Functional analysis of developmental control genes in California poppy (*Eschscholzia californica* Cham.)

Funktionelle Analyse von Entwicklungsgenen im Kalifornischen Mohn (*Eschscholzia californica* Cham.)

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

Dem Fachbereich Biologie/Chemie

der Universität Bremen

vorgelegt von

Matthias Lange

aus Bremen

Bremen 2010

Reviewer: **Prof. Dr. Annette Becker**
Evolutionary Developmental Genetics
University of Bremen

Prof. Dr. Uwe Nehls
Section of Botany, Department of Ecology
University of Bremen

Examiner: **Prof. Dr. Barbara Reinhold-Hurek**
Department of Micobe-Plant Interactions
University of Bremen

Dr. Andrea Krause
Department of Micobe-Plant Interactions
University of Bremen

Staff: **Sabrina Lange**
University of Bremen

Student: **Nils Engelke**
University of Bremen

"I am a firm believer,

that without speculation there is

no good and original observation."

Charles Darwin,

Letter to A. R. Wallace,

December 1857

Danksagungen

Ich möchte mich bei Prof. Dr. Annette Becker dafür bedanken, dass Sie es mir ermöglicht hat, an einem spannenden Thema zu arbeiten! Ich möchte Ihr außerdem meinen Dank aussprechen für die vielseitige Unterstützung meiner Arbeit sowie für die Förderung meiner wissenschaftlichen Entwicklung.

Prof. Dr. Uwe Nehls möchte ich danken für die Übernahme des Zweitgutachtens meiner Dissertation. Ebenfalls Dank geht an Frau Prof. Dr. Reinhold-Hurek, sowie an Dr. Andrea Krause für die Einwilligung bei meinem Dissertationskolloquium als Mitglied des Prüfungsausschusses teil zu nehmen.

Nochmals danken möchte ich Prof. Dr. Reinhold-Hurek für die Möglichkeiten in Ihrem Labor zu arbeiten. Hier sei auch Janina Oetjen noch mal im Besonderen gedankt für Ihre zuverlässige Hilfe bei den Experimenten.

Ein besonderer Dank geht an die Mitglieder der AG Becker für all die schönen Momente während der langjährigen Arbeit im Labor, sowie während unserer schönen Betriebsausflüge. Mein Dank geht an Svetlana, Sabrina, Aravinda und Robert, die die Zeit hier mit mir geteilt haben, und immer für ein helfendes oder aufmunterndes Gespräch bereit waren. Ich danke auch Anna, Dawit und den vielen Studenten, die hier bei uns im Labor waren und zur guten Atmosphäre beigetragen haben. Vielen Dank an Tina, im Besonderen für die Korrekturen an meiner schriftlichen Dissertation und diversen Manuskripten.

Den unzähligen netten Mitarbeitern an der Universität Bremen möchte ich hiermit meinen Dank aussprechen, denn alle zusammen ermöglichen es letzten Endes, dass wir unsere Arbeit verrichten können.

Ich möchte im besonderen Svetlana danken, dass Sie alle Freude und alles Leid in dieser Zeit mitgetragen hat.

Vielen Dank auch an meine Eltern, die mir immer eine Stütze im Leben wahren und mir nicht nur einmal aus einer schwierigen Situation geholfen haben.

Table of contents

Abbreviations

3-AT	3-amino-1,2,4-triazole
Aa	Asparagus asparagoides
ACT	ACTIN
A	adenine
Ade	adenine
AEKI	Atomic Energy Research Institute
ANA	Amborellales, Nymphaeales, Austrobaileyales
Af	Aquilegia formosa
AG	AGAMOUS
AGL	AGAMOUS-like
Amb	Amborella trichopoda
AP	APETALA
APG	Angiosperm Phylogeny Group
At	Arabidopsis thaliana
BAC	bacterial artificial chromosome
BNQ	BANQUO
BiFC	bimolecular fluorescence complementation
bp	base pairs
ca	carpels
CArG-box	5'-CC(A/T)$_6$GG-3' consensus sequence
cDNA	complementary DNA
cfc	cellulose fortified cells
CLV	CLAVATA
cox3	CYTOCHROME OXIDASE 3
CRC	CRABS CLAW
DEF	DEFICIENS
DL	DROOPING LEAF
DSB	double strand break
Ec	Eschscholzia californica

EcAuxRF1	Eschscholzia californica auxin response factor 1
EMSA	Electrophoretic mobility shift assay
EREBP	ethylene responsive element binding protein
ESca	Eschscholzia californica
euAP3 motif	eudicot APETALA3 motif
FAE	formaldehyde-acetic acid-ethanol solution
FAR	FARINELLI
FIL	FILAMENTOUS FLOWER
g	gynoecium
GGM	Gnetum gnemon MADS
GLO	GLOBOSA
GTR	general time reversal
Gy	gray
His	histidine
HMG-box	high mobility group-box
HMW	high molecular weight
HOX	homeobox transcription factor
ig	inner gynoecium
INO	INNER NO OUTER
Kb	kilobase pairs
LB medium	lysogeny broth medium
Leu	leucine
LEU	LEUNIG
lpc	lignified parenchymatic cells
lr	lateral ridges
MADS-box	‘MCM1-AGAMOUS-DEFICIENS-SRF’-box transcription factor
Mb	megabase pairs
MIKC	MADS, Intervening, Keratin-like and C-terminal domain structure
min	minutes
miRNA	microRNA
mr	middle ridges

MYA	million years ago
og	outer gynoecium
ONPG	*ortho*-Nitrophenyl-β-galactoside
ORF	open reading frame
OsMADS	Oryza sativa MADS
ov	ovule
p	placental tissue
paleoAP3 motif	paleo APETALA3 motif
PCR	polymerase chain reaction
pe	petals
PFGE	pulse field gel electrophoresis
Ph	Petunia hybrida
PLE	PLENA
PI	PISTILLATA
PI motif	PISTILLATA- motif
PPI	protein-protein interaction
QP	glutamine-proline
r	replum
RACE	rapid amplification of cDNA ends
RAGE	rapid amplification of genomic DNA ends
rbcL	Ribulose-1,5-bisphosphate carboxylase oxygenase, large subunit
RBL	REBELOTE
RT-PCR	reverse transcriptase-polymerase chain reaction
SD	synthetic dropout
se	sepals
sec	seconds
SEP	SEPALLATA
SEU	SEUSS
SHP1/2	SHATTERPROOF 1 and 2
si-1	silky-1
sir	sirene

SLiM	short linear interaction motif
spw-1	superwoman-1
SQUA	SQUAMOSA
SQN	SQUINT
SSR	simple sequence repeat
st	stamens
STK	SEEDSTICK
SUP	SUPERMAN
T	thymine
ta-siRNA	transacting-small interfering RNA
TBE buffer	Tris Borate EDTA buffer
TM6	TOMATO MADSBOX GENE 6
Trp	tryptophan
UPT	ULTRAPETALA
UTR	untranslated region
vb	vascular bundle
VIGS	virus induced gene silencing
wt	wild type
WUS	WUSCHEL
YAB	YABBY

Summary

The development of plant organs is organized by a hierarchy of regulatory genes interacting with each other to build the plant's structures. How exactly floral organs and their morphology has evolved in angiosperms is investigated in the field of evolutionary development (evo-devo). Evo-devo research requires model species from widely distributed phylogenetic lineages, and the presented results were conducted in *Eschscholzia californica* (California poppy), a member of the Papaveraceae in one of the earliest diverging lineages of the basal eudicots. *E. californica* has recently been established as a model system for evo-devo studies and for this purpose new genetic resources were established.

A bacterial artificial chromosome library of California poppy genomic DNA was constructed and characterized to facilitate the molecular analysis of developmental control genes. Further, fast neutron mutagenesis was conducted to create a mutant population of California poppy in order to discover unknown genes involved in flower development of *E. californica*.

A new floral homeotic mutant, called *sirene* (*sir*), was uncovered within the help of the mutant population. *sir* plants showed homeotic conversions of petals into sepals and stamens are transformed into carpels. The molecular basis of the sir mutation was found to be based on a mutation in the C-terminal domain of the B class gene *SIR* (*EScaGLO*). This mutation altered the protein-protein interaction behaviour of SIR and multimeric complexes composed of SIR, EScaAG1 (C class), and EScaAGL9 (E class) were not assembled in the mutant. The failure to form those multimers most likely resulted in the failure of organ specification in stamens and petals.

The *E. californica* ortholog of the carpel developmental control gene *CRABS CLAW* (*CRC*) of *Arabidopsis thaliana*, *EcCRC*, was investigated in detail to uncover its function in *E. californica* carpel development. It was shown that *EcCRC* has very specific functions in regulating the development of carpel marginal tissues in the gynoecium, along with widely conserved functions shared by most angiosperm species. The *CRC*-like genes in general show an enormous diversity of functions in the different lineages of flowering plants. To understand the causes of this functional diversity might be very important to follow the evolution of the gynoecium, considered as being one the most important inventions of angiosperms that laid the ground for their evolutionary success.

Zusammenfassung

Die Entwicklung pflanzlicher Organe wird durch eine Hierarchie von Entwicklungs-Genen gesteuert, die durch das Zusammenwirken untereinander die Pflanzenstrukturen ermöglichen. Wie genau sich diese Blütenorgane und die Morphologie der Blüte als Ganzes im Laufe der Evolution entwickelt haben, ist der Forschungsgegenstand der Evolutionären Entwicklungsbiologie (Evo-Devo). Die Forschung im Bereich Evo-Devo erfordert Modell-Organismen aus einem weiten Bereich der Angiospermen-Systematik. Die vorliegenden Untersuchungen wurden deshalb mit *Eschscholzia californica* (Kalifornischer Mohn) ausgeführt, einem Mitglied der Familie der Papaveraceae, welche wiederum zu den frühesten Linien der eudicotylen Pflanzen gehört. *E. californica* ist eine neuartige Modellpflanze für evolutionäre Entwicklungsbiologie und für diese Forschungen wurden neue genetische Methoden etabliert.

Eine genomische Bibliothek des Kalifornischen Mohns wurde hergestellt und charakterisiert, um die molekulare Analyse von Entwicklungsgenen zu beschleunigen. Zusätzlich wurde eine Mutantenpopulation mit Hilfe der Fast Neutron-Mutagenese aufgebaut. Diese soll es ermöglichen, bisher unbekannte Kontrollgene der Blüten-entwicklung in *E. californica* zu entdecken.

Mit Hilfe der etablierten Mutantenpopulation war es möglich, eine florale Mutante des Kalifornischen Mohn zu entdecken, welche *sirene* (*sir)* genannt wurde. *sir* Mutanten zeigen homeotische Transformationen der Kronblätter (Petalen) in Kelchblätter (Sepalen) sowie eine Umwandlung der Staubblätter (Stamina) in Fruchtblätter (Karpelle). Die Ursache der *sir* Mutation konnte im C-terminalen Ende des homeotischen Gens *SIR* (*EScaGLO*) lokalisiert werden. Diese spezifische Mutation hat zur Folge, das sich SIR Proteine nicht mehr mit dem C-Klasse Protein EScaAG1 sowie dem E-Klasse Protein EScaAGL9 zu höheren Proteinkomplexen zusammen lagern können. Dies hat vermutlich zur Folge, dass die Identität von Kron- und Staubblättern in der *sir* Mutante nicht festgelegt werden kann und sich deshalb diese Organe nicht bilden.

Ein dem *CRABS CLAW* (*CRC*) aus *Arabidopsis thaliana* orthologes Karpell-entwicklungsgen, genannt *EcCRC*, wurde in *E. californica* detailliert funktionell unter-sucht. Es konnte gezeigt werden, das *EcCRC* sehr spezifische Funktionen ausübt in der Entwicklung jener Gewebe, die sich von den Karpell-Marginalen ableiten. Zusätzlich hat *EcCRC* Aufgaben in der Entwicklung des Gynoeciums, die in vielen Blütenpflanzen konserviert sind. Darüber hinaus zeigen die *CRC*-ähnlichen Gene eine besondere

funktionelle Vielfalt zwischen den verschiedenen Linien der Blütenpflanzen. Die Gründe dieser Vielseitigkeit an Genfunktionen werden als wichtig erachtet für die Evolution des Fruchtknotens (Gynoecium). Dieser stellt eine der entscheidenden Neuerungen der Blütenpflanzen dar, denen sie einen Großteil ihres evolutionären Erfolges verdanken.

1. Introduction

1.1 Evolutionary development (evo-devo)

1.1.1 Evolutionary developmental genetics of plants

The enormous variation of plant flower morphologies has always caught the attention of botanists and evolutionary biologists (Darwin, 1862). Many hypotheses have been launched to explain a particular astonishing phenomenon of plant structures, their great diversity in morphology and simultaneously, the appearance of a common underlying blueprint to all these structures. A modern attempt to understand this phenomenon comes from the field of evolutionary development, or evo-devo (Hall, 2003). It tries to integrate evolution and development, focusing on the function and molecular evolution of developmental genes that organize the ontogeny of organisms. A key to the evo-devo concept is that developmental control genes, often encoding highly conserved transcription factors or signalling molecules, are organized in regulatory networks that show hierarchical gene interactions. Genes with a higher hierarchical position in development generally occupy more central roles in the regulatory network whereas genes that are more peripherally located in the network have a lower hierarchical position.

A good example of central regulatory genes is given by the vertebrate *HOX* genes which specify the main body axis in the developing embryos in these organisms. An important evolutionary characteristic of genes like the *HOX* genes is the degree of pleiotropy by which these genes can be characterized. Mutational changes in the coding sequence can affect several developmental pathways simultaneously. For instance, mutations in *HOX* gene coding sequences can lead to homeotic transformation of one body part into another and are often embryo-lethal (Lemons and McGinnis, 2006). On the other hand, mutations in *cis*-regulatory regions are less likely to be deleterious because *cis*-regulatory mutations usually affect only a subset of the overall expression domain of these genes which can lead to altered expression patterns and allow phenotypic evolution (Stern and Orgogozo, 2008; Gompel and Prud'homme, 2009).

Another aspect of evo-devo deals with the dynamics of gene duplications. In plant evolutionary developmental studies, it became apparent that the majority of developmental control genes belong to transcription factor families originated by segmental or whole genome duplications (Bowers et al., 2003; Blanc and Wolfe, 2004; Chapman et al., 2004). Gene duplication can lead to neofunctionalization or subfunctionalization, i.e. the duplicated copy can evolve novel functions or become restricted to a subset of functions of the progenitor gene (Lynch and Conery, 2000). Also, duplicated genes usually retain partial redundancy in function, particularly directly after duplication, but this redundancy can remain in duplicated gene pairs over very long periods (Wagner, 1999; Moore and Purugganan, 2005). Gene duplications likely effect the organization of regulatory networks. Sub- and neofunctionalization allow the developmental networks to adjust to different environmental conditions whereas functional redundancy buffers the genome against harmful mutations and adds stability to regulatory networks (Blanc and Wolfe, 2004; Moore and Purugganan, 2005).

A prominent example of gene duplications in plants are the MADS-box transcription factors which have expanded by gene and/or genome duplications to different extents during the evolution of angiosperms (Theissen, 2001; Parenicova et al., 2003; Pinyopich et al., 2003). The developmental function of duplicated MADS-box genes has been elucidated in several species and ranges from total redundancy in the case of the *SHATTERPROOF1/2 (SHP1/2)* duplicated genes in *A. thaliana*, to a clear subfunctionalization between *PLENA (PLE)* and *FARINELLI (FAR)* genes in *Antirrhinum majus*, although remnant overlapping functions still exist between these both paralogs (Davies et al., 1999; Pinyopich et al., 2003).

1.1.2 *Eschscholzia californica* as a versatile model species

The California poppy, *Eschscholzia californica*, is a species native to the western parts of North America where it occupies a wide range of habitats, often living in naturally disturbed, open environments. *E. californica* lives from the sea level up to 2000 m on different soil formations, on river terraces and on the hillside, but also in human disturbed environments and along roadsides in California. The plant has also been introduced to

Chile, Tasmania, Australia and New Zealand where it is now a natural part of ecosystems (Cook, 1962). In these new locations, *E. californica* has proven to have a huge potential as an invasive species and was therefore chosen as a model species for invasion biology of plants (Leger and Rice, 2003, 2007).

E. californica, as a member of the family Papaveraceae, has long been known to contain specific compounds that had already been used as medicine by North-American natives. The secondary metabolites of California poppy belong to different types of alkaloids; aporphines, pavines, protoberberines, protopines and benzophenanthridines (Hauschild et al., 1998; Fabre et al., 2000). Many of these alkaloids have been characterized in detail, but the biosyntheses of benzophenanthridine-type alkaloids have been completely elucidated at the enzymatic level (Haider et al., 1997; Hauschild et al., 1998) and an alkaloid detoxifying enzyme has also been characterized recently(Vogel et al., 2010). Due to the pharmaceutical importance of their alkaloid contents, cell cultures of *E. californica* are continuously engineered to produce commercially important agents like benzyl-isoquinoline alkaloids (Pauli and Kutchan, 1998; Inui et al., 2007; Takemura et al., 2010).

E. californica also serves as a model for studying the molecular basis of development in early diverging eudicots and for the comparative genetics of developmental processes. The latter approach requires particularly informative taxa with respect to morphology or phylogenetic position. *E. californica,* as a member of the earliest diverging lineage of the Ranunculales, represents an important evolutionary link, and its position points in two directions; i) towards the core eudicots and ii) to different groups of highly evolved monocots like grasses or orchids (Fig.1)(APG, 2003). Moreover, the developmental characteristics of *E. californica* can be compared to more basal angiosperm lineages like the magnoliids or the species of the most basal angiosperm ANA grade (Amborellales, Nymphaeales and Austrobaileyales) (Fig.1) (Soltis et al., 2004; Frohlich and Chase, 2007). Besides these facts, the following characteristics are making *E. californica* a well-suited model plant for evolutionary developmental genetics. It is an easily cultivated annual to perennial plant with a life-cycle of three month under optimal growth conditions (Becker et al., 2005; Wege et al., 2007).

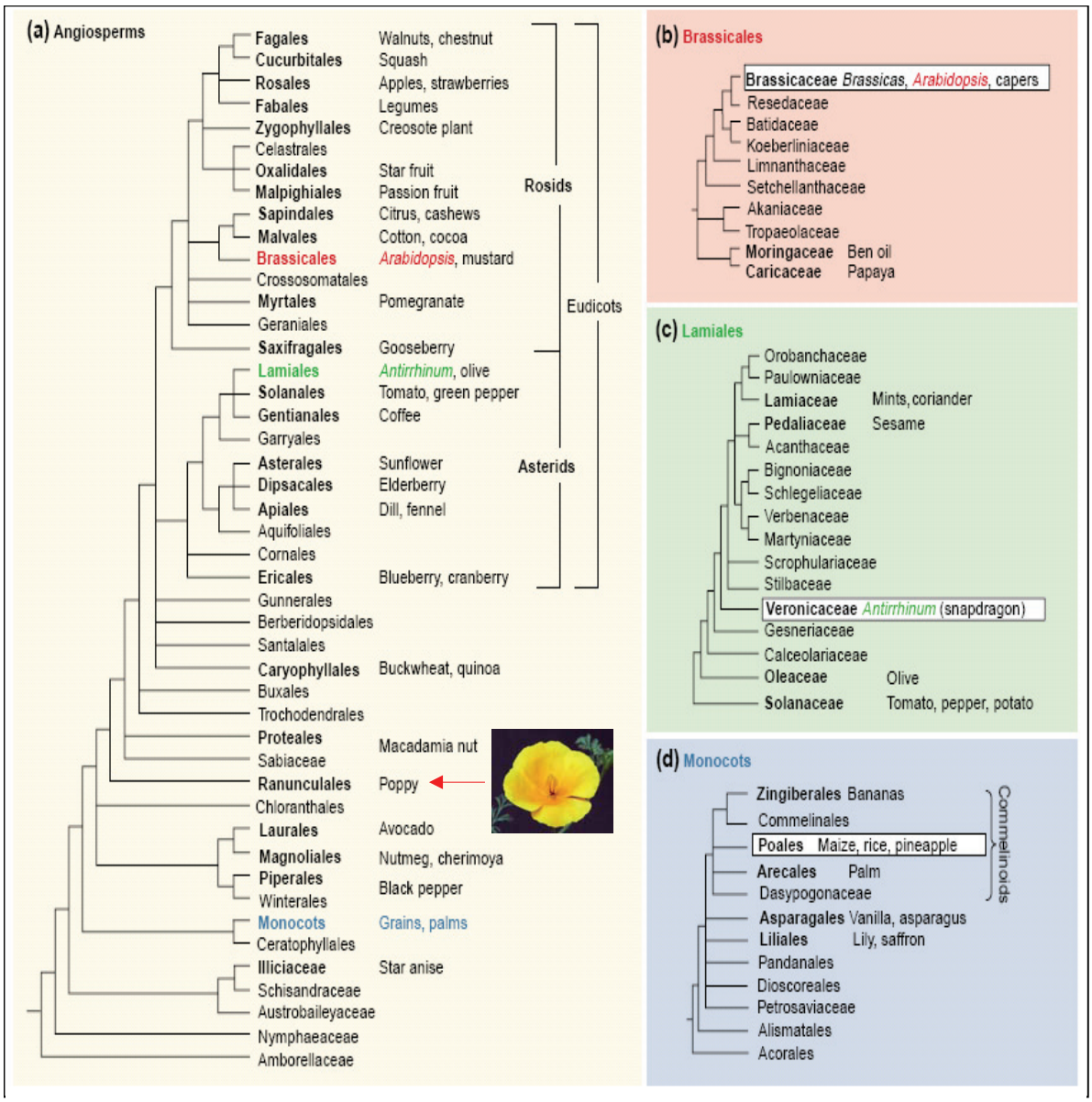

Figure 1: Summary tree of the angiosperm phylogeny

In **(a)**, the angiosperm phylogeny is given with the 'core' eudicots composed of two major clades named rosids and asterids is highlighted, and known model species are additionally coloured. In the inset, a typical flower of *E. californica* is placed. In **(b)** and **(c)**, the orders of Brassicales and Lamiales are shown. In **(d)**, maize and other Poaceae are part of the order Poales, which is nested within the commelinoid clade of monocots. Besides the model species, common names of exemplary agricultural important crops are shown for some plant groups but it is not considered a complete list. Modified from (Soltis and FGP, 2002).

The California poppy is amenable to genetic transformation and therefore allows a detailed characterization of gene functions (Park and Facchini, 2000). *E. californica* has a diploid genome consisting of 12 chromosomes with a size of 1115 Mb per haploid chromosome set (Bennett et al., 2000). Additionally, the genome of *E. californica* has

independently duplicated after the split that separated higher eudicots and Ranunculales, with the possibility to study the evolutionary dynamics of gene duplications (Cui et al., 2006). The Floral Genome Project has made an extensive effort to compile large amounts of expressed sequence tags and next-generation sequencing data that are publicly available (Soltis and FGP, 2002; Carlson et al., 2006; Wall et al., 2009). Of particularly interest is the fact that the miRNA content of different tissues has been surveyed in transcriptome analysis. Several classes of miRNAs and ta-siRNAs have been identified and provide an additional gateway for future evolutionary developmental studies (Barakat et al., 2007). In the field of leaf development, *E. californica* has already been established as a model species for evo-devo genetics (Busch and Gleissberg, 2003; Groot et al., 2005).

1.2 Molecular genetics of flower development

1.2.1 Floral organ specification and the ABCE model

Most angiosperm flowers consist of four distinct organs arranged in whorls. The outermost sterile perianth is composed of sepals and petals. The central flower whorls are occupied by male and female reproductive organs which are stamens and the carpels. In many cases, the carpels are fused and thereby build up the gynoecium which bears the ovules and provides a cavity for the developing seeds. The genetics of flower development is one of the best studied systems in plant developmental biology. Careful observations of floral homeotic mutants and their molecular characterization revealed a model of how the different organs of flowers are genetically determined. This is known as the ABC model whose hallmarks are three distinct genetic functions, A, B, and C that specify each floral organ (Fig.2) (Bowman et al., 1991; Coen and Meyerowitz, 1991).

In detail, the floral homeotic genes mainly belong to the MADS-box family of transcription factors, a large gene family that is involved in many plant developmental programs (Riechmann and Meyerowitz, 1997). In the following paragraphs, the floral developmental program of *A. thaliana* will be described as it is the model organism from which most of the data for the ABC model were obtained. The ABC functions act alone or combinatorial to specify floral organ identity. In this manner, sepals are specified solely by the action of the A class genes *APETALA1 (AP1)* and *AP2*, the latter belonging to the

AP2/EREBPs (APETALA2/ethylene-responsive element binding protein) family of transcription factors (Okamuro et al., 1997) and being the only non-MADS-box gene in the ABC model. Petals are specified by combinatorial action of the B function genes *AP3* and *PISTILLATA (PI)* together with A function genes. The C class gene is *AGAMOUS (AG)* which acts collectively with B class genes to specify stamen organ identity and *AG* alone specifies the identity of carpels (Coen and Meyerowitz, 1991) (Fig.4). A major extension of the classic ABC model was the discovery of the four redundantly acting E class *SEPALLATA (SEP)* genes, which are additionally required to specify all four floral organs (Pelaz et al., 2000; Theissen and Saedler, 2001; Ditta et al., 2004).

MADS-box genes other than *AG* also play an important role in carpel development and these genes mainly belong to the *AG* subclass of MADS-box transcription factors. The two genes that can independently of *AG* promote carpel development in *A. thaliana* are *SHP1/2*. Recently, *SHP* genes were also recognized to be more specifically occupied for the development of stigma, style and valve marginal tissues (Colombo et al., 2010). Both *SHP* genes are also involved in the specification of ovule identity, and share this function redundantly with *AG* and *SEEDSTICK (STK)*, another member of this subclass (Pinyopich et al., 2003). *STK* is required for the development of the funiculus, a stalk-like structure that connects the ovules to the carpel wall within the gynoecium (Pinyopich et al., 2003).

Predictions of the originally described ABC model were comparatively tested in several other angiosperm taxa. Extensive functional conservation, however, has only been found for B and C class genes (e.g. (Pnueli et al., 1994; Samach et al., 1997; Kramer et al., 1998; Ambrose et al., 2000; Drea et al., 2007). B and C function genes were also identified in gymnosperms and, given their expression in the corresponding reproductive organs, it was suggested that one of the ancestral functions of B and C genes included male and female reproductive development. The evolutionary origin of the ABC model is likely to be found in a sex determination system already present in the ancestors of angiosperms and gymnosperms (Mouradov et al., 1999; Theissen and Becker, 2004).

Contrarily, clearly comparable A class function homologs of *A. thaliana* were difficult to find in any other species surveyed (Gutierrez-Cortines and Davies, 2000; Litt, 2007). Furthermore, it was suggested that a specific A function is not required for perianth organ

identity outside of the Brassicaceae and can therefore be omitted from the original ABC model (Litt, 2007).

1.2.2 MADS-box gene function at the molecular level

MADS-box proteins consist of a highly conserved N-terminal DNA-binding MADS domain (M), a less conserved intervening domain (I) and a keratin-like domain (K) that contains three amphipathic α-helices involved in protein-protein interactions. At the C-terminal end of MADS box proteins, a variable C-terminal (C) domain is located which are specific for the different subgroups of this gene family. The MADS box proteins of the ABC model belong to the subgroup of MIKC^{C} type of MADS-box proteins (Yang and Jack, 2004).

MADS-box proteins form protein dimers and higher-order multimeric complexes to carry out their biological function as transcriptional regulators. The molecular mechanism of floral organ specification in *A. thaliana* was hypothesized to be mediated by 'floral quartets' consisting of MADS-box protein tetramers that act in a combinatorial manner (Theissen and Saedler, 2001).

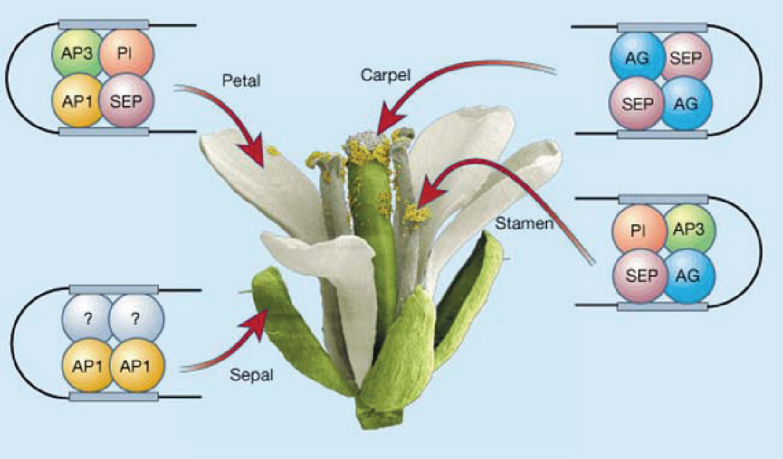

Figure 2: ABCE model and floral quartets

Assumed molecular action of floral organ specification in which tetramers of MADS-box transcription factor proteins act in combination to specify the identity of the four floral organs. Protein complexes bind to DNA on CArG-boxes in the promoter of target genes. For details see text, modified from (Theissen and Saedler, 2001).

To specify sepals, two A- and probably two E class proteins (AP1/AP1/SEP/SEP) need to assemble into higher-order complexes. For petal specification, tetramers of A, B and E proteins (AP1/AP3/PI/SEP) build up complexes. For stamens specification, complexes of B, C and E proteins (AP3/PI/AG/SEP) are necessary, and to specify carpels, tetramers of C and E class proteins (AG/AG/SEP/SEP) are required (Fig.2) (Honma and Goto, 2001). Some of the postulated higher-order complexes have been shown to occur in yeast protein interaction assays (Davies et al., 1996; Egea-Cortines et al., 1999; Honma and Goto, 2001; Favaro et al., 2003; Shchennikova et al., 2004).

It was determined that multimeric protein complexes need to bind to two specific DNA motifs, named 'CArG-boxes'. These boxes consist of the consensus nucleotide sequence motif 5'-CC (A/T)$_6$ GG-3' and are present in the promoter of genes that are regulated by MADS-box genes. The relative orientation of CArG-boxes present in the promoters of target genes differs which is thought to be important for target gene specificity of different MADS-box protein multimers (Melzer et al., 2008; Melzer and Theissen, 2009). Moreover, varying bases surrounding individual CArG-boxes are likely to be involved in target gene specificity by affecting DNA binding affinity of different higher-order MADS-box protein complexes (Kaufmann et al., 2009). However, the occurrence of these tetrameric complexes *in planta* is still rarely confirmed and has not yet been correlated to any mutant phenotype observed.

1.3 B-class MADS-box genes specify petals and stamens

1.3.1 *DEF* and *GLO* -like genes

Beside *A. thaliana*, B function genes were first identified in the model species *A. majus* where they have been named *DEFICIENS (DEF)* and *GLOBOSA (GLO)*, the representative orthologs of *AP3* and *PI*, respectively (Sommer et al., 1990; Tröbner et al., 1992). Phylogenetic analysis revealed that B class MADS box proteins contain the typical MIKC structure but are variable in their C-terminus. GLO proteins contain a highly conserved amino acid sequence motif within the C-terminal domain, present in almost all proteins investigated, called the PI motif (Kramer and Irish, 2000). A similar motif, the PI-derived motif, is present in the C-terminus of all DEF proteins. DEF proteins contain

additional conserved amino acid sequence motifs in the C-terminus, called the euAP3-motif, present only in the core eudicots or an ancient paleoAP3 motif in all other angiosperm DEF proteins (Fig.3) (Kramer and Irish, 2000).

These two different *DEF* gene lineages presumably arose by a gene duplication event just before the radiation of the higher eudicots (Kramer and Irish, 2000). Not all higher eudicot species have retained both duplicated *DEF* copies in their genome of which the paleoAP3 containing genes have evolved into the *TM6* lineage, named after the *TOMATO MADS BOX GENE 6* (Pnueli et al., 1991). *TM6* genes are particularly present in the family of Solanaceae; however, they are neither present in *A. thaliana* nor in *A. majus* (Fig.3).

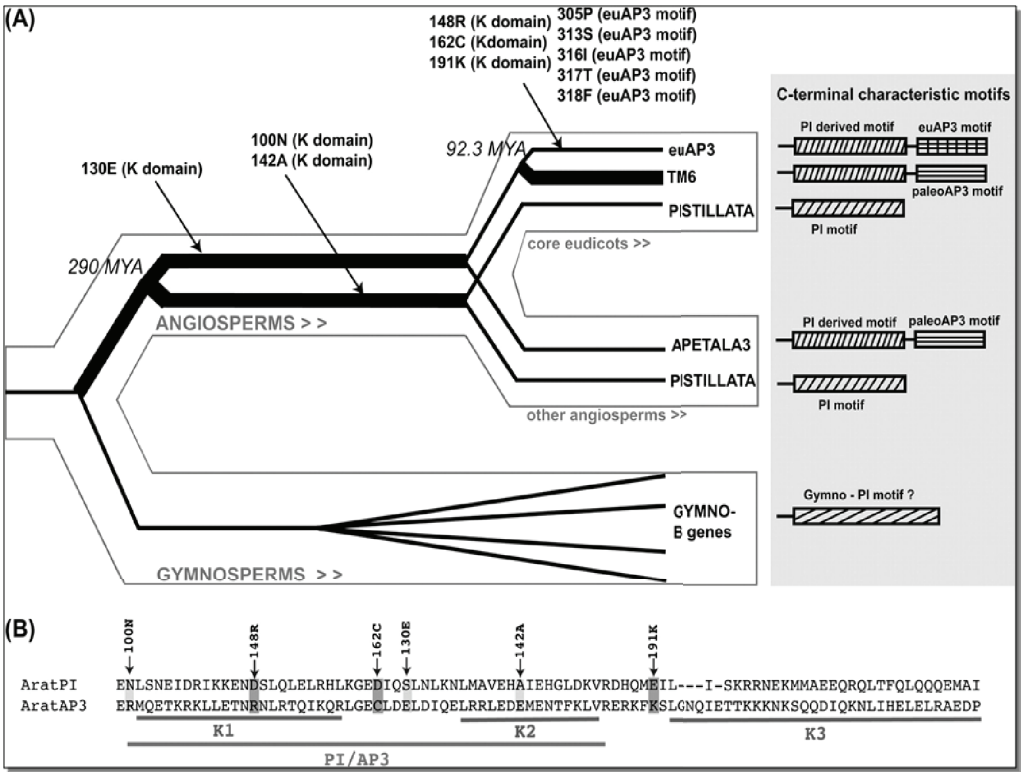

Figure 3: Simplified phylogenetic tree of DEF (AP3) and GLO (PI) proteins

(A) The estimated evolutionary history of AP3 and PI proteins from angiosperms and gymnosperms are illustrated. The C-terminal motifs are shown on the right. Amino acids in critical positions of AP3 and PI proteins that were subject to positive selection are indicated by arrows. The estimated time of gene duplication events is shown at the split of PI/AP3 and euAP3/paleoAP3 lineages. **(B)** Three amphipathic helices K1-K3 are shown with the positions of amino acids under positive selection and the region responsible for PI/AP3 protein heterodimerization. Modified according to (Hernandez-Hernandez et al., 2007).

The reason why some species of higher eudicots, especially in asterids, have retained both *DEF* and *TM6* gene copies while others lost exclusively their paleoAP3 containing *TM6* paralogous genes remains enigmatic. This is particularly remarkable when considering that genes which contain the paleoAP3 motif are the functionally ancient *DEF* genes present in all other angiosperm lineages.

The functional significance of either of the conserved motifs in the C-terminus of DEF and GLO proteins could not definitely been shown to date, basically due to inconsistent experimental results. *In vitro* protein interaction assays clearly showed that DEF and GLO C-termini are required for higher-order protein complex formation (Egea-Cortines et al., 1999). Furthermore, complementation experiments in the respective mutants of *A. thaliana* with domain swapped and C-terminal truncated B class proteins revealed that the C-terminus is essential for proper protein function (Lamb and Irish, 2003). On the other hand, similar experiments did not confirm that the C-terminus of DEF- and GLO-like proteins is required for floral organ identity function (Berbel et al., 2005; Piwarzyk et al., 2007). Due to this incongruence, the evolutionary and functional importance of the conserved C-terminal motifs of B class genes remain poorly understood.

1.3.2 Evolution of functional interactions of B class proteins

All floral homeotic MADS-box proteins have three amphipathic helices within their K domain, named K1-K3 and these were shown to be important for DEF and GLO protein heterodimerization via conserved hydrophobic positions within these helices (Yang et al., 2003; Yang and Jack, 2004). Several studies on DEF/GLO protein interactions elucidated that in higher eudicots like *A. thaliana* and *A. majus* no homodimers but only hetero-dimers are formed (Schwarz-Sommer et al., 1992; Goto and Meyerowitz, 1994; Davies et al., 1996; Krizek and Meyerowitz, 1996; McGonigle et al., 1996; Riechmann et al., 1996b). DEF/GLO heterodimers bind to CArG boxes which are present in the promoters of downstream target genes regulated by B genes in eudicots (Schwarz-Sommer et al., 1992; Tröbner et al., 1992; Riechmann et al., 1996b; Riechmann et al., 1996a).

A specific feature of DEF/GLO heterodimers is that they control their own transcription in a positive autoregulatory feedback mechanism during organ development (Jack et al., 1992; Goto and Meyerowitz, 1994). This function was imposed to be direct in the case of

AP3 in *A. thaliana* due to the presence of two adjacent CArG boxes in the *AP3* promoter that are indeed bound by AP3/PI heterodimers (Riechmann et al., 1996a; Hill et al., 1998; Tilly et al., 1998; Honma and Goto, 2000). In case of *PI*, the autoregulation was postulated to be indirect, because the responsible *PI* promoter does not contain any CArG boxes (Chen et al., 2000; Honma and Goto, 2000).

Besides their ability to form dimers, it is likely that DEF and GLO hetero- or homodimers take part in higher-order transcription factor complexes *in planta* which are hypothesized to regulate petal and stamen development (Fig.2) (Honma and Goto, 2001; Theissen and Saedler, 2001). Indeed, the DNA binding ability of MADS-box protein complexes consisting of the *A. thaliana* SEP3-SEP3-SEP3-SEP3 homotetramers and of SEP3-SEP3-AP3-PI heterotetramers has been shown by protein-DNA binding assays (Melzer et al., 2008). But how exactly multimer formation, promoter specificity and transcriptional activation are realized in the cell is still poorly understood.

The dimerization capabilities of B class proteins were determined in several studies to reveal the degree of functional conservation across divergent taxa. In basal eudicots and monocots, the formation of DEF or GLO homodimers was uncovered with some of them also showing DNA binding activity and even evidence of a GLO-like DNA binding homomultimer exists in orchids (Tzeng and Yang, 2001; Hsu and Yang, 2002; Kramer et al., 2003; Tzeng et al., 2004; Drea et al., 2007; Kramer et al., 2007; Tsai et al., 2008). Furthermore, the ancestral B class proteins were hypothesized to carry out their functional repertoire as obligate homodimers which was shown for GGM2 B class protein homodimers in the gymnosperm *Gnetum gnemon* (Winter et al., 2002). This obligate homodimerization has evolved into obligate heterodimerization of DEF/GLO proteins in higher eudicots and monocot grasses with some homodimerization ability retained in other monocots and basal eudicots (Riechmann et al., 1996b; Winter et al., 2002; Whipple et al., 2004; Drea et al., 2007; Kramer et al., 2007; Tsai et al., 2008). The question whether B class homodimers still have a residual function in eudicots or monocots remains open, since the functional specialisation from homo- to heterodimerization most probably has been accompanied by structural modifications in protein sequences of participating protein partners and may also have influenced the expression domains of B class genes via

auto- or cross-regulatory mechanisms (Winter et al., 2002; Hernandez-Hernandez et al., 2007; Lenser et al., 2009).

Addressing these questions, Kim and colleagues (2004) extensively studied the diversification of protein sequences during angiosperm evolution. Both, DEF and GLO proteins in the most basal angiosperm *Amborella trichopoda* show several features that are unique to the angiosperm DEF/GLO proteins, e. g. a prominent 'DEAER' amino acid motif in the C-terminus and altered amino acid residues in otherwise highly conserved positions like the KEN motif in GLO proteins. Also, critical residues for B class heterodimerization, like 97E and 98N in PI and 98N and 102R in AP3, were not found in the respective *A. trichopoda* proteins. These results together led to the suggestion that protein interactions among B class proteins might be very flexible in terms of homo- and heterodimerization capabilities and dynamic with respect to higher-order protein formation in *A. trichopoda* and to a lesser extent in other basal angiosperms (Kim et al., 2004). Deletions and other modifications in the K domain and possibly also in the C-terminus of DEF and GLO proteins might than have aided in the canalization of their functional properties, possibly resulting in obligate heterodimerization of B class proteins in core eudicots and monocot grasses. Furthermore, the regulation of B class higher-order protein interactions within the sepal and petal developmental programs might have been facilitated by protein sequence co-evolution of participating B class proteins (Vandenbussche et al., 2003; Kim et al., 2004).

1.3.3 Evolution of petal and stamen developmental programs

The B gene clade originated early in land plant evolution approximately 300-400 MYA in the common ancestor of gymnosperms and angiosperms (Theissen et al., 2000). Ferns, the sister group to seed plants, do not contain orthologs of floral homeotic genes, and therefore it was suggested that the B clade arose in the common ancestor of seed plants (Theissen et al., 2000). Putative B class gene orthologs have been isolated from gymnosperms, e.g. from the gnetophyte *G. gnemon* (gnemon) and the conifers *Picea abies* (Norway spruce), and *Pinus radiata* (Monterey pine) (Mouradov et al., 1999; Sundstrom et al., 1999; Winter et al., 1999). Approximately 290 MYA, close at the base of the angiosperms, but after gymnosperms and angiosperms have split, a duplication of B genes

was proposed that led to the present *DEF* and *GLO* lineages (Hernandez-Hernandez et al., 2007). Within the angiosperms, *DEF* and *GLO*-like genes have duplicated several times independently, e.g. *GLO*-like genes in monocot grasses and lilies and *DEF*-like genes in orchids and in the basal eudicot order Ranunculales (Kramer and Irish, 1999, 2000; Theissen et al., 2000; Munster et al., 2001; Kanno et al., 2003; Kramer et al., 2003; Tsai et al., 2004). In case of the core eudicots, DEF gene duplication led to the aforementioned differential *AP3/TM6* gene lineages which contain either euAP3 or paleoAP3 motifs in the C-terminal domain pointing towards functional deviations of the two duplicated gene copies (Fig.3).

According to their role in specifying petal and stamen organ identity, *DEF* and *GLO* expression can be found primarily in petals and stamens (Jack et al., 1992; Schwarz-Sommer et al., 1992; Tröbner et al., 1992; Goto and Meyerowitz, 1994). In the eudicot model species *A. thaliana* and *A. majus*, shortly after sepal primordia arise in the floral meristem, *DEF* expression can be detected in places where petals and stamens will develop later. The expression continues throughout stages of floral organ development. An additional domain of low expression of *DEF* can be found at the base of the sepals and in carpels of *A. majus* (Schwarz-Sommer et al., 1992). *AP3* expression, on the other hand, can also be found in developing *A. thaliana* ovules. The expression of *GLO* and *PI* genes is very similar to *DEF* and *AP3* in whorls 2 and 3, where petals and stamens will develop. They are also expressed at low levels in early stages of developing carpels (Jack et al., 1992). Other core eudicot species were found to have a similar spatially and temporally restricted domain of *DEF* and *GLO* expression, e.g. tomato, petunia and primrose (Tsuchimoto et al., 2000; de Martino et al., 2006; Li et al., 2008).

In the basal eudicot order Ranunculales, the expression patterns of B class genes have been determined for a wide range of families and species. Here, the expression domains of *DEF* and *GLO* genes appear generally broader, e.g. expression domains slide laterally from petals into the sepals or centrally from stamens into the carpels (Kramer and Irish, 2000; Kramer et al., 2003; Drea et al., 2007; Kramer et al., 2007). A comparable situation is found in monocots, where the B gene expression has been explored in lineages of grasses, lilies and orchids. Here, a generally less restricted domains of expression activity compared to core eudicots can be observed for B class genes (Ambrose et al., 2000; Munster et al., 2001;

Tzeng and Yang, 2001; Hsu and Yang, 2002; Nagasawa et al., 2003; Tsai et al., 2004). Further deviations from the conserved expression domain of B genes in petals and stamens was evident in some magnoliid dicot species which revealed that *DEF* and *GLO* genes are not universally expressed cooperatively in stamens as the classic ABC model would predict (Kramer and Irish, 2000).

These observations led to the proposal that the conservation of B class gene expression domains is less strict outside the core eudicots with their domains allowed to 'slide' or 'shift' over their boundaries as opposed to a more canalized expression pattern in higher eudicots (Fig.4) (Bowman, 1997; Kramer and Irish, 2000). This scenario is supported by expression analysis of B class genes in basal angiosperms, where they are expressed in all four floral organs and sometimes even in vegetative leaves (Kim et al., 2005). While B gene expression is not restricted to specific floral organs in *A. trichopoda*, *Nuphar advena* (cow lily), and *Illicum floridanum* (Florida anisetree), they do form a gradient of decreasing strength towards the margins of their domains (Fig.4). This results in an intergrading morphology of spirally arranged floral organs (Buzgo et al., 2004; Kim et al., 2005). Therefore, the 'fading border' model was launched as an explanation of how the spatial structure of gene expression affect floral organ morphology in basal angiosperms (Buzgo et al., 2004; Soltis et al., 2007).

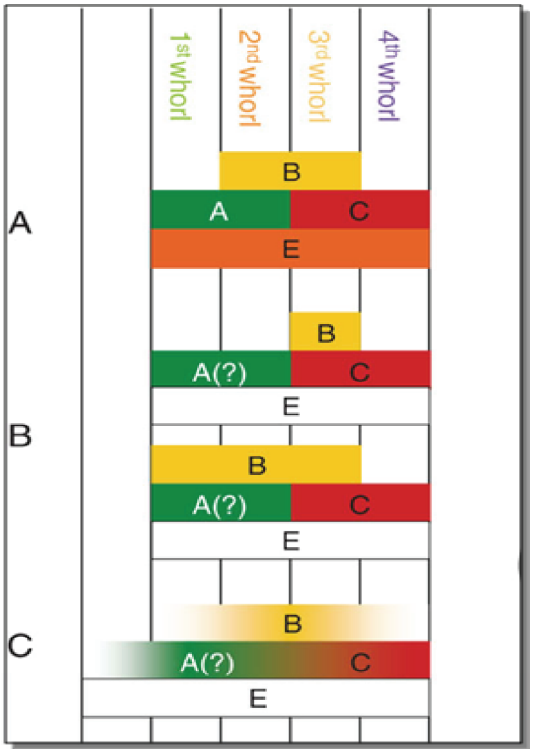

Figure 4: Expression domains of ABCE function regulatory genes in the four floral whorls of different plant groups.

(A) In higher eudicots, B class genes show a clear boundary of expression in the 2nd and 3rd whorl. **(B)** The sliding boundary of constricted or extended B gene expression as derived from basal eudicots and magnoliids is shown. **(C)** The fading border model of class gene expression in basal angiosperms is highlighted. Modified from (Soltis et al., 2007).

1.4 Carpel developmental genetics

1.4.1 Fundamentals of carpel structure

The female reproductive organs of angiosperms are complex structures, called gynoecium when composed of more than one carpel. Two types of gynoecia basically exist: apocarpous gynoecia with unfused carpels that form independent pistils within the flower and syncarpous gynoecia with fused carpels, which are considered a key innovation of angiosperms (Endress and Igersheim, 1999; Endress, 2001). It is of major commercial interest to understand the evolution of gynoecium development, since angiosperm carpels and their derived structures are an important source of food and drugs for human living and health. Most of the discoveries of carpel or gynoecium development were made in the advanced higher eudicot plant *A. thaliana* and in the monocot grasses *Oryza sativa* (rice) and *Zea mays* (maize) (e.g. (Mena et al., 1996; Alvarez and Smyth, 1999; Alvarez and Smyth, 2002; Yamaguchi et al., 2004; Yamaguchi et al., 2006). But with respect to the complexity of gynoecium structures and their development, these groups alone might not be sufficient to explain the full evolutionary history of gynoecium development, and there is need of more comparative data derived from model species covering a wider phylogenetic range like *E. californica*.

The gynoecium can be divided into several parts based on an apical-basal axis. The most basally located structure is the ovary which bears the ovules and can be separated by septa into one or more locules or cavities. On top of the ovary, a structure called style is positioned which contains the transmitting tract. And most apically, the stigma develops with specialized tissue cells at its top. The pollen attaches to the stigmatic cells and the pollen tubes subsequently grow via the transmitting tract to the ovules for fertilization. The stages of prefertilization flower development and postfertilization fruit development have been described in detail for California poppy by Becker and colleagues (Becker et al., 2005).

1.4.2 *CRABS CLAW*-like YABBY transcription factors in carpel development

The YABBY genes comprise a small family of plant specific transcriptional regulators which are present only in seed plants (Floyd and Bowman, 2007). YABBY genes encode for proteins that contain a C_2C_2 zinc finger domain near their N-terminal end and a helix-loop-helix domain, with sequence homology to the known DNA binding HMG-box (high mobility group), a common part of regulatory proteins in vertebrates (Bowman, 2000). Six YABBY genes are known from *A. thaliana*, whereas eight members were discovered in rice (Alvarez and Smyth, 1999; Bowman and Smyth, 1999; Eshed et al., 1999; Sawa et al., 1999; Siegfried et al., 1999; Villanueva et al., 1999; Yamaguchi et al., 2004; Toriba et al., 2007). All YABBY genes in eudicots are involved in promoting abaxial cell fate in lateral organs (Bowman, 2000). In contrast to eudicots, YABBY genes in monocot grasses differ in this respect, because maize YABBY genes are expressed adaxially and members in rice, wheat and sorghum are not expressed in a polar manner at all (Jang et al., 2004; Juarez et al., 2004; Zhao et al., 2006; Ishikawa et al., 2009).

The YABBY genes *CRABS CLAW (CRC)* and *INNER NO OUTER (INO)*, initially analysed in *A. thaliana*, are involved in the development of female reproductive organs. *INO* acts exclusively in the specification of the outer integument of ovules in *A. thaliana* (Villanueva et al., 1999). *CRC* has a particular function in carpel growth and fusion and is the major determinant of nectary development (Alvarez and Smyth, 1999; Bowman and Smyth, 1999; Eshed et al., 1999). *CRC*-like genes are also the most widely characterized YABBY members with respect to phylogenetic distribution and number of species, where representatives of higher eudicots as well as monocots and basal angiosperms were analysed (Yamaguchi et al., 2004; Fourquin et al., 2005; Lee et al., 2005b). *CRC*-like genes are consistently expressed at the abaxial side of developing carpels from basal angiosperms to higher eudicots (Bowman and Smyth, 1999; Fourquin et al., 2005; Lee et al., 2005a). Exceptions to this are the *CRC*-like genes in the highly specialized grass lineage, where they have acquired additional roles. The *CRC* ortholog from rice, *DROOPING LEAF (DL)*, is a major determinant of carpel identity, floral meristem termination and leaf midrib development. *DL*, like other YABBY genes in monocot grasses, has lost the polar

expression domain within the carpel and leaf primordia. Instead, the expression of *DL* is extended to the adaxial side of carpels and the leaf midrib (Yamaguchi et al., 2004).

2. Generation of genetic tools for functional analysis of genes in *E. californica*

2.1 Objectives

Comparative developmental genetic studies require the precise analysis of gene functions. In model plants like *A. thaliana*, diverse genetic tools to investigate and to manipulate gene function are often available but less so in the new model organism *E. californica*.

We aimed to expand the repertoire of functional tools for *E. californica* by constructing a bacterial artificial chromosome (BAC) library of nuclear DNA with the emphasis to aid cloning strategies of coding sequences and in particular of *cis*-regulatory regions of developmental genes. The latter are probably involved in the evolution of gene function by mutational changes in orthologous genes whose function has been described in other model plants. But since the focus of the thesis was not only to compare the functional characteristics of genes already described in other species, a second project was launched. It consists of a population of poppy plants, mutagenized with fast neutrons and subject to subsequent mendelian crossing that can be screened for alterations of the flower and organ morphology. With this approach it was expected to significantly enhance the discovery rate of genes involved in flower and carpel development in *E. californica* which are not known so far or very specific to this lineage.

Both projects were thought to strengthen the role of *E. californica* as a model plant for evolutionary developmental genetics of flower organs by extending the experimental repertoire. The mutant population was also expected to benefit from the BAC library, i.e. that the molecular analysis of promising mutant phenotypes would be aided by the analysis of the mutated genomic locus carried out with the help of BAC library screening.

2.2 Material and Methods

2.2.1 Construction of a bacterial artificial chromosome library

E. californica variety 'Aurantiaca Orange King' were grown under conditions previously described (Wege et al., 2007). The above-ground organs of three week old seedlings were harvested, snap frozen and stored at -80°C. High molecular weight (HMW) DNA was used for construction of the BAC library. HMW-DNA isolation and BAC library construction was carried out by Rx Biosciences (Rockville, USA). For this, HMW-DNA of *E. californica* was partially digested with *Hind*III and size separated by pulse field gel electrophoreses (PFGE). DNA fragments in the size range between 100 kb and 250 kb were ligated into the pINDIGO-BAC-5 *Hind*III *cloning ready* vector system (Epicentre Biotechnologies, Hess. Oldendorf, Germany), transformed into *EPI300™* cells (Epicentre Biotechnologies, Hess. Oldendorf, Germany) and selected on 12,5 µg chloramphenicol/ml.

BAC clones were picked manually, cultured in LB media [13 mM KH_2PO_4, 36 mM K_2HPO_4, 1.7 mM sodium citrate, 6.8 mM $(NH_4)_2SO_4$, 0.4 mM $MgSO_4$, 4.4 % v/v glycerol and 12,5 µg chloramphenicol/ml], and subsequently stored at -80°C. For membrane spotting, carried out by the ADIS facility at the Max-Planck Institute for Plant Breeding Research (Cologne, Germany), the clones were picked by a QPix2 colony picker (Genetix, New Milton, UK) and macroarrayed on Amersham Hybond™-N+ membranes (GE Healthcare Europe GmbH) with the Biorobotics MircroGridII system (Isogen Life Science, De Meern, NL).

2.2.2 Estimation of BAC library genome coverage

BAC plasmids were isolated from bacterial cultures, grown in LB media containing 12,5 µg chloramphenicol/ml, with the EasyPrepPro miniprep kit (Biozym Scientific GmbH, Hess. Oldendorf, Germany). The inserts of the BAC vector were released by *Not*I digestion, separated on a 1 % (w/v) agarose gel in 0,5X TBE buffer with PFGE on a CHEFII apparatus (BioRAD Laboratories, Munich, Germany) using the following conditions: 5s-15s linear switch ramp, 120° angle at 6V/cm for 16 h at 12°C, and documented.

2.2.3 Estimation of organelle content in the library

The screening of the BAC library for organelle contamination was conducted by Anna Menke, Developmental Genetics Group, University of Bremen, Germany. Hereafter is a brief description of the methodological work.

The BAC library was screened with *E. californica* gene specific probes. A 316 bp fragment of the mitochondrial gene *cox3* (*CYTOCHROME OXIDASE 3*), a 392 bp fragment of the chloroplast *rbcL* gene (large subunit of *RuBisCO*) and a 303 bp fragment of the nuclear gene *EcAuxRF1* were amplified from genomic DNA for probe generation. The PCR fragments of *cox3, rbcL* and *EcAuxRF1* were labeled with digoxigenin (DIG). Hybridization of the BAC filter was done according to the Roche DIG Application Manual for Filter Hybridization (Roche Applied Science, Mannheim, Germany). Hybridizations were performed over night at 43,3°C for *rbcL*, 43,8°C for *cox3*, and 43,4°C for *EcAuxRF1*, followed by low and high stringency washes. Signals were detected by CDP-star (Roche Applied Science) and luminescence of the BAC filter was measured with a LAS-3000 chemiluminescence detector (Fuji Photo Film Europe GmbH, Düsseldorf, Germany). (Primer sequences and accession numbers are given in chapter 6.1).

2.2.4 Fast neutron mutagenesis

Dry seeds from *E. californica* were irradiated with fast neutrons at three dosages: 20 gray (Gy), 40 Gy, and 60 Gy at the AEKI Irradiation Facility in Budapest, Hungary under the expertise of Joe K. Palfalfi. Seeds exposed to fast neutron irradiation were surface sterilized with 2% sodium hypochloride and transferred to Petri dishes with B5-agar. To ensure synchronous germination, seeds were vernalized at 4°C for 4 days. Germination and survival assays were carried out in growth rooms at 20 °C with 16 hours light at 70 µmol s^{-1} m^{-2}. Seed germination and seedling growth was observed over a period of 21 days in 3-day intervals, and the numbers of germinating seeds at each dose rate and the seedlings survival over the three week period was recorded.

2.2.5 Plant crossings

Seeds irradiated with 40 Gy were used for all further experiments and cultivated (Wege et al., 2007). Two F_0 plants were crossed with each other to produce the F_1 generation. F_1 sibling plants were than inbreeded to produce homozygous recessive genotypes in the F_2 generation. Inbred lines of all F_2 crosses were grown and screened for flower developmental mutants.

2.3 Results and Discussion

2.3.1 Construction and characterization of the genomic BAC library

To facilitate reverse genetic techniques a genomic bacterial artificial chromosome library of *E. californica* was constructed which is intended to help in cloning and characterization of open reading frames and regulatory regions of developmental control genes. Additionally, mutant plants identified in the fast neutron mutant population will be characterized with the help of the BAC library.

The construction of the BAC library was essentially done by Rx Biosciences (Rockville, Maryland, USA). Genomic DNA of the above-ground of young plants was isolated as high-molecular weight DNA and partially digested. Fragments were size selected with pulse field gel electrophoresis and subsequently used to construct the BAC library. A total of 22,565 BAC clones were obtained which were kept in 384 well plates at -80°C for long term storage. The whole library was replicated from hand-picked 96-well-plates and spotted on nylon membranes by the Automatic Sequencing Facility (ADIS) at the Max-Planck Institute for Plant Breeding Research in Cologne, Germany. The identification of particular clones can be facilitated by screening the spotted library with standard hybridization techniques using sequence specific probes.

In order to characterize the *E. californica* BAC library, the average insert size of BAC plasmids was assessed by picking 137 randomly selected BAC clones. The BAC plasmids were digested with *Not*I to release the insert from the vector pINDIGO-BAC-5 (Fig.5A). The average insert size of the BAC plasmids was calculated to be 131 kb. The insert size

distribution of BAC inserts was also estimated (Fig.5B). Only three of 137 clones showed no insert (2.2 %), whereas 18 clones (13.1 %) have inserts smaller than 50 kb. 95 clones (69.3 %) are in the size range between 50 kb and 200 kb, and 21 clones (15.3 %) had insert sizes larger than 200 kb (Fig.5B).

By calculating the size of separated fragments, an average insert size of 131 kb was estimated for the BAC library. The genome coverage of the entire BAC library was calculated with a genome size of 1078 Mb and alternatively with 502 Mb per haploid genome (Bennett et al., 2000; Cui et al., 2006). Given the two estimates of genome size and the obtained average sizes of BAC inserts, the genome coverage can be calculated as being 2.8-fold or 5.9-fold, respectively. The discrepancy of genome size estimations in the literature impedes a precise calculation of the probability to observe any genomic DNA sequence in the library (Bennett et al., 2000; Cui et al., 2006). However, the calculation was made with the larger genome size estimation of 1078 Mb. With an average BAC insert size of 131 kb it is possible to observe any genomic fragment with a probability of 93.6 % if considering BAC clones having organelle DNA inserted (Clarke and Carbon, 1976).

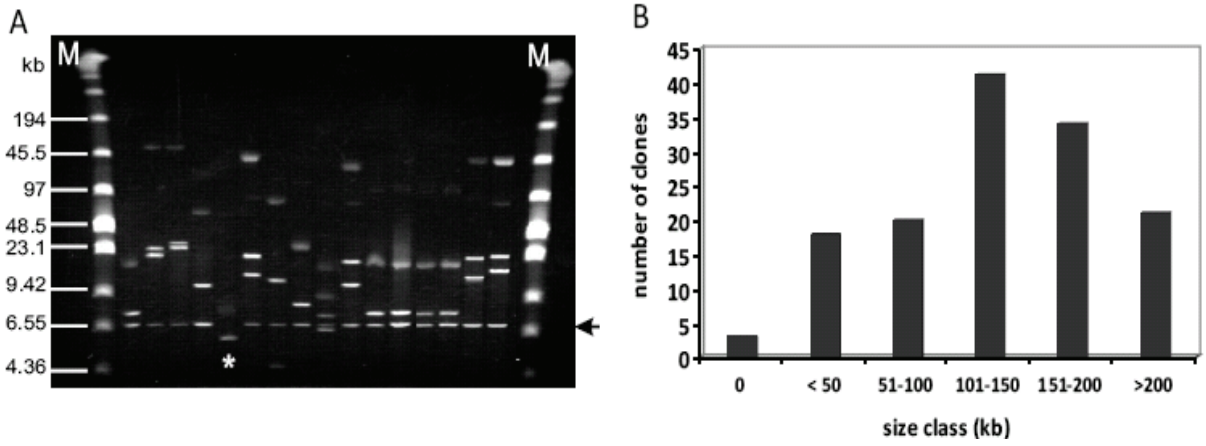

Figure 5: Summary of BAC library inserts size assessment

(A) Sixteen randomly selected BAC clones, digested with *Not*I and separated with pulse field gel electrophoresis (PFGE). The arrow marks the vector band and the star denotes a digested BAC without *pINDIGO-BAC5* vector band. **(B)** The total insert size distribution of 137 BAC clones is depicted as the number of clones in different size categories of BAC vector inserts.

The organelle content is an important feature of the overall quality of BAC libraries. Specific probes of the mitochondrial gene *cox3* (*CYTOCHROME OXIDASE3)* and the

chloroplast gene *rbcL* (the large subunit of the enzyme *RIBULOSE BIS-PHOSPHATE CARBOXILASE/OXIDASE*) were designed to assess the amount of organellar DNA contamination in the library. A total of six hybridization signals for the mitochondrial gene *cox3* was detected which would suggest a contamination rate of 0.03% BAC plasmids having DNA of mitochondrial origin. The chloroplast gene specific probe of *rbcL* revealed 57 signals when the library was probed. This approximately corresponds to 0.25% BAC clones with chloroplast DNA content in the library.

Angiosperm chloroplast genomes are usually in the size range between 120 kb-160 kb (reviewed in(Palmer, 1985) whereas plant mitochondrial genomes can be much larger and vary in size more between species. The probability of finding any chloroplast and mitochondrial sequence in the library was also estimated. Since the organelle genome sizes of *E. californica* are not known, a genome size of 120 kb for chloroplasts was assumed which is at the lower size limit of angiosperm chloroplast genomes (Palmer, 1985). With 57 BAC clones identified having chloroplast DNA inserted and an average BAC insert size of 131 kb, it was assumed that almost any BAC clone with chloroplast DNA can be detected using a single chloroplast gene-specific probe. The genome sizes of mitochondria in flowering plants are more variable and range between 200 kb and 2400 kb. Six hybridization signals with a mitochondrial DNA specific probe were detected and therefore, with a single mitochondrial gene specific probe, it is likely to find this particular gene in the library with a rate between 65% to 5% of all mitochondria containing BAC clones in the library, depending on the actual mitochondrial genome sizes, when calculated with 200 kb to 2400 kb, respectively.

This low contamination rate of the library with organelle DNA shows that the method used to isolate HMW-DNA from poppy tissue was extremely efficient in excluding organelle DNA, specifically mitochondrial DNA, from further processing of the library.

2.3.2 Fast neutron mutagenesis and subsequent gene identification

The identification of novel regulatory genes of floral organ development in poppy requires the molecular analysis of mutant plants. In order to obtain mutants affected in flower development, poppy seeds were mutagenized with fast neutron irradiation to saturate the

genome with mutations, and a strategy for screening the resultant mutant population for relevant mutant phenotypes was developed.

It has long been known that ionizing radiation like fast neutrons induce DNA double strand breaks (DSB) (McClintock, 1931). Such break points are targeted by the plant's endogenous DSB repair system (Britt, 1999; Gorbunova and Levy, 1999). Upon induction of DSB, deletions and insertions are usually introduced into the genomic locus by the endogenous double strand break repair system that mainly acts via non-homologous end-joining, an error prone mechanism (Rinehart et al., 1997), and the genomic loci affected by DSB frequently show a complex rearrangement pattern. The genomic alterations at the DSB loci can span a range between a few base pairs up to 10 kb, thereby most likely producing mutations such as whole gene deletion, mutations in coding genes that lead to aberrant proteins or reorganized promoters that regulate the expression patterns of genes (Shirley et al., 1992; Bruggemann et al., 1996).

The dosages that are applied to seeds in fast neutron treatments are given in gray (Gy), which is the amount of radiation energy absorbed by any form of matter. This spanned from 4,5 Gy in *Lactuca sativa* (lettuce), to 8 Gy in *Lotus japonicus* and *Glycine max* (soybean) up to 60 Gy in *A. thaliana* (Okubara et al., 1994; Li et al., 2001; Men et al., 2002; Hoffmann et al., 2007). Therefore, the optimal dosage of fast neutron irradiation on *E. californica* seeds was experimentally assayed by observing seed germination and survival of seeds treated with 20 Gy, 40 Gy, and 60 Gy. Seeds were plated onto plant agar and the germination of seeds was recorded for a three week period. From the seeds that germinated, between 80-90% did so within the first 3 days of the assay, the survival rate was determined as the number of seedlings that were still growing after a three week period.

The germination rate of seeds increased with irradiation intensity from 34.9% in control plants to 54.7% in the 60 Gy treatments. This possibly reflects reduced seed dormancy in mutagenized seeds since dormancy appears to be a quantitative trait genetically controlled by many loci and also by environmental factors (Foley and Fennimore, 1998). On the other hand, the survival of germinated seedlings was negatively correlated with increasing fast neutron dosage. The high survival rate in the 20 Gy treated plants suggests mild