Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 573

Evolution of a Family of Plant Genes with Regulatory Functions in Development; Studies on *Picea abies* and *Lycopodium annotinum*

BY

MATS SVENSSON

ACTA UNIVERSITATIS UPSALIENSIS
UPPSALA 2000

Dissertation for the Degree of Doctor of Philosophy in Physiological Botany presented at Uppsala University in 2000

Abstract

Svensson, M., 2000. Evolution of a Family of Plant Genes with Regulatory Functions in Development; Studies on *Picea abies* and *Lycopodium annotinum.* Acta Universitatis Upsaliensis. *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 573. 45 pp. Uppsala. ISBN 91-554-4826-7.

This work is focused on the molecular genetic basis for morphological change in evolution. Genes belonging to the MADS-box gene family, which includes members, that determine angiosperm floral organ identity, were isolated and characterised from two non-angiosperm plants; Norway spruce (*Picea abies*) and the club moss (*Lycopodium annotinum*).

The exon/intron organisation of the isolated genes was determined, and its significance as an independent test of the position of a gene within the gene family tree evaluated.

Norway spruce genes that are closely related to the angiosperm floral organ identity genes were identified. One Norway spruce gene, *DAL2*, is an ortholog to angiosperm C-class MADS-box genes that specify stamen and carpel identity. The expression of *DAL2* in male and female cones suggests that orthologous genes in conifers and angiosperms determine the identities of pollen- and seed-bearing structures. Constitutive expression of *DAL2* in the angiosperm Arabidopsis resulted in homeotic conversions very similar to those resulting from constitutive expression of the Arabidopsis C-class gene.

Angiosperm B-class MADS-box genes determine petal and stamen identity. The isolated Norway spruce B-class orthologs: *DAL11, DAL12*, and *DAL13* are expressed in the developing male cones exclusively, suggesting a conserved function of B-class related genes in the determination of pollen forming organs among seed plants.

No orthologs to the floral organ identity genes could be isolated from the club moss, suggesting that the origin of these gene classes may be coupled to the origin of the pollen and the seed.

The club moss MADS-box genes, *LAMB2*, *LAMB4* and *LAMB6*, conform structurally to plant type MADS-box genes, whereas *LAMB1* is divergent in details. The genes *LAMB3* and *LAMB5* encode shorter proteins.

LAMB1 expression is restricted to reproductive structures, whereas *LAMB2*, *LAMB4*, *LAMB5* and *LAMB6* are broadly expressed. The implications from these expression patterns on the ancestral function of plant type MADS-box genes are discussed.

Mats Svensson, Department of Physiological Botany, EBC, Uppsala University, SE-752 36, Uppsala, Sweden

ISSN 1104-232X
ISBN 91-554-4826-7
Printed in Sweden by University Printers, Uppsala 2000

…Lycopodium? That's a good one!

E. MEYEROWITZ

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals:

I. **Tandre K., Svenson M., Svensson M. E., Engström P.** (1998) Conservation of gene structure and activity in the regulation of reproductive organ development of conifers and angiosperms. *Plant J.* **15**:615-623.

II. **Sundström J., Carlsbecker A., Svensson M. E., Svenson M., Johanson U., Theißen G., Engström P.** (1999) MADS-box genes active in developing pollen cones of Norway spruce (*Picea abies*) are homologous to the B-class floral homeotic genes in angiosperms. *Dev. Genet.* **25**:253-266.

III. **Svensson M. E., Johannesson H., Engström P.** (2000) The *LAMB1* gene from the clubmoss, *Lycopodium annotinum*, is a divergent MADS-box gene, expressed specifically in sporogenic structures. *Gene* **253**:31-43.

IV. **Svensson M. E., Engström P.** Closely related MADS-box genes in the club moss, *Lycopodium annotinum*, show broad expression patterns and are structurally similar to typical seed plant MADS-box genes. (in manuscript)

Table of contents

	Abbreviations	6
1.	**Homeotic genes and the evolution of plant developmnet**	7
1.i	Interaction of floral homeotic genes, the ABC-model	8
1.ii	Floral homeotic genes belonging to the MADS-box gene family	9
1.iii	The expression patterns of MADS-box genes reflect their function	9
1.iv	Conservation of function between orthologous genes	9
1.v	The MADS-box genes constitute a large gene family in eukaryotes	10
1.vi	The MADS-domain	10
1.vii	Subfamilies of MADS-box genes	11
1.iix	Functional domains of MIKC-genes	11
1.ix	Exon/intron organisation of MIKC-genes (I, II, IV)	12
1.x	Divergent MADS-box genes with a MIKC exon/intron pattern (IV)	13
1.xi	The role of MIKC-genes in angiosperm development	15
1.xii	Phylogeny of function in plant MADS-box genes	16
2.	**MADS-box genes in gymnosperms (I, II)**	16
2.i	*DAL2* – a Norway spruce C-class gene (I, II)	17
2.ii	*DAL2* can act as a C-class gene in an angiosperm context (I)	17
2.iii	*DAL2* and the functional diversification of the *AG*-clade	18
2.iv	Putative B-class genes from Norway spruce (II)	19
2.v	Exon/intron organisation of *DAL11*, *DAL12*, and *DAL13* (II)	20
2.vi	Expression of Norway spruce B-class genes in male cones (II)	20
2.vii.	B- and C-genes in *Gnetum*	21
3.	**MADS-box genes in non seed plants, *e.g.* Lycopsids (III, IV)**	21
3.i	Lycopsids constitute the sister group to all other extant vascular plants	22
3.ii	*LAMB1* is a divergent MIKC-gene from *Lycopodium annotinum* (III)	22
3.iii	Typical MIKC-genes from *L. annotinum* (IV)	24
3.iv	The phylogenetic position of the *L. annotinum* MADS-box genes (III,IV)	25
3.v	Exon/intron organisation of *LAMB2*, *LAMB3*, *LAMB4*, and *LAMB6* (IV)	25
3.vi	A leucine-zipper in the K- and C-region of *LAMB4* (IV)	26
3.vii	Expression patterns of *L. annotinum* MIKC-genes (III, IV)	26
3.iix.	MADS-box genes in ferns	27
3.ix	Evolutionary considerations (III,IV)	27
4.	**Homeotic genes and organ homology**	28
4.i	Homology as a neo-Darwinian paradigm	30
4.ii	Homeotic genes and homology	30
4.iii	Homeotic gene expression does not define morphological homology	30
4.iv	Homologies between reproductive structures of seed plants	32
4.v	B- and C genes and seed plant reproductive organ homologies (I, II)	33
4.vi	ABC-genes and floral organ homology	34
4.vii	Lycopsid microphyll homology	35
5.	**Concluding remarks and suggestions**	35
	Acknowledgements	37
	References	38

Abbreviations

aa	amino acids
bp	base pairs
mya	million years ago
RACE	rapid amplification of cDNA ends
RT-PCR	reverse transcriptase polymerase chain reaction

1. Homeotic genes and the evolution of plant development

When the discipline of molecular evolution emerged, two important notions were made. Firstly, there appeared to be no necessary connection between the number of genes, and the morphological complexity of an organism (Ohno, 1972). A recent comparison between the number of genes in the nematode, *Caenorhabditis elegans* and the fruit fly, *Drosophila melanogaster* genomes, ca 20000 *vs*. 12000 (Meyerowitz, 1999), is in accordance with this notion. Secondly, evolutionary change in coding sequences appeared to a large extent to be neutral with respect to protein function (Kimura, 1977), and hence have no effect on morphology. These findings suggest that change in the *cis*-regulatory elements of genes, rather than in their coding sequences would be the predominant mechanism for morphological evolution (Doebley and Lukens, 1998).

Development can be modeled as composed of modules showing varying degree of connection /dissociation relative each other (*e.g.* Raff, 1996). That developmental modules are (more or less) dissociated from each other is of great evolutionary importance, because mutations may consequently affect one module without having effects on the entire organism. For naturalists such as Cuvier, the belief that the organism represents an integrated whole, and that no individual detail could be changed without affecting this whole, served as the major argument against the evolutionary theory (see Gould, 1992). At the morphological level a developmental module may be represented by for example cell types, tissues or structures such as a meristem or different organ primordia in plants, or tagma, segments, and developing organs in arthropods. Evolutionary change of a module may involve quantitative changes in the interaction between components within the module, or the recruitment of submodules into the module, probably by the changes in *cis*-regulatory elements.

Large phenotypic changes are less likely to confer a selective advantage, than small (*e.g.* Fisher, 1930). Doebley and Lukens (1998) argue that a transcription factor, unlike many other proteins, generally is confined to a particular developmental module, rather than having global developmental roles. This suggests that mutations directly affecting transcription factor activity would be less pleiotropic than many other types of mutations. Accordingly, Doebley and Lukens demonstrate that mutations in transcriptional regulators often have less pleiotropic effects than mutations in signal transduction genes. As a synthesis of the presented arguments, Doebley and Lukens propose that the *cis*-regulatory elements of transcription factors are likely to be most important in the evolution of novel phenotypes.

One type of transcription factors that is especially interesting from the perspective of morphological evolution determines organ specific gene expression. Mutations in such genes may result in homeotic changes, *i.e.* that one structure in an organism becomes replaced by another structure. One of the most exciting biological discoveries in the second half of the 20th century biology was that of the *Drosophila melanogaster* homeotic mutants *ultrabithorax* and *antennapedia* affecting genes that indeed encode transcription factors. Both these genes encode proteins with a conserved DNA binding motif, the homeodomain. Since, a large number of transcription factors with key regulatory roles possessing this motif have been isolated (for a review on these discoveries, see Gehring, 1994). Hence, within *D.*

melanogaster, and in other species, a family of homeobox genes exists, and different members of this gene family have different developmental roles.

Members of a gene family within a species are derived from a common ancestor through gene duplication events, and are referred to as paralogous genes (homologous by gene duplication, Fitch, 1970). The same paralogs in two different species are referred to as orthologous genes (homologous by speciation, Fitch, 1970).

A duplication event of a homeotic gene may lead to functional divergence between the resulting paralogs, a process that might be coupled to the formation of two dissociated developmental modules. For example, repeated morphological features, may end up being controlled by different developmental regulators, and hence evolve independently from each other. Lewis already in 1978 proposed a similar scheme for the evolution of the insect body from a centipede-like ancestor, by the diversification of homeotic genes. Although wrong in detail (see Akam *et al.*, 1994), Lewis general idea is still attractive. Thus, even if changes in *cis*-regulatory elements may very well be the main cause of phenotypic change, gene duplication and diversification do occur, and the diversification of the homeotic transcription factor encoding genes, may be connected to the origin of new morphological complexity.

Reconstructing the evolutionary history of a class of homeotic genes, may thus be an important contribution to the study of the evolution of developmental processes and hence of morphology.

1.i Interaction of floral homeotic genes, the ABC-model

Homeotically altered plants of different kinds have been known to botanists for centuries (*e.g.* Svedelius, 1909), and an important task for plant geneticists was to characterise the molecular genetic basis for homeosis in plants, as had been done in *Drosophila melanogaster.*

When the modern work on homeosis in Arabidopsis, *Arabidopsis thaliana* and snapdragon, *Antirrhinum majus* began, three important types of floral homeotic mutants were recognised, the A-, B- and C-mutants (*e.g.* Bowman *et al*., 1991; Coen and Meyerowitz, 1991). In the A-mutants, sepals and petals do not form. Instead, carpels and stamens respectively are formed in their place (but see below). In the B-mutants, sepals and carpels replace petals and stamens respectively, and in the C-mutants, petals and sepals replace stamens and carpels. The C-mutants furthermore have a “flower within a flower” phenotype, new whorls of petals, petals, sepals and so forth, are formed in the floral center. There is a difference between *A. majus* and Arabidopsis here: in Arabidopsis the inner sepals actually belong to the inner flower, whereas in *A. majus* a true fourth whorl of more or less sepaloid organs is present (for discussion, see Coen and Meyerowitz, 1991 and Davies *et al*., 1999). A model, based on these mutants, and of their double- and triple-mutant combinations, explains how the corresponding A-, B-, and C-function genes interact in a combinatorial manner to determine organ identity. A-genes direct sepal development, A+B genes direct petal development, B+C genes direct stamen development, and C-genes alone direct carpel development. Furthermore, A- and C-functions act as each others antagonists. In the A-mutant, C-function is ectopically active within the first two whorls, and vice versa (Coen and Meyerowitz, 1991).

The A function is somewhat loosely defined. In accordance with the ABC-model, the described A-mutant phenotype is found also in gain of function mutants of the *A. majus* C-class gene (Bradley *et al*., 1993; Lönnig and Saedler, 1994). Loss-of-function mutants in genes referred to as A-class genes may, however, also show other phenotypes. Mutant alleles of the A-class gene *APETALA2* (*AP2*) in Arabidopsis display a range of floral phenotypes, some conforming to the described A-mutant phenotype, but more common is the lack of second whorl organs (Bowman *et al.*, 1991). Plants carrying mutant alleles of the gene *APETALA1* (*AP1*) in Arabidopsis do not form sepals and petals, and it is hence an A-class gene, but bracts and secondary flowers, respectively, may be formed in their place (Mandel *et al*., 1991).

1.ii Floral homeotic genes belonging to the MADS-box gene family

In 1990 it was recognised that the homeotic genes *AGAMOUS* from Arabidopsis, and *DEFICIENS* from *A. majus*, both contained a region of 168 bp that was highly similar in sequence. Similar regions were also found in *MCM1* from *Saccharomyces cerevisiae*, and *SRF* from *Homo sapiens*. This element was called the MADS-box (Schwarz-Sommer *et al.*, 1990), an acronym referring to these four genes, in which it was first identified. MCM1 was known to be a DNA-binding transcription factor (Passmore, 1988), and evidence suggests that all MADS-box genes encode transcription factors. Of the genes that have been shown to act in the ABC-model of floral organ identity determination, all, but the A-gene *AP2* from Arabidopsis (Jofuku *et al.*, 1994), are MADS-box genes.

Two B-genes have been identified in *A. majus*, *DEFICIENS*, *DEF* (Sommer *et al.*, 1990) and *GLOBOSA, GLO* (Tröbner *et al.*, 1992), as well as in Arabidopsis, where the corresponding genes are called *APETALA3, AP3* (Jack *et al.*, 1992) and *PISTILLATA, PI* (Goto and Meyerowitz, 1994). On the other hand, there is only one C-class gene in Arabidopsis, *AGAMOUS* (*AG*) (Yanofsky *et al.*, 1990). *PLENA* (*PLE*) in *A. majus* confers the same function (Bradley *et al.*, 1993).

1.iii The expression patterns of MADS-box genes reflect their function

The expression patterns of MADS-box genes within the floral primordia reflect their function; for example C-genes are expressed in stamen- and carpel primordia, and later in developing stamens and carpels. The B-class gene expression patterns are also consistent with a role in the development of the organs, which are affected in their mutant phenotypes: *AP3* and *PI* are mainly expressed in whorls two and three. Constitutive expression of the floral homeotic genes in plants, show phenotypes expected from the ABC-model (Krizek and Meyerowitz, 1996; Mandel *et al*., 1992; Mizukami and Ma, 1992). For example ectopic expression in Arabidopsis of the C-class gene *AG* results in an A-mutant like phenoptype: sepals transformed to carpels in the first whorl and staminoid, or absent, organs in the second whorl (Mizukami and Ma, 1992).

1.iv Conservation of function between orthologous genes

Coen and Meyerowitz (1991) recognised that the floral homeotic genes with the same functions in Arabidopsis and *A. majus* were those that showed the highest sequence similarity. Of the genes mentioned above, *AG* is most similar in sequence to *PLE, AP3* to *DEF*, and *PI* to *GLO*. This suggests a conservation of function

between orthologous genes. Subsequent studies in *e.g.* Petunia gave consistent results. However, the Petunia *DEF*-ortholog *pMADS1*, was found to be essential for petal development only, but redundant for stamen development (Angenent *et al*., 1992; 1995; Kush *et al*., 1993; Tsuchimoto *et al*., 1993; van der Krol *et al*., 1993; van der Krol and Chua, 1993). Hence, although the ABC-model was found to be valid for a range of angiosperm species, differences in details are apparent.

Phylogenetic analyses on MADS-box genes from several angiosperm species confirmed that the MADS-box genes with the corresponding function in different species were orthologs. Thus, the gene duplication events resulting in the floral homeotic genes predate the origin of these angiosperm species (Doyle, 1994; Tandre *et al.*, 1995).

1.v The MADS-box genes constitute a large gene family in eukaryotes

A large number of MADS-box genes have been isolated in particular from flowering plants, and shown to function in various developmental processes (*e.g.* Ma *et al.*, 1991; Huang *et al.*, 1995; Rounsley *et al.*, 1995; Riechmann and Meyerowitz, 1997). MADS-box genes have also been isolated from animals and fungi (for a review on non-plant MADS-box genes, see Shore and Sharrocks, 1995). In animals, MADS-box genes of the *SRF* and *MEF2*-classes have roles in the control of cellular differentiation. MADS-box genes in the budding yeast, *Saccharomyces cerevisiae*, participate in the regulation of processes as diverse as arginine metabolism (*ARG80,* first described as *ARGR1*, Dubois *et al*., 1987) and the mating type switch (*MCM1*). MADS-box genes have not so far been described in protists or prokaryotes, although the *UspA*-like genes from *Escherichia coli* and *Haemophilus influenzae* have been suggested as possible relatives to MADS-box genes (for a discussion, see Alvarez-Buylla *et al.*, 2000; Theißen *et al.*, 2000).

1.vi The MADS-domain

The MADS-box is a DNA-sequence element, generally defined as an 168 bp motif, as mentioned, or as an 180 bp motif (*e.g.* Theißen *et al*., 1996), that encodes a DNA-binding protein motif, the MADS-domain.

The interaction between the MADS-domain and DNA is best understood in the human proteins SRF, and MEF2A, as well as the yeast protein MCM1, since the structures of these DNA-protein complexes have been determined (Pellegrini *et al.*, 1995; Tan *et al.,* 1998; Huang *et al.*, 2000; Santelli and Richmond, 2000). The proteins form homodimers that bind specifically to a sequence element called the CArG-box, with the sequence CC $(A/T)_6$GG (Phan-Dinh-Tuy *et al.*, 1988). The MADS-domains of the SRF, MEF2A, and MCM1 protein homodimers bind the center of the CArG sequence through one amphipatic α-helix from each monomer making contact predominantly with the minor groove. These α-helices contain some of the most conserved amino acids in the MADS-domain, indicating that DNA binding of different members within the gene family occurs in a similar manner. The binding causes a bending of DNA, which in the case of MCM1 has been shown to enhance interaction of other transcription factors with the adjacent major grooves. (Tan *et al*., 1998) The DNA-binding α-helices are also involved in dimerisation, together with a more C-terminal β-sheet structure (Pellegrini *et al.*, 1995, Tan *et al.*, 1998).

1.vii Subfamilies of MADS-box genes

Outside the MADS-box, MADS-box genes from plants, animals and fungi display different structural characteristics. Most MADS-box genes characterised from plants conform to an architecture that distinguishes them from known MADS-box genes in other organisms. These plant MADS-box genes possess a second sequence element, the K-box (named after similarities to keratin, Ma *et al.*, 1991), encoding a 65 aa motif. The K-box shows a high degree of conservation, although lower than the MADS-box. An intervening region, the I-region (also called the linker region, L-region), separates the K-box from the MADS-box. The I-region is less conserved than the K-domain and also varies somewhat in length, between 27 and 42 amino acids on the protein level (using the definition of the MADS-domain as consisting of 56 amino acids, Riechmann and Meyerowitz 1997). Downstream of the K-domain follows the C-terminal region which is the least conserved part of these transcription factors both in length and sequence (Ma *et al.*, 1991; Pnueli *et al.*, 1991; Riechmann and Meyerowitz, 1997).

Genes conforming to the described architecture are referred to as MIKC-genes (Münster *et al.*, 1997). Many MIKC-proteins have the MADS-domain at the N-terminal end, but in *AG*-like genes, an N-terminal region is present, typically extending 16 or 17 aa upstream the MADS-domain (Ma *et al.*, 1991; Mandel *et al.*, 1992; Tsuchimoto *et al.*, 1993; Davies *et al.,* 1999).

A number of *Arabidopsis* MADS-box genes, have been shown to lack a K-box, and most of these group in a clade separate from the MIKC-genes (Alvarez-Buylla *et al.*, 2000, III). A recent study suggests that this clade is most closely related to the genes *MCM1* and *ARG80* from yeast and the *SRF* genes from animals. The MIKC-genes, on the other hand, appear to be related to the *MEF2*-type genes from animals and the genes *YBR1245* and *RLM1* from yeast (Alvarez-Buylla *et al.*, 2000). These two clades are characterised by MADS-domains containing diagnostic amino acids, referred to as TypeI and TypeII MADS-domains. MIKC-genes encode TypeII MADS-domains. Hence, the split between genes with TypeI and TypeII MADS-boxes predate the separation of the lineage leading to fungi and animals from that leading to plants.

1.iix Functional domains of MIKC-genes

Like MCM1 and SRF, MIKC-proteins bind DNA as dimers at CArG-boxes. Both homo- and heterodimerisation of MADS-domain proteins have been demonstrated (*e.g.* Schwarz-Sommer *et al.*, 1992, Tröbner *et al.*, 1992; Huang *et al.*, 1993; Huang *et al.*, 1996, Davies *et al.*, 1996; Riechmann *et al.,* 1996). The MADS-domain and a region extending towards the C-terminal, variable in length from encompassing only the first amino acids of the I-region to a portion of the K-domain, constitutes the core that is needed for DNA-binding and dimerisation (Huang *et al.,* 1996; Riechmann *et al.*, 1996). This is similar to the situation in SRF, where the MADS-domain and the adjacent downstream region constitute the DNA-binding and dimerisation core of the protein (Pellegrini *et al.*, 1995).

The major function of the I-region appears to be that of determining dimerisation specificity, as demonstrated by domain swapping experiments with AP1, AP3, PI and AG (Riechmann *et al.*, 1996). To be precise, a region slightly larger than the I-region is needed for specific dimerisation, encompassing also the

C-terminal region of the MADS-domain (*i.e.* AG homodimerisation) or the N-terminal region of the K-domain (*i.e.* AP3/PI heterodimerisation).

A similar function was expected for the K-domain, since its predicted coiled-coil structure with hydrophobic faces implies a role in protein-protein interaction (Ma *et al.*, 1991). It has been demonstrated that the K-domain of AG is sufficient to mediate specific dimerisation with other K-domains (Fan *et al.*, 1997). However, the K-domains of certain proteins with which AG is known to dimerise could not interact in these experiments. Thus, the K-domain appears to be involved in dimerisation, but in a manner not well understood.

The highly variable C-region has not until recently been functionally characterised. In AP1, the C-terminal region has been demonstrated to function as a transcription activation domain (Cho *et al.*, 1999). The C-region appears to be of importance in the formation of a ternary complex between DEF/GLO heterodimers and SQUAMOSA homodimers in *A. majus* (Egea-Cortines *et al.*, 1999). It thus appears that the C-region is involved in the interaction with ternary factors, being either other MADS-domain proteins or factors closer to the basal transcriptional machinery.

1.ix Exon/intron organisation of MIKC-genes (I, II, IV)

Typically, the coding region of MIKC-genes contains six or seven introns, that I refer to as introns M, A, B, C, D, E, and F (I, II, IV). Intron M is situated within the I-region, close to the end of the MADS-box. Together with intron A, situated within the I-region, close to the 5' end of the K-box, this M defines the exon that corresponds to the major part of the I-region. The K-box of *AG*, and most other MIKC-genes is built up by three exons. In *AG* the exon defined by intron A and B is 62 bp long, that by B and C 100 bp long, and that defined by C and D 42 bp. The K-box is followed by an exon also spanning 42 bp, defined by introns D and E. Intron F is positioned further toward the 3'-end, in AG this intron is situated at the 3'-end of the coding sequence (Yanofsky *et al.*, 1990).

Certain genes may lack some of these introns, for example *PI* lacks intron M, *AGL5* lacks intron B, and *AGL2* lacks intron D. Genes possessing the same introns as *AG*, may differ in the exact positions of these. Some of these changes appear to be unique to a particular gene, but other are clade-specific. This indicates that the details in the exon/intron organisation may be useful as an independent test of the phylogenetic association of a particular gene. In this study, I focus on introns M, A, B, C, D and E (see fig. 1, see I, II, IV for sources of data).

Independent origins of the introns M, A, B, C, D and E is highly unlikely. A phylogenetic position of a gene that would imply loss and regain of any of these introns is hence likely to be incorrect. However, shifts in the exact positions of these introns may be prone to homoplasy. Alternative splicing was encountered among the genes presented in this report (II, IV) and a shift in intron position might hence in certain cases be regarded as a quantitavtive, rather than qualitative change: acceptor/donor sites used at low frequency become the "correct" sites. This simple mechanism suggests that identical shifts in intron positions may occur independently. Alternative splicing has also been documented to occur in other MIKC-genes, *e.g.* the *A. majus* gene *FAR* (Davies *et al.*, 1999).

The genes *AP1*, *SQUA*, and *AGL8* share a shift of intron A three bp toward the 5' end relative its position in *AG*. They further have intron M shifted three bases

toward the 3'-end relative *AG*. An identical shift in intron M occurs in *AGL9*, *AGL2*, *AGL4*, and *AGL3* that belong to a more inclusive clade together with *AP1*, *SQUA* and *AGL8*. This more inclusive clade also includes the closely related genes *AGL6* and *AGL13*, which, however, conform to *AG* in the position of intron M. The relation between the *AP1*, *AGL9* and *AGL6* subclades differs among analyses or is unresolved (*e.g.* I, II, III, IV). If the *AP1*-like position is derived, and the *AG*-like position ancestral, the intron M position in itself suggests a basal position for *AGL6*, because independent acquisitions of this position would otherwise be demanded in the *AP1*- and *AGL9* subclades. The present phylogenetic analyses indicate that identical shifts in the position of intron M have occurred several times. For example, the *AGL17* clade has members with both *AP1*- and *AG*-like intron M positions. In short, homoplasies in the shifts of intron positions exist to some extent. The close paralogs *AGL2* and *AGL4* are united in the lack of intron D. The paralogous genes *AGL17*, *ANR1*, *F2809.80* and *F20D10.60* are united by a dramatic shift of intron E, 42bp towards the 3'-end, doubling the size of the exon defined by D and E, relative *AG*. This shift is also present in their homolog *DEFH125* from *A. majus*.

The B-class genes *AP3*, *PI*, *DEF* and *GLO* are united by a shift of intron E three bases to the 3' end, and apart from *PI* that lacks intron M, a shift of intron M six bases toward the 3'-end (see section 2.v).

The *AG*-like pattern of intron positions occurs in genes closely related to *AG*: *AGL1*, and *AGL11*. Within the AG-clade some deviations are documented; *AGL5* lacks intron B.

The *AG*-like pattern also occurs in *AGL24*, *AGL6*, *F7H19.130*, and *AGL20* that do not appear to be closely related to *AG*, or (except *F7H19.130* and *AGL20*), to each other. That the *AG*-pattern is suitable as a reference is probably not a coincidence; it may represent an ancestral exon/intron organisation. In this work, data supporting this hypothesis is presented.

1.x Divergent MADS-box genes with a MIKC exon/intron pattern (IV)

The exon/intron organisations of genes encoding TypeII MADS-domains, *AGL12*, *AGL27* (=*F22K20.15*), *FLF* (=*AGL25*), but reportedly lacking K-boxes (Alvarez-Buylla *et al*., 2000), as well as of the divergent gene *F28K20.7* are similar to that of MIKC-genes. *AGL12* has a exon/intron organisation conforming to that of *AG*, regarding introns M, A, B, C, D, and E. The introns B, C, D, and E of *AGL27* also conform to the positions in *e.g. AG.* Intron A, however, is lacking in this gene, and intron M is shifted by four bp toward the 5'-end, compared to *e.g. AG*. In *FLF*, introns A, B, C, D, and E conform to the *AG*-pattern, whereas intron M is shifted by three bp toward the 3'-end in relation to it position in *AG*. In *F28K20.7*, introns B and E are missing, intron A is shifted by 12 bp toward the 5'-end, and intron D is shifted by 21 bp toward the 3'-end. Intron M is shifted by 6 bp towards the 3'-end in this gene. The position of Intron C, however, conforms to the corresponding position in *e.g. AG*.

The similarities in exon/intron organisation within the K-domain region (introns A-E) to typical MIKC-genes, show that the genes *AGL12*, *AGL27* (=*F22K20.15*), *FLF* (=*AGL25*) and *F28K20.7*, do have regions homologous to K-domains. However, the exon/intron pattern of *F28K20.7* is highly divergent.

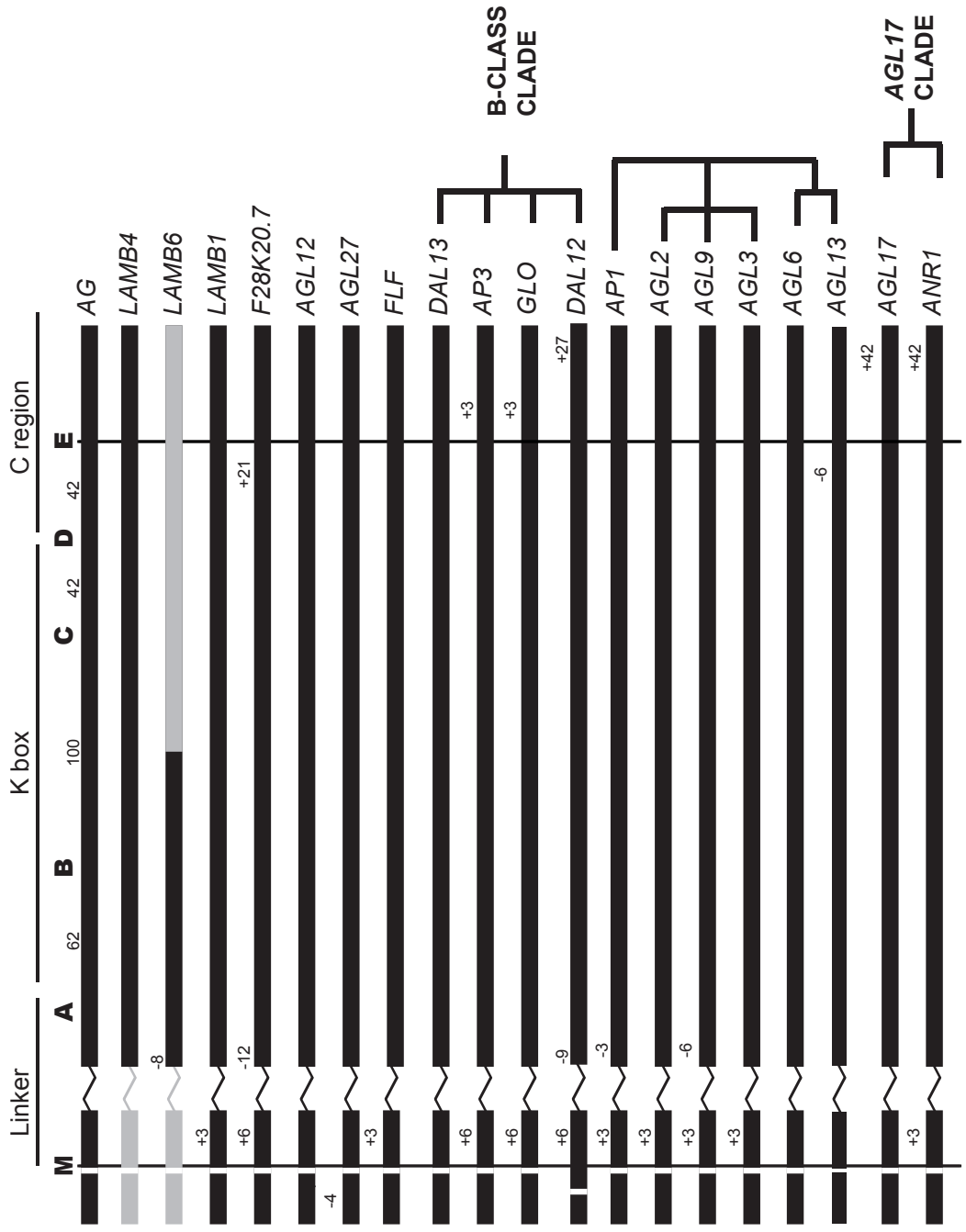

Figure 1. Exon/intron organisation of MIKC-genes. Intron positions are indicated by triangles. Exon lengths are indicated by numbers above the *AG*-sequence, and shifts in position relative *AG* by positive (downstream) or negative (upstream) numbers above each triangle, corresponding to a shifted intron. Clades are indicated to the right.

1.xi The role of MIKC-genes in angiosperm development

A large number of MIKC-genes have been isolated from angiosperms, and their function is best known in *Arabidopsis thaliana*, *Antirrhinum majus* (*e.g*, Schwarz-Sommer *et al.*, 1990; Huijser *et al.*, 1992; Tröbner *et al.*, 1992; Lönnig and Saedler, 1994; Zachgo *et al*., 1997; Egea-Cortines *et al.*, 1999), Tomato, *Lycopersicon esculentum* (*e.g.* Pnueli *et al.*, 1991; 1994; Hareven *et al.*, 1994), *Petunia* (*e.g.*, Angenent *et al.*, 1992; 1994; 1995; Kush *et al.*, 1993; Tsuchimoto *et al.*, 1993; van der Krol *et al.*, 1993; van der Krol and Chua, 1993; Colombo *et al.*, 1995), and *Gerbera hybrida* (Yu *et al.*, 1999). Unless otherwise stated, the following discussion will refer to work done in Arabidopsis.

Some MADS-box genes participate in the regulation of vegetative development. *AGL12*, *AGL14*, *AGL17* (Rounsley *et al.*, 1995), and *ANR1* (Zhang and Forde, 1998) have all been shown to be expressed in the root. Of these, *ANR1* is functionally best characterized and it is involved in the formation of lateral roots as a response to environmental [NO_3^-] (Zhang and Forde, 1998). The gene *AGL3* in Arabidopsis is expressed in all above ground organs (Huang *et al.*, 1995), but its function in these tissues is poorly understood.

A broad array of MADS-box genes are known to have regulatory functions in floral development, from the onset of flowering, the establishment of floral meristem identity, the determination of floral organ identity, as well as in fruit and seed development. One of the earliest changes in MADS-box gene expression related to the transition to flowering is the down regulation of *FLF* that encodes a repressor of flowering. *FLF* has been demonstrated to be regulated by genes in flowering induction pathways (Sheldon *et al.*, 1999).

The MADS-box gene *AP1*, together with other genes, mediates the determination of floral meristem identity. One of these other genes is *LEAFY*, that encodes a transcription factor unrelated to MADS-proteins (Weigel *et al.*, 1992). The *lfy* mutation shows transition of flowers into shoots, although abnormal flowers may arise late in development. The double mutant *lfy ap1* shows an enhanced vegetative phenotype (Bowman *et al.*, 1993). Plants ectopically expressing *AP1* flower early and have shoots transformed into flowers (Mandel and Yanofsky, 1995). Hence, *AP1* is important in the establishment of floral meristem identity. The gene *CAL* is very similar in sequence to *AP1*, and a *cal ap1* double mutant has, in the place of flowers, "cauliflowers" with a reiterated floral meristem (Kempin *et al.*, 1995), but may finally form *ap1*-like flowers (described above). A third MADS-box gene closely related to *AP1* and *CAL* is *FUL* (=*AGL8*). *FUL* is expressed in the inflorescence stem and meristem, as well as in the cauline leaves. In the floral primordia it becomes downregulated by *AP1*, that initially is expressed throughout the floral meristem (Gu *et al.*, 1998). A triple mutant *ap1calful* shows a very dramatic vegetative phenotype (Ferrandiz *et al.*, 2000), that under normal growth conditions never flowers. In short, transition to flowering involves a network of interacting transcription factors of MADS- and other types, showing partial redundancy.

At later stages, the onset of *AG* expression (the C-function gene) in the two inner whorls downregulates *AP1*, that specifically becomes expressed in whorl one and two. Here, *AP1* promotes sepal and petal development, and in this sense functions as an A-gene.

Temporal and spatial expression patterns of the gene *AGL9* in Arabidopsis, and the orthologous genes *TM5* in tomato (Pnueli *et al.*, 1994) and *FBP2* in *Petunia*, (Angenent *et al.*, 1994) suggest a function in the developmental sequence between the activity of floral identity genes and the B- and C- genes of the inner three whorls. In a recent study it was shown that *agl2agl4agl9* triple mutants form indeterminate flowers with all whorls occupied by sepals (hence the genes are also referred to as *SEPALLATA 1, 2, 3*), but that the initial expression of B- and C-genes remains unaffected in the triple-mutant (Pelaz *et al.*, 2000).

The gene *FUL* (*AGL8*), whose role in inflorescence- and flower development is mentioned above, also has a function in carpel development. The loss of function mutant show reduced growth of the siliques (Gu *et al.*, 1998).

The Arabidopsis MADS-box gene *AGL11* is specifically expressed in developing ovules and associated placental tissue (Rounsley *et al.*, 1995). In petunia, the *AGL11*-orthologs *FBP7* and *FBP11* have been shown to be necessary and sufficient, within the floral context, for the initiation of ovule formation, and have been proposed to represent D-function genes (Angenent *et al.*, 1995, Colombo *et al.*, 1995). Also other MADS-box genes are expressed in the ovule. The B-class gene *AP3* is expressed in the integuments, but viable seeds are produced by *ap3* mutants, and hence the gene is not essential for ovule development (Jack *et al.*, 1992).

1.xii Phylogeny of function in plant MADS-box genes

In the discussion above, examples of the conservation of function between orthologous genes were encountered, *e.g.* the function in the inner three whorls of the orthologous genes *AGL9*, *TM5*, and *FBP2*. Paralogous genes may also have similar functions. The partial redundancy in function between the close paralogs *AP1*, *CAL*, and *FUL*, as well as between the close paralogs *AGL2*, *AGL4* and *AGL9*, provide examples of gene duplication products that probably originally had identical functions, but that are undergoing functional divergence. The entire MADS box-gene family most likely has evolved in this manner, through gene duplication and subsequent functional divergence of the resulting genes. This leads to the question of what function(s) the common ancestor(s) of the different clades/functional groups recognised among angiosperm MIKC-genes had. Do orthologs to *e.g.* the C-class MADS-genes that define stamen and carpel organ identity genes exist only in plants with stamens and carpels, or do such genes exist also in other plants? When did the different paralogs of MADS-box genes defining structures unique to angiosperms evolve, or in more general terms, in what respect is the evolution of organ identity genes related to the evolution of morphological novelties?
Such questions constitute the basis for the work presented here.

2. MADS-box genes in gymnosperms (I, II)

A major breakthrough in the MADS-box gene research was the demonstration by Tandre *et al.*, 1995 that the conifer Norway spruce, *Picea abies*, contains MADS-box genes, which belong to the clades that correspond to functional groups in angiosperms. The gene *DAL1* was shown to group close to the *AGL6/ AGL2*-like genes, and *DAL3* grouped together with the gene *TM3* from tomato, a gene subsequently shown to be orthologous to *AGL14* in Arabidopsis (II). Most interesting, however, was the finding that *DAL2* appeared to be closely related to *AG*, and other angiosperm C-class genes.

MADS-box genes have later been isolated also from other conifer species, *e.g. Pinus radiata* (Mouradov *et al.*, 1996; 1997; 1998; 1999), some of, which are included in the phylogenetic analysis, presented in (II).

2.i *DAL2* – a Norway spruce C-class gene (I, II)

The gene *DAL2* (Tandre *et al.*, 1995) shows extensive sequence similarity to C-class MADS-box genes, *e.g. AG* from *Arabidopsis.* Within the MADS-domain (defined as 56 aa), only one amino acid differs between *DAL2* and *AG* (Tandre *et al.*, 1995). Phylogenetic analyses support the association between *DAL2* and the angiosperm C-class genes. As an independent test of the phylogenetic position of *DAL2*, a partial characterisation of its exon/intron organisation was conducted. This showed that the introns A, B, C, D and E conform to the *AG*-like positions. However, this pattern is likely to be ancestral among MIKC-genes and hence of limited phylogenetic significance (see below). *DAL2* is also closely related to the Arabidopsis genes *AGL1*, *AGL5*, *AGL11*, and their orthologs in other angiosperm species (Tandre *et al.*, 1995, I, II, III, IV). All these genes have in common a role in gynoecial development. *AGL11* is expressed in ovules (Rounsley *et al.*, 1995), *AGL1* and its close paralog *AGL5* are specifically expressed in the carpel (Ma *et al.*, 1991, Flanagan *et al.*, 1996, Savidge *et al.*, 1995).

The gynoecial expression of *AG* and all its close paralogs, and the expression of *DAL2* in female cones (I), suggest a conservation of a function in the development of ovuliferous structures between *DAL2* and the entire *AG*-clade. The expression of *DAL2* also in the pollen forming organs, the male cones (Tandre *et al.*, 1995, II), however, is specifically similar to the staminal expression of *AG* and its orthologs in other angiosperm species, but not to the *AG* paralogs within the clade (and their orthologs in other species).

In situ hybridisation experiments were made in order to get a morphologically detailed expression pattern of *DAL2* (I, II). *DAL2 in situ* hybridisation signal was detected in the ovuliferous scale, but no expression was found in the bract, subtending the ovuliferous scale, in the primary axis of the cone, or in the apical meristem (I). *DAL2* expression could be detected also at early stages of the development of the ovuliferous scale, in the axil of bracts, with no discernible scale primordia. Cones with these developing ovuliferous scales present close to the apical meristem also have developmentally older ovuliferous scales further toward the base. At late stages, *DAL2* expression was found to be restricted to the part of the scale on which ovules later will develop (I). *In situ* hybridisation experiments conducted on male cones (II) revealed expression of *DAL2* specifically in he pre-pollen mother cells on the abaxial side of the microsporophylls.

2.ii *DAL2* can act as a C-class gene in an angiosperm context (I)

To study if *DAL2* could act developmentally in a manner similar to *AG*, a construct, of *DAL2* positioned in the sense orientation downstream the 35S promoter was introduced in *Arabidopsis* via *Agrobacterium tumefaciens* mediated transformation (I). *DAL2* lacks the N-terminal extension characteristic of C-class genes in angiosperms, and in order to elucidate whether this motif is of importance for a functional C-class gene, a fusion construct between the N terminal of the *Brassica napus* AG-homolog, *BAG1*, and *DAL2* was made, called NDAL2.

Arabidopsis plants constitutively expressing *DAL2* constructs show either a weak phenotype with minor alterations in floral organ shape or a strong phenotype with homeotic alterations similar to those of plants transformed with a corresponding *AG*-construct (I). More specifically, ovule-like structures and stigmatic tissue were formed on the margins of the sepals, and the petals were transformed into filamentous or stamen-like organs. This was seen both in two plants transformed with the 35S::*DAL2* construct, and one plant transformed with the 35S::*NDAL2* construct.

Similar results were obtained in a separate study, where the *DAL2*-ortholog, from black spruce, *SAG1* (*Picea mariana*) was ectopically expressed in Arabidopsis (Rutledge *et al*., 1998). This indicates a highly conserved tertiary structure between DAL2 (and SAG1) and AG, suggesting that the context in which DAL2 acts during cone development, may be very similar to that of AG during flower development.

2.iii *DAL2* and the functional diversification of the *AG*-clade

Two genes that are structurally similar but that have distinct functions in different parts of the plant, may still confer the same phenotype when constitutively expressed. If constitutive expression of a sister gene to both these genes, derived from a different species, again result in the same phenotype, it can not be determined whether this sister gene acts in a developmental context similar to the first, or to the second gene.

Phylogenetic reconstructions suggest that *DAL2* is a sister gene to the entire *AG* clade, including *AGL1*, *AGL5*, as well as *AGL11* (I, II). *AGL1* and *AGL5* was only recently functionally characterised (Liljegren *et al.*, 2000). These two genes are closely related and show functional redundancy, with no dramatic single gene mutant phenotypes. An *agl1agl5* mutant lacks dehiscence zone development (and hence the genes are now known as *SHATTERPROOF 1* and *2*), but has no effect on floral organ identity. Constitutive expression of *AGL1* or *AGL5* causes floral homeotic alterations, similar to those observed to result from constitutive expression of *AG*, although their wild type function is restricted to the carpel.

Constitutive expression of *AGL1* and of *AGL5* results in homeotic alterations, that are similar to those resulting from constitutive expression of *AG*, in spite of the fact that *AGL1* and *AGL5* have developmental functions distinct from those of *AG*. This makes a conclusion of conserved function between *DAL2* and *AG*, based upon the same homeotic conversions somewhat less convincing. However, ectopic expression of *AGL1* and *AGL5* also results in alterations restricted to the fruit, and not seen upon ectopic expression of *AG*, or *DAL2*. This suggests that the function of *AG* and *DAL2* indeed may be conserved, whereas the *AGL1* and *AGL5* function may be derived.

The diversification of the *AG*-family within the angiosperms may reflect new developmental processes associated with the origin of the angiosperm gynoecium (I). It can not, however, at present be excluded that one or more of the paralogs may exert functions distinct from *AG*, but similar to functions of *DAL2* in spruce development. Hence the diversification within the *AG*-family in angiosperms may also reflect a "division of labor", rather than the origin of new functions. *AGL1* and *AGL5* appear to regulate an angiosperm-specific function, whereas the D-function of *AGL11* (as inferred from the D-function of the *Petunia* orthologs *FBP7* and *FBP11*, and the ovule specific expression pattern of *AGL11*) regulates the formation of a

structure shared between angiosperms and other seed plants. Accordingly, Theißen *et al.* (2000) proposed that *DAL2* may be considered to be a C/D- function gene, and the *AG/AGL11* split would thus represent the "division of labor" alternative. No expression of *DAL2* was observed in ovules however, which would be expected from this hypothesis.

The picture is still a bit more complex, because several species have been shown to contain more than one *AG* paralog. In *Petunia*, tobacco, and *A. majus*, respectively, the *AG* homologs *FBP6, NTPLE* and *PLE* form a clade separate from a second clade of *AG*-homologs, *PMADS3, NAG*, and *FAR*, from the same species (Davies *et al.*,1999, for similar examples in other species see *e.g.* Kater *et al*., 1998 and Theißen *et al*., 1995). These close paralogs may have different functions. For example, ectopic expression of *PLE* in tobacco plants gives carpels in whorl one and stamens in whorl two (Davies *et al*., 1996). Constitutive expression of *FAR*, however, generally leaves the first whorl unaffected, but causes a more extreme conversion of whorl two into stamens. The *ple*-mutant is a typical C-mutant, whereas the *far*-mutant does not lead to homeotic changes, but to male sterility (Davies *et al.*, 1999). This indicates that constitutive expression of *DAL2* in different angiosperm species may result in different phenotypes, and the choice of Arabidopsis, in this perspective is of course arbitrary.

2.iv Putative B-class genes from Norway spruce (II)

The phylogenetic analyses placing the conifer MADS-box genes *DAL1, DAL2* and *DAL3*, within clades of angiosperm genes, suggested that orthologs also to other angiosperm MADS-box genes, such as the B-class genes, should exist in conifers (Tandre *et al*., 1995, I).

Three conifer MADS-box genes *DAL11, DAL12* and *DAL13*, were isolated using a screen of a Norway spruce cDNA library, and/or RACE. In a phylogenetic analysis, based on MADS- and K-box sequences, the genes appear in the B-class gene clade (II). However the B-clade is unstable among analyses that differ in details of the matrix, as compared to, for example, the C-clade, that is less affected by such differences. This is also reflected in the low bootstrap support for the B-clade in the tree (II). In general, the comparatively low support for the B-class clade, may be attributed to a higher evolutionary rate within this lineage (see Purugganan, 1997).

An independent source of data that supports the association of *DAL11*, *DAL12* and *DAL13* to the angiosperm B-class clade is their possession of specific protein sequence motifs in the C-terminal region. These motifs were identified in an extensive comparative study of angiosperm B-class genes, and are referred to as the PI and AP3 motifs respectively (Kramer *et al.*, 1998). The AP3 motif can be divided into the EuAP3-motif and the PaleoAP3 motif. The PaleoAP3 motif (YGxHDLRLA) occurs in AP3-like genes of basal angiosperms, whereas the derived EuAP3-motif is a characteristic of “higher” eudicots (Kramer *et al.*, 1998). The PI motif (MPFxFRVQPxQPNLQE) is present in all angiosperm B-class genes (and hence supports the monophyly of these). DAL11, DAL12 and DAL13 all possess the PI-motif, whereas a PaleoAP3 motif is present only in DAL12. In itself, this suggests *DAL12* to be a member of the *AP3* clade, and *DAL11* and *DAL13* to be members of the *PI*-clade. Most phylogenetic analyses support a close relation between *DAL11* and *DAL13*, whereas the topology uniting these two genes, *DAL12* and the angiosperm B-class genes varies. The tree presented in (II) suggests that all

three spruce genes belong to the PI-clade, a position that would require the existence (or loss) of unknown *AP3*-like spruce genes, and hence is not very likely.

2.v Exon/intron organisation of *DAL11*, *DAL12*, and *DAL13* (II)

Another source of data we have suggested as an independent test of the phylogenetic position of MIKC-genes is details in the exon/intron organisation (see section 1.ix). Angiosperm B-class genes have a diagnostic exon/intron organisation. Compared to the positions in *AG*, intron M in *AP3*, *DEF*, and *GLO* is shifted six bases toward the 3'-end (*PI* lacks this intron). In *DEF*, *GLO*, *PI* and *AP3*, the exon between intron D and E has a length of 45 bp, rather than 42 bp, as in *AG*. In *AP3* and *DEF*, as well as in *AG*, the exon separating introns C and D is 42bp long. In *PI* and *GLO* this exon is 30 bp long, making introns D and E appear as shifted 12 bp toward the 5'end (I), unless a corresponding gap is introduced in the exon between intron C and D (II). In other words, *PI*, *GLO*, *DEF*, and *AP3* differ from *AG* in that intron E is shifted three bp toward the 3' end.

Whereas in *DAL13* all the introns M, A, B, C, D, and E, were characterised, in *DAL11*, only introns C, D, and E were covered. The exon/intron organisation of *DAL11* and *DAL13*, does not conform to that of angiosperm B-class genes, but rather to that of *AG*-like genes, including *DAL2* (I). This is in conflict with the position of *DAL11*, and *DAL13* among the B-class genes, only if the *AG*-like exon/intron pattern is a unique derived character (synapomorphy) of the C-clade, but not if it is a shared ancestral, plesiomorphic, trait among MIKC-genes. The latter is very likely, as will be discussed in greater detail below. Thus the exon/intron organisation does not provide any information as to whether *DAL11* and *DAL13* are related to the B-class genes, but given such a relation, it suggests that *DAL11* and *DAL13* are positioned at the base of the B-class clade. *DAL12* has a unique exon/intron organisation, where intron D is missing, intron A is shifted 9 bp to the 5'-end, and intron E is shifted 27 bases toward the 3' end, a position not shared with any other MIKC-genes. Unlike *DAL13*, *DAL12* has intron M at a position shared with the B-class genes, shifted six bases toward the 3'-end relative the corresponding position in *AG*. The position of intron M is the most variable among the discussed introns, and homoplasies at this position are not unlikely.

2.vi The expression of conifer B-class genes is restricted to male cones (II)

Of all tissues examined, including vegetative buds, cambium and female cones, *DAL11*, *DAL12*, and *DAL13* are expressed exclusively in pollen cones. The expression of *DAL11* is rather uniform in the developing male cone both prior to and after winter dormancy. *In situ* hybridisation experiments reveal that *DAL11* is expressed in the apical meristem, in the microsporophyll primordia and in the central pith of recently initiated pollen cones, but not in the subtending bract or in the basal "crown" region. At later stages, expression decreases in the pollen mother cells. *DAL12* and *DAL13*, however, show more restricted expression patterns, both temporally and spatially. Northern blot analyses show that the expression of *DAL12* is high in cones after meristem termination, before winter dormancy, but there appeared to be no expression when growth resumes in spring. *In situ* hybridisation experiments show that *DAL12* is predominantly expressed in the procambium. A Northern blot shows that the expression of *DAL13* decreases earlier before winter dormancy, at a stage earlier than that of *DAL12*, but unlike *DAL12*, the expression of

DAL13 increases again in spring. In pollen cones before apical meristem termination, *DAL13* was expressed in the apical meristem, and in the microsporophyll primordia. In the microsporophyll primordia, expression was lowest in the prepollen mother cells. After apical meristem termination, *DAL13* expression is high only in the tissues surrounding the developing pollen mother cells. After winter dormancy, the expression is restricted to the microsporangial wall.

Hence, the genes *DAL11*, *DAL12*, and *DAL13* are expressed specifically in developing male cones. This, together with their phylogenetic position close to the angiosperm B-class genes, indicates that the B-function in the determination of pollen producing reproductive structures is conserved between angiosperms and gymnosperms. The expression of *DAL2* also in male cones (Tandre *et al.*, 1995, II) suggests that also the interaction between B- and C-genes in the determination of pollen organ identity, might be conserved between gymnosperms and angiosperms.

2.vii B- and C-genes in *Gnetum*

A number of MIKC-genes have also been isolated the gnetophyte, *Gnetum gnemon* (Winter *et al.*, 1999). Two of these, *GGM3* and *GGM2* have been reported to be orthologs of C- and B-class genes respectively, and their expression patterns are consistent with such a notion. Interestingly, the *Gnetum* genes appear as being more closely related to Norway spruce genes, than to angiosperm genes, thus contradicting the anthophyte theory, where gnetophytes are considered the extant closest sister group of angiosperms. However, because of the instability of the B-class clade, among different analyses, it might be equivocal to draw such conclusions based upon the phylogenetic position of *GGM2.*

3. MADS-box genes in non seed plants, *e.g.* Lycopsids (III, IV)

When orthologs to the floral homeotic genes were shown to exist in gymnosperms, we asked whether these genes emerged in the common ancestor of seed plants specifically, or earlier in evolution. Purugganan (1997) assumed, from molecular clock estimates, that the divergence of the floral homeotic gene lineages dated back to approximately 480 mya, in the common ancestor of all land plants, including liverworts.

However, if the origin of floral homeotic genes is tightly coupled to the origin of ovules and pollen bearing organs, B- or C- orthologs would not be expected to be present in non-seed vascular plants, *e.g.* ferns and lycopsids. Alternatively, such orthologs may be present, but have other developmental roles, perhaps related to those in seed plants, such as in sporogenesis, or, alternatively, in the determination of male and female gemetophyte structures.

In order to investigate the MADS-box gene family in an earlier branch of the plant evolutionary tree, and throw light on the question of ancestral expression patterns of MIKC-genes, I here present the isolation of MIKC-genes from a lycopsid, *Lycopodium annotinum*.

3.i Lycopsids constitute the sister group to all other extant vascular plants

The earliest sporophytes of vascular plants known are those of rhyniophytes of the late Silurian. They have protostelic stems (similar to roots in ”higher” plants) that ultimately end in sporangia. They lack roots and leaves (Stewart and Rothwell, 1993; Kenrick and Crane, 1997). The Devonian gametophyte *Sciadophyton*, with its branching stem containing vascular tissue, ultimately ending in the reproductive organs, suggests that the earliest vascular plant may have had a close to isomorphic life cycle (Kenrick and Crane, 1997).

In the early Devonian, vascular plants were split in to two main groups. One group maintained the plesiomorphic condition of terminal sporangia, this is the branch leading to *e.g.* angiosperms. The other group bore lateral sporangia. Early representatives of this group, the zosterophylls, had in common with the rhyniophytes the absence of roots and leaves, and the stems were protostelic and branched dichotomously (Kenrick and Crane, 1997). The zosterophylls became extinct in the early Carboniferous (Kenrick and Crane, 1997). Today the only representatives of the lateral sporangiate group are the lycopsids, which appeared already in the lower Devonian (Kenrick and Crane, 1997). The lycopsids differ from the zosterophylls in that they have roots and leaf like structures, called microphylls.

The position of lycopsids as the sister group to all other extant vascular plants is supported by cladistic analyses on morphological datasets (Kenrick and Crane, 1997), as well as by molecular phylogenies based on *cox3* (Malek et al., 1996) and mitochondrial 19S rDNA sequences (Duff and Nickrent, 1999), and the gene order in a 30-kb region of the cpDNA. Lycopsids (including *Lycopodium*, *Selaginella* and *Isoetes*) share the gene order with bryophytes, whereas other vascular plants are united by the reverse gene order (Raubeson and Jansen, 1992).

We have chosen to isolate MADS-box genes from *Lycopodium annotinum*, the most common lycopsid in Sweden. Only sporophytes have been available for study. The shoot system of the *Lycopodium annotinum* sporophyte consists of a creeping main axis, which dichotomises unequally to produce erect side branches. After up to 7 or, in Arctic areas, 12 years of growth these terminate in the reproductive structure, a strobilus bearing sporangia (Kukkonen, 1994). Roots are formed continuously along the main axis.

3.ii *LAMB1* is a divergent MIKC-gene from *Lycopodium annotinum* (III)

A screen of a *Lycopodium annotinum* strobili cDNA library for MADS-box genes, using a fragment of a MADS-box isolated through PCR using degenerate MADS-primers, resulted in the isolation of a cDNA containing a region identical in sequence to the probe. The MADS-box gene, corresponding to the cDNA was named *LAMB1* (*Lycopodium annotinum* MADS Box gene 1). The MADS-box of *LAMB1* was found to be different from those previously identified, by the presence of an inserted amino acid within its C-terminal end. The exact position of this insertion is uncertain. It may be the A, P, A or T at positions 49, 50, 51, or 52, respectively. These positions are consistent with an insertion that would not disrupt the MADS-domain structure, because they correspond to a region that in MEF2A and SRF has been shown to form a loop within the β-sheet structure involved in MADS-domain dimerisation (Pellegrini *et al.*, 1995, Huang *et al.*, 2000, see also section 1.vi).

LAMB1 was shown to contain a region with high sequence similarity to K-domains. However, outside a central core region the similarity to other K-domains was low, and in some clustalW analyses gaps appeared within these less conserved regions. The I-region of *LAMB1*, with its length of 171 nucleotides (taking into account that the inserted amino acid within the MADS-domain suggests a 61 aa MADS-domain, rather than 60 aa, and assuming that no gaps occur within the 5'-region of the K-box) is longer than those of all other MIKC-genes.

PCR, using genomic DNA as a template was performed in order to investigate the exon/intron organisation of *LAMB1*. The analyses revealed seven introns. Of these, intron 1 is situated in the 5' untranslated region. Intron 2 is situated close to the end of the MADS-box, and hence corresponds to the conserved intron M of angiosperm and conifer MADS-box genes. The I-region contains three additional introns, compared to other MIKC-genes. Interestingly, Clustal W analyses at the amino acid level produce gaps in typical MIKC-genes corresponding approximately in position to these introns. Hence, instead of one exon as in typical MIKC-genes, four exons are residing entirely within the I-region, of 48, 13, 68, and 28 bp respectively. Intron 6 is situated in the I-region, nine bases 5' of the putative K-box, and thus corresponds exactly to the position of intron A in *AG*. This supports the alignment of the *LAMB1* K-domain, to other K-domains in the less conserved N-terminal region. Intron 7 is located within the K-box, and corresponds to the conserved intron B. The conserved introns C, D, E, and F that occur at the 3' end of the K-box, and within the C-terminal of other MIKC-genes appear to be absent in *LAMB1*. However, structural features of the C-terminal suggests that it starts after a K-domain of typical length (see below), with no gaps in the less conserved 3' end of the *LAMB1* K-box.

Is the exon/intron organisation of *LAMB1* primitive or derived? If the organisation of *LAMB1* is primitive, the shorter I-regions of other MIKC-genes may have evolved through the loss of exons by use of the acceptor/donor sites of two neighboring introns, including the intervening I-region exons into a larger intron. Intron M is very long in most MIKC-genes, and this might perhaps reflect such an evolutionary history of this intron.

Secondary structure prediction analyses of the I- as well as the K-region of LAMB1 revealed that this entire region is likely to have a mostly α-helical structure. Plotting the amino acids into a helical wheel showed that these helices would be likely to contain hydrophobic as well as hydrophilic faces, consistent with a function in dimerisation, as in other MIKC-genes (see section 1.iix).

Between amino acid positions 80 and 101 the I-region shows limited, but notable, similarities to a different group of plant transcription factors, the HD-ZIP proteins (Ruberti *et al.*, 1991; Mattsson *et al.,* 1992; Schena *et al.,* 1992). Although the similarity occurs in a region that is a putative amphipathic helix, the region is not leucine-zipper-like. Leucine zippers are built up by a series of amino acid heptads, each given a number starting from the N-terminal part. Within the heptads, the amino acids are given the letters a-g. Applying this nomenclature on the LAMB1 sequence, the following amino acids appear to be shared between *LAMB1* and the HD-ZIP genes (whose corresponding position is given within brackets, Mattsson 1995): the Q at position 80 (1c), the D at position 84 (1g), EY position 86-87 (2b-c), K position 89 (2e). The latter four amino acids are characteristic of the Class 2 HD-Zip genes, whereas the first two occur in all HD-Zip genes. Three additional amino

acids appear in a second constellation identical between LAMB1 and Class 2 HD-Zips, the E at position 92 (3b), the N at position 98 (4a) and L at position 101 (4d). The significance of this similarity is not clear, but may reflect a common evolutionary origin of these regions in *LAMB1* and the HD-Zip genes. It should be mentioned that a HD-Zip gene recently was isolated from *L. annotinum* (my unpublished results).

The region C-terminal to the K-box in LAMB1 spans 855 bp, and hence it is longer than in any other MIKC-gene known. In contrast to the C-termini of other known MADS-box genes, it is built up by repeated sequence elements. The entire C-terminus can be divided into four regions that upon alignment at the amino acid level show similarities to each other. These repeats 1 to 4 are of different lengths: repeat 1 is 38 amino acids long, repeats 2 and 3 are 99 amino acids long each, and repeat 4 is 49 amino acids long. No amino acid positions are completely conserved among all four repeats, but all repeats share sequence similarities with one or more of the other sequences.

Compared to plant MADS-domain proteins, long C-terminal regions are also found in the non-plant MEF2 class of MADS domain proteins. In MEF2C this long C-terminus has been shown to function as a transcriptional activation domain (Molkentin *et al.*, 1996). The C-terminus of AP1 has also been shown to function in transcriptional activation (Cho *et al.*,1999), and a similar role for the LAMB1 C-terminal is therefore possible. The region is glutamine-rich, a characteristic of certain transactivation domains (Latchman, 1998). Glutamine and asparagine together constitute 16% of the amino acids in the C-terminal region. Acidic amino acids comprise 14% of the C-terminal amino acids. However, preliminary experiments examining the capacity of fusions between the GAL4 DNA-binding domain and segments of the *LAMB1* cDNA to activate transcription of the *lacZ* reporter gene in a yeast system has not supported a transactivating function of the C-terminal part of LAMB1, and its function remains unknown.

3.iii Typical MIKC-genes from *L. annotinum* (IV)

Degenerate PCR on genomic DNA from *Lycopodium annotinum* was undertaken in order to isolate additional MADS-boxes. Based on the sequence of four fragments obtained in these degenerate PCR reactions, primers were designed and used in 3'-RACE experiments. Products from these experiments, in combination with subsequent 5'-RACE and RT-PCR experiments were used to reconstruct cDNA sequences of five novel MIKC-genes from *Lycopodium annotinum*: *LAMB2*, *LAMB3*, *LAMB4*, *LAMB5*, and *LAMB6*. Three of these genes, *LAMB2*, *LAMB4*, and *LAMB6*, conform very closely in architecture to seed plant MIKC-genes. The I-regions of LAMB2 and LAMB6 are 34 amino acids, and that of LAMB4 is 30 amino acids long. These lengths are very similar to for example the I-region of AG, that is 31 aa long, but different from the 57 aa I-region of LAMB1. LAMB3 and LAMB5 represent shorter proteins, but are highly similar in sequence to LAMB2, LAMB4, and LAMB6. The LAMB5 deduced protein ends eight amino acids downstream the MADS-domain. The LAMB3 deduced protein is longer, and includes a 30 aa I-region, as well as a region corresponding to the first exon of the K-box. A phylogenetic analysis confirms that *LAMB2*, *LAMB3*, *LAMB4*, *LAMB5*, and *LAMB6* are more closely related to each other, than to other MADS-box genes.

3.iv The phylogenetic position of the *L.annotinum* MADS box-genes (III, IV)

Phylogenetic analyses suggest that *LAMB1* may represent an early branch in the phylogenetic tree of MIKC-genes. One analysis (III) is based on a nucleotide matrix of MADS- and K-box sequences that includes a majority of the available Arabidopsis MADS-box genes, also those encoding what subsequently was defined as TypeI MADS-domains (Alvarez-Buylla *et al.*, 2000). The non-plant MADS-box genes included in this analysis were *MCM1*, and human and *Drosophila SRF*-genes, all encoding TypeI MADS-domains. In the analysis the Arabidopsis genes encoding a TypeI MADS-domain appear as a sister clade of the MIKC-genes. One of the K-box lacking genes, *TM021BO4.16* (*=AGL39*), that by Alvarez-Buylla *et al*., 2000, was suggested to possibly represent a "mixed" type MADS-domain, group among the MIKC-genes. In the tree, *LAMB1* is situated at a branch between the TypeI genes and the MIKC-genes, indicating a basal position of *LAMB1* among the latter.

The phylogenetic analysis in (IV) is based on an amino acid matrix of MADS- and K-domains. The analysis is based only on the deduced peptide sequences from genes encoding TypeII-domains, including *MEF2*-genes from man and *Drosophila*, as well as the yeast genes *YBR1245*, and *RLM1*. Also in this analysis, the consensus tree of which is shown in fig. 2, *LAMB1* appears at the very base of the MIKC-genes. *F28K20.7* appear at a branch, positioned between *LAMB1* and an unresolved clade including the rest of the MIKC-genes. *LAMB2, LAMB3, LAMB4, LAMB5*, and *LAMB6* appear as a clade (the *LAMB2* clade) with unresolved relation to other MIKC-genes. That *LAMB1* and the *LAMB2* clade appear at different positions in the phylogeny in relation to the fern and seed plant MIKC-genes indicates that the split between *LAMB1* and the *LAMB2* clade occurred in the common ancestor of lycopsids and other vascular plants. If the position of *LAMB1* at the very base of the MIKC-gene tree is correct, it may imply that *LAMB1* represents a lineage lost in the ancestor of seed-plants, *i.e.* Arabidopsis, the only plant from which a close to full set of MADS-box genes have been identified. Alternatively, orthologs to *LAMB1* exist, but the phylogenetic analyses have failed in identifying these.

In the analyses, sequences interpreted to represent K-boxes were included from genes that by Alvarez-Buylla *et al.* (2000), was interpreted not to encode K-domains. Arguments that these genes do possess regions homologous to K-boxes are presented in section 1.x., and in (IV). A highly derived putative K-box is found in the gene *F28K20.7.*

3.v Exon/intron organisation of *LAMB2, LAMB3, LAMB4,* and *LAMB6* (IV)

Two RT-PCR clones of *LAMB2* differed in length as compared to the other *LAMB2* clones. One clone contained an insertion of 41 bp at a position two bases downstream from the MADS-box, and the second contained a seven bp deletion at the corresponding position. This position corresponds to the intron M, matching the corresponding position in *AG* perfectly. These variants are hence likely to be products of alternative splicing. Sequences spanning over positions corresponding to the conserved introns were amplified from genomic DNA. However, no introns were present in the resulting PCR-products. This apparent paradox may be explained by the existence of both intron-less and intron containing *LAMB2* copies in the genome, the former being more easily amplified in these PCR-experiments.

The exon/intron organisation of the *LAMB4* gene was examined in a region spanning the 3'-end of the I-region and the entire C-region within the *LAMB4* sequence. The experiments revealed the existence of six introns, A, B, C, D, E, and F. An alignment with other MIKC-genes revealed that the positions of the introns A, B, C, D, and E perfectly matched those of the corresponding introns in *AG* .

This finding is consistent with the *AG*-like exon/intron organisation being an ancestral state among MIKC-genes, with important implications as regards the significance of using exon/intron organisation as a test of the phylogenetic position of MIKC-genes (see sections 1.ix, 1.x, and 2.v). The ancestral position of the *AG*-pattern is further supported by the fact that introns A-E, as well as intron M, in the phylogenetically diverse Arabidopsis genes *AGL6, AGL12, AGL15, F7H19.130, AGL20,* and *AGL24*, match this pattern. The ancestral state of intron M is equivocal, however, for example *LAMB1* conforms to the *AP1*-like, rather than the *AG*-like, position. The present phylogenetic analyses indicate that identical shifts in the position of intron M have occurred several times (see fig. 1).

LAMB3 encodes a shorter protein than *LAMB2, LAMB4,* and *LAMB6*, that does not contain a complete K-domain. Hence this gene lacks the introns B, C, D, and E, that in other genes reside within and downstream the K-box. However, intron A is present within the 3'-end of the I-region, at a position corresponding to that in *LAMB4*, and *AG*.

LAMB6 differs in its structure from *LAMB3* and *LAMB4*, but is similar to *LAMB2* in that it contains an insertion within the 3'-end of the I-region, corresponding to four amino acids. Interestingly, intron A resides within this insertion and is shifted eight bp toward the 5'-end relative the position in *LAMB4.*

3.vi A leucine-zipper in the K- and C- region of *LAMB4* (IV)

In *LAMB4*, the last exon of the K-box, and the first exon of the C-region, both 42 bp in length, together encode a leucine zipper-like structure. Corresponding leucine zippers, encoded by the corresponding exons appear also to exist also in other MIKC-genes, in *e.g. AGL14*, in Arabidopsis. This indicates that these two exons form a functional unit, and that the boundary between the K-box, and the C-region is somewhat artificially assigned, at least in this case.

3.vii Expression patterns of *L. annotinum* MIKC-genes (III, IV)

To study the expression of the *LAMB*-genes, main shoot apices, side shoot apices, strobili, and roots were collected at intervals of 5-20 days during the growth season, from May to September. Strobili were sectioned and examined with light microscopy in order to determine the developmental stage of the structure of the different dates. In June, masses of sporogeneous cells are present in the developing sporangia. In the samples from the July 10th, different stages of meiosis could be observed, and spore tetrads are found in samples from July 22nd.

Since the quantities of RNA obtained from several samples were low, RT-PCR was chosen to examine the expression patterns. The result showed a striking difference between *LAMB1*, on the one hand, and *LAMB2, LAMB4, LAMB5,* as well as *LAMB6,* on the other. Whereas expression of *LAMB1* could be detected in developing strobili only, the other genes showed broader expression patterns. *LAMB2*, *LAMB4,* and *LAMB6* were all more strongly expressed in vegetative tissues, including roots, than in strobili. Especially for *LAMB2,* the expression in

strobili appeared to be low. *LAMB5*, however, showed the strongest signal in strobili and main apices. The broad expression patterns of the *LAMB2*, *LAMB4*, *LAMB5*, and *LAMB6* can be taken to support the notion that the broad expression patterns of MIKC-genes, represents an ancestral condition (Theißen *et al*., 2000). The expression pattern of *LAMB1*, however, being restricted to strobili, does not only show that MIKC-genes with a more restricted expression pattern also occur in lycopsids. It also suggests that it may indeed be an ancestral condition, because *LAMB1* generally group at the very base of the MIKC-genes in phylogenetic analyses.

3.iix MADS-box genes in ferns

A number of genes have been isolated from ferns, mainly from *Ceratopteris richardii* and *C. pteroides*, which are leptosporangiate ferns, but also from the eusporangiate fern *Ophioglossum pedunculosum* (Münster *et al*., 1997; Hasebe *et al*., 1998). The *Ceratopteris* genes generally group into three distinct clades in phylogenetic analyses, the *CRM1*, *CRM3* and *CRM6*-like genes (Theißen *et al.*, 2000, note that *CRM6*=*CerMADS2*). None of the genes appear to be orthologs to any seed plant genes, but the fern clades generally appear as dispersed among the angiosperm clades, rather than forming a separate monophyletic group. Some structural features have been suggested to associate certain *Ceratopteris* genes with specific angiosperm genes. For example, *CRM6*-like genes have an N-terminal extension (Hasebe *et al.*, 1998; Theißen *et al*., 2000), as previously reported for genes closely related to *AG* (see section 1.vii). This is weak phylogenetic evidence, however, especially since *DAL2* lacks such an extension. The *CRM3* gene has on similarly rather weak grounds been suggested to be related to the B-class genes, because of its possession of a sequence element with some similarities to the PaleoAP3 Motif (Kramer *et al.*, 1998). Apart from the genes *OPM3* and *OPM4*, the *Ophioglossum* genes appear to be closely related to the *Ceratopteris* genes (Theißen *et al*., 2000). In contrast to most seed plant MADS-box genes, the *Ceratopteris* genes generally appear to have very broad expression patterns, including both sporophytic, and gametophytic tissue. *CRM9* and *CMADS1*, however, show mainly sporophytic expression (Hasebe *et al*., 1998; Theißen *et al*., 2000), and the expression of *CMADS4* within the sporophyte is concentrated to roots (Hasebe *et al*., 1998).

3.ix. Evolutionary considerations (III, IV)

The reproductive expression of *LAMB1* is interesting from the perspective that many of the Arabidopsis MIKC-genes appear to participate in processes related to floral development, and, thus, this expression pattern may reflect a retention of an ancestral function of MIKC-genes in reproductive development. The complex morphologies seen in vascular plants might be regarded as structures added on the sporophyte, from the ancestral state of a simple sporangium, as found in extant charophycean algae (see Albert, 1999). Hence, an ancestral function in reproductive development for a class of morphological key regulators in land plant sporophytes is an attractive hypothesis. However, *LAMB1* does not appear to be more closely related to angiosperm MIKC-genes with specific functions in reproductive development, than to genes involved in other developmental processes. On the contrary, *LAMB2*, *LAMB3*, *LAMB4*, *LAMB5*, and *LAMB6* may be more closely

related to the floral homeotic genes in Arabidopsis reproductive development, than *LAMB1*. Hence, the hypothesis that the angiosperm MIKC-genes, involved in reproductive development represents a retention of a primitive (though diversified) function shared with *LAMB1*, is at the present equivocal. An alternative hypothesis states that a broad expression pattern is a primitive feature of MIKC-genes, as have been previously proposed based upon the fern MIKC-gene expression patterns (Theißen *et al*., 2000). According to this hypothesis, the function of the MIKC-genes have evolved from controlling more general developmental processes, to more restricted ones, as the gene family has diversified (Theißen *et al*., 2000).

Furthermore, the fossil evidence suggests that the earliest vascular plants may have had a close to isomorphic life cycle (section 3.ii). It is hence very likely that sporophyte developmental programs were recruited from gametophyte specific programs in the common ancestor of extant charophycean algae, bryophytes, and vascular plants rather than invented exclusively within the sporophyte context. This implies that to a large extent, the same key regulatory genes probably were active in both the gametophyte and sporophyte of primitive vascular plants. The expression pattern in both phases of the life cycle of *Ceratopteris* may be a retention of this ancestral condition. If so, it is likely that also the *L. annotinum* MIKC-genes are expressed in the gametophyte generation. The species *Lycopodium cernuum* and *Lycopodium obscurum* have spores that are easy to germinate (Bruce, 1976; Whittier, 1976), and hence would provide an oppurtunity for studying the gametophyte generation.

4. Homeotic genes and organ homology

The expression patterns of MADS-box genes have been proposed to reflect organ homology (e.g. Doyle, 1994; Albert *et al*., 1998; Winter *et al*., 1999; Theißen *et al*., 2000). In this section, I will discuss this idea, first in general terms, and then its relation to the data presented in this work.

The concept of homology was first put forward by Richard Owen in the 19th century. Owen belonged to a school of morphologists that claimed that different morphologies are functional variations, of a non-functional archetypal idea, manifesting itself as "homology" between organs (see *e.g.* Young, 1992; Gould, 1993). Johann Wolfgang von Goethe, some decades earlier, supported a similar idea, that all the organs in a flower, the sepals, petals, stamens and carpels constitute modified leaves, by observations of homeotic mutants (see Singer, 1959). Thus, homologies were not only seen between organisms, but also, in serial homology, between structures within a single organism. Since Darwin (1859), homologies are defined as similarities in structure caused by a common ancestry (as opposed to structures being similar due to independent adaptation to a similar environment). In cladistic terms, homology is equivalent to synapomorphy. Homologies are corroborated when intermediate stages between structures, predicted from the homology proposition, are found, for example in the fossil record. A classical example is the intermediate forms between the reptilian jaw joint and the mammalian middle ear (Crompton and Parker, 1978). Various assignments of homology, especially in plants, have been highly controversial, for example the integuments of the ovule, the derived structures of grass embryos, the angiosperm floral organs (sepals, petals, stamens, and carpels), or the microphylls of lycopsids.

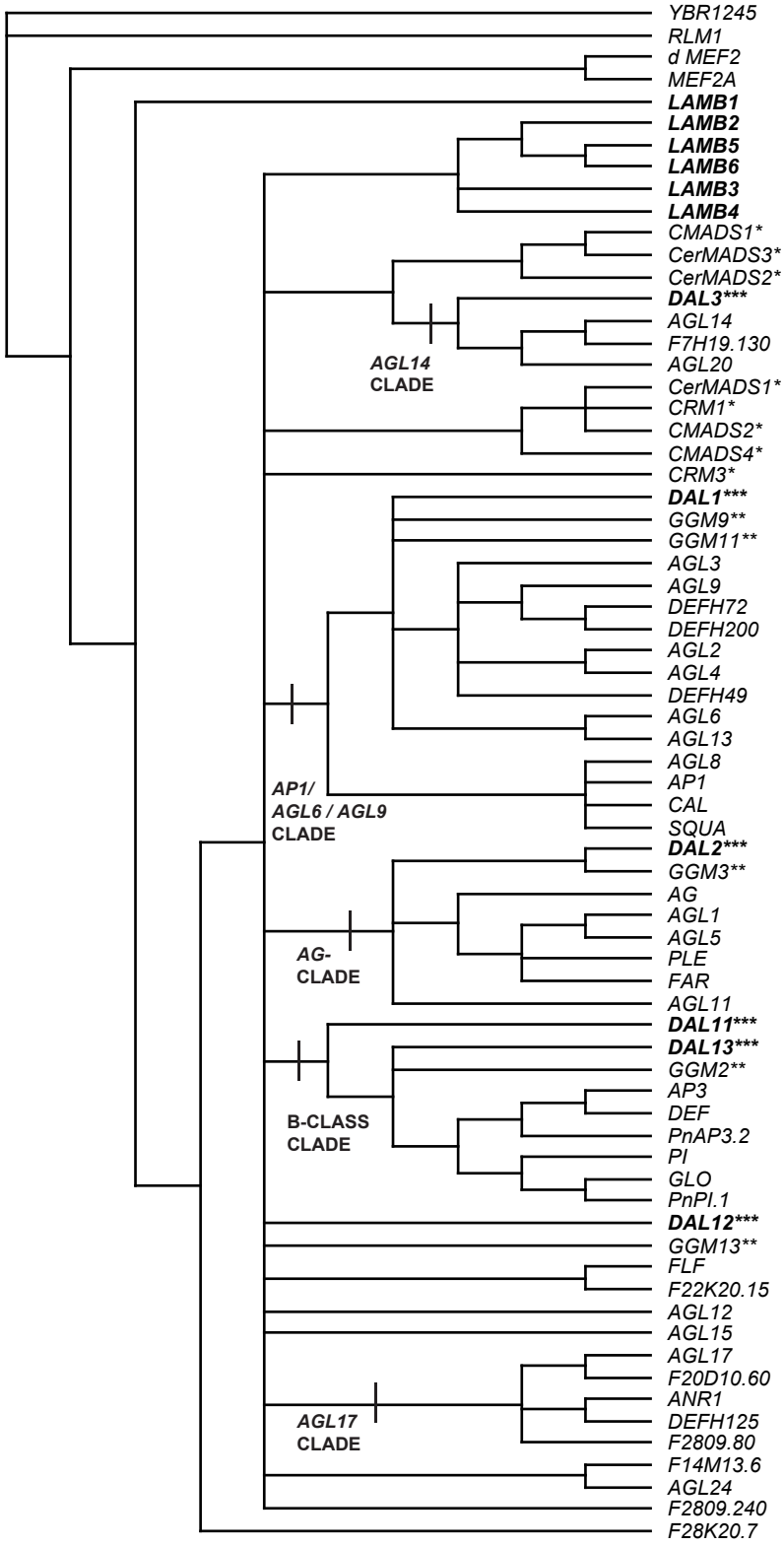

Figure 2. A phylogenetic analysis on the deduced peptide sequences of MADS-box genes from *Saccharomyces cerevisiae* (*YBR 1245* and *RLM1*), *Homo sapiens* (*MEF2*) and Drosophila melanogaster (*dMEF2*). *Lycopodium annotinum* genes are marked with bold face. *) MADS-box genes from *Ceratopteris richardii*. **) MADS-box genes from the gymnosperm: *Gnetum gnemon*. The *Picea abies* genes are bold and marked with ***. This analysis is presented in (IV).

4.i. Homology as a neo-Darwinian paradigm

The neodarwinian model of evolution, which implies that new form evolves through gradual modification of pre-existing form, suggests that homologies should be possible to identify. If we do not find the intermediate forms, this might be explained to an incompleteness of the fossil record. Several authors (*e.g.* McKinney and McNamara, 1991) have proposed that most morphological evolution in animals could be attributed to changes in the relative growth rate of body parts (heterochrony), a hypothesis consistent with homologous structures gradually changing in form. The hypothetical processes suggested to explain evolutionary changes in plant form (overtopping, recurvation, planation etc.) in the telome theory (Zimmermann, 1965) are conceptually similar. If, on the other hand, evolutionary novelties originated in a non-Darwinian, saltatory manner, structures could originate without being homologous to a pre-existing structure. This has sometimes been proposed to be of importance in "macro evolution". For example, Theißen *et al.* (2000), in their review on MADS-box gene evolution, assumes that "novel structures or complete new body plans" have originated in a non-Darwinian manner.

4.ii Homeotic genes and homology

If evolutionary changes at the morphological level are due to differences in relative growth of body parts, most evolutionary changes at the genetic level would involve regulatory genes affecting these processes in a quantitative manner. This probably involves mutations at *cis*-regulatory elements causing quantitative changes in gene expression, most likely confined to an organ specific module (see section 1). Changes often referred to as heterochrony, resulting in different proportions of the body in animals, or branching patterns in plants, may involve genes affecting a larger part of the organism, however. Such changes are likely to also involve expression levels of proteins directly participating in hormone production/sensation, and not only transcription factors.

Some components of a specific module are likely to be highly conserved among species, particularly those that are important for the dissociation of one module from other modules. Therefore these components may serve as indicators of modular/organ homology, and may reveal organ homology between highly diverged structures. Candidates for such components are the homeotic genes encoding transcription factors that determine organ identity. Homologous homeotic genes would thus be expressed in homologous organs. This logic is evidently applicable to orthologous genes, but it may as well be significant for paralogous genes. Serially homologous organs, or “paralogous” organs may be developmentally regulated by paralogous (or the same) genes.

4.iii Homeotic gene expression does not define morphological homology

The use of homologous genes to address organ homology may not always be straightforward. Homologous regulatory genes may act in homologous processes in non-homologous structures. This is illustrated by the orthologs to the gene *distal-less* being involved in the formation of non-homologous appendages in vertebrates, arthropods and echinoderms (Abouheif *et al.,* 1997; Panganiban *et al.,* 1997). So, even if *distal-less* may have a conserved, homologous, function in the formation of protruding structures, it can not be used to identify whether particular appendages are homologous.

Pax-6 homologs have been shown to regulate eye formation in both insects and vertebrates. This might support homology between all eyes. Gehring and Ikeo (1999) propose, in a discussion of general interest, that the evolution of eye-morphologies, have occurred through the intercalation of new developmental steps in an ancestral simple regulatory pathway in which *Pax-6* control rhodopsin-formation. Obviously, these intercalated steps would be different in different evolutionary branches. This clearly illustrates that an organ might be homologous at one level, but not another, and that use of homeotic genes to address homologies must be precise with respect to this level (for a discussion, see also Bolker and Raff, 1996). For example, the involvement of *Pax-6* homologs in formation of eyes in cephalopods and vertebrates does not indicate that the *camera architecture* of the two eyes is due to homology.

The notion that a number of developmental regulatory pathways show redundancy (*e.g.* Pickett *et al.,* 1995; Tauz, 1992) indicates that it is possible for a regulatory gene to be replaced by a non-homologous in evolution, via a state where they exist in parallel. The partially redundant interaction between transcription factors of unrelated types in the A-function and the determination of floral meristem identity, *LFY, AP2, AP1*, may represent a system with the potential to evolve in this manner. The existence of overlapping regulatory pathways in plants may be explained by the partly similar, partly distinct responses by plant development on environmental factors. If this explanation is correct, one might expect redundancy to be more frequent in developmentally plastic processes, such as onset of flowering, than in processes defining non-plastic traits, such as flower morphology. This hypothesis has not been tested. It is tempting to speculate that the A-function appear to be less conserved among angiosperm species (see *e.g.* Theißen *et al*., 2000), than the B- and C-functions, because it is more closely associated with the plastic process of the onset of flowering.

Homeotic changes have generally been considered not to be an important process in evolution, because such changes often have deleterious effects on the survival of the individual. Using a very broad definition of homeosis, including also cell-types, or physiological features (Sattler, 1988), homeosis has already been suggested to be a component in the evolution of novelties (as in the incorporation of sub modules within a module). However, it is very likely that homeotic changes in the more strict sense also have played some role in evolution. Homeosis may result in "hybrid", "chimeric" or "intermediate" organs making assignments of homology between structures difficult or even pointless (Sattler, 1988; 1991). Homeosis does not necessarily mean saltational, non-Darwinian evolution, however. In certain instances, a homeotic change might represent a slight morphological change in an adaptive perspective, and the formation of "intermediate" organs is conceivable as a gradual process. One example from the animal kingdom is the demonstration by Averof and Patel (1997) that the evolution of thoracic limbs into feeding appendages in crustaceans correlates with a reduced expression pattern of the homeotic genes *Ubx* and *AbdA*.

Because of the morphological simplicity and great plasticity of plant development, rearrangements between plant parts probably have a higher chance of being maintained in evolution compared to animals. Hempel *et al*. (1998) showed that primordia on the Arabidopsis stem, are not strictly committed to, but rather biased toward a particular developmental state. Continuous light was shown to be

able to induce floral development of already initiated primordia, that would otherwise form side-shoots at such a late stage, that chimeric structures were formed. Tomato plants grown at low temperature display floral homeotic changes, coupled to changes in floral homeotic MADS-box gene expression levels (Lozano *et al*., 1998). These two examples demonstrate that homeosis may lie within the reaction norm of wild type plant development.

The perianths of lilies (and several other monocotyledons) that contain two whorls of petal like organs (tepals) have probably evolved through the extension of the B-function into the first whorl (Theißen *et al*., 2000). In *Rumex* (a eudicot) the perianth has undergone the opposite transition into two sepaloid whorls, by withdrawal of the B-function from whorl two (Ainsworth *et al*., 1995). Hence both the lily, and the *Rumex* perianth appear to have evolved through homeosis (Theißen *et al*., 2000).

A striking example of cooption of a MADS-box gene to an apparently unrelated function must be mentioned; *NMH7* from alfalfa (*Medicago sativa*). This gene is phylogenetically nested within the homeotic B-class genes (Kramer *et al.*, 1998; Theißen *et al.*, 1996; II), and has a consistent exon/intron organisation (acc no. AF042068), but is involved in root-nodule formation (Heard and Dunn, 1995), a structure apparently with no homology to floral organs. It would be of great interest to see if there are any similarities in the downstream genes being regulated by *NMH7* and by floral B-class genes.

Gene-expression data in questions of morphological homology should be used as a tool for proposing/testing hypothesis, not as a definition. If two conflicting hypothesis state that structure A is homologous to structure B or structure C, respectively, the expression of an orthologous homeotic gene in A and B, but not in C, would support that A is homologous to B, but not to C.

4.iv Homologies between reproductive structures of seed plants

Pollen sacks and ovules in angiosperms are born on stamens and carpels respectively, both within the flower, while in conifers they are born on separate cones. The female cones of conifers consist of spirally arranged ovuliferous scales, each subtended by a bract. This organisation suggests that the bract may be a leaf, and the ovuliferous scale a branch in its axil. Florin (1951) supported such a view, by proposing homology between the ovuliferous scale with the ovuliferous branches of the extinct *Cordaitales*, a plant group that is known from the Carboniferous and Permian. The cordaitean female cones consisted of two orders of branches, of which the second bore ovules. According to Florin the ovuliferous scale had evolved from secondary branches that had undergone reduction in the number of ovules, from several to *e.g.* two. Recurvation of the ovules made their micropyle point toward, rather than away from, the primary axis. Finally, the fusion of the constituent parts of the secondary branch made it condense and scale-like, rather than forming a loose branching structure. In the *Voltziales* of the Carboniferous-Triassic periods, several possible intermediate stages in the evolution of a branching system to an ovuliferous scale were recognised.

The male conifer cones, on the other hand, consist of a simple determinate shoot. A bract subtends this shoot that bears spirally arranged microsporophylls. Since micro- and megasporangia (*i.e.* pollen and seed) are derived from a single monosporangiate structure, seed- and pollen cones may be homologous at some

level. The male cone, because of the branching order, have been proposed to be homologous to the ovuliferous scale, rather than to the entire female cone (Banks, 1972).

The origin of the angiosperm flower has remained obscure. It has not been possible to test the different variants and combinations of the two main hypothesis, whether the stamens and carpels are simple sporophylls or compound branching systems (the euanthium *vs.* pseudanthium hypothesis, see Friis and Endress, 1990, for a review).

4.v B- and C genes and seed plant reproductive organ homologies (I, II)

The discussion above on the B- and C-functions being conserved between angiosperms and conifers is based on their expression in organs on which homologous structures form; pollen and ovules respectively. The notion of a conserved, homologous function, is not based upon a proposal of homology between the organs *per se*.

The expression of *DAL11*, *DAL12*, and *DAL13* in the male cone, and *PI* and *AP3* orthologs in developing angiosperm stamens indicate homology between these organs. Because these *DAL* genes differ in their exact expression patterns (see above), the data do not provide suggestions as to whether the stamen is homologous to the microsporophyll, or to the entire male cone. Thus, the expression patterns do not provide arguments in the pseudanthium/euanthium controversy.

The expression of *DAL2* in the ovuliferous scale (II) might be taken as an indication of homology between this structure and the angiosperm carpel. This would support the pseudanthial theory of carpel origin: the carpel being homologous to a branching structure. The interpretation is problematic, however, because of the diversification of the *AG*-family within angiosperms (see above).

Extending the comparison to serial homologies, the expression of *DAL2* in the microsporophyll of the male cone (II), and in the ovuliferous scale of the female cone (I), might be taken to suggest homology between these structures. This is in apparent conflict with the ovuliferous scale being homologous to the entire male cone, rather than to the microsporophyll. However, late *DAL2* expression is strongest in the part of the ovuliferous scale on which ovules are to be formed (I), and this may be the region homologous to the microsporophyll. According to one, highly speculative, interpretation consistent with both Florin and this homology proposition based on the *DAL2* expression, the ovuliferous scale may be a fused organ between a branching structure and a sporophyll (II).

If the ancestral function of a C-class gene, or C/D-class gene (Theißen *et al.*, 2000) is to promote ovule development (or pollen development in concert with B-class genes) on the developing organ on which it is expressed, evolutionary change in the expression pattern of the C-class gene may confer this property to an other part of the plant. Such a homeotic change needs not to be deleterious to the plant, and may have played a role in the evolution of seed plant reproductive structures. A disjunction between orthologous gene expression and organ homology is conceivable in this case, *i.e.* the pollen- and ovule bearing organs of conifers and angiosperms may well be non-homologous, although they express particular orthologous homeotic genes.

4.vi ABC-genes and floral organ homology

Inference of serial homology within the angiosperm flower from MADS-box gene expression would suggest that stamens and carpels are homologous (C-function), that stamens and petals are homologous (B-function) and that petals and sepals are homologous (A-function). Homology, at some level of organisation, between stamens and carpels, like the male and female cone, is reasonable. Micro- and macrosporangia of seed plants are likely to be derived from one single type of sporangium in a homosporous ancestor. Fossil progymnosperms support this hypothesis (Stewart and Rothwell, 1993).

A- and B-gene expression, indicate that core eudicot petals may be homologous to both sepals and stamens, and hence may be "hybrid" organs. Baum (1998; 1999) presented a similar hypothesis, that from a primitive flower with a sepal-like perianth, petals may have evolved through the extension of B-function into the perianth. Albert *et al.* (1998) state that petals are homologous to both sepals and stamens, but not clearly as a hypothesis of evolution (they use their term "process homology"). Instead, they speculate that the basal angiosperm flower had a simple, petaloid perianth expressing both A- and B-genes. It now appears as if the genus *Amborella* is the most basal angiosperm identified, followed by water lilies, *Nymphaeaceae* (Matthews and Donoghue, 1999; Soltis *et al.,* 1999; Qiu *et al.,* 1999). This supports a simple perianth as a basal condition among angiosperms, and in at least *Nymphaeaceae* it is clearly petaloid.

The hypothesis that some basal angiosperms may have petals derived from sepaloid organs (bracteopetals, Thaktajan, 1991), but unrelated to stamens, would be corroborated if an absence of B-function in these could be proven.

B-mutants in the grasses, maize and rice, conform to the ABC-model (Schmidt *et al.,* 1998; Kang *et al.,* 1998; Theißen *et al.,* 2000). This is consistent with homology between the derived structures of the grass flower, the palea and the lodicule, and sepals and petals respectively. It further suggests that the ABC-system was established before the split between eudicots and monocots.

A mutant form of *Anemone nemorosa* fails to form petals and stamens, but instead cauline leaves are present in their place (Svedelius, 1909). This is probably a B-mutant (Svensson, 1996), with the cauline leaves behaving like sepals. These cauline leaves may be homologous to sepals, determined by an A-function organ identity gene. These sepals may have become enlarged, possibly by recruiting developmental pathways specifically involved in the formation of vegetative leaves, enabling floral photosynthesis in this rhizomic plant. Alternatively, the organs may be non-homologous to sepals, but are ordinary vegetative leaves, that form ectopically in the mutant flower in the absence of other organ identity signals. Detailed study of this, and other known mutant forms of *A. nemorosa*, may provide information on the ABC-system within basal eudicots, which recently has been given some attention by expression studies of B-class genes (Kramer and Irish, 1999).

The B-class gene *AP3* is expressed in the integuments (see section 1.xi.). The significance of a homology proposal based on this expression patterns depends on whether the regulatory function of *AP3* in ovule development is distinct from, or related to that in flower development. If the latter is the case this may be a reflection of flowers and ovules having evolved as developmental modules interconnected as a reiterative series (see Albert *et al.,* 1998).

4.vii Lycopsid microphyll homology

Homology between microphylls of lycopsids and megaphylls (leaves of ferns and seed plants) is in conflict with their phylogenetic distribution (Crane and Kenrick, 1996; Kenrick and Crane, 1997). Microphylls are generally believed to have evolved from enations, as found on the stems of several zosterophyll genera (Bower 1935, reviewed in *e.g.* Gifford and Foster, 1989; Stewart and Rothwell, 1993; Raven *et al.*, 1999). A recent hypothesis is that microphylls evolved from sporangia. Microphylls and sporangia, unlike enations, show similar distribution on the stem and are vascularized. Furthermore, enations are not found on the zosterophylls (a paraphyletic group) most closely related to lycopsids (Crane and Kenrick, 1996; Kenrick and Crane, 1997). The hypothesis that microphylls and sporangia are homologous appears to be the most likely at the present.

Studies on genetic mechanisms for organ identity determination may contribute to the evaluation of these conflicting hypothesis. For example, is the development of microphylls controlled by genes homologous to those that control development of true leaves or to those that control sporangial development?

5. Concluding remarks and suggestions

The present data it suggests that the floral homeotic B- and C-class genes emerged in the common ancestor of all seed plants, but after this lineage split from ferns, the closest extant relatives of seed plants. The expression patterns of the conifer B-class orthologs *DAL11*, *DAL12*, *DAL13* are restricted to male cones. The conifer C-class ortholog *DAL2* is expressed in both male and female cones. These data suggest that the function of B- and C-class genes as key regulators of male- and female sporophytic structures respectively, was established in the common ancestor of extant seed plants.

The structurally divergent MIKC-gene *LAMB1*, isolated from the club moss, *Lycopodium annotinum*, may represent one of the earliest lineages in the MIKC-gene family. This gene has an expression pattern restricted to strobili, suggesting that a function in reproductive development may have been ancestral for MIKC-genes. The genes *LAMB2*, *LAMB4*, *LAMB5*, and *LAMB6* from *L. annotinum* have a more typical MIKC-organisation. These genes have broad expression patterns, like MIKC-genes from ferns. It remains unclear whether the reproductive expression patterns of floral homeotic genes represent an ancestral condition shared with *LAMB1*, or whether the valid hypothesis is that the diverse functions of MIKC-genes are derived from more general developmental roles, reflected by broad expression patterns.

Different clades of MIKC-genes have diagnostic details in their exon/intron organisation. *DAL2* conforms to other C-class genes (*e.g. AG*) in this respect. However, also *DAL11* and *DAL13* conform to this pattern, a finding that is not in conflict with the association of these genes to the B-class genes, if the pattern is ancestral. *LAMB4* has an exon/intron organisation that in its details resembles that of *AG*, indicating that this organisation might be ancestral.

The basal branching pattern of MIKC-genes is equivocal, perhaps due to a very rapid rate of gene duplication events occurring simultaneously with the separation of the major lineages of land plants. Isolation and characterisation of MIKC-genes from lineages branching off earlier in the evolution of plants, such as

bryophytes and charophycean algae, may help both to resolve the phylogeny and to address the structural uniqueness of *LAMB1*. Isolation of genes from other lycopsids, such as *Selaginella* that belong to a lineage separated from *Lycopodium* since the Devonian (Stewart and Rothwell, 1993) may also be informative. The basal branching pattern of the MADS-box gene family as a whole is also problematic: for example, TypeI MADS-box genes need to be isolated from more plant species, and protists need to be examined.

Within a not too far future, we anticipate the genomes, or very inclusive cDNA-collections, of several grasses (maize, rice, and other crop species), commercially important eudicots, including a range of floral morphologies and growth habits, from weeds to trees. Phylogenetically important species may be thoroughly investigated because their simple morphologies make them useful model systems, such as the moss *Physcomitrella*, the liver wort, *Marchantia*, or algae, such as *Chlamydomonas* and *Volvox*. Hence we might expect as large samples of MADS-box genes, from several species, as at the present is only available in Arabidopsis. Furthermore, expression studies of these genes, and other functional analyses, will be conducted at a very different scale. The study of evolution of developmental mechanisms and its connection to the origin of morphological novelties is a field that will be understood at much greater depth than at the present.

Acknowledgements
I would like to thank professor Peter Engström for taking me on as a PhD student and sustaining the funding of the projects.

I deeply want to thank Charlotta Thornberg for excellent technical assistance and fikapauser.
I am also greatful to Afsaneh, Eva and Marie, who have been sequencing quite a lot of samples from me. Thanks!

Jan Bergsten, Anders Rådén and Anders Lindholm, thanx for driving the car to the forest!

All the Funky Lab/ MADS-people are thanked:
Karolina Tandre, Annelie Carlsbecker (thanks a lot for the support!), Jens Sundström, Anders Kvarnheden, Victor Albert, Deyue Yu, Liz Izquerdo, Henrik Johannesson, Yan Wang, and Jelka L.

All the other PhD students are of course warmly thanked for putting some color to the world and so on:
Alessia, Anna, Coby, Eva, Joel, Johannes, Katarina, Mattias (thanks for support and interesting discussions, and the very much needed boat trip), Ingela, and Sandra.

I also thank Annika Sundås-Larsson and Eva Sundberg, as well as…
…I simply thank the present and former workers at the department, in particular Cellskapet. All my friends at the other departments at EBC are warmly thanked.

Birgitta is thanked for taking care of us all.
Gun-Britt is thanked for her contribution to the courses.
Agneta is thanked for taking care of *e.g.* the shopping list.
Marie Svenson, sorry Marie Englund, is thanked for always being nice and supporting.
Thanks to Jin-Long for the calligraphy kit!
Thanks to Tage for his interest in knowledge.

Björn and Kauko, thank you.
I thank my long time biology friends,
Tim Olsson (thanks for all interesting e-mails!), Olle Israelsson (thanks for the *Xenoturbella*-adventures), Anders Normann, and Anders Lindroth.

I thank all my new and old friends, and my family…

This work is dedicated to the memory of my mother, Gunnel Svensson.

REFERENCES

Abouheif, E., Akam, M., Dickinson, W. J., Holland P. W. H., Meyer, A., Nipam, H.P., Raff, R. A., Roth, V. L., Wray, G. A. (1997) Homology and developmental genes. *Trends Genet.* **13**:432-433.

Ainsworth, C., Crossley, S., Buchanan-Wollaston, V., Thangavelu, M., Parker, J. (1995) Male and female flowers of the dioecious plant sorrel show different patterns of MADS box gene expression. *Plant Cell* **7**:1583-1598.

Akam, M., Averof, M., Castelli-Gair, J., Dawes, R., Falciani, F., Ferrier, D. (1994) The evolving role of Hox genes in arthropods. *Dev Suppl* **1994**:209-215.

Albert, V. A. (1999) Shoot apical meristems and floral patterning: an evolutionary perspective. *Trends Plant Sci* **4**:84-86.

Albert, V. A., Gustafsson, M. H. G., Di Laurenzio, L. (1998) Ontogenetic systematics, molecular developmental genetics, and the angiosperm petal. In: Soltis DE, Soltis PS, Doyle JJ (eds), Molecular Systematics of Plants II, pp. 349–374. Kluwer Academic Publishers, Boston.

Alvarez-Buylla, E. R., Pelaz, S., Liljegren, S. J., Gold, S. E., Burgeff, C., Ditta, G. S., Ribas de Pouplana, L., Martinez-Castilla, L., Yanofsky, M. F. (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. U S A.*, **97**:5328-5333.

Angenent, G. C., Busscher, M., Franken, J., Mol, J. N., van Tunen, A. J. (1992) Differential expression of two MADS box genes in wild-type and mutant petunia flowers. *Plant Cell* **4**:983-993.

Angenent, G. C., Franken, J., Busscher, M., Weiss, D., Van-Tunen, A. J. (1994) Co-suppression of the petunia homeotic gene *fbp2* affects the identity of the generative meristem. *Plant J.* **5**:33-44.

Angenent, G.C., Franken, J., Busscher, M., van Dijken, A., van Went, J. L., Dons, H. J., van Tunen, A. J. (1995) A novel class of MADS box genes is involved in ovule development in petunia. *Plant Cell* **7**:1569-1582.

Averof, M., Patel, N. H. (1997) Crustacean appendage evolution associated with changes in Hox gene expression. *Nature* **388**:682-686.

Banks, H. P. (1972) Evolution and plants of the past. Macmillan press, London.

Baum, D. A. (1998) The evolution of plant development. *Curr. Opin. Plant Biol*. **1**:79-86.

Baum, D. A., Whitlock, B. A. (1999) Plant development: Genetic clues to petal evolution. *Curr. Biol.* **1999**:R525-527.

Bolker, J. A., Raff, R. A. (1996) Developmental genetics and traditional homology. *Bioessays.* **18**:489-494.

Bower, F. O. (1935) Primitive land plants. London: Macmillan.

Bowman, J. L., Smyth, D. R., Meyerowitz, E. M. (1991) Genetic interactions among floral homeotic genes of Arabidopsis. *Development* **112**:1-20.

Bowman, J. L., Alvarez, J., Weigel, D., Meyerowitz, E. M., Smyth, D. R. (1993) Control of flower development in Arabidopsis thaliana by *APETALA1* and interacting genes. *Development* **119**:721-743.

Bradley, D., Carpenter, R., Sommer, H., Hartley, N., Coen, E. (1993) Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *plena* locus of *Antirrhinum. Cell* **72**:85-95.

Bruce, J. G. (1976) Gametophytes and subgeneric concepts in *Lycopodium*. *Amer. J. Bot.* **63**:919-924.

Cho, S., Jang, S., Chae, S., Chung, K. M., Moon, Y. H., An, G., Jang, S. K. (1999) Analysis of the C-terminal region of *Arabidopsis thaliana APETALA1* as a transcription activation domain. *Plant Mol. Biol.* **40**:419-429.

Coen, E.S., Meyerowitz, E.M. (1991) The war of the whorls: genetic interactions controlling flower development. *Nature* **353**:31-37.

Colombo, L., Franken, J., Koetje, E., van Went, J., Dons, H. J., Angenent, G. C., van Tunen, A. J. (1995) The petunia MADS box gene *FBP11* determines ovule identity. *Plant Cell* **7**:1859-1868.

Crane, P. R., Kenrick, P. (1997) Diverted development of reproductive organs: a source of morphological innovation in land plants. *Pl. Syst. Evol*. **206**:161-174.

Crompton, A.W., Parker, P. (1978). Evolution of the mammalian masticatory apparatus. *American Scientist*. **66**:192-201.

Davies B., Egea-Cortines, M., de Andrade Silva, E., Saedler, H., Sommer, H. (1996) Multiple interactions amongst floral homeotic MADS box proteins. *EMBO J.* **15**:4330-4343.

Davies, B., Motte, P., Keck, E., Saedler, H., Sommer, H., Schwarz-Sommer, Z. (1999) *PLENA* and *FARINELLI*: redundancy and regulatory interactions between two *Antirrhinum* MADS-box factors controlling flower development. *EMBO J.* **18**:4023-4034.

Doebley, J., Lukens, L. (1998) Transcriptional regulators and the evolution of plant form. *Plant Cell* **10**:1075-1082.

Doyle, J.J. (1994) Evolution of a plant homeotic multigene family: toward connecting molecular systematics and molecular developmental genetics. *Syst. Biol.* **43**:307-328.

Dubois, E., Bercy, J., Messenguy, F. (1987) Characterization of two genes, *ARGRI* and *ARGRIII* required for specific regulation of arginine metabolism in yeast. *Mol. Gen. Genet*. **207**:142-148.

Duff, R. J., Nickrent, D. L. (1999) Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. *Am. J. Bot*. **86**:372-386.

Egea-Cortines, M., Saedler, H., Sommer, H. (1999) Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus. EMBO J.* **18**:5370-5379.

Endress, P. K. (1994) Floral structure and evolution of primitive angiosperms: recent advances. *Pl. Syst. Evol.* **192**:79-97.

Fan, H. Y., Hu, Y., Tudor, M., Ma, H. (1997) Specific interactions between the K-domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. *Plant J.* **12**:999-1010.

Ferrandiz, C., Gu, Q., Martienssen, R., Yanofsky, M. F. (2000) Redundant regulation of meristem identity and plant architecture by *FRUITFULL, APETALA1* and *CAULIFLOWER*. *Development.* **127**:725-734.

Fisher, R. A. (1930) The genetical theory of natural selection. Calerdon press, Oxford.

Fitch, W. M. (1970) Distinguishing homologous from analogous proteins. *Systematic Zoology* **19**:99-113.

Flanagan, C. A., Ma, H. (1994) Spatially and temporally regulated expression of the MADS-box gene *AGL2* in wild-type and mutant Arabidopsis flowers. *Plant Mol. Biol.*, **26**:581-95.

Flanagan, C. A., Hu Y., Ma H. (1996) Specific expression of the *AGL1* MADS-box gene suggests regulatory functions in Arabidopsis gynoecium and ovule development. *Plant J.* **10**:343-353.

Florin, R. (1951) Evolution in cordaites and conifers. *Acta Horti Bergiani* **15**:285-388.

Friis, E. M., Endress, P. K. (1990) Origin and evolution of angiosperm flowers. In: *Advances in botanical research*. Ed. J. A. Callow. **17**:100-162.

Gehring, W. J. (1994) A history of the Homeobox. In: Denis Duboule (ed.) Guidebook to the homeobox genes. pp. 3-9. Oxford university press, New York.

Gehring, W. J., Ikeo K. (1999) *Pax 6*: mastering eye morphogenesis and eye evolution. *Trends Genet.* **15:**371-377.

Gifford, E.M., Foster, A. S. (1989) Morphology and evolution of vascular plants. 3rd ed. New York: Freeman.

Goto, K., Meyerowitz, E. M. (1994) Function and regulation of the Arabidopsis floral homeotic gene *PISTILLATA*. *Genes Dev* **8**:1548-1560.

Gould, S. J. (1992) Ontogeny and phylogeny, revisited and reunited. *Bioessays.* **14**:275-279.

Gould, S. J. (1993) Eight little piggies, reflections in natural history. W. W. Norton & Company, New York, London.

Gu, Q., Ferrandiz, C., Yanofsky, M. F., Martienssen, R. (1998) The *FRUITFULL* MADS-box gene mediates cell differentiation during Arabidopsis fruit development. *Development.* **125**:1509-1517.

Hareven, D., Gutfinger, L., Pnueli, L., Bauch, L., Cohen, O., Lifschitz, E.(1994) The floral system of tomato. *Euphytica* **79**:235-243.

Hasebe, M., Wen, C. K., Kato, M., Banks, J. A.. (1998) Characterization of MADS homeotic genes in the fern *Ceratopteris richardii*. *Proc. Natl. Acad. Sci. U S A.* **95**:6222-6227.

Heard, J., Dunn, K. (1995) Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. *Proc. Natl. Acad. Sci. U S A* **92**:5273-5277

Hempel, F. D., Zambryski, P. C., Feldman, L. J. (1998) Photoinduction of flower identity in vegetatively biased primordia. *Plant Cell* **10**:1663-1676.

Huang, K., Louis, J. M., Donaldson, L, Lim, F. L, Sharrocks, A. D., Clore, G. M. (2000) Solution structure of the MEF2A-DNA complex: structural basis for the modulation of DNA bending and specificity by MADS-box transcription factors. *EMBO J.* **19**:2615-2628.

Huang, H., Tudor, M., Su, T., Zhang, Y., Hu, Y., Ma, H. (1996) DNA binding properties of two Arabidopsis MADS domain proteins: binding consensus and dimer formation. *Plant Cell* **8**:81-94.

Huang, H., Tudor, M., Weiss, C. A., Hu, Y., Ma, H. (1995) The Arabidopsis MADS-box gene *AGL3* is widely expressed and encodes a sequence-specific DNA-binding protein. *Plant Mol. Biol.* **28**:549-567.

Huijser, P., Klein, J., Lonnig, W. E., Meijer, H., Saedler, H., Sommer, H. (1992) Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene *squamosa* in *Antirrhinum majus*. *EMBO J.* **11**:1239-1249.

Jack, T., Brockman, L. L., Meyerowitz, E. M. (1992) The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell* **68**:683-697.

Jack, T., Fox, G. L., Meyerowitz, E. M. (1994) Arabidopsis homeotic gene *APETALA3* ectopic expression: transcriptional and posttranscriptional regulation determine floral organ identity. *Cell* **76**:703-716.

Kang, H. G., Jeon, J. S., Lee, S., An, G. (1998) Identification of class B and class C floral organ identity genes from rice plants. *Plant Mol Biol.* **38**:1021-1029.

Kater, M. K., Colombo, L., Franken, J., Busscher, M., Masiero, S., Van Lookeren Campagne, M., Angenent, G. C. (1998) Multiple *AGAMOUS* homologs from cucumber and petunia differ in their ability to induce reproductive organ fate. *Plant Cell* **10**:171-182.

Kempin, S. A., Savidge, B., Yanofsky, M. F. (1995) Molecular basis of the cauliflower phenotype in Arabidopsis. *Science* **267**:522-525.

Kenrick, P., Crane, P. R. (1997) The origin and early diversification of Land Plants. Smithsonian institution press, Washington and London.

Kimura, M. (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**:275-276.

Kramer, E. M., Dorit, R. L, Irish, V. F. (1998) Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. *Genetics*. **149**:765-83.

Kramer, E. M., Irish, V. F. (1999) Evolution of genetic mechanisms controlling petal development. *Nature* **399**:144-148.

Krizek, B. A., Meyerowitz, E. M. (1996) The *Arabidopsis* homeotic genes *APETALA3* and *PISTILLATA* are sufficient to provide the B class organ identity function. *Development* **122**:11-22.

Kukkonen, I., (1994) Notes on the treatment of the family *Lycopodiaceae* for Flora Nordica. *Ann. Bot. Fenn*. **31**:197-202.

Kush, A., Brunelle, A., Shevell, D., Chua, N. H.. (1993) The cDNA sequence of two MADS box proteins in Petunia. *Plant Physiol*. **102**:1051-1052.

Latchman, D. S. (1998). Eukaryotic Transcription Factors. 3d ed. Academic Press, London.

Lewis, E. B. (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* **276**:565-570.

Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L., Yanofsky, M. F. (2000) *SHATTERPROOF* MADS-box genes control seed dispersal in Arabidopsis. *Nature* **404**:766-770.

Lönnig, W. E., Saedler, H. (1994) The homeotic Macho mutant of Antirrhinum majus reverts to wild-type or mutates to the homeotic plena phenotype. *Mol. Gen. Genet.* **245**:636-643.

Lozano, R., Angosto, T., Gomez, P., Payan, C., Capel, J., Huijser, P., Salinas, J., Martinez-Zapater, J. M. (1998) Tomato flower abnormalities induced by low temperatures are associated with changes of expression of MADS-Box genes. *Plant Physiol.* **117**:91-100.

Ma, H., Yanofsky, M. F., Meyerowitz, E. M. (1991) *AGL1-AGL6*, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev.* **5**:484-495.

Malek, O., Lättig K. , Hiesel R. , Brennicke A., Knoop, V. (1996). RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO J*. **15**:1403-1411.

Mandel, M. A., Bowman, J. L., Kempin, S. A., Ma, H., Meyerowitz, E. M., Yanofsky, M. F. (1992) Manipulation of flower structure in transgenic tobacco. *Cell* **71**:133-143.

Mandel, M. A., Gustafson-Brown, C., Savidge, B., Yanofsky, M. F. (1992) Molecular characterization of the Arabidopsis floral homeotic gene *APETALA1*. *Nature* **360**:273-277.

Mandel, M. A., Yanofsky, M. F. (1995) A gene triggering flower formation in Arabidopsis. *Nature* **377**:522-524.

Mathews, S., Donoghue, M. J. (1999) The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* **286**:947-950.

Mattsson, J., Söderman, E., Svenson, M., Borkird, C., Engström, P. (1992). A new homeobox-leucine zipper gene from *Arabidopsis thaliana*. *Plant Mol. Biol.* **18**:1019-1022.

Mattsson, J. (1995). Isolation and characterization of homeodomain-leucine zipper encoding genes in plants. Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology, Acta Universitatis Upsaliensis, Uppsala.

McKinney, M. L., McNamara, K. J. (1991) Heterochrony. Plenum Press, New York and London.

Meyerowitz, E. M. (1999) Plants, animals and the logic of development. *Trends Cell Biol.* **9:**M65-68.

Mizukami, Y., Ma, H. (1992) Ectopic expression of the floral homeotic gene *AGAMOUS* in transgenic Arabidopsis plants alters floral organ identity. *Cell* **71**:119-131.

Mizukami, Y., Huang, H., Tudor, M., Hu, Y., Ma, H. (1996) Functional domains of the floral regulator *AGAMOUS*: characterization of the DNA binding domain and analysis of dominant negative mutations. *Plant Cell* **8**:831-845.

Molkentin, J. D., Black, B. L., Martin, J. F., Olson, E. N. (1996). Mutational analysis of the DNA binding, dimerisation and transcriptional activation domains of MEF2C. *Mol. Cell. Biol.* **16**:2627-2636.

Mouradov, A., Glassick, T., Vivian-Smith, A., Teasdale, R. (1996) Isolation of a MADS-box gene family from *Pinus radiata. Plant physiol*. **110**:1047.

Mouradov, A., Hamdorf, B., Teasdale, R. D. (1997) Isolation and characterization of a new MADS-box cDNA from *Pinus radiata. Plant Physiol* **113**:664.

Mouradov, A., Glassick, T. V., Hamdorf, B. A., Murphy, L. C., Marla, S. S., Yang, Y., Teasdale, R. D. (1998) Family of MADS-Box genes expressed early in male and female reproductive structures of monterey pine. *Plant Physiol*. **117**:55-62.

Mouradov, A., Hamdorf, B., Teasdale, R. D., Kim, J. T., Winter, K. U., Theissen, G. (1999) A *DEF/GLO*-like MADS-box gene from a gymnosperm: *Pinus radiata* contains an ortholog of angiosperm B class floral homeotic genes. *Dev Genet*. **25**:245-252.

Münster, T., Pahnke, J., Di Rosa, A., Kim, J. T., Martin, W., Saedler, H., Theißen, G. (1997) Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proc. Natl. Acad. Sci. U S A* **94**:2415-2420.

Ohno, S. (1972) An argument for the Genetic Simplicity of Man and other Mammals. *Journal of human evolution* **1**:651-662.

Panganiban, G., Irvine, S. M., Lowe, C., Roehl, H., Corley, L. S., Sherbon, B., Grenier, J. K., Fallon, J. F., Kimble, J., Walker, M., Wray, G. A., Swalla, B. J., Martindale, M. Q., Carroll, S. B. (1997)The origin and evolution of animal appendages.*Proc. Natl. Acad. Sci. U S A* **94**:5162-5166.

Passmore, S., Maine, G. T., Elble, R., Christ, C., Tye, B.-K. (1988) *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MATα cells. *J. Mol. Biol.* **204**:593-606.

Phan-Dinh-Tuy, F., Tuil, D., Schweighoffer, F., Pinset, C., Kahn, A., Minty, A. (1988) The 'CC.Ar.GG' box. A protein-binding site common to transcription-regulatory regions of the cardiac actin, c-fos and interleukin-2 receptor genes. *Eur. J. Biochem.* **173**:507-515.

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

Pellegrini, L., Tan, S., Richmond, T. J. (1995) Structure of serum response core bound to DNA. *Nature* **376**, 490-497.

Pickett, F. B. and Meeks-Wagner, D. R. (1995) Seeing double: appreciating genetic redundancy. *The Plant Cell* **7**:1347-1356.

Pidkowich, M. S., Klenz, J. E., Haughn, G. W. (1999) The making of a flower: control of floral meristem identity in Arabidopsis. *Trends Plant Sci*. **4**:64-70.

Pnueli, L., Hareven, D., Broday, L., Hurwitz, C., Lifschitz, E. (1994) The *TM5* MADS box gene mediates organ differentiation in the three inner whorls of tomato flowers. *Plant Cell* **6**: 175-186.

Pnueli, L., Abu-Abeid, M., Zamir. D., Nacken. W., Schwarz-Sommer, Z., Lifschitz, E. (1991) The MADSbox gene family in tomato: temporal expression during floral development, conserved secondary structures and homology with homeotic genes from Antirrhinum and Arabidopsis. *Plant J.*, **1**:255-266

Purugganan, M. D. (1997) The MADS-box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. *J. Mol. Evol.* **45**:392-396.

Qiu, Y. L., Lee, J., Bernasconi-Quadroni, F., Soltis, D. E., Soltis, P. S., Zanis, M., Zimmer, E. A., Chen, Z., Savolainen, V., Chase, M. W. (1999) The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**:404-407.

Raff, R. A. (1996) The Shape of Life. The University of Chicago press, Chicago and London.

Raubeson, L. A., Jansen, R. K. (1992) Chloroplast DNA Evidence on the Ancient Evolutionary Split in Vascular Land Plants. *Science* **255**:1697-1699.

Raven, P. H., Evert, R. F., Eichhorn, S. E. (1999) Biology of plants. 6th ed. W. H. Freeman and Company. New York.

Riechmann, J. L., Krizek, B. A., Meyerowitz, E. (1996) Dimerization specificity of Arabidopsis MADS domain hometotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc. Natl. Acad. Sci. USA* **93**: 4793-4798.

Riechmann, J. L., Meyerowitz, E. M. (1997) MADS domain proteins in plant development. *Biol. Chem.***378**:1079-1101.

Rounsley, S. D., Ditta, G. S., Yanofsky, M. F. (1995) Diverse roles for MADS-box genes in Arabidopsis development *Plant Cell* **7**:1259-1269.

Rutledge, R., Regan, S., Nicolas, O., Fobert P., Cote, C., Bosnich, W., Kauffeldt, C., Sunohara, G., Seguin, A., Stewart, D. (1998) Characterization of an *AGAMOUS* homolog from the conifer black spruce (*Picea mariana*) that produces floral homeotic conversions when expressed in Arabidopsis. *Plant J.* **15**:625-634.

Santelli, E., Richmond, T. J. (2000) Crystal structure of MEF2A core bound to DNA at 1.5 Å resolution. *J. Mol. Biol.* **297**:437-449.

Sattler, R. (1988) Homeosis in plants *Amer. J. Bot.* **75**:106-117.

Sattler, R. (1991) Process morphology: structural dynamics in development and evolution, *Can. J. Bot.* **70**:708-714.

Savidge, B., Rounsley, S. D., Yanofsky, M. F. (1995) Temporal relationship between the transcription of two Arabidopsis MADS box genes and the floral organ identity genes. *Plant Cell* **7**:721-733.

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., Sommer, H. (1990) Genetic Control of Flower Development by Homeotic Genes in *Antirrhinum majus*, *Science* **250**:931-936.

Sheldon, C. C., Burn, J. E., Perez, P. P., Metzger, J., Edwards, J. A., Peacock, W. J., Dennis, E. S. (1999). The *FLF* MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. *Plant Cell* **11**:445-458.

Shore, P., Sharrocks, A. D. (1995) The MADS-box family of transcription factors. *Eur J Biochem*, **229**:1-13.

Singer, C. (1959) A short history of scientific ideas to 1900. Oxford university press, London.

Soltis, P. S., Soltis, D. E., Chase, M. W. (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**:402-404.

Sommer, H., Beltran, J. P, Huijser, P., Pape, H., Lönnig, W. E., Saedler, H., Schwarz-Sommer, Z. (1990) *Deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *EMBO J.* **9**:605-613.

Stewart, W. N., Rothwell, G. W. (1993) Paleobotany and the Evolution of Plants 2nd. ed. Cambridge University Press, Cambridge.

Svedelius, N. (1909) Om några svenska monstrositetsformer af *Anemone nemorosa*. *Svensk Bot. Tidskr*. **3**:47-63.

Svensson, M. (1996) Homeotiska mutanter kastar nytt ljus över blommorfologin hos vitsippa, *Anemone nemorosa Svensk Bot. Tidskr*. **90**:231-237.

Takhtajan, A. L. (1991) Evolutionary trends in flowering plants. Columbia University Press, New York.

Tan, S., Richmond, T.J. (1998) Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex, *Nature* **391**:660-666.

Tandre, K., Albert, V. A., Sundås, A., Engström, P. (1995) Conifer homologs to genes that control floral development in angiosperms. *Plant Mol Biol.* **27:**69-78

Tauz, D. (1992) Redundancies, development and the flow of information. *BioEssays*. **14**:263-266.

Theißen, G., Strater, T., Fischer, A., Saedler, H. (1995) Structural characterization, chromosomal localization and phylogenetic evaluation of two pairs of *AGAMOUS*-like MADS-box genes from maize. *Gene* **156**:155-166.

Theißen, G., Kim, J., Saedler, H. (1996) Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J. Mol. Evol.* **43**: 484–516.

Theißen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J. T., Münster, T., Winter, K. U., Saedler, H. (2000) A short history of MADS-box genes in plants. *Plant Mol. Biol.* **42**:115-149.

Tröbner, W., Ramirez, L., Motte, P., Hue, I., Huijser, P., Lonnig, W. E., Saedler, H., Sommer H., Schwarz-Sommer, Z. (1992) *GLOBOSA*: a homeotic gene which interacts with *DEFICIENS* in the control of *Antirrhinum* floral organogenesis. *EMBO J*. **11**: 4693-4704.

Tsuchimoto, S., van der Krol, A. R., Chua, N. H. (1993) Ectopic expression of *pMADS3* in transgenic petunia phenocopies the petunia blind mutant. *Plant Cell* **5**: 843-853.

van der Krol, A. R., Brunelle, A., Tsuchimoto, S., Chua, N. H. (1993) Functional analysis of petunia floral homeotic MADS box gene *pMADS1*. *Genes Dev.* **7**:1214-1228.

van der Krol, A. R., Chua, N. H. (1993) Flower development in *Petunia*. *Plant Cell* **5**:1195-1203.

Weigel D., Alvarez, J., Smyth, D. R., Yanofsky, M. F., Meyerowitz, E. M. (1992) *LEAFY* controls floral meristem identity in Arabidopsis. *Cell* **69**:843-859.

Whittier, D. P., (1976) Gametophytes of *Lycopodium obscurum* as grown in axenic culture. *Can. J. Bot.* **55**:563-567.

Winter, K. U., Becker, A., Münster, T., Kim, J. T., Saedler, H., Theißen, G. (1999) MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc. Natl. Acad. Sci. U S A* **96**:7342-7347.

Yanofsky, M. F, Ma, H., Bowman, J. L., Drews, G. N., Feldmann, K. A., Meyerowitz E. M. (1990) The protein encoded by the Arabidopsis homeotic gene *AGAMOUS* resembles transcription factors. *Nature* **346**:35-39.

Young, D. (1992) The discovery of evolution. Natural History Museum Publications, London and Cambridge University Press, Cambridge.

Yu, D., Kotilainen, M., Pollanen, E., Mehto, M., Elomaa, P., Helariutta, Y., Albert, V. A., Teeri, T. H. (1999) Organ identity genes and modified patterns of flower development in *Gerbera hybrida*. *Plant J.* **17**:51-62.

Zachgo, S., Saedler, H., Schwarz-Sommer, Z. (1997) Pollen-specific expression of DEFH125, a MADS-box transcription factor in *Antirrhinum* with unusual features**.** *Plant J.* **11**:1043-1050.

Zhang, H., Forde, B. G. (1998) An Arabidopsis MADS Box gene That Controls Nutrient-Induced Changes in Root Architecture. *Science* **279**:407-409.

Zimmermann, W. (1965) Die Telomtheorie. Gustav Fischer Verlag, Stuttgart.

Errata list
p. 34
line 6: macrosporangia should be megasporangia
line 11: Baum (1999) should be Baum and Whitlock (1999)
line 19: ”This supports a simple perianth as a basal condition among angiosperms, and in at least *Nymphaeaceae* it is clearly petaloid.” should be ”This supports a spiral perianth as a basal condition among angiosperms, and in at least *Nymphaeaceae* it contains clearly petaloid organs.”

Errors can be discussed by e-mail: mats.svensson@ebc.uu.se