The molecular basis of maternal control in seed development

Genetic and molecular analysis of maternal effects in seed development: molecular mapping of *cap2* and expression and functional analyses of *AtLDC* and *AtHD2C*

Nirma Škrbo

Thesis for the degree of *Cand. Scient.*
Program for Molecular Genetics
Department of Molecular Biosciences
University of Oslo, September 2005

Acknowledgements

The work presented in this thesis was carried out at the Department of Molecular Biosciences, University of Oslo, from September 2003 to September 2005. Supervisors have been Professor Reidunn B. Aalen and Dr. Paul E. Grini.

I would like to thank Professor Reidunn B. Aalen for giving me the opportunity to study in her group and for her continuous support and enthusiasm.

Special thanks to Dr. Paul E. Grini for all help, inspiration and encouragement. His enthusiasm has been a great source of motivation throughout this work. I am very grateful for his supervision in the laboratory and for his great help in the writing process.

Many thanks to my fellow students and everyone in our group for fruitful discussions in the laboratory and creating a great study environment both scientifically and socially. Special thanks to Mirela Karabeg for being such a good and supportive friend. Thanks to Bitte Stenvik for our good friendship and all the fun.

I am also very thankful to Solveig Hauge Engebretsen and Roy Fallet for help and assistance in the laboratory and the phytotrone.

Last, but not least, the warmest thanks to my family, my Thorjørn, parents and my sister for their continuous support, encouragement and understanding. Thank you for always being there for me!

Contents

Abbreviations

aa	amino acids
Agrobacterium	*Agrobacterium tumefaciens*
AMV	alfalfa mosaic virus
an	*Angustifolia*
ap1	*apetala-1*
Arabidopsis	*Arabidopsis thaliana*
ATPEN6	*Arabidopsis* pentacyclic triterpene synthase gene
Att	lambda specific attachment site (sequence)
BAC	bacterial artificial chromosome
BGA	*BORGIA*
bp	base pair
BR	Brassinosteroid
BSA	bovine serum albumin
C24	Columbia 24
cap	*capulet*
cap1	*capulet 1*
cap2	*capulet 2*
CAPS	cleaved amplified polymorphic sequence
cDNA	complementary DNA
CDS	coding sequence
cer5	*eceriferum-5*
CLF	*CURLY LEAF*
ClH	chloral hydrate
cM	centi-Morgan
Col	Columbia
CZE	chalazal endosperm
dATP	2’-deoxyadenosine 5’-triphosphate
DCL1	*DICER-LIKE1*
dCTP	deoxycytidine 5’-triphosphate
ddm1	*decrease in DNA methylation1*
ddNTP	Dideoxyribonucleotide
Df	Degrees of freedom
dGTP	2’-deoxyguanosine 5’-triphosphate
dis1	*distorted-1*
DMAPP	dimethylallyl diphosphate
DME	*DEMETER*
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
Drosophila	*Drosophila melanogaster*
dsRNA	double stranded RNA
DTT	Dithiothreitol
dTTP	2'-deoxythymidine 5'-triphosphate
E(Z)	*ENHANCER OF ZESTE*
E. coli	*Escherichia coli*
EDD1	*EMBRYO-DEFECTIVE DEVELOPMENT*

EDTA	ethylene diamine tetraacetic acid
EMF2	*EMBRYONIC FLOWER2*
EMS	ethyl methane sulfonate
ESC	*EXTRA SEX COMBS*
EtBr	ethidium bromide
EtOH	Ethanol
FAA	Formaldehyde
FK	*FACKEL*
FBP	*Floral Binding Protein*
FBX	loading buffer
FIE	*FERTILIZATION INDEPENDENT ENDOSPERM*
FIS	*FERTILIZATION INDEPENDENT SEED*
gl2	*glabra-2*
GPP	geranyl diphosphate
GUS	*β-glucuronidase*
H3K9	the lysine residue (K) in position 9 at N-terminal tail of histone H3
HDAC	histone deacetylases
hemz	Hemizygous
homz	Homozygous
Hyg^R	hygromycin resistant
Hyg^S	hygromycin sensitive
IHF	Integration Host Factor
in trans	on two homologous chromosomes
InDel	insertion-deletion
IPP	isopentenyl diphosphate
IPTG	isopropyl-beta-D-thiogalactopyranoside
kb	kilo base
KCl	potassium chloride
LA-medium	Luria Broth medium with agar
lacZ	beta galactosidase gene in lac operon
LB	left border
LB-medium	Luria Broth medium
LDC	Lysine Decarboxylase
L*er*	Landsberg *erecta*
LiDS	lithium dodecyl sulfate
LP	left primer
M_2	EMS mutants
Mb	mega base
MCE	micropylar endosperm
MEA	*MEDEA*
MEP	methylerythritol phosphate pathway
MET1 a/s	*DNA methyltransferase1 antisense*
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulfate
mm1	multiple marker chromosome 1
mRNA	messenger ribonucleic acid
MS-2	Murashige Skoog plant growth medium 2 (Murashige and Skoog, 1962)

MSI1	*MULTICOPY SUPPRESSOR OF IRA 1*
MVA	mevalonate pathway
N_2	Nitrogen
NaAc	sodium acetate
NaCl	sodium chloride
NaH_2PO_4	sodium dihydrogen phosphate
nos	*nopaline synthase*
*npt*II	*neomycin phosphotransferase* gene encoding kanamycin resistance
OD_{600}	optical density at wavelenght 600 nm
ORF	open reading frame
PcG	Polycomb group
PCR	polymerase chain reaction
PEN	peripheral endosperm
PHE1	*PHERES1*
pLDC	Promoter of lysine decarboxylase gene
PRL	*PROLIFERA*
RB	right border
RNA	ribonucleic acid
RNAi	ribonucleic acid interference
RSY3	*RASPBERRY3*
RT	reverse transcriptase
S.O.C	S.O.C. cell growth medium
SDS	sodium dodecyl sulfate
SLP	*SCHLEPPERLESS*
SIN1	*SHORT INTEGUMENTS1*
SMT1	*STEROL METHYLTRANSFERASE1*
SNP	single nucleotide polymorphisms
SPEP	SDS, EDTA, phosphate buffer solution
SPS	solanesyl diphosphate synthase
SSC	NaCl and sodium citrate solution
SSLP	single sequence length polymorphism
T_1	first transformant generation
T_2	second transformant generation
TAE	tris-acetate
TAIR	The Arabidopsis Information Resource
Taq	*Thermus aquaticus*
T-DNA	transfer-DNA
Tm	melting temperature
UTR	untranslated region
UV	ultra violet
WD-40	Tryptophan–aspartate dipeptid repeat motif
wt	wild-type
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
YAC	yeast artificial chromosome
YEB-medium	yeast extract broth

Abstract

The female gametophyte of higher plants gives rise to the diploid embryo and the triploid endosperm which develop to produce the mature seed. Seed development is a concerted interplay of the embryo, endosperm and the surrounding diploid maternal tissue. In addition, it is highly dependent on the contribution from genetic programs executed in the gametophytic generations. What role the gametophytic maternal factors play in this process is still largely unknown.

This thesis describes two approaches to identify novel genes involved in seed development. A forward genetic approach addresses the molecular nature of the maternal effect mutant *capulet2* (*cap2*) by molecular mapping and a reverse genetics approach analyze the role in seed development of candidate genes from a promoter trap screen.

The *capulet2* gametophytic maternal-effect mutant was found in a linkage based screen preformed to identify gametophytic mutants in *Arabidopsis* (Grini et al., 1999). *cap2* embryo and endosperm development is blocked at a very early stages, and heterozygous plants display a 50% reduced seed set. To investigate the molecular nature of the *CAP2* gene, a map-based cloning approach was performed. Using PCR-based molecular markers the *cap2* mutation was mapped to a genetic interval of 4238 basepairs, on the tip of the right arm of chromosome 1. This interval spanned parts of two genes, one involved in monoterpenoid biosynthesis and the other putatively involved in triterpenoid biosynthesis. Neither of these two genes could be verified to be responsible for the *cap2* phenotype by complementation analysis. However the mapping interval of *cap2* was reduced from more than 1 Mb to less than 100 kb.

In a reverse genetic approach two candidate genes (*AtHD2C* and *AtLDC*) selected from a collection of promoter trap lines were analyzed to elucidate their role in seed development. Reporter gene expression studies, expression analysis, and the analysis of T-DNA insertion lines revealed that the candidate genes were expressed in the seed, but also in other organs. The promoter reporter line of *AtLDC* was found to have a similar but in some respects also different expression patterns in the seed than the original promoter trap line. The *AtHD2C* gene was found to be redundant as no phenotype could be observed in knock out alleles of the gene.

1. Introduction

Seed development in angiosperms is initiated by a double fertilization event in which two haploid sperm cells fuse with the haploid egg cell and the diploid central cell of the female gametophyte. During seed development the fertilized egg cell develops into the embryo and the fertilized central cell give rise to the endosperm. The sporophytic ovule integuments expand to form the seed coat and the sporophytic ovary develops into the fruit. On a genetic level, seed development depends upon contribution from genetic programs executed in both the sporophytic and the gametophytic generations, and requires a highly coordinated gene expression pattern in embryo, endosperm as well as maternal tissue. Moreover, genes expressed in the gametophytic maternal tissue (i.e. in the egg cell and the central cell) prior to fertilization, appears to play a crucial role in determining the developmental path of both embryo and endosperm. Whereas the role of sporophytic acting genes contributing to this process has been well studied over the past years, the role of gametophytic maternal factors is still largely unknown. The main focus of this thesis is the analysis and characterization of genes involved in seed development. A special emphasis is however put on haplo-phase specific genes required in the gametophytic phase for embryo and endosperm development, and a substantial part of this work is the molecular mapping of a gametophytic maternal effect mutant required for embryo and endosperm development.

1.1 How to study the gametophytic phase of the plant life cycle?

The gametophytes play a central role in the angiosperm reproductive process in which the female gametophyte is an essential structure. The female gametophyte carries the egg and the central cell that gives rise to the embryo and the endosperm, respectively, and participates in several reproductive processes like directing the pollen tube to the ovule and mediating fertilization of the egg cell and central cell. Upon fertilization, genes expressed in the female gametophyte participate to induce seed development, and possibly also play a role in controlling embryo and endosperm development (Drews et al., 1998; Drews and Yadegari, 2002). The gametophytic phase and the reproduction process have been studied over the past years and much has been learned about these processes by biochemical, histological, cytological and functional approaches (Redei, 1964; Tanksley et al., 1981;

Heslop-Harrison and Heslop-Harrison, 1986; Hill and Lord, 1987; Regan and Moffatt, 1990; Chibi et al., 1994; Schneitz et al., 1995; Wu et al., 1995). However, relatively little is known about molecular and genetic basis for processes and pathways involved in gametophytic development and function.

Several efforts have been made in the past few years to identify mutants that affect the gametophytic phase of the plant life cycle (Chen and McCormick, 1996; Chaudhury et al., 1997; Feldmann et al., 1997; Bonhomme et al., 1998; Christensen et al., 1998; Howden et al., 1998; Grini et al., 1999; Grini et al., 2002; Lalanne et al., 2004; Rotman et al., 2005). A genetic analysis of gametophytic mutants is challenging for several reasons; the gametophytes are restricted to a small number of cells and direct scoring for a phenotype can be a difficult task. The male and female gametophyte development occurs inside sporophytic tissue making it inaccessible for direct inspection (Page and Grossniklaus, 2002). In order to overcome these challenges, the combination of forward and reverse genetics approaches provides powerful tools. Both forward and reverse genetics approaches were used in this thesis to identify and characterize genes involved in seed development in the model plant *Arabidopsis thaliana*.

1.1.1 Forward genetics

The classical genetic approach to identify genes with a specific function is called “forward genetics”. Forward genetics primary relies on induction of mutations, identification of mutants with desired phenotypes and subsequent identification of the reflecting specific gene functions of the corresponding genes. Mutations can typically be introduced in *Arabidopsis* genome as single nucleotide changes using chemical mutagens, such as ethyl methane sulfonate (EMS) that generate predominantly G/C to A/T point mutations, or as small deletions using physical mutagens such as X-rays or UV irradiation. In addition, insertional mutagenesis using DNA elements that are able to insert at random within the genome, such as transposons or T-DNA is shown to be a successful approach to create loss-of-function mutations in plants (Bouchez and Hofte, 1998; Østergaard and Yanofsky, 2004). While chemical or radiation mutagenesis offers many advantages for isolating a collection of mutants that are defective for a particular process, the use of insertional mutagenesis in principle provides a more rapid way to clone a mutated gene. This is because the sequence of

the inserted element is known, and thus is used as a molecular tag for the gene in which it is inserted. The disrupted gene can easily be recovered using standard cloning and PCR-based strategies. However, a much larger population of plants must be screened to obtain the same number of alleles as for instance using chemical mutagenesis (Østergaard and Yanofsky, 2004). In addition, analysis of gametophyte T-DNA mutant collections suggest that T-DNA insertional mutagenesis can be problematic for identification and characterization of mutants that affect female gametophyte development (Bonhomme et al., 1998).

To isolate the gene that corresponds to a chemically or radiation-derived mutant, a positional (or map-based) cloning technique is the most effective strategy. This method was used in the first part of this thesis to map the EMS-induced gametophytic maternal-effect mutant *capulet2 (cap2)* (Grini et al., 2002) and will be thoroughly described in section 1.7.

Mutations that affect the female gametophyte can be identified as lethal in which female gametophytes carrying the mutation either abort embryo and endosperm development or are nonfunctional. Two basis types of screens, segregation distortion and seed set screens, have been used to identify female gametophyte mutants (Drews et al., 1998).

Mutants that disrupt female gametogenesis are expected to result in semisterility and thus can be identified by a phenotype showing a reduced seed set compared to wild-type plants. In plants heterozygous for a female gametophyte mutation, about half of the ovules harbor mutant and nonfunctional female gametophytes. These ovules fail to undergo seed development and eventually desiccate. Consequently, self-pollination of such heterozygous female gametophyte mutants would result in plants with siliques containing 50% normal seeds and 50% aborted seeds (reviewed in Drews et al., 1998; Chaudhury and Berger, 2001; Drews and Yadegari, 2002). Thus, one way to identify female gametophytic mutants is to screen for lines with siliques containing 50% normal seeds and 50% aborted ovules. However, chemical mutagenesis dramatically reduces fertility of mutants in general and hence, the semisterile phenotype is not suitable as a basis for a forward genetic screen. Other factors that can reduce a seed set are inappropriate environmental conditions, chromosomal rearrangements and sporophytic mutations. A second screening step, segregation distortion

screen, is therefore necessary for the final identification of a female gametophyte mutant (reviewed in Drews et al., 1998; Grini et al., 2002; Page and Grossniklaus, 2002).
The segregation distortion screen is based on the fact that a mutation affecting the development or the function of a gametophyte is transmitted to the progeny at a reduced frequency. The reduced transmission through one or both sexes results in a distorted, non-Mendelian segregation ratio of the mutation (and linked loci), therefore segregation ratio distortion can be used as a reliable indicator of a gametophytic defect. The segregation distortion screens are facilitated by following the segregation of either linked visible markers (e.g. recessive mutants) or linked T-DNA/transposons that carry a dominant antibiotic resistance gene. The progeny of a hemizygous T-DNA/transposon mutagenized line is expected to have a segregation ratio of 3:1 for antibiotic-resistant to antibiotic-sensitive seedlings. Any deviation of this Mendelian ratio would indicate that the gene providing antibiotic resistance is transmitted at a reduced frequency through one or both sexes, and that the insertion has disrupted a gene required either for gametophyte or embryo viability. The progeny of a hemizygous T-DNA/transposon line that segregate at ratio of 2:1 for resistant to sensitive seedlings indicate that the insertion has affected a gene required for embryo viability, whereas a ratio of 1:1 would result from lethality or failure to transmit through either female or male gametophyte (Figure 1). In order to further determine if the mutation is transmitted through the female or male gametophyte, reciprocal crosses and a seed-set screen have to be performed (reviewed in Drews et al., 1998; Page and Grossniklaus, 2002; Pagnussat et al., 2005).

A screening strategy using segregation distortion of a recessive visible marker closely linked to a newly induced gametophytic mutation is based on the fact that a recessive visible marker closely linked *in trans* to the locus of a newly induced gametophytic mutation will show aberrant segregation ratio of the marker. In other words, the wild-type marker allele that is closely linked to the gametophytic mutation would not be passed on to the next generation through the female gametophyte, subsequently increasing frequency of the recessive marker allele and the number of plants with the marker phenotype. This screening strategy was used by Grini et al. (1999) and Grini et al. (2002) to identify the *cap2* mutant.

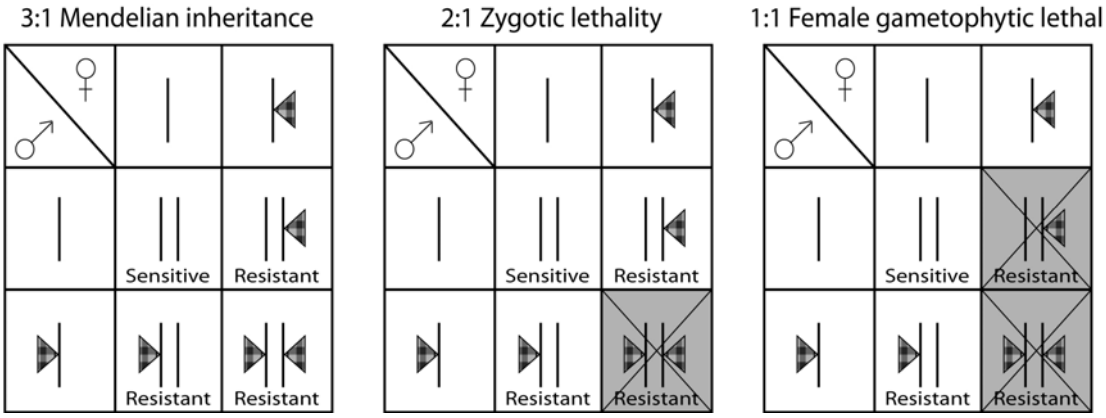

Figure 1. Screening for distorted antibiotic resistance gene segregation (redrawn after Page and Grossniklaus, 2002). If the T-DNA/transposon segregates as a dominant Mendelian locus, the expected ratio of antibiotic resistant to antibiotic sensitive seedlings would be 3:1. If the T-DNA/transposon insertion affect a gene required for embryo viability, the expected ratio for antibiotic resistant to antibiotic sensitive seedlings would be 2:1. For an insertional mutation, which affects exclusively one sex with full penetrance, a 1:1 segregation ratio of resistant to sensitive seedlings is expected.

Classical genetic approaches to gene identification, as described above, are based on generation of mutations that disrupt expression of genes involved in a specific process of interest. However, not all genes can be identified by this approach for two main reasons. First, mutations in many genes will not lead to an easy identifiable phenotype because many genes are functionally redundant, and the presence of one or several other members of a gene family can provide the same function. Second, many genes can function at different stages of development. Mutations in these genes can result in an early lethal phenotype and may mask the role of the gene in a later stage of development (Springer, 2000).

A second common method of gene identification does not rely on the generation of a mutant phenotype. This approach assumes that a promoterless reporter gene (e.g. *β-glucuronidase*, *GUS*), randomly inserted in the plant genome, will be transcribed if inserted in vicinity of regulatory elements controlling the expression of an endogenous gene. The expression pattern of the reporter gene would reflect the expression pattern of the tagged gene. There are two basic types of gene traps: enhancer trap and promoter trap. While in an enhancer trap the reporter gene is fused to a minimal promoter that is unable to drive its expression, but can be activated by neighboring enhancer elements, a promoter trap contain promoterless reporter gene so that expression can occur only when it is inserted within a transcriptional

unit and in the correct orientation (Springer, 2000). Gene trap provides a powerful tool for gene identification. Genes are identified by screening for tissue-specific reporter gene expression and does not require a mutant phenotype. The advantage of this method is that genes that are either expressed in many tissues during multiple developmental stages or are functionally redundant, can easily be identified. A second part of the work presented in this thesis is based on data from a promoter trap screen (Stangeland et al., 2003).

1.1.2 Reverse genetics

While forward genetics starts with the mutant and then leads to the gene, reverse genetics starts with a gene of interest and studies its expression and function in the cell or organism. Typically the temporal and spatial expression pattern is studied by PCR or hybridization techniques. The expression pattern of the gene of interest can also be studied by cloning its promoter region in front of a promoterless reporter gene. The reporter gene will be regulated by this promoter and the reporter protein expressed according to the expression pattern of the gene of interest. Alternatively, a direct way to obtain information on the function of a gene is to generate loss-of-function mutations and study the phenotype of the mutant. The availability of large collections of plants mutagenized by an insertion element, such as T-DNA or transposon, has made the process of finding a mutant line with a knockout in the gene of interest more straightforward (see section 1.2). Another approach that has successfully been used in *Arabidopsis* to silence the function of a target gene is RNA interference (RNAi), a knockdown approach. The method is based on a process by which a double-stranded RNA (dsRNA) silences specifically the expression of the homologous gene, through degradation of their related mRNA. The primary advantages of RNAi are the ease of generating short dsRNAs that mediate RNAi, and the flexibility of the inhibition: one can spatially and temporally control the interference reaction (reviewed in Steinmetz and Davis, 2004). Several stock centers with the available T-DNA, transposon insertion and RNAi collections have been established and can be searched for lines carrying the inserted element in a particular location of the genome. At the Salk Institute (http://signal.salk.edu) more than 225 000 independent insertion events have been created, and the precise location has been determined for T-DNA insertions in approximately 90 000 lines (Alonso, 2003). Information from this database has been used in this work. Several lines were ordered to study the T-

DNA knockout mutants of the putative *cap2* alleles and a histone deacetylase gene, which study was based on the promoter trap screen mentioned above.

In this work a reverse genetics approach was used to study the involvement of two genes in seed development. The genes were selected from a collection of sequenced promoter trap lines (Stangeland et al., 2003), based on the expression pattern of a promoter trap line and postulated function of the gene (see Results), in addition to current knowledge regarding the genetic basis of the seed development. An expression profile was made, and promoter::*GUS* and mutant analysis was performed for these two genes.

1.2 *Arabidopsis thaliana* as a model organism

Arabidopsis thaliana is a small dicotyledonous angiosperm (flowering plant) belonging to the *Brassica* family (Arabidopsis Genome Initiative, 2000). *Arabidopsis* is considered to be an excellent biological model for many aspects of plant biology, as it is related to several hundred thousand plant species due to a relatively recent evolutionary radiation of flowering plants from a common ancestor (Somerville and Koornneef, 2002). *Arabidopsis* holds many advantages as a model organism compared to other plant species. The small size and simple growth requirements make it easy to grow under laboratory conditions. For a flowering plant, it has a relatively short life cycle and a new generation is obtained in only six weeks. *Arabidopsis* is self-pollinating and produces thousands of seeds from a single individual, thus rapidly giving many progenies from a single mutant or transgenic plant. In addition, controlled crossing of plants can be performed (Meinke et al., 1998; Somerville and Koornneef, 2002). *Arabidopsis* has the smallest known plant genome (approximately 125 Mb) and contains fewer repetitive sequences than any known higher plant, which makes *Arabidopsis* an excellent model system for classical genetics. The complete sequencing of the Columbia *Arabidopsis* genome in year 2000 (Arabidopsis Genome Initiative, 2000) and low coverage shotgun sequencing of the Landsberg ecotype (Cereon Genomics (Monsanto)) (Jander et al., 2002) have resulted in more than 50 000 polymorphisms between these two most commonly used ecotypes. This greatly facilitates molecular studies and isolation of mutant genes by map-based cloning (Page and Grossniklaus, 2002; Somerville and Koornneef, 2002).

Transgenes can be easily introduced into *Arabidopsis* with high efficiency using the techniques based on the natural transformation mechanism of the soil bacteria *Agrobacterium tumefaciens* (Barghchi, 1995). This method is of great importance for molecular and expressional analysis of a given gene by reverse genetic approach. These methodical advantages, together with the complete sequence of *Arabidopsis* genome, have resulted in thousands of transgenic lines carrying random T-DNA insertions throughout the genome and DNA microarrays containing probes for all of the genes known to be expressed in the plant (Meinke et al., 1998; Somerville and Koornneef, 2002). A large collection of characterized mutations and transgenic plants is also available, in which genes involved in near to every major biochemical pathway have been knocked-out (Somerville and Koornneef, 2002). This provides a powerful foundation for functional studies of *Arabidopsis* genes and facilitates the identification and analysis of new genes and gene families.

1.3 The life cycle of Arabidopsis thaliana

The life cycle of plants alternates between a multicellular haploid generation, called the gametophyte, and a multicellular diploid generation, called the sporophyte. The first step in the life cycle is fertilization, followed by embryonic development, seed germination and vegetative growth, reproductive development and finally senescence (Hartwell et al., 2000). In flowering plants, the gametophyte generation has been reduced to consist only of two small structures inside the sexual organs of the sporophyte. After meiosis the haploid spores undergo mitosis and differentiate into either a pollen grain (male gametophyte) or an embryo sac (female gametophyte) (Goldberg et al., 1994). The major function of the gametophyte generation is to produce haploid male or female gametes, sperm and egg cells, respectively. Fusion of the egg cell with the sperm cell gives rise to the sporophyte, thus completing the life cycle (reviewed in Drews et al., 1998).

1.3.1 Female gametophyte development

The female gametophyte, also referred to as the embryo sac or megagametophyte, develops in the ovule, found within the carpel's ovary. The ovule is the site of female meiosis (production of functional megaspore-megasporogenesis), formation of the embryo sac (megagametogenesis), double fertilization and embryogenesis. It consists of tissues derived from both generations of the plant life cycle, the diploid sporophyte and the haploid

gametophyte. Surrounded by several cell layers of sporophytic origin (inner and outer integuments), the multicellular female gametophyte develops in the center of the ovule to produce the gamete classes. After double fertilization, the ovule develops into a seed containing the fertilization products, the embryo and the endosperm (reviewed in Reiser and Fischer, 1993; Grossniklaus and Schneitz, 1998).

The female gametophyte is generated from the haploid functional megaspore via process called megagametogenesis. In this process the functional megaspore enlarges and undergoes three successive nuclear divisions without cytokinesis, producing an eight-nucleated coenocytic megagametophyte that contains four nuclei at both poles. Subsequently, cell walls form around these nuclei, resulting in a cellularized female gametophyte. During cellularization, two nuclei, one from each pole, migrate toward the center of the developing female gametophyte, and fuse to form the diploid endosperm nucleus. These events results in a seven-celled structure that represents a mature female gametophyte, consisting of three antipodal cells at the chalazal (proximal) end, one egg cell and two synergid cells at the micropylar (distal) end, and one central cell. These cells comprise four groups that functions in nutrition of the female gametophyte, fertilization, embryogenesis and nutrition of the embryo. In Arabidopsis, the three antipodal cells undergoes cell death immediately before fertilization (reviewed in Reiser and Fischer, 1993; Drews et al., 1998; Grossniklaus and Schneitz, 1998; Yadegari and Drews, 2004).

1.3.2 Double fertilization and early seed development

The male gametophyte, or pollen, develops in the anther from microspores. A mature male gametophyte consists of two haploid sperm cells enclosed by a haploid vegetative cell (reviewed in McCormick, 1993). Sexual reproduction is initiated when the male gametophyte is transferred from the anther to the carpel's stigma. Soon after pollination, the male gametophyte forms a pollen tube that grows through the carpel's sporophytic tissue to reach the female gametophyte. The pollen tube enters the female gametophyte through the micropylar end by growing into one of the synergid cells. The penetrated synergid cell undergoes cell death either before or upon pollen tube arrival. Immediately after the pollen tube arrives at the synergid, its contents, including the two sperm cells, are released into the degenerating synergid cytoplasm. One of the sperm cells migrates to the egg cell and the other to the central cell. Double fertilization occurs when the plasma membranes of the male

gametes fuse with those of the haploid egg cell and diploid central cell, and the sperm nuclei are transmitted into these cells for karyogamy. Fertilization of the egg cell and the central cell gives rise to the diploid embryo and the triploid endosperm, respectively, which develop to produce the mature seed (reviewed in Drews et al., 1998; Drews and Yadegari, 2002; Yadegari and Drews, 2004). Fertilization also initiates changes in maternal tissue. The sporophytic ovary develops into a fruit and the ovule integuments differentiate to form the protective seed coat (Drews and Yadegari, 2002; Gehring et al., 2004).

1.4 Seed development

1.4.1 Embryo development

Once the egg cell is fertilized, the unicellular zygote undergoes a progressive transition to become the embryo and produce a new multicellular generation, the seedling, which displays the basic body plan and organization of the plant (Jürgens et al., 1995; Laux and Jürgens, 1997). An apical-basal pattern is established along the main body axis of the embryo. It is a top-to-bottom array of distinct elements, including the shoot meristem, cotyledons, hypocotyl, root and root meristem (Jürgens et al., 1995; Laux and Jürgens, 1997) (Figure 2). The shoot and the root meristem are two populations of stem cells present in the embryo that will form the mature plant body after germination. The cotyledon, on the other hand, is a terminal differentiated organ that functions mainly during seed germination by accumulating food reserves, utilized by the seedling for growth and development (Goldberg et al., 1994)**.**.

Plant embryogenesis can be divided into three general phases where different developmental and physiological events occur: 1) postfertilization-proembryo, 2) globular-heart transition and 3) organ expansion and maturation (Goldberg et al., 1994) (Figure 2). The basic body plan is established during the first phase of embryogenesis and becomes completely evident by the time the dicot embryos reach the heart stage. Subsequent events include further growth of the embryo, morphogenesis, cell differentiation and preparation for dormancy (Laux and Jürgens, 1997).

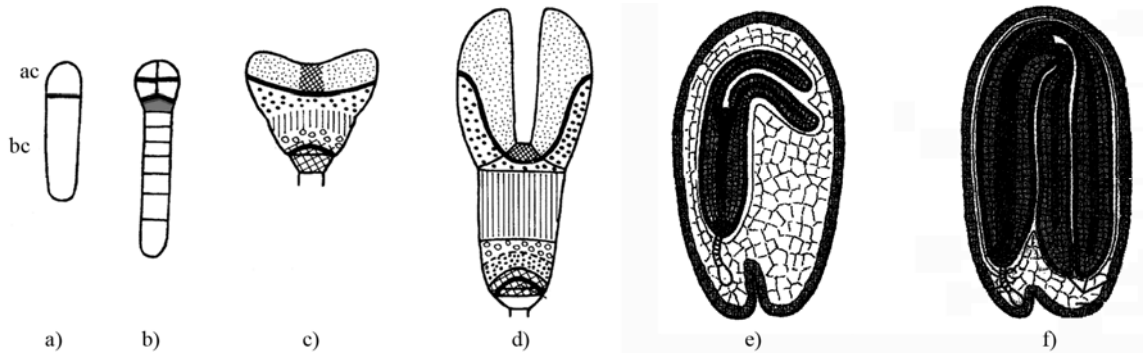

Figure 2. *Arabidopsis* embryo development (redrawn after Goldberg *et al*., 1994 and Jürgens *et al*., 1995). a) Two-cell proembryo stage. The first division of the zygote following fertilization is asymmetric and gives rise to a small apical (ac) and a large basal (bc) cell. b) Eight-cell proembryo stage. (c) Heart embryo stage. d) Torpedo embryo stage. e) Walking-stick embryo stage. f) Mature embryo. a) and b) show postfertilization-proembryo phase, c) shows a globular-heart phase and d)-f) show organ expansion and maturation phase in embryogenesis.

1.4.2 Endosperm development

After the diploid central cell has been fertilized, the endosperm development is initiated by successive divisions of the triploid nucleus without cytokinesis, forming a large multinucleate structure, a so called syncytium, lining the central cell (Berger, 1999; Brown et al., 1999; Olsen, 2004). Endosperm development can be divided into nine distinct developmental substages based on the total number of nuclei. During these stages eight successive rounds of syncytial mitoses occur. At the final stage, the syncytial endosperm contains about 200 nuclei (Boisnard-Lorig et al., 2001). The syncytial division occurs as rapid coordinated waves, in domains that are patterned along the anterior-posterior (AP) axis of the syncytium (Brown et al., 1999; Boisnard-Lorig et al., 2001; Berger, 2003). This polarity axis is divided into three regions that become distinct as the seed grows: 1) the micropylar endosperm (MCE), the region surrounding the developing embryo at the anterior pole, 2) the peripheral endosperm (PEN) in the central chamber, which is the largest domain, containing the largest number of nuclei and 3) the chalazal endosperm (CZE) at the posterior pole, containing few nuclei. The mitotic activity of MCE, PEN and CZE occur independently of each other, with nuclei dividing synchronously within each domain. However, no mitotic divisions seems to occur in the CZE after the third round of division, suggesting that these nuclei undergo endoreduplication resulting in greater ploidy in the CZE than in other mitotic domains (Boisnard-Lorig et al., 2001). Cellularization begins in the micropylar region and occurs simultaneously with the initiation of cotyledon development

within heart embryo stage. The cellularization spreads then as a wave across the peripheral region toward the chalazal region. At the heart embryo stage most of the endosperm is cellularised with the exception of the nuclei at the chalazal pole that remains syncytial until late stages of seed maturation. After cellularization cell divisions are rare in the endosperm, while the embryonic cells divide rapidly and the embryo eventually occupies most of the seed. Finally, most endosperm cells die during seed maturation with the exception of the outer aleurone cell layer (Berger, 1999; Brown et al., 1999; Sørensen et al., 2002; reviewed in Olsen, 2004).

1.5 Maternal control of seed development

1.5.1 Sporophytic maternal effects

Seed development is controlled in part by maternal-effect genes, which are expressed in the female gametophyte prior to fertilization and/or in the diploid maternal tissue surrounding the developing embryo and endosperm. The basic body plan of the plant is established early during embryogenesis and the apical-basal axis of the embryo is always aligned parallel to the chalazal-micropylar axis of the ovule, suggesting that the polarity of the embryo is controlled by genes expressed in the surrounding maternal tissue (the diploid sporophytic tissue of the ovule) (Chaudhury et al., 1998; Chaudhury and Berger, 2001; Drews and Yadegari, 2002; Laux et al., 2004). These maternal-effect genes expressed in the diploid sporophytic maternal tissue are referred to as sporophytic maternal-effect genes. Several sporophytic maternal mutations have been reported, in which the defects in embryo and/or endosperm development are due to the genotype of the sporophytic maternal tissue, and not dependent on the genotypes of these two structures. In such sporophytic maternal mutants, a recessive homozygous mutant embryo would be able to grow as long as the sporophyte is heterozygous. In contrast, a heterozygous embryo would fail to develop normally if the sporophyte mother is homozygous for the mutation (Drews et al., 1998; Chaudhury and Berger, 2001; Drews and Yadegari, 2002).

Two essential ovule specific MADS box genes *Floral Binding Protein 7* (*FBP7*) and *FBP11*, isolated from Petunia, have been shown to play a role in the development of the seed coat and are required for normal endosperm development (Colombo et al., 1997). Another group

of sporophytic maternal-effect genes has been described in barley. Mutations in these genes result in endosperm mutants that produce shrunken seeds regardless of the genotype of the pollen source (Felker et al., 1985). In Arabidopsis, the *SHORT INTEGUMENTS1* (*SIN1*) gene is required in the maternal sporophyte for normal patterning in the embryo (Ray et al., 1996). The *SIN1* has recently been renamed *DCL1* (*DICER-LIKE1*) because of its sequence similarity to the *Dicer* group of genes, required for RNA silencing in *Drosophila* and *Caenorhabditis* (Schauer et al., 2002). It has been suggested that maternal *DCL1* functions early in *Arabidopsis* development, presumably through posttranscriptional regulation of specific mRNA molecules (Golden et al., 2002).

By contrast, embryo and endosperm development can also be influenced by maternal-effect genes expressed within the haploid female gametophyte. These genes are referred as gametophytic maternal-effect genes. The gametophytic maternal-effect genes and mutations are the main focus of this thesis and will hence be thoroughly presented in the next two sections (1.5.2 and 1.6).

1.5.2 Female-gametophytic maternal effects

A "maternal effect" refers to a reciprocal cross that results in different phenotypes, which are exclusively determined by the genotype of the female parent (Autran et al., 2005). A gametophytic maternal effect can be caused by various mechanisms. First, it can be caused by mutations in genes that are expressed during the gametophyte development in the egg and/or the central cell, and whose gene products are required for embryo and endosperm development after fertilization. Mutations in these kinds of genes represent the true gametophytic maternal effects (see section 1.6). Second, it can be caused by haplo-insufficiency in the endosperm, i.e. mutations in genes which require at least two wild-type copies for endosperm development. Third, mutations in genes whose paternally contributed alleles are imprinted will also have a gametophytic maternal effect (Grossniklaus et al., 2001; Autran et al., 2005).

Female gametophytic mutants typically are identified by phenotypes that show a reduced number of normal seeds in the silique compared to that of wild-type plants (see section 1.1.1). Both haplo-insufficient mutants and mutants affecting paternally imprinted genes

show the same ratio of normal and aborted seeds as the female gametophytic mutants. However, mutations that cause the gametophytic maternal effects can be distinguished from mutations in paternally imprinted genes by molecular analysis. Transcripts that are produced only before fertilization and not after, would suggest strongly that the gene falls into the true gametophytic maternal effect class. By contrast, if transcripts are produced only after fertilization and not before, it would imply that the gene is paternally imprinted. To determine whether a mutant falls into the haplo-insufficient endosperm class, crosses with diploid pollen donors can be used (reviewed in Drews and Yadegari, 2002; Autran et al., 2005).

Recent studies have led to identification of mutations in several important *Arabidopsis* genes that cause distinctive female-gametophytic maternal effects on embryo and endosperm development. These include *PROLIFERA* (*PRL*) (Springer et al., 1995) and mutants of the *FERTILIZATION INDEPENDENT SEED* (*FIS*) class of genes that has attracted wide spread attention in recent years. The *FIS* class consists of three previously identified genes, *FIS1/MEDEA* (*MEA*), *FIS2* and *FIS3/FERTILIZATION INDEPENDENT ENDOSPERM* (*FIE*) (Ohad et al., 1996; Chaudhury et al., 1997; Grossniklaus et al., 1998) and two newly isolated members, *MULTICOPY SUPPRESSOR OF IRA 1* (*MSI1*) and *BORGIA* (*BGA*), which is still an unidentified gene (Köhler et al., 2003; Guitton et al., 2004a). Mutation in these genes results in abnormal embryo and endosperm development that cannot be rescued by pollination with wild-type pollen. Mutations in *FIS* class genes are female gametophyte mutations that result in the spontaneous initiation of endosperm development in the absence of fertilization. This suggests that the *FIS* genes normally control repression of genes required for initiation of endosperm development in unfertilized ovule (Grossniklaus et al., 2001; Guitton et al., 2004). In addition, *MEA*, *FIS2* and *FIE* have been proposed to control organization of the endosperm anterior-posterior axis and the endosperm nuclei proliferation (Sørensen et al., 2001). *MEA*, *FIS2*, *FIE* and *MSI1* encode proteins that have homologies with Polycomb group (PcG) proteins, which in *Drosophila* form multiprotein complexes. The PcG complexes have a general repressive effect on gene expression, most likely through the effect of chromatin remodeling (Pirrotta, 1998). In animals, PcG complexes regulate homeotic genes and genes responsible for cell proliferation (Chaudhury and Berger, 2001). *MEA* encodes a SET-domain polycomb protein that shares homology with the *Drosophila*

protein ENHANCER OF ZEST (E[Z]) (Grossniklaus et al., 1998). *FIS2* encodes a Zn-finger protein related to *Drosophila* PcG protein SUPRESSOR OF ZESTE 12 (SU[Z]12) (Luo et al., 1999), while *FIE* encodes a WD-repeat polycomb protein related to *Drosophila* EXTRA SEX COMBS (ESC) (Ohad et al., 1999). *MSI1* encodes a WD-40 domain protein that is homologous to the *Drosophila* histone binding protein p55 (Hennig et al., 2003). As their homologues in *Drosophila*, MEA and FIS have been shown to interact physically (Luo et al., 2000; Spillane et al., 2000; Yadegari et al., 2000). However, no interaction has been shown between FIS2 and either FIE or MEA. The lack of the detectable interaction between FIS2-MEA and FIS2-FIE suggests that FIS2 may act in the early formation of the complex and not directly with the complex, or the interaction could be mediated by additional proteins (Chaudhury and Berger, 2001; Drews and Yadegari, 2002). Recently, CURLY LEAF (CLF) protein was observed to interact with EMBRYONIC FLOWER2 (EMF2) protein, which share sequence similarity to MEA and FIS2, respectively. The interaction was mediated by the VEFS (VRN2-EMF2-FIS2-SU(Z)12) box domain, suggesting that FIS2 and MEA also might interact via the VEFS box. The VEFS box of FIS2 was specifically tested against MEA in a two-hybrid assay, and the result demonstrated a possible interaction between the two proteins (Chanvivattana et al., 2004). Moreover, MSI1 was shown to interact physically with FIE (Köhler et al., 2003a). MSI1, FIE and MEA have also been shown to be parts of a protein complex *in vivo*, that based on molecular weights, probably containing several other components (Köhler et al., 2003a). By analogy to the *Drosophila* PcG complex that interacts with histone deacetylases (HDAC) (Tie et al., 2001), it has been proposed that the MEA, FIS2, FIE and MSI1 proteins (FIE-MEA complex) also in *Arabidopsis* form a large histone deacetylase complex. This complex may repress transcription of a number of target genes, by establishing a repressive chromatin state at these loci. A large family of HDAC is present in *Arabidopsis*, yet no members has been identified in the FIE-MEA complex (Chaudhury and Berger, 2001; Berger, 2003; Guitton and Berger, 2005). One of the target genes, *PHERES1* (*PHE1*), encoding a MADS-box class transcription factor, was identified recently (Köhler et al., 2003b). The MEA-FIE complex is shown to be associated with the *PHE1* promoter, directly controlling its expression (Köhler et al., 2003b). Furthermore, it was shown that *PHE1* is an imprinted gene, with a paternally expressed and a maternally repressed allele, and that MEA is required for its maternal repression (Köhler et al., 2005).

Genes belonging to the *FIS*-group are expressed in the female gametophyte before fertilization. After fertilization, expression is entirely from the maternal allele during early seed development, but later it becomes bi-allelic. The gametophytic maternal effect of these genes on seed development could therefore be due paternal imprinting or haplo-insufficiency (reviewed in Drews and Yadegari, 2002; Lohe and Chaudhury, 2002; Schauer et al., 2002; Schubert and Goodrich, 2003).

Haplo-insufficiency has been ruled out for the *MEA* gene through the use of crosses with diploid pollen donors (Grossniklaus et al., 1998). It was shown that the paternally inherited *MEA* allele is transcriptionally silenced, specifically within the endosperm, and hence is an imprinted gene. However, in the embryo (at the torpedo stage and later) and in the postembryonic vegetative tissues, the paternally inherited MEA allele was shown to be transctiptionally active (Kinoshita et al., 1999). Moreover, pollen with the mutations in *decrease in DNA methylation1* (*ddm1*) gene, which reduce genomic DNA methylation to 30%, rescues the maternal effect of *mea* by zygotically reactivating the *MEA* function from the paternally inherited wild-type allele (Vielle-Calzada et al., 1999). For the *FIS2* gene, no diploid pollen crosses have been carried out (Drews and Yadegari, 2002). The *fis2* female gametophyte can generate viable seeds when pollinated with hypomethylated pollen, using a transgenic DNA methyltransferase1 antisense (*MET1 a/s*), in which DNA methylation levels are approximately 15% of the wild-type level (Luo et al., 2000). The *mea* mutants pollinated with the hypomethylated *MET1 a/s* pollen produced also 98.8% viable seeds with normal endosperm end embryo development. However, the rescue of female *mea* and *fis2* by hypomethylated *MET1 a/s* pollen is not due to a functional paternally derived *MEA* and *FIS2* allele. This is because pollen with demethylated DNA that carries a mutation in either *mea* or *fis2* also restores seed viability (Luo et al., 2000). This suggests that the rescue of *mea* and *fis2* by hypomethylation of the pollen genome is the result of activation of silenced genes other than *FIS*. These genes may have a role in the early developing seed and acts as modifiers of the maternal effect of *fis* mutations (Luo et al., 2000; Berger, 2003).

In contrast to *mea*, maternal *fie* cannot be rescued by *ddm1 FIE* paternal allele (Yadegari et al., 2000). However, maternal *fie1* seeds are rescued if the pollen is hypomethylated (*MET1 a/s*) and has a wild-type *FIE* allele (Vinkenoog et al., 2000). The *bga* and *msi1* mutants can

also be rescued by pollination with hypomethylated *MET1 a/s* pollen (Guitton et al., 2004). Hence, by using *met1* mutants it has been shown that DNA methylation may regulate imprinting of *FIS*-class of genes.

Two recent studies have shown that, in *Arabidopsis*, hypermethylation of imprinted loci represent the default state, whereas transcriptional initiation of maternally inherited alleles requires an active mechanism that removes the inhibitory modifications (reviewed in Autran et al., 2005). Recent evidence suggests that inhibitory modifications of the maternal allele of *MEA* are removed in the female gametophyte's central cell by a protein called DEMETER (DME) (Choi et al., 2002). *DME* encodes a DNA glycosylase that is required to form nicks, and possibly remove 5'-methylcytosine residues from the promoter of *MEA* (Xiao et al., 2003). *DME* is expressed in the central cell before fertilization and its activity is necessary for *MEA* expression in the developing endosperm. *DME* activate the maternal copy of *MEA*, while the paternal copy remains inactive, most likely because *DME* is not expressed during male gametogenesis (Choi et al., 2002). Recent studies have shown that the DME and the MET1 methyltransferase act antagonistic on the *MEA* promoter in the central cell to control endosperm imprinting and seed viability. While MET1 methyltransferase suppresses the maternal *MEA* allele expression by directly methylating the *MEA* promoter, DME excises 5'-methylcytosine residues from the promoter, leading to activation of the maternal *MEA* transcription in the central cell (Xiao et al., 2003). Finally, mutations in *dme* are not able to rescue *fie* and *fis2* mutant phenotypes, supporting the idea that all *FIS* genes might not be regulated by the same imprinting mechanism (Choi et al., 2002).

Not all gametophytic maternal effect genes are paternally inactive in early embryo and endosperm development. The *PROLIFERA* gene is one example. The *PRL* gene encodes a homologue of the DNA replication licensing factor Mcm7. The Mcm7 is required for the initiation of DNA replication along with other factors (Springer et al., 1995). The PRL protein has been shown to accumulate during embryo sac development, and the accumulation of maternal protein is required for successful seed development following fertilization. The *PRL* gene is expressed from both paternal and maternal alleles, in both embryo and endosperm. This rule out imprinting as the explanation for the maternal effect

(Springer et al., 2000). Hence, *PRL* is a clear example of a true gametophytic maternal effect gene.

1.6 Novel gametophytic maternal-effect mutants in *Arabidopsis*: *capulet* (*cap*) mutants

Two new gametophytic maternal-effect mutants, *capulet1* (*cap1*) and *cap2* have recently been isolated, presenting a novel type of gametophytic maternal effect phenotype (Grini et al., 2002). The *cap* mutants, termed after Shakespeare's *Romeo and Juliet*, have a gametophytic maternal effect on embryo and endosperm development. In the *cap1* mutant, both embryo and endosperm development are arrested at early stages. In the *cap2* mutant, which is the main focus in this thesis, endosperm development is blocked at a very early stage, while the embryos can develop to the early heart stage (Figure 3). Like other female gametophytic mutants, heterozygous plants display a reduced number (ideally ~50%) of normal seeds in a silique compared to wild-type.

The *cap* mutant phenotype is obtained only when the mutant allele is transmitted through the female gametophyte. The mutations cannot be rescued by pollination with wild-type pollen. Moreover, the *cap* mutant phenotypes do not depend on *CAPULET* (*CAP*) gene dosage in the endosperm and embryo, since the mutant phenotypes are not influenced by crosses with diploid pollen. In contrast to the *fis* mutants, *cap1* and *cap2* mutations do not result in fertilization-independent seed development, suggesting that the *CAP* genes are not required for prefertilization repression of embryo and endosperm development (Grini et al., 2002). In addition, unlike the *FIS* genes, pollination with the *MET1 a/s* or *ddm1* pollen fails to restore *cap* seed development. This suggests that the maternal effect of the *cap* mutants is not due to imprinting of the paternal *CAP* alleles.

To further investigate the role of the *CAP* genes in the endosperm development, Grini *et al.* (2002) analyzed the expression pattern of a *FIS2* promoter *GUS* fusion transgene (*FIS2::GUS*) in the *cap* mutant background. In wild-type plants, no *GUS* expression is observed in the endosperm when the *FIS2::GUS* is introduced via pollen. This is due to imprinting of the *FIS2* allele in pollen (Luo et al., 2000). When wild-type, *cap1/CAP1* and *cap2/CAP2* plants were pollinated with the *FIS2::GUS* transgene pollen, no *GUS* expression

was observed during endosperm development, indicating that *cap1* and *cap2* mutations are not required to remove the imprinting-mediated expression barriers from the *FIS2* promoter. Furthermore, when *cap1/CAP1;FIS2::GUS* and *cap2/CAP2;FIS2::GUS* plants were pollinated with wild-type pollen, the *FIS2::GUS* transgene was expressed in both *cap1* and *cap2* endosperm, showing that *cap1* and *cap2* embryo sacs are able to activate the *FIS2::GUS* transgene (Grini et al., 2002). These results indicate that *CAP1* and *CAP2* are not required for *FIS2::GUS* expression in the endosperm.

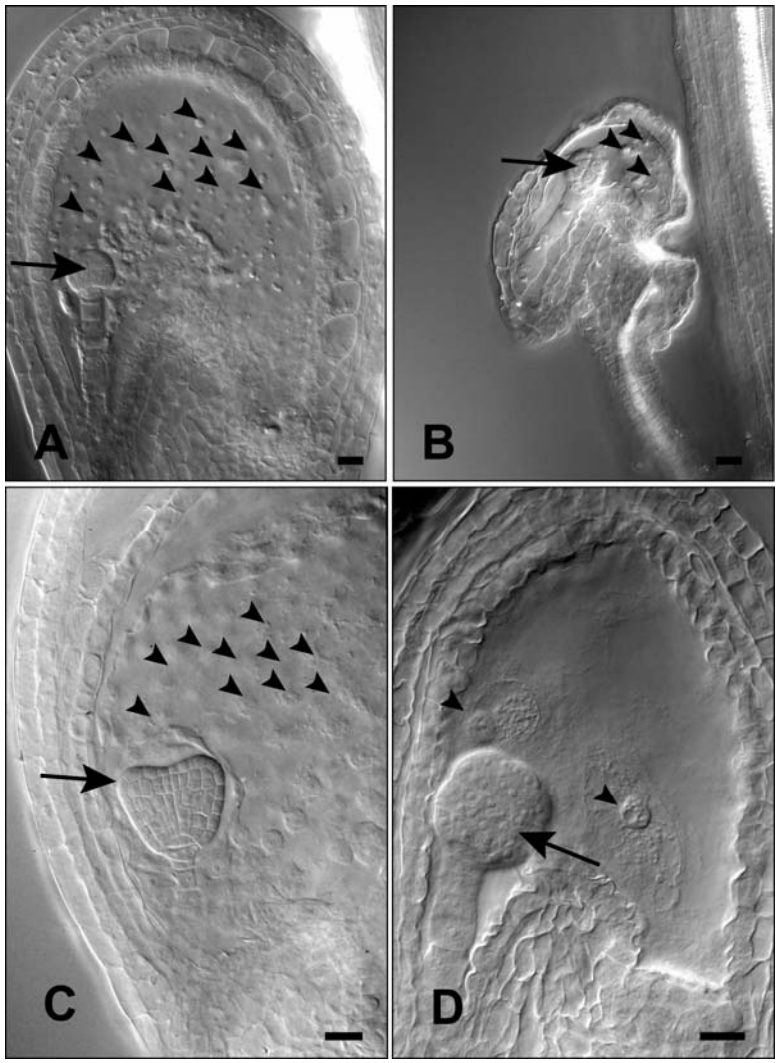

Figure 3. Embryo and endosperm development in *cap2* and wild type. A, B are taken from the hybrid F_2 Landsberg Columbia mapping population. C,D are taken from heterozygous *cap2* in Landsberg. A, C wild type. B, D *cap2*. A)WT in early globular stage. Many endosperm nuclei are visible (arrowheads). B) *cap2* from same silique as A. Seed size is drastically reduced and only a few endosperm nuclei are visible. C) WT in heart stage. D) *cap2* in comparable stage as C. Only two enlarged endosperm nuclei are visible. Both embryo and endosperm is arrested at this stage. Arrowheads, endosperm nuclei; Arrow, embryo. Bars = 20µm.

Based on the present data, the *CAP* genes could represent novel true gametophytic maternal effect genes that are expressed in the female gametophyte itself and whose products are required for embryo and endosperm development. The molecular basis for the gametophytic maternal effects of these genes remains to be resolved. This will be determined by molecular cloning and characterization of the *CAP* genes. Thus, one of the main goals of this thesis was to map the *cap2* mutation in order to further determine the role of the *CAP2* gene in seed development.

1.7 Molecular mapping; positional cloning

Positional cloning, also called map-based cloning was used in this thesis to map the *cap2* mutation. This is a widely used approach to identify the gene defined by a mutation. The method is based on identification of the mutated gene by looking for linkage to molecular or genetic markers, whose physical location in the genome is already known (Lukowitz et al., 2000; Jander et al., 2002). In addition to markers, a mapping population that segregates for the mutation of interest is also required. The mapping population is mostly generated by crossing the mutant plant in a specific ecotype (Landsberg in this work) to wild-type plants of a different ecotype (Columbia in this work). The positional cloning of *Arabidopsis* genes has become more easy to perform, and has been greatly improved by completion of the Columbia (Col) *Arabidopsis* genome sequencing project (The Arabidopsis Genome Initiative, 2000) and low coverage shotgun sequencing of the Landsberg *erecta* (L*er*) ecotype carried out at Cereon Genomics (Monsanto) (Jander et al., 2002). Data from Col and L*er* genomes has made it possible to develop a database of DNA polymorphisms that can be used as genetic markers. Stretches of L*er* shotgun sequence were compared with Col genomic sequence from BACs in the Cereon Genomics project. The sequence variations between these two ecotypes were classified into two types: single nucleotide polymorphisms (SNPs), differences in one single nucleotide at specific location in genome, and insertion-deletion (InDel) differences, where one ecotype has an insertion of a number of nucleotides relative to the other (Jander et al., 2002). This work resulted in 56,670 randomly distributed polymorphisms of which 37,344 SNPs, 18,579 InDels, and 747 Large InDels, all available now in Monsanto Arabidopsis Polymorphism and L*er* Sequence Collection (http://www.arabidopsis.org/cereon/). The information found in this database was used in this work to create new molecular markers for mapping of the *cap2* mutation.

Mapping of the *cap2* mutation was performed with PCR-based molecular markers that detected sequence variations (polymorphisms) between Col and L*er* ecotypes at a particular locus. PCR-based cleaved amplified polymorphic sequences (CAPS) and simple sequence length polymorphisms (SSLP) were two types of molecular marker used in mapping. The CAPS method is based on detection of SNPs that affect restriction sites in one of the ecotypes. The sequence containing the SNP is amplified by marker-specific primers and the polymorphism is detected by loss or gain of restriction enzyme recognition site in one of the ecotypes, but not in the other (Konieczny and Ausubel, 1993). The SSLP method is based on amplification of regions of microsatellite DNA that differ in length between two ecotypes (Bell and Ecker, 1994) (Figure 4).

In general, the molecular mapping of the *cap2* mutation was based on screening of mapping population for plants that were recombinants within a previously defined mapping interval, flanked by two molecular markers. By "recombinant" is meant a plant that has different ecotype specific genotype in two markers found in a mapping interval. Since recombination events between *cap2* and the molecular markers can be recorded from both sides of the mutation, the mapping interval can be made. Knowing that the *cap2* mutation is present in L*er* ecotype we can determine from which side of the mutation the recombination has occurred. This can be done by determining the phenotype of the recombinant plant and comparing its phenotype with the genotypes found for the two markers. If a plant, for example, shows *cap2* phenotype and has Col/Col and L*er*/Col genotypes in markers flanking the mutation from the left and the right side, respectively, we can conclude that the mutation is located to the right of the Col/Col marker (since the mutation is present in L*er* ecotype). A plant with the same markers genotype but a wild-type phenotype would lead to a conclusion that the mutation is to the left of the L*er*/Col marker (Figure 5). Double recombinants are not taken into account, since they have the same genotype in both flanking markers. As long as there are polymorphisms that can be used to create new molecular markers in the mapping interval, and as long as there is a recombination event left, the interval can be narrowed down by "walking" towards the mutation site from both sides.

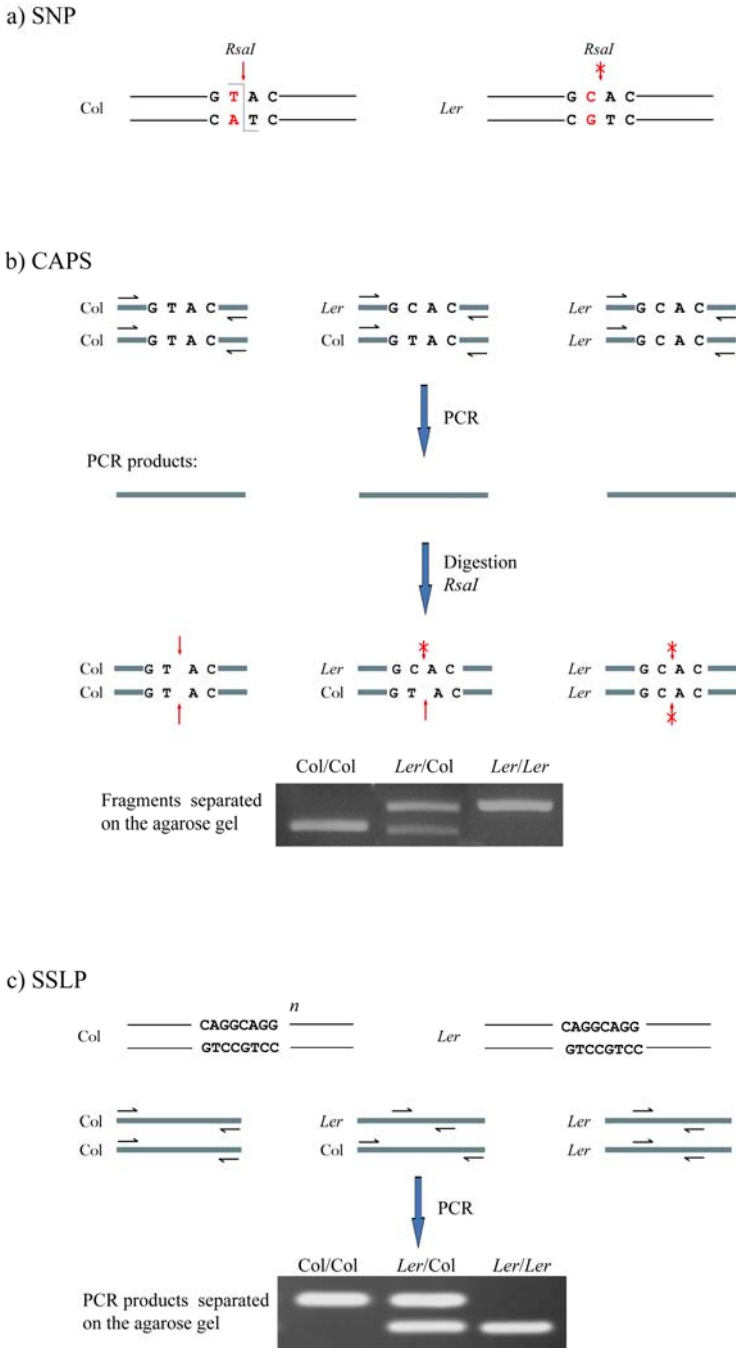

Figure 4. Schematic illustration showing genotyping of CAPS and SSLP markers. a) A single nucleotide polymorphisms (SNP) at a particular position in Col and L*er* ecotypes. The polymorphism affects *RsaI* restriction site in L*er* ecotype; L*er* can not be digested by *RsaI*, while Col is digested. b) Primers designed to amplify a genomic fragment containing the SNP. Grey lines represent homolog chromosomes. The *RsaI* restriction sequence is written in 5'→3' direction for each homologue chromosome. PCR products are digested with *RsaI* restriction enzyme and separated on the agarose gel. Three different genotypes are detected: Col/Col (shown as one short bond on the gel), L*er*/Col (shown as two bonds on the gel: shorter, Col (digested), and longer, L*er* (undigested)) and L*er*/L*er* genotype, undigested PCR bond. c) A microsatellite DNA region at the particular locus in Col has more repeats than in L*er* ecotype. Primers are designed to anneal outside the microsatellite regions. PCR products are separated on the agarose gel and the three different ecotypes can be detected by different fragments length.

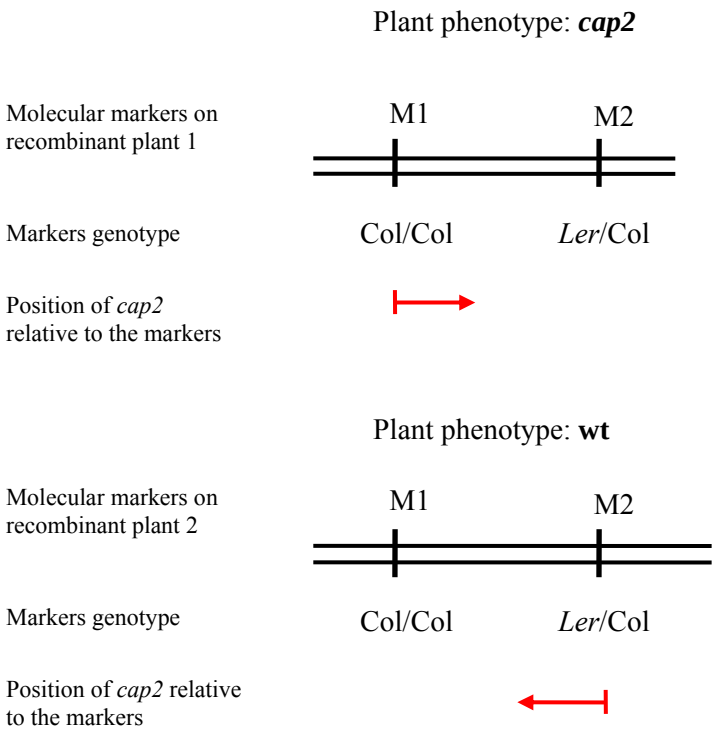

Figure 5. An illustration showing two plants from the *cap2* mapping population having the same genotype Col/Col and L*er*/Col in molecular markers M1 and M2, respectively, but different phenotypes. Plant 1 is *cap2* and plant 2 is the wt. The arrow is showing direction of *cap2* mutation relative to these two markers. In the case of plant 1, the genotype in marker M1 (Col/Col) leads to conclusion that the mutation is located to the right of M1 marker, since this plant has *cap2* phenotype and the mutation is present in L*er* ecotype. In the case of plant 2 that have the wt phenotype, the genotype in the marker M2 (L*er*/Col) indicates that the mutation is positioned to the left of this marker, i.e. in that part of the genome having the Col/Col genotype.

2. Materials and methods

2.1 DNA methods

2.1.1 Isolation of genomic DNA from *Arabidopsis*

Isolation of genomic DNA from *Arabidopsis* was done using Aquapure Genomic DNA Isolation Kit (BioRad). The protocol described by manufacturer was modified and scaled down as described in the following.

2-4 rosette or stem leaves were frozen in liquid nitrogen (N_2). The tissue was ground to a fine powder using a drill. 215 µl of Cell Lysis Solution was added to the tissue and the tube was vortexed for 1-3 seconds to wet the tissue. After incubating the cell lysate for 60 min. at 65°C, and cooling it down to room temperature, 72 µl of Protein Precipitation Solution was added to the cell lysate. The samples were vortexed at high speed for 20 seconds to mix the Protein Precipitation Solution uniformly with the cell lysate. The samples were centrifuged at 3000 x g for 10 min., and the supernatant containing the DNA was poured into an eppendorf tube containing 215 µl 100% isopropanol. After mixing by inverting 50 times, the samples were centrifuged at 3000 x g for 5 min. The supernatant was removed and the tubes with DNA pellet were dried for 10 min. at room temperature to allow all isopropanol to evaporate. The DNA pellet was washed with 215 µl 70% ethanol (EtOH) and centrifuged at 3000 x g for 5 min. The EtOH was removed, and the pellet was dried at 37°C for 10 min. to allow all EtOH to evaporate. The DNA pellet was dissolved in 50 µl DNA Hydration Solution by incubating the samples for 60 min. at 65°C. When cell debris was still present in the dissolved DNA samples, the samples were centrifuged at 16.000 x g for 10 min. The supernatant were transferred to a clean tube, and the DNA samples were stored at 4°C.

2.1.2 Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) was used to amplify desired fragments of DNA for genotyping of *cap2* mapping population, amplification of genes and fragments of interest for cloning, screening of bacterial colonies after transformation and genotyping of T-DNA lines. A target sequence is amplified in three major PCR steps. In the first step, the template DNA is denatured due to high temperature (94°-95°C). The primers hybridize in the next step

(54°-65°C depending on the melting temperature (T_m) of the primer pair) to the complementary sequences on opposite DNA strands and flank the target sequence that is to be amplified. The optimal annealing temperature has to be determined empirically. In reactions performed in this work the annealing temperature was about 5°-10°C higher than T_m of the primers in order to avoid amplification of non-specific DNA fragments. During the last step (72°C or 68°C depending on the enzyme) the annealed primers are extended by a heat-stable DNA polymerase. A repetitive series of 25-40 cycles, involving these three steps results in exponential accumulation of a specific DNA fragment.

Standard set-up for PCR was 1x reaction buffer, 200 µM dNTP mix, 0.2-0.4 µM of each primer, 0.5-1 U DNA polymerase (DyNAzyme™ II DNA Polymerase (Finnzymes), *Taq* DNA Polymerase (New England BioLabs) or BD Advantage™ 2 Polymerase Mix (Clontech)), 1 ng of plasmid or 10-100 ng of genomic template, and Milli-Q H_2O to adjust to desired volume. DyNAzyme and *Taq* DNA Polymerase are two standard thermostable DNA polymerases that possess 5'→3' DNA polymerase activity and 5'→3' exonuclease activity, but not 3'→5' exonuclease (proofreading) activity. They were used for screening and genotyping. BD Advantage™ 2 Polymerase Mix contains *Taq* DNA polymerase, BD TaqStart™ Antibody (to provide automatic hot-start PCR) and a small amount of a proofreading polymerase. This was used for amplification of fragments that were to be cloned, because of the enzyme's high efficiency and accuracy of amplification.

Amplification of some polymorphic DNA markers was not successful using the buffer provided with the DyNAzyme or *Taq* DNA Polymerase (Table 1). The optimal buffer for a particular PCR may vary, depending on the target and the primer sequences, concentration of several components in the reaction, such as dNTP, primers, concentration of Mg^{2+} and H^+ (pH value) (Cha and Thilly, 1993; Roux, 1995). *Taq* polymerase requires Mg^{2+} ions for its activity (Linz et al., 1990). Mg^{2+} ions form soluble complex with dNTPs (in a 1:1 molar ratio) to produce the actual substrate that the polymerase recognizes (Roche, 1999). To work properly, Taq polymerase requires also free Mg^{2+} ions. An increase in the concentration of dNTPs decreases the concentration of free Mg^{2+} ions available to influence polymerase function (Roux, 1995). Each PCR has an optimal ion concentration, pH value, primer

concentration and sequence, hybridization temperature and number of cycles (Linz et al., 1990).

Amplification of two CAPS (Cleaved Amplified Polymorphic Sequences) markers, G17311 and GL2, was optimized using twelve different buffers with varying concentrations of Mg^{2+}, K^+ and H^+ ions (Table 2) and also tested in a temperature gradient. The concentration of dNTPs and primers was not changed in the optimization reactions.

G17311 and GL2 markers were successfully amplified using *Taq* DNA Polymerase (New England Biolabs) together with buffer 12.

All programs used were variation of the general program shown in Table 3. Annealing temperature, elongation time and number of cycles used with the particular primer pair are shown in Appendix 1 and Appendix 2. Primer sequences are shown in Appendix 3.

Table 1. Concentrations of Mg^{2+} and K^+ and pH values for buffers provided with DyNAzyme or *Taq* DNA Polymerase.

1x Reaction buffer	[Mg^{2+}] (mM)	[K^+] (mM)	[TrisHCl] (mM)	pH
DyNAzyme	1.5	50	10	8.8
Taq DNA Polymerase	2	10	20	8.8

Table 2. Concentrations of Mg^{2+} and K^+ and pH values for different buffers used in optimization.

Buffer no.	[$MgCl_2$] (mM)	[KCl] (mM)	[TrisHCl] (mM)	pH
1	1.5	25	10	8.3
2	1.5	75	10	8.3
3	3.5	25	10	8.3
4	3.5	75	10	8.3
5	1.5	25	10	8.8
6	1.5	75	10	8.8
7	3.5	25	10	8.8
8	3.5	75	10	8.8
9	1.5	25	10	9.2
10	1.5	75	10	9.2
11	3.5	25	10	9.2
12	3.5	75	10	9.2

Table 3. The standard PCR program.

	Temperature	Time	Number of cycles
Initial denaturation	94°C	4 min	
Denaturation	94°C	30 sec	
Annealing	54° - 64°C*	30 sec	25-40*
Elongation	72°C or 68°C**	30 sec - 6 min*	
Final Elongation	72°C or 68°C**	7 min	

*see Appendix 1 and Appendix 2.
**72°C is an optimal elongation temperature for DyNAzyme and *Taq* DNA Polymerase. 68°C is an optimal elongation temperature for BD Advantage™ 2 Polymerase.

2.1.3 Reverse transcriptase PCR (RT-PCR)

Reverse Transcriptase PCR (RT-PCR) is a highly sensitive method for determining gene expression at the RNA level. The RT-PCR was performed to confirm the expression of endogenous genes, *LYSINE DECARBOXYLASE* (*LDC)*, *SOLANESYL DIPHOSPHATE SYNTHASE* (*SPS*) and *PENTACYCLIC TRITERPENE SYNTHASE* (*ATPEN6*) in different *Arabidopsis* organs (flowers, seedlings, rosette leaves and siliques). 1µl of first-strand cDNA synthesized in the reverse transcriptase reaction (see section 2.3.2) was used as the template in the RT-PCR. The reaction was performed as shown in Table 3 and Appendix 2. All primer pairs were designed to span intron regions so that amplification of RNA could be distinguished from amplification of any contaminating genomic DNA. Actin RT-PCR was done as template control, since the actin gene is highly expressed at a relative constant level in all tissues.

2.1.4 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate and identify DNA fragments according to size. PCR products and digested DNA fragments were analyzed on 1-2% (w/v) agarose (SeaKem® LE agarose, BioWhittacer Molecular Applications) gel, depending on expected size of fragments to be separated. Fragments differing less than 50 bp from each other were separated on 3% agarose gel (1/3 mass of NuSieve 3:1 agarose, 2/3 mass of SeaKem® LE agarose, BioWhittacer Molecular Applications). For visualizing DNA fragments, the gel was added 0.6 µg/ml ethidium bromide (EtBr). EtBr intercalates to DNA, and fluorescence under UV light (Sharp et al., 1973). The samples were loaded with 1/6 volume of loading buffer (FBX), and run in 1x TAE buffer (40 mM Tris-acetate, 1 mM EDTA) at 60-100 V for 30-60 min. To determine the size of the migrated DNA fragments, standard DNA ladder and

markers were used (GeneRuler™ 1kb DNA Ladder, GeneRuler™ 100kb DNA Ladder and λ DNA/ *Hind*III Marker, Fermentas).

2.1.5 Restriction enzyme digestion of DNA

Digestion of DNA with various restriction endonucleases was performed in 30 µl reactions containing 20 µl of PCR product or 1-2 µl of plasmid DNA, 1\10 volume of buffer provided with the enzyme, 1\100 volume of BSA (if not contained in the buffer), and 5-10 U of restriction enzyme. The reaction was incubated at temperature recommended by the respective endonuclease manufacturer. It was important to use sufficient amount of enzyme, especially in genotyping of *cap2* plant population in order to ensure complete digestion, as incomplete digestion could lead to misleading conclusions regarding genotype.

2.1.6 Control digestion of DNA

Three different genotypes (plant ecotype) may be found in a given CAPS marker of the *cap2* mapping population: Columbia (Col) homozygote, Landsberg *erecta* (L*er*) homozygote, and Col/L*er* heterozygote. The genotype in a particular CAPS marker was determined by digesting the amplified sequence with the proper restriction enzyme. Wild-type plants with known genotype (homozygote Col, homozygote L*er* and heterozygote Col/L*er*) were used as controls. DNA from these plants was isolated and the CAPS markers amplified. The sequence of each CAPS marker was cleaved with appropriate enzyme in order to check the cleaving pattern and band size for the three different genotypes in a given CAPS marker.

2.1.7 Isolation and purification of DNA-fragments from agarose gel

DNA fragments of interest were cut from 1% agarose gel using a clean scalpel and thereafter purified according to the Wizard® SV Gel and PCR Clean-Up System (Promega). The kit is provided with columns containing silica membranes that bind DNA in the presence of chaotropic salts (Membrane Binding Solution). After washing the column twice with the Membrane Wash Solution containing EtOH, the column was dried for 10 min. at 37°C to allow evaporation of EtOH left in the column. This step is not according to the manufacturer's protocol, but it gives better DNA yield since the EtOH can influence the measurements of DNA concentration. The DNA was eluted in 50 µl Nuclease-Free Water provided with the kit.

2.1.8 Purification of DNA

PCR fragments not run on the gel were purified in order to remove residues of primers, nucleotides, polymerases and salts using the same kit as for gel extraction (Wizard® SV Gel and PCR Clean-Up System (Promega)). The procedure is based on the same principles as for gel extraction.

2.1.9 Concentration of DNA

The DNA purified from the gel had often lower concentration than needed for the Gateway cloning. Therefore the purified and diluted DNA had to be concentrated prior to cloning. The method used for this purpose is based on precipitation with isopropanol.

1/10 volume of 3 M sodium acetate (NaAc) was added to the sample to adjust the concentration of monovalent cations. DNA was precipitated by adding 1 volume of isopropanol to the sample and incubating at -20°C for 30min. After centrifugation at 12.000 x g for 15 min. the isopropanol was poured off, and the sample was dried for 10 min. at 37°C to allow evaporation of isopropanol. The DNA was resuspended in nuclease free water or TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) to desired concentration.

2.1.10 Quantification of DNA

Quantification of DNA was done with Hoefer® DyNAquant 200 fluorometer (Amersham Pharmacia Biotech) as described by the manufacturer. Bisbenzimide, commonly known as Hoechst 33258, is a fluorescent dye that binds with high specificity to the minor groove of DNA. The Hoechst 33258 dye has different fluorescence characteristics when bound to DNA compared to when no DNA is present. This is used in DNA quantification.

2.1.11 Cloning of PCR product using Gateway Cloning Technology

The Gateway® Cloning Technology (Invitrogen Life Technologies) is based on the site-specific recombination system used by bacteriophage lambda (λ) to integrate into *E. coli* chromosome and switch between the lytic and lysogenic life cycles (Landy, 1989). The principle of this recombination system is based on rapid and efficient way to transfer DNA sequences into and between different vectors. The two major components involved in λ-based recombination system are the specific DNA recombination sequences (*att* sites) and

the enzymes that catalyze the recombination reaction by binding to the *att* sites (Landy, 1989).

The fragments to be cloned were amplified using primers labeled with attB1 and attB2 Gateway® sequences (see Appendix 3). The wild-type λ *att* recombination sites have been modified in the Gateway® cloning system. These modifications ensure specificity and efficiency of the Gateway® reactions by maintaining reading frame and orientation of the DNA segment during recombination (Invitrogen, 2003). This means that the modified *att*B recombination sites, *att*B1 and *att*B2 (25 bp) will recombine with modified *att*P1 and *att*P2 (200 bp) sites respectively, but not the other way around. The same also applies for *att*L (100bp) and *att*R (125bp) sites, where *att*L1 sites react only with *att*R1 sites, and *att*L2 sites react only with *att*R2 sites.

In the first Gateway® recombination reaction, BP reaction, the purified *att*B-flanked PCR product was recombined into a donor vector containing *att*P sites to create an entry clone. The promoter region of lysine decarboxylase (*pLDC*) gene (1509 bp) was recombined into pDONR™207 (Invitrogen). *SPS* gene (3600 bp) was recombined into pDONR™201 (Invitrogen) and *ATPEN6* gene (5537 bp) was recombined into pDONR™/Zeo (Invitrogen).

150 ng of donor vector was mixed with 100 fmol of PCR product, 1x BP reaction buffer and BP Clonase™ enzyme mix containing Integrase (Int) encoded by the phage λ, and Integration Host Factor (IHF) encoded by *E. coli* (Landy, 1989). The reaction was incubated over night at the room temperature. After incubation, proteinase K was added to deactivate the clonase enzymes, and the sample was incubated at 37°C for 10 min. 1 µl of recombination mix was used to transform DH5α competent gyrA+ *E. coli* cells. The transformed cells were plated on LA-plates containing the selective antibiotic, and were subject to double selection. Cells not containing the entry clone could not grow on the antibiotic medium, and the cells containing the donor vector without the insert died because of the *ccdB* gene carried on the donor vector (see also 2.2.2). The product of this reaction is an entry clone containing the insert of interest flanked by *att*L sites produced in recombination between *att*B and *att*P sites.

The second Gateway® recombination reaction, LR reaction, was used to create an expression clone. In this reaction the insert of interest was moved from the entry clone into a destination vector. The destination vector contains a gateway cassette. The cassette is a chloramphenicol resistance gene, and the *ccdB* gene flanked by modified sequences of the *att*R sites. In this reaction the *pLDC* fragment was moved from pDONR™207 vector into a pPZP211G Gateway destination vector in front of promoterless *GUS* gene. The pPZP211G vector (Butenko et al., 2003) contains C.1 Gateway cassette in front of the *GUS* gene, spectinomycin bacterial selectable marker and a plant selectable marker *nptII*. The SPS and ATPEN6 fragments were moved from pDONR™201 and pDONR™/Zeo respectively into pGSC1704 Gateway destination vector for complementation analysis. The pGSC1704 vector (Butenko et al., 2003) contains spectinomycin bacterial selectable marker and hygromycin resistance gene within the T-DNA for plant selection.

150 ng of entry clone was mixed with 150 ng of destination vector, 1x LR reaction buffer and LR Clonase™ enzyme mix containing the phage λ integrase (Int) and exitionase (Xis) enzymes and the *E. coli* integration host factor (IHF) encoded by *E. coli* (Landy, 1989). The reaction was incubated over night at room temperature. Proteinase K was added after incubation to deactivate the clonase enzymes, and the sample was incubated at 37°C for 10 min. 1 µl of recombination mix was used to transform DH5α competent gyrA+ *E. coli* cells. The transformed cells were plated on LA-plates containing the selective antibiotic (see 2.2.2).

2.1.12 TOPO TA Cloning®

TOPO TA Cloning® (Invitrogen) was used to subclone the probes used in northern hybridization. This is a highly efficient and rapid cloning system that provides direct insertion of *Taq* polymerase-amplified PCR product into a plasmid vector. The vectors supplied with the kit are linearized and have single, overhanging 3′ deoxythymidine (T) residue and covalently bound topoisomerase I enzyme. PCR product amplified with a *Taq* polymerase has a singe deoxyadenosine (A) overhang on the 3′ ends added by the *Taq* polymerase. This allows PCR insert to base-pair with 3'-T overhang on the vector, and ligate efficiently with it. The PCR products were cubcloned into pCR®2.1-TOPO® vector

(Invitrogen), and the cloning reaction was set up according to the TOPO TA Cloning® manual.

2.1.13 Isolation of plasmid DNA from bacteria cultures

The method is based on principle that treatment of a bacterial culture with alkali and SDS opens the cell wall, denatures chromosomal DNA and proteins, and realizes plasmid DNA into the supernatant. By adding the neutralization solution (acetate) to the cell lysate, denaturized proteins and chromosomal DNA will precipitate and can be pelleted. Plasmids, which are retained in the supernatant, can be precipitated with isopropanol.

Isolation of plasmid DNA was done using Quantum Prep® Plasmid Miniprep Kit (BioRad) and Wizard® *Plus* SV Minipreps DNA Purification System (Promega). The method was performed as described in protocol from manufacturer. Both miniprep kits are supplied with a columns containing DNA binding matrix. The supernatant containing the plasmids was transferred to column that bound the DNA. After washing, the DNA was eluted from the column with nuclease free water or TE buffer. The eluted DNA was stored at -20°C.

2.1.14 Sequencing

For sequencing of DNA fragments, the dideoxy, or chain termination method (Sanger et al., 1977) was used. The method involves the synthesis of new strands of DNA complementary to a single-stranded template. Strand synthesis does not proceed indefinitely because the reaction mixture contains small amounts of dideoxynucleotides (ddNTPs), that blocks further elongation because they have a hydrogen atom rather than hydroxyl group attached to their 3'-carbon. The termination points are therefore nucleotide specific but occur randomly along the length of the target DNA. The fragments to be sequenced were cloned in a vector, and 350-500 ng of a plasmid DNA was concentrated in 11 µl. Automated sequencing was performed on the MegaBACE 1000 using the DYEnamic ET Dye Terminator Cycle Sequencing Kit (Amersham Pharmacia Biotech) with suitable primers (2µM). Usually one forward and one reverse primer was used. MegaBACE 1000 separates the different DNA fragments by capillary electrophoresis, and the different fluorescent tags are attached to each of the different ddNTPs, for detecting which ddNTP has terminated the synthesis. The sequencing reactions were prepared and purified as recommended by the manufacturer.

2.1.15 Bioinformatics analysis

Bioinformatics was used in this thesis for primer design, alignment of sequences and generally for basic analysis of data and sequences. The Arabidopsis Information Resource (TAIR) (http://arabidopsis.org) was used for searching for *Arabidopsis* BACs, genes and sequences. The single nucleotide polymorphisms between Col and L*er* ecotypes were found using information on TAIR and Cereon Genomics (Monsanto) (http://www.arabidopsis.org/cereon/). Vector NTI Suite 8.0 (InforMax) was used for primer design and the restriction analysis. Alignment and analysis of sequencing results was done using Contig Express (InforMax).

2.2 Bacterial methods

2.2.1 Transformation of bacteria

Transformation of *Escherichia coli*

All vectors were transformed into DH5α competent *E. coli* cells by heat shock transformation, except TOPO® Cloning vectors, which were transformed into DH5α cells using a rapid transformation protocol.

The heat shock transformations were done by adding 2-4 µl of plasmid to 50 µl of competent cells and incubating for 30 min. on ice. The cells were heat shocked for 30 seconds at 42°C and transferred immediately to ice for 2 min. 450 µl of S.O.C medium (2% Tryptone, 0.5% Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) was added and the cells were incubated for one hour at 37°C by horizontally shaking (200 rpm). After incubation, 50 µl and 100 µl of cells were plated on pre-warmed LA-plates (10 g/l Peptone, 5 g/l Yeast Extract, 10 g/l NaCl, 5 g/l agar) containing selective antibiotic, and incubated over night at 37°C.

TOPO® Cloning reaction was transformed into DH5α competent *E. coli* cells using a rapid transformation protocol, which is only recommended for transformations using ampicillin selection. 4 µl of TOPO® Cloning reaction was added to 50 µl *E. coli* cells and incubated on ice for 5 min. The cells were spread on a pre-warmed LA-plate containing 100 µg/ml of ampicillin and incubated over night at 37°C. For selection of positive colonies, 40 µl of X-

gal (40 mg/ml) was spread on the LA-plate and incubated for 30 min. at 37°C prior to cell-spreading.

Transformation of *Agrobacterium tumefaciens*

The *Agrobacterium tumefaciens* strain C58 pGV2260 was used for transformation of wild-type *Arabidopsis*. This strain was transformed by both pPZP211G-GAWI and pGSC1704 Ti-vectors.

The competent *Agrobacterium* cells were transformed by electro-transformation. 1 µl of plasmid (0.5-1 ng/µl) was added to the cells and incubated on ice for 30-60 sec. The mixture was moved to a cold electroporation cuvette (Bio-Rad) and shocked at 25 µF, 200Ω, and 2.4 kV. The S.O.C medium was added immediately to the cells, and incubated one hour at 28°C with shaking. For selection of transformants, the culture was plated on YEB-plates (5 g/l Beef Extract, 1 g/l Yeast Extract, 1 g/l Peptone, 5 g/l sucrose, pH 7.4, 2 ml/l $MgSO_4$ added after autoclaving, 15 g/l agar) containing selective antibiotic. The plates were sealed with a parafilm and incubated for 2-3 days at 28°C.

2.2.2 Selection of transformants

E .coli

All vectors contain an antibiotic resistance gene that makes the selection of transformed cells possible. Only the cells transformed by plasmid carrying an antibiotic resistance gene will grow on LB-plates added antibiotic. Entry clones, produced in BP recombination reaction where pDONR™207 was used as a donor vector, were selected on LA-plates containing 7 µg/ml of gentamycin. In the reaction where pDONR™201 was used as a donor vector, the entry clones were selected on LA-plates added 50 µg/ml of kanamycin, while the pDONR™/Zeo entry clones were selected on LB-medium containing 25 µg/ml of zeocin. Expression clones produced in LR recombination reaction were selected on LA-plates containing either 50 µg/ml of spectinomycin or 100 µg/ml of streptomycin when both pGSC1704 and pPZP211G-GAWI destination vectors were used.

In addition to the antibiotic resistant gene, vectors used in Gateway® cloning contain also *ccdB* gene, which encodes for a protein that is toxic for the DH5α, *E. coli gyrA+* strain. The ccdB protein binds to the *DNA gyrase subunit A*, the product of the *gyrA* gene in *E. coli*,

turning it into a cellular poison (Bahassi et al., 1999). Thus, when an *E. coli* gyrA+ strain is transformed by such vectors, the ccdB gene product blocks bacterial growth. However, if the cassette harboring the *ccdB* gene in the Gateway® vector is recombined out and replaced by a foreign DNA fragment, this recombinant plasmid no longer interferes with host viability.

Blue/white selection was used to screen positive transformants when TOPO TA Cloning® was used. The pCR2.1-TOPO vector used in TOPO TA Cloning® contains two antibiotic resistance markers, ampicillin and kanamycin, for selection of bacterial colonies containing the vector. In addition to the antibiotic resistance markers, the vector contains also a portion of the *lacZ* gene (lac-α subunit) encoding the enzyme *β-galactosidase*. DH5α strain contains the rest of the *lacZ* gene, *lacZ*ΔM15, which express a fragment that complements the lac-α subunit encoded by the vector. Although neither the lac-α subunit encoded by the vector nor the segment of *lacZ* gene encoded by DH5α cells is active alone, the polypeptide encoded by each region associate in host cell to form active β-*galactosidase* enzyme (Ullmann et al., 1967). The pCR2.1-TOPO vector contains a polylinker in *lacZ*α coding region. Successful cloning of a DNA fragment into the polylinker of pCR2.1-TOPO vector will interrupt the reading frame of the *lacZ* gene and complementation will not occur. When spreading 40 µl of 40 mg/ml X-Gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) substrate on top of the agar, the functional enzyme β-*galactosidase* will hydrolyze the X-Gal in bacteria not containing the insert, forming an intense blue precipitate. DH5α strain does not require IPTG to induce expression from *lac* promoter even though the strain expresses the Lac repressor. The copy number of most plasmids exceeds the repressor number in the cell (Invitrogen). However, the bacteria containing the pCR2.1-TOPO vector with inserted DNA fragment will not be able to make the functional enzyme, and the colonies will remain white.

Agrobacterium tumefaciens

Agrobacterium cells containing pPZP211G-GAWI and pGSC1704 vectors with their respective inserts were selected on YEB- medium containing streptomycin, carbenicillin and rifampicin, each in concentration of 100mg/l.

2.2.3 Growth and storage of bacteria

E. coli cultures were grown in LB-medium (10 g/l Peptone, 5 g/l Yeast Extract, 10 g/l NaCl) containing selective antibiotic with same concentration as used in LA-plates. The cultures were grown under shaking at 37°C.

Cultures of *Agrobacterium* were grown in YEB-medium (5 g/l Beef Extract, 1 g/l Yeast Extract, 1 g/l Peptone, 5 g/l sucrose, 2 ml/l $MgSO_4$ added after autoclaving) at 28°C under shaking. Initially, the cultures were grown as 5 ml cultures for one day. 20 ml of YEB-medium was added next day, and the cultures were grown for one more day, thereafter the volume was increased to final 250 ml. The 250 ml cultures were grown further to an OD_{600} of 1.2.

For long-term storage, 1.5 ml of culture was centrifuged and the cell pellet was resuspended in 1.5 ml of 25% glycerol. The cells were stored at -80°C.

2.2.4 Screening of transformants

Screening of transformants was done by PCR on plasmids isolated from over-night cultures preferably using one vector-specific, and one gene-specific primer. Sometimes a set of gene-specific primers was used, and a PCR-product was cleaved by a specific restriction enzyme. As en extra control, the entire plasmids were also cleaved by a specific restriction enzyme. Finally, all the positive plasmids were sequenced to make sure that the vector contained the right insert.

2.3 RNA methods

Northern blotting and reverse transcriptase (RT) reaction (first-strand cDNA synthesis) were two RNA methods used in this thesis. Northern blotting was used to examine expression and relative mRNA levels of lysine decarboxylase (*LDC*) and histone deacetylase (AtHD2C) genes in different plant tissues. Reverse transcriptase reaction was performed to generate first-strand cDNA that was used as a template in reverse transcriptase-PCR (RT-PCR). The first strand cDNA was also used to amplify hybridization probe templates for Northern blotting.

2.3.1 mRNA isolation

The plant tissue used for mRNA isolation was C24 ecotype of *Arabidopsis*. Poly-A mRNA was extracted from 100 mg N_2 frozen flowers, seedlings, rosette leaves and siliques using magnetic oligo-dT_{25} beads (GenoPrep mRNA beads; GenoVision) as described by manufacturer. The plant tissue was homogenized for 1-2 min. in lysis/RNA-binding buffer (100 mM Tris-HCl, pH 8.0, 500 mM LiCl, 10 mM EDTA, pH 8.0, 1% LiDS, 1 mM DTT). The homogenized material was centrifuged and the supernatant containing the mRNA was added to the magnetic oligo-dT beads and incubated on ice for 5 min. Oligo-dT residues base-pair with poly-A tail on mRNA, and when placed on magnet, the beads hybridized to mRNA will bind to the magnet wall and the solution can be removed. After washing the beads two times in middle wash buffer with SDS (150 mM LiCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5% SDS) and three times in middle wash buffer without SDS they were divided into two parts.

The mRNA used for Northern blotting was eluted from the magnetic beads by incubating on 65°C for 3 min. The concentration was measured spectrophotometrically (GeneQuant, Pharmacia) and the mRNA was precipitated with ice cold 250 µl 96% EtOH + 10 µl 3M NaAc pH 5.2 at -20°C for 30 min. The mRNA was centrifuged for 40 min. at maximum speed at 4°C, and the pallet was dissolved in formaldehyde loading buffer to a concentration of 100ng/µl.

The mRNA hybridized to the magnetic beads that was used in RT reaction was washed with 1xRT-buffer (50 mM Tris, pH 8.3, 7 mM $MgCl_2$, 40 mM KCl, 0.1 mg/ml BSA), resuspended in 1x RT buffer and kept on ice prior to RT reaction.

2.3.2 Reverse transcriptase (RT) reaction

In the reverse transcriptase reaction, the enzyme AMV reverse transcriptase, use the mRNA bound to the magnetic beads as a template to synthesize first-strand cDNA as an extension to oligo-dT residues.

1xRT buffer was removed from the magnetic beads, and 42°C pre-warmed solution containing 20.5 µl dH_2O, 6 µl 5X RT-buffer (Promega), 1.9 µl dNTPs (4 x 2 mM), 0.8 µl

RNasin (Promega, 40 U/µl) was added to the beads. The reaction mix was kept warm on 42°C before adding 0.8 µl AMV reverse transcriptase enzyme (Promega, 10 U/µl), followed by incubation at 42°C for 1 hour. After washing with 500 µl dH_2O, the beads were added 100 µl dH_2O and incubated at 65°C for 3 min. to elute mRNA from the newly synthesized first-strand cDNA. The beads holding the first-strand cDNA bound to oligo-dT were resuspended in 20 µl 10 mM Tris pH 7.5 and kept at 4°C.

2.3.3 Northern gel

The mRNA was size-fractionated on a denaturing northern gel (1.7% agarose, 1x running buffer (RB), 6% formaldehyde). Before loading the mRNA on the gel, it was heat-denatured at 65°C for 5 min. to remove any secondary structures. Equal amounts of purified mRNA (850 ng) was loaded on the gel, and run for about 4 hours at 60V in northern gel buffer (1xRT, 6% formaldehyde).

2.3.4 Northern blotting

(Sambrook and Russel, 2001)

After separating the mRNA on the gel, the gel was placed upside down on a clean glass plate and covered with a piece of Hybond-N+ nylon membrane (Amersham Life Science). Three peaces of 3MM paper (wetted in 10X DEPC-SSC) and a stack of paper towels was placed over the membrane. The paper towels were topped with a glass plate and weights. Finally, everything was covered with a plastic foil to avoid evaporation, and left over night. The mRNA was transferred to the membrane by capillary action. In this setup, the paper towels act as a blotter, pulling the buffer and the mRNA through the gel. In this way the mRNA become trapped in the membrane. The membrane was baked at 60°C for 10 min., and UV-crosslinked at 70000 micro joules and 254 nm shortwave UV light for 1 min. (Hoefer Scientific Instruments). Shortwave UV light causes the nitrogenous bases in RNA, mostly uracil, to become highly reactive and to form covalent linkages to amine groups on the surface of the membrane (Ambion, 2005).

2.3.5 Probe labeling

Hybridization probe templates for *LDC* and *AtHD2C* mRNA were amplified from oligo-dT beads with first-strand cDNA, while the template for 18S rRNA probe was genomic DNA

from *Arabidopsis*. The PCR fragments were extracted from agarosegel and 25 ng of DNA was radioactively labeled with ^{32}P dCTP using Rediprime II, random prime labeling system (Amersham Biosciences). In this reaction the DNA was denatured by heating to 100°C, cooled on ice for 5 min. and added to Rediprime II reaction tube containing buffered solution of dATP, dGTP, dTTP, Klenow polymerase enzyme and random primers. 5 µl ^{32}P dCTP was added to the reaction and incubated at 37°C for 10 min. The reaction was stopped by adding 5 µl of 0.2M EDTA. For use in hybridization, the labeled DNA was denaturized by heating to 100°C for 5 min.

2.3.6 Hybridization

(Galau et al., 1986)

The specific mRNAs and their intensity were detected by hybridizing the membrane by a specific single stranded radioactive probe. As a control for equal loading of each sample, the 18S rRNA was hybridized to each blot. High temperature and buffer salt concentration was used for stringency to reduce non-specific binding to related sequences. The pre-hybridization solution and the hybridization solution also contained heterologous salmon sperm DNA, witch reduces non-specific binding.

Prior to pre-hybridizing for 3 hours in pre-warmed pre-hybridization solution (0.7x SSC, 1x SPEP, 5x Denhardts, 100 µg/ml heterologous DNA), the new membrane was wetted for 30 min. in wetting solution (0.7x SSC, 1x SPEP). The probe was added to the pre-hybridization solution and boiled for 10 min. to denature it. The pre-hybridization solution was removed and hybridization solution with the probe was added to the membrane and hybridized over night at 68°C.

The membrane was first rinsed twice with pre-warmed wash solution (0.7x SSC, 1x SPEP, 1x Denhardts), and then washed for 45 min. at 68°C until background signal was reduced.

2.3.7 Autoradiography

After washing, the membrane was wrapped in plastic foil and placed in an exposure cassette with a Storage Phosphor Screen (Amersham Biosciences) on top of it. Expression data were quantified with a Typhoon 9400 phosphorimager (Amersham Biosciences), running the

ImageQuant Software. The Storage Phosphor Screen retains energy from beta particles (emitted by ^{32}P), X-rays, and gamma rays, and upon laser-induced stimulation, light is emitted from the screen in proportion to the amount of radioactivity in the sample. The image is then immediately scanned into the computer's database.

2.3.8 Stripping of membrane

In order to hybridize the membrane with the new probe or to store it for a longer period, the old probe has to be removed by stripping the membrane using 0.1% SDS and 1mM EDTA. The membrane was incubated with shaking for 30 min. in stripping solution, pre-warmed to boiling temperature. After stripping, the membrane was wrapped in plastic foil, and kept in a plastic folder at 4°C.

2.4 Arabidopsis methods

2.4.1 Plant strains

Arabidopsis plants used in mapping of *cap2* and plants transformed with the At1g78510 and At1g78500 complementation constructs were dihybrid between Landsberg *erecta* and *Columbia* ecotypes. *pLDC::GUS* transgene construct was transformed into C24 ecotype.

2.4.2 Building the *cap2* mapping population

Initially, *Arabidopsis thaliana* multiple marker 1 (*mm1*) line, carrying five visible recessive marker mutations on chromosome 1 was crossed with the wild-type Landsberg *erecta*. The generated seeds, heterozygous for *mm1* line were mutagenized with ethyl methane sulfonate (EMS) as described previously (Grini et al., 1999). In order to isolate the EMS-induced gametophytic mutants, the M_2 plants were screened for distorted segregation of the *mm1* morphological markers. Lines segregating >40% of two adjacent *mm1* markers were rescreened and examined for aborted sees development (for details, see Grini *et al*., 1999).

For further genetic analysis and fine-scale mapping of *cap2* mutant, split *mm1* line, homozygous for *ap1* and *gl2* markers was used. This line was backcrossed to L*er* three times to get a line containing *cap2* without any visible marker mutations. *cap2* mapping population was made by crossing the heterozygote mutants as male partner to wild-type Col plants (Figure 6). F_1 plants showing the *cap2* phenotype were selected for further self-fertilization.

The resulting F_2 progeny constituted the mapping population. The mapping population plants were dihybrid between Col and L*er* ecotypes where the *cap2* mutation was in L*er* ecotype (Figure 6). The first recombination between these two ecotypes occurred in flowers of F_1 generation.

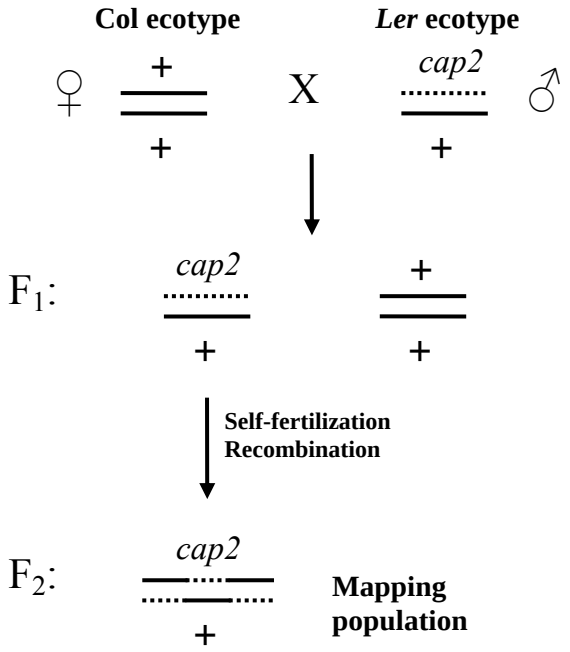

Figure 6. Schematic outline of the outcross performed to generate the *cap2* mapping population. Col wild-typed plants were pollinated by heterozygous *cap2* pollen in L*er* ecotype. F_1 progeny was phenotyped and plants with the *cap2* phenotype were selected for further self-crossing. Generated F_2 progeny of this self-cross comprised the mapping population.

2.4.3 Transformation of *Arabidopsis thaliana*

(Bechtold et al. (1993) modified by Clough and Bent (1998))

This method is based on *Agrobacterium tumefaciens* ability to infect plants and incorporate a segment of its Ti plasmid, called T-DNA, into the host cell genome. *Agrobacterium tumefaciens* strain C58 pGV2260 containing either pPZP211G-GAWI or pGSC1704 vectors with their respective inserts was grown under shaking at 28°C in 250 ml YEB-medium (5 g/l Beef Extract, 1 g/l Yeast Extract, 1 g/l Peptone, 5 g/l sucrose, pH 7.4, 2 ml/l $MgSO_4$ added after autoclaving), to an OD_{600} of 1.2. The YEB-medium contained the antibiotics

streptomycin (100mg/l), carbenicillin (100mg/l) and rifampicin (100mg/l) for selection of all constructs. The cells were centrifuged at room temperature for 10 min. at 5000 rpm and resuspended in freshly made 5% sucrose solution to an OD_{600} of 0.8. To get a better attachment of bacteria to the plants, Silwet L-77 was added to a concentration of 0.05%.

Arabidopsis plants were grown until flowering and all fully developed flowers were removed because they already were self-pollinated. Transformation of these flowers would hence not give the positive transformants. The entire plant containing immature flower buds was soaked in *Agrobacterium* solution for 5 min. Thereafter the plants were incubated over night in the dark under plastic to maintain high humidity. After 4-6 weeks the plants were harvested several times.

2.4.4 Seed sterilization and growth conditions

Prior to sowing and sterilization, the seeds were dried at 37°C for 4-10 days. Seeds used in mapping population and for transformation with *Agrobacterium* were sown directly on soil and germinated at 22°C with 16 hours light (light intensity of 100µE/m^2s) and 8 hours dark.

The surface sterilization method was used on transgenic seeds. For surface sterilization, seeds were added 70% EtOH, mixed by inverting the tube and incubated for 5 min. The EtOH was removed and bleach solution (20% chlorine, 0.1% Tween20) was added, mixed by inverting and incubated for 5 min. Finally, the seeds were washed twice in wash solution (sdH_2O, 0.001% Tween20), resuspended in 1ml sdH_2O and spread on MS-2 medium (0.5 g/l MES, 4.43 g/l Murashige and Skoog salts, 20 g/l sucrose, 5 g/l Bacto agar) (Murashige and Skoog, 1962). The MS-2 plates were sealed with surgical tape and transferred to the growth room after over night incubation at 4°C. They were cultivated at 22°C with 16 hours light (light intensity of 100µE/m^2s) and 8 hours dark. Seedlings were transferred to soil after 2-3 weeks and kept at high humidity for one additional week. Temperature and light conditions were the same.

Transgenic seeds were selected on MS-2 medium containing antibiotics. By adding antibiotics to the MS-2 medium, the wild-type seedlings will die at dicotyledonous stage, while the transformants will continue to grow and develop. *pLDC::GUS* transgene plants

were selected on MS-2 medium containing 50 µg/ml kanamycin, while plants containing solanesyl diphosphate synthase or pentacyclic triterpene synthase transgenes were selected on MS-2 medium containing 20 µg/ml hygromycin.

2.4.5 Dissection of siliques and microscopy

To be able to study seed phenotype and *GUS* expression in seeds under microscope and magnifying glass, siliques were dissected with needles and carpel walls were removed. Ovules that in this way remained connected to the placenta were placed on microscope slides in 4-5 drops of ClH solution (8:2:1 chloral hydrate:water:glycerol). The microscope slides were kept at 4°C.

The recombinant plants phenotype and *GUS* expression in seeds, flowers and rosette leaves was examined with Zeiss Axionplan Imaging2 microscope system equipped with Nomarski optics and cooled LCD imaging facilities.

2.4.6 Histochemical GUS assay

(Grini et al., 2002)

Histochemical GUS assay was used to analyze *pLDC::GUS* transformants for *GUS* expression. The *GUS* activity is obtained by a reaction where the GUS enzyme hydrolyzes substrate X-Gluc (5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid) to a water-soluble indoxyl intermediate that is further dimerized into a dibromo-dichloro-indigo blue precipitate at the site of enzyme activity by oxidative reaction (Jefferson, 1989). This dimerization is stimulated by atmospheric oxygen, and can be enhanced by using an oxidation catalyst such as a potassium-ferricyanide/ferrocyanide mixture (Lojda, 1970).

GUS-assay was performed on dissected siliques at different developmental stages, flowers and rosette leaves. The plant tissue was fixed in ice-cold 90% acetone for 10 min., rinsed in staining buffer without substrate (X-Gluc) for 10 min. and incubated in staining buffer (50 mM NaH_2PO_4, pH 7.2; 2 mM potassium-ferrocyanide; 2 mM potassium-ferricyanide; 0.1% Triton X-100; 2 mM X-Gluc) at 37° over night. The tissue was rinsed for 5 min. in a series of 10% EtOH, 30% EtOH and 50% EtOH to remove excess color and chlorophyll, and further fixed in FAA (10:7:2:1 EtOH:distilled water:acetic acid:37% formaldehyde) on ice

for 30 min. After fixation in FAA the tissue was hydrated for 5 min. in a series of 50% EtOH, 30% EtOH, and 10% EtOH, rinsed for 1 min. in 50mM NaH_2PO_4 buffer and kept on microscope slides in ClH solution (8:2:1 chloral hydrate:water:glycerol).

2.4.7 Segregation analysis

Genetic segregation analysis is a useful method to determine the number of T-DNA insertions and was used in this work to investigate number of complementation constructs per transformant line. T_2 seeds from the transformed lines carrying the construct were plated on MS-2 medium supplemented with hygromycin. Seedlings were scored for hygromycin resistance (Hyg^R) or sensitivity (Hyg^S) two to three weeks after germination. Those that did not develop beyond the dicotylous stage were considered to be hygromycin sensitive and did not contain any T-DNA. If a T_1 plant is hemizygous for one T-DNA insertion, and the T-DNA segregates as one Mendelian locus, the T_2 progeny would segregate on a hygromycin plate with a ratio of 75% of Hyg^R to 25% of Hyg^S (or 3:1 for Hyg^R:Hyg^S) seedlings. If the two T-DNA loci were integrated, the progeny would segregate at 93.75%:6.25% for Hyg^R:Hyg^S (or 15:1 for Hyg^R:Hyg^S). If there were three T-DNA insertions, the ratio would be 98.44%:1.56% (or 63:1 for Hyg^R:Hyg^S), and so on. Hence, the general ratio between Hyg^R:Hyg^S is $(1-0.25^n):0.25^n$, where n is number of T-DNA loci.

The number of Hyg^R and Hyg^S seedlings was counted and the hypothesis was tested with the chi-square test of significance.

2.4.8 The chi-square test

The chi-square test is a statistical test used to check whether observations fit with the expected results. In genetics, it is the commonly used test for evaluating if data fits with expected ratios. In order to perform the test, a null hypothesis must be formulated. The null hypothesis will typically be:

H_0: The difference between the observed and expected ratios is random, or they are not significantly different.

Calculation of the chi square statistics is done according to:

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

The degrees of freedom (*df*) is one less than the number of classes involved. The value for χ^2 is then used together with the *df* to determine a P value. This can be done either from tables or using an online calculator (e.g.: www.fourmilab.ch/rpkp/experiments/analysis/chiCalc.html). The size of the P value will then tell the probability of H_0 being true, i.e. that the deviation between the observed and expected ratio is random. Commonly a P value of 0.05 is used as limit for accepting or rejecting the null hypothesis. If P is smaller than 0.05, H_0 should be rejected, which means that the observed differences between the expected and observed ratios is significant. If P is larger than 0.05, it is more than 95% probability for the observed difference being random, meaning that the observed ratio is not significantly different from the expected ratio.

3. Results

3.1 Part A: Molecular mapping and characterization of the *cap2* gametophytic maternal-effect mutant

3.1.1 Initial mapping

Initially, the position of *cap2* was roughly mapped near the bottom end of the chromosome I, between *ap1* and *gl2* morphological markers as the self crosses of *cap2* mutants gave increased frequencies of these two flanking markers (42.4 % for *ap1* and 49.3% for *gl2*) (Grini et al., 2002). Genetic distance of *cap2* from the *ap1* and *gl2* markers was calculated from marker segregation data, and determined to be ~18 cM south of *ap1* and 2 cM north of *gl2* (Grini et al., 2002) (Figure 7).

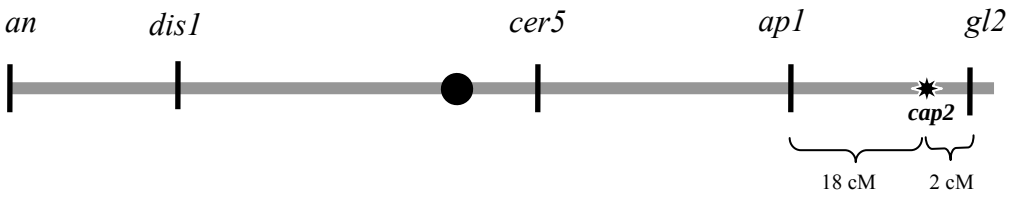

Figure 7. The multiple marker chromosome 1 (*mm1*) carrying five visible markers: *angustifolia* (*an*), *distorted-1* (*dis1*), *eceriferum-5* (*cer5*), *apetala-1*(*ap1*) and *glabra-2* (*gl2*). The position of *cap2* mutation was roughly mapped and determined to be 18 cM south of *ap1* and 2 cM north of *gl2*. The black circle represents the centromere.

In the initial mapping done by Grini et al. 2002, the *cap2* population consisted of 102 plants. 14 recombinants were found for marker nF5I14 (on BAC F5I14), 5 recombinants for marker nga111 (on BAC F28P22), 2 recombinants for marker ADH (on BAC T14N5), and no recombinants for dSNP142 (on BAC F18B13), all on the centromere side. On the telomere side 2 recombinants were found for dSNP142 marker (Figure 8). The ADH-SNP142 interval spanned ~1.2 Mb, which corresponded to ~3-5 cM (4 recombinants/104 recombinants total x 100 = 3.8 cM). In *Arabidopsis* a genetic distance of 1 cM corresponds, on average, to a physical distance of about 250 kb (Lukowitz et al., 2000). According to this, the genetic distance of 3.8 cM would correspond to ~0.95 Mb (3.8 x 250 kb = 950 kb or 0.95 Mb). The

estimated genetic distance of 3.8 cM is an approximate distance and is reduced by double crossovers, which are not included in the estimate.

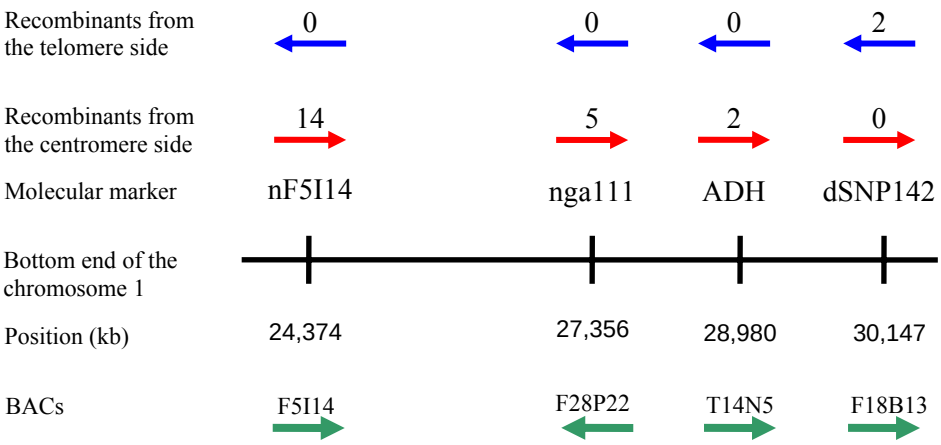

Figure 8. Schematic illustration of markers on chromosome 1 used in initial mapping of *cap2*. The exact position of SNP corresponding to each marker and BACs (green arrows, showing direction of the BAC sequences) where the markers are located is shown. Blue and red arrows show number of recombinants for each marker from the telomere and centromere sides, respectively.

3.1.2 Confirming the initial interval

In order to map and identify the *CAP2* locus, the mapping interval had to be reduced. For this purpose a new mapping population was grown as described in Materials and methods (Figure 6). The T_2 mapping population consisted of 470 plants. DNA was isolated and the whole population was genotyped for PCR markers nga111 (on BAC F28P22), ADH (on BAC T14N5), G17311 (on BAC F5I6 and BAC T21F11) and GL2 (on BAC F19K16) in order to confirm the interval defined by Grini *et al.* (Grini et al., 2002) (Figure 9). nga111 is a SSLP marker, while ADH, G17311 and GL2 are CAPS markers. Primers for marker amplification and appropriate restriction enzymes for cleaving of CAPS markers were available on The *Arabidopsis* Information Resource (TAIR) website (www.arabidopsis.org). The initial mapping interval was flanked on the telomere side by the dSNP142 marker. This marker is positioned between GL2 and G17311 markers. GL2 was also used as a visible marker in the initial screening for gametophytic mutants (Figure 7), where the *cap2* was mapped within the *ap-gl2* interval (Grini et al., 2002).

A total of 31 plants were identified where a recombination event had occurred between nga111 and G17311 markers. After phenotyping the recombinant plants, 23 recombinants were found for marker nga111 and 11 recombinants for marker ADH from the centromere side. From the telomere side, 8 recombinants were found for marker G17311 and 7 recombinants for marker GL2 (Figure 9). These results confirmed the ADH-dSNP142 interval, and indicated that the *cap2* was positioned between ADH and GL2 markers. The ADH-GL2 interval spanned ~1.06 Mb and included one YAC and 12 BAC clones. The interval contained 308 loci according to TAIR.

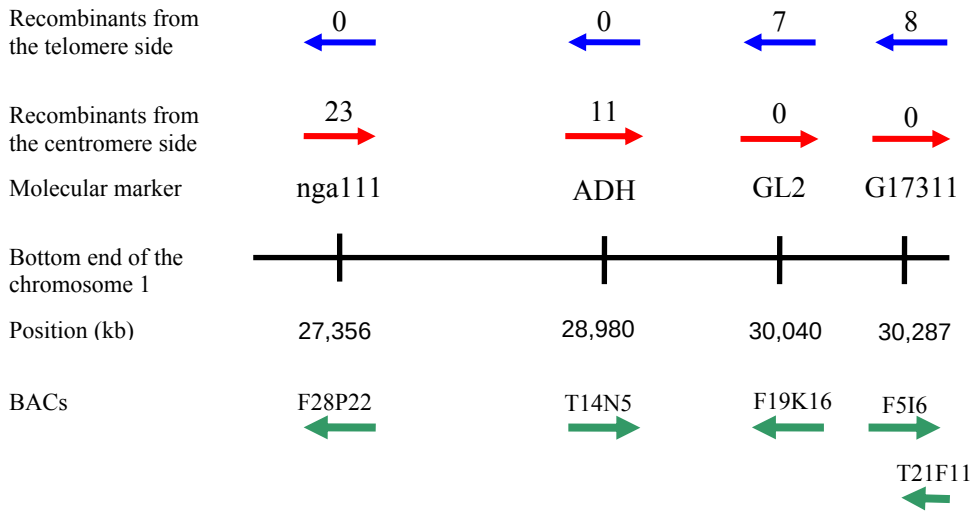

Figure 9. Schematic illustration of markers on chromosome 1 used to confirm the initial interval in the *cap2* mapping. The exact position of SNP corresponding to each marker and BACs (green arrows, showing direction of the BAC sequences) where the markers are located is shown. Blue and red arrows show number of recombinants for each marker from the telomere and centromere sides, respectively.

3.1.3 Mapping with the markers made from SNPs available in the TAIR database

The recombinants were further genotyped for four additional markers within the ADH-GL2 interval that were made from polymorphisms available on the TAIR website. The sequences flanking the polymorphism were provided at the TAIR database. Using the dCAPs Finder program on the TAIR website we could find enzymes that digested in an ecotype specific manner for a particular locus. The sequences flanking the polymorphic site were imported to the Vector NTI (VNTI) software, and CAPS markers were made by designing the primers that would amplify the region of ~1-2 kb containing the polymorphism. Generally, enzymes were chosen that cut only once in the amplified region of Columbia ecotype resulting in two

fragments. The Landsberg ecotype contained no cleavage site in the amplified region. The markers available from the TAIR database were SNP253 (on BAC F3F9), SNP10847 (on BAC F3F9 and BAC T30F21), SNP10870 (on YAC YUP8H12R) and SNP10876 (on YAC YUP8H12R), positioned from the centromere to the telomere side, respectively.

Approaching the SNP253 marker from the centromere side resulted in loss of 6 recombinants in the interval between ADH and SNP253 markers. There were 5 recombinants remaining for the SNP253 marker. 3 additional recombinants were lost when we moved toward the SNP10847 marker from the centromere side, resulting in 2 recombinants remaining for the SNP10847 marker. No recombinants were detected for the SNP10870 marker from the centromere side. From the telomere side, no recombinants were lost in the interval between GL2 and SNP10876 markers, resulting in 7 recombinants remaining for the marker SNP10876. In the interval between SNP10876 and SNP10870 markers 2 recombinants were lost from the telomere side, resulting in 5 remaining recombinants for the SNP10870 marker. No recombinants were detected for the SNP10847 marker from the telomere side (Figure 10). The resulting SNP10847-SNP10870 interval spanned ~204 kb and included three BACs (F3F9, T30F21 and F9K20) and the YAC, YUP8H12R. The interval contained 59 loci according to TAIR.

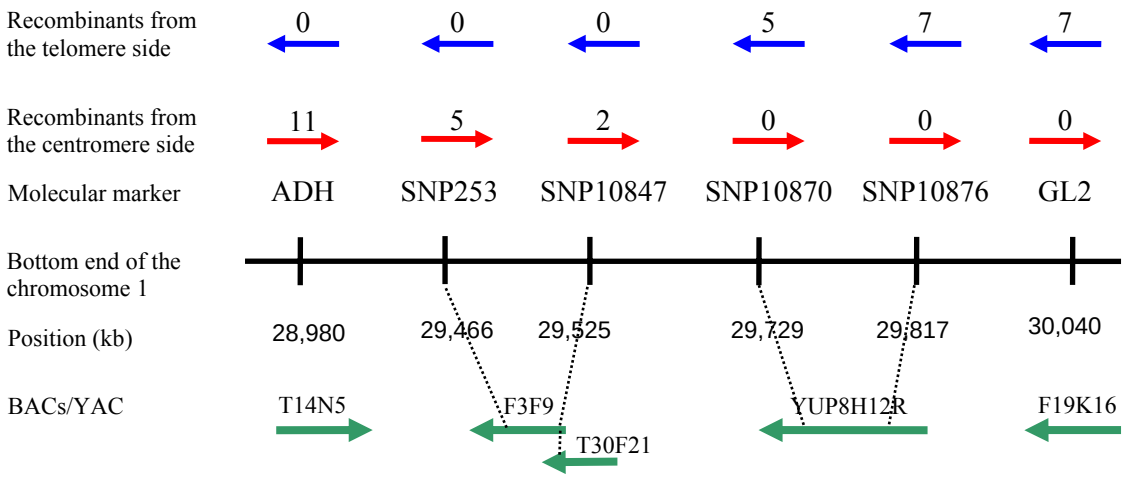

Figure 10. Schematic illustration of markers on chromosome 1 positioned in the *cap2* mapping interval available from the TAIR database. The exact position of SNP corresponding to each marker and BACs/YAC (green arrows, showing direction of the BAC/YAC sequences) where the markers are located is shown. The dotted lines show the locations of the four new markers on BAC F3F9 (and BAC T30F21) and YUP8H12R. Blue and red arrows show number of recombinants for each marker from the telomere and centromere sides, respectively.

3.1.4 Making new PCR markers

As the TAIR database did not have any other available CAPS and SSLP markers within the SNP10847-SNP10870 interval, we searched for the new markers in the *Monsanto Arabidopsis Polymorphism and Ler Sequence Collection* (http://www.arabidopsis.org/cereon/). The SNPs in this database were presented in form of the single nucleotide difference between Col and L*er* at the particular locus and their exact position on the BACs or YAC. Sequences flanking the SNP were also provided as well as the list of restriction enzymes that recognize and cleave within the sequence containing the polymorphism in Col ecotype. Using all these information the new CAPS markers were made covering the remaining mapping interval. Primers amplifying the polymorphism were created using VNTI software. The appropriate enzyme was chosen considering the number of recognition sites within the amplified sequence. The enzyme that cleaved only within the polymorphic site was preferred, such that cleaving would generate two fragments in Col ecotype, and none in L*er*. If none of the enzymes from the list cleaved only within the polymorphic site, but also at other sites within the amplified sequence, enzymes and primers were chosen so that cleavage of Col and L*er* would result in different fragments that could be distinguished on an agarose gel.

3.1.5 Walking on BACs

In order to narrow down the SNP10847-SNP10870 interval and locate the *cap2* mutation to one BAC or YAC we searched on the three BACs and the YAC contained in the mapping interval for the suitable SNPs that were distributed 40-70 kb south and north from SNP10847 and SNP10870 respectively (see Figure 11). This interval corresponded to BAC T30F21 and BAC F9K20 (Figure 11). The next step was to find SNPs that were positioned near the ends of both T30F21 and F9K2, in order to make a series of markers spanning the BAC conting.

Four CAPS markers distributed at intervals of ~23-66 kb were made within the SNP10847-SNP10870 interval. Markers CER446094 and CER446080 were located at BAC T30F21 in an interval of ~23 kb. The other two markers, CER433278 and CER433242 were located on BAC F9K20 in an interval of ~41 kb (Figure 11).

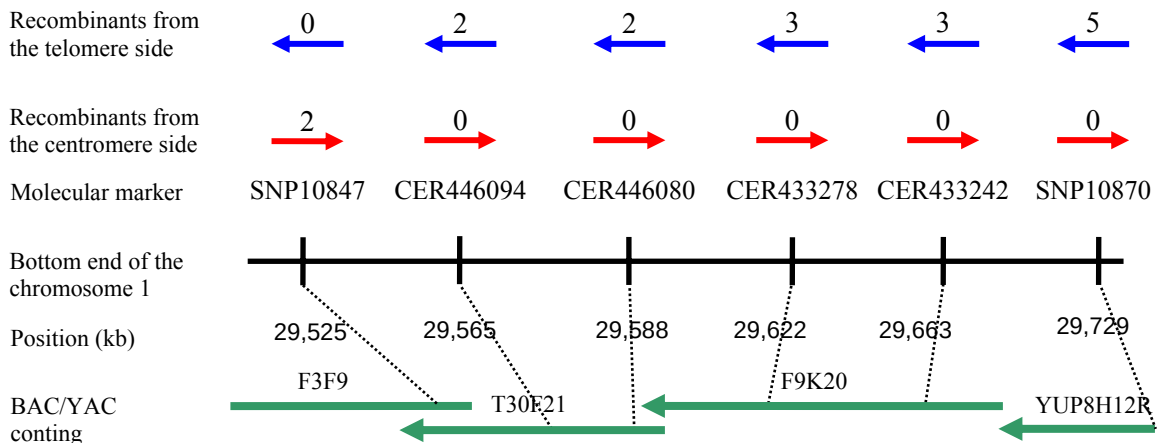

Figure 11. Schematic illustration of the four new markers made from the SNPs presented at the Monsanto database and two flanking markers from the TAIR database. The exact position of SNP corresponding to each marker is shown. BAC/YAC conting (green arrows, showing direction of the BAC/YAC sequences) in this mapping interval and the position of the markers on corresponding BACs/YAC (dotted lines) is illustrated. Blue and red arrows show number of recombinants for each marker from the telomere and centromere sides, respectively.

After genotyping, 3 recombinants were found for marker CER433242, meaning that 2 recombinants were lost in the interval between SNP10870 and CER433242 coming from the telomere side. No recombinants were lost moving further from the CER433242 toward the CER433278 marker, meaning that 3 recombinants remained for the CER433278 marker. 1 recombinant was lost in the interval between markers CER433278 and CER446080, resulting in 2 recombinants for the marker CER446080. Finally, no recombinants were lost moving towards the CER446094 from the CER446080 marker, meaning that 2 recombinants remained for the marker CER446094. All recombinants were detected from the telomere side, while none were detected from the centromere side. No recombinants were found for the marker SNP10847 from the telomere side, and none were found for the marker CER446094 from the centromere side. The new resulting interval was between the SGCSNP10847 and CER446094 markers. This interval spanned ~40 kb and included 12 loci according to TAIR.

3.1.6 Fine-scale mapping; walking on genes

Since there were two recombinants left from the SNP10847 marker and two recombinants left from the CER446094 marker, the mapping interval could be narrowed down by walking along the BAC T30F21. For this purpose, five additional CAPS markers were made: CER446110, CER446109, CER446106, CER446104 and CER446100. The markers spanned the BAC T30F21 at intervals of ~2-9 kb (Figure 12).

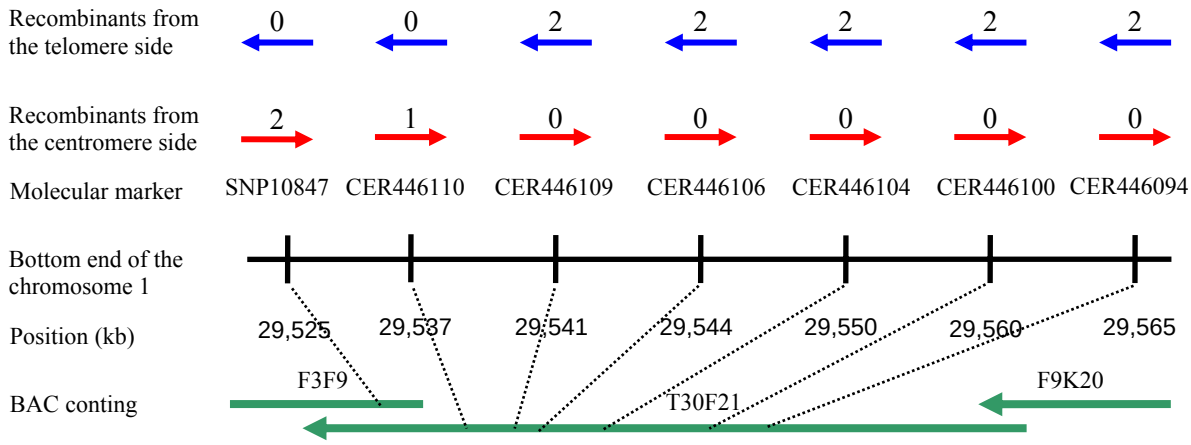

Figure 12. Schematic illustration of markers made from the SNPs presented at the Monsanto database and the flanking SNP10847 from the TAIR database. The exact position of SNP corresponding to each marker is shown. BAC conting (green arrows, showing direction of the BAC sequences) in this mapping interval and the position of the markers on corresponding BAC (dotted lines) is illustrated. Blue and red arrows show number of recombinants for each marker from the telomere and centromere sides, respectively.

All markers in the interval between CER446094 and CER446109 (including CER446109) detected 2 recombinants from the telomere side. This means that no recombinants were lost in the interval between CER446094 and CER446109 markers. No recombinants were detected for the marker CER446110 from the telomere side. From the centromere side, 1 recombinant was lost in the interval between SNP10874 and CER446110, resulting in 1 recombinant for the marker CER446110. No recombinants were detected for the marker CER446109 from the telomere side (Figure 12). The new results indicated that *cap2* mutation was located on BAC T30F21, flanked by markers CER446110 and CER446109 (Figure 12). The interval between two corresponding SNPs spanned 4238 bp and included two genes. These two genes were At1g78500, encoding a putative pentacyclic triterpene synthase (*ATPEN6*) and At1g78510, encoding a solanesyl diphosphate synthase (*SPS*)

(Figure 13). The At1g78500 gene is positioned between 55571-59102 bp on BAC T30F21 and contains the CER446110 polymorphism in its third intron. The CER446110 polymorphism is located at position 58353 bp on BAC T30F21 (Figure 13 and Figure 14). The At1g78510 gene is positioned between 53331-55463 bp on BAC T30F21 and contains the CER446109 polymorphism in its second intron. The CER446109 polymorphism is located at position 54115 bp on BAC T30F21 (Figure 13 and Figure 14).

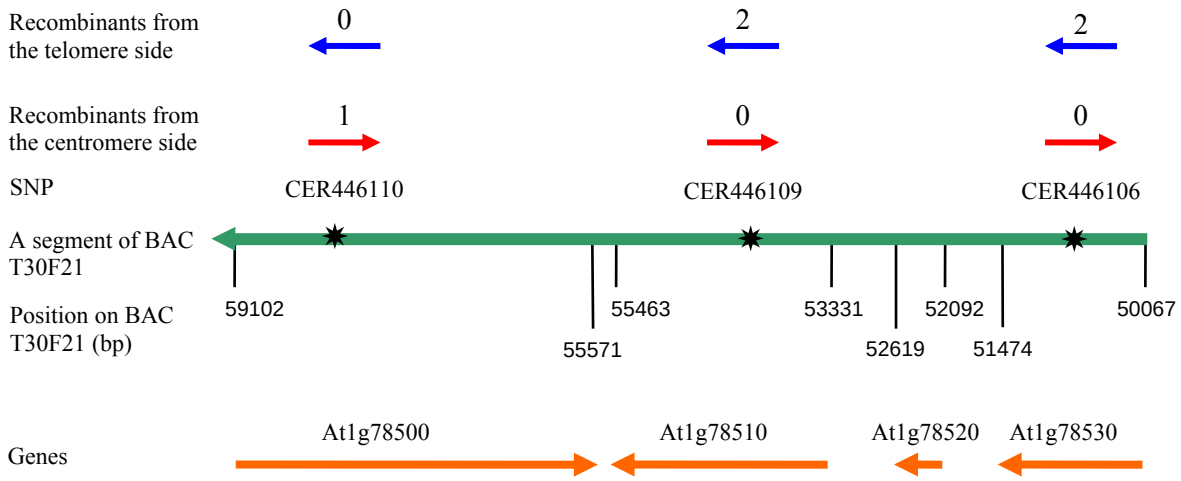

Figure 13. Schematic illustration of genes (orange arrows, showing direction of transcription) spanning the segment of BAC T30F21 carrying the three SNPs in the region between 59120 and 50067 bp. The stars indicate position of the SNPs on the BAC T30F21 (green arrow, showing direction of BAC sequence) and the position in the corresponding gene. The numbers (bp) show the start and end of the genes. Blue and red arrows show number of recombinants for each marker from the telomere and centromere sides, respectively.

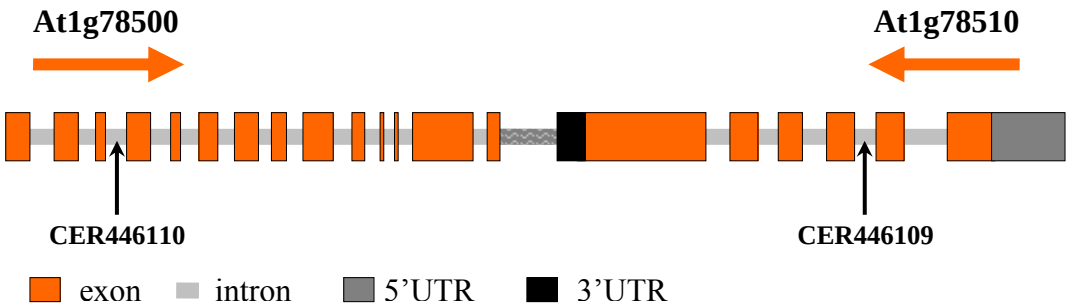

Figure 14. Schematic illustration of annotated At1g78500 and At1g78510 genes. Orange arrows show direction of transcription. Exons are presented as orange boxes, introns are in light grey, 5'UTR is marked in dark grey and 3'UTR is black. Vertical arrows show the position of CER446110 and CER446109 markers in At1g78500 and At1g78510 genes, respectively.

3.1.7 Characterization of the At1g78500 and At1g78510 genes

At1g78500 is a 3532 bp long gene that consists of 14 exons and 13 introns (Figure 14). The putative mRNA (accession nr. NM 106497) has 2304 bp long open reading frame (ORF) and encodes pentacyclic triterpene synthase, a putative protein of 767 aa (accession nr. NP 177971). The annotation is not verified by any full length cDNA. Pentacyclic triterpene synthase has a catalytic activity in pentacyclic triterpenoid biosynthesis (sterol biosynthesis). The cellular component in which the protein is located and its functions is unknown (www.arabidopsis.org).

The pentacyclic triterpene synthase consists of two conserved domains; two prenyltransferase and squalene oxidase repeat (www.ncbi.nlm.nih.gov). *Arabidopsis* triterpene synthases (ATLUP 1-5) catalyze and converse 2,3-oxidosqalene into various pentacyclic triterpenes (e.g. cycloartenol (plant sterol precursor), stigmasterol (plant sterol)) (Husselstein-Muller et al., 2001). The At1g78500 belongs to *ATPEN 1-7* putative triterpene synthases gene group whose deduced polypeptide sequences have about 55% identity with the *ATLUP* family and could encode triterpene synthases (Husselstein-Muller et al., 2001).

At1g78510 is 2133 bp long gene consisting of 6 exons and 5 introns (Figure 14). The gene is transcribed into a 1600 bp long mRNA (accession nr. NM 106498), supported by cDNA. 1218 bp of the mRNA is a translated region and encodes for solanesyl diphosphate synthase (SPS) protein (accession nr. NP 177972). The SPS protein consists of 406 aa and has dimethylallyltranstransferase activity in monoterpenoid biosynthesis. The action of SPS enzyme is localized in plastids where it catalyzes the reaction in which dimethylallyl diphosphate (DMAPP) react with isopentenyl diphosphate (IPP) to generate geranyl diphosphate (GPP) and diphosphate (www.arabidopsis.org; www.ncbi.nlm.nih.gov; Gene Ontology Consortium GO:0004161). Geranyl diphosphate is a key precursor in monoterpene biosynthesis that in plants occurs in plastids (Burke et al., 1999; Chappell, 2002; Kasahara et al., 2002).

The solanesyl diphosphate synthase has one conserved domain, Trans IPPS HT (Trans-Isoprenyl Diphosphate Synthases (Trans_IPPS), head-to-tail (HT) (1'-4) condensation reactions). This conserved domain includes all trans-isoprenyl diphosphate synthases that are

involved in the synthesis of various chain length (C10, C15, C20, C25, C30, C35, C40, C45, and C50) linear isoprenyl diphosphates from precursors, IPP and DMAPP. They catalyze the successive 1'-4 condensation of the 5-carbon IPP to allylic substrates geranyl-, farnesyl-, or geranylgeranyl-diphosphate (www.ncbi.nlm.nih.gov).

The possible involvement of At1g78500 and At1g78510 in seed development will be discussed in the Discussion chapter.

3.1.8 Expression analysis of At1g78500 and At1g78510 genes by RT-PCR

RT-PCR was used to investigate whether At1g78500 and At1g78510 genes were expressed, and if they were, in which organs. The mRNA isolated from flowers, seedlings, rosette leaves and siliques of Columbia ecotype was reverse transcribed into cDNA. Generated first-strand cDNA was used as template in the RT-PCR. Primers designed to amplify a fragment of At1g78500 cDNA annealed to sequences corresponding to exon 2 and exon 10, respectively. The resulting PCR product was 1297 bp long. Primers that amplified a fragment of At1g78510 cDNA annealed to sequences corresponding to exon 1 and exon 6, respectively, generating 831 bp long PCR product. Any genomic contamination in the cDNA sample would be detected, as the DNA amplification with these primers would generate longer fragments due to the presence of introns. Amplification of genomic At1g78500 and At1g78510 sequences would result in 2071 bp and 1368 bp long PCR products, respectively.

To get an indication of which organs having the highest level of mRNA, a semi-quantitative RT-PCR was done. The same amount of template was used for each sample and each reaction was run with two different number of cycles. The first reaction was carried out with 30 PCR cycles for both genes. No product was observed for the primer pair annealing in exon 2 and exon 10 of At1g78500 cDNA (data not shown). The reaction was performed one more time using lower annealing temperature, also resulting in no PCR product (data not shown). To confirm that the At1g78500 gene was active in at least some of the organs, the number of PCR cycles was increased to 40, resulting in a product of the expected size (Figure 15). This indicates that the weakly expressed transcript of At1g78500 gene requires more than 30 cycles of PCR in order to be detected. The results presented in Figure 15 indicate that At1g78500 mRNA is present in flowers, seedlings and siliques, but not in

rosette leaves. It also appears that mRNA level is highest in siliques and lowest in seedlings. The results presented in Figure 15 show clear bands of the expected size for primers annealing in exon 1 and exon 6 of At1g78510 cDNA when the reaction was carried out with 30 cycles. The results indicate that At1g78510 gene is expressed in flowers, rosette leaves, seedlings and siliques. Since 30 cycles gave bands with the same intensity, the number of cycles was reduced to 25 in order to possibly see different expression levels between different organs. The results in Figure 15c indicate that there possibly may be higher expression of At1g78510 gene in rosette leaves and seedlings than in flowers and siliques. However, to be able to determine the expression level in different organs more precisely, a more quantitative method, such as Northern blotting or real time PCR, has to be used.

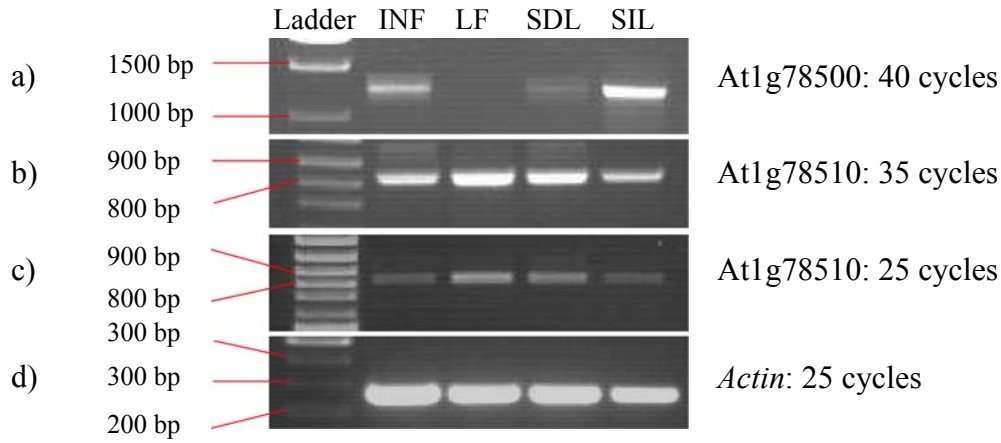

Figure 15. Expression analysis of the At1g78500 (a) and At1g78510 (b,c) genes in flowers (INF-inflorescence), rosette leaves (LF), seedlings (SDL) and siliques (SIL) by RT-PCR. Right-hand side designations refer to the gene specific primers used to amplify mRNA and number of RT-PCR cycles in each reaction. Actin primers (d) were used as template control.

3.1.9 Complementation

To investigate whether At1g78500 or At1g78510 gene could rescue *cap2* mutants by restoring the wild-type development pattern in embryo and endosperm, two complementation constructs were made using the Gateway Cloning System and transformed into *cap2* plants.

3.1.9.1 Complementation constructs

At1g78510 complementation construct was made by amplifying a 3600 bp fragment containing At1g78510 CDS. Att-primers, Gene15 F-attB1 and Gene15 R-attB2 were designed to amplify 1173 bp of promoter region and 684 bp downstream of the coding region (Figure 16). The promoter region included 462 bp of the CDS of the neighboring upstream At1g78520 to ensure that the whole promoter region is covered. At1g78500 complementation construct was made by amplifying 5537 bp fragment containing At1g78500 CDS. Att-primers, Gene16 F-attB1 and Gene16 R-attB2 were designed to amplify 1331 bp of promoter region and 674 bp downstream of the coding region (including also 418 bp of the At1g78510 CDS) (Figure 16). The PCR fragments, amplified from wild-type (wt) DNA, were recombined into the pGSC1704 destination vector using the Gateway Cloning System (Invitogen Life Technologies). The constructs, verified by sequencing of borders and restriction digests, were introduced into *Agrobacterium tumefaciens* strain C58C1 pGV2260. *cap2*/+ *Arabidopsis* plants were transformed by floral dipping in order to rescue the wt phenotype.

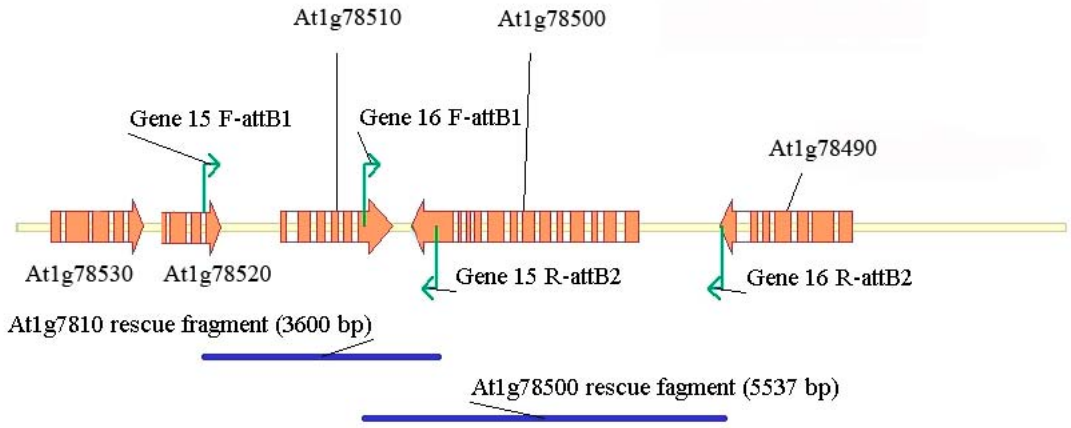

Figure 16. Illustration from Vector NTI showing annealing position of primers (green arrows) designed to amplify the At1g78510 and At1g78500 rescue fragments (blue lines). The orange box/arrows represent genes. Annotation is taken from BAC T30F21 which is inverted compared to sequence from AGI.

3.1.9.2 Expectation from the rescue experiment

The gametophytic maternal-effect mutation in *cap2* plants is caused by a mutation in a gene that is required to be expressed in the female gametophyte for normal embryo and endosperm development. Only T_1 *cap2* seeds in which the rescue T-DNA has been inserted prior to fertilization would be able to develop and grow. It was shown previously that the female gametophyte is the primary, if not only, target of T-DNA integration when the floral dipping is used as the T-DNA transformation method (Bechtold et al., 2000; Desfeux et al., 2000). Since the T-DNA is not integrated in the pollen, but only in the ovule, T_1 seed progeny selected on hygromycin medium contains four different classes presenting four different genotypes. The two main parameters are: 1) Whether the transformed ovule is *cap2* or *CAP2*; 2) Whether the transformed ovule is fertilized by *cap2* or *CAP2* pollen. If an ovule carrying *cap2* gametophyte is transformed, then the *cap2*/R[1] ovule fertilized with *cap2*/- pollen would give rise to a progeny class with the *cap2*/*cap2*;R/- genotype[1]. The *cap2*/R ovules fertilized with wt (*CAP2*/-) pollen would result in seeds with the *cap2*/*CAP2*;R/- genotype. Transformed *CAP2*/R ovules fertilized with *cap2*/- or *CAP2*/- pollen produce *CAP2*/*cap2*;R/- and *CAP2*/*CAP2*;R/- genotype classes, respectively (Figure 17). *cap2*/*CAP2*;R/- and *CAP2*/*cap2*;R/- genotypes are considered two different genotypes as *cap2* have different parental origin (maternal/paternal).

♂ \ ♀	***cap2* R**	***CAP2* R**
***cap2* -**	*cap2*/*cap2*;R/-	*CAP2*/*cap2*;R/-
***CAP2* -**	*cap2*/*CAP2*;R/-	*CAP2*/*CAP2*;R/-

Figure 17. Four genotypes presenting four different T_1 seed classes that can be generated by pollination of *cap2* R and *CAP2* R ovules (♀) with *cap2* - and *CAP2* - pollen (♂).

If the inserted T-DNA construct complements the *cap2* mutation, four possible scenarios (depending on the genotypes of the T_1 plants) can be obtained from the self-cross of the T_1

[1] R denote T-DNA insertion, - denote no T-DNA insertion.

plants (Figure 18). Moreover, different gamete classes producing these scenarios have different frequencies that depend on the distance between *CAP2* locus and the position of the T-DNA insertion. The two parental classes have the gamete frequency of (1-p), while the frequency of the two recombinant classes is p, where p is the genetic distance or recombination frequency between the *CAP2* locus and the T-DNA locus. If the T-DNA and *CAP2* loci are not linked, then p = 0.5 and (1-p) = p = 0.5, which means that the frequency of the parental and the recombinant gametes are the same. In this case we would expect to get 50% *cap2* and 50% *CAP2* seed phenotype in siliques of the *cap2*/*cap2*;R/- T_2 plants (Figure 18). The *cap2*/*CAP2*;R/- and *CAP2*/*cap2*;R/- plants would both produce 25% *cap2* and 75% *CAP2* seeds, while there would be 100% wt seeds in the T_2 progeny of the *CAP2*/*CAP2*;R/- plants (Figure 18b-d). If the T-DNA is inserted at genetic distance less than 50 cM ($p < 0.5$) from the *CAP2* locus, the recombination frequency goes towards zero. As a consequence the frequency of the parental gametes increases and the frequency of the recombinant gametes decreases. Thus, the frequency of the parental gametes increases in inverse ratio with the distance between the *CAP2* and T-DNA loci. At complete linkage, when p = 0, the expected distribution between *cap2* and *CAP2* seeds from the self-pollinated T_1 plants would be as follows: 50% *cap2* and 50% *CAP2* seeds from the T_1 plants with the *cap2*/*cap2*;R/- genotype, 0% *cap2* and 100% *CAP2* seeds from the T_1 plants with the *cap2*/*CAP2*;R/- genotype and 50% *cap2* and 50% *CAP2* seeds from the T_1 plants with the *CAP2*/*cap2*;R/- genotype. At intermediate linkage, when p = 0.3 the expected distribution between *cap2* and *CAP2* seeds from the self-pollinated T_1 plants would be: 50% *cap2* and 50% *CAP2* seeds from the *cap2*/*cap2*;R/- T_1 plants, 15% *cap2* and 85% *CAP2* seeds from the *cap2*/*CAP2*;R/- T_1 plants, and 35% *cap2* and 65% *CAP2* seeds from the *CAP2*/*cap2*;R/- plants. The *CAP2*/*CAP2*;R/- plants always produce only *CAP2* seeds. Thus, when the *CAP2* and T-DNA loci are linked ($p < 0.5$), six different scenarios (ratios between *cap2* and *CAP2* seeds) could be obtained, depending on the T_1 genotype and the p-value.

As seen from Figure 18a-d the frequency of *cap2* could in general range from 0-50%. Thus, in the case of the *cap2*/*cap2*;R/- self-cross when p = 0.5, p = 0 and p = 0.3 and *CAP2*/*cap2*;R/- self-cross when p = 0.3, the expected ratio between *CAP2* and *cap2* seed phenotypes would be 50%:50%, which is the same as expected ratio in *cap2* plants when complementation is not obtained.

a) T_1: *cap2*/*cap2*;R/- self-cross

♂ \ ♀	½(1-p) *cap2* R	½(1-p) *cap2* -	½p *cap2* -	½p *cap2* R
cap2 R	*cap2*/*cap2*;R/R	*cap2*/*cap2*;-/R	*cap2*/*cap2*;-/R	*cap2*/*cap2*;R/R
cap2 -	*cap2*/*cap2*;R/-	*cap2*/*cap2*;-/-	*cap2*/*cap2*;-/-	*cap2*/*cap2*;R/-
cap2 -	*cap2*/*cap2*;R/-	*cap2*/*cap2*;-/-	*cap2*/*cap2*;-/-	*cap2*/*cap2*;R/-
cap2 R	*cap2*/*cap2*;R/R	*cap2*/*cap2*;-/R	*cap2*/*cap2*;-/R	*cap2*/*cap2*;R/R

p	*CAP2*:*cap2* T_2 seeds
0.5	50:50
0	50:50
0.3	50:50

b) T_1: *cap2*/*CAP2*;R/- self-cross

♂ \ ♀	½(1-p) *cap2* R	½(1-p) *CAP2* -	½p *cap2* -	½p *CAP2* R
cap2 R	*cap2*/*cap2*;R/R	*CAP2*/*cap2*;-/R	*cap2*/*cap2*;-/R	*CAP2*/*cap2*;R/R
CAP2 -	*cap2*/*CAP2*;R/-	*CAP2*/*CAP2*;-/-	*cap2*/*CAP2*;-/-	*CAP2*/*CAP2*;R/-
cap2 -	*cap2*/*cap2*;R/-	*CAP2*/*cap2*;-/-	*cap2*/*cap2*;-/-	*CAP2*/*cap2*;R/-
CAP2 R	*cap2*/*CAP2*;R/R	*CAP2*/*CAP2*;-/R	*cap2*/*CAP2*;-/R	*CAP2*/*CAP2*;R/R

p	*CAP2*:*cap2* T_2 seeds
0.5	75:25
0	100:0
0.3	85:15

c) T_1: *CAP2*/*cap2*;R/- self-cross

♂ \ ♀	½(1-p) *CAP2* R	½(1-p) *cap2* -	½p *CAP2* -	½p *cap2* R
CAP2 R	*CAP2*/*CAP2*;R/R	*cap2*/*CAP2*;-/R	*CAP2*/*CAP2*;-/R	*cap2*/*CAP2*;R/R
cap2 -	*CAP2*/*cap2*;R/-	*cap2*/*cap2*;-/-	*CAP2*/*cap2*;-/-	*cap2*/*cap2*;R/-
CAP2 -	*CAP2*/*CAP2*;R/-	*cap2*/*CAP2*;-/-	*CAP2*/*CAP2*;-/-	*cap2*/*CAP2*;R/-
cap2 R	*CAP2*/*cap2*;R/R	*cap2*/*cap2*;-/R	*CAP2*/*cap2*;-/R	*cap2*/*cap2*;R/R

p	*CAP2*:*cap2* T_2 seeds
0.5	75:25
0	50:50
0.3	65:35

d) T_1: *CAP2*/*CAP2*;R/- self-cross

♂ \ ♀	½(1-p) *CAP2* R	½(1-p) *CAP2* -	½p *CAP2* R	½p *CAP2* -
CAP2 R	*CAP2*/*CAP2*;R/R	*CAP2*/*CAP2*;-/R	*CAP2*/*CAP2*;R/R	*CAP2*/*CAP2*;-/R
CAP2 -	*CAP2*/*CAP2*;R/-	*CAP2*/*CAP2*;-/-	*CAP2*/*CAP2*;R/-	*CAP2*/*CAP2*;-/-
CAP2 R	*CAP2*/*CAP2*;R/R	*CAP2*/*CAP2*;-/R	*CAP2*/*CAP2*;R/R	*CAP2*/*CAP2*;-/R
CAP2 -	*CAP2*/*CAP2*;R/-	*CAP2*/*CAP2*;-/-	*CAP2*/*CA2*;R/-	*CAP2*/*CAP2*;-/-

p	*CAP2*:*cap2* T_2 seeds
0.5	100:0
0	100:0
0.3	100:0

Figure 18. Four possible scenarios (a, b, c and d) obtained from the self-cross of T_1 plants in which the inserted T-DNA construct complements the *cap2* mutation. T_2 gametes, their frequencies and the genotypes of the T_2 seed progenies are presented in four Punnett squares. ½(1-p) and ½p represents parental and recombinant gametes frequencies, respectively, where p is the genetic distance or recombination frequency between the *CAP2* and the T-DNA loci. Shaded columns represent progeny-classes that are lethal due to maternal *cap2* allele and no T-DNA complementation insertion. Table on the right side of each Punnett square shows expected ratio of *CAP2* to *cap2* seed phenotype in progeny obtained from that particular cross when the *CAP2* and the T-DNA loci are unlinked (p=0.5), at complete linkage (p=0) and at intermediate linkage (p=0.3).

For both linked and unlinked situations it was assumed that one T-DNA was inserted per genome in the T_0 plants. If the T_0 plants carried more than one copy of the T-DNA insertion, the ratios in the T_1 and T_2 would be more. In general multiple insertions would produce higher rescue frequencies. Assuming that a plant carries three T-DNA insertions that segregate as independent loci, eight different gamete classes would be produced. Seven of these eight classes would carry at least one T-DNA copy. Hence, self-cross would result in progeny seeds whereof $1/8^{th}$ (12.5%) would not carry any T-DNA insertion. Among these 12.5%, half of the seeds would carry *cap2* allele and half would be wt. Hence, the expected frequency of *cap2* seed phenotype in the case of rescue would range from zero to 6.25% in the progeny produced by self-crossing of plants carrying three T-DNA insertions. This in contrast to the frequency of *cap2* that would range from 0-50% in a progeny carrying only one T-DNA insertion.

The above calculations are made under the assumption that the rescue construct is fully penetrant.

3.1.9.3 Analysis of the primary transformants

In the T_1 generation, 8 independent transformants harboring the At1g78500 complementation construct and 2 independent transformants harboring the At1g78510 complementation construct were obtained. To confirm that transformants indeed carried the inserted transgene, a control-PCR was done using one gene specific and one hygromycin primer (see Appendix 1 and Appendix 3) (data not shown).

Siliques from all T_1 transformants were inspected for *cap2* phenotype (Table 4). Aborted seeds in siliques from plants with the At1g78500 transgene, varied from 3.4% to 39.5%. For the two plants carrying the At1g78510 transgene, the arrested seeds accounted for 40.5% and 71.7%. Our expectation is that half of the T_1 progeny (transformants) resulting from the *cap2*/*CAP2* plants carries one copy of the mutant *cap2* allele (Figure 18a-c). The other half carries two wt (*CAP2*) alleles (Figure 18d).

In order to determine whether any of the T_1 transformants carrying the *cap2* allele was complemented by the rescue construct, the ratio of wt to aborted seeds in each transformant

was compared with expected ratios for the four possible self-crosses using a chi-square test. It was assumed that only one T-DNA was inserted per genome in the T_0 plants. The chi-square test was done considering three different recombination frequencies between the *CAP2* and T-DNA loci in each self-cross: p = 0.5 (for unlinked loci), p = 0 and p = 0.3 (for linked loci). The following four T_1 self-crosses were tested: *cap2*/*cap2*;R/-, *cap2*/*CAP2*;R/-, *CAP2*/*cap2*;R/- and *CAP2*/*CAP2*;R/-. Probabilities (P-values) for the chi-square tests are presented in Table 5. The P-values for the *CAP2*/*CAP2*;R/- self-cross are not presented, since the chi-square gives no result (0 in the denominator) as the expected ratio is 100:0 in all cases.

For the eight lines harboring the At1g78500 complementation construct seven showed significant similarity with expected ratio for complementation for at least one possible genotype in the chi-square test. Each of these lines could hence represent complementation of the *cap2* phenotype. For line 16.4.2.1, the chi-square test showed significant similarity between expected and observed ratios for *cap2*/*CAP2*;R/- T_1 genotype with p = 0.5 and *CAP2*/*cap2*;R/- with p = 0.5 and p = 0.3. Line 16.5.1.3 showed significance for *cap2*/*CAP2*;R/- T_1 genotype with p = 0.3. Line 16.5.2.1 showed significance for *CAP2*/*cap2*;R/- T_1 genotype with p = 0.3. Line 16.5.3.1 showed significance for *cap2*/*CAP2*;R/- T_1 genotype with p = 0.5 and 0.3 and for *CAP2*/*cap2*;R/- with p = 0.5. Line 16.5.4.1 showed significance for *CAP2*/*cap2*;R/- T_1 genotype with p = 0.3. Line 16.5.4.2 showed significance for *cap2*/*CAP2*;R/- T_1 genotype with p = 0.3 and *CAP2*/*cap2*;R/- with p = 0.5 and p = 0.3. Line 16.5.5.1 showed significance for *cap2*/*CAP2*;R/- T_1 genotype with p = 0.3. The line 16.5.1.1 gave no significant similarity with any expected ratio for complementation.

For the two lines harboring the At1g78510 complementation construct, the line 15.6.5.1 showed significant similarity with expected ratio for complementation for *cap2*/*cap2*;R/- T_1 genotype for all three recombination frequencies tested, and for *CAP2*/*cap2*;R/- T_1 genotype with p = 0. The expected ratio between wt and aborted seeds in *cap2* mutants is 50:50, which is the same as expected ratio when complementation is obtained for *cap2*/*cap2*;R/- T_1 genotype (p = 0, p = 0.3 and p = 0.5) and for *CAP2*/*cap2*;R/- with p = 0. Thus, it is not

possible to say if this plant is rescued or not from this test. Line 15.6.5.2 gave no significant similarity with any expected ratio for complementation.

Table 4. Percentage of developed torpedo stage (green) and arrested (white/brown) seeds in transgene T_1 plants carrying the At1g78500 and At1g78510 transgenes. N means total number of seeds counted per line. The wild-type L*er* was used as control.

Line	Developed torpedo stage (green seeds) (%)	Arrested (white/brown seeds) (%)	N
At1g78500			
16.4.2.1	69.5	30.5	105
16.5.1.1	96.5	3.4	114
16.5.1.3	91.9	8.1	173
16.5.2.1	66.1	33.9	121
16.5.3.1	81.0	19.0	105
16.5.4.1	60.5	39.5	76
16.5.4.2	70.0	30.0	90
16.5.5.1	86.5	13.5	104
At1g78510			
15.6.5.1	59.5	40.5	232
15.6.5.2	28.3	71.7	92
Wt	92.8	7.2	110

Table 5. P-values for chi-square test of complementation analysis for two complementation constructs (At1g78500 and At1g78510) with three different recombination frequencies (p = 0.5, 0 and 0.3) between *CAP2* and T-DNA loci. The null hypothesis (H_0) tested in the chi-square tests was: the difference between the observed and expected distribution if complementation occurred is random. The P-values that are written in boldface allow you to accept the H_0.

T_1 self-cross:	*cap2*/*cap2*;R/-			*cap2*/*CAP2*;R/-			*CAP2*/*cap2*;R/-		
Recomb. freq.:	p=0.5	p=0	p=0.3	p=0.5	p=0	p=0.3	p=0.5	p=0	p=0.3
CAP2:*cap2*	50:50	50:50	50:50	75:25	100:0*	85:15	75:25	50:50	65:35
At1g78500									
16.4.2.1	0.0001	0.0001	0.0001	**0.206**	0	0	**0.206**	0.0001	**0.3429**
16.5.1.1	0	0	0	0	0	0.0013	0	0	0
16.5.1.3	0	0	0	0.0001	0	**0.0531**	0.0001	0	0
16.5.2.1	0.0013	0.0013	0.0013	0.0402	0	0	0.0402	0.0013	**0.8151**
16.5.3.1	0	0	0	**0.1761**	0	**0.2570**	**0.1761**	0	0.0008
16.5.4.1	0.0353	0.0353	0.0353	0.0008	0	0	0.0008	0.0353	**0.3483**
16.5.4.2	0.0001	0.0001	0.0001	**0.2483**	0	0	**0.2483**	0	**0.2945**
16.5.5.1	0	0	0	0.0077	0	**0.6666**	0.0077	0	0
At1g78510									
15.6.5.1	**0.0579**	**0.0579**	**0.0579**	0.0003	0	0	0.0003	**0.0579**	0
15.6.5.2	0	0	0	0	0	0	0	0	0

*In order to avoid 0 in the denominator in the calculation, the test was performed on the ratio 99.9:0.1.

Since the *CAP2* locus is located at the tip of chromosome I, it is unlikely that the T-DNA is inserted close to the *CAP2* locus (closer than 50 cM, corresponding to 12.5 Mb), than anywhere else in the genome (either in another part of chromosome 1 or in any of the other four chromosomes). Chromosome I is 31.5 Mb long and the four other chromosomes are together 82.9 Mb (Chang et al., 2001), meaning that the possibility is about 8 times[2] larger that the T-DNA is inserted elsewhere in the genome, and not linked to the *CAP2* locus. Therefore if the chi-square test shows significance for more than one p- value it is more likely that the ratio where p = 0.5 (not linked loci) is correct.

In summary the results from the analysis of the primary transformants indicate that in the seven lines harboring the At1g78500 construct and in the one line harboring the At1g78510 construct, complementation could have been accomplished. This is under the assumption that the plants carried one T-DNA insertion. However, as it is not known if the genotypes of the transgenic T_1 plants are *cap2/CAP2, cap2/cap2 or CAP2/CAP2*, we can not make any certain conclusion whether the complementation is indeed achieved by using this test. A segregation analysis has to be performed to establish the number of T-DNA insertions in each of the transformants. Such an analysis may reveal if the obtained ratio between *CAP2*:*cap2* seed phenotypes in the T_1 plants is due to complementation.

3.1.10 Segregation analysis

Segregation analysis was performed on T_2 seeds in order to determine the number of complementation constructs (T-DNA) inserted per transformant line. The *hygromycin* gene of the T-DNA gives the hygromycin resistance to the plants. The hygromycin resistance is a dominant trait, meaning that both hemizygous and homozygous plants will be resistant to hygromycin and will be able to germinate on medium supplemented with this antibiotic. The segregation ratio of hygromycin resistant (Hyg^R) to hygromycin sensitive (Hyg^S) seedlings indicates how many T-DNA loci have been integrated in the plant genome.

T_2 seeds from the transformed lines carrying the rescue construct were plated on MS-2 medium supplemented with hygromycin. After three weeks the number of Hyg^R and Hyg^S

[2] The entire genome as reported by Chang et al. (2001) is 114.4 Mb, subtracting 12.5 Mb (50 cM) gives 101.9 Mb, which is about 8 times more than 12.5 Mb.

seedlings was counted and the number of T-DNA insertions tested using the chi-square test. It was tested if the T-DNA was segregated as one Mendelian locus when the expected ratio was 3:1 for Hyg^R:Hyg^S (one integrated T-DNA), 15:1 for Hyg^R:Hyg^S (two integrated T-DNAs) and 63:1 for Hyg^R:Hyg^S (three integrated T-DNAs). The results from the chi-square test are presented in Table 6.

The chi-square test shows that the observations are not in agreement with ratio 3:1; for most transformant lines also not for 15:1, but for all lines in good agreement with 63:1 (highly significant for all lines). Hence the transformants did most likely contain three or more T-DNA insertions or were homozygous for the T-DNA.

As described previously in section 3.1.9.2, one would expect the frequency of *cap2* seed phenotype to range from zero to 6.25% in the progeny produced by self-crossing of plants carrying three T-DNA insertions. Thus, considering the results of the T_1 plants phenotyping (Table 4), it is revealed that all lines but one (16.5.1.1) have the frequency of the *cap2* seed phenotype larger than 6.25%. This indicates that all the lines (except 16.5.1.1) could not be complemented, contrary to what we assumed from the first analysis (section 3.1.9.3).

Table 6. Observed number and percentages of Hyg^R (green) and Hyg^S (white) T_2 seedlings of the rescue lines. P-values for chi-square test for three different expected segregation ratios Hyg^R: Hyg^S (3:1, 15:1 and 63:1) are also shown. The null hypothesis (H_0) tested in the chi square tests was: the difference between the observed and expected distribution is random. The P-values that are written in boldface allow you to accept the H_0, hence the observations are in accordance with the expected ratio.

	Hyg^R #green	Hyg^S #white	Hyg^R %	Hyg^S %	P-value exp. 3:1	P-value exp. 15:1	P-value exp. 63:1
At1g78500							
16.4.2.1	108	2	98.2	1.8	0	**0.158**	**0.977**
16.5.1.1	67	1	98.5	1.5	0	**0.266**	**0.998**
16.5.1.3	80	0	100.0	0.0	0	**0.059**	**0.530**
16.5.2.1	120	0	100.0	0.0	0	0.018	**0.387**
16.5.3.1	95	1	99.0	1.0	0	**0.108**	**0.919**
16.5.4.1	125	0	100.0	0.0	0	0.016	**0.372**
16.5.4.2	128	0	100.0	0.0	0	0.014	**0.362**
16.5.5.1	153	0	100.0	0.0	0	0.006	**0.297**
At1g78510							
15.6.5.1	97	0	100.0	0.0	0	0.039	**0.463**
15.6.5.2	130	2	98.5	1.5	0	**0.080**	**0.999**

In addition to the 16.5.1.1 line that carried 3.4% aborted seeds in its siliques, the line 16.5.1.3 that carried 8.1% *cap2* seeds could also be considered as a possibly complemented line. Both of these lines harbor the At1g7500 complementation construct. In order to confirm if any of these two lines had the *cap2* genotype, and if the low frequency of the *cap2* seeds was due to the complementation, the lines have been outcrossed with wild-type. If the progeny shows *cap2* phenotype (50% *CAP2* and 50% *cap2* seeds in the siliques) it would indicate that the corresponding line had the *cap2* genotype, and hence was complemented. This work is in progress but is not a part of this thesis. It is however expected that half of the T1 generation do not carry the mutant *cap2 allele*, and lines 16.5.1.1 and 16.5.1.3 are thus most probably CAP2 wild types.

3.1.11 Study of T-DNA knockout mutants; analysis of putative *cap2* alleles

Since transformation of *Arabidopsis* with the rescue constructs resulted in few primary transformants and a negative complementation result, three lines of putative *cap2* mutant alleles (SALK_036693, SALK_060682, and SALK_126948) were identified and ordered from the SALK Insertion Sequence Database (http://signal.salk.edu/).

SALK_036693 and SALK_060682 are reported to carry a T-DNA insertion in the fourteenth and ninth exon of At1g78500 gene, respectively. SALK_126948 is reported to carry a T-DNA insertion in the first exon of At1g78510 gene. Twenty-four T_3 plants from SALK_036693, six plants from SALK_060682 and twenty-eight from SALK_126948 line were genotyped by two PCRs (Figure 19) in order to find plants hemizygous or homozygous for the T-DNA insertion that later would be screened for *cap2* phenotype. For primer pairs used in genotyping PCRs, the reaction conditions and the products length, see Appendix 1.

In the first reaction LP and RP gene specific primers would amplify the PCR fragment if the T-DNA insert is not present in the locus (Figure 19a). If the T-DNA is inserted between the LP and RP primers, no PCR product would be generated because the T-DNA insert is ~ 4.4 kb and the fragment would probably be too long to be amplified with the extensive time used. In the second reaction the RP gene specific primer is used together with LBa1 primer that anneals in the left border of the T-DNA (Figure 19b). The PCR product would be

generated only when the T-DNA is inserted, otherwise the LBa1 primer would not have any sequence for annealing.

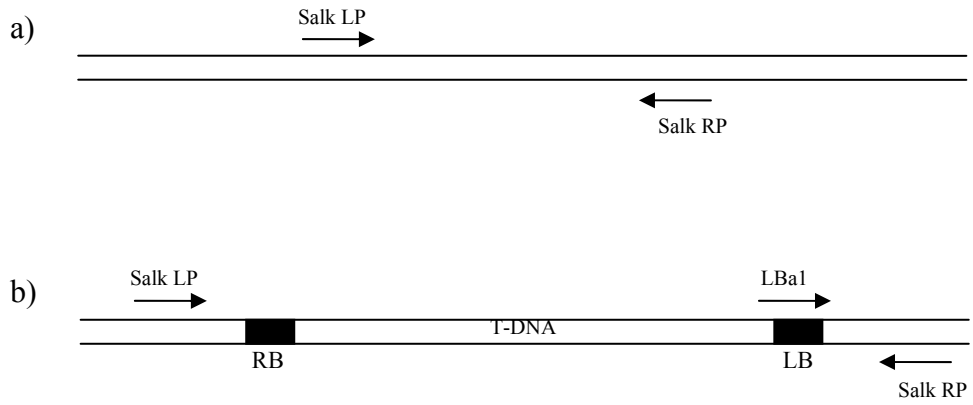

Figure 19. Two PCRs carried out in genotyping of SALK insertion lines. If the T-DNA is not inserted in the locus (a), amplification with the gene specific primers LP and LP will result in a PCR product. If the T-DNA is inserted in the locus (b), only reaction using one gene specific primer (RP) and one T-DNA specific primer (LBa1) will produce a PCR product.

Wild-type plants (plants without the T-DNA insertion) would be detected by the first reaction, because only this reaction would result in a PCR product. Both reactions would result in a PCR product for plants that are hemizygous for the T-DNA. For plants that are homozygous for the insertion, only the second reaction would result in a PCR product.

If the T-DNA was inserted in the *CAP2* locus, homozygous SALK plants were not expected to be obtained, since an ovule carrying the maternal *cap2* allele in the egg- or central cell would not develop into a functional seed after fertilization. Hence, the ratio of wt to hemizygous (carrying the paternal T-DNA insertion) to homozygous plants would be 1:1:0 (Figure 20).

♂ \ ♀	**T-DNA**	***CAP2***
T-DNA	T-DNA/T-DNA	T-DNA/*CAP2*
CAP2	T-DNA/*CAP2*	*CAP2*/*CAP2*

Figure 20. Punnett square illustrating hypothetical SALK plants genotypes generated in a self-cross of hemizygous plants containing the T-DNA insertion in the *CAP2* locus. Shaded column represents genotypes that lead to lethality due to T-DNA insertion in the maternal *CAP2* locus. Thus, the expected ratio of wt:hemizygous:homozygous SALK plants is 1:1:0.

The results from the genotyping analysis are presented in Table 7. Five homozygous plants were detected among twenty-eight SALK_036693 plants (carrying the T-DNA in the At1g78500) and eleven homozygous among twenty-four SALK_126948 plants (carrying the T-DNA in the At1g78510). This result suggests that none of the At1g78500 and At1g78510 genes correspond to the *CAP2* locus. Moreover, the results show that the ratio between wt, hemizygous and homozygous SALK_036693 plants is 0.9:2.4:0.7. If the T-DNA insertion does not induce a gametophytic mutation, the T-DNA would segregate as a Mendelian locus with the ratio 1:2:1 for wt, hemizygous and homozygous plants. The observed ratio seems to be in agreement with the Mendelian distribution[3]. This also supports that the T-DNA insertion is not located in the *CAP2* locus.

In the SALK_126948 line, the ratio between wt, hemizygous and homozygous plants is 1.2:1:1.8. The genotyping results for this SALK line show a distortion of the Mendelian 1:2:1 ratio[4]. The large deviation in the observations from the expected ratio is most probably due to small number of plants. The high frequency of homozygous plants could be due to a gametophytic lethal mutation linked *in trans* to the T-DNA.

[3] The distributions were compared using a chi-square test, using H_0: the difference between the observed and expected Mendelian distribution is random. Observed distribution: 6:17:5, expected distribution: 7:14:7. Chi-square test gave $\chi^2 = 1.357$ with 2 degrees of freedom, giving P = 0.51. Null hypothesis should be accepted; the difference is random and the observed distribution is possibly Mendelian.

[4] Observed distribution: 7:6:11, expected distribution: 6:12:6. Chi-square test gave $\chi^2 = 7.333$ with 2 degrees of freedom, giving p = 0.0256 and H_0 (same as above) is rejected.

In the SALK_060682 line, carrying the T-DNA in the At1g78500 gene, no homozygous plants were found among the six plants tested. Six plants are too few to get statistically significant results from a statistical test of possible ratio. The ratio between hemizygous, wt and homozygous plants in this SALK line was 1:1:0, indicating that the insertion of the T-DNA might have induced gametophytic or embryo lethal mutation in the At1g78500 locus. However, a chi-square test shows that the distribution may also be in agreement with the Mendelian[5] distribution.

Table 7. Total number of plants, number (#) of wild-type (wt), hemizygous (hemz) and homozygous (homz) and the ratios wt:hemz:homz for each SALK line.

SALK line	Locus	# plants	# wt	# hemz	# homz	wt:hemz:homz
SALK_036693	At1g78500	28	6	17	5	0.9:2.4:0.7
SALK_126948	At1g78510	24	7	6	11	1.2:1:1.8
SALK_060682	At1g78500	6	3	3	0	1:1:0

All homozygous and hemizygous plants from the three SALK lines were examined for the *cap2* phenotype, but no *cap2* phenotype was detected in any of these plants. This further suggests that the insertions in the three SALK lines are not located in the *CAP2* locus and that neither At1g78500 nor At1g78510 corresponds to *CAP2* gene.

3.1.12 Summary of part A

In the first part of this thesis the main goal was to map the *cap2* gametophytic maternal-effect mutation in order to get the molecular identity of the *CAP2* gene and determine its role in embryo and endosperm development. For this work a mapping population of L*er* and Col ecotypes consisting of 470 plants was generated. Twenty new molecular markers were made based on known polymorphisms from the TAIR database and single nucleotide polymorphisms found in the Monsanto database. This allowed us to map the *cap2* mutation to a genetic interval of 4238 bp. This interval spanned parts of two genes, one involved in monoterpenoid biosynthesis and the other putatively involved in triterpenoid biosynthesis. These two candidate genes were subcloned and introduced into *cap2* plants to complement

[5] Observed distribution: 3:3:0, expected distribution: 1.5:3:1.5. Chi-square test gave $\chi^2 = 3.0$ with 2 degrees of freedom, giving $P = 0.223$ and H_0 (same as above) is accepted.

the mutation. The analysis of the first generation of transgene plants was promising, but a more thorough analysis of the next generation could not verify that complementation had been achieved. In addition, three SALK lines carrying a T-DNA insertion in exons of these two genes were found not to have a *cap2* phenotype and to segregate plants homozygous for the insertion. A more thorough discussion of these results and future experiments will be presented in the discussion chapter.

3.2 Part B: Molecular analysis of promoter trap lines

The second part of this thesis presents a reverse genetic approach to analyze two candidate genes that may play a role in seed development. The genes presented in this work were selected from a collection of promoter trap lines expressing the *GUS* reporter gene in seed organs (Stangeland et al., 2003). The promoterless reporter gene was introduced in the *Arabidopsis* genome by the random insertion of a T-DNA construct. Because the expression of the *GUS* gene reflects activities of tagged genes, the lines were further screened for tissue-specific *GUS* expression in the seed. This screen identified 16 lines showing *GUS* expression in the endosperm, embryo and other seed organs at different developmental stages. In order to identify these seed specific genes in *Arabidopsis*, the position and pattern of the T-DNA integration in 16 lines was analyzed and determined (Stangeland et al., 2005). The tagged genes in two particular marker lines (GNOCCHI/760 and LINGUINE6/619) were chosen for further investigation based on their gene identities and *GUS* expression pattern.

The GNOCCHI/760 line showed *GUS* expression in the chalazal endosperm and embryo at the embryo globular stage. The *GUS* activity was also detected at the early heart stage, but specifically in the chalazal endosperm (Stangeland et al., 2003). The T-DNA in this line was found to be inserted in the third intron of At2g28305 gene (Stangeland et al., 2005), encoding a short protein of 213 aa that belongs to the lysine decarboxylases (LDC) protein family. Lysine decarboxylase catalyzes the reaction generating cadaverine by decarboxylation of L-lysine. Cadaverine is a diamine that is shown to possible play an important role in root development of germinating soybean (*Glycine max*) seeds (Gamarnik and Frydman, 1991). This gene was chosen for further analysis because of the characteristic *GUS* expression in the chalazal endosperm found in the GNOCCHI/760 line. Similar *GUS* expression pattern in chalazal endosperm has been detected in lines carrying promoter *GUS*

construct of the maternal-effect *FIS* genes: *MEA* and *FIS2* (*MEA*::*GUS* and *FIS2*::*GUS*) (Luo et al., 2000) and the *PHE1* gene (*PHE1*::*GUS*) that is maternally repressed by *MEA* (Köhler et al., 2003b). In addition, the *GUS* expression pattern in GNOCCHI/760 line was similar to the expression pattern of the *GFP* reporter gene in KS117 marker line (Sørensen et al., 2001). In this line the expression of the tagged gene was shown to be under maternal control by *FIS* genes. *FIS* genes were also shown to play major role in setting of anterior-posterior polar axis during endosperm development (Sørensen et al., 2001). Hence, the characteristic expression in the chalazal endosperm may suggest a link between *FIS*, *PHE1* and the tagged *LDC* gene in endosperm development and organization.

The LINGUINE6/619 line showed very weak expression in the basal part of the embryo and the surrounding endosperm from the globular to the walking stick stage (Stangeland et al., 2003). The T-DNA in this line was integrated near the transcription start of At5g03740 gene, encoding a histone deacetylase (Stangeland et al., 2005). This gene was selected for further analysis because of its putative role in gene regulation. A significant regulatory level of gene expression occurs by specific modification of the chromatin components, such as acetylation, methylation and phosphorylation in specific amino acid residues of histone tails. These modifications affect gene expression via induction of changes in chromatin structure and lead to activation or repression of the associated DNA (Strahl and Allis, 2000; Jenuwein and Allis, 2001). Acetylation of lysine 9 in the histone tail of histone H3 (H3K9) has a general activating effect on transcription. H3K9 deacetylation and methylation, and the associated H3K9-specific histone deacetylases and histone methyltransferases are regulators of chromatin assembly in a negative manner leading to an inactive chromatin state (heterochromatin) and gene silencing (reviewed in Jenuwein and Allis, 2001). In addition, the *Drosophila* PcG proteins ESC (homologues to *FIE*) and E(Z) (homologues to *MEA*) are present in a complex that is shown to interact with histone deacetylases (Tie et al., 2001). It has been proposed that PcG complexes establish repressive chromatin environments at certain target loci through histone deacetylation and histone methylation (reviewed in Francis and Kingston, 2001). Hence, the tagged histone deacetylase gene may play a key role in repression of gene expression in seed.

3.2.1 Expression analysis of lysine decarboxylase gene (*LDC*) At2g28305

The expression pattern of *LDC* was investigated by RT-PCR, Northern hybridization and promoter-*GUS* fusion. The promoter-*GUS* fusion was done to analyze *LDC* expression during seed development and to compare the expression pattern of the endogenous *LDC* gene with the expression pattern of the tagged *LDC* gene in the GNOCCHI/760 line.

3.2.2 RT-PCR analysis of the *LDC* reveals expression in all major tissues

mRNA isolated from flowers, seedlings, rosette leaves and siliques of *Arabidopsis* ecotype C24 was reverse transcribed into a first-strand cDNA that was used as template in RT-PCR. Primers 146-760REV and 150-760DIR were designed to anneal in different exons, amplifying a fragment of 569 bp. The PCR was carried out with 25 cycles (see Appendix 2). Negative control-PCR was carried out using the same primer set and C24 DNA as a template (data not shown).

The results presented in Figure 21 a show clear bands of the expected size and indicate that the *LDC* gene is expressed in flowers, seedlings, rosette leaves and siliques. It appears also that the expression of *LDC* is higher in flowers and rosette leaves than in seedlings and siliques.

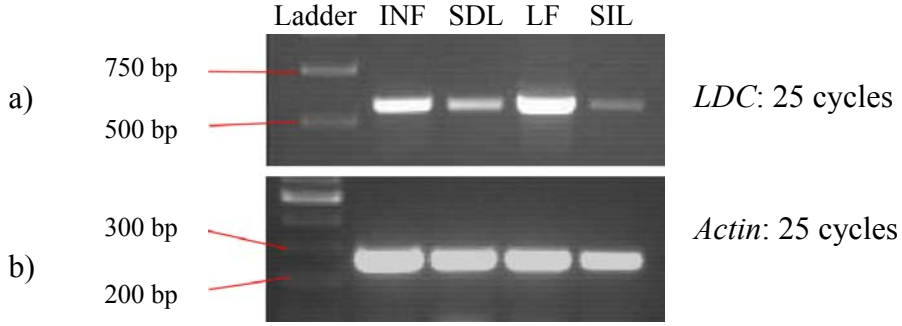

Figure 21. RT-PCR analysis. Expression analysis of (a) *LDC* gene in flowers (INF-inflorescence), seedlings (SDL), rosette leaves (LF) and siliques (SIL). Right-hand side designations refer to the gene specific primers and number of cycles in RT-PCR reactions. Actin primers (b) were used as template control.

3.2.3 Northern hybridization shows that *LDC* expression is most prominent in flowers and rosette leaves

The expression pattern of the *LDC* gene was additionally investigated by Northern blotting. The mRNA isolated from wild-type *Arabidopsis* flowers, seedlings, rosette leaves and siliques at different developmental stages was blotted onto a membrane and hybridized with a *LDC* probe. The probe template was amplified by PCR using 146-760REV and 150-760DIR primers that generated 569 bp long PCR product. Strong expression of *LDC* is observed in flowers and rosette leaves. In seedlings and siliques the expression seems to be reduced to some extent compared to flowers and rosette leaves (Figure 22). The results from Northern hybridization confirm the results from the RT-PCR expression analysis.

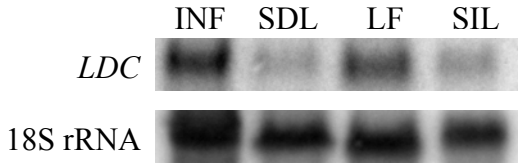

Figure 22. Expression pattern of *LDC* by Northern analysis. A northern blot on mRNA extracted from flowers (INF-inflorescence), seedlings (SDL), rosette leaves (LF) and siliques (SIL) of wild-type C24 *Arabidopsis* was hybridized with *LDC* probe. 18S rRNA probe was used as a loading control.

3.2.4 Promoter-*GUS* construct for the *LDC* gene

The expression pattern of the gene of interest can be studied by cloning its promoter region in front of a promoterless *GUS* gene. The *GUS* gene will be regulated by this promoter and the GUS protein expressed according to the expression pattern of the gene of interest. To investigate whether the *GUS* expression in the line GNOCCHI/760 was regulated by *LDC* promoter and if it revealed the expression pattern of the tagged *LDC* gene, the promoter-*GUS* fusion construct was made and transformed into *Arabidopsis* plants.

The transgene was made by amplifying 1509 bp of the promoter region and 5' UTR of the lysine decarboxylase gene (*pLDC*) using primers labeled with *att*B1 and *att*B2 Gateway sequences in the 5' end, 760-attB1 and 760-attB2 respectively (Figure 23). The PCR fragment was recombined into a pPZP211G-GAWI Gateway destination vector (see Materials and methods and Figure 24). The construct, verified by sequencing, was

introduced into *Agrobacterium tumefaciens* strain C58C1 pGV2260 and transformed into *Arabidopsis* ecotype C24 plants by the floral dip method.

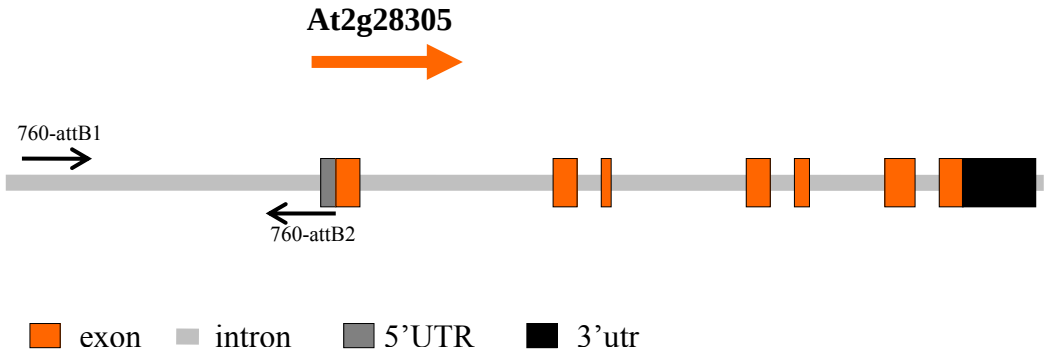

Figure 23. Schematic illustration of the annotated *LDC* gene (At2g28305) and primers used in amplifying 1509 bp of its promoter region, including 5' UTR (marked in grey). The orange arrow shows the direction of transcription.

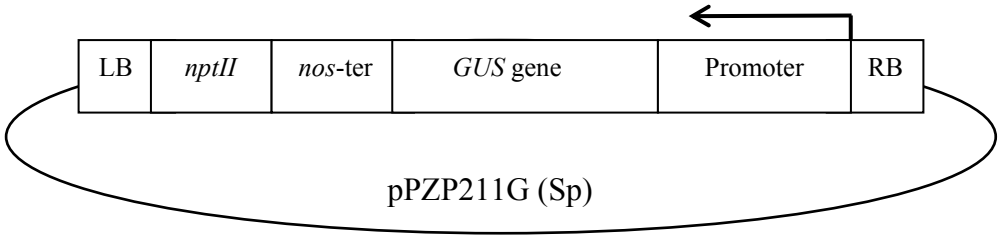

Figure 24. The promoter region of the *LDC* cloned in front of a promoterless *GUS* gene and a *nos* terminator in the vector pPZP211G. The vector is a Ti vector containing right and left border sequences (RB and LB), spectinomycin (Sp) resistance gene for selection in bacteria and kanamycin (*nptII*) resistance gene for selection of transformants. The arrow shows direction of transcription.

3.2.5 Analysis of the *GUS* expression pattern from the lysine decarboxylase promoter

To confirm that the *pLDC::GUS* T_1 transformants indeed carried the inserted transgene, two control-PCRs were done using two *pLDC* specific primers in combination with one *GUS* specific primer (see Appendix 1 and Appendix 3) (data not shown). In addition, positive and negative control-PCR was run using the same primer pairs. The pPZP211G harboring the

pLDC::GUS transgene was used as a positive control template, while C24 DNA was used as a negative control template (data not shown).

The T_1 transformants confirmed to carry the *pLDC::GUS* transgene were analyzed for *GUS* expression using a histochemical GUS detection assay (Grini et al., 2002). The GUS-assay was performed on flowers, rosette leaves and dissected siliques at different developmental stage. The *pLDC::GUS* expression was detected in the vascular tissue of flower petals and stamen filaments (Figure 25). In rosette leaves, the weak expression was found in the vascular tissue and veins (Figure 25b). In early post fertilization stages, the *pLDC::GUS* was expressed in maternal chalazal tissue (Figure 25c). In seed at the early globular embryo stage the *pLDC::GUS* expression was observed in syncytial micropylar endosperm (Figure 25d). This was the earliest developmental stage, after fertilization, where the *GUS* expression was observed in the fertilization product of the seed. No expression was detected in the embryo at this stage. At the heart embryo stage the *pLDC::GUS* expression was found in the embryo proper, with the strongest expression in the epidermis and cells facing the epidermis (Figure 25e). At this stage the weak *pLDC::GUS* expression was also found in endosperm, with the equal intensity in micropylar, central and chalazal endosperm. The expression was stronger at this stage in aleurone cells, the cells lining the maternal seed coat (Figure 25e). The *pLDC::GUS* expression was also found at near mature stage, in the cotyledons of the bent cotyledon embryo stage (Figure 25f). In addition, a strong *pLDC::GUS* expression was detected in maternal chalazal tissue of the seed at all developmental stages after fertilization (Figure 25c-f). The observed expression pattern of the *pLDC::GUS* transgene was different from the *GUS* expression pattern in the GNOCCHI reporter line (Stangeland et al., 2003; Stangeland et al., 2005). The expression of this transgene confirmed the expression of *LDC* in the endosperm and embryo, even though the characteristic expression pattern in chalazal endosperm in the GNOCCHI line could not be verified.

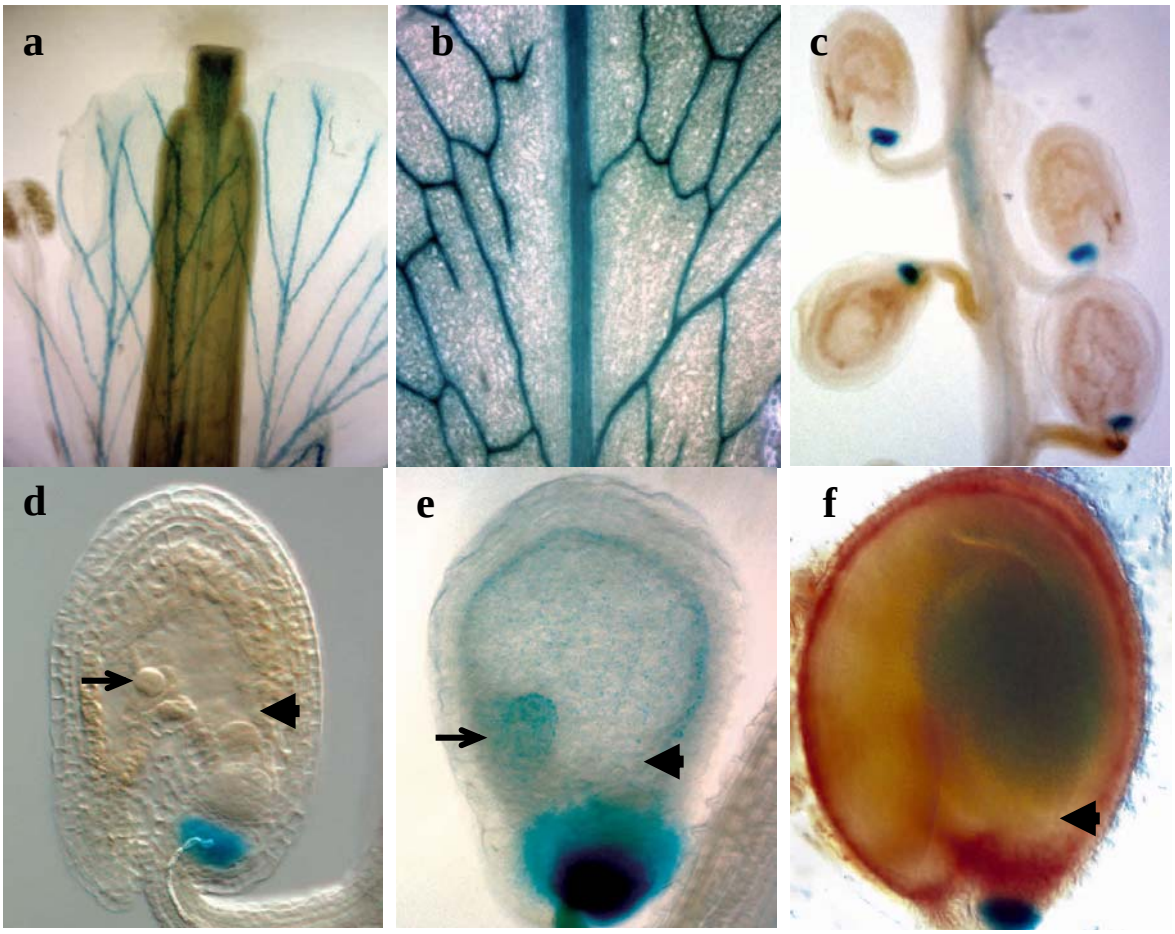

Figure 25. Expression pattern of *LDC* by *pLDC::GUS* transgene in T_1 transformants. a) *pLDC::GUS* expression in vascular tissue of flower petals and stamen filaments. b) Segment of *pLDC::GUS* expression in rosette leaf vascular tissue. c) Expression of *pLDC::GUS* in maternal chalazal tissue. d) Seed at globular embryo stage. Weak *pLDC::GUS* expression in micropylar central endosperm. e) Seed at heart embryo stage. *pLDC::GUS* expression in embryo and endosperm. f) Seed at bent cotyledon embryo stage. *pLDC::GUS* expression in cotyledons. Arrow shows embryo and arrowhead chalazal endosperm.

3.2.6 Expression analysis of the histone deacetylase gene (*AtHD2C*) At5g03740 by Northern hybridization

The expression pattern of *AtHD2C* gene was investigated by Northern hybridization. mRNA isolated from flowers, seedlings, rosette leaves and siliques at different developmental stages was blotted onto a membrane and hybridized with *AtHD2C* probe. The probe template was amplified by PCR using HDAC forward and HDAC reverse primers that generated 566 bp long PCR product. The results from the Northern hybridization (Figure 26) show that the *AtHD2C* is mainly expressed in flowers, but some weaker expression is also observed in

seedlings, siliques and leaves. However, the *GUS* expression in the LINGUINE6/619 reporter line was not detected in seedlings and leaves (Stangeland et al., 2003).

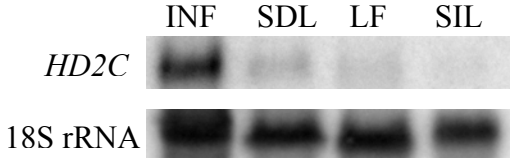

Figure 26. Expression pattern of *AtHD2C* by Northern analysis. A northern blot on mRNA extracted from flowers (INF-inflorescence), seedlings (SDL), rosette leaves (LF) and siliques (SIL) of wild-type C24 *Arabidopsis* was hybridized with *HD2C* probe. 18S rRNA probe was used as a loading control.

3.2.7 Study of T-DNA knockout mutants of histone deacetylase gene

In order to investigate the role of the *AtHD2C* gene in the seed development, two lines (SALK_039774 and SALK_039784) reported to carry a T-DNA insertion in the seventh exon of *AtHD2C* gene were ordered from the SALK Insertion Sequence Database (http://signal.salk.edu). The lines were genotyped for the T-DNA insertion and examined for any detectable seed abnormalities. Fourteen T_3 plants from SALK_039784 and ten plants from SALK_039774 were genotyped by two PCRs. When HDAC LP and HDAC RP gene specific primers were used to amplify SALK_039774 lines, the PCR fragment of expected size was amplified for each line (data not shown). Amplification with the T-DNA left border specific primer LBa1 and the gene specific primer HDAC RP did not give any PCR product (data not shown), indicating that the SALK_039774 line most likely did not have any T-DNA insertion in the *AtHD2C* gene. The PCR analysis of fourteen SALK_039784 lines using gene specific primers (HDAC LP and HDAC RP) and the T-DNA left border specific primer (LBa1) in combination with the gene specific primer (HDAC RP) resulted in 9 wt plants, 4 hemizygous and 1 homozygous for the T-DNA insertion. If the T-DNA segregates as one Mendelian locus the expected ratio between wt, hemizygous and homozygous plants would be 1:2:1. The chi-square test gave a P-value of 0.0028 (rejecting H_0 [6]), showing that the observed ratio is significantly different from the Mendelian ratio.

[6] H_0: The difference between the observed and expected ratio is random

If the insertion of the T-DNA induced either a female or male gametophytic mutation, the plant homozygous for the T-DNA insertion would not have been developed and the ratio of wt plants to hemizygous to homozygous would have been 1:1:0. The chi-square test for this expected ratio gave a P-value of 0.0008[7], indicating that the observed ratio is also significantly different from this expected ratio. However, the chi-square test will reject H_0 when there are observations in a class that is expected to have none (homozygous). If we omit the one homozygous observation and perform the test again using ratio 9:4:0, the chi-square test give a P-value of 0.37 (keeping H_0), suggesting that the T-DNA insertion in the *AtHD2C* could have induced a gametophytic mutation. One observed homozygous can also mean that the T-DNA insertion (mutation) in this locus is not fully penetrant. However, when the phenotype was examined, no visible phenotype in seeds or other organs was detected. The involvement of the *AtHD2C* gene in seed development can thus not be verified at this stage and must be analyzed in subsequent generations.

[7] In order to avoid 0 in the denominator in the calculation, the test was performed on the ration 1:1:0.01.

4. Discussion

The general objective of this thesis was the identification and characterization of genetic components involved in seed development. In particular, so called gametophytic maternal-effects was the main focus since this mode of genetic regulation of seed development is poorly understood. Forward and reverse genetics approaches were employed to identify the molecular nature of the gametophytic maternal *cap2* mutant and further to analyze and characterize *AtHD2C* and *AtLDC*, two genes putatively involved in seed development. In the first part of this thesis the main effort was to map and identify the *cap2* gametophytic maternal-effect mutation in order to identify the molecular identity of the *CAP2* gene and determine its role in seed development. In the second part of this work various expression and functional analyses of the *AtHD2C* and *AtLDC* genes was carried out to investigate the putative role of these two genes in seed development.

The first part of this discussion will focus on the results from the molecular analysis of the promoter trap lines. The second part discusses the results from the molecular mapping of the *cap2* mutation, possible involvement of the mapped genes in seed development and the outcome of the complementation analysis.

4.1 Molecular analysis of promoter trap lines

4.1.1 The *AtLDC* gene is expressed in the seed

The *LDC* promoter-*GUS* fusion construct was made with the aim to investigate whether the *GUS* expression pattern observed in the GNOCCHI/760 marker line was a reflection of the expression pattern of the tagged *LDC* gene. The *pLDC::GUS* expression was detected in flowers, rosette leaves and seeds, as described previously. The *LDC* expression pattern in these organs was confirmed by both RT-PCR and Northern hybridization. In addition, the RT-PCR and Northern hybridization showed that the *LDC* was also expressed in seedlings. However, the expression was stronger in flowers and rosette leaves compared to siliques and seedlings. These results are consistent with the observed *pLDC::GUS* expression pattern in flowers, rosette leaves and seeds. The weak expression in siliques may reflect the restricted

expression of *pLDC::GUS* to specific cells of the seeds in the siliques, as shown previously (Figure 25).

The expression pattern of the *LDC* gene was different from that observed in the GNOCCHI/760 line (Stangeland et al., 2003; Stangeland et al., 2005). In this marker line the *GUS* activity was not detected in flowers, leaves and seedlings. In addition, no obvious expression of the *LDC::GUS* transgene was detected in the chalazal endosperm. In the GNOCCHI/760 line, the T-DNA harboring the *GUS* gene was inserted in the third intron of the *LDC* gene. It was found that the fusion LDC-GUS transcript confirmed by Northern hybridization and sequencing of a cDNA fragment, can not be translated into a fusion protein. However, the *LDC* exons may function as untranslated leader sequences, where translation of the GUS starts from the ATG of the *GUS* gene at a sufficient frequency for histochemical detection (Stangeland et al., 2005). Therefore, lack of *GUS* expression in seedlings, flowers and leaves of the GNOCCHI/760 line, contrary to the *pLDC-GUS* transgene, may be due to limited use of the ATG start codon of the *GUS* gene. This is because the ATG codon of the *GUS* gene is not in frame with the upstream ATG of the *LDC* gene in the LDC-GUS fusion transcript (Stangeland et al., 2005).

The different expression of the *GUS* gene in the GNOCCHI/760 line compared to what was observed from the *pLDC::GUS* transgene can be explained by different theories. One possibility is that the *LDC* promoter fragment in the *pLDC::GUS* transgene did not encompass all regulatory elements required, and thus not sufficient to drive the expression of the *GUS* gene alone. Other regulatory sequences (upstream promoter elements) may also be responsible for the *LDC* expression *in planta*. Furthermore, an enhancer element, located some distance upstream or downstream from the *LDC* gene may be responsible for its increased transcription *in planta*. The *pLDC::GUS* transgene may also have been inserted in a genomic location not supporting transcriptional activity, such as facultative heterochromatin regions. The latter is however unlikely, since several independent lines showed a similar expression pattern.

The enhancer trap KS117 marker line and its corresponding *AtFH5* tagged gene (belonging to the *Arabidopsis* formin homology family) is one previously described example where the

enhancer activity detected by the reporter line is not entirely consistent with the endogenous gene expression pattern. The KS177 marker line is characterized by the specific expression of the mGFP5 in the chalazal endosperm at the posterior pole (Sørensen et al., 2001). In this marker line the mGFP5 expression starts at the 8 nuclei (stage III) syncytial endosperm and is detected in MCE, PEN and CZE endosperm. The expression remains present in the entire syncytial endosperm up to the stage VIII (in which the endosperm contains 90-100 nuclei). After this stage, the mGFP5 expression gradually decreases in the MCE and PEN but remains high in the CZE endosperm. By contrast, the expression of the *AtFH5* promoter-reporter gene (*AtFH5::H2B-YFP*) transgene is detected only in the posterior (CZE) endosperm at the 16 nuclei stage and remains restricted to the posterior pole throughout the syncytial stage (Ingouff et al., 2005).

4.1.2 Analysis of the *AtHD2C* histone deacetylase gene

The *AtHD2C* gene was analyzed due to its putative requirement in seed development. The Northern blot analysis showed that the *AtHD2C* was mainly expressed in flowers. Expression was also detected in siliques, seedlings and rosette leaves, however only weakly. In the LINGUINE6/619 marker line, harboring the T-DNA insertion in the upstream region of the *AtHD2C* gene, the *GUS* activity was detected in the basal part of the embryo and the surrounding endosperm from the globular to the walking stick embryo stage (Stangeland et al., 2005). However, the *GUS* expression in this line was not found in seedlings and leaves as detected for the *AtHD2C* gene by Northern hybridization.

A T-DNA insertion line, SALK_039784, reported to carry a T-DNA insertion in the seventh exon was analyzed for the defect in seed development. Plants were genotyped in order to fined hemizygous or homozygous for the T-DNA insertion. These were thereafter screened for the visible phenotype in seeds. Homozygous plants are not expected to be obtained if a T-DNA insertion induces a gametophytic mutation. This is because an ovule carrying the maternal mutated allele will not develop into a functional seed after fertilization. Thus, one would expect a ratio 1:1:0 for wt:hemizygous:homozygous plants (see Results, Figure 20) if a T-DNA knockout mutant line carries a T-DNA insertion in a locus for the maternal inherited trait. If a T-DNA insertion does not induce a gametophytic mutation, the T-DNA would segregate as a Mendelian locus with the ratio 1:2:1 for wt:hemizygous:homozygous.

The chi-square test was performed to test if the observed ratio between wt, hemizygous and homozygous plants was in agreement with the Mendelian distribution (1:2:1) or distribution for the gametophytic mutation (1:1:0). The statistical analysis showed that the T-DNA insertion in this line could possibly have induced the gametophytic mutation. However, the phenotyping did not reveal any visible defect in seed or other organs. It is well known that the extensive gene duplication in *Arabidopsis* has caused important functional redundancies in this plant (Vision et al., 2000). Therefore, the lack of visible phenotype could be due to possible redundancy between *AtHD2C* and some other members of the *HD2* gene family. Experiments involving double or multi-mutant combinations might aid in the elucidation of HD2 family function.

4.2 Molecular mapping and characterization of the *capulet2* gametophytic maternal-effect mutant

4.2.1 Molecular mapping of *cap2* identified two candidate genes

The *cap2* gametophytic maternal-effect mutant was originally identified in a linkage based screen using a line marked with multiple visible markers (Grini et al., 1999; Grini et al., 2002). The position of *cap2* was roughly mapped near the bottom end of the chromosome 1, between *ap1* and *gl2* visible recessive markers. The genetic distance of *cap2* from the *ap1* and *gl2* markers was calculated from marker segregation data, and determined to be ~18 cM south of *ap1* and 2 cM north of *gl2* (Grini *et al*., 2002). In the initial mapping of the *cap2* mutant done by Grini *et al.* (2002) the mapped interval spanned ~1.2 Mb and was flanked by ADH and SNP142 molecular markers. The first task of this work was to confirm the initially determined interval in order to proceed with mapping within this interval and finally determine the putative locus of the *cap2* mutation. A new mapping population, consisting of 470 plants was created for this purpose. Furthermore, to ensure that the *cap2* was not positioned outside the ADH-SNP142 interval, two new molecular markers, nga111 and G17311 were chosen as the flanking markers at the centromere and telomere side, respectively (Figure 27). The nga111 marker was located about 1.6 Mb south of the ADH marker and the G17311 marker was located about 140 kb north of the SNP142 marker. Thus,

the initial mapping interval in this work (nga111-G17311) spanned about 2.9 Mb. A total of 31 plants were found to be recombinants within this interval.

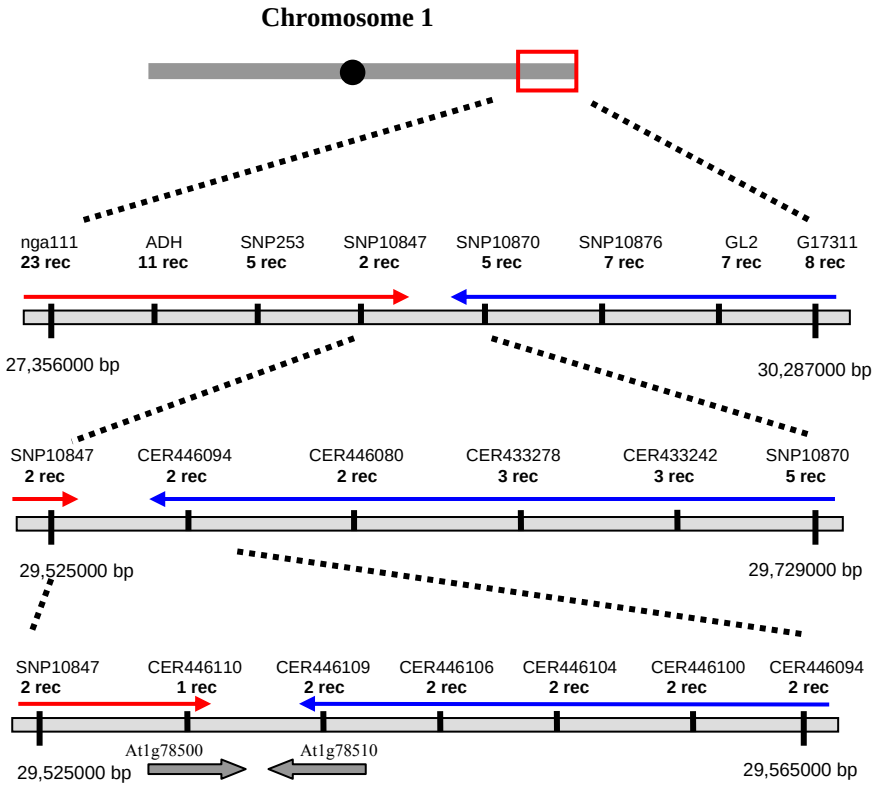

Figure 27. Schematic illustration of molecular mapping of *cap2*. Top panel shows mapping interval on the right end of chromosome 1 (red rectangle). All molecular markers used in the mapping are presented. Red and blue arrows indicate the recombination event from the centromere side and the telomere side, respectively. Number of recombinants found in each molecular marker is written under the marker name. Distance between two markers flanking the mapping interval is marked in bp. Two genes spanning the final interval are marked as grey arrows, showing the direction of transcription.

Since the positional cloning is a process that narrows down the genetic distance between a mutation and a molecular marker by successively excluding all other parts of the genome (Lukowitz et al., 2000), new molecular markers need to be developed continuously to the vicinity of the mutation as the fine mapping progresses. In this work twenty new molecular markers were developed and used together with four pre-existing markers from the TAIR database to finally map the *cap2* mutation to a genetic interval of 4238 bp. The final interval contained one recombinant in the CER446110 marker flanking the interval from the

centromere side, and two recombinants in the CER446109 marker flanking the interval from the telomere side. No recombinants were found in CER446110 marker from the telomere side, and none was found in the marker CER446109 from the centromere side. The final mapping interval included parts of two genes: At1g78500, encoding a putative pentacyclic triterpene synthase, and At1g78510, encoding solanesyl diphosphate synthase. In order to determine if any of these two genes could be the *CAP2* locus, a complementation analysis was performed and will be discussed later. Figure 27 summarize the whole mapping process and illustrates all main markers used in *cap2* mapping together with number of recombinants detected for each molecular marker as the interval was gradually reduced.

4.2.2 The RT-PCR analysis detects expression of both At1g78500 and At1g78510 in siliques and flowers

If the *cap2* phenotype was caused by a lesion in one of the At1g78500 and At1g78510 genes, we would expect to detect expression of that gene in siliques and flowers. Thus expression of the At1g78500 and At1g78510 was examined by semi-quantitative RT-PCR in which all samples had the same amount of template. In addition, each reaction was run using the lowest number of cycles that could amplify the product. The expression of the At1g78500 and At1g78510 was examined in flowers, siliques, seedlings and rosette leaves of Columbia ecotype.

The At1g78500 transcript could not be amplified using 30 PCR cycles, but a product was detected when the reaction was run with 40 cycles. This indicates that the At1g78500 is a weakly expressed gene. In addition, the expression was detected in flowers, siliques and seedlings, but not in rosette leaves. It appeared that the cDNA level (corresponding to the mRNA level) was highest in siliques and lowest in seedlings. These results seem to be consistent with out present knowledge of *cap2* and its putative function in the seed.

Expression of the At1g78510 gene was confirmed in reaction run with 25 cycles. The expression was detected in flowers, siliques, seedlings and rosette leaves. It seemed that the expression of the At1g78510 was higher in rosette leaves and seedlings than in flowers and siliques. These observations suggest that if the At1g78510 gene encodes the *CAP2* locus it is not seed specific. However, the *CAP2* gene could also be expressed and have function in

other organs besides seeds. The expression of the maternal-effect *FIE* in vegetative tissues suggested that this gene was likely to play a role in vegetative development (Ohad et al., 1999; Sørensen et al., 2001). This may also be the case for this putative *CAP2* gene.

4.2.3 Characterization of the At1g78500 and At1g78510 genes and their possible involvement in seed development

As the final mapping of the *cap2* mutation resulted in two genes, it was of interest to investigate what function these two genes have *in planta* and if they possibly could be involved in control of seed development. For this reason, this part of the discussion will focus on whether any of the At1g78500 and At1g78510 genes could be the putative *CAP2* locus in view of their known function in *Arabidopsis*.

Pentacyclic triterpene synthase, encoded by At1g78500, has a catalytic activity in pentacyclic triterpenoid biosynthesis (sterol biosynthesis), whereas solanesyl diphosphate synthase, encoded by At1g78510, has dimethylallyltranstransferase activity in monoterpenoid biosynthesis. Thus, both of these genes encode proteins that have function in terpene biosynthesis.

Terpenes (or isoprenoids) form a large group of lipids that are synthesized by the condensation of the two basic five carbon units, isopentenyl diphosphate and its isomer dimethylallyl diphosphate (Lichtenthaler et al., 1997; Kasahara et al., 2002). The isoprenoid biosynthetic pathway is used to synthesize a wide variety of molecules, typically lipophilic, with a vast range of biochemical functions. In plants, the isoprenoids play important roles as quinones in electro transport chain, as components of membranes (sterol), in subcellular targeting and regulation (prenylation of proteins), as photosynthetic pigments (carotenoids, side chain of chlorophyll), as hormones (gibberellins, brassinosteroids, abscisic acid, cytokinins), and plants defense compounds as well as attractants for pollination (monoterpenes, sesquiterpens and diterpenes) (Laule et al., 2003).

Monoterpenes and triterpenes are a group of isoprenoids that contain 10 and 30 carbons, respectively (Fox and Whitesell, 1997). In plants the biosynthesis of these two groups of isoprenoids take place by two independent pathways, the methylerythritol phosphate (MEP)

pathway and the mevalonate (MVA) pathway, localizes in different cellular compartments. The MEP pathway is localized to the chloroplast and is generally responsible for the formation of mono- and di-terpenoids, plastoquinones, and the prenyl group of chlorophylls. On the other hand, the MVA pathway is localized in cytoplasm and, in plants plays an essential role in the biosynthesis of triterpenoids, sterols and sesqiuterpenoids (Lichtenthaler et al., 1997; Kasahara et al., 2002; reviewed in Chappell, 2002).

As mentioned previously, At1g78510 is involved in monoterpenoid biosynthesis which is localized to the chloroplast. Several lines of evidence suggest that the chloroplast is a site of important functions for the plant cell (amino acid biosynthesis, fatty acid biosynthesis, biosynthesis of gibberellins) and that its development and differentiation are essential for proper embryo development (Uwer et al., 1998; Albert et al., 1999; Apuya et al., 2001; Despres et al., 2001; Apuya et al., 2002).

During *Arabidopsis* embryo development, differentiated chloroplasts are observed beginning at the heart stage, and their number continues to increase until the end of embryo growth (Mansfield and Briarty, 1991). Several embryo mutants identified, suggested that chloroplast development and differentiation are essential for the transition from globular to heart embryo stage. Mutation in the nuclear *EMBRYO-DEFECTIVE DEVELOPMENT* (*EDD1*) gene encoding a glycyl-tRNA synthetase, which is targeted to chloroplast, leads to arrested embryonic development at the globular-heart stage (Uwer et al., 1998). The *EMB509* gene, which is also required for the normal post-globular development of the embryo, encodes a putative protein with ankyrin repeats that plays a role in chloroplast development (Albert et al., 1999; Despres et al., 2001). Moreover, mutation in the nuclear-encoded chaperonin-60α protein (*schlepperless* (*slp*)) required for proper folding of chloroplast-bound protein leads to a defect in plastid development that result in cotyledon shortening and embryo arrest before the heart stage (Apuya et al., 2001). The *raspberry3* (*rsy3*) mutation, which causes arrest of embryogenesis at the globular stage, is located in a nuclear gene that encodes a novel protein, which is proposed to be required for chloroplast differentiation (Apuya et al., 2002). All these observations indicate the importance of the function and integrity of chloroplasts in the development of the embryo. A defect in chloroplast differentiation during early embryo formation may prevent the formation of important biosynthetic precursors that are required

in subsequent metabolic steps for the production of embryo signaling molecules (Apuya et al., 2002). Mutation in the At1g78510 gene may affect chloroplast development that further would possibly prevent proper embryo development and lead to the *cap2* phenotype.

On the other hand, a role of the At1g78500 gene in seed development could also be proposed considering that this gene is putatively involved in triterpenoid (sterol) biosynthesis. Sterols are isoprenoid-derived lipids that have diverse and essential function in all eukaryotes. They are major components modulating the permeability and fluidity of eukaryotic cell membranes. In addition to their structural role, sterols function as biosynthetic precursors of steroid hormones (such as brassinosteroids (BRs) in plants) (Clouse, 2002). The role of animal steroids in the regulation of embryonic and postembryonic development along with adult homeostasis is well known (Beato et al., 1995).

Two *Arabidopsis* mutants, *sterol methyltransferase1* (*smt1*) (Diener et al., 2000) and *fackel* (*fk*) (Jang et al., 2000; Schrick et al., 2000), showing several defects in embryogenesis, have been identified to be affected in early steps of sterol biosynthesis. SMT1 is a C-24 methyltransferase responsible for the conversion of cycloartenol to 24-methylene-cycloartenol in the first step of sterol biosynthesis, whereas FK acts downstream of MST1 with C-14 sterol reductase activity. Mutants affected in these steps have reduced sterol and BR levels and are affected uniquely at specific stages of embryo development. Both *smt1* and *fk* show a lack of asymmetrical divisions in cells in the center of the globular embryo, and mutant embryos arrest at this stage. The fact that the disrupted genes in these two embryo mutants are involved in sterol synthesis might indicate that the mutation in the putatively pentacyclic triterpene synthase encoding the At1g78500 gene could be responsible for the *cap2* phenotype.

4.2.4 Complementation experiment and analysis of putative *cap2* T-DNA knockout mutants did not resolve the question about molecular identity of the *CAP2* gene

Complementation demonstrates that the introduction of a wild type allele into the mutant plant restore or rescue the mutant phenotype. This is a clear proof that the isolated and the cloned gene is the one that was mutated in the mutant line. *cap2* plants were transformed with two independent complementation constructs, one containing the coding sequence of

the At1g78500 gene with 1331 bp promoter region and a downstream sequence including 418 bp of the At1g78510 CDS. The other construct contained the coding sequence of the At1g78510 gene with 1173 bp promoter region including 462 bp of the upstream At1g7820 gene and 684 bp downstream sequence (see Results, Figure 16).

Eight independent transformants harboring the At1g78500 construct and two harboring the At1g78500 construct were obtained. Siliques from all T1 transformants were examined for the *cap2* phenotype. The results showed that the percentage of arrested seeds in plants harboring the At1g78500 complementation construct varied from 3.4% to 39.5%. Two plants carrying the At1g78510 construct had 40.5% and 71.7% arrested seeds in their siliques.

It is not straightforward to determine if complementation has been accomplished or not in gametophytic mutants having a reduced seed phenotype. Several factors influence the complementation results and have to be taken into account when phenotyping the transformants. As described previously in the Results, section 3.1.9.2 T_1 *cap2* seeds in which the rescue T-DNA has been inserted prior to fertilization would be able to develop and grow. Since the female gametophyte is the primary target of T-DNA insertion (Bechtold et al., 2000; Desfeux et al., 2000) and the complementation construct is introduced in heterozygotic *cap2*/*CAP2* plants, only four T_1 progeny classes presenting four different genotypes would be selected on hygromycin medium. The genotypes are defined by two main parameters: 1) whether the transformed ovule is *cap2* or *CAP2* and 2) whether the transformed ovule if fertilized by *cap2* or *CAP2* pollen (see Results, Figure 17). If the inserted wild type gene complements the *cap2* mutation four scenarios, depending on the genotypes of the T_1 plants, can be obtained from the self-cross of the T_1 plants. The rescue frequency (ratio of *CAP2* to *cap2* seeds) in these crosses is dependent on gamete frequencies (parental and recombinant), which on the other hand depend on genetic distance or recombinant frequency between the *CAP2* and the T-DNA loci. Therefore, assuming that only one T-DNA has been inserted per genome in T_0 plants, the frequency of *cap2* seeds could range from zero to 50 % (see Results, section, 3.1.9.2)

Statistical analysis of the phenotyping results indicated that in seven out of eight transformants harboring the At1g78500 construct, complementation could have been

accomplished, given that they carry one T-DNA insertion. In addition, only in one transformant line out of two, harboring the At1g78510 construct, the results indicated that complementation could have been achieved.

Since the results from the complementation experiment were considered under assumption that only one T-DNA was inserted per genome in the T_0 plants, the additional verification of the results had to be done by segregation analysis. The segregation analysis was performed to determine the number of inserted T-DNA per transformant line, and then to reveal if the observed ratio between *CAP2* and *cap2* seeds in the T1 plants was due to complementation. As described previously in section, 3.1.9.2, multiple T-DNA insertions would produce higher rescue frequencies, and the ratio between *CAP2* and *cap2* seeds would be more complex than when the single T-DNA copy is inserted. In the progeny produced by self-crossing of plants carrying three T-DNA insertions the expected frequency of *cap2* seed phenotype would range from zero to 6.25% if the complementation is obtained.

The number of expected T-DNA insertions was calculated according to general ratio $(1-0.25^n):0.25^n$ for $Hyg^R:Hyg^S$, where n is number of T-DNA loci. The chi-square test was used to check how our observations fitted with the expected results. The segregation analysis indicated that three or more T-DNA copies were inserted per genome in each transformant line. In addition, in all lines but one (16.5.1.1) the frequency of the *cap2* seed phenotype was larger than 6.25%. Hence, our assumption that the complementation could have been accomplished was weakened for all of the lines, except for the line 16.5.1.1. In addition to line 16.5.1.1, line 16.5.1.3 with the *cap2* seed frequency of 8.1% could also be considered as a possibly complemented line. However, it is not possible to say if the low frequency of the aborted seeds in these two lines is due to complementation or if they carry the wt *CAP2* alleles. In order to check that, the lines should be crossed with the wt pollen, and the progeny examined for the *cap2* phenotype. If the progeny exhibit 50% of *cap2* seed phenotype, it would indicate that the corresponding lines had *cap2* genotype and that At1g78510 complement *cap2*.

Since neither the At1g78500 nor At1g78510 rescue construct did clearly show to rescue the *cap2* mutation by restoring the wild-type seed phenotype in transformant lines, three T-DNA

insertion lines carrying T-DNA insertions in the At1g78500 and At1g78510 genes were ordered to investigate if they would show *cap2* phenotype. Two T-DNA insertion lines were reported to carry a T-DNA insertion in the ninth and fourteenth exon of the At1g78500 gene (SALK_060682 and SALK_036693, respectively), while the third one (SALK_126948) carried a T-DNA insertion the first exon of the At1g78510 gene. Plants from each line were first genotyped in order to find hemizygous or homozygous for the T-DNA insertion. These were thereafter inspected for the *cap2* phenotype.

If any of the At1g78500 and the At1g78510 genes is the *CAP2* locus, homozygous plants are not expected to be obtained, since an ovule carrying the maternal *cap2* allele will not develop into a functional seed after fertilization. However homozygous plants were found among both the SALK_036693 lines (carrying the T-DNA in the At1g78500 gene) and the SALK_126948 lines (carrying the T-DNA in the At1g78510 gene). In addition, the hemizygous and homozygous plants did not have any visible phenotype in seeds. These results indicated that the T-DNA insertions in the analyzed SALK lines were probably not located in the *CAP2* locus, excluding both the At1g78500 and the At1g78510 gene as the putative *CAP2* locus.

Taken together, both complementation analysis and analysis of the T-DNA insertion lines strongly indicate that none of the At1g78500 and the At1g78510 genes is *CAP2* locus. However there are several aspects that may be considered when discussing why the two rescue constructs did not complement the *cap2* mutation, and why the T-DNA insertion lines did not have any apparent phenotype. One possibility is that enhancer elements, located some distance upstream or downstream from the At1g78500 and At1g78510 genes could be responsible for their transcription. If this was the case, the promoter region included in each rescue construct was probably not sufficient to drive expression of its gene in the transgene plants. Consequently, none of the construct was able to rescue the *cap2* phenotype. Another possibility is that the rescue construct is not fully penetrant, meaning that the expression of the transgene is too low to be able to rescue the mutant phenotype. The lack of seed phenotype in the T-DNA insertion lines could be explained by the T-DNA insertions in the three SALK lines being alleles that do not have the same effect as the *cap2* alleles on the *CAP2* function, resulting in plants with normal seed phenotype. Moreover, the T-DNA

insertion in the ninth and fourteenth exon of the At1g78500 gene in the SALK_060682 and SALK_036693 line, respectively, may not affect and alter the essential part of the protein, producing the functional protein and no visible phenotype. In addition, the T-DNA insertions reported to be positioned in exons may actually be located in flanking introns and thus may not affect gene function if the affected intron is correctly spliced. Although genotyping PCR reactions produced fragments of expected lengths, sequencing of genomic T-DNA borders must be performed to exclude this possibility.

4.2.5 Concluding remark and further work

Despite above-mentioned possibilities that could explain why the identity of the *CAP2* gene was not revealed by the complementation experiment and the analysis of the T-DNA insertion lines, it is most likely that neither the At1g78500 nor the At1g78510 gene is the *CAP2* locus. This is rather surprising since the molecular mapping seemed to be successful. The number of recombinants was decreasing as we moved along the mapping interval. Recombinants that were lost as the mapping progressed along the interval, was always genotyped twice for the two markers flanking the recombination to confirm that the genotyping was correct. As long as there was a recombination event left from the both side of the interval, the genetic distance between the flanking markers and the *cap2* mutation was successfully narrowed down by developing new molecular markers.

To determine the phenotype of recombinants was a difficult task in itself. It was not always easy to determine if a plant was a true *cap2* mutant, since in some of the recombinant plants the frequency of the *cap2* seed phenotype was less than 50%, but more than 0-7%, which is the frequency in the wild-type. Variations in the phenotype among the recombinant plants might be due to different phenotype expression in Col and L*er* ecotype, since the mapping population was hybrid between these two ecotypes. The variable phenotype could also be due to various environmental conditions in the growth chambers.

Although all isolation of plant material and isolation DNA always was performed repeatedly for important recombinant samples, we cannot exclude contamination as a possible cause for our mapping result. This could be on the level of plant material, seeds and DNA, but could also include cross pollination and contamination of the mapping population.

The schematic illustration of the molecular mapping presented in Figure 27 shows that 23 recombinants were reduced to one recombinant within the interval between the nga111 and the CER446110 marker, which spanned 2.181 Mb on the centromere side. The number of recombinants was evenly reduced in the interval that spanned 699 kb on the telomere side, between the G17311 and the CER446080 marker, starting with eight recombinants in the G17311 marker, and further reducing to two in the CER446080 marker. However, the last recombination event at the telomere side occurred within this interval, between the CER433278 and the CER446080 marker. After that no recombinants were lost in the 47 kb long interval between the CER446080 and the CER446109 marker (Figure 27).

As long as the mapping interval was decreasing by reducing the number of recombinants from both sides, it was the indication that the mapping proceeded in right direction. Hence, the interval between the CER446110 on the centromere side and the CER433278 on the telomere side may be considered as a reliable and verified mapping interval. A new mapping population should be grown and genotyped for the CER446110 and the CER433278 markers first. In order to avoid uncertainty regarding the phenotyping of *cap2* mutants, one could only select recombinants that show wt phenotype and further genotype them for the other six markers located within that interval. This mapping would be easer to perform since the molecular markers are already established within the interval, and the doubts regarding the phenotype would be avoided by choosing only wt plants for further work. The interval between the CER446080 and the CER446109 marker spans 85 kb and contains 28 loci according to TAIR database. Among these 28 loci, *CAP2* candidate genes include a transcription factor and a ribosomal protein L13 that is targeted to chloroplast. L13 is reported to exhibit developmental arrest of embryos (www.arabidopsis.org) in homozygous mutants, and although recessive this feature should be examined.

Taken the power of the *Arabidopsis* insertion databases, another strategy to identify the molecular nature of *CAP2* is to screen T-DNA insertion lines for all 28 loci in the 85 Kb interval. Both the former and the latter strategies have been initiated, and will hopefully be successful in identifying the molecular nature of *capulet2*.

4.3 The molecular basis of maternal control in seed development

Epigenetic processes such as chromatin remodeling and DNA methylation affect imprinting of specific genes involved in maternal control of seed development. *MEA*, *FIS2* and *FIE* are all maternally expressed genes, required for seed viability. Mutations in these genes lead to maternal-effect seed abortion phenotype and initiate endosperm formation in the absence of fertilization. It was shown that the *FIS*-genes are targets of imprinting and that the genes they control in developing endosperm are also regulated by DNA methylation and chromatin remodeling genes. The *FIS* genes were proposed to act in a large chromatin remodeling complex that may repress transcription of number of target genes by establishing a repressive chromatin environment at these loci (Chaudhury and Berger, 2001; Chaudhury et al., 2001).

The *cap2* mutant has a gametophytic maternal-effect on embryo and endosperm development, where the endosperm development is blocked at a very early stage, while the embryos can develop to the early heart stage (Grini et al., 2002). The maternal-effect and aborted embryo in the *cap2* mutant resemble the *fis* mutants phenotype. Hence, the *CAP2* gene could be of the same molecular nature as the *FIS*-genes and their relatives.

In this thesis the *cap2* was mapped to an interval containing two genes, one involved in monoterpenoid and the other in triterpenoid biosynthesis. It could not be confirmed in the time course of this thesis that the *cap2* phenotype was caused by lesion in one of these loci. Furthermore, components in the monoterpene and triterpene pathways do not fit in with our present knowledge of the cloned maternal-effect genes as described previously. However, it is possible that either At1g78500 or At1g78510 could function in maternal control of seed development as described previously in this discussion.

The *LDC* analyzed in this work was originally selected due to its expression in the chalazal endosperm which is a typical feature of the *FIS* class of genes, and thus possibly play a role in similar processes. We could not confirm the chalazal expression pattern reported previously, but the gene is indeed expressed in a specific manner in the seed.

The *AtHD2C* analysis in this thesis is in line with the hypothesis that chromatin remodeling complexes play a major role in maternal control of seed development. The specific expression pattern in embryo supports the involvement of *AtHD2C* in seed development although no obvious phenotype could be found in the inspected plants. Double mutant analysis with other histone deacetylases or mutants such as *capulet2* or *fis* class mutants combined with forward and reverse genetic approaches will give insight into the roles of these genes in seed development and new knowledge on the molecular basis for maternal control of seed development.

References

Albert S, Despres B, Guilleminot J, Bechtold N, Pelletier G, Delseny M, Devic M (1999) The EMB 506 gene encodes a novel ankyrin repeat containing protein that is essential for the normal development of *Arabidopsis* embryos. Plant Journal **17:** 169-179

Alonso JM (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science **301:** 1849-1849

Ambion (2005) Membrane Transfer and Crosslinking for RNA. Technical Bulletin #169. http://www.ambion.com/techlib/tb/tb_169.html

Apuya NR, Yadegari R, Fischer RL, Harada JH, Goldberg RB (2002) RASPBERRY3 gene encodes a novel protein important for embryo development. Plant Physiology **129:** 691-705

Apuya NR, Yadegari R, Fischer RL, Harada JJ, Zimmerman JL, Goldberg RB (2001) The *Arabidopsis* embryo mutant schlepperless has a defect in the chaperonin-60 alpha gene. Plant Physiology **126:** 717-730

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796-815

Autran D, Huanca-Mamani W, Vielle-Calzada JP (2005) Genomic imprinting in plants: the epigenetic version of an Oedipus complex. Current Opinion in Plant Biology **8:** 19-25

Bahassi EM, O'Dea MH, Allali N, Messens J, Gellert M, Couturier M (1999) Interactions of CcdB with DNA gyrase - Inactivation of GyrA, poisoning of the gyrase-DNA complex, and the antidote action of CcdA. Journal of Biological Chemistry **274:** 10936-10944

Barghchi M (1995) High-frequency and efficient Agrobacterium-mediated transformation of Arabidopsis thaliana ecotypes "C24" and "Landsberg *erecta*" using *Agrobacterium tumefaciens*. Methods Mol Biol. **44:** 135-147

Beato M, Herrlich P, Schutz G (1995) Steroid-hormone receptors - many actors in search of a plot. Cell **83:** 851-857

Bechtold N, Ellis J, Pelletier G (1993) In-planta *agrobacterium*-mediated gene-transfer by infiltration of adult *Arabidopsis-thaliana* plants. Comptes Rendus De L Academie Des Sciences Serie III-Sciences De La Vie-Life Sciences **316:** 1194-1199

Bechtold N, Jaudeau B, Jolivet S, Maba B, Vezon D, Voisin R, Pelletier G (2000) The maternal chromosome set is the target of the T-DNA in the in planta transformation of *Arabidopsis thaliana*. Genetics **155:** 1875-1887

Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *arabidopsis*. Genomics **19:** 137-144

Berger F (1999) Endosperm development. Current Opinion in Plant Biology **2:** 28-32

Berger F (2003) Endosperm: the crossroad of seed development. Current Opinion in Plant Biology **6:** 42-50

Boisnard-Lorig C, Colon-Carmona A, Bauch W, Hodge S, Doerner P, Bancharel E, Dumas C, Haseloff J, Berger F (2001) Dynamic analyses of the expression of the HISTONE: YFP fusion protein in *Arabidopsis* show that syncytial endosperm is divided in mitotic domains. Plant Cell **13:** 495-509

Bonhomme S, Horlow C, Vezon D, de Laissardiere S, Guyon A, Ferault M, Marchand M, Bechtold N, Pelletier G (1998) T-DNA mediated disruption of essential

gametophytic genes in *Arabidopsis* unexpectedly rare and cannot be inferred from segregation distortion alone. Molecular and General Genetics **260:** 444-452

Bouchez D, Hofte H (1998) Functional genomics in plants. Plant Physiology **118:** 725-732

Brown RC, Lemmon BE, Nguyen H, Olsen OA (1999) Development of endosperm in *Arabidopsis thaliana*. Sexual Plant Reproduction **12:** 32-42

Burke CC, Wildung MR, Croteau R (1999) Geranyl diphosphate synthase: Cloning, expression, and characterization of this prenyltransferase as a heterodimer. Proceedings of the National Academy of Sciences of the United States of America **96:** 13062-13067

Butenko MA, Patterson SE, Grini PE, Stenvik GE, Amundsen SS, Mandal A, Aalen RB (2003) Influorescene deficient in abscission controls floral organ abscission in *Arabidopsis* and identifies a novel family of putative ligands in plants. Plant Cell **15:** 2296-2307

Cha RS, Thilly WG (1993) Specificity, efficiency, and fidelity of pcr. PCR-Methods and Applications **3:** S18-S29

Chang YL, Tao QZ, Scheuring C, Ding KJ, Meksem K, Zhang HB (2001) An integrated map of *Arabidopsis thaliana* for functional analysis of its genome sequence. Genetics **159:** 1231-1242

Chanvivattana Y, Bishopp A, Schubert D, Stock C, Moon YH, Sung ZR, Goodrich J (2004) Interaction of polycomb-group proteins controlling flowering in *Arabidopsis*. Development **131:** 5263-5276

Chappell J (2002) The genetics and molecular genetics of terpene and sterol origami. Current Opinion in Plant Biology **5:** 151-157

Chaudhury AM, Berger F (2001) Maternal control of seed development. Seminars in Cell & Developmental Biology **12:** 381-386

Chaudhury AM, Craig S, Dennis ES, Peacock WJ (1998) Ovule and embryo development, apomixis and fertilization. Current Opinion in Plant Biology **1:** 26-31

Chaudhury AM, Koltunow A, Payne T, Luo M, Tucker MR, Dennis ES, Peacock WJ (2001) Control of early seed development. Annual Review of Cell and Developmental Biology **17:** 677-699

Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, Peacock WJ (1997) Fertilization-independent seed development in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America **94:** 4223-4228

Chen YCS, McCormick S (1996) Sidecar pollen, an *Arabidopsis thaliana* male gametophytic mutant with aberrant cell divisions during pollen development. Development **122:** 3243-3253

Chibi F, Matilla AJ, Angosto T, Garrido D (1994) Changes in polyamine synthesis during anther development and pollen gernination in tobacco (*Nicotina tabacum*). Physiologia Plantarum **92:** 61-68

Choi YH, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. Cell **110:** 33-42

Christensen CA, Subramanian S, Drews GN (1998) Identification of gametophytic mutations affecting female gametophyte development in *Arabidopsis*. Developmental Biology **202:** 136-151

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant Journal **16:** 735-743

Clouse SD (2002) *Arabidopsis* mutants reveal multiple roles for sterols in plant development. Plant Cell **14:** 1995-2000

Colombo L, Franken J, VanderKrol AR, Wittich PE, Dons HJM, Angenent GC (1997) Downregulation of ovule-specific MADS box genes from petunia results in maternally controlled defects in seed development. Plant Cell **9:** 703-715

Desfeux C, Clough SJ, Bent AF (2000) Female reproductive tissues are the primary target of *Agrobacterium*-mediated transformation by the *Arabidopsis* floral-dip method. Plant Physiology **123:** 895-904

Despres B, Delseny M, Devic M (2001) Partial complementation of embryo defective mutations: a general strategy to elucidate gene function. Plant Cell **27:** 149-159

Diener AC, Li HX, Zhou WX, Whoriskey WJ, Nes WD, Fink GR (2000) STEROL METHYLTRANSFERASE 1 controls the level of cholesterol in plants. Plant Cell **12:** 853-870

Drews GN, Lee D, Christensen GA (1998) Genetic analysis of female gametophyte development and function. Plant Cell **10:** 5-17

Drews GN, Yadegari R (2002) Development and function of the angiosperm female gametophyte. Annual Review of Genetics **36:** 99-124

Feldmann KA, Coury DA, Christianson ML (1997) Exceptional segregation of a selectable marker (Kan(R)) in *Arabidopsis* identifies genes important for gametophytic growth and development. Genetics **147:** 1411-1422

Felker FC, Peterson DM, Nelson OE (1985) Anatomy of immature grains of 8 maternal effect shrunken endosperm barley mutants. American Journal of Botany **72:** 248-256

Fox MA, Whitesell JK (1997) Organic Chemistry, second edition. Jones and Bartlett Publishers, Sudbury, Massachusetts

Francis NJ, Kingston RE (2001) Mechanisms of transcriptional memory. Nature Reviews Molecular Cell Biology **2:** 409-421

Galau GA, Hughes DW, Dure L (1986) Abscisic-acid induction of cloned cotton late embryogenesis-abundant (LEA) messenger-RNAs. Plant Molecular Biology **7:** 155-170

Gamarnik A, Frydman RB (1991) Cadaverine, an essential diamine for the normal root development of germinating soybean (*Glycine Max*) seeds. Plant Physiology **97:** 778-785

Gehring M, Choi Y, Fischer RL (2004) Imprinting and seed development. Plant Cell **16:** S203-S213

Goldberg RB, Depaiva G, Yadegari R (1994) Plant embryogenesis - zygote to seed. Science **266:** 605-614

Golden TA, Schauer SE, Lang JD, Pien S, Mushegian AR, Grossniklaus U, Meinke DW, Ray A (2002) SHORT INTEGUMENTS1/SUSPENSOR1/CARPEL FACTORY, a Dicer homolog, is a maternal effect gene required for embryo development in *Arabidopsis*. Plant Physiology **130:** 808-822

Grini PE, Jurgens G, Hulskamp M (2002) Embryo and endosperm development is disrupted in the female gametophytic capulet mutants of *Arabidopsis*. Genetics **162:** 1911-1925

Grini PE, Schnittger A, Schwarz H, Zimmermann I, Schwab B, Jurgens G, Hulskamp M (1999) Isolation of ethyl methanesulfonate-induced gametophytic mutants in *Arabidopsis thaliana* by a segregation distortion assay using the multimarker chromosome 1. Genetics **151:** 849-863

Grossniklaus U, Schneitz K (1998) The molecular and genetic basis of ovule and megagametophyte development. Seminars in Cell & Developmental Biology **9:** 227-238

Grossniklaus U, Spillane C, Page DR, Köhler C (2001) Genomic imprinting and seed development: endosperm formation with and without sex. Current Opinion in Plant Biology **4:** 21-27

Grossniklaus U, Vielle-Calzada JP, Hoeppner MA, Gagliano WB (1998) Maternal control of embryogenesis by medea, a Polycomb group gene in *Arabidopsis*. Science **280:** 446-450

Guitton AE, Berger F (2005) Control of reproduction by Polycomb Group complexes in animals and plants. International Journal of Development Biology **49:** 707-716

Guitton AE, Page DR, Chambrier P, Lionnet C, Faure JE, Grossniklaus U, Berger F (2004) Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in *Arabidopsis thaliana*. Development **131:** 2971-2981

Hartwell L, Hood L, Goldberg ML, Reynolds AE, Silver LM, Veres R (2000) Genetics; From Genes to Genomes. McGraw-Hill Higher Education

Hennig L, Taranto P, Walser M, Schonrock N, Gruissem W (2003) *Arabidopsis* MSI1 is required for epigenetic maintenance of reproductive development. Development **130:** 2555-2565

Heslop-Harrison J, Heslop-Harrison Y (1986) Pollen-tube chemotropism: fact or delusion? *In* M Cresti, D Romano, eds, Biology of Reproduction and Cell Motility in Plants and Animals. University of Siena Press, Siena

Hill JP, Lord EM (1987) Dynamics of pollen tube growth in the wild radish *Raphanus raphanistrum* (*Brassicaceae*). II. morphology, cytochemistry and ultrastructure of transmitting tissues, and path of pollen tube growth. American Journal of Botany **74:** 988-997

Howden R, Park SK, Moore JM, Orme J, Grossniklaus U, Twell D (1998) Selection of T-DNA-tagged male and female gametophytic mutants by segregation distortion in Arabidopsis. Genetics **149:** 621-631

Husselstein-Muller T, Schaller H, Benveniste P (2001) Molecular cloning and expression in yeast of 2,3-oxidosqualene-triterpenoid cyclases from *Arabidopsis thaliana*. Plant Molecular Biology **45:** 75-92

Ingouff M, Gerald JNF, Guerin C, Robert H, Sørensen MB, Van Damme D, Geelen D, Blanchoin L, Berger F (2005) Plant formin AtFH5 is an evolutionarily conserved actin nucleator involved in cytokinesis. Nature Cell Biology **7:** 374-U334

Invitrogen (2003) Gateway Technology - A universal technology to clone DNA sequences for functional analysis and expression in multiple systems. Invitrogen Life Technologies, 60 pp. Catalog nos. 12535-019 and 12535-027 Version D.

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002) *Arabidopsis* map-based cloning in the post-genome era. Plant Physiology **129:** 440-450

Jang JC, Fujioka S, Tasaka M, Seto H, Takatsuto S, Ishii A, Aida M, Yoshida S, Sheen J (2000) A critical role of sterols in embryonic patterning and meristem programming revealed by the fackel mutants of *Arabidopsis thaliana*. Genes and Development **14:** 1485-1497

Jefferson RA (1989) The GUS reporter gene system. Nature **342:** 837-838

Jenuwein T, Allis CD (2001) Translating the histone code. Science **293:** 1074-1080

Jürgens G, Mayer U, Busch M, Lukowitz W, Laux T (1995) Pattern-formation in the *Arabidopsis* embryo - a genetic perspective. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **350:** 19-25

Kasahara H, Hanada A, Kuzuyama T, Takagi M, Kamiya Y, Yamaguchi S (2002) Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of gibberellins in *Arabidopsis*. Journal of Biological Chemistry **277:** 45188-45194

Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL (1999) Imprinting of the MEDEA polycomb gene in the *Arabidopsis* endosperm. Plant Cell **11:** 1945-1952

Konieczny A, Ausubel FM (1993) A procedure for mapping *Arabidopsis* mutations using codominant ecotype-specific PCR-based markers. Plant Journal **4:** 403-410

Köhler C, Hennig L, Bouveret R, Gheyselinck J, Grossniklaus U, Gruissem W (2003a) *Arabidopsis* MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. EMBO Journal **22:** 4804-4814

Köhler C, Hennig L, Spillane C, Pien S, Gruissem W, Grossniklaus U (2003b) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. Genes & Development **17:** 1540-1553

Köhler C, Page DR, Gagliardini V, Grossniklaus U (2005) The *Arabidopsis thaliana* MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. Nature Genetics **37:** 28-30

Lalanne E, Michaelidis C, Moore JM, Gagliano W, Johnson A, Patel R, Howden R, Vielle-Calzada JP, Grossniklaus U, Twell D (2004) Analysis of transposon insertion mutants highlights the diversity of mechanisms underlying male progamic development in *Arabidopsis*. Genetics **167:** 1975-1986

Landy A (1989) Dynamic, structural, and regulatory aspects of Lambda site-specific recombination. Annual Review of Biochemistry **58:** 913-949

Laule O, Furholz A, Chang HS, Zhu T, Wang X, Heifetz PB, Gruissem W, Lange BM (2003) Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America **100:** 6866-6871

Laux T, Jürgens G (1997) Embryogenesis: A new start in life. Plant Cell **9:** 989-1000

Laux T, Wurschum T, Breuninger H (2004) Genetic regulation of embryonic pattern formation. Plant Cell **16:** S190-S202

Lichtenthaler HK, Schwender J, Disch A, Rohmer M (1997) Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway. Febs Letters **400:** 271-274

Linz U, Delleing U, Rubsamenwaigmann H (1990) Systematic studies on parameters influencing the performance of the polymerase chain-reaction. Journal of Clinical Chemistry and Clinical Biochemistry **28:** 5-13

Lohe AR, Chaudhury A (2002) Genetic and epigenetic processes in seed development. Current Opinion in Plant Biology **5:** 19-25

Lojda Z (1970) Indigogenic methods for glycosidases. 2. an improved method for beta-d-galactosidase and its application to localization studies of enzymes in intestine and in other tissues. Histochemie **23:** 266-288

Lukowitz W, Gillmor CS, Scheible WR (2000) Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you. Plant Physiology **123:** 795-805

Luo M, Bilodeau P, Dennis ES, Peacock WJ, Chaudhury A (2000) Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing *Arabidopsis* seeds. Proceedings of the National Academy of Sciences of the United States of America **97:** 10637-10642

Luo M, Bilodeau P, Koltunow A, Dennis ES, Peacock WJ, Chaudhury AM (1999) Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America **96:** 296-301

Mansfield SG, Briarty LG (1991) Early embryogenesis in *Arabidopsis-Thaliana*. 2. The developing embryo. Canadian Journal of Botany **69:** 461-476

McCormick S (1993) Male gametophyte development. Plant Cell **5:** 1265-1275

Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M (1998) *Arabidopsis thaliana*: A model plant for genome analysis. Science **282:** 662-682

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiologia Plantarum **15:** 473-&

Ohad N, Margossian L, Hsu YC, Williams C, Repetti P, Fischer RL (1996) A mutation that allows endosperm development without fertilization. Proceedings of the National Academy of Sciences of the United States of America **93:** 5319-5324

Ohad N, Yadegari R, Margossian L, Hannon M, Michaeli D, Harada JJ, Goldberg RB, Fischer RL (1999) Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. Plant Cell **11:** 407-415

Olsen OA (2004) Nuclear endosperm development in cereals and *Arabidopsis thaliana*. Plant Cell **16:** S214-S227

Østergaard L, Yanofsky M (2004) Establishing gene function by mutagenesis in *Arabidopsis thaliana*. Plant Journal **39:** 682-696

Page DR, Grossniklaus L (2002) The art and design of genetic screens: *Arabidopsis thaliana*. Nature Reviews Genetics **3:** 124-136

Pagnussat GC, Yu HJ, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie LF, Ye D, Sundaresan V (2005) Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. Development **132:** 603-614

Pirrotta V (1998) Polycombing the genome: PcG, trxG, and chromatin silencing. Cell **93:** 333-336

Ray S, Golden T, Ray A (1996) Maternal effects of the short integument mutation on embryo development in *Arabidopsis*. Developmental Biology **180:** 365-369

Redei GP (1964) Non-Mendelian megagametogenesis in *Arabidopsis*. Genetics **51:** 857-872

Regan SM, Moffatt BA (1990) Cytochemical analysis of pollen development in wild-type *Arabidopsis* and a male-sterile mutant. Plant Cell **2:** 877-889

Reiser L, Fischer RL (1993) The ovule and the embryo sac. Plant Cell **5:** 1291-1301

Roche (1999) PCR Application Manual, 2nd edition. Roche Diagnostics GmbH, Mannheim. http://www.roche-applied-science.com/PROD_INF/MANUALS/pcr_man/start.html

Rotman N, Durbarry A, Wardle A, Yang WC, Chaboud A, Faure JE, Berger F, Twell D (2005) A novel class of MYB factors controls sperm-cell formation in plants. Current Biology **15:** 244-248

Roux KH (1995) Optimization and troubleshooting in PCR. PCR-Methods and Applications **4:** S185-S194

Sambrook J, Russel DW (2001) Molecular cloning: A laboratory manual, third edition. Cold Spring Harbor Laboratory Press

Sanger F, Niklen S, Coulson A (1977) DNA sequencing with chain terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America **74:** 5463-5467

Schauer SE, Jacobsen SE, Meinke DW, Ray A (2002) DICER-LIKE1: blind men and elephants in *Arabidopsis* development. Trends in Plant Science **7:** 487-491

Schneitz K, Hullskamp M, Pruitt RE (1995) Wild-type ovule development in *Arabidopsis thaliana*: a light microscope study of cleared whole-mount tissue. Plant Journal **7:** 731-749

Schrick K, Mayer U, Horrichs A, Kuhnt C, Bellini C, Dangl J, Schmidt J, Jurgens G (2000) FACKEL is a sterol C-14 reductase required for organized cell division and expansion in *Arabidopsis* embryogenesis. Genes and Development **14:** 1471-1484

Schubert D, Goodrich J (2003) Plant epigenetics: MEDEA's children take centre stage. Current Biology **13:** R638-R640

Sharp PA, Sugden B, Sambrook J (1973) Detection of 2 restriction endonuclease activities in *Haemophilus-Parainfluenzae* using analytical agarose-ethidium bromide electrophoresis. Biochemistry **12:** 3055-3063

Somerville C, Koornneef M (2002) Timeline - A fortunate choice: the history of *Arabidopsis* as a model plant. Nature Reviews Genetics **3:** 883-889

Spillane C, MacDougall C, Stock C, Köhler C, Vielle-Calzada JP, Nunes SM, Grossniklaus U, Goodrich J (2000) Interaction of the *Arabidopsis* Polycomb group proteins FIE and MEA mediates their common phenotypes. Current Biology **10:** 1535-1538

Springer PS (2000) Gene traps: Tools for plant development and genomics. Plant Cell **12:** 1007-1020

Springer PS, Holding DR, Groover A, Yordan C, Martienssen RA (2000) The essential Mcm7 protein PROLIFERA is localized to the nucleus of dividing cells during the G(1) phase and is required maternally for early *Arabidopsis* development. Development **127:** 1815-1822

Springer PS, McCombie WR, Sundaresan V, Martienssen RA (1995) Gene trap tagging of prolifera, an essential mcm2-3-5-like gene in *Arabidopsis*. Science **268:** 877-880

Stangeland B, Nestestog R, Grini PE, Skrbo N, Berg A, Salehian Z, Mandal A, Aalen R (2005) Molecular analysis of *Arabidopsis* endosperm and embryo promoter trap lines: Reporter-gene expression can result from T-DNA insertions in antisense orientation, in introns and in intergenic regions in addition to sense ionsertion at the 5' end of genes. Journal of Experimental Botany **In press**

Stangeland B, Salehian Z, Aalen R, Mandal A, Olsen OA (2003) Isolation of GUS marker lines for genes expressed in Arabidopsis endosperm, embryo and maternal tissues. Journal of Experimental Botany **54:** 279-290

Steinmetz LM, Davis RW (2004) Maximizing the potential of functional genomics. Nature Review Genetics **5:** 190-201

Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature **403:** 41-45

Sørensen MB, Chaudhury AM, Robert H, Bancharel E, Berger F (2001) Polycomb group genes control pattern formation in plant seed. Current Biology **11:** 277-281

Sørensen MB, Mayer U, Lukowitz W, Robert H, Chambrier P, Jurgens G, Somerville C, Lepiniec L, Berger F (2002) Cellularisation in the endosperm of *Arabidopsis thaliana* is coupled to mitosis and shares multiple components with cytokinesis. Development **129:** 5567-5576

Tanksley SD, Zamir D, Rick CM (1981) Evidence for extencive overlap of sporophytic and gametophytic gene expression in *Lycopersicon esculentum*. Science **213:** 453-455

Tie F, Furuyama T, Prasad-Sinha J, Jane E, Harte PJ (2001) The *Drosophila* Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. Development **128:** 275-286

Ullmann A, Jacob F, Monod J (1967) Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of beta-galactosidase structural gene of Escherichia Coli. Journal of Molecular Biology **24:** 339-&

Uwer U, Willmitzer T, Altmann T (1998) Inactivation of a glycyl-tRNA synthetase leads to an arrest in plant embryo development. Plant Cell **10:** 1277-1294

Vielle-Calzada JP, Thomas J, Spillane C, Coluccio A, Hoeppner MA, Grossniklaus U (1999) Maintenance of genomic imprinting at the Arabidopsis medea locus requires zygotic DDM1 activity. Genes & Development **13:** 2971-2982

Vinkenoog R, Spielman M, Adams S, Fischer RL, Dickinson HG, Scott RJ (2000) Hypomethylation promotes autonomous endosperm development and rescues postfertilization lethality in fie mutants. Plant Cell **12:** 2271-2282

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in *Arabidopsis*. Science **290:** 2114-2117

Wu HM, Wang H, Cheung AY (1995) A pollen tube growth stimulatory glycoprotein is deglycosylated by pollen tubes and displays a glycosylation gradient in the flower. Cell **82:** 395-403

Xiao WY, Gehring M, Choi Y, Margossian L, Pu H, Harada JJ, Goldberg RB, Pennell RI, Fischer RL (2003) Imprinting of the MEA polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. Developmental Cell **5:** 891-901

Yadegari R, Drews GN (2004) Female gametophyte development. Plant Cell **16:** S133-S141

Yadegari R, Kinoshita T, Lotan O, Cohen G, Katz A, Choi Y, Katz A, Nakashima K, Harada JJ, Goldberg RB, Fischer RL, Ohad N (2000) Mutations in the FIE and MEA genes that encode interacting polycomb proteins cause parent-of-origin effects on seed development by distinct mechanisms. Plant Cell **12:** 2367-2381

Appendix 1

Primers for amplification of genomic DNA

Product length, annealing temperature, elongation time and number of PCR cycles used for different primer pairs in different experiments.

Primers	Experiment	Product length (bp)	Annealing temp. (°C)	Elongation time (sec)	Number of cycles
nga111 forward nga111 reverse	Mapping of *cap2*	Col: 128 L*er*: 162	60	45	40
ADH forward ADH reverse	Mapping of *cap2*	1291	58	120	33
SNP253/5 SNP253/6	Mapping of *cap2*	549	60	60	30
SNP10847 primer1forward SNP10847 primer1reverse	Mapping of *cap2*	235	57	45	35
SNP10870 primer2 forward SNP10870 primer2 reverse	Mapping of *cap2*	453	58	45	30
SNP10876 primer5 forward SNP10876 primer5 reverse	Mapping of *cap2*	418	58	45	30
GL2 forward GL2 reverse	Mapping of *cap2*	1800	60	120	33
G17311 forward G17311 reverse	Mapping of *cap2*	1600	60	120	33
CER446094F CER446094R	Mapping of *cap2*	811	63	75	30
CER446080F CER446080R	Mapping of *cap2*	526	61	60	30
CER433278F CER433278R	Mapping of *cap2*	778	63	75	30
CER433242F CER433242R	Mapping of *cap2*	593	61	60	30
CER446110 forward2 CER446110 reverse2	Mapping of *cap2*	558	60	60	30
CER446109 F3-ny CER446109 R3-ny	Mapping of *cap2*	478	64	45	33
CER446106 F1-ny CER446106 13-ny	Mapping of *cap2*	1048	62	90	30

CER446104 forward2 CER446104 reverse2	Mapping of *cap2*	708	58	75	30
CER446100 forward1 CER446100 reverse1	Mapping of *cap2*	452	60	45	30
760 nestedprimer 1 760 nestedprimer 2	Amplification of *pLDC* for construction of the p*LDC::GUS* transgene	1550	58	120	30
760 attB1 760 attB2	Amplification of *pLDC* for construction of the p*LDC::GUS* transgene	1509	63	120	30
Prom 760-kontr 1F Prom 760-kontr 1R	Control PCR confirming the presence of *pLDC* fragment in the expression vector	596	54	60	33
Prom 760-kontr 2F Prom 760-kontr 2R	Control PCR confirming the presence of the *pLDC* fragment in the expression vector	757	54	60	33
760:gus1 sense (*pLDC*) 760:gus1/2 antisense (*GUS*)	Control PCR confirming the presence of the *pLDC::GUS* transgene in the T_1 transformants	~1430	56	120	30
760:gus2 sense (*pLDC*) 760:gus1/2 antisense	Control PCR confirming the presence of the *pLDC::GUS* transgene in the T_1 transformants	~1010	56	90	30
Gene16 F-attB1 Gene16 R-attB2	Amplification of At1g78500 rescue fragment	5537	60	345	30
Gen15 F-attB1 Gene15 R-attB2	Amplification of At1g78510 rescue fragment	3600	60	240	30
Gen16-kontr 1F Gen16-kontr 1R	Control PCR confirming the presence of the At1g78500 rescue fragment in the expression vector	863	57	30	33

Gen16-kontr 2F Gen16-kontr 2R	Control PCR confirming the presence of the At1g78500 rescue fragment in the expression vector	778	57	30	33
Gen15-kontr 1F Gen15-kontr 1R	Control PCR confirming the presence of the At1g78510 rescue fragment in the expression vector	803	56	30	33
Gen15-kontr 2F Gen15-kontr 2R	Control PCR confirming the presence of the At1g78510 rescue fragment in the expression vector	621	54	30	33
Gen16-kontr 1F pipp 2314 (Hyg)	Control PCR confirming the presence of the At1g78500 rescue fragment in the T_1 transformants	~1300	57	90	33
Gen 15-kontr 1F pipp 2314 (Hyg)	Control PCR confirming the presence of the At1g78510 rescue fragment in the T_1 transformants	~1200	57	90	33
HPT2-774R HPT 21F	Control PCR confirming the presence of the hygromycin gene in the T_1 transformants	~900	58	60	33
146-760REV 150-760DIR	Amplification of *LDC* probe	569	58	45	33
HDAC forward HDAC reverse	Amplification of *AtHD2C* probe	566	60	45	33
B-NS3 -F NS4-R	Amplification of 18S rRNA probe	600	58	45	33
SALK_060682 LP SALK_060682 RP	Genotyping of the SALK_060682 line	932	57	60	33

SALK_036693 LP SALK_036693 RP	Genotyping of the SALK_036693 line	927	57	60	33
SALK_126948 S LP SALK_126948 AS RP	Genotyping of the SALK_126948 line	1077	59	75	33
SALK_060682 RP LBa1 (PAULext)	Genotyping of the SALK_060682 line	~750	57	60	33
SALK_036693 RP LBa1 (PAULext)	Genotyping of the SALK_036693 line	~800	57	60	33
SALK_126948 AS RP LBa1 (PAULext)	Genotyping of the SALK_126948 line	~700	58	60	33
HD2C RP exon HD2C LP exon	Genotyping of the SALK_039784 line	954	60	60	33
HD2C RP exon LBa1 (PAULext)	Genotyping of the SALK_039784 line	~700	58	60	33

Appendix 2

Primers for RT-PCR
Product length, annealing temperature, elongation time and number of cycles used for different primer pairs in the RT-PCR.

Primer pair	RT-PCR	Product length (bp)	Annealing temp. (°C)	Extension time	Number of cycles
RT-PCR Gen16F RT-PCR Gen16R	At1g78500	1297	55	75	30
RT-PCR Gen15F RT-PCR Gen15R	At1g78510	831	57	60	30
146-760REV 150-760DIR	*LDC*	569	58	45	25
act2int2_sense act2int2_antisense	*Actin*	~250	58	30	25

Appendix 3

Primer name, primer sequence and Tm

Primer	Sequence 5'→3' (attB sequences = bold)	Tm (°C)
146-760REV	GGACTTAGCCCTATTTGCTCCATCTCCCAGC	61
150-760DIR	GCTGGTAATAAAGTCAGTTACAAAGATGCTGCTATCG	58
760 attB1	**GGGGACAAGTTTGTACAAAAAAGCAGGCT** ATGAACCAAATGTGGAATTG	63
760 attB2	**GGGGACCACTTTGTACAAGAAAGCTGGGT** ACACAAAGTTTTGTTTTAAG	63
760 nestedprimer 1	GTACCCATTAGGCTATTATAC	43
760 nestedprimer 2	CTCTTCACACAAAGTTTTGT	41
760:gus1 sense (*pLDC*)	CCCAAAAAATCGTGTTTATATCAAGTGC	50
760:gus1/2 antisense (*GUS*)	GTTTTCGTCGGTAATCACCATTCCC	53
760:gus2 sense (*pLDC*)	TCTTTAAACTGGCTAGGTGATAAGTCGG	53
act2int2_antisense	CCGCAAGATCAAGACGAAGGATAGC	54
act2int2_sense	CCCTGAGGAGCACCCAGTTCTACTC	58
ADH forward	AAAAATGGCAACACTTTGAC	41
ADH reverse	GCGTGACCATCAAGACTAAT	45
attL1	TCGCGTTAACGCTAGCATGGATCTC	54
attL2	GTAACATCAGAGATTTTGAGACAC	47
B-NS3-F	GCAAGGTCGTGTGCCAGCAGCC	57
CER433242F	AGATACCGTTACTATTCTTGG	43
CER433242R	TTTGCAATATTTGACACAGG	41
CER433278F	TCTCGGAGGACAATCTTAGAAGCG	52
CER433278R	TCTGTACCTGCGATGTCGCG	51
CER446080F	AAGAGTCAAATCACTTTCGC	43
CER446080R	TCAGGTATCTTACAAATGGC	43
CER446094F	CACATCTTGTTTCCTGTTTAAGGC	49
CER446094R	CGTGGACAACACATAATGCG	47
CER446100 forward1	TTTTCTGTTGACAGGTGATGTTAGG	49
CER446100 reverse1	ATCCAACTCGCTGACCAAGC	49
CER446104 forward2	GAGAAGAACTTCACAACTATTCC	47
CER446104 reverse2	ACAAATCTTCCATTTCTTCG	41
CER446106 F1-ny	CGTGTATTTGAGTGGTTTCTTTAGC	49
CER446106 R1-ny	TCCAATGCCCACTTTGTACG	47

CER446109 F3-ny	TGAGCTAGTAGCTGTTGATCTGC	50
CER446109 R3-ny	CTCTTCGCATGTCACTCTCG	49
CER446110 forward2	TCTGCTTCTACTCCCAGAATCC	50
CER446110 reverse2	AGTGGAGATAGGAGATGCGG	49
G17311 forward	TACATCGCCGTGAGTCTGTC	49
G17311 reverse	GCGAGTCCTGCGTAGATCCA	51
Gen15-kontr 1F	TTGATTGTCGTGTTCTTTTGAGTGG	49
Gen15-kontr 1R	CTTGAGGCTAAGATTTCACC	45
Gen15-kontr 2F	TAAAGAATCTACAGTGTCTACC	44
Gen15-kontr 2R	ATTGATGGTTCTTATTATGG	38
Gen16-kontr 1F	GCCATTTTCAGCAGAGTTGAGC	50
Gen16-kontr 1R	TGGATAATGGCGATTTTCCG	45
Gen16-kontr 2F	TATCGGAAAGGCATAACATTAAGC	47
Gen16-kontr 2R	AATCTCATTTCCTGCTTCCG	45
Gene15 F-attB1	**GGGGACAAGTTTGTACAAAAAAGCAGGCT** TCGCATGTTCTCATGTTGATTGTCG	66
Gene15 R-attB2	**GGGGACCACTTTGTACAAGAAAGCTGGGT** AACTGTGAGCCGATTCGTAAAGCG	68
Gene16 F-attB1	**GGGGACAAGTTTGTACAAAAAAGCAGGCT** TTTAGTGGCTGCGAGCACAAAAG	67
Gene16 R-attB2	**GGGGACCACTTTGTACAAGAAAGCTGGGT** CCGAGATCATTCGTTCACCATTCC	68
GL2 forward	GTAGCTCTCTCACAACTGTG	47
GL2 reverse	ATAGGAAACTGGGACTCTGG	47
HD2C LP exon	AAGAGAAGGTTACTGCTGAATCTGACAGTGAGG	58
HD2C RP exon	CCATTTCCGAAGTAAAGGTTCTGGTGC	55
HDAC forward	TAGAACCTTCTCCTTCTCCTTCGACGCATCGT	59
HDAC reverse	TGGCAACTGAAAGTTCACCTGTTTCGCA	55
HPT 21F	CGCGACGTCTGTCGAGAAGTTTC	54
HPT2-774R	TCCTGCAAGCTCCGGATGCC	53
LBa1 (PAULext)	TGGTTCACGTAGTGGGCCATCG	53
nga111 forward	TGTTTTTTAGGACAAATGGCG	43
nga111 reverse	CTCCAGTTGGAAGCTAAAGGG	49
NS4-R	CTTCCGTCAATTCCTTTAAG	43
pipp 2314 (Hyg)	CAGCTATTTACCCGCAGGAC	49
Prom 760-kontr 1F	AAAATACAATTTAAGAAGAGAGG	41
Prom 760-kontr 1R	ATGCACTTGATATAAACACG	41
Prom 760-kontr 2F	ATCGTGTTTATATCAAGTGC	41

Prom 760-kontr 2R	TATATACCTAGCTCATACGC	43
RT-PCR Gen15F	GATGACGTCATGTCGGAATATAGATTTAGG	54
RT-PCR Gen15R	TCAGGCTCAACTCTGCTGAAAATGG	53
RT-PCR Gen16F	AGAAGAGGGATACTCTATTTCAAGG	49
RT-PCR Gen16R	GAAAATAGAGAAGTAAATTAACGGC	46
SALK_036693 LP	CACAGAAGGGAAGAGGTTGAAAAGTTA	52
SALK_036693 RP	AATCTACAGTGTCTACCTCGAAGTGGCT	55
SALK_060682 LP	ATATAGCCGTTTCATATGCTTGCTTGTT	50
SALK_060682 RP	TGATGTGGTAAAGATAGTACTCATGCTCG	54
SALK_126948 AS RP	ATGATCTCATGTGGGCTTGTCTCGTCC	56
SALK_126948 S LP	AAACGCCGGTGTTCTGTTGTTAGTTCC	55
SNP10847 primer1forward	ACATATCGTCAGGAAGTTG	42
SNP10847 primer1reverse	GAAAAACGGGAAGGTAAT	38
SNP10870 primer2 forward	GGTTGCGCCAATTATCGTATTC	48
SNP10870 primer2 reverse	CCCTAAAAATTCACCCCCAATC	48
SNP10876 primer5 forward	TCGTGACACCAAGTAATGTCCTC	50
SNP10876 primer5 reverse	GTGGTGATTGGTGAACCCAG	49
SNP253/5	CAACAATGGCGTCTGGGC	47
SNP253/6	AACAAACCACCATCCACGG	46