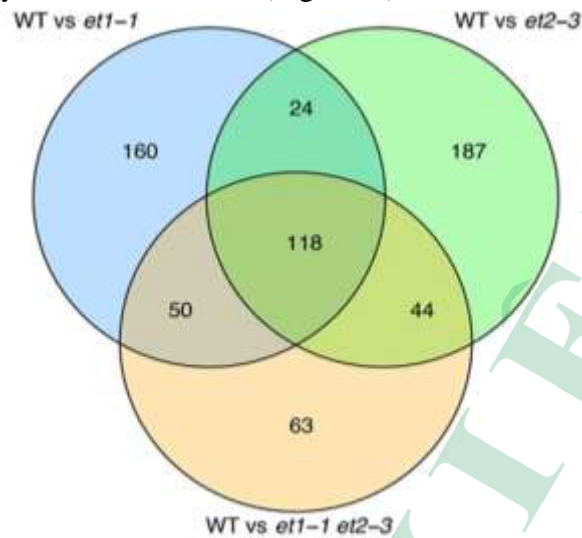

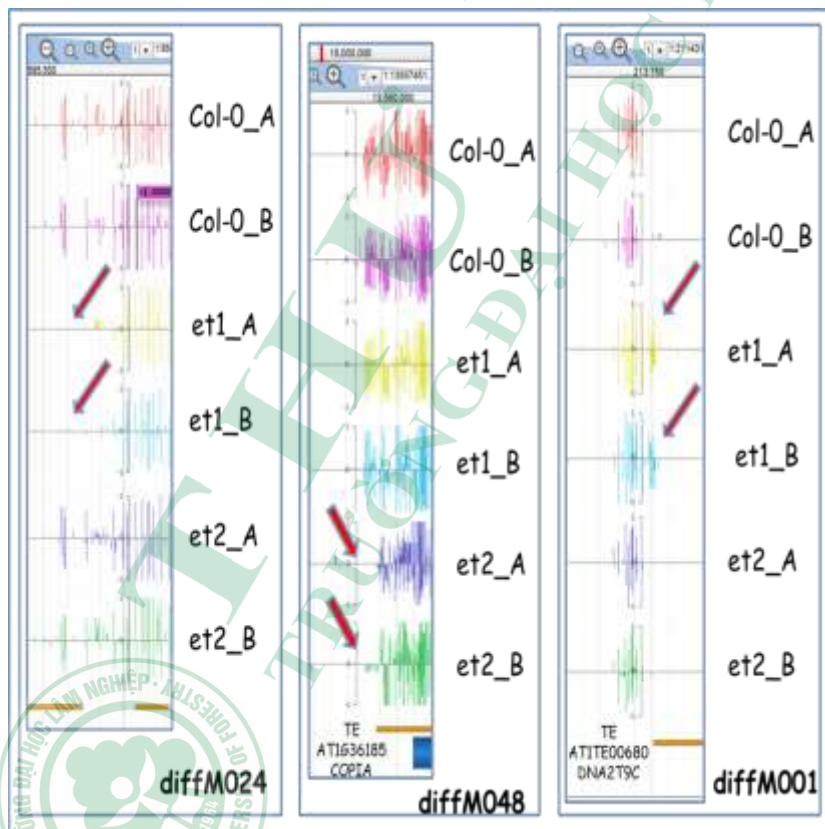*Fig. 2.10. Overlap of hDMRs in all et mutants vs Col-0 (WT)*

Fig. 2.11. Differential methylation in selected DMRs reveals the occurrence of various combinations between wild type and the two mutants.

Taken together, all possible combinations concerning the methylation patterns in wild type and the two mutants are possible and have been detected as summarized in table 2.1. The methylation context CG is the most frequent one, followed by CHG with very rare cases of the CHH context. Up- and down methylation in mutants occurs with similar frequency. The observation suggests that ET1 and ET2 have common and gene specific target regions.

Tab. 2.1. All combinations of methylation patterns between wild type and both et mutants can be found. ET1 and ET2 have common AND gene specific target regions.
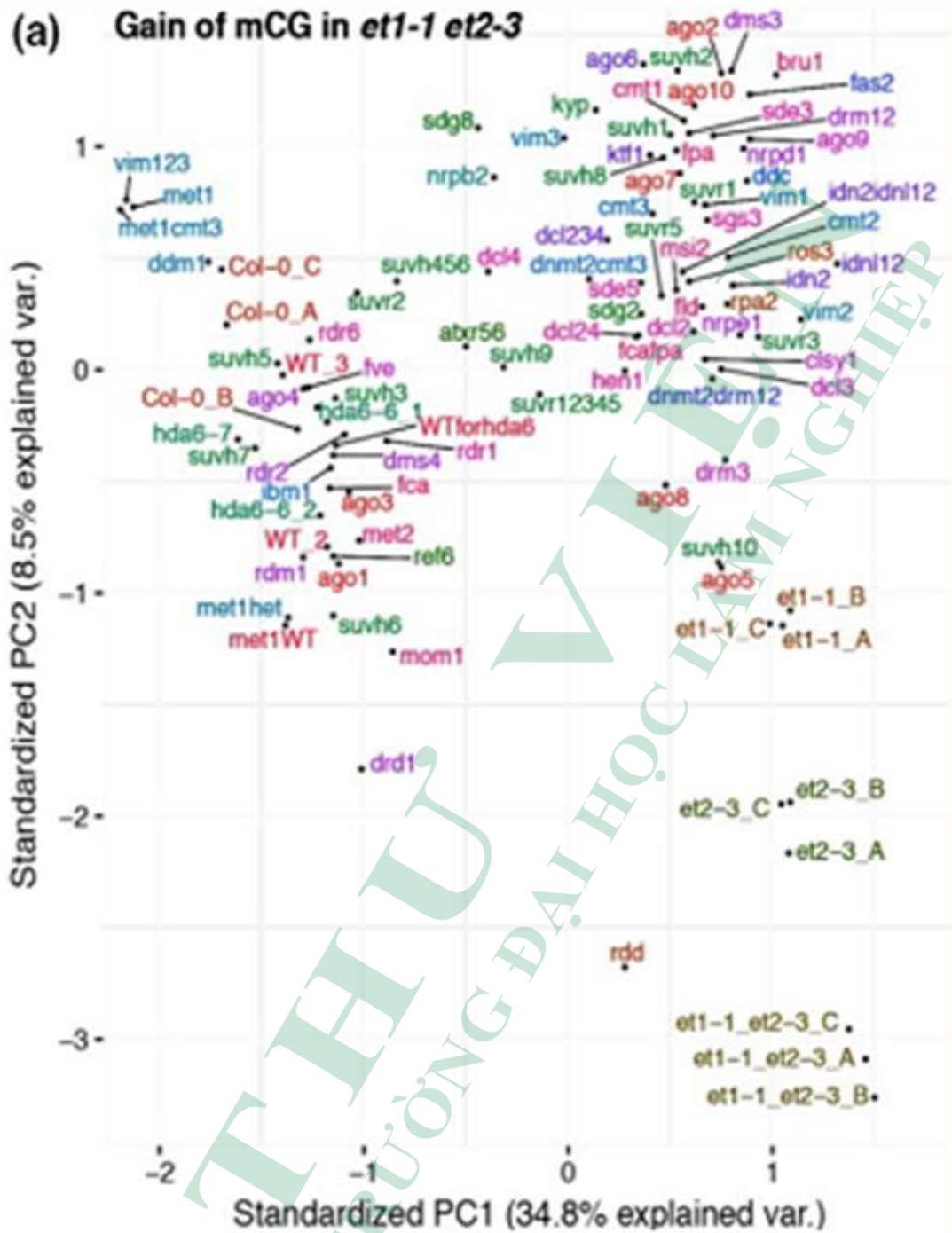
Col	*et1-1*	*et2-3*	
+	-	-	√
+	+	-	√
+	-	+	√
-	+	+	√
	+	-	√
-	-	+	√

For the most of hDMRs, DNA methylation in the *et1-1/et2-3* double mutant was similarity DNA methylation in one of the single mutants (*et1-1* or *et2-3*), or showed additive effects because they corroborated the combination of overlapping and specific function of ET1 and ET2 that we had already derived from the hDMR overlap analysis. Whereas, intriguingly, a small subset of hDMRs that showed loss of methylation in either *et1-1* or *et2-3* did not show methylation changes in the double mutant, suggesting epistatic interaction of ET1 and ET2 at these loci.

To gain insights into the methylation pathway that ET1 and ET2 might be involved in, we next used the hDMR between Col-0 and the *et1-1/et2-3* double mutant as a proxy to investigate DNA methylation at the same loci in a collection of previously published epigenetic mutants (Stroud et al., 2013). As CG methylation was the most prominently affected in *et1-1/et2-3*, we focused our analysis on this context. Analysis of hDMRs with gain of methylation in *et1-1/et2-3* revealed the closest similarity to methylation patterns of rdd mutant plants (Fig. 2.12a). rdd is a triple mutant defective for ROS1, DM2 and DML3 (Penterman et al., 2007b).

Regions which reduced CG methylation in *et1-1/et2-3* (Fig. 2.12b) compared to Col-0 showed the closest similarity to met1 (DNA METHYLTRANSFERASE 1; Kankel et al., 2003) and the triple mutant vim1 vim2 vim3 (VARIANT IN METHYLATION; Shook & Richards, 2014), both defective for CG-specific maintenance of methylation.

66

(a) Gain of mCG in *et1-1 et2-3*
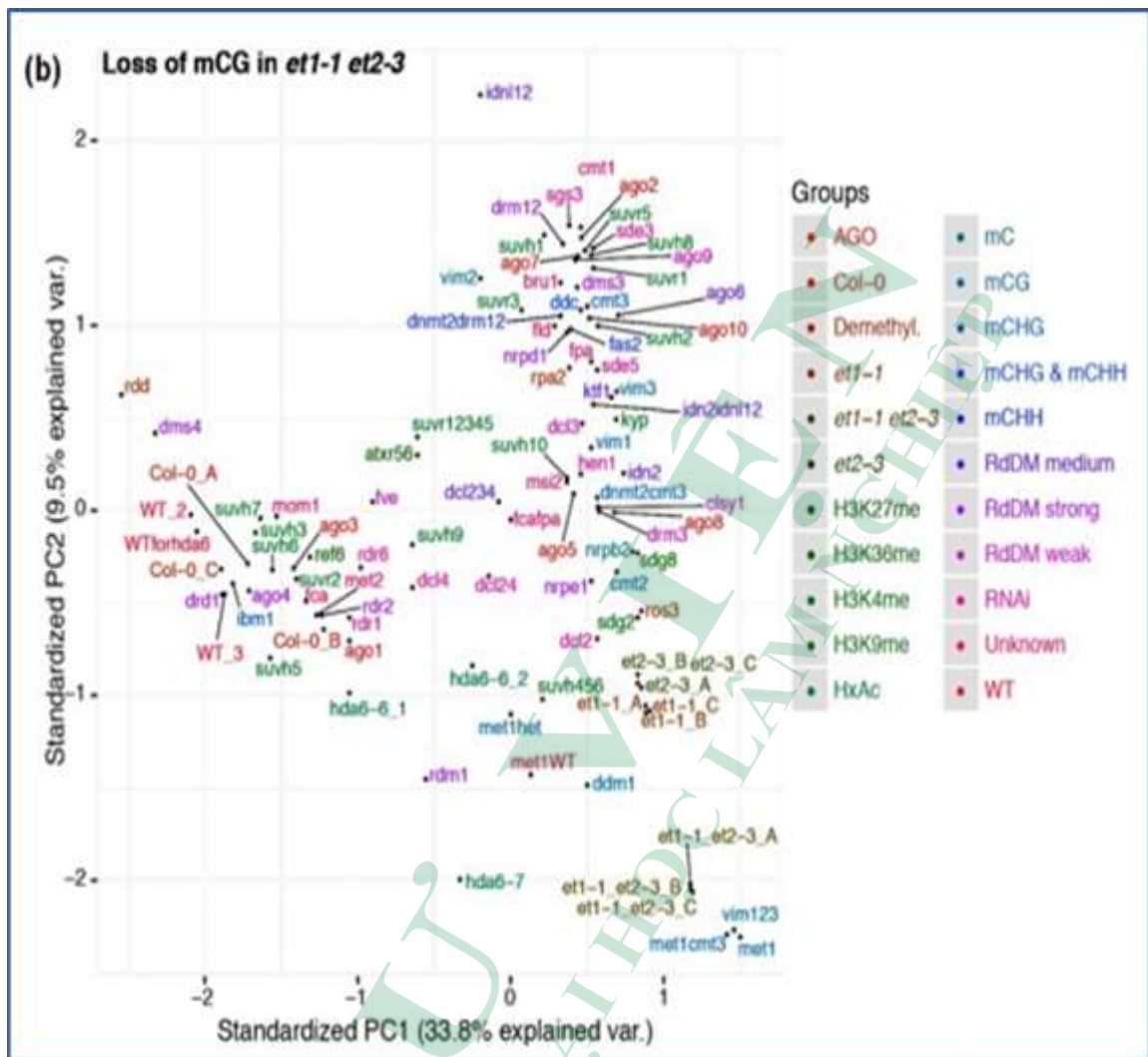
Fig.2.12. Principal component (PC) analysis of highly differential methylated regions (hDMRs) in Arabidopsis thaliana et1-1/et2-3 vs Col-0 (WT) and other epigenetic mutants (according Stroud et al., 2013).

(a) hDMRs with gain of methylation in *et1-1/et2-3* of CG context. (b) hDMRs with loss of methylation in *et1-1/et2-3* of CG context (Francesca Tedeschi, Bui Thi Mai Huong, 2019).

Although prior *in vitro* studies (Ivanov et al., 2012) showed that ETs bind to DNA irrespective of the sequence context, we investigated whether any sequence feature could be identified using the set of identified DMRs. Therefore, we choose the 136 hDMRs detected in *et2-3* flower tissue which showed gain of methylation. We applied the motif-based sequence analysis tool MEME (Bailey et al., 2006) on these potential ET2 DNA binding motifs. No motif could be identified, suggesting that the DNA binding is not sequence-specific. The DIMONT approach (Grau et al., 2013), which includes sorting of the sequences according to intensities, did not reveal any binding pattern either.

Importantly, when using a motif length of 10 (bgOrder = 0, motifOrder = 0, other parameters = default), we detected two adjacent pyrimidines (TT, CT, TC) as a recurring motif (Fig. 2.13).
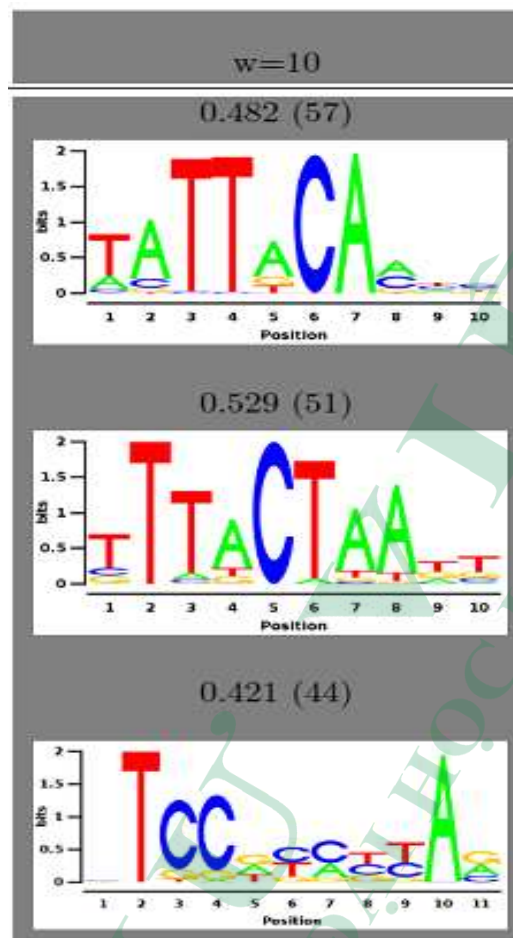
Fig. 2.13. Identification of common motifs in et2-3 highly differential methylated regions (hDMRs) (N=136) utilising the DIMONT approach.

The number of DMRs included in each motif is given in brackets. DMRs were extendet. +/- 50 bps to include regions before and after DMRs. Motif length was set to w010, bgOrder=0, motifOrder=0, other parameters = default.

2.5. DISCUSSION

2.5.1. Genome wide methylation

In collaboration with Dr. C. Becker at MPI Tübingen it became possible to compare the genome-wide methylation status between wild type and both mutants *et1-1* and *et2-3*. Using the iPlant visualisation tool in total 527 differentially methylated regions could be detected. As shown in Fig. 2.2 these regions are rather equally distributed along the five chromosomes. At first, the data show a remarkable reproducibility for detected genomic regions in three completely independent experimental replicates. Despite of the obviously

69

high reliability of the described results, it is currently difficult to interpret the results in detail. The methylation context CG is the one which is mostly affected followed by the other symmetric context CHG whereas the CHH context is very rarely affected. Up- and down methylation in mutants occurs with similar frequency. Concerning the methylated *versus* non-methylated status of different genomic regions compared in wild type, *et1-1* and *et2-3* mutants all combinations have been detected (Tab. 2.1) indicating a high complexity of the underlying regulatory mechanisms. As revealed by the transcriptome analysis, ET1 and ET2 have both common and gene specific target regions in terms of methylation.

A putatively interesting special case deserving a more detailed discussion concerns the differentially methylated region diffM133 (Fig. 2.8), which is non-methylated in both mutants. The differentially methylated region precisely overlaps with the position of a HELITRON transposable element. The transposon is located in the putative promoter region of a gene encoding an E3-SCF protein also known as Arabidopsis SKP1 Homologue (ASK1). Remarkably, this gene product in cooperation with UFO and LEAFY is involved in flower development where it is required to regulated the B function (Zhao et al., 2001). Thus, this invites the speculation that this differentially methylated region contributes to the above described flower phenotypes of *et* mutants.

Another selected example concerns the differentially methylated region diffM105 (Fig. 2.9) containing a β-glucosidase encoding gene. This gene is methylated in wild type and *et2-3* but not in the *et1-1* mutant further indicating that ET1 and ET2 have specific target genes. In preliminary results, an adjacent gene encoding a glucose-6-phosphate translocator was found to be differentially expressed.

2.5.2. Identified DMRs in the et mutants indicate similarities to mutants with impaired demethylation

Based on our hypothesis that ET factors act on DNA by single strand cleavage, gain of methylation is expected for ET-target regions in the ET loss of function mutants. Therefore, identified DMRs were separated for gain and loss of methylation and analysed individually. Approximately one-third of the identified hDMRs showed gain of methylation in the mutants and two-thirds loss of methylation. The detectable hypomethylation might result from complex feedback regulation, also reported for ros1 (Zhu et al., 2007) and dme mutant plants (Ortega-Galisteo et al., 2008). The evolutionary neo-functionalisation of the ET2-based gene duplication might explain the identified ET1 and ET2 specific DMR and DEGs.

PCA of hypomethylated hDMRs in et mutants indicated similarities with met1 and vim123 mutants (Kim et al.,2014; Shook&Richards, 2014). This observation indicated the unspecific loss of methylation in genomic regions which are under control of the DNA methylation maintenance pathway. The hypermethylated hDMRs exhibit a high similarity to DMRs detected in the genome of the rdd mutant, a triple mutant defective for ROS1 and

DeMeter Like-2 and 3 (DML2 and DML3) (Penterman et al., 2007b), all involved in demethylation of DNA. Moreover, selected regions known to be affected by active demethylation such as AT1G26400, AT1G34245, AT5G10140, AT5G20240 and AT2G36490 (Fig.2.5; Fig.2.6; Fig.2.7) were found to be hypermethylated in all sequence contexts in the *et* mutants. Together this implies either a function of ETs in demethylation via deregulating ROS1, DME and DML2/3 or ETs represent another novel component of the active demethylation pathway. Because DME and DMLs are not found among the differentially expressed genes in *et* mutants, we favour the second view. As regulation at the post-translational level or upregulation in a different context cannot be excluded, the detailed molecular mechanism remains to be investigated.

2.5.3. Transposons and cell specification

Recently, it has been proposed that companion cells such as the vegetative cell of the male gametophyte and the central cell of the female gametophyte undergo active de-methylation followed by the activation of transposable elements. Subsequently, transposon-derived small interfering RNAs can move to the gametes, the sperm cell and the egg cell, to reinforce silencing of transposons in gametes, zygote and the derived next generation. This has been indicated as a fundamental biological process of reproductive biology in plant, and likely in animals as well (Slotkin et al., 2009; Calarco and Martienssen, 2011). In the female gametophyte DEMETER is required for this active demethylation process (Choi et al, 2002); a corresponding gene product in the male gametophyte has been predicted, but is still unknown. This might lead to the proposal that ETs acting as regulators of DNA methylation might be involved in this basic reproductive pathway.
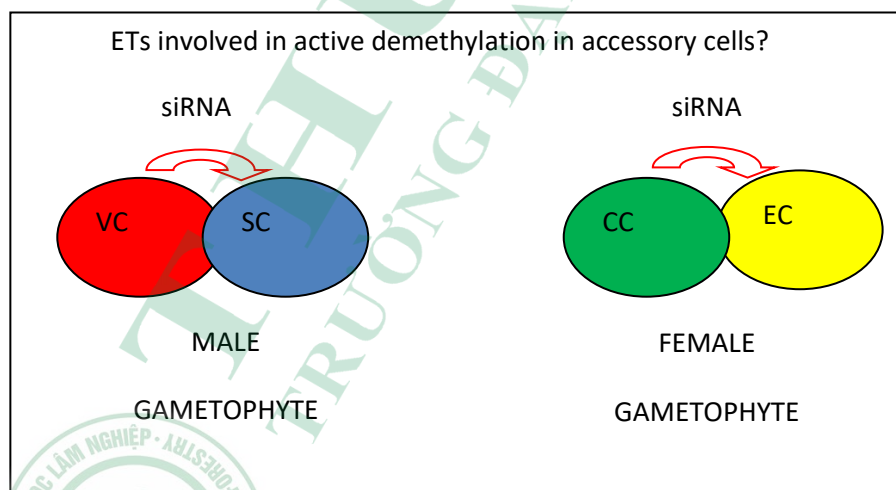

Fig. 2.14. Hypothetical function of ETs in accessory reproductive cells in plant gametophytes.

VC-vegetative cell, SC-sperm cell, CC-central cell, EC-egg cell.

2.6. CONCLUSSION

The results of DNA methylation analyses showed that *et1-1* revealed 352 highly differentially methylated regions, *et2-3* and *et1-1/et2-3* revealed 373 and 275 highly differentially methylated regions, respectively. And the highly differentially methylated regions of *et* mutants dectected loss of methylations mainly in the symmetric CG context. There are two situations of methylation *et* mutants: The hypermethylation of *et* mutants at AT1G26400 and AT1G34245 belong to regions for RdDM pathway, whereas, Demethylation of *et* mutants located on chromosome 1 such as Helitron 1 element and psORF (AT5G35935) Which identified as the second region belong to NERD pathway. For the most of hDMRs, *et1-1*, *et2-3* and *et1-1/et2-3* had overlapping and specific target regions. Summarizing the described results it is concluded that ET factors represent novel plant specific epigenetic regulators of reproductive tissue development acting on DNA-methylation.

REFERENCES

1. Bailey TL.,Williams N., Misleh C., Li W.W. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research 34*: W369–W373

2. Becker, C., J. Hagmann, J. Müller, D. Koenig, O. Stegle, K. Borgwardtand D. Weigel(2011). Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* 480: 245-249.

3. Calarco, J. P. and R. A. Martienssen (2011). Genome reprogramming and small interfering RNA in the Arabidopsis germline. *Curr.Opin.Genet.Dev.* 21: 134-139.

4. Choi, Y., M. Gehring, L. Johnson, M. Hannon, J. J. Harada, R. B. Goldberg, S. E. Jacobsen and R. L. Fischer (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. *Cell* 110: 33-42.

5. Choi, Y., J. J. Harada, R. B. Goldberg and R. L. Fischer (2004). An invariant aspartic acid in the DNA glycosylase domain of DEMETER is necessary for transcriptional activation of the imprinted MEDEA gene. *Proc.Natl. Acad. Sci.* 101: 7481-7486.

6. Choi Y., M. Gehring, L. Johnson, M. Hannon, J.J. Harada, R.B. Goldberg, S.E. Jacobsen, R.L. Fischer (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. *Cell* 110:33–42.

7. Cokus, S. J., S. Feng, X. Zhang, Z. Chen, B. Merriman, C. D. Haudenschild, S. Pradhan, S. F. Nelson, M. Pellegrini and S. E. Jacobsen (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452: 215-219.

8. Gehring, M., W. Reik and S. Henikoff (2009). DNA demethylation by DNA repair. *Trends Genet.* 25: 82–90.

9. Gong Z., Morales-Ruiz T., Ariza R.R., Roldan-Arjona T., David L., Zhu J.K. 2002. ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell* 111: 803–814.

10. Grau J., Posch S., Grosse I., Keilwagen J. 2013. A general approach for discriminative de novo motif discovery from high-throughput data. *Nucleic Acids Research* 41: e197.

11. Haince, J. F., M. Rouleau and G. G. Poirier (2006). Gene expression needs a break to unwind before carrying on. *Science* 312: 1752-1753.

12. Hagmann J., Becker C., Muller J., Stegle O., Meyer R.C.,Wang G., Schneeberger K., Fitz J., Altmann T., Bergelson J. et al. 2015. Century-scale methylome stability in a recently diverged Arabidopsis thaliana lineage. *PLoS Genetics* 11: e1004920

13. He, X. J., T. Chen and J. K. Zhu (2011). Regulation and function of DNA methylation in plants and animals. *Cell Research* 21: 442-465.

14. Ivanov R., J. Tiedemann, A. Czihal, H. Baumlein (2012). Transcriptional regulator AtET2 is required for the induction of dormancy during late seed development. *Journal of Plant Physiology* 169: 501–508.

15. Ivanov R., J. Tiedemann, A. Czihal, A. Schallau, L. H. Diep, H.P. Mock, B. Claus, A. Tewes, H. Baumlein (2008). EFFECTOR OF TRANSCRIPTION2 is involved in xylem differentiation and includes a functional DNA single strand cutting domain. *Developmental Biology* 313:93–106.

16. Jaenisch, R. and A. Bird (2003). Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature Genetics* 33: 245-254.

17. Ju, B. G., V. V. Lunyak, V. Perissi, I. Garcia-Bassets, D. W. Rose, C. K. Glass and M. G. Rosenfeld (2006). A topoisomerase IIß-mediated dsDNA break required for regulated transcription. *Science* 312: 1798-1802.

18. Kankel M.W., D.E. Ramsey, T.L. Stokes, S.K. Flowers, J.R. Haag, J.A. Jeddeloh, N.C. Riddle, M.L. Verbsky, E.J. Richards (2003). Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics* 163: 1109–1122.

19. Kuhlmann M., M.F. Mette (2012). Developmentally non-redundant SET domain proteins SUVH2 and SUVH9 are required for transcriptional gene silencing in Arabidopsis thaliana. *Plant Molecular Biology* 79: 623–633.

20. Köhler, C., P. Wolff and C. Spillane (2012). Epigenetic mechanisms underlying genomic imprinting in plants. *Annu. Rev. Plant Biol.* 63: 331-352.

21. Law J. A. and S. E. Jacobsen (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11: 204–220.

22. Lister, R., M. Pelizzola , R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren and J. R. Ecker (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462: 315-322.

23. Matzke, M. A. and R. A. Mosher (2007). RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* 15: 394-408.

74

24. Morales-Ruiz T., A.P. Ortega-Galisteo, M.I. Ponferrada-Marin, M.I, Ariza R.R. Martinez-Macias, T. Roldan-Arjona (2006). DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proceedings of the National Academy of Sciences*, USA 103: 6853–6858.

25. Ortega-Galisteo A.P., T. Morales-Ruiz, R.R. Ariza, T. Roldan-Arjona (2008). Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Molecular Biology* 67: 671–681.

26. Ossowski, S., K. Schneeberger, R. M. Clark, C. Lanz, N. Warthmann, D. Ossowski, S., K. Schneeberger, R. M. Clark, C. Lanz, N. Warthmann and D. Weigel (2008). Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* 18: 2024-2033.

27. Penterman J., R.Uzawa, R.L. Fischer (2007a). Genetic interactions between DNA demethylation and methylation in Arabidopsis. *Plant Physiology* 145: 1549–1557.

28. Penterman J., D. Zilberman , J.H. Huh, T.Ballinger, S. Henikoff, R.L. Fischer (2007b). DNA demethylation in the Arabidopsis genome. *Proceedings of the National Academy of Sciences*, USA 104: 6752–6757.

29. Pontier D., C.Picart, F.Roudier, D. Garcia, S. Lahmy, J. Azevedo, E. Alart, M. Laudie, W.M. Karlowski, R. Cooke (2012). NERD, a plant-specific GWp.rotein,defines an additional RNAi-dependent chromatin-based pathway in Arabidopsis. *Molecular Cell* 48: 121–132.

30. Schneeberger, K., J. Hagmann, S. Ossowski, N. Warthmann, S. Gesing, O. KohlbacherandD. Weigel (2009). Simultaneous alignment of short reads against multiple genomes. *Genome Biol.* 10: R98.1-R98.12.

31. Slotkin, R. K., M.Vaughn, F. Borges, M. Smyth D.R., J.L. Bowman, E.M. Meyerowitz (1990). Early flower development in Arabidopsis. *Plant Cell* 2: 755–767.

32. Tanurdzić, J. D. Becker, J. A. Feijó and R. A. Martienssen (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136: 461-472.

33. Stroud H., Greenberg M.V., Feng S., Bernatavichute Y.V., Jacobsen S.E. 2013. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* 152: 352–364.

34. Viswanathan, C. and Z. Jian-Kang (2009). RNA-directed DNA methylation and demethylation in plants. *Sci. China in Life Sci.* 52: 331–343.

35. Wierzbicki, A. T., J. R. Haag and C. S. Pikaard (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135: 635–648.

36. Xiao, W., M. Gehring, Y. Choi, L. Margossian, H. Pu, J. J. Harada, R. B. Goldberg, R. I. Pennell and R. L. Fischer (2003). Imprinting of the MEA polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Developmental Cell* 5: 891-901.

37. Zhai J., J. Liu, B. Liu, P. Li, B.C. Meyers, X. Chen, X. Cao (2008). Small RNA-directed epigenetic natural variation in Arabidopsis thaliana. *PLoS Genetics* 4: e1000056.

38. Zhao, D., Q. Yu , M. Chen and H. Ma (2001). The ASK1 gene regulates B function gene expression in cooperation with UFO and LEAFY in Arabidopsis. *Development* 128: 2735-2746.

39. Zhu J., A. Kapoor, V.V. Sridhar, F. Agius, J.K. Zhu (2007). The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in Arabidopsis. *Current Biology* 17:54–59.

Chapter 3

THE IDENTIFICATION OF PUTATIVE TARGET GENES BY TRANSCRIPTOME ANALYSIS USING DEEP SEQUENCING TECHNOLOGY

Bui Thi Mai Huong[1], Helmut Bäumlein[2], Lothar Altschmied[2]

1: Vietnam National University of Forestry

2: Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

3.1. ABSTRACT

To further investigate molecular basis of ATET1 and ATET2 action and to identify putative target genes as well as to find correlation between transcriptomeand methylom we have performed a deep RNA sequencing analysis of wild type compared to *et1-1*, *et2-3* and *et1-1/et2-3*. Using cDNA libraries for Next generation sequencing were received from slightly modified TruSeq RNA v2 protocol (Illumina) we have got the results of the 337, 330 and 486 differentially expressed genes for *et1-1*, *et2-3* and *et1-1/et2-3* compared to Col-0, respectively. And the upregulated genes were coincided with the loss of CG methylation, whereas the dowregulated genes were associated with the gain of methylation at *et1-1*, *et2-3* and *et1-1/et2-3*. Moreover, the complementation of the *et1-1* mutant after transformation of a 4.5kb genomic wild type fragment was used confirm the function of ET1 as well as demonstrate the relation between genotypes and phenotypes.

3.2. BACKGROUND

RNA sequencing is the application of next-generation sequencing techniques to study RNA which usues next-generation sequencing to reveal the presence and quantity of RNA in a biological sample at a given moment, analyzing the continuously changing cellular transcriptome. (Chu Y et al, 2012; Wang Z et al, 2009).

Specifically, RNA-seq facilitates the analysis of mutations and changes in gene expression or differences in gene expression in different groups or treatments.

RNA-Seq is a recently developed approach to transcriptome profiling that uses deep-sequencing technologies. The key aims of transcriptomics are: to catalogue all species of transcript, including mRNAs, non-coding RNAs and small RNAs; to determine the transcriptional structure of genes, in terms of their start sites, 5′ and 3′ ends, splicing patterns and other post-transcriptional modifications; and to quantify the changing expression levels of each transcript during development and under different conditions.

Unlike hybridization-based approaches, RNA-seq detected transcripts that correspond to existing genomic sequence. This makes RNA-Seq particularly attractive for non-model organisms with genomic sequences that are yet to be determined (Vera J.C et al, 2008). RNA-Seq can exhibit the precise location of transcription boundaries. In addition, RNA-Seq can also exhibit sequence variations (SNPs) in the transcribed regions (Cloonan N, et al., 2008; Morin R, et al., 2008).

Unlike DNA microarrays approaches, RNA-Seq does not have limit for quantification, These quantification correlates with the number of sequences obtained. Therefore, it has a large range of expression levels over which transcripts can be detected. For example, a greater than 9,000-fold range was estimated in a study that analysed 16 million mapped reads in *Saccharomyces cerevisiae*, and 40 million mouse sequence reads (Nagalakshmi U, et al., 2008; Mortazavi A, et al., 2008).

By using transcriptomic analyses Jing Zhou was provided the understanding of DEGs involved in drought stress at the transcriptome level in Giant Juncao with different drought and recovery conditions (Jing Zhou et al.,2021). Moreover, transcriptomic analysis helped to find key transcription factors associated to drought tolerance in wild papaya genoype (Estrella-Maldonado Humberto et al.,2021). And transcriptomic analysis was to determine molecular factors lipid hydrolysis, secondary cell-walls and oxidative protection associated with heat tolerance in thermal bentgrass (Yi Xu and Bingru Huang, 2018).

3.3. MATERIALS AND METHODS

3.3.1. Materials

Arabidopsis thaliana ecotypes Columbia-0 (Col) and *Wassilewskija-2* (Ws) were obtained from Gene Regulation Group (IPK, Gatersleben, Germany) and used throughout this study as wild type control experiments. T-DNA insertion lines *et1-1; et2-3; et1-1/et2-1; et1-1/et2-3* were generated as showed in chapter 1.

3.3.2. Bacterial strains

Several bacterial strains were used for different purposes such as DNA cloning, plasmid DNA amplification, sequencing etc… as showed in chapter 1.

3.3.3. Enzymes, markers, antibiotics

Enzymes, markers, antibiotics, other chemicals were used for different experiments as showed in chapter 1.

3.3.4. Commercial kits

RevertAid first strand cDNA synthesis kit, DNA labelling kit (Fermentas, Vilnius, Lithuania); RNeasy kit, Aquick PCR purification kit, QIAquick gel extraction kit, Qiagen plasmid purification kit mini (Qiagen, Hilden, Germany); TruSeq RNA library Prep Kit v2 (Illumina).

3.3.5. Vectors

Various vectors were used for DNA amplification, cloning genes into plants and other purposes as showed in chapter 1.

3.3.6. Primers and oligonucleotides

Primers for PCR and sequencing

Primer name	Sequence 5'-3'	Tm(°C)
ET1-1HuF	AAG AGA GAC GAC TAC ATT CGA ACT AAT C	68
ET1-1HuR	AGT ACC ATC TTC TAG TAA GAC TCC ACA AG	66
LBa1	TGGTTCACGTAGTGGGCCATCG	66

3.3.7. Methods

3.3.7.1. RNA deep sequencing

cDNA libraries for Next Generation Sequencing were created following a slightly modified TruSeq RNA v2 protocol (Illumina). Starting from 0.4 – 4 µg total RNA with a RIN factor ≥8 (Agilent) in 50 µl DEPC treated water, polyA+ RNA was isolated via affinity chromatography on oligo-(dT) magnetic beads and fragmented at elevated temperature (94°C, 8 min) using divalent cations. First strand cDNA was synthesized (25°C, 10 min; 42°C, 50 min; 70°C, 15 min) using random hexamer primers and Superscript II reverse transcriptase (Invitrogen), followed by second strand synthesis (16°C, 60 min) and purification on magnetic AMPure XP beads (Beckman; PEG precipitation on bead surface, 2x EtOH wash, elution in 10 mM Tris-HCl, pH 8.5). After blunting of cDNA fragments (30°C, 30 min), purification on AMPure XP beads, A-tailing (37°C, 30 min), and ligation of Y-shaped adapters containing the respective index sequence as well as the P5 and P7 sequences for hybridization to the inner surfaces of an Illumina flowcell, the libraries were purified on AMPure XP beads and amplified using the P5 and P7 sequences as primers (98°C, 30 sec; 15x [98°C, 10 sec; 60°C, 30 sec; 72°C, 30 sec]; 72°C, 5 min). QiaQuick (Qiagen) purified libraries were applied to a 2% agarose gel stained with SYBR-Gold (Life Technologies; illumination with a Dark Reader [Clare Chemical Research]). After electrophoresis regions between 300 and 400 bp were cut from the gel and cDNA libraries were purified via MinElute spin columns (Qiagen). The average fragment length of cDNA libraries were determined on an Agilent Bioanalyzer and their concentrations were calculated from qPCR reactions with cDNA libraries of known concentrations (known cluster densities on Illumina flowcells) as references. Libraries were denatured and diluted as recommended by Illumina, applied to a flowcell and sequenced. This work has been performed in collaboration with Dr. L. Altschmied, IPK Gatersleben.

3.3.7.2. Complementation assay

An ET1 genomic fragment including 1 kb upstream and 500 bp downstream sequence was PCR amplified using platinum Taq polymerase and resequenced. The gene fragment cloned into pDONR/Zeo using the BP reaction and further transferred into the pBGW destination vector using the LR reaction. Finally, the gene was transformed into *Agrobacterium tumefaciens* strain pGV2260 and used for Arabidopsis transformation (Col wild type and *et1-1* mutant). The presence of the transgene in *et1-1* plants was verified by PCR.

3.4. RESULTS

3.4.1. Comparative transcriptome analysis of et mutants

To identify putative target genes we have performed a deep RNA sequencing analysis of wild type compared to *et1-1*, *et2-3* and *et1-1/et2-3*. We used 10 days old seedlings. Triplicate strand-specific cDNA libraries of *Col-0*, *et1-1*, *et2-3* and the double mutant *et1-1/et2-3* carried out between 13.7 and 23.3 million short reads, of which, after adapter and quality trimming, 7.2–12.3 million reads mapped in sense orientation onto annotated, nuclear gene models in the genome of *A. thaliana* (TAIR10, Table S2). PCA (principal component analysis) of mapped read counts revealed reliable separation of the mutant samples and showed that mRNA abundance of the double mutant was more similar to *et2-3* than to *et1-1* (Fig. 3.1a). DEGs were identified for pairwise comparisons between Col-0 and *et* mutants. In total, 337, 330 and 486 DEGs were detected for *et1-1, et2-3* and *et1-1/ et2-3* comparing to Col-0, respectively (Fig. 3.1b) with a false discovery rate (FDR) ≤ 0.01 and an absolute log2 fold change (lg2FC) ≥ 1. The number of transcripts downregulated in *et* mutants (*et1-1*, 193; *et2-3*, 240; *et1-1/ et2-3*, 329) were always larger than the number of upregulated ones (*et1-1*, 144; *et2-3*, 90; *et1-1/et2-3*, 157). There are 185 transcripts differentially regulated in the *et1-1/et2-3* double mutant, which demonstrate that interactions between regulatory pathways influenced by ET1 and ET2 define gene sets not affected in the single mutants. All transcripts affected in two (*et1-1* and *et2-3*, 142; *et1-1* and *et1-1/et2-3*, 174; *et2-3* and *et1-1/et2-3*, 256) or all three mutants (129) showed an overlapping function in the different lines. In 129 transcripts significantly influenced in all three mutants showed that there are 56 transcripts, the effects of *et1-1* and *et2-3* were additive, while there are 72 transcripts the influence of one mutation was modulated by the other. This suggested epistatic interactions, similar to what we observed for DNA methylation effects.
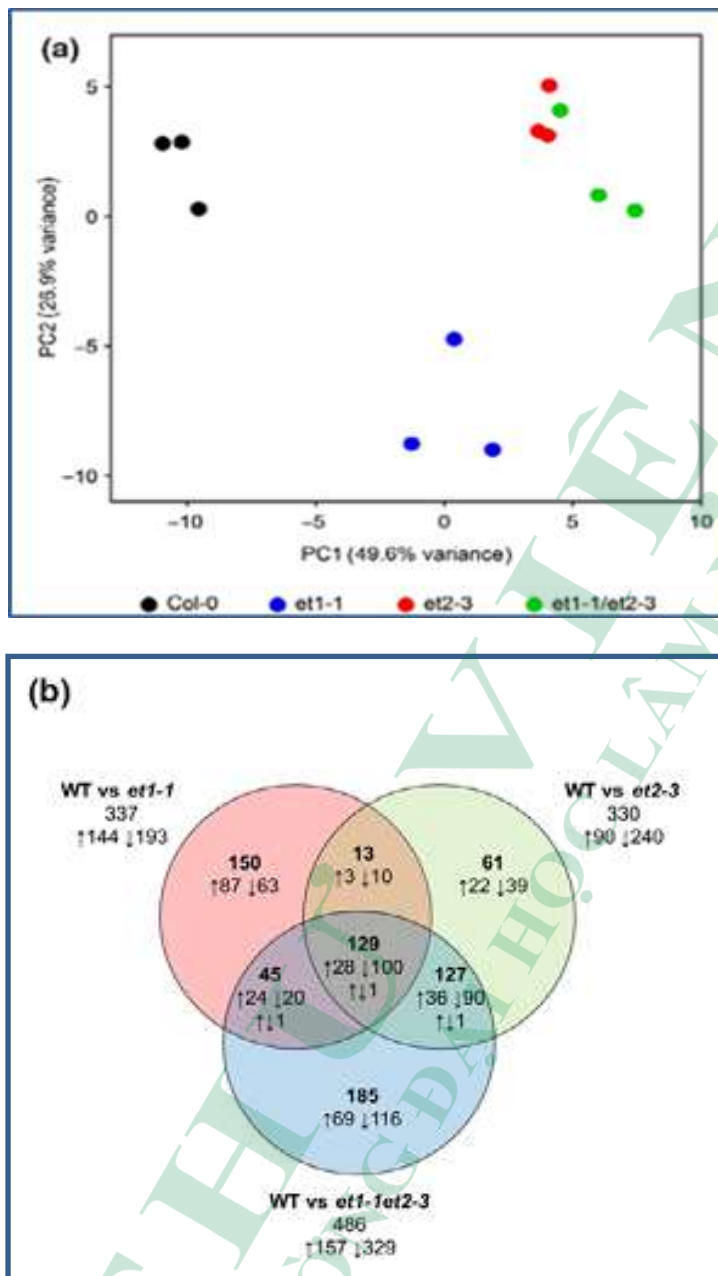
Fig. 3.1. Differentially expressed genes (DEGs) in Arabidopsis thaliana et mutant flowers.

(a) Principal component (PC) analysis of flower transcriptome. The plot shows the transcriptome data of *et1-1*, *et2-3* and *et1-1/et2-3* mutants and Col-0 (WT) in triplicate. (b) Venn diagram of DEGs between Col-0 (WT) and mutant flowers *(Francesca Tedeschi, Bui Thi Mai Huong, 2019)*.

Their differential expression relative to wild type has been mapped on the genome. As shown in Fig. 3.2 about 2/3 of these genes are downregulated genes in *et* mutants.
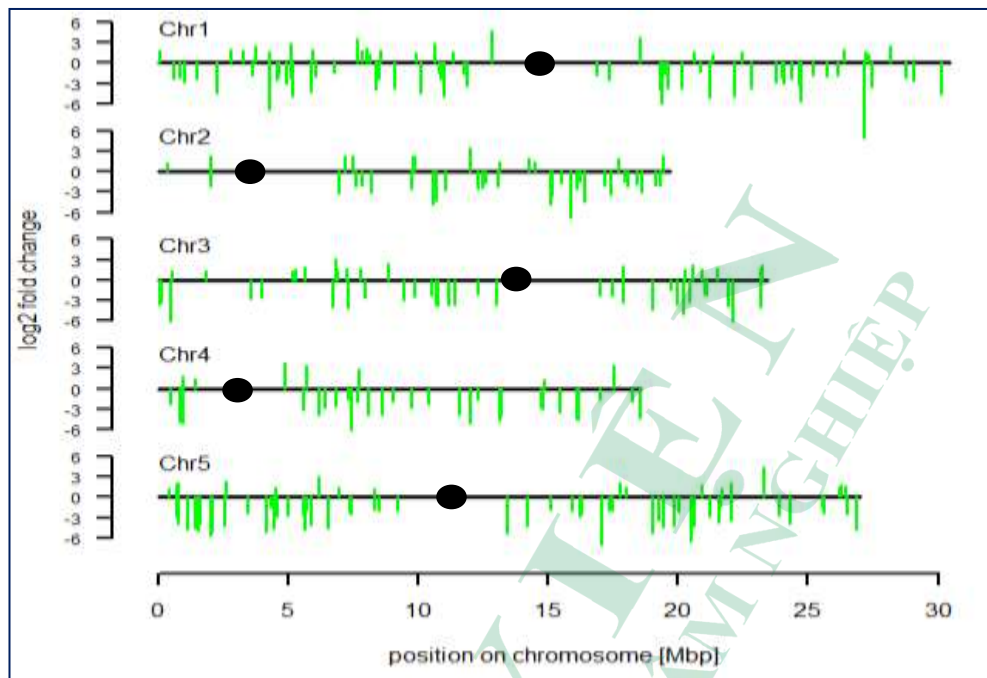
Fig. 3.2. Positions and levels of expression of differentially expressed genes in et mutants compared to wild type. About 2/3 of the genes are downregulated in et mutants.

Tab. 3.1. List of up-regulated and down-regulated genes in et1-1, et2-3 and et1-1/et2-3 mutant relative to wild type.

AGI	log2R	Annotation
et1-1		
AT2G09187	3.851	Athila6 transposable element gene
AT4G20370	3.573	TSF, PEBP family protein
AT1G53480	3.371	mto 1 responding down 1
AT1G08930	3.229	Major facilitator superfamily protein
AT4G13340	3,21	LRR family protein
AT5G48850	3.153	Tetratricopeptide repeat (TPR)-like superfamily protein
AT1G65480	3.091	FT, PEBP family protein
AT5G05365	3.072	Heavy metal transport/detoxification superfamily protein
AT2G44460	2.819	Beta glucosidase 28
AT3G62150	2.732	P-glycoprotein 21
AT4G37800	2.536	Xyloglucan endotransglucosylase/hydrolase 7
AT3G02380	-2.362	CONSTANS-like 2

AGI	log2R	Annotation
AT2G21320	-2.447	B-box zinc finger family protein
AT1G66725	-2.538	MIR163; miRNA
AT5G43630	-2.552	Zinc knuckle (CCHC-type) family protein
AT4G26170	-2.662	ET1 (TAIR:AT5G56780.1)
AT5G28030	-2.680	L-cysteine desulfhydrase 1
AT3G17609	-2.768	HY5-homologue
AT1G02820	-2.971	Late embryogenesis abundant 3 (LEA3) family protein
AT3G09450	-3.348	Fusaric acid resistance protein (TAIR:AT2G28780.1)
AT4G34550	-4.297	F-box family protein
	et2-3	
AT4G20370	3.160	TSF, PEBP family protein
AT1G07050	2.874	CCT motif family protein
AT1G08930	2.871	Major facilitator superfamily protein
AT5G56780	2.835	Effector of transcription2
T4G15690	2.722	Thioredoxin superfamily protein
AT1G56150	2.628	SAUR-like auxin-responsive protein family
AT3G18550	2.615	TCP family transcription factor
AT1G65480	2.454	FT, PEBP family protein
AT1G15010	2.401	Unknown protein AT2G01300.1
AT3G13061	-2.696	Other RNA, put. nat. antisense RNA
AT5G58770	-2.737	Undecaprenyl pyrophosphate synthetase family protein
AT3G48320	-2.797	Cytochrome P450, family 71, subfamily A, polypeptide 21
AT3G56290	-2.807	Unknown protein
AT5G66300	-2.812	NAC domain containing protein 105, VND3
AT3G58990	-2.835	Isopropylmalate isomerase 1
AT1G29920	-3.109	Chlorophyll A/B-binding protein 2
AT1G02820	-4.125	Late embryogenesis abundant 3 (LEA3) family protein
AT4G34550	-4.519	F-box family protein (TAIR:AT2G16365.3
AT3G09450	-4.747	Fusaric acid resistance protein (TAIR:AT2G28780.1
	et1-1/et2-3	
AT5G64120	3,61	Peroxidase superfamily protein

AGI	log2R	Annotation
AT1G65480	3,52	FT, PEBP family protein
AT4G20370	3,42	TSF, PEBP family protein
AT3G57460	3,13	Catalytics;metal ion binding
AT1G07050	2,74	CCT motif family protein
AT1G08930	2,72	Major facilitator superfamily protein
AT3G11340	2,71	UDP-Glycosyltransferase superfamily protein
AT1G56150	2,66	SAUR-like auxin-responsive protein family
AT2G09187	2,57	Athila6 transposable element gene
AT3G26210	2,56	Cytochrome P450, family 71, subfamily B, polypeptide 23
AT4G26170	-2,79	ET1 (TAIR:AT5G56780.1)
AT2G31380	-2,80	Salt tolerance homologue
AT3G09450	-2,85	Fusaric acid resistance protein (TAIR:AT2G28780.1)
AT5G52310	-2,98	Low-temperature-responsive protein 78 (LTI78/RD29A)
AT1G29920	-3,05	Chlorophyll A/B-binding protein 2
AT1G18330	-3,05	Homeodomain-like superfamily protein
AT2G42540	-3,06	Cold-regulated 15a
AT1G02820	-3,71	Late embryogenesis abundant 3 (LEA3) family protein
AT4G25470	-3,73	C-repeat/DRE binding factor 2
AT4G34550	-4,40	F-box family protein (TAIR:AT2G16365.3)

Genes with overlapping regulation in *et1-1*, *et2-3* and *et1-1/et2-3* given in purple; Genes with overlapping regulation in both mutants *et1-1* and *et2-3* given in red; in *et1-1* and *et1-1/et2-3* given in yellow; in *et2-3* and *et1-1/et2-3* given in blue. Note the strong upregulation of various transposable elements exclusively in the *et1-1* mutant.

For example a list of some **up**- and **down**-regulated genes of *et1-1* line and *et2-3* line and *et1-1/et2-3* line are given in table 3.1.

3.4.2. Functional analysis of DEGs (Differentially expressed genes)

Based on the focus of our study, they were selected several genes for individual inspection. This includes the ET-gene family and the top 10 DEGs (Table 3.2). We inspected the top 10 DEGs for correlation of DNA methylation difference. hDMR and DMR lists from Table 4.2 were used, and these genes were tested for minor changes in DNA methylation among the triplicates (vDMR, visually detected differential methylated regions and single methylation polymorphisms SMPs), visually detected single methylation polymorphisms). ET1 expression is lower than ET2 and ET2 shows a peak of expression in flower tissues. In the *et1-1* T-DNA insertion line as well as in the double mutant, 0–2 reads per million (RPM) are detected which were located downstream of the

insertion site, confirming the absence of functional mRNA. Consequently ET1was found in the list of downregulated genes in *et1-1*. An increase of DNA methylation upstream of the second intron associated with the *et1-1* T-DNA insertion was detectable (DMR1409). In the *et2-3* mutant expression of ET1 was not significantly different from that in Col-0.

In contrast to the qPCR results, ET2 was found in the top 10 lists of upregulated genes in *et2-3* and the double mutant (lg2FC = 2.8). This might be caused by the integrated pROK2-derived T-DNA in the used SALK_151861 line leading to 35S promotor-driven ectopic transcription (Daxinger et al., 2008). Inspection of reads and subsequent sequencing of the *et2-3* ET2 gene revealed a 24 bp deletion at position 1203 in the third exon and confirmed the T-DNA insertion located in the first exon 85 bp after the start ATG. A potential alternative translation start 869 bp after start ATG of the gene might lead to expression of a truncated version without DNA cleavage domain. Therefore, absence of functional full-length ET2 mRNA in the analysed *et2-3* T-DNA insertion plants could be confirmed. The differential expression was associated with the DMRs 1890 and 1891, located within the coding region of ET2 showing reduction of methylation.

Table 3.2. Top 10 list of differential expressed genes in
et mutant flowers obtained by RNA sequencing

AGI	*et1-1*/Col-0	DNA methylation	*Annotation*
AT4G34550	-4.297	SMP coding region	F-box family protein
AT3G09450	-3.348	SMP coding region	Fusaric acid resistance protein (TAIR:AT2G28780.1)
AT1G02820	-2.971	No methylation	Late embryogenesis abundant 3 (LEA3) family protein
AT3G17609	-2.768	No methylation	HY5-homologue
AT5G28030	-2.680	SMP coding region	L-cysteine desulfhydrase 1
AT4G26170	-2.662	DMR1409 coding region	ET1 (TAIR:AT5G56780.1)
AT5G43630	-2.552	SMP coding region	Zinc knuckle (CCHC-type) family protein
AT1G66725	-2.538	No methylation	MIR163; miRNA
AT2G21320	-2.447	No methylation	B-box zinc finger family protein
AT3G02380	-2.362	SMP coding region	CONSTANS-like 2
AT4G37800	2.536		Xyloglucan endotransglucosylase/ hydrolase 7
AT3G62150	2.732	SMP	P-glycoprotein 21

AGI	*et1-1*/Col-0	DNA methylation	*Annotation*
AT2G44460	2.819	vDMR coding, promotor	Beta glucosidase 28
AT5G05365	3.072	No methylation	Heavy metal transport/detoxification superfamily protein
AT1G65480	3.091	SMP promotor	FT, PEBP family protein
AT5G48850	3.153	SMP coding region	Tetratricopeptide repeat (TPR)-like superfamily protein
AT1G08930	3.229	vDMR 3' region	Major facilitator superfamily protein
AT1G53480	3.371	vDMR coding	mto 1 responding down 1
AT4G20370	3.573	vDMR promotor	TSF, PEBP family protein
AT2G09187	3.851	hDMR243 Athila6	Athila6 transposable element gene
AGI	*et2-3*/Col-0	DNA methylation	*Annotation*
AT3G09450	-4.747	SMP coding region	Fusaric acid resistance protein (TAIR:AT2G28780.1)
AT4G34550	-4.519	SMP coding region	F-box family protein (TAIR:AT2G16365.3)
AT1G02820	-4.125	No methylation	Late embryogenesis abundant 3 (LEA3) family protein
AT1G29920	-3.109	vDMR coding	Chlorophyll A/B-binding protein 2
AT3G58990	-2.835	SMP 30 region	Isopropylmalate isomerase 1
AT5G66300	-2.812	hDMR865 coding 5'	NAC domain containing protein 105, VND3
AT3G56290	-2.807		Unknown protein
AT3G48320	-2.797	SMPs coding region	Cytochrome P450, family 71, subfamily A, polypeptide 21
AT5G58770	-2.737	SMPs coding region	Undecaprenyl pyrophosphate synthetase family protein
AT3G13061	-2.696	SMPs coding region	Other RNA, put. nat. antisense RNA
AT1G15010	2.401		Unknown protein AT2G01300.1
AT1G65480	2.454	SMP promotor	FT, PEBP family protein
AT3G18550	2.615		TCP family transcription factor
AT1G56150	2.628	SMP coding region	SAUR-like auxin-responsive protein family

AGI	*et1-1*/Col-0	DNA methylation	*Annotation*
AT4G15690	2.722	No methylation	Thioredoxin superfamily protein
AT5G56780	2.835	SMP, DMR1890	Effector of transcription2
AT1G08930	2.871	vDMR 3'	Major facilitator superfamily protein
AT1G07050	2.874	No methylation	CCT motif family protein
AT5G65080	2.936	SMP 3'	K-box region,MADS-box transcription factor family protein
AT4G20370	3.160		TSF, PEBP family protein
AGI	***et1-1/et2-3*/Col-0**	**DNA methylation**	*Annotation*
AT4G34550	-4.399	SMP coding region	F-box family protein (TAIR:AT2G16365.3)
AT4G25470	-3.732	No methylation	C-repeat/DRE binding factor 2
AT1G02820	-3.710	No methylation	Late embryogenesis abundant 3 (LEA3) family protein
AT2G42540	-3.065	vDMR promotor/5'	Cold-regulated 15a
AT1G18330	-3.051		Homeodomain-like superfamily protein
AT1G29920	-3.049	vDMR coding region	Chlorophyll A/B-binding protein 2
AT5G52310	-2.984		Low-temperature-responsive protein 78 (LTI78/RD29A)
AT3G09450	-2.851	SMP coding region	Fusaric acid resistance protein (TAIR:AT2G28780.1)
AT2G31380	-2.803		Salt tolerance homologue
AT4G26170	-2.796	vDMR coding region	ET1 (TAIR:AT5G56780.1)
AT3G26210	2.557	SMP Promotor	Cytochrome P450, family 71, subfamily B, polypeptide 23
AT2G09187	2.568	hDMR243	Athila6 transposable element gene
AT1G56150	2.663	SMP coding region	SAUR-like auxin-responsive protein family
AT3G11340	2.710	vDMR coding region	UDP-Glycosyltransferase superfamily protein
AT1G08930	2.719	vDMR 3'	Major facilitator superfamily protein

AGI	*et1-1*/Col-0	DNA methylation	*Annotation*
AT1G07050	2.742	No methylation	CCT motif family protein
AT3G57460	3.129	hDMR528 Promotor	Catalytics;metal ion binding
AT4G20370	3.415	vDMR promotor	TSF, PEBP family protein
AT1G65480	3.521	SMP promotor	FT, PEBP family protein
AT5G64120	3.612	vDMR coding region	Peroxidase superfamily protein
AT1G07050	2.742	No methylation	CCT motif family protein

SMP: single methylation polymorphism, DMR: differential methylated region, vDMR: visual detected differential methylated region, hDMR: differential methylated region with high significance.

The upregulated gene in *et1-1* and the *et1-1/ et2-3* double mutant was AT2G09187 (lg2FC = 6.54 in *et1-1* and lg2FC = 6.52 in the double mutant), it is a transposable element gene and matching the annotated transposable element AT2TE15880 from the Athila6A family. It was confirmed upregulation in the *et1-1* mutant, *et2-3* mutant and *et1-1/et2-3* mutant by qRT-PCR (Fig. 3.3a). The upregulation of AT2TE15880 is specific for the *et1-1* mutant, indicating a functional difference between both ET mutants concerning the regulation of this transposon. The induced transcript overlapped with the highly differentially methylated region hDMR165 (Fig. 3.3b), with CG methylation loss specific to *et1-1* and *et1-1/et2-3* but not in *et2-3*.

The downregulated gene in *et2-3* and *et1-1/ et2-3* flowers, which showed lg2FC = -2.8 higher transcript abundance is VND3 (AT5G66300, Fig. 3.3c). VND3 is a VASCULAR-RELATED NAC-DOMAIN transcription factor (Yamaguchi et al., 2010; Zhou et al., 2014) associated with xylem vessel formation (Ivanov et al., 2008). Transcriptional suppression of VND3 was associated with a gain of methylation (hDMR865) at the transcriptional start site of the gene (Fig. 4.3d).

The gene not belongs to the top 10 DEGs (sense), that was AT1G64790 (antisense) showed lg2FC = 4.2 upregulated in all mutants (Fig. 3.3e), was associated with hDMR153 (Fig. 3.3f). This region is referred to as RITA (AT1G64795, encoded in antisense orientation upstream of ILYTHIA, AT1G64790), already described as a metastable DMR (Havecker et al., 2012). As mentioned above, psORF (AT5G35935), hypomethylated in both *et* mutants (hDMR750), was found to be transcriptionally activated in the mutants (Fig. 3.3g,h).
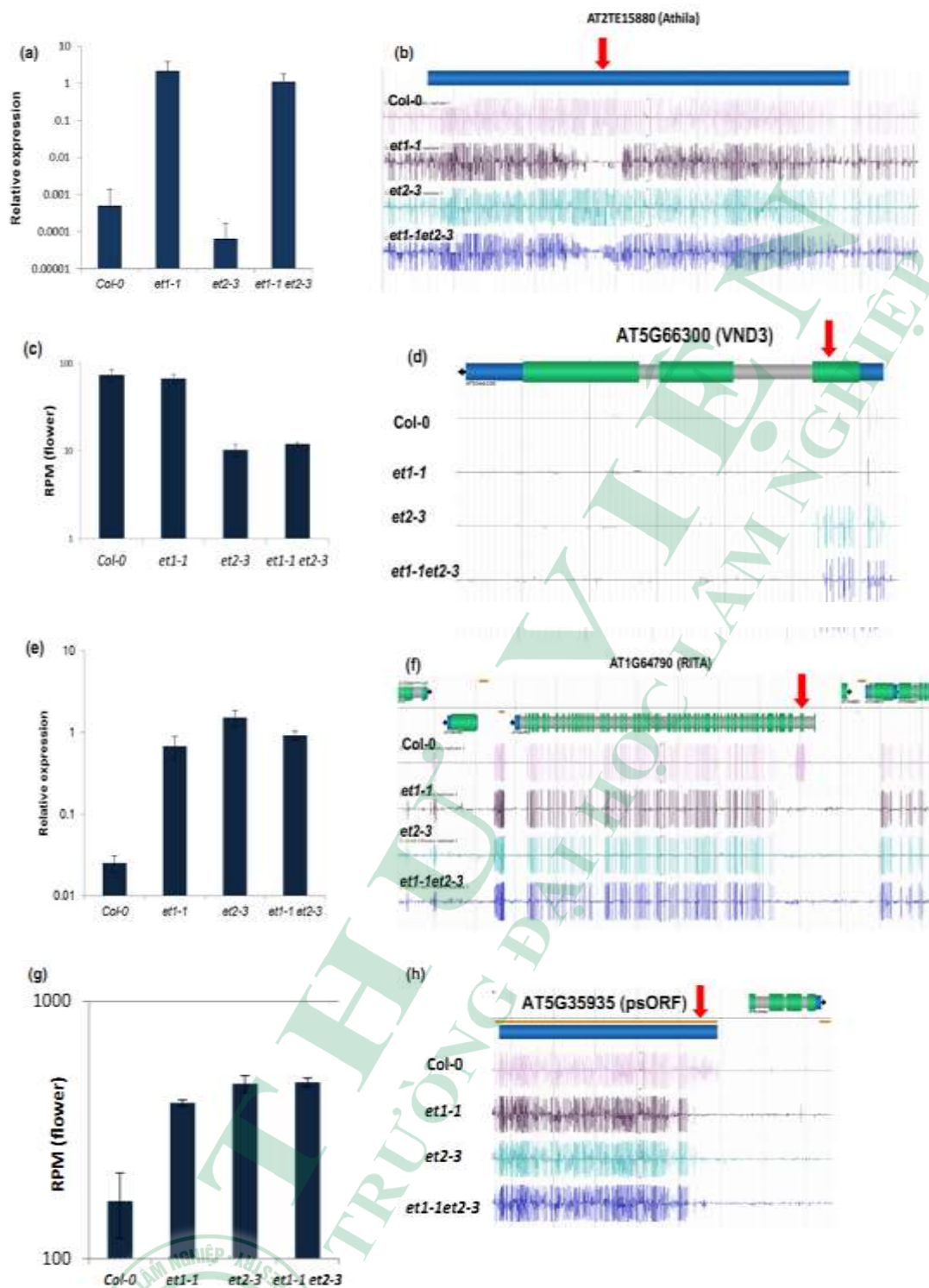
Fig. 3.3. Selected differentially expressed genes (DEGs) in Arabidopsis thaliana et mutant flowers and correlation with differentially methylated regions (DMRs).

(a) Relative expression of AT2G09187 (transcribed from AT2TE15880, Athila6A) confirmed by real-time RT-qPCR from shoot apical meristem(SAM). Bars indicate the

mean of three independent samples with SE. (b) DNA methylation signature in the region of AT2TE15880 encoding the Athila6A retroelement. Red arrow indicates hDMR165. (c) Expression analysis of flower tissue of AT5G66300 (VND3) derived fromRNA sequencing. Displayed are reads per million from three independent experiments. (d) DNAmethylation signature in the region of AT5G66300 encoding VND3 with hDMR865 (red arrow). (e) Relative expression of AT1G64795 (RITA) confirmed by real-time RT-qPCR fromSAM. (f) DNAmethylation signature in the region of AT1G64790 annotated as ILYTHIA. AT1G64795 (RITA) transcripts are antisense orientated to ILYTHIA and covering hDMR153. Blue regions, untranslated regions; green regions, translated regions; grey regions, introns; red arrow points towards the respective hDMR865. Methylation signature shown is as a representative from three independent replicates. (g) Expression analysis of flower tissue of AT5G35935 (psORF) derived fromRNA sequencing. Displayed are reads per million from three independent experiments. (h) DNA methylation signature in the region of AT5G35935 annotated as psORF; red arrow points towards the respective hDMR750 (Francesca Tedeschi, Bui Thi Mai Huong., 2019).

3.5. Complementation ET1

Genetic studies often rely on the analysis of phenotypic consequences of certain genotypic mutant alleles. Prior to a more detailed discussion of the phenotypic effects of the mutants, it is required to demonstrate the relation between genotypes and phenotypes. Principal possibilities include the analysis of multiple alleles and/or the phenotypic complementation of the genotype. The first option has been applied for *ET2* by the analysis of the mutant alleles, *et2-1* and *et2-3*, which obviously exhibit similar phenotypes. The second approach has been used for *ET1*. In this case, a 4.5 kb genomic wild type fragment has been PCR amplified, re-sequenced and transformed into the *et1-1* mutant (Fig.3.4). The phenotypic characterisation of the transformed lines indicates the partial correction of the mutant phenotypes (Fig. 3.5).
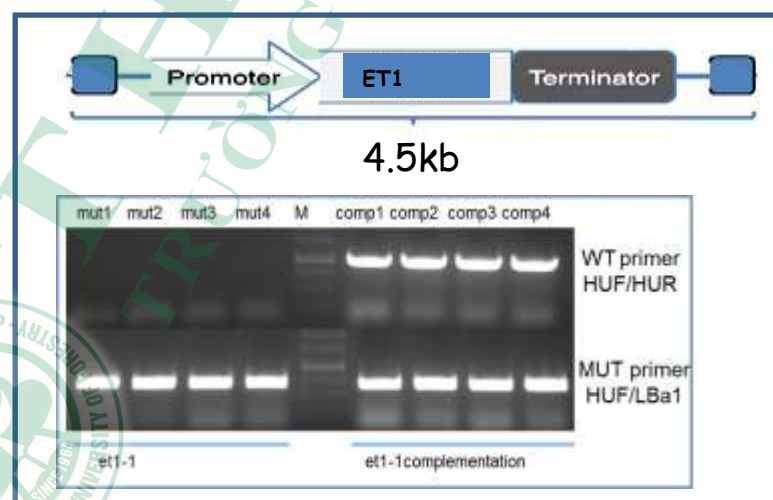

Fig. 3.4. Complementation of the et1-1 mutant after transformation
of a 4.5 kb genomic wild type fragment.

PCR analysis of four homozygous *et1-1* lines (left panel) and four complementation ET1/et1-1 using ET1-1HUF and ET1-1HUR as gene specific primers and ET1-1HUF and LBa1 to detect the T-DNA insertion.
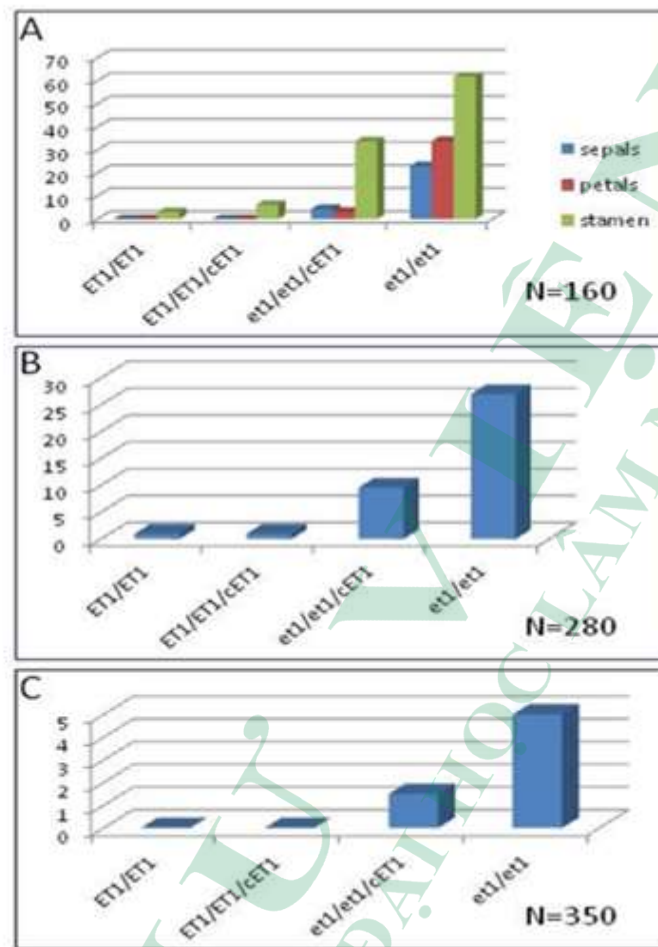

Fig. 3.5. Partial phenotypic complementation of the et1-1 mutant after transformation of a 4.5 kb genomic wild type fragment.

Partial correction of effects on distorted flower organ numbers (A), partial complementation of the described effects on enlarged nucleoli in endosperm nuclei (B) and partial correction of the effects on non-fused polar nuclei in the gametophyte (C).

In ET1 Complementation, for the majority of upregulated genes in *et1-1*, namely AT5G48850 (SDI1), AT1G65480 (FT), AT2G44460 (BGLU28), AT4G31800 (WRKY18), AT5G40360 (MYB115), AT2G09187 (Athila6A) and the 50 located antisense transcripts of AT1G64790 (ILLITHYA) RITA (Havecker et al., 2012), transcript level could not be restored to the Col-0 level by transgenic insertion of ProET1:ET1. For genes found to be downregulated, such as AT1G26770 (EXPA10), AT1G02820 (LEA) and AT4G27330 (SPL) restored the partial correction in Col-0-like expression (Fig. 3.6).
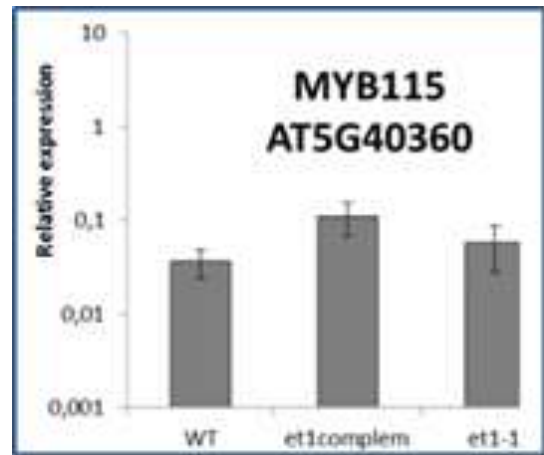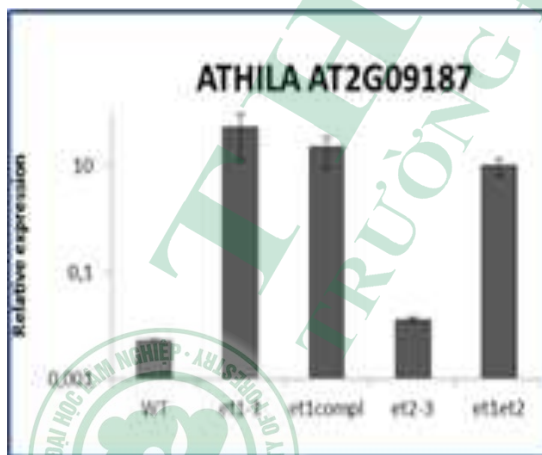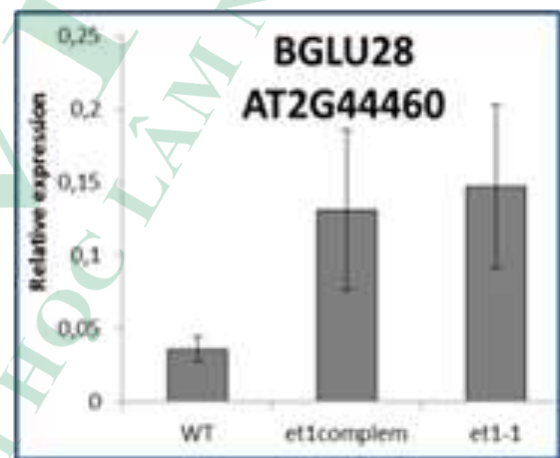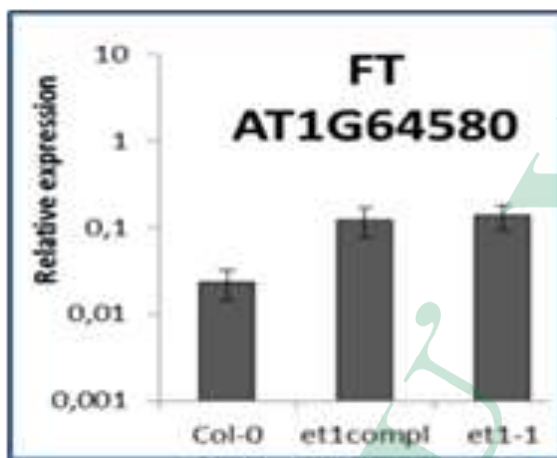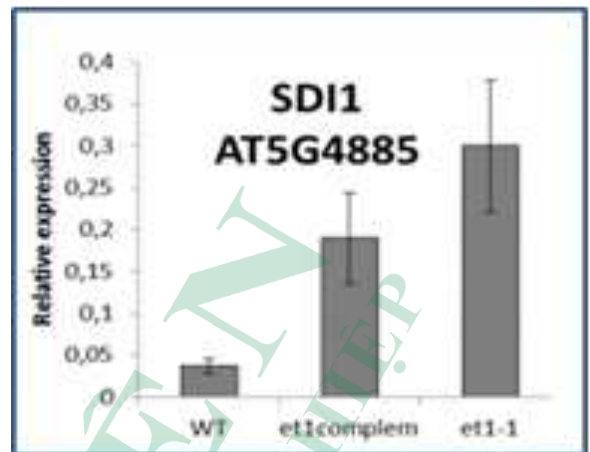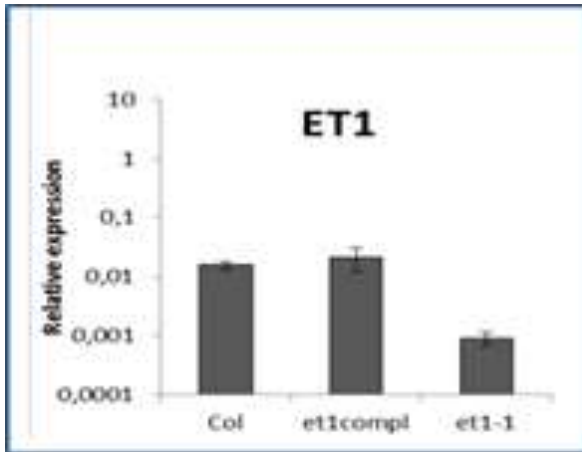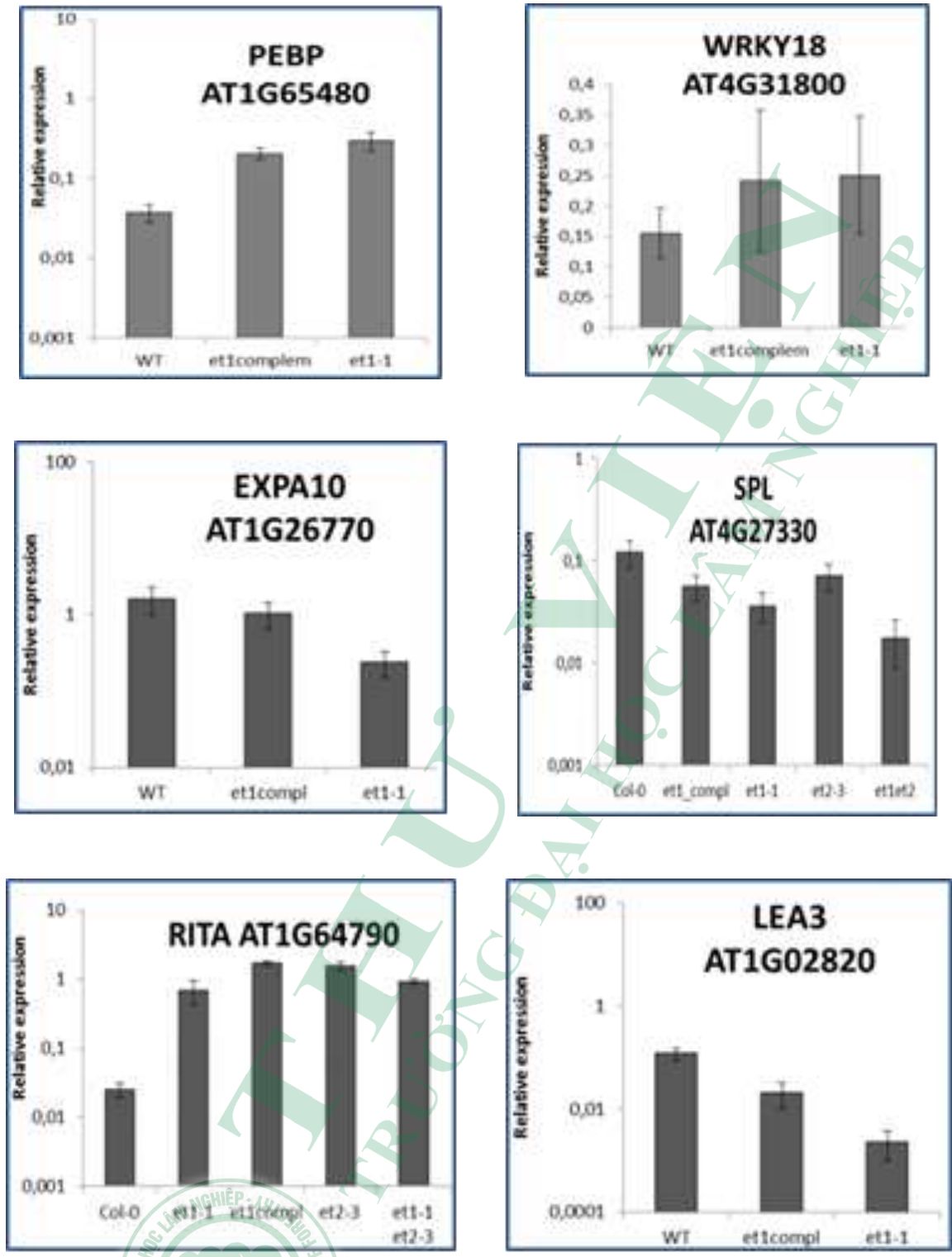
Fig. 3.6. qPCR analysis of ET1 Complementation.

Displayed are realtime qRT-PCR analysis testing selecting genes for relative mRNA abundance in the Et1/ ProET1:ET1 complemented plants (et1compl). Relative mRNA

abundance was compared to Col (Col-0, wildtype) and *et1-1*. PCRs were performed in three technical and three biologcal replicates from flower tissue. Error bars indicate the standard error .

3.6. DISCUSSION

3.6.1. RNA deep sequencing

Based on the assumption that ETs act as regulators of genomic DNA methylation with further influence on gene transcription, an RNA deep sequencing approach has been performed. In total, 337, 330 and 486 DEGs were found for the comparisons of Col-0 vs *et1-1,* Col-0 vs *et2-3*, and Col-0 vs *et1-1/ et2-3*, respectively differentially expressed genes could be identified and their expression was mapped on the genome (Fig. 3.2). About 2/3 of these differentially expressed genes are down-regulated in the mutants. This corresponds well with the initial hypothesis that ET factors might act as de-methylators. It is broadly accepted that hypo-methylated promoters lead to stronger expressed genes, whereas hyper-methylated promoters lead to suppressed gene expression. Thus, a missing or reduced de-methylation in the *et* mutants would lead to hypermethylation and suppression of gene transcription. Following this interpretation, the 1/3 up-regulated genes might be seen as secondary effects in a dynamically changing methylation and transcription pattern.

As shown in Tab. 3.1 up- and down-regulated genes strongly overlap between *et* mutants (*et1-1*, *et2-3* and *et1-1/et2-3*). With the most of genes were found to be differentially expressed in *et2-3* are also found as differentially expressed genes in the *et1-1* mutant. The most difference in the transcriptome of *et1-1* and *et2-3* mutants (see Tab. 3.1) concerns the expression of transposable element (AT2G09187) strongly upregulated in *et1-1* mutant with log2 ratios of about 3.85. The transposable elements are retrotransposon which belong to the Athila subfamily of the Ty3/Gypsy family (for a review see Wicker et al., 2007). They represent over 2.7% of the total Arabidopsis genome and are a major building block of the centromere (Slotkin, 2010). Athila retrotransposons share a large internal region of up to 10.5 kb flanked by two about 1.8 kb long terminal repeats (LTR). The internal region encodes the capsid structural protein *gag* and the protein *pol* which carries the protease, reverse transcriptase and integrase domains essential for element duplication. The silencing of Athila retrotransposons has come to the forefront of Arabidopsis small RNA regulation, the control of centromere core as well as potentially playing a role in speciation (Slotkin et al., 2009). Taken together, the data demonstrate that ET1 and ET2 have both overlapping as well as gene-specific functions with ET1 being specific for the regulation, most likely the suppression, of retrotransposons especially of the Athila subfamily.

94

3.6.2. Rare overlap between DEGs and regions with strong methylation difference (hDMRs)

Using our definition of hDMRs and DEGs, an overlap of regions and expression was barely detectable: AT2G09187 (overlapping with the Athila6A retrotransposon AT2TE15880) with hDMR165, AT5G56780 (AtET2) with hDMR517 and AT5G66300 (VND3) overlapping with hDMR865. Such rare overlap between DMRs and DEGs has been reported by several other studies (Havecker et al., 2012; Kawakatsu et al., 2016). The position of the DMR, located in a gene body or promotor, together with its genomic environment make it difficult to predict whether the change in methylation is a cause or consequence of differential expression. The rare overlap might indicate that our criteria defining an hDMR definition are very strict. As shown in the top list of DEGs, there are small regions of differential methylation as well as a number of SMPs in the genomic environment of DEGs (Table 3.2). These might lead to a difference in the amount of detectable transcripts of the respective genes in the mutants.

The over-representation of detectable hDMRs associated with coding regions indicates a mechanism which might be associated with histone modifications such as H3K9 acetylation (H3K9ac) and H3K4 trimethylation (H3K4me3) (Ha et al., 2011), and these chromatin marks might also improve the recognition of putative ET binding sites. Such influence has been described for the histone acetyltransferase IDM1 which is required for demethylation (Qian et al., 2012). In addition, the identification of the NERD-pathway target genes (Pontier et al., 2012) HELITRON1 (AT1TE93275) and psORF (AT5G35935) might indicate an association of H3K4 histone modification with ET function.

3.6.3. Correlation between transcriptome and methylom and phenotype in et mutants

The typical mature gametophyte of Arabidopsis consists of two synergids, the egg cell and the homodiploid central cell with the two nuclei fused together. However, the most obvious gametophytic phenotype observed in *et* mutants displayed the distorted fusion of the two polar nuclei. A relatively large numbers of mutations described previously share this phenotype. Such as, in the BiP1, BiP2, BiP3 (other name HSP70) mutant lines exhibited non-fused polar nuclei (Maruyama et al. 2010). We found that in *et* mutants showed hypermethylation of an adjacent gene encoded HSP70-molecular chaperons in the ER- exhibit non-fused polar nuclei, which overlaped AT3G09440 (HSP70) with log2FC=-1,29 (Fig. 3.7). Thus, This invites the speculation that ETs mediate changes in the methylation status of such genes and could contribute to the phenotype of non-fused polar nuclei. However, to answer all question we need further investigations.
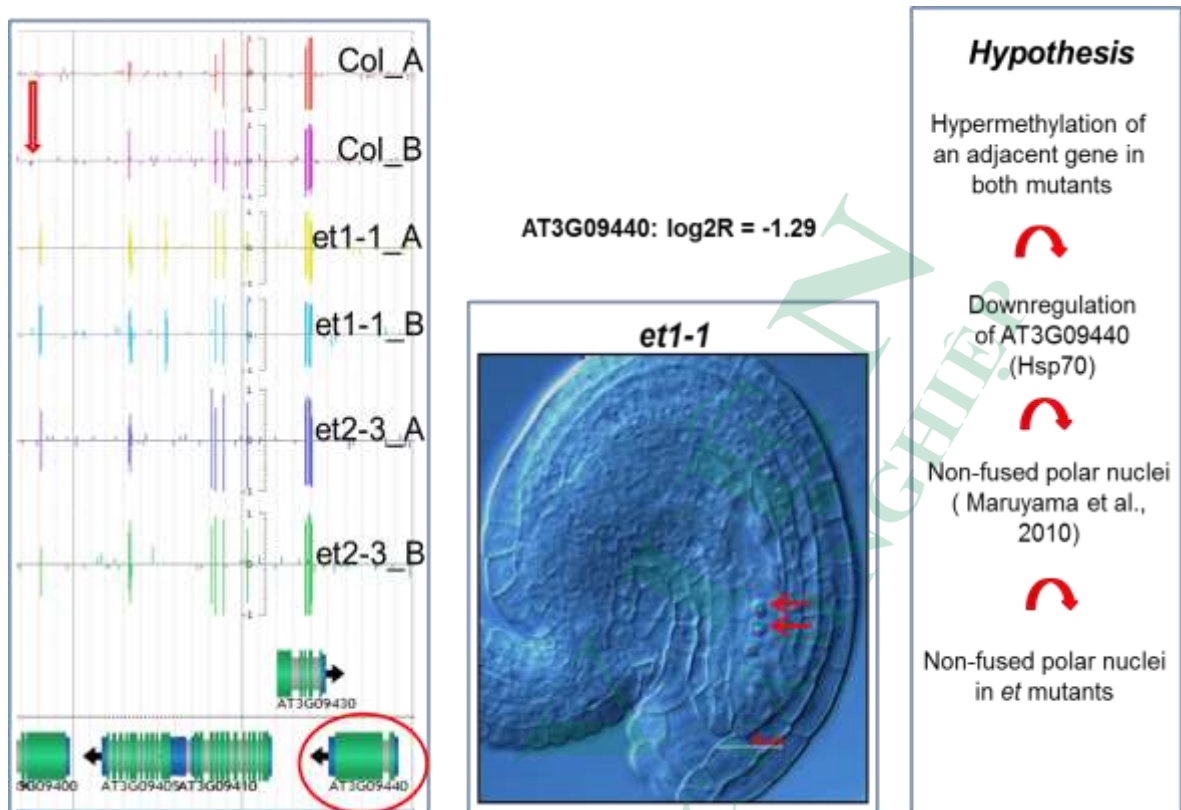
Fig.3.7. Methylation, transcription and phenotype: Hsp70 was downregulated as well as hypermethylated region in et1-1, could be related to non-fused polar nuclei

(Maruyama et al.,2010).

Specially, *et* mutants show a precocious germination phenotype both *in vivo* and *in vitro* (Fig. 2.14, Fig. 2.15 chapter 2). Usually germination is initiated when the seedling penetrates the seed coat with the radicula ahead. But, the germination of *et* mutants occurs with cotyledons first. It speculate that ETs are involved in epigenetic pathway involved in the regulation of the ratio between the phytohormones GA and ABA. This hypothesis was clarified by the hypermethylation of an adjacent gene LTP3 (AT5G59320), and this gene also downregulated with log2FC=-7.35 which was tested by RT-PCR (Fig.3.8) in *et* mutants. In addition, the gene encoding a MYB96 protein (AT5G62470) is significantly hypermethylated in the *et* mutants (Guo et al.,2013). This view fits with the well established knowledge that the ratio of both hormones is crucial for the maintenance of dormancy *versus* initiation of seed germination, with ABA favouring dormancy and GA triggering germination. Assuming that ET acts as an epigenetic repressor of GA activity (see above), indeed one would predict an early germination phenotype for *et* mutants as observed.
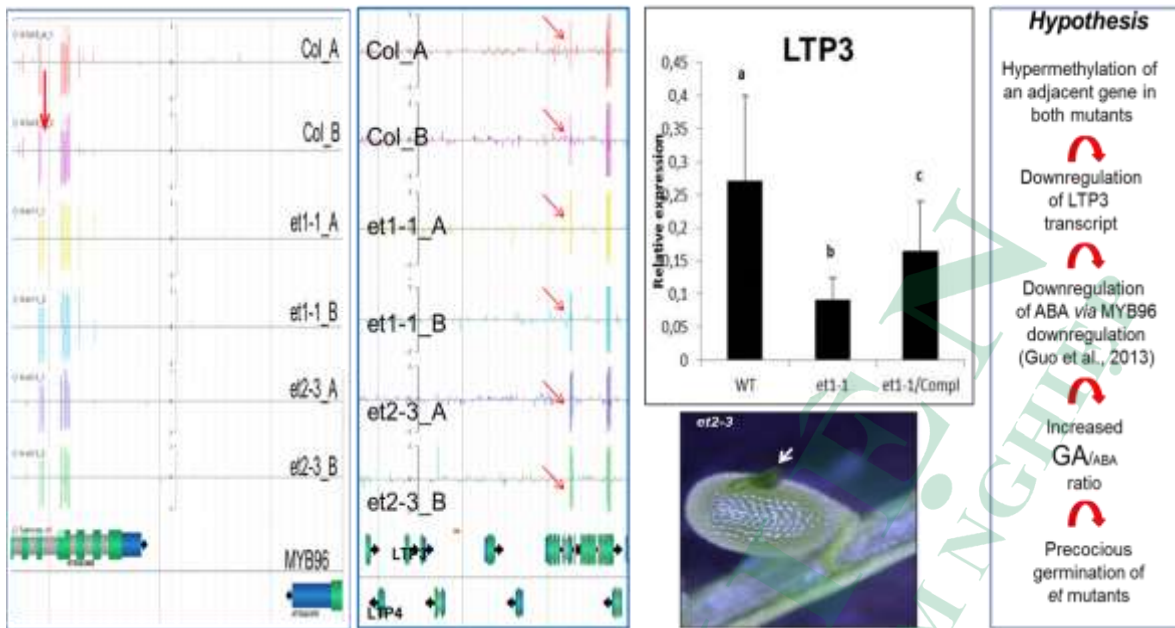
Fig.3.8. Methylation - LTP3 transcription - MYB96/ABA - precocious germination.

LTP3 was downregulated and hypermethylated region in *et1-1* which could be associated with downregulation of ABA via MYB96 (Guo et al.,2013) caused to precocious germination of *et* mutants.

3.7. CONCLUSSION

The results of transcriptome analyses showed that in total 337, 330 and 486 DEGs were detected for *et1-1*, *et2-3* and *et1-1/et2-3*, respectively. The number of transcripts downregulated in *et* mutants were larger than the number of upregulated genes. In all transcripts of *et1-1*, *et2-3* and *et1-1/et2-3* had overlapping and specific target transcripts. For the most of upregulated genes were associated with hypomethylated regions, whereas the downregulated genes coincided with hypermethylated regions. In ET1 complementation, the majority of upregulated genes in *et1-1* which could not be restored as the Col-0 level, while some downregulated genes were restored partial level like Col-0 expression. Summaring DNA methylation, transcriptome and phenotype analyses in flowers and seedling of *et* mutants revealed ET specific differentially expressed genes and ET specific differentially methylated regions. Loss of ET function results in pleiotropic developmental defects.

REFERENCES

1. Chu Y., R. C. David (2012). RNA sequencing: Platform selection, Experimental design, and data interpretation. *Nucleic Acid Ther*. 22 (4):271-274.

2. Cloonan N., R. Alistair, K. Gabriel, B. Brooke, J. F.Geoffrey, K. B.Mellissa, F.T.Darrin, L.S.Anita, W.Shivangi, B.Graeme, J.R.Alan, C.P.Andrew, J.B. Stephen, C.L. Clarence, S.R.Swati, E. P.Heather, M. M., Kevin J. M.Jonathan, M. G.Sean (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods*. 5:613–619.

3. Estrella-Maldonado H., A. G.Ramírez, G. F.Ortiz, O. M.Vega, E.Góngora-Castillo, J. M.Santamaría (2021). Transcriptomic analysis reveals key transcription factors associated to drought tolerance in a wild papaya (*Carica papaya*) genotype. *PLoS One*. 16(1): e0245855.

4. Francesca T., P. Rizzo, B.T.M. Huong, A.Czihal, T. Rutten, L. Altschmied, S. Scharfenberg, I. Grosse, C. Becker, D. Weigel, H.Bäumlein and M. Kuhlmann (2019). EFFECTOR OF TRANSCRIPTION factors are novel plant-specific regulators associated with genomic DNA methylation in Arabidopsis. *New Phytologist* (2019) 221: 261–278. doi: 10.1111/nph.15439

5. Guo L., H. Yang, X. Zhang and S. Yang (2013). Lipid transfer protein 3 as a target of MYB96 mediates freezing and drought stress in Arabidopsis. *Journal of Experimental Botany*: 64(6): 1755–1767.

6. Ha M., W.H. Li, Z.J.Chen (2011). Coordinated histone modifications are associated with gene expression variation within and between species. *Genome Research* 21: 590–598.

7. Havecker E.R., L.M.Wallbridge, P.Fedito, T.J.Hardcastle, D.C.Baulcombe (2012). Metastable differentially methylated regions within Arabidopsis inbred populations are associated with modified expression of non-coding transcripts. *PLoS ONE* 7: e45242.

8. Ivanov R, J.Tiedemann, A.Czihal, A.Schallau, L.H.Diep , H.P.Mock, B.Claus, A.Tewes, H.Baumlein (2008). EFFECTOR OF TRANSCRIPTION2 is involved in xylem differentiation and includes a functional DNA single strand cutting domain. Developmental Biology 313:93–106. *PLoS ONE* 7: e45242.

9. Kawakatsu T., J.R.Nery, R.Castanon, J.R.Ecker (2017). Dynamic DNA methylation reconfiguration during seed development and germination. *Genome Biology* 18: 171.

10. Maruyama, D., T. Endo and S. Nishikawa (2010). BiP-mediated polar nuclei fusion is essential for the regulation of endosperm nuclei proliferation in *Arabidopsis thaliana*. *Proc. Natl. Acad.Sci*. 107: 1684-1689.

11. Morin R., M. Bainbridge, A. Fejes, M. Hirst, M. Krzywinski, T. Pugh, H. Donald, R. Varhol, S. Jones, M. Marra (2008). Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* 45:81–94.

12. Mortazavi A, B.A.Williams, K.McCue, L.Schaeffer, B.Wold (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5:621–628.

13. Nagalakshmi U., Z.Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, M. Snyder (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320:1344–1349.

14. Pontier D., C. Picart , F. Roudier, D. Garcia, S. Lahmy, J. Azevedo , E. Alart, M. Laudie, W.M. Karlowski, R. Cooke (2012). NERD, a plant-specific GW protein defines an additional RNAi-dependent chromatin-based pathway in Arabidopsis. *Molecular Cell* 48: 121–132.

15. Qian W., D. Miki , H. Zhang , Y. Liu , X. Zhang, K.Tang, Y.Kan, H. La, X. Li, S. Li (2012). A histone acetyltransferase regulates active DNA demethylation in Arabidopsis. *Science* 336: 1445–1448.

16. Slotkin, R. K. (2010). The epigenetic control of Athila family of retrotransposon in Arabidopsis. *Epigenetics* 5: 483-490.

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

18. Vera J. C., C.W. Wheat, H.W. Fescemyer, M.J. Frilander, D.L. Crawford, I. Hanski, J.H. Marden (2008). Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol. Ecol* 17:1636–1647.

19. Wang Z., M. Gerstein, M. Snyder (2009). RNAseq: a revolutionary tool for transcriptomics. *Nat Rev Genet*.10 (1):57-63.

20. 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). An unified classification system for eukaryotic transposable elements. *Nature Rev.Genet*.8: 973-982.

21. Yamaguchi M., M.Ohtani, N.Mitsuda, M.Kubo, M.Ohme-Takagi, H.Fukuda, T. Demura (2010). VND-INTERACTING2, a NAC domain transcription factor, negatively regulates xylem vessel formation in Arabidopsis. *Plant Cell* 22: 1249–1263.

22. Yi Xu and Bingru Huang (2018). Transcriptomic analysis reveals unique molecular factors for lipid hydrolysis, secondary cell-walls and oxidative protection associated with thermotolerance in perennial grass. *BMC Genomics*. 2018; 19: (70). doi: 10.1186/s12864-018-4437-z

23. Zhou J., S.Chen, W.Shi, R.David-Schwartz, S.Li, F.Yang, Z.Lin (2021). Transcriptome profiling reveals the effects of drought tolerance in Giant Juncao. *BMC Plant Biol*.2021;21(2).doi: 10.1186/s12870-020-02785-7.

24. Zhou J., R.Zhong, Z.H. Ye (2014). Arabidopsis NAC domain proteins, VND1 to VND5, are transcriptional regulators of secondary wall biosynthesis in vessels. *PLoS ONE* 9: e105726.