Functional Diversification among MADS-Box Genes and the Evolution of Conifer Seed Cone Development

ERIKA GROTH
Dissertation presented at Uppsala University to be publicly examined in Lindahlsalen, Evolutionsbiologiskt centrum, Norbyvägen 18A, Uppsala, Friday, September 17, 2010 at 10:00 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract

MADS-box genes are important regulators of reproductive development in seed plants, including both flowering plants and conifers. In this thesis the evolution of the AGAMOUS subfamily of MADS-box genes, and what the ancestral function of this group of genes might have been in the early seed plants about 300 million years ago, was addressed by the discovery of two novel conifer genes, both basal to all previously known AGAMOUS subfamily genes. DAL20, the most basal of these genes, was exclusively expressed in roots, unlike all previously known AGAMOUS subfamily genes. I also studied the evolutionary mechanisms leading to functional diversification of duplicated genes in two different subfamilies of MADS-box genes; the AGAMOUS and AGL6 subfamilies. Focus was on studying changes in gene expression pattern, representing changes in the transcriptional regulation between the genes, and on comparing the functional properties of the gene products, representing changes in the protein-coding sequence between the genes. Duplicated genes in the AGL6 subfamily were found to have evolved by both mechanisms. In the AGAMOUS subfamily I found duplicated spruce genes; DAL2 and DAL20, that appear to have functionally diversified mainly by changes in the transcriptional regulation. Conifer AGAMOUS subfamily genes were also used in a comparative developmental-genetics approach to evaluate hypotheses, based on the morphology of fossil and extant conifer seed cones, on the identity of the female reproductive organ, the ovuliferous scale, and the evolution of seed cone morphology in the conifer families Pinaceae, Taxodiaceae and Cupressaceae. Seed cones in these families have been hypothesized to have homologous ovule-bearing organs, but I found substantial differences in the expression patterns of orthologous AGAMOUS subfamily genes in seed cones of these families that are not compatible with this hypothesis, indicating that the evolutionary history of conifer seed cones is more diverse than previously thought.

Keywords: conifer, seed cone, evo-devo, morphology, plant development, plant evolution, gene family, gene evolution, AGAMOUS, MADS-box, transcription factor, Picea, Cryptomeria, Thuja, Juniperus

Erika Groth, Department of Organism Biology, Physiological Botany, Norbyvägen 18D, Uppsala University, SE-75236 Uppsala, Sweden

© Erika Groth 2010

ISSN 1651-6214
ISBN 978-91-554-7850-6
urn:nbn:se:uu:diva-128937 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-128937)
“Though this be madness, yet there is method in it”

William Shakespeare (Hamlet)

Till min familj
List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I  Groth, E., Tandre, K., Engström, P., Vergara-Silva, F. AGAMOUS subfamily MADS-box genes and the evolution of seed cone morphology in Cupressaceae and Taxodiaceae. Submitted manuscript

II Tandre, K., Groth, E., Liu, S., Engström, P. Identification and characterization of basal AGAMOUS subfamily MADS-box genes from Norway spruce (Picea abies); implications for the evolution of AGAMOUS subfamily genes in seed plants. Manuscript

III Groth, E., Tandre, K., Engström, P. Molecular and functional evolution of the AGAMOUS subfamily MADS-domain proteins. Manuscript

IV Groth, E., Carlsbecker, A., Tandre, K., Engström, P. Functional divergence by multiple mechanisms between the paralogous sister genes DAL1 and DAL14 in the AGL6 subfamily of MADS-box genes in the conifer Picea abies. Manuscript

Author contributions for paper I: All authors participated in the planning of the project. E.G. collected the Cupressaceae/Taxodiaceae tissues, isolated the genes and performed the expression analyses (RT-PCR and RNA in situ hybridization). K.T. performed the phylogenetic analysis. E.G. wrote the manuscript, with contributions and comments from K.T., P.E. and F.V.-S.

Author contributions for papers II-IV are stated in the manuscripts.
Contents

Thank you everyone who has helped to make this book a reality! .............. 9

Evolution of reproductive development in conifers.............................. 11
  Introduction to conifers and their evolution.................................. 11
  Evolution and development of the conifer seed cone based on the
  morphology of fossil and extant cones.......................................... 15

MADS-box genes regulate plant development........................................ 19
  MADS-box genes in angiosperms: The (A)BC(DE) model............... 19
  MADS-box genes in gymnosperms.................................................. 24
  Using a developmental genetics approach to evaluate paleobotany-based
  hypotheses on the morphological evolution of Taxodiaceae and
  Cupressaceae s.str. seed cones (paper I)........................................ 26
  Evolution of MADS-box genes....................................................... 29
  What was the function of the ancestral AG subfamily gene? (paper II) 31
  Evolution of the AG subfamily of MADS-domain proteins in conifers and
  angiosperms (paper III).................................................................. 32
  MADS-domain proteins are transcription factors............................. 34
  Evolution of the AGL6 subfamily of MADS-box genes in seed plants
  (paper IV).......................................................................................... 35

Sammanfattning på svenska (Summary in Swedish)................................. 38

References.............................................................................................. 42
Abbreviations and Nomenclature

AGL AGAMOUS-like
DAL DEFICIENS-AGAMOUS-like
cDNA Complementary DNA
DNA Deoxyribonucleic acid
EST Expressed sequence tag
I-domain Intervening domain
K-domain Keratin-like domain
MADS MCM1, _AG_, DEF, SRF
mRNA Messenger RNA
mya Million years ago
PCR Polymerase chain reaction
RACE Rapid amplification of cDNA ends
RAM Root apical meristem
RNA Ribonucleic acid
RT-PCR Reverse transcriptase PCR
qRT-PCR Quantitative (real time) reverse transcriptase PCR
SEM Scanning electron microscopy
s.l. _sensu lato_ (in the wider sense)
s.str. _sensu strictu_ (in the narrow sense)

The following nomenclature is used in this thesis:

Protein names are written in upper case letters, e.g. AGAMOUS (AG).
Gene names are written in upper case italic letters, e.g. _AGAMOUS_ (AG).
Mutant names are written in lower case italic letters, e.g. _agamous_ (ag).

**Box** refers to a DNA section that encodes a protein domain with a specific structure and/or function, e.g. the **MADS-box** is the part of a gene that codes for the **MADS-domain** of the corresponding protein.
Thank you everyone who has helped to make this book a reality!

“Actually it is not so much courage that is called for but endurance. I should place independence first and the endurance, neither of which are virtues but acquired habits. Independence is regarded as an unsociable habit, especially in a woman”

Lucy Evelyn Cheesman, English Entomologist (1881-1969)

Now, finally, this book is finished! It will take many long years until we start to approach a true understanding all the way down on the molecular scale of how plants are made, how they work and, most curiously of all, how they ended up the way they are. This thesis will not take us all the way, but it will bring us one step further down the path towards that goal.

Producing this book has taken a long time and a huge amount of work. Without the help and support of many people this book would still not be finished. Thank you everyone!

First, I would like to thank the people that have been my advisors during these years. My main advisor Peter Engström for accepting me as a PhD student and by doing so allowing me to spend years working with this fascinating subject, for your clear logical thinking and for, particularly during the last few years, allowing me to work independently. Because of this I ended up learning much more than I would have otherwise. My additional advisors Karolina (Lina) Tandre and during the first years also Annika Sundás-Larsson for your enthusiasm and support during these years. It really helped! Especially thank you to Karolina for all the phylogenetic analyses.

To all the other past and present people in the Spruce group (or more accurately the Conifer group), thank you for your help and support, and for all your previous work that laid the foundation for my work. Most notably (but not limited to) Francisco (Paco) Vergara-Silva for opening the door for new groups of study species, Marie Englund for technical advice and for donating some of your *Picea* material for my RNA *in situ* hybridizations, Annelie Carlsbecker and Jens Sundström for encouragement, technical advice and access to your previously collected *Picea* material, Agneta Ottosson for transforming and taking care of my *Arabidopsis* plants and for making huge amounts of yeast medium, Agneta (again), David Clapham and others in-
volved for your efforts with the spruces at Ultuna (even though that project isn’t included in this book), Jan Naef for your yeast work (and for having patience with me as an advisor) and Shuoran Liu for your help with my plasmids and the expression studies of DAL18 and DAL20.

A large part of this work would not have been possible without the ‘yeast library’ of Arabidopsis MADS-clones for yeast-2-hybrid that Richard Immink and coworkers provided me with. Thank you for the clones and for the technical advice!

Thank you Botaniska trädgården here in Uppsala and Bergianska trädgården in Stockholm for allowing me to collect samples from your collections (and thank you to all the trees I’ve collected from!). As a molecular biologist I spend most of my time in the lab, so I’ve really appreciated my little reoccurring field excursions to collect material, both in the gardens and in the ‘wild’ nature surrounding Uppsala.

To the people that I’ve worked with at FysBot, EBC and GC/SLU, thank you for being nice, supportive and helpful. It made all the hard work seem much easier. I especially want to thank my fellow PhD students for the friendly atmosphere and for all our discussions about science, the university and everything else. During my time as a PhD student I’ve been teaching biology labs in 7 (!) different undergraduate and master level courses, and I am particularly grateful for the well organized assistance of Marianne Svarvare at IBG during all these courses. I also want to thank the technicians at EBC for all your work building and rebuilding darkrooms for me, and the staff at the other departments who have generously allowed and helped me to use their equipment for my experiments when our own was insufficient, broken or otherwise unavailable. Thank you also to Stefan Gunnarsson for your assistance with the microscopes.

Ett stort tack till mamma, pappa och Daniel för ert stöd! Ni har alltid ställt upp för mig och visat intresse för vad jag sysslar med. Förhopningsvis har det blivit i alla fall lite klarare nu vad jag egentligen sysslat med alla dessa år. Tack för att ni hjälper mig hålla mig rotad i världen utanför Uppsala och universitetet.

Travel expenses have been covered by Gertrud Thelins resestipendium, Eliassons fond, Scandinavian Plant Physiology Society, Rektors resebidrag från Wallenbergs-stiftelsen and Liljewalchs resestipendium. My last year as a PhD student was partly founded by a stipend from the foundation Olof O:son Wijks minne för botanik. I also received economic support from Lånefonden at Norrlands Nation. Thank you all.

So now that you’ve read ‘the only part of a thesis that everyone reads’, I hope and recommend that you will continue and read the rest of the book. I hope that you will find it interesting and that it can inspire you to share some of the fascination that I feel about the evolution of life and of the very little studied non-flowering plants in particular. Who knows, maybe you will even learn something new…
Evolution of reproductive development in conifers

Introduction to conifers and their evolution

The conifers are one of the four now living groups of gymnosperms. The others are Gnetales, consisting of the three genera *Welwitschia*, *Ephedra* and *Gnetum*, the cycads and the maidenhair tree (*Ginkgo biloba*), which is the sole living survivor in its group. There are few extant gymnosperm species compared to the huge number of angiosperm species, but ecologically conifers dominate large areas like the taiga region covering much of Eurasia and North America. Like the angiosperms they have seeds and pollen, but they do not have flowers. One of the main criteria separating gymnosperms from angiosperms is that gymnosperms have naked seeds, i.e. the ovules are exposed to the wind-blown pollen at the time of pollination.

Technically an ovule is an unfertilized seed, but since it is in practice very difficult to see exactly when fertilization occurs the limit for when to switch from using the term ovule to using seed tend to be set rather arbitrarily. It should be noted that the ovules of gymnosperms are not necessarily exposed at all times. In some species the seeds become covered and completely enclosed in various structures that facilitate seed dispersal by animals in a way that is functionally similar, although structurally very different, to the angiosperm fruits (Tomlinson and Takaso 2002; Farjon 2008). Conifer seed and pollen cones develop on the same tree, as for example in *Picea abies*, or on separate trees, like in *Juniperus communis*, but either way the cones are unisexual.

The angiosperm flower is a much younger structure than the conifer cone. The oldest fossil flowers are from the Early Cretaceous (Friis et al. 2006; Friis et al. 2010), but conifers are present in the fossil record from the Pennsylvanian (Taylor et al. 2009) (fig. 1). In other words, conifer cones are about twice as old as flowers, although the last common ancestor of conifers and flowering plants (and of the other extant seed plant groups) lived at the latest during the early Pennsylvanian (Taylor et al. 2009). Extant seed plants are considered to be monophyletic (Chaw et al. 1997; Rydin et al. 2002; Bateman et al. 2006), although it should be remembered that this group actually also includes extinct groups of gymnosperms like the bennettites, the glossopterids and the cordaites (Beck 1988; Willis and McElwain 2002;
Bateman et al. 2006). The relationship between the main groups of seed plants is however still unresolved (see e.g. Rydin et al. 2002). This means that the flower evolved from some sort of gymnosperm reproductive structure, but from what kind of structure and how that happened is still a matter of debate. Many hypotheses have been suggested (reviewed in Frohlich and Chase 2007), but there is still far to go until any consensus is reached on the matter.

The seed itself, which was one of the most important inventions in all of plant evolution, is far older than both cones and flowers and first appears in the fossil record in the Devonian (fig. 1) (Rothwell and Scheckler 1988; Taylor et al. 2009). The seed is a remarkable dispersal unit which contains multiple generations within its seed coat, made from the ovule integument; the megasporangium (nucellus) of the parental sporophyte, the entire megagametophyte which develops from the megaspore and finally the embryo, representing the sporophyte of next generation. The earliest fossil found that could be interpreted as a primitive version of an ovule is *Runcaria* from the middle Devonian (approximately 385 mya) (Gerrienne et al. 2004; Gerrienne and Meyer-Berthaud 2007; Taylor et al. 2009). The early ovules developed on the inner surface of lobed structures called cupules (Rothwell and Scheckler 1988; Gerrienne et al. 2004; Taylor et al. 2009). A cupule is basically the slightly modified, heavily branched tip of a branching axis. It is not known how the ovules, if the conifer ovules are homologous to the Devonian and early Carboniferous ovules, made the transition from developing directly on a branching axis within a cupule to developing in a cone.

Due to new technologies and the focus on model organisms, we now know much about the molecular regulation of floral development, but still only very little is known about the molecular regulation of cone development. This is mainly because of the technical difficulties involved when working with wind pollinated, difficult to genetically manipulate, large trees with long generation times and large, unsequenced genomes. Recently a Swedish initiative has been announced where the Knut and Alice Wallenberg Foundation will fund the sequencing of the genome of Norway spruce, which will be the first gymnosperm genome sequenced (Travis 2010).

Two categories of naturally occurring abnormal ‘mutant’ seed cones have been described, although no molecular analysis of either category is yet published. Bisexual conifer cones (also known as bisporangiate strobili) have been described sporadically. In all cases described the lower bracts were replaced by microsporophylls and the distal part of the cone was a (more or less) normal seed cone with axillary ovuliferous scales (Coulter and Chamberlain 1910; Littlefield 1931; Caron and Powell 1990; Tabor 1990). Conifer shoots that are intermediate structures between seed cones and vegetative shoots have also been described in a number of different species, for example cones with ovuliferous scales replaced by needles (referred to in Coulter and Chamberlain 1910) and shoots that switches identity during develop-
ment between being a vegetative shoot and a seed cone, like the *Picea abies* variety *acrocona* (*acrocona* spruces can be found in public gardens in Uppsala).

The taxonomic classification of conifers into families varies somewhat, but nine extant conifer families are commonly recognized: Pinaceae (11 genera), Araucariaceae (3 genera), Podocarpaceae (18 genera), Phyllocladaceae (1 genera), Sciadopityaceae (1 genera), Cephalotaxaceae (1 genera), Taxaceae (5 genera), Cupressaceae s.str. (21 genera) and Taxodiaceae (9 genera) (see for example Chaw et al. 1997; Gadek et al. 2000; Farjon 2008). Cupressaceae s.str. and Taxodiaceae are often combined into one big family, Cupressaceae s.l. (Gadek et al. 2000; Farjon 2008). By far most is known about Pinaceae, which includes well known and economically important forest trees like spruces (*Picea*), pines (*Pinus*), hemlocks (*Tsuga*), larches (*Larix*), (true) cedars (*Cedrus*), firs (*Abies*) and Douglas fir (*Pseudotsuga*). Pine and spruce together forms the main basis of the Swedish forest industry, which constitutes about 11% of Sweden’s total exports and is the world’s second largest exporter (after Canada) of sawn timber products, paper and pulp taken together (‘The Forest Industry-A natural part of Sweden’, Skogsindustrierna, Stockholm, Sweden). Also Taxodiaceae contains some famous conifers like the gigantic coast redwoods (*Sequoia*) and giant sequoias (*Sequoiadendron*) of the western USA. For common names of the main plant groups and species in this thesis see table 1.

**Table 1. Common names of some of the most important species and plant groups referred to in this thesis. For genera that contain only one species the second part of the species name is placed in brackets.**

<table>
<thead>
<tr>
<th>Scientific name</th>
<th>English common name(s)</th>
<th>Swedish common name(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gymnosperms</td>
<td>Gymnosperms</td>
<td>Nakenfröiga växter</td>
</tr>
<tr>
<td>Angiosperms</td>
<td>Flowering plants</td>
<td>Gömfröiga växter, Blomväxter</td>
</tr>
<tr>
<td>Coniferales</td>
<td>Conifers</td>
<td>Barträd</td>
</tr>
<tr>
<td>Pinaceae</td>
<td>Pine family</td>
<td>Tallväxter</td>
</tr>
<tr>
<td>Cupressaceae</td>
<td>Cypress family</td>
<td>Cypressväxter</td>
</tr>
<tr>
<td></td>
<td>Swamp cypress family, Redwood family</td>
<td>Sumpcyypressväxter</td>
</tr>
<tr>
<td>Taxodiaceae</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Picea</em></td>
<td>Spruces</td>
<td>Granar</td>
</tr>
<tr>
<td><em>Picea abies</em></td>
<td>Norway spruce</td>
<td>Gran</td>
</tr>
<tr>
<td><em>Picea mariana</em></td>
<td>Black spruce</td>
<td>Svartgran</td>
</tr>
<tr>
<td><em>Pinus</em></td>
<td>Pines</td>
<td>Tollar</td>
</tr>
<tr>
<td><em>Juniperus</em></td>
<td>Junipers</td>
<td>Enar</td>
</tr>
<tr>
<td><em>Juniperus communis</em></td>
<td>Common juniper</td>
<td>En</td>
</tr>
<tr>
<td><em>Cryptomeria (japonica)</em></td>
<td>Japanese cedar, Sugi</td>
<td>Kryptomeria</td>
</tr>
<tr>
<td><em>Thujaops (dolabrata)</em></td>
<td>Hiba arborvitae</td>
<td>Hiba</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>Thale cress</td>
<td>Backtrav</td>
</tr>
<tr>
<td><em>Antirrhinum majus</em></td>
<td>Snapdragon</td>
<td>Lejongap</td>
</tr>
<tr>
<td><em>Oryza sativa</em></td>
<td>Rice</td>
<td>Ris</td>
</tr>
</tbody>
</table>
Almost all the early conifer families were extinct by end of the Permian (Farjon 2008; Taylor et al. 2009), but during the Mesozoic the conifers started to expand and the modern conifer families first appear in the fossil record during this period (Willis and McElwain 2002; Farjon 2008; Taylor et al. 2009). The Mezozoic is often referred to as the Age of the Dinosaurs, but it would be equally appropriate to refer to it as the Age of the Conifers.

Figure 1. A simplified geological timescale showing the periods and epochs of the last half billion years of Earth’s history (not drawn to scale). Some of the main highlights of the evolution of conifers and angiosperms are (roughly) indicated on the scale based on fossil evidence (Kenrick and Crane 1997; Taylor et al. 2009). Important events in the evolution of MADS-box genes are also indicated (minimum age). This geological timescale is based on the 2009 Geological Time Scale of the Geological Society of America.
Evolution and development of the conifer seed cone based on the morphology of fossil and extant cones

A ‘stereotypical conifer seed cone’ (fig. 2), like a pine or spruce cone, has a main axis with a series of complexes attached, often arranged in a spiral. These complexes are initiated at the apical end of the cone, i.e. the youngest are at the top and they get progressively older in developmental terms as you follow the cone axis towards the attachment point at the base. The cones are compound strobili, meaning that these complexes consist of two kinds of appendages. The abaxial one (the lower one counted from the tip of the cone) is the sterile bract. The adaxial one (the upper one) is the ovule-bearing scale, also known as the fertile scale or the ovuliferous scale. This scale is the reproductive organ of the conifer cone and it develops in the axil of the bract and can be partially fused to the bract. The ovules develop on the adaxial surface of the ovuliferous scale.

However, already in the 19th century scientists started to take a closer look at the seed cones of conifer families other than Pinaceae and things quickly became confused (Wordsdell 1900). One cause of confusion is the inconsistent use of terminology, but the main problem was that the variation in cone morphology is big. All seed cones have a main axis, at least a few bracts and ovules. It was, and still is, the ovuliferous scale that is the problem. Many studies of conifer seed cones have started with a preconceived notion of what should be there, and the investigator then tried very hard to fit the actual cone into the preconceived mold. Mainly for this reason the studies of Barry Tomlinson with coworkers from the late 1980’s onwards, where they use SEM to study the early developmental stages of conifer seed cones in a number of families, have revolutionized the understanding of the development of conifer cones and have challenged previous ideas of the homologous relationships and functions of cone structures.

In Cupressaceae s.str. the seed cones have a cone axis with one kind of appendage (bracts) and the ovules developing directly from the cone axis either axially to the bracts, alternating with the bracts or at the tip of the cone axis (see examples in fig. 2) (Takaso and Tomlinson 1989b; Tomlinson et al. 1993; Farjon and Ortiz Garcia 2002; Tomlinson and Takaso 2002; Schultz et al. 2003). In Libocedrus (Cupressaceae s.str.) there is an additional appendage (the scale or ligular structure), that appears shortly after the ovules have been initiated (Tomlinson et al. 1993), but this is otherwise a feature more common in Taxodiaceae. In Juniperus the uppermost bracts start expanding after the ovules have initiated and develop after pollination into the juniper ‘berry’, enclosing the seeds and forming a seed dispersal unit adapted for dispersal by animals (Farjon and Ortiz Garcia 2002; Tomlinson and Takaso 2002; Farjon 2008).

In Taxodiaceae the seed cones are very similar to those of Cupressaceae s.str., but some species have an additional appendage developing either to-
gether with the ovule from the same organ primordium in *Cryptomeria* (fig. 2) (Takaso and Tomlinson 1989a) or appearing between the ovules and the bract after ovule initiation (Takaso and Tomlinson 1990; Tomlinson and Takaso 2002; Farjon and Ortiz Garcia 2003). In *Cunninghamia* there is a one-to-one relationship between ovules and scales (lobes) (Farjon and Ortiz Garcia 2003), but not in *Taxodium* or *Glyptostrobus* (Takaso and Tomlinson 1990). Interestingly in *Taxodium* and *Glyptostrobus* the additional appendages (referred to as scales or lobes) develop also in the axil of bracts where no ovules develop, indicating that there is no absolute connection between this appendage and the ovules (Takaso and Tomlinson 1990). The Taxodiaceae ovules develop on the base of the bract or axillary to the bract, sometimes on a shallow tissue cushion (Takaso and Tomlinson 1990; Takaso and Tomlinson 1992; Tomlinson and Takaso 2002). In both Cupressaceae s.str. and Taxodiaceae many species have cones with multiple rows of ovules developing in connection to each bract (Takaso and Tomlinson 1989b; Takaso and Tomlinson 1992; Farjon and Ortiz Garcia 2002; Tomlinson and Takaso 2002). In *Sciadopitys* (Sciadopityaceae) the seed cones are similar to those of Taxodiaceae, but here the ovules develop from additional appendages (scales) that develop axillary to the bracts (Takaso and Tomlinson 1991), apparently making the cone a true compound structure.

In the common yew (*Taxus baccata*, Taxaceae) the single ovule develops terminally on a dwarf shoot (Tomlinson and Takaso 2002; Farjon 2008). At the base of the ovule there are a few bracts, but there is nothing that can be called an ovuliferous scale in the traditional (i.e. Pinaceae) sense of the word. Later a structure called the aril appears at the base of the ovule, and develops further into a fleshy, bright red structure that almost entirely surrounds the seeds, creating a berry-like structure that is eaten by birds (Farjon 2008). In the related *Cephalotaxus* (Cephalotaxaceae) the ovules develop in the axil of sterile bracts (Tomlinson and Takaso 2002), also without any morphologically distinguishable ovuliferous scale present.

The Podocarpaceae seed cones are the most strongly adapted to seed dispersal by animals, and hence do not look much like other conifer cones. The single ovule develops from the adaxial side of a structure known as the epimatium (Tomlinson and Takaso 2002). The epimatium also surrounds and later often encloses the ovule, forming either alone or together with the bracts below (that sometimes fuse to form a fruitlike structure known as the receptacle) after substantial changes in shape and color a variety of remarkably fruitlike structures depending on the species (Tomlinson and Takaso 2002; Farjon 2008). The epimatium has been suggested to be homologous to the ovuliferous scale of Pinaceae cones, but this is still a matter of debate (Tomlinson and Takaso 2002). In *Phyllocladus* (Phyllocladaceae), which is thought to be closely related to the podocarps, the ovules develop in the axil of sterile bracts (Tomlinson and Takaso 2002). Little is known about cone
development in Araucariaceae, but in this family the ovules develop on the adaxial side of the bracts (Tomlinson and Takaso 2002).

Are all conifer seed cones homologous? The main hypothesis for the evolution leading to the modern conifer seed cones is the hypothesis suggested by Rudolf Florin (1951; 1954), mainly based on decades of detailed studies of Paleozoic and Mesozoic fossil conifer cones. The early conifers in Florin’s interpretation had dwarf shoots growing in the axil of bracts (fig. 2). These dwarf shoots were basically simple cones (simple strobili), with both sterile scales and fertile scales carrying ovules. According to Florin’s hypothesis the ovules of modern cones in all conifers (except the taxads) develop on ovuliferous scales that are the fused, reduced and flattened remains of the dwarf shoots of the early conifers (fig. 2). Actually Florin distinguished between a sterile part of the modern ovuliferous scale (originating from the sterile scales of the original dwarf shoot) and a fertile part (originating from the old fertile scales) (Florin 1951), but this distinction has been dropped by later generations of scientists. The bracts of modern cones he considered to be derived from the bracts subtending the dwarf shoots of the fossil cones. The entire cones of modern conifers are therefore interpreted as compound strobili, with the ovuliferous scale being derived from the dwarf shoot. In all cases of modern conifer cones where no obvious ovuliferous scales are visible in the cone his explanation was that the ovuliferous scale has fused with the bract, while still retaining its reproductive identity, except for the taxad cones which he considered to be simple strobili (Florin 1951).

Florin only had access to impression and/or compression fossils from the Northern hemisphere (Florin 1951). Over the years since Florin the collections of Paleozoic and Mesozoic fossil conifer seed cones have increased, including some permineralized cones that give information about internal and 3D structures, and new interpretations have introduced some modifications to Florin’s initial ideas (Schweitzer 1963; Clement-Westerhof 1988; Hernandez-Castillo et al. 2001). Despite this, and although simple (as opposed to compound) Paleozoic conifer seed cones from Gondwana have been found that have axillary ovules subtended by bracts or leaves without any sign of the dwarf shoot (see fig. 2C for example) (Clement-Westerhof 1988; Taylor et al. 2009), Florin’s hypothesis of the reduction of an axillary dwarf shoot leading to the modern conifer ovuliferous scale and modern conifer seed cones being compound structures has remained almost entirely unchallenged for more than half a century. However, based on developmental studies Tomlinson, Takaso and Cameron have suggested (1993) that the seed cones of Libocedrus, and maybe other Cupressaceae as well, are simple cones rather than the compound cones of Florin’s hypothesis.
Figure 2. Schematic presentations of Paleozoic and modern conifer seed cones. The ovules are black, the dwarf shoots/ovuliferous scales are dark grey and the sterile organs are white. A) An early Paleozoic dwarf shoot subtended by a bract (Clement-Westerhof 1988). B) A later Paleozoic flattened dwarf shoot with partly fused scales, subtended by a bract. (Schweitzer 1963; Clement-Westerhof 1988). C) The bract and ovule of Ferugliocladus, a putative simple conifer seed cone from Paleozoic Gondwana (Clement-Westerhof 1988). D) A modern spruce cone (Pinaceae) at a pre-pollination developmental stage. E) A modern post-dormant Cryptomeria seed cone (Taxodiaceae) (Takaso and Tomlinson 1989a, and paper I). F) A modern Thujaopsis seed cone to the left (Florin 1951; Jagel 2002, and paper I) and a modern Juniperus seed cone to the right (Tomlinson and Takaso 2002; Schultz et al. 2003, and paper I) (both Cupressaceae s.str.). Neither kind of Cupressaceae cone shows any visible sign of a scale between the ovules and the bracts. The arrow in the Juniperus branch points towards the branch tip. The current main hypothesis for the evolution of conifer seed cones is summarized with block arrows. ov: ovule; br: bract; ss: sterile scale; fs: fertile scale. The drawings are not to scale.
MADS-box genes regulate plant development

MADS-box genes in angiosperms: The (A)BC(DE) model

Homeotic mutants are mutant organisms where one distinct morphological structure (like an organ) is replaced by another developing where and when you would expect the original structure to develop. Homeotic flower mutants in plants have been well known for millennia, and appreciated for their beauty.

Based on mutant analysis of homeotic flower mutants in *Antirrhinum majus* (Schwarz-Sommer et al. 1990) and *Arabidopsis thaliana* (Bowman et al. 1989, 1991) the initial ABC model of floral organ identity specification was formulated about 20 years ago (fig. 3A) (for an early version of the model see Bowman et al. 1989; for the textbook ABC model see Coen and Meyerowitz 1991). According to this model there are three classes of functions named the A, B and C functions that are necessary for the specification of the correct identity of the floral organs. Where, in the floral meristem, the gene(s) performing the A function alone are active sepals develop. The genes performing the A and B functions together specify petals, the genes performing the B and C functions together specify stamens and where the gene(s) performing the C function are active alone carpels develop. The C function also involves the termination of the floral meristem after the development of the carpels, preventing the development of more floral organs. The A and C functions are mutually antagonistic (Coen and Meyerowitz 1991; Drews et al. 1991). One conclusion of this is that the C function is necessary for reproductive organs to develop.

Later the D function; specifying ovule identity, was added based on work in *Petunia* (Colombo et al. 1995). The last function to be added to the model was the E function (Theissen 2001) which does not by itself specify the identity of any individual floral organ, but is necessary for correct identity of petals, stamens, carpels and ovules, i.e. it provides the floral context which is required for the B, C and D functions (Pelaz et al. 2000; Honma and Goto 2001; Favaro et al. 2003; Vandenbussche et al. 2003) and it might also be required for sepal development (Ditta et al. 2004). According to this updated version of the ABC model then the floral organ identities are specified as follows: A(+E) function specify sepals, A+B+E functions together specify
petals, B+C+E functions specify stamens, C+E functions specify carpels and D+E functions specify ovules (fig. 3B).

In *Arabidopsis* the A function is performed by the MADS-box gene *APETALA1* (*AP1*) and *APETALA2* (*AP2*), the only non-MADS-box gene among the *Arabidopsis* genes performing these kinds of functions, the B function is performed by the paralogous MADS-box genes *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), the C function is performed by *AGAMOUS* (*AG*) and the E function is performed redundantly by several paralogous members of the *SEPALATA* (*SEP*) subfamily of MADS-box genes (fig. 3B) (for a review see Theissen 2001). It has been shown that two of the aspects of the C function in *Arabidopsis*; the specification of reproductive organ identity and the floral meristem termination, both performed by *AG*, can be separated (Mizukami and Ma 1995). The D function in *Arabidopsis* is performed redundantly by the four paralogous *AG* subfamily genes *AG*, *SHATTERPROOF1* (*SHP1*), *SHATTERPROOF2* (*SHP2*) and *SEEDSTICK* (*STK*) together, although *STK* is the only one of these genes that is exclusively active in the ovule (fig. 3C) (Pinyopich et al. 2003). The MADS-box genes constitute a large eukaryotic gene family (Alvarez-Buylla et al. 2000). MADS is an acronym referring to the four founding proteins; the yeast protein MCM1, AG, DEFICIENS (DEF) and the mammalian serum response factor (SRF) (Schwarz-Sommer et al. 1990).

The first *Antirrhinum*-based version of the model only included two functions, the ones nowadays referred to as the B and C functions (Schwarz-Sommer et al. 1990). In *Antirrhinum* the B function is performed by *DEF* and *GLOBOSA* (*GLO*), both belonging to the same subfamily of MADS-box genes as *AP3* and *PI*. The C function is performed by *PLENA* (*PLE*), which is the orthologue of *SHP1* and *SHP2* in the *AG* subfamily (Causier et al. 2005). There is indirect evidence suggesting that the *SEP* subfamily genes *DEFH200*, *DEFH72* and *DEFH84* in *Antirrhinum* may perform the E function, but so far this has not been confirmed by mutant analysis (reviewed in Davies et al. 2006). There is no recessive mutant that displays the expected phenotypic deviations of an A function mutant, i.e. homeotic conversions in the flower of sepals to carpels and petals to stamens, in *Antirrhinum*.

There have been multiple gene duplication events, combined with gene losses in some plant lineages, in the MADS-box gene family in angiosperms, resulting in a complicated evolutionary history with multiple genes within each gene subfamily in many species. Sub- and neofunctionalizations have followed the gene duplications, and the result of all this is that the genes of each gene subfamily in each species together perform a range of functions, often related to reproduction, but the B and C functions according to the (A)BC model still appear to be preserved in one or more of the genes of the *AP3/PI* and *AG* gene subfamilies, respectively, in each species investigated so far.
However, it is not necessarily a clear case of orthologous genes performing homologous functions. As an example we will look more closely at a few of the gene duplications in the *AG* subfamily. The gene duplication that resulted in the *AG* subclade, containing *AG*, *SHP1* and *SHP2* from *Arabidopsis*, *PLE* and *FARINELLI* (*FAR*) from *Antirrhinum* and *pMADS3* and *FLORAL BINDING PROTEIN6* (*FBP6*) from *Petunia*, and the STK subclade containing *STK* from *Arabidopsis* and *FBP7* and *FBP11* from *Petunia*, occurred early in angiosperm evolution, since both subclades are represented in the basal angiosperm *Nymphaea* (Kramer et al. 2004). It is often assumed that *AG* subclade genes perform the C function and STK subclade genes perform the D function, but in reality it is not always that simple. In *Petunia* *FBP7* and -11 are apparently together sufficient to specify ovule identity (Angenent et al. 1995; Colombo et al. 1995), but in *Arabidopsis* all four *AG* subfamily genes; *AG*, *SHP1*, *SHP2* and STK act redundantly in the ovule identity specification (Pinyopich et al. 2003) (fig. 3C).

A later gene duplication within the *AG* subclade resulted in two groups; one containing *AG*, *FAR* and *pMADS3* and the other containing *SHP1* and -2, *PLE* and *FBP6* (Kramer et al. 2004; Causier et al. 2005; Rijpkema et al. 2006; Zahn et al. 2006). Notice that the C function is performed by the paralogous genes *AG* and *PLE* in *Arabidopsis* and *Antirrhinum*, respectively (Causier et al. 2005). In *Petunia* the C function appears to be performed mainly by *pMADS3* (Tsuchimoto et al. 1993; Kater et al. 1998). The *AG* orthologue *FAR* acts partly redundantly to *PLE*, mostly in the repression of the expression of the genes performing the B function in the forth whorl and to a lesser degree in stamen organ identity, but mostly *FAR* is involved in male reproduction (Davies et al. 1999). The functionally redundant genes *SHP1* and -2 in *Arabidopsis* are the result of an even later gene duplication event (Kramer et al. 2004; Zahn et al. 2006). *SHP1* and -2 have a unique role in fruit dehiscence, but they also act redundantly with *AG* in the specification of carpel identity (Pinyopich et al. 2003) (fig. 3C).

In basal angiosperms the expression pattern in flowers is broader for the *PI/AP3*, and to a lesser degree also the *AG*, subfamily genes than for the corresponding genes in monocots and eudicots (Kim et al. 2005). Based both on the gradual morphological transitions between organs and the gene expression patterns, modifications of the ABC-model in basal angiosperms has been suggested (reviewed in Soltis et al. 2006; Soltis et al. 2008), particularly the ‘fading borders’ model (Buzgo et al. 2004; Kim et al. 2005).

The A function has been strongly questioned, and it is highly unclear if it actually exists as defined in the model, or whether it is just a side effect of the function of the genes supposed to perform the A function in the specification of the floral meristem identity (for recent reviews see Davies et al. 2006; Litt 2007; Causier et al. 2010). The regulation of the expression of the genes performing the B and C functions is also far more complicated than the simple antagonistic relationship between the genes performing the A and
C functions as suggested by the original model (Davies et al. 2006; Causier et al. 2010).

A substantial functional overlap between $AP1$, $AP2$ and the $SEP$ subfamily genes have been found, which confuses the boundary between the A and E functions, for example both classes of genes are involved in the specification of floral meristem identity (for the $AP1$, $AP2$ and orthologous genes in other species see the recent reviews listed above and for the $SEP$ subfamily genes see Uimari et al. (2004), reviewed in Teeri et al. (2006), and Ditta et al. (2004)) and the constitutive expression of either $API$ or a $SEP$ subfamily gene together with $AP3$ and $PI$ is sufficient to convert leaves into petals (the conversion was most complete when both $AP1$ and two different $SEP$ genes; $SEP2$ and $SEP3$, were included) (Honma and Goto 2001; Pelaz et al. 2001).

It is known that MADS-domain proteins can interact both as dimers and as trimers with other MADS-domain proteins (see e.g. Egea-Cortines et al. 1999; de Folter et al. 2005; Immink et al. 2009). In a yeast-4-hybrid experiment a tetramer between PI, AP3, AG and SEP3 from *Arabidopsis* has been demonstrated (Honma and Goto 2001), and when these four genes are together constitutively expressed in *Arabidopsis* cauline leaves are converted into stamenoid organs (Honma and Goto 2001). This result in particular inspired Günter Theissen and coworkers to propose a model, the floral quartet model, which is basically an extension of the genetic ABCE model in *Arabidopsis*, where each function is (presumably) performed by a tetramer of MADS-domain proteins with the two participating dimers in each tetramer recognizing different DNA motifs (Theissen and Saedler 2001). It is an appealing model in its simplicity, but it does not really add much in terms of mechanistically explaining how floral organ development is regulated, there is little actual evidence supporting it *in planta* and the floral quartet model only partly fits the available empirical data. A couple of recent studies has provided some support for this model by showing that a tetramer of SEP3 proteins, or a tetramer of two SEP3 proteins, AP3 and PI, can bind to two separate DNA sites by bending the intervening DNA (Melzer and Theissen 2009; Melzer et al. 2009).

In addition to the interactions between different MADS-domain proteins it is also known that MADS-domain proteins from yeast, mammals and plants can physically interact with other kinds of proteins (Mueller and Nordheim 1991; Honma and Goto 2001; Masiero et al. 2002; Páez-Valencia et al. 2008), although this is often neglected in the theory building that has been based on protein-protein interaction studies. For example AG has been shown to interact with the leucine rich repeat protein FLOR1 in the FLOR1-VSP1 complex (Gamboa et al. 2001; Acevedo et al. 2004) and the sister proteins AG, SHP1, SHP2 and STK can all form trimers with SEP3 and the homeodomain factor BEL1 (Brambilla et al. 2007), a fact which is considered to be important during ovule development. The *Antirrhinum* AG subfamily proteins PLE and FAR, as well as the SEP subfamily proteins
DEFH72 and DEFH200, interact with the non-MADS-domain protein and putative transcription factor MIP1 (Causier et al. 2003). Interestingly, one protein that is notably lacking among the interaction partners detected so far is AP2, the supposed main A function protein in *Arabidopsis*.

Figure 3. The ABC(DE) model of floral organ identity specification in dicots. A) The classical textbook ABC model (Coen and Meyerowitz 1991). The protein names included for each function are those of *A. majus*. B) The ABCDE model of floral organ identity determination based mainly on mutant analysis in *Arabidopsis*, *Antirrhinum* and *Petunia* (Coen and Meyerowitz 1991; Theissen 2001) is probably the currently most widely accepted version of the model. The protein names included for each function are those of *A. thaliana*. C) The functional roles of the *Arabidopsis* AG subfamily genes AG, SHP1, SHP2 and STK as determined by genetic studies. Modified from Pinyopich et al. (2003).
MADS-box genes in gymnosperms

The first gymnosperm MADS-box genes identified were DEFICIENS-AGAMOUS-LIKE1 (DAL1), DAL2 and DAL3, belonging to the AGAMOUS-LIKE6 (AGL6), AG and TM3/SOCl subfamilies of MADS-box genes, respectively (Tandre et al. 1995). This discovery demonstrated for the first time that MADS-box genes homologous to genes required for the development of the flower were present also in gymnosperms (Tandre et al. 1995). Using directed candidate gene approaches and in some cases also tissue specific EST libraries in a variety of gymnosperms representing all four extant orders, orthologues of most of the key genes involved in floral meristem development and floral organ identity specification have been identified. These include MADS-box genes representing a number of different subfamilies, most importantly the AG subfamily and the PI/AP3 subfamily. Notably missing so far are the AP1 and SEP subfamilies, but representatives of the related AGL6 subfamily of MADS-box genes are present in gymnosperms (Zahn et al. 2005). Other important MADS-domain subfamilies known to be represented in gymnosperms are the SOCl subfamily (Tandre et al. 1995; Walden et al. 1998; Winter et al. 1999) and the B_sister subfamily (Becker et al. 2002). Unfortunately, there are no mutants available for any of the gymnosperm MADS-box genes. Because of this the function of these genes can only be studied using indirect methods and therefore the specific function of these genes in the gymnosperms is still unclear.

There are several paralogous genes in the PI/AP3 subfamily in gymnosperms and these genes are generally expressed in the male, but not female, reproductive structures in gymnosperms (Mouradov et al. 1999; Sundström et al. 1999; Winter et al. 1999; Becker et al. 2000; Fukui et al. 2001; Sundström and Engström 2002; Becker et al. 2003). The spatio-temporal expression pattern of the gymnosperm, particularly the Pinaceae, PI/AP3 subfamily genes implies a function in the sex determination and development of the male cones (Mouradov et al. 1999; Sundström and Engström 2002; see also Becker et al. 2003) and likely also in the organ identity specification of the male reproductive organ; the microsporophyll (Sundström and Engström 2002; see also Becker et al. 2003). Comparative studies of the dimerization properties of PI/AP3 subfamily proteins from two different gymnosperm orders; Gnetales and conifers, have indicated that the dimerization properties might not be conserved. Using PI/AP3 subfamily proteins from Gnetum gnemon, GGM2 formed homodimers but did not form heterodimers with PI, AP3, DEF, GLO or GGM15 (Winter et al. 2002). However, the sister proteins DAL11 and DAL13 (both orthologues of GGM2) from Picea abies not only formed homodimers, but also heterodimers with each other and with AP3 and possibly also DAL12, a parologue of DAL11 and DAL13 and an orthologue of GGM15 (Sundström and Engström 2002).
The situation in the AG subfamily is quite different. In this case only one representative has been identified in each gymnosperm species studied until now (a cycad, ginkgo, a Gnetum species and several species of conifers) (Tandre et al. 1995; Rutledge et al. 1998; Winter et al. 1999; Jager et al. 2003; Liu et al. 2003; Zhang et al. 2004; Futamura et al. 2008), and the predicted protein sequences of these genes are very similar. This indicates that unlike the AG genes in angiosperms and also unlike other MADS-box gene subfamilies in gymnosperms, this gene seems to have a very high degree of sequence conservation and not to be duplicated. The gene isolation methods used do not exclude that there are undetected duplicated gene copies in the genome, but if that is the case they most likely have a divergent sequence or are not expressed in cones.

AG subfamily genes in gymnosperms are expressed in developing male and female reproductive organs and ovules (Tandre et al. 1995; Rutledge et al. 1998; Tandre et al. 1998; Winter et al. 1999; Jager et al. 2003; Liu et al. 2003; Zhang et al. 2004), which suggests that they have conserved functions in reproductive organ development and are probably also, as in angiosperms, required for reproductive organ identity specification (Tandre et al. 1998). The expression pattern of the two conifer AG subfamily genes DAL2 and SAG1 in Norway spruce and Black spruce, respectively, has been studied in particular detail in developing ovuliferous scales where, in both cases, this gene is expressed throughout the developing ovuliferous scale already from early developmental stages (Rutledge et al. 1998; Tandre et al. 1998). Further indirect support for the conservation of the functional capacity of the gene products of these gymnosperm genes is provided by the phenotypic deviations (very similar to those caused by the constitutive expression of AG) caused by constitutively expressing the genes in Arabidopsis thaliana (Rutledge et al. 1998; Tandre et al. 1998), and particularly by the fact that expression of the cycad gene CyAG can rescue the ag mutant (Zhang et al. 2004).

In the AGL6 subfamily there appears to have been a gene duplication early in the history of gymnosperms, resulting in two paralogous gene lineages (Li et al. 2010). The most well studied AGL6-like gene in gymnosperms is DAL1 from Picea abies. DAL1 is hypothesized to be involved in the juvenile-to-adult transition, i.e. the transition from reproductively incompetent to reproductively competent (Carlsbecker et al. 2004). This suggestion was based mainly on the DAL1 expression pattern with the expression level increasing with both the age of the tree and with the successive initiation of vegetative structures within the tree; i.e. the relative DAL1 expression level within the tree is highest in vegetative shoots in the apical part of the tree, and further supported by the extremely early flowering phenotype of Arabidopsis thaliana constitutively expressing DAL1 (Carlsbecker et al. 2004). Both DAL1 and other AGL6-like genes in gymnosperms have been shown to be active in cones (Liu and Podila 1997; Mouradov et al. 1998; Shindo et al. 2004).
1999; Winter et al. 1999; Becker et al. 2003; Carlsbecker et al. 2004; Brenner et al. 2005), but any general function of this category of genes in cones is currently unclear.

There also appears to be gymnosperm specific subfamilies of MADS-box genes, like the DAL10 subfamily (Carlsbecker et al. 2003). DAL10 from Norway spruce has been hypothesized to be involved in specifying the reproductive identity of shoot primordia (Carlsbecker et al. 2003), upstream of the putative function of the PI/AP3 subfamily genes DAL11 and/or DAL12 in specifying the developing cones as male rather than female (Sundström and Engström 2002) and of the putative functions of the AG subfamily gene DAL2 and the PI/AP3 subfamily gene DAL13 in specifying the reproductive organ identity within the context of the developing cones (Tandre et al. 1998; Sundström and Engström 2002).

Using a developmental genetics approach to evaluate paleobotany-based hypotheses on the morphological evolution of Taxodiaceae and Cupressaceae s.str. seed cones (paper I)

The conifer families Taxodiaceae and Cupressaceae s.str. are closely related, and Cupressaceae s.str. appears to have evolved from a taxodiaceous ancestor (see e.g. Gadek et al. 2000). The seed cones of these families seemingly reflect this history with the taxodiaceous seed cones having a more complex morphology with scales developing (usually) between the ovules and the sterile bract and the Cupressaceae s.str. seed cones apparently lacking scales. The morphology of both categories of seed cones does not fit the predicted morphology based on Florin’s hypothesis, which says that in addition to the bract there should be an ovuliferous scale from which the ovules develop (Florin 1951, 1954). Florin was aware of this problem, and his solution was to add specific hypotheses regarding the identity of the ‘ovuliferous scale’ in various species representing these families (Florin 1951).

In Cryptomeria japonica (Taxodiaceae) he hypothesized that the scale, known as the tooth, that appears after ovule initiation, is an ovuliferous scale homologous to that of Pinaceae seed cones, despite the fact that it appears after the ovule it is supposed to give rise to (Florin 1951). In the Cupressaceae s.str. species Thujopsis dolabrata and Juniperus communis he hypothesized that an ovuliferous scale homologous to that in Pinaceae still exists as a functional unit, but that it is completely fused to the subtending bract and therefore not morphologically distinguishable (Florin 1951). He did not address the problem that in these species the ovules do not actually develop from the bracts, which would be predicted from his hypothesis.
We decided to take advantage of what is known about ovuliferous scale development at the molecular level in spruce (Rutledge et al. 1998; Tandre et al. 1998) to test Florin’s hypotheses; i.e. that ovuliferous scales, homologous to the Pinaceae ovuliferous scale, are present in Cryptomeria japonica, Thujaopsis dolabrata and Juniperus communis, and specifically that these ovule-bearing organs are present where he stated for each species.

In order to obtain molecular markers for developing conifer ovuliferous scales in the Taxodiaceae and Cupressaceae species I isolated, using RACE with a number of different degenerate primers, cDNA clones representing three new genes, the DAL2/SAG1 orthologues CjMADS4, TdMADS3 and JcMADS2 from C. japonica, T. dolabrata and J. communis, respectively (paper 1). Using RNA in situ hybridization I then analyzed the expression pattern of these genes in developing seed cones of different developmental stages to determine if these molecular markers were expressed where Florin had hypothesized that the ovuliferous scales should be.

If Florin was correct one would expect TdMADS3 in T. dolabrata to be expressed in the adaxial (i.e. upper) part of the bracts subtending the row(s) of ovules and JcMADS2 in the related species J. communis to be expressed in the adaxial part of the uppermost tier of bracts, which are alternating with rather than subtending the ovules. However, what I found was that in pre-dormant seed cones, where ovules were already differentiated, both TdMADS3 and JcMADS2 were expressed in the ovules but not in the bracts (paper 1). These gene expression data are not consistent with Florin’s hypotheses regarding the identity of the ovuliferous scale in these two species. We therefore concluded that Cupressaceae s.str. seed cones, particularly the seed cones of Juniperus and Thujaopsis, really do not have any ovuliferous scale, or any other identifiable ovule-bearing organ. Our data can not exclude the possibility that there is an ovule-bearing organ where the DAL2/SAG1 orthologue is not expressed in these cones, but given what is known about AG subfamily genes we find this less likely.

In post-dormant seed cones I did detect expression of both TdMADS3 and JcMADS2 in the bracts (paper 1) representing essentially the expression pattern predicted from Florin’s hypotheses (1951), but this gene expression appears long after the ovules have already developed, and hence it can not be related to an organ producing ovules, i.e. an ovuliferous scale. We therefore find it most likely that this DAL2/SAG1 orthologue gene expression during the late developmental phase reflects a novel function of these genes in bract development, particularly since in both these species the bracts undergo substantial developmental changes at this stage. It is also known that angiosperm AG subfamily genes have been recruited to perform apparently novel functions in fruit development (Liljegren et al. 2000; Itkin et al. 2009; Pan et al. 2009; Vrebalov et al. 2009).

In C. japonica seed cones one would expect CjMADS4 to be expressed in the common ovule-tooth primordium at the early developmental stage, pri-
arily basal to the developing ovule, and later in the entire developing tooth if Florin was correct. In fact, *CjMADS4* is expressed throughout the common ovule-tooth organ primordium in pre-dormant cones, although at the highest level in the developing integument (paper I). However, in the developing tooth of post-dormant *C. japonica* seed cones *CjMADS4* is expressed only in a very limited region at the adaxial side of the lower half of the tooth, close to the ovule. This is a very different expression pattern than that of *DAL2* or *SAG1*, which are both expressed throughout the developing ovuliferous scale. Based on this we conclude that Florin was wrong; the tooth of *Cryptomeria* is not an ovuliferous scale *sensu* Pinaceae.

In conclusion, our developmental genetics data suggest that neither Taxodiaceae nor Cupressaceae s.str. seed cones have any ovuliferous scales, carrying the ovules, that are homologous to the Pinaceae ovuliferous scale. If this is correct, it implies that the evolution of conifer seed cones has been more divergent than previously thought.

We find this to be the simplest explanation to our data, but there are alternative explanations. One is that the ovuliferous scales in Cupressaceae s.str. have been reduced to consist of only the ovules (see e.g. the discussion in Farjon and Ortiz Garcia 2003; Schultz et al. 2003; Schulz and Stützel 2007), which would explain our expression data in the pre-dormant cones, but not in the post-dormant cones. It might also be noted that actually the part of the tooth that is supposed to be derived from the fertile scales of the dwarf shoot (Florin 1951) match reasonably well the expression pattern of *CjMADS4* in the developing tooth, and hence this gene expression might be taken as an argument for Florin’s hypothesis rather than against it. If this is the case, it then becomes necessary to explain why *DAL2* and *SAG1* are expressed in the entire scale covering both the presumably fertile and sterile parts, rather than as expected only in the fertile part. We do not currently know if one gene expression pattern is more ancient than the other or if the scales in these different families have originated independently. A concept which should be mentioned in this context is heterochrony, i.e. that changes over generations in the timing of developmental events during the ontogeny result in altered adult morphology of later generations compared to previous generations (Gould 1977, 2000). Such a process might ‘explain’ how an ‘ovuliferous scale’ giving rise to the ovules might end up developing after the ovules and hence no longer functionally be an ovuliferous scale while still being homologous to ovuliferous scales in other species. This might provide an alternative explanation to the gene expression patterns in post-dormant Cupressaceae s.str. seed cones, although it is then necessary to explain how and why the timing of the development of this organ shifted with respect to the organs that should develop from it.
Evolution of MADS-box genes

In animals and fungi there are only one or a few MADS-box genes in each species, but during plant evolution this gene family has expanded enormously (Alvarez-Buylla et al. 2000). There are four main groups of living land plants: the bryophytes (mosses, liverworts and hornworts), the lycophytes (Lycopodiaceae, Isoetes and Selaginella), a diverse group consisting of ferns, horsetails and wisk ferns and the seed plants (Kenrick and Crane 1997; Pryer et al. 2001).

There are two main groups of MADS-box genes in eukaryotes, type I and type II (Alvarez-Buylla et al. 2000, fig. 1 and 4). Both groups have expanded in plants (Alvarez-Buylla et al. 2000), but still little is known about the function of plant type I MADS-box genes. Much more is known about the type II genes, and in particular the plant specific kind of type II genes that have a second conserved domain called the K-domain, named after a predicted structural similarity to keratin (Ma et al. 1991, fig. 1 and 4). This category of genes is called MIKC-type MADS-box genes after their protein domain structure (fig. 4), with the MADS-domain first followed by a variable intervening (linker) domain (the I-domain) and then the K-domain. After the K-domain there is a C-terminal domain of varying length. The K-domain is considered to consist of multiple repeats of the heptad sequence form \((A-b-c-D-e-f-g)_n\), where more than 75% of \(A\) and \(D\) positions are occupied by apolar residues and the other positions are mostly polar or charged residues (Steinert and Roop 1988). These areas of repeats are occasionally interrupted by linker regions. This primary structure favors a coiled coil secondary structure (Steinert and Roop 1988). However, so far this secondary structure is only a theoretical prediction, and no MADS-domain protein with a K-domain has ever had its structure actually determined by X-ray crystallography. Consequently there are different, only partly compatible, predictions of the helices of the K-domain in the literature (Ma et al. 1991; Yang et al. 2003, see also paper III). The K-domain is thought to be involved in protein-protein interactions in particularly AP3/PI subfamily MIKC-type MADS-domain proteins (Davies et al. 1996; Yang et al. 2003; Yang and Jack 2004) but it is also sufficient for protein-protein interactions between AG and each of SEP1, SEP2, SEP3 and AGL6 (Fan et al. 1997). The C-terminal domain has been shown to, in some MADS-domain proteins, be required to activate transcription of downstream target genes (e.g. Honma and Goto 2001; Cho et al. 2004) and it has also been implicated in the formation of higher order (trimer) protein complexes (Egea-Cortines et al. 1999), although it might actually be the predicted K3 helix (Yang et al. 2003) spanning the border between the K-domain and the C-terminal domain as originally defined (Ma et al. 1991) that is the most important region for this function (Immink et al. 2009; Melzer and Theissen 2009).
MIKC-type genes are known from charophycean green algae (Tanabe et al. 2005), liverwort (Zobell et al. 2010), moss (Henschel et al. 2002; Riese et al. 2005), lycophytes (Svensson et al. 2000; Tanabe et al. 2003), ferns (Münster et al. 1997; Hasebe et al. 1998; Münster et al. 2002), gymnosperms and angiosperms. Charophyceae is believed to be the closest living relatives of the land plants (Kenrick and Crane 1997; Taylor et al. 2009). No MIKC-type MADS-box genes have been found in the genomes of the non-charophycean green algae *Chlamydomonas reinhardtii* or the red algae *Cyanidioschyzon merolae*, although they each have a single MADS-box gene without any K-box (Tanabe et al. 2005).

**Figure 4.** Almost all the type II MADS-domain proteins in plants, but not in animals or fungi, have the MIKC domain structure. Some MIKC-type proteins also have an N-terminal domain before the MADS-domain. At the end of the C-terminal domain of the MIKC type proteins is one or two short subclass specific motif(s) of unknown function. Plant type I MADS-domain proteins have a different domain structure of varying length, lacking the K-domain.

Only a few MADS-box genes have been found in all the non-seed plants studied so far. This suggests that the large expansion of the MADS-box gene family occurred sometime between the last common ancestor of seed plants and ferns and the last common ancestor of the extant seed plant orders (fig. 1), the latter living no later than the early Pennsylvanian (Beck 1988). This pattern is in contrast with two of the main transcription factor families active in plant development; the bHLH protein family and the homeodomain protein family, where the great expansion probably happened early in land plant evolution since there is a steep rise from only a few genes in chlorophytan green algae to large gene numbers in moss, representing most of the gene subfamilies present in seed plants (Mukherjee et al. 2009; Pires and Dolan 2010). This indicates that the evolutionary pattern of MADS-box genes is not simply a matter of seed plants having larger families of transcription factors than non-seed plants, but that it is probably directly linked to the evolution of seed plants, and therefore MADS-box genes are particularly interesting to study in relation to the evolution of development and the evolution of reproductive structures in seed plants.
What was the function of the ancestral AG subfamily gene? (paper II)

AG-subfamily MADS-box genes have been identified in representatives of the angiosperms, cycads, conifers, ginkgoales and gnetales, but no representatives have been found in ferns, lycophytes or bryophytes. For bryophytes and lycophytes the absence of representatives of this subfamily of MADS-box genes is further demonstrated by the whole genome sequencing of a moss (*Physcomitrella*) and a lycophyte (*Selaginella*). The most likely interpretation of this is that the AG subfamily originated, seemingly suddenly, by gene duplication early in the seed plant lineage.

In paper II we utilized the fact that in the last few years the number of conifer EST clones available in public databases has increased enormously. In this paper we present the two new *Picea abies* genes DAL18 and DAL20, which were isolated based on sequence similarity to novel kinds of conifer MADS-box genes found in a public EST database. These two genes are the founding members of two new basal subclades; the DAL18 clade and the DAL20 clade, within the AG subfamily. We found representatives of both these subfamilies in the public sequence databases also from other conifer species, but not from any angiosperm species. More importantly, no recognizable orthologues exist in the sequenced genomes of *Arabidopsis* or rice.

Given the basal position of these genes within the AG subfamily, the most likely explanation is that the DAL18 and DAL20 clade genes have been lost in the angiosperms, probably at the latest before the split between monocots and dicots. The finding of these novel kinds of AG subfamily genes was totally unexpected, particularly since no such genes appear to exist in angiosperms. To us this meant an unexpected opportunity to address the question of what the ancestral function within the AG subfamily was, and to this aim we analyzed the expression pattern of DAL18 and DAL20. Previously studied genes in this subfamily have functions in reproduction, particularly but not exclusively in the development of the male and female reproductive organs and the ovules. The previously known AG subfamily gene from *Picea abies*; DAL2, has an expression pattern that fits the expectation given this kind of function. We found that DAL2 is exclusively expressed in developing male and female cones (consistent with previous results, see Tandre et al. 1995; Tandre et al. 1998) and in developing seeds.

DAL18 turned out to be expressed at a very low level, making the study of its expression pattern difficult. The gene was originally isolated from post-dormant seed cones, showing that it is expressed there, where it could potentially overlap in expression with DAL2. Using qRT-PCR we also managed to confirm low-level expression in the stem cambial region, including differentiating vascular tissue. This is supported by the fact that the EST clone from *Picea glauca* that the isolation of DAL18 was based on was isolated from differentiating secondary xylem from the stem. The expression level was
unfortunately so low in all tissues examined that no reliable comparisons of expression levels or more detailed spatial analysis using RNA in situ hybridization could be performed.

DAL20 was isolated from roots and was found to be exclusively expressed in roots, and we found that the expression was not limited to the root tip indicating that although the gene appears to be expressed in the root tip it is not exclusively expressed in the meristematic region of the RAM. The Picea and Pinus EST clones we found belonging to the DAL20 clade were also in all cases isolated from roots.

The lack of overlap in the spatio-temporal expression pattern of DAL18 and DAL20 means that we can not from this easily resolve what the ancestral function of the AG subfamily was, although since the DAL20 clade is the most basal of the two and DAL20 clade genes appear to be exclusively expressed in roots, this might be a good candidate for further studies. Particularly since XAL1, the founding member of the putative sister clade of the AG subfamily (Martinez-Castilla and Alvarez-Buylla 2003; Tapia-Lopez et al. 2008); the XAL1 subfamily, is mainly expressed in roots where it regulates cell cycle duration and cell proliferation, although XAL1 is also involved in the photoperiodic pathway for flowering time regulation (Tapia-Lopez et al. 2008). If DAL20 turns out to have the same function in roots as XAL1 or as DAL5; the XAL1 orthologue from Picea abies (Carlsbecker 2002), this would indicate that this function represents an inherited ancestral function. Unfortunately, without mutant analysis in conifers it will be difficult to show what DAL20 is actually doing in roots.

Although it is still not clear what the ancestral function was, the fact that DAL20 is exclusively expressed in vegetative tissues and that DAL18 also appears to be mainly, although not exclusively, expressed in vegetative tissues strongly indicates that neither the C nor D function of the ABCDE model was the ancestral function of the AG subfamily genes.

Evolution of the AG subfamily of MADS-domain proteins in conifers and angiosperms (paper III)

The fact that the paralogous genes DAL20, DAL18 and DAL2 from the same species have different gene expression patterns indicates that these genes have evolved by changes in the transcriptional regulation. We wanted to address the question if the AG subfamily genes in angiosperms and conifers have also evolved by changes in the protein coding region, causing the gene products to have altered functional properties. To this aim we used a comparative approach where we studied the dimerization properties of a number of AG subfamily proteins together with XAL1 in yeast cells. This means that we are comparing the capacity, under the same experimental conditions, of proteins (baits) to physically interact with a large, but fixed, set of potential
dimerization partners (preys), that in most cases will never actually come in contact with the baits in planta because they are expressed in different species or in different tissues of the same species. We used two angiosperm AG subfamily proteins (AG and STK from Arabidopsis thaliana), four conifer AG subfamily proteins (DAL2, DAL18 and DAL20 from Picea abies and CjMADS4 from Cryptomeria japonica) and XAL1 from A. thaliana as baits in a large comparative yeast-2-hybrid assay where we screened 119 MADS-domain proteins, mainly from A. thaliana and P. abies, for protein dimerizations with these baits (paper III).

I found (paper III) that, although there were minor differences, the dimerization properties of AG, STK, DAL2, DAL20 and XAL1 are very similar, indicating a very high degree of conservation of the dimerization properties in the AG and XAL1 subfamilies in seed plants over more than 300 million years. This despite the fact that these genes have different functions in planta, as demonstrated by expression analyses and in some cases also mutant analyses. All these proteins interacted specifically with one or more proteins of the SEP/AGL6 and the ANR1 subfamilies of MIKC-type MADS-domain proteins as well as three related proteins in the α clade of type I MADS-domain proteins (for the phylogenetic classifications see Parenicova et al. 2003), and in addition AG also interacted with AGL15 of the AGL15/AGL18 subfamily of MIKC-type MADS-domain proteins and the conifer proteins DAL2 and DAL20 both interacted with the conifer protein DAL12 in the PI/AP3 subfamily of MIKC-type MADS-domain proteins. Similar interaction properties had already previously been demonstrated for AG, SHP1, SHP2, STK and XAL1 using a similar approach with only MADS-domain proteins from A. thaliana (de Folter et al. 2005), which hinted towards this conclusion, but the fact that I have now shown that also the spruce proteins DAL2 and DAL20, especially with DAL20 representing the most basal subclade within the entire AG subfamily, also has very similar dimerization properties to those of the angiosperm proteins confirms that the similarity between the dimerization properties between the A. thaliana AG subfamily proteins and XAL1 are not due to parallel evolution but are the result of conserved functional properties in this entire clade dating back to the last common ancestor of angiosperms and conifers.

The conifer proteins DAL18 and CjMADS4, however, showed drastically different dimerization properties compared to all the other baits used in our study; they did not interact with any of the preys used (paper III). This indicates that while the protein-coding sequence seems to be highly conserved by functional constraints in most of the genes in this clade, there have been potentially functionally important changes in the protein-coding sequence in two separate subclades of AG subfamily genes in conifers; the basal DAL18 subclade and along the gene lineage leading to the DAL2 orthologue CjMADS4 in Taxodiaceae. Given the very high degree of sequence similarity of the predicted proteins representing CjMADS4, TdMADS3 and JcMADS2
we expect that TdMADS3 and JcMADS2 have very similar dimerization properties as CjMADS4. It should be noted that few of the proteins among the prey used in our study were from conifers, and except for CjMADS4 all prey used were from *A. thaliana* or *P. abies*. Therefore we do not know if the protein-protein interaction properties relevant for the actual *in planta* interactions are different between DAL2 and CjMADS4, hence DAL2 and CjMADS4 might still have conserved functions in their respective species.

Of particular interest is the fact that, like the two AG subfamily proteins from *A. thaliana*, the *P. abies* proteins DAL2 and DAL20 as well as XAL1 from *A. thaliana* interacted with SEP1 in our screen (*paper III*). Only the two AG subfamily proteins from *A. thaliana* interacted with SEP3 (*paper III*), which is considered to function as a central ‘glue’ in protein trimers, binding two other MADS-domain proteins together (Honma and Goto 2001; Immink et al. 2009). These results indicate two things. Firstly, that the dimerization capacity of AG subfamily proteins and SEP subfamily proteins represents ancient dimerization properties dating back to the last common ancestor of angiosperms and gymnosperms, regardless of whether only the ancestral AGL6 subfamily gene or both the ancestral SEP and AGL6 subfamily genes were present in this ancestral plant. Secondly, that this capacity may have been further modified with respect to the specific dimerization capacity of the AG subfamily proteins and the SEP3 clade proteins (Zahn et al. 2005) during the evolution of the angiosperms.

A conclusion from the fact that the paralogous spruce genes DAL2 and DAL20 have very different expression patterns; DAL2 is expressed exclusively in reproductive structures and DAL20 exclusively in roots (*paper II*), combined with the very similar dimerization properties of the DAL2 and DAL20 proteins found in *paper III* is that these paralogous genes appear to have functionally diverged mainly by changes in the transcriptional regulation of the genes, rather than by changes in the protein-coding regions of the genes.

**MADS-domain proteins are transcription factors**

MADS-box genes are a eukaryotic family of genes coding for transcription factors. For a protein to be classified as a transcription factor it has to, either alone or as part of a protein complex, be able to do two things; bind to DNA and directly regulate gene transcription. The MADS-domain has been shown to bind to a DNA sequence known as the CArG box in animals (Pellegrini et al. 1995), fungi (Tan and Richmond 1998) and plants (Mueller and Nordheim 1991; Huang et al. 1993), fulfilling the first criterion. However, while MADS-domain proteins certainly bind to DNA, and while differences in DNA binding specificity have been shown between different MADS-domain proteins (Mueller and Nordheim 1991; Pellegrini et al. 1995; Tan and Richmond 1998), there are also indications that differences in DNA binding
specificity might not be the most important aspect of the functional specificity of the proteins. Some of the early work on MADS-box genes in plants involved cutting and pasting bits and pieces of different MADS-box genes together creating chimeric genes and then in various comparative studies expressing these chimeric genes in a plant, where they are translated into chimeric proteins. One of the results obtained by this method is that the MADS-domain can be swapped between different MADS-domain proteins without apparently affecting the ability to specify the floral organ identity when the chimeric genes are expressed in *Arabidopsis* (Riechmann and Meyerowitz 1997). Another result was that the region of the MADS-domain proteins that provides the functional specificity is apparently only partly the main DNA-binding domain; for AG and AP1 the DNA-binding domain seems to provide functional specificity (the MADS- and I-domains were found to be the most important domains in this aspect) but for AP3 and PI the I- and K-domains were found to define the functional specificity (Krizek and Meyerowitz 1996; see also Riechmann et al. 1996).

Proving that these proteins directly regulate the transcription of downstream target genes is more difficult than showing that the protein can bind to DNA. There are a number of potential mechanisms for this, involving for example modifying the chromatin structure, directly interacting with RNA polymerase II or the basal transcriptional complex, preventing DNA binding of the polymerase or other transcription factors or facilitating transcriptional elongation of the mRNA after transcription has been initiated. Unfortunately it is still not known how MADS-domain transcription factors activate or repress the transcription of their primary target genes. In some cases, like AP1 (Honma and Goto 2001; Cho et al. 2004), the SEP proteins (see e.g. Honma and Goto 2001; de Folter et al. 2005) and several representatives of the AGL6 subfamily (Moon et al. 1999; de Folter et al. 2005; Ohmori et al. 2009; Thompson et al. 2009, see also paper IV), it has been shown that the C-terminal domain is required for transcriptional activation, showing that at least some MADS-domain proteins can activate gene transcription alone, and they can do it in yeast cells where other MIKC-type proteins are not available.

**Evolution of the AGL6 subfamily of MADS-box genes in seed plants (paper IV)**

The *AP1* and *SEP* subfamilies of MADS-box genes appear to be angiosperm specific, but the *AGL6* subfamily, which is closely related to the *AP1* and *SEP* subfamilies, is known to be present in both gymnosperms and angiosperms (e.g. Zahn et al. 2005). *Arabidopsis thaliana* has two paralogous *AGL6* subfamily genes; *AGL6* and *AGL13*. *Picea abies* also has two paralogous *AGL6* subfamily genes; *DAL1* and *DAL14*, but these are derived from a separate gene duplication that occurred early in the history of gymnosperms
(see phylogenetic analysis in Carlsbecker 2002; Carlsbecker et al. 2003; Li et al. 2010). These genes must therefore have been retained in the genome for hundreds of millions of years as paralogous genes, most likely because they have at least partly evolved divergent functions. These genes are therefore suitable to use in a study of the evolutionary mechanisms leading to new gene functions in duplicated genes.

There are two evolutionary mechanisms by which the function of paralogous genes resulting from a gene duplication can change. Functional evolution of genes can occur via changes in the regulatory sequences causing different transcriptional regulation of the two genes, leading to different spatio-temporal gene expression patterns. Functional evolution of genes can also occur via changes in the protein-coding sequence, leading to gene products with different functional properties.

Functional evolution by the first of these two mechanisms is by far the easiest to detect, since there are a variety of methods available for analyzing mRNA expression patterns. Using RT-PCR we analyzed the expression pattern of \textit{DAL1} and \textit{DAL14} in a variety of different tissues and organs of \textit{Picea abies}, and found that while the expression pattern appears to be overlapping, there are also substantial differences, for example \textit{DAL14}, but not \textit{DAL1}, was expressed in roots (paper IV). This result indicates that the transcriptional regulation of these two genes is at least partly different, meaning that the first of the two mechanisms have played an important role in the functional diversification of these two paralogous genes.

The effect of the second evolutionary mechanism stated above is more difficult to demonstrate. We used a combination of two different methods that are both based on comparisons of the functional properties of the gene products of these genes in heterologous systems. First we compared the resulting phenotypic alterations caused by the constitutive expression of \textit{DAL1} (Carlsbecker et al. 2004) and \textit{DAL14} (paper IV) in the angiosperm \textit{Arabidopsis thaliana}. Both genes caused very similar phenotypic alterations when constitutively expressed in \textit{Arabidopsis}, and we only detected quantitative rather than qualitative differences between \textit{Arabidopsis} plants expressing \textit{DAL1} and plants expressing \textit{DAL14}. This suggests that in this context the functional properties of the \textit{DAL1} and \textit{DAL14} gene products are very similar.

Next, we used the same comparative yeast-2-hybrid method as for the comparison of the AG subfamily proteins described above, but this time using AGL6, AGL13, DAL1 and DAL14 as baits (paper IV). Since AP1, SEP and AGL6 proteins in angiosperms are all known to, via the C-terminal domain, activate transcription of reporter genes when used as baits in yeast, also in the absence of any prey, I first tested this capacity in AGL6, AGL13, DAL1 and DAL14. I found that AGL6, DAL1 and to a lesser degree AGL13, but not DAL14, when fused to the GAL4 DNA binding domain, activated transcription in yeast. In the case of DAL1 I mapped this capacity to the C-terminal domain. The fact that also the conifer AGL6 subfamily protein
DAL1, like the angiosperm SEP/AP1/AGL6 proteins, is capable of activating transcription in yeast indicates that this represents an ancestral functional capacity for transcriptional activation inherited from the last common ancestor of angiosperms and conifers. It further suggests that these proteins activate transcription by physically interacting with one or more proteins, not MADS-domain proteins, that are part of a conserved transcriptional regulation mechanism between plants and fungi. The fact that DAL14 did not activate the transcription of any of our reporter genes in yeast, further suggests either that DAL14 has lost the capacity to activate gene transcription or that the transcriptional activation properties of DAL14 have changed in a way that prevents DAL14 from performing this function in yeast. Either way, this represents a difference in functional properties between DAL1 and DAL14.

In the comparative yeast-2-hybrid screen I found large differences in the interaction capacity between all four baits used, particularly between the two members in each paralogous pair (paper IV). DAL1 and AGL6 both interacted with a large number of dimerization partners whereas DAL14 and AGL13 both interacted with only a small handful of proteins, showing that there are substantial differences in dimerization properties both between the paralogous proteins DAL1 and DAL14 and between the paralogous proteins AGL6 and AGL13, representing the products of two independent gene duplications. There was very little overlap among the interaction partners detected for any of the baits used.

The conclusion from all these results put together is that both changes in the transcriptional regulation and changes in the protein-coding regions causing different functional properties have contributed to the functional diversification of the paralogous genes DAL1 and DAL14.

Note that this does not say much about the actual protein-protein interactions of DAL1 and DAL14 in Norway spruce, since they are not ever going to be co-localized with most of the preys used in our study (they are mostly from Arabidopsis!). I did however detect some protein-protein interactions among spruce proteins that might have functional significance in spruce. DAL1 interacted both with itself and with DAL14, but it also interacted with DAL10, presumably involved in specifying the reproductive identity of the shoots that will develop into cones (Carlsbecker et al. 2003), with DAL11 and DAL12 which are believed to be involved in determining the male identity of developing pollen cones and DAL13 which is believed to, together with the gene product of the putative C function gene DAL2, specify the organ identity of the male reproductive organ, the microsporophyll (Sundström et al. 1999; Sundström and Engström 2002). DAL1 is co-expressed with all of these genes in developing pollen cones (Sundström et al. 1999; Sundström and Engström 2002; Carlsbecker et al. 2003; Carlsbecker et al. 2004). Also DAL14 interacted with DAL11 and DAL12, but in this case the issue of co-expression is less clear, although there is a possibility of co-expression in post-dormant pollen cones.
De äldsta fossila barrträd man hittat är drygt 300 miljoner år gamla. De äldsta fossila bevisen för blomväxter dyker upp först drygt 150 miljoner år efter de äldsta barrträden och drygt 250 miljoner år efter de första fröna. Hos barrträden utvecklades fröämnen (fröanlagen) i kotor ända från början.

Enligt den hypotes som den svenske paleobotanikern Rudolf Florin lade fram på 1950-talet, baserat på många års studier av fossila kottar, är fröfjället i moderna honkottar den reducerade resten av ett dvärgskott som växte i vecket mellan det sterila fjället och kotteaxeln i de tidiga barrträdkottarna.

Fröfjället bär fröämnen, och är därmed det honliga reproduktionsorganet. Dvärgskotten i de tidiga kottarna hade vanligen massor av sterila fjäll och några få fertila fjäll som bar frön. Över årmiljonerna sammansmälte de fertila och de sterila fjällen i det lilla dvärgskottet och kvar blev dagens fröfjäll. Trots att senare fynd lett till vissa modifikationer, så har Florins hypotes för honkottarnas evolution hos barrträden stått stadig ända sedan 50-talet.

Ett problem med hypotesen är dock att i de flesta barrträdfamiljer förutom tallfamiljen saknas fröfjället i kottarna, vilket borde innebära att de är enkla kottar (med bara en sorts fjäll) och inte sammansatta kottar (med två sorters fjäll, varav fröfjället egentligen är resten av ett sekundärt skott) som barrträdkottar ska vara.

I denna avhandling har jag förutom grankottar (Picea abies, tallfamiljen) tittat närmare på kottarna i representanter för två mycket närbesläktade barrträdfamiljer; sumpcypressfamiljen och cypressfamiljen. Dessa två familjer är så nära släkt (och så lika varandra) att de ofta slås ihop till en familj. De arter jag har studerat är de två japanska barrträden kryptomeria (Cryptomeria japonica, sumpcypressfamiljen) och hiba (Thujopsis dolabrata, cypressfamiljen) och så vår enda svenska representant i cypressfamiljen, nämligen en (Juniperus communis). Kottarna i cypressfamiljen saknar någon som helst form av synligt fröfjäll, vilket Florin förklarade med att fröfjället har sammansmält med det sterila fjället, och numera utgör den övre halvan av det sterila fjället. I sumpcypressfamiljen, från vilken cypressfamiljen ska ha bildats, finns det fjäll vars position i den mogna kotten i alla fall mer eller mindre stämmer med den förväntade positionen för ett fröfjäll, men dessa
fjäll utvecklas efter själva fröämnen (som de ju ska utgöra basen för om de verkliga är fröfjäll).

En vanlig anledning till att honkottarna i många barrträd inte alls liknar tallfamiljens kottar är att de är anpassade för att locka djur som ska sprida fröna. Den bärliknande struktur som i dagligt tal kallas enbär är kanske det mest bekanta exemplet på detta. Själva ”bäret” bildas av att det översta lagret av sterila fjäll börjar växa och sväller upp efter pollineringen, och innesluter fröna som sitter längst upp kring spetsen av kotten. ”Riktiga bär”, som blåbär eller lingon, har samma syfte, men är egentligen frukter som bildas av pistillen, där fröna redan sitter inneslutna, i blomväxter.

MADS-box gener är bland annat inblandade i att specificera organidentiteten hos blommans organ. MADS-boxgener är en eukaryot genfamilj vars genprodukter (MADS-domänproteiner) reglerar uttrycket av andra gener. De jobbar normalt inte ensamma utan verkar i grupp, ofta med andra MADS-domänproteiner. Speciellt viktiga för blommans utveckling är AGAMOUS (AG)-gruppen av MADS-boxgener, för utan dem bildas inga reproduktionsorgan i blommen. Utan SEPELLA (SEP)-gruppen av MADS-boxgener bildas överhuvudtaget inga blomorgan, utan alla dessa ersätts av bladliknande strukturer.

AG-gener behövs alltså för att bilda reproduktionsorganen och även själva fröämnet i blommorna, och dessa gener finns även i nakenfröiga växter (gymnospermer), inklusive barrträd, där de är uttryckta i kottar. I både vår vanliga svenska gran och svartgran (Picea mariana) har man visat att AG-genen är konstant uttryckta i fröfjällen, och senare även i fröämnen. Baserat på detta drog man slutsatsen att hela fröfjället i gran är ett reproduktivt organ. I den här studien har jag isolerat AG-gener (en per art) från kryptomeria, hiba och en. Denna gen har jag sedan, i varje art, använt i uttrycksstudier som molekylär markör för det honliga reproduktiva organet, för att kontrollera om Florins hypoteser angående identiteten på de fröbärande organen i dessa arter var korrekta. Till skillnad från i granens fröfjäll, där AG-genen är uttryckt i hela fjället, så är AG-genen i kryptomeria bara uttryckt i ett litet, begränsat område i det fjäll som utvecklas mellan fröämnet och det sterila fjället, vilket vi tolkar som att detta fjäll inte är ett homologt organ till fröfjällen i tallfamiljen. I cypressarterna hiba och en fann jag att i mycket unga kottar så är AG-genen endast uttryckt i fröämnet, och uttryck i det enda, sterila fjället (där ett fuserat fröfjäll borde finnas) saknades helt, vilket tyder på att Florin hade fel; det finns inte något ”gömt” fröfjäll i dessa kottar. Där emot visade sig AG-genen vara uttryckt i det sterila fjället i äldre kottar, ungefär vid pollineringen, vilket är en tid på året som detta fjäll genomgår stora förändringar i båda arterna och i enkottar bildar själva ”bäret”. Det finns olika möjliga tolkningar av detta, men den vi anser vara mest sannolik är att detta är ett uttryck för en ny genfunktion i dessa kottar. Detta innebär först och främst att den minsta gemensamma nämnaren när det gäller uttryck av AG-gener i barrträdens honkottar är uttryck i fröämnet och inte i fröfjället.
Dessutom innebär det att dessa kottars evolutionära historia är mer kompli- 
cerad än Florins hypotes anger, bl.a. kan sekundära fjäll ha uppstått obero-
ende av varandra i flera barrträdsfamiljer.

AG-generna finns som sagt i fröväxter men så vitt vi vet inte i sporväxter, 
vilket innebär att denna typ av gener troligen uppstod i en gemensam förfä-
der till de nuvarande fröväxterna, en förfader som levde för mer än 300 mil-
joner år sedan. Vi presenterar här två nya gener; DAL18 och DAL20, från 
gran. Dessa gener representerar två nya undergrupper inom AG-generna, och 
båda dessa undergrupper är basala till alla tidigare kända gener i AG-
gruppen. Dessa tillhör alltså ”släktgrenar” på ”genfamiljens släktträd” som 
gått skilja vägar från de övriga generna mycket tidigt i AG-gruppens histo-
ria. DAL20-gruppen är den äldsta av dessa släktgrenar, och varken DAL18- 
och DAL20-grupperna verkar längre finnas kvar hos blomväxterna. Dessa 
nya gener erbjuder en unik och oväntad möjlighet att ta reda på vad funktionen 
hos den ursprungliga AG-genen var för hundratals miljoner år sedan. Vi stu-
derade därför dessa geners uttrycksmönster, och kunde visa att till skillnad 
från de tidigare kända AG-generna, som t.ex. DAL2 i gran, som alla är ut-
tryckta i reproduktiva strukturer, så är DAL20 uteslutande uttryckt i rötter 
medan DAL18 är huvudslakligen uttryckt (fast på mycket låg nivå) i kambi-
umregionen och (på en ännu lägre nivå) i honkottar efter första vintervilan. 
Dessa uttrycksmönster liknar mer de som rapporterats för AG-genernas sys-
tergrupp; XAL1-gener, och innebär troligen att AG-genernas organidentitets-
bestämmende funktion i reproduktionsorganen inte tillhörde den ursprungli-
ga funktionen för denna gengrupp.

Vi gjorde också en jämförande studie av proteinernas förmåga att binda 
till över hundra andra MADS-domänproteiner i en jäst-2-hybrid studie, där 
AG, STK (en systergruppen till AG) och XAL1 från blomväxter backtrav jäm-
fördes med DAL2, DAL18 och DAL20 från gran och CjMADS4 från kryp-
tomeria. AG, STK, DAL2 och även DAL20 och XAL1 visade sig i vårt sys-
tem kunna interagera med huvudsakligen samma proteiner som tillhör ett 
litet antal undergrupper av MADS-domän proteiner, inklusive SEP-proteiner 
från backtrav, medan både CjMADS4 och DAL18 båda avvek starkt genom 
att inte interagera med ett enda av proteinerna i studien. Detta tyder på att 
överlag så har proteinerna i AG och XAL1 gruppen inte förändrats särskilt 
mycket på mer än 300 miljoner år, inte ens proteiner med ett avvikande ut-
trycksmönster som XAL1 och DAL20, men under barträdens evolution så 
har det inom två grupper av AG-generna, DAL18-gruppen och den evolution-
ära linje som leder till CjMADS4, oberoende av varandra skett förändringar 
av proteinerna som kan ha funktionell betydelse. En intressant observation är 
också att de paraloga generna DAL20 och DAL2 i gran, som har helt olika 
upptäcksmönster (DAL20 är uttryckt i rötter och DAL2 i kottar) har mycket 
lika protein-protein interaktionsförmåga, vilket tyder på att de huvudsakligen 
evolverat genom förändringar i deras transkriptionella reglering.

40
Inga SEP-gener, och inte heller de närläktade AP1-generna som är nödvändiga för bildandet av blommristemet, har någonsin hittats i gymnospermer. AGL6-gruppen, som är en systergrupp till SEP- och AP1-generna, finns däremot i gymnospermer. I gran tros AGL6-genen DAL1 reglera övergången från juvenil till vuxen, d.v.s. reproduktivt kompetent fas. DAL1 är också aktiv i själva kottarnas utveckling, men där är dess funktion mer oklar. DAL1 har en syster; DAL14, som härrör från en genduplikation tidigt i gymnospermernas historia. Vi studerade närmare vilka evolutionära mekanismer som varit inblandade i att särskilja funktionerna hos de två generna efter genduplikationen. Vi visade först att DAL1 och DAL14 har överlappande, men inte identiska, uttrycksmönster i gran. Vi studerade även om själva proteinerna har förändrats så mycket att de har olika funktionell kapacitet. För detta ändamål använde vi två olika metoder; en jämförande studie av den resulterande fenotypen när DAL1 och DAL14 från gran uttrycks i backtrav kombinerat med samma jäst-2-hybrid upplägg som vi använt i den tidigare studien, men denna gång jämförande AGL6 och AGL13 från backtrav och DAL1 och DAL14 från gran. Våra resultat från dessa studier visar att både förändringar i regleringen av transkriptionen, vilket lett till olika uttrycksmönster, och förändringar i själva proteinsekvensen, vilket både förändrat proteinernas förmåga att interagera med andra proteiner och att (på okänd väg genom en gemensam eukaryot mekanism) initiera transkription av nedströmsgener, har spelat stor roll för att skilja funktionerna åt hos de två paraloga generna DAL1 och DAL14 i gran efter den genduplikation som gav upphov till dessa syster gener. Liknande processer verkar ha pågått parallellt hos de närbesläktade paraloga generna AGL6 och AGL13 i backtrav.
References


Mizukami, Y., Ma, H. 1995. Separation of AG function in floral meristem determinacy from that in reproductive organ identity by expressing antisense AG RNA. Plant Molecular Biology 28 (5):767-784.


cones of Norway spruce (*Picea abies*) are homologous to the B-class floral homeotic genes in angiosperms. Developmental Genetics 25 (3):253-266.


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 754

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)