AN ABSTRACT OF THE THESIS OF

Amy M. Brunner for the degree of Doctor of Philosophy in Forest Science presented on June 5, 1998. Title: Structure and Expression of Two *Populus trichocarpa* Homologs of the Floral Homeotic Gene *AGAMOUS.*

Signature redacted for privacy.

Abstract approved: _____

Steven H. Strauss

Because of their small genomes, facile clonal propagation, fast growth, and susceptibility to *Agrobacterium* transformation, poplars (genus *Populus*) are widely considered model systems for the application of molecular genetics and biotechnology in forestry. However, a major concern over commercial use of genetically engineered trees is the release of transgenes into wild populations. The goal of this study was to characterize two genes that are expected to be critical for female and male reproductive development in poplars, and thus, could be used for genetic engineering of reproductive sterility. This trait would mitigate ecological risks associated with commercial deployment of transgenic trees by preventing the spread of transgenes via pollen and seed. A secondary goal was to use these genes as probes to help understand the reproductive biology of poplars, whose two-whorled, unisexual flowers are distinct from those of any previously studied plant species.

We isolated and characterized two closely related genes from *P. trichocarpa* (black cottonwood), a native tree of the Pacific Northwest. These genes are homologous to *AGAMOUS (AG)*, a gene controlling reproductive development in the model herbaceous plant *Arabidopsis*. The proteins encoded by *PTAG1* and *PTAG2* are 89% identical, and phylogenetic analysis suggests that they are most closely related to genes which specify the identity of both stamens and carpels in herbaceous plants.

Gene structure is conserved between *PTAG1, PTAG2, AG*, and the *Antirrhinum AG* ortholog, *PLENA*.

The floral RNA expression patterns of the *PTAG* genes are also very similar to those of *AG* and *PLENA*. *In situ* hybridization studies revealed that *PTAG1* and *PTAG2* are expressed in the center of both female and male floral meristems before reproductive organ primordia have initiated, and in developing stamen and carpels. Unlike *AG*, *PLENA,* and other close *AG* homologs, *PTAG* transcripts are detected in vegetative tissue. These results suggest that *PTAG1* and *PTAG2* may function in a largely redundant manner to specify reproductive organs in *Populus*. Therefore, inhibiting the endogenous genes or proteins is likely to be an effective way to genetically engineer reproductive sterility. However, weak vegetative expression may preclude use of their promoters to ablate floral tissues.

Structure and Expression of Two *Populus trichocarpa* Homologs of the Floral Homeotic Gene *AGAMOUS*

by

Amy M. Brunner

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

DOCTOR OF PHILOSOPHY

Presented June 5, 1998
Commencement June 1999

Doctor of Philosophy thesis of Amy M. Brunner presented on June 5, 1998

APPROVED:

Signature redacted for privacy.

Major Professor, representing Forest Science

Signature redacted for privacy.

Chair of Department of Forest Science

Signature redacted for privacy.

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Signature redacted for privacy.

Amy M. Brunner, Author

Acknowledgments

I thank my major professor, Steve Strauss, for his support, guidance, and relentless inspiration for perfection, and my committee for their time and advice. I am grateful to Will Rottmann for daily discussions and advice during my first two years, and to Lorraine Sheppard for helping with techniques and sharing her poplar experience. Kostya Krutoviskii generously provided his expertise in performing phylogenetic analyses.

Sheila Vollmer helped facilitate many aspects of this project and is responsible for the poplar floral diagrams. Victoria Hawkins provided technical assistance one summer as part of an undergraduate research experience program. Information on the linkage of *PTAG1* and *PTAG2* was provided by the work of Steve DiFazio, Stefano Leonardi, and Toby Bradshaw. I am indebted to Virginia Gewin and Gretchen Bracher for their time and expertise in preparing the composite figure of *in situ* hybridization slides. Finally, I wish to thank members of the Strauss Lab for their help collecting and dissecting poplar inflorescences.

Table of Contents

Table of Contents, Continued

List of Figures

List of Tables

Dedication

For Joy and Esther, two independent women from different generations whose example taught me what strength and unselfishness are, to think for myself , to embrace challenges, to live by my own standards, and to not be limited by convention. Thanks for these valuable life lessons, some of which were excellent training for becoming a scientist (on my own terms, of course!).

STRUCTURE AND EXPRESSION OF TWO *POPULUS TRICHOCARPA* HOMOLOGS OF THE FLORAL HOMEOTIC GENE *AGAMOUS*

Chapter I: Introduction

OVERVIEW

This project was conceived not only as a prerequisite to a practical application, but also as a step towards understanding the genetic control of flower development in a dioecious tree. Several characteristics of poplar argue for its designation as a model system for genetic and molecular analyses of long-lived, woody plants. Concurrently, the commercial importance of poplars is increasing markedly. These two facets are converging so that the first genetically modified tree in commercial use may be a transgenic poplar. I initiated this study in order to ultimately provide "tools" for the genetic engineering of reproductive sterility in poplars. This trait would serve as an ecological safety measure --- a means to contain transgenes, preventing their dispersal via pollen and seed.

Guiding this project were the rapidly advancing studies of the genes and mechanisms underlying floral development in the model herbaceous plants, *Arabidopsis thaliana* and *Antirrhinum majus*. In addition, homologs of genes cloned in these two species have been identified in several dicots, monocots and gymnosperms. Although these genes exhibit broad conservation in sequence, expression pattern and function across phylogenetically distant species, there are also important variances. However, the extent of these differences and how they relate to the diversity of floral form is still poorly understood. Poplar flowers and their development are distinct in several ways, particularly in comparison to the species in which floral homeotic genes have been studied. Thus, it would not be suprising to find that the poplar genes controlling floral morphogenesis exhibit some divergence in function and regulation. Studying the poplar

genes will undoubtedly add to our understanding of the evolutionary diversity of floral homeotic genes.

The objective of this project was to isolate and characterize the *Populus trichocarpa* (black cottonwood) homolog of the *Arabidopsis* floral homeotic gene *AGAMOUS (AG)*. We chose to pursue study of the poplar *AG* homolog because *AG* is necessary for both male and female reproductive development and is only expressed in floral tissues. In addition, homologs from other species exhibit similar functions and expression patterns. This suggested that the poplar homolog could be used to engineer reproductive sterility using two different strategies. Specifically, its cDNA could be used to inhibit expression of the endogenous gene, and its promoter could be used to drive expression of a cytotoxic gene, thereby ablating reproductive tissues. Subsequently, two closely related *P. trichocarpa* genes, equally similar to *AG*, were identified and analyzed. This work is presented in Chapter II.

The remainder of this chapter provides the scientific background necessary to place this project in the context of both forestry and basic research on the molecular genetics of flower development. In the realm of forestry, I consider fast-growing plantations with an emphasis on poplars. Next, the role of biotechnology in the commercial application of poplars as well as the reasons and methods for genetic engineering of reproductive sterility are addressed. My review of the genetics of floral development highlights *AG*, its orthologs, and additional genes identified by phylogenetic analysis to belong to the *AG* subfamily. Finally, poplar floral development is discussed. Chapter III contains conclusions and suggestions for future experiments. I also present some thoughts on extrapolating from studies in herbaceous annuals to trees.

LITERATURE REVIEW

Intensively Managed Forest Tree Plantations

For over a decade, reports of deforestation as well as the need to preserve old growth and other native forests have received major attention in international news media. These issues has captured the public's attention and has helped stimulate the development of new forest management practices, including reductions in the amount of wood available from public lands. In the Pacific Northwest, 10 to 12 million acres have been removed from production in the past decade (Polak 1997). Few would disagree with the movement to restore and protect forests. However, this issue is often portrayed in the overly simplistic and unrealistic light of preservation versus harvest, or restoring native forests versus growing trees for profit.

The use of wood products in commerce is desirable and likely to increase in the future. Analyses show that, in most cases, alternatives to wood are more costly and their production is more harmful to the environment (Koch 1992, Sedjo and Botkin 1997). In addition, wood is an essential and basic resource for much of the world's population. Two out of five people rely on fuelwood or charcoal as their main or sole source of domestic energy for cooking and heating (FAO 1997). Furthermore, human population is projected to grow from 5.7 billion to over 7 billion by 2010, insuring that world demand for wood will continue to increase. Developing economies are also augmenting the demand for wood and paper products. In China, a shortage in timber supply of 25% or more is projected by the year 2010 (Zhang et al. 1997). World demand for wood is forecast to rise about 25% over the next 25 years, but wood supply only 15% (Polak 1997). While preserving forests and habitat regionally, North American timberland withdrawals could result in the undesirable consequence of increased use of less productive forests in other regions, where environmental protection may be less rigorous and species loss greater (Koch 1992, Perez-Garcia 1995).

How do we meet the demand for wood products, while promoting the conservation of global natural forests? Several have argued persuasively that high-yield, intensive forest plantations are a part of the answer (e.g., Sedjo and Botkin 1997, Libby 1993, Gladstone and Ledig 1990). Worldwide forest plantation area nearly doubled between 1980 and 1995, and plantations are expected to play an increasingly important role in national forest programs (FAO 1997). This approach relies on genetic improvement of fast-growing species appropriate for the location; not infrequently, the species of choice is an exotic (e.g., radiata pine in New Zealand and eucalypts in Brazil and South Africa). In some cases, clonal forestry is employed, allowing rapid deployment of elite clones matched to site. Most high-yield forests readily produce 10 cubic meters per hectare (m^3/ha) annually and yields of 20 m^3/ha or greater are achieved in some regions, while the growth rate of usable timber in natural forests ranges from 1 to 3 m^3/ha/year (Sedjo and Botkin 1997).

Consequently, plantations could allow a substantial reduction in the amount of land needed to meet major timber and fiber needs, enabling conservation of more natural forest lands. For example, while plantations in Brazil represent only 1.2 % of the total forest area, they account for 60 % of the country's industrial wood production. Similarly, New Zealand plantations cover 16.1% of the forest area, yet they supply 93% of the industrial wood (FAO 1997). Sedjo and Botkin (1997) further contend that use of non-timber fiber crops, such as kenaf and begasse, has both economic and environmental drawbacks. In particular, the annual regime of tilling, cultivating, planting, fertilizing, application of herbicides and pesticides, and harvest make non-timber fiber crops a far less environmentally benign system than a tree plantation with a multi-year rotation.

Recent studies indicate that forests store much more carbon than previously thought, suggesting that better forest management could remove substantial amounts of carbon dioxide from the atmosphere (Moffat 1997). Fast-growing plantations that are harvested and replanted sequester significant amounts of carbon, because there is less decay and wood is used in long-lived products, creating a new carbon sink. By slowing

deforestation, promoting forest regeneration, and increasing the area in plantations and agroforestry systems, approximately 12-15 % of the projected CO_2 emissions from now until 2050 could be offset (FAO 1997). Moreover, plantations could further reduce the rate of CO_2 buildup if they serve as sources of biomass fuel, partially displacing the use of fossil fuels (Perlack et al. 1995, Ranney and Mann 1994).

Correspondingly, the establishment of short-rotation (e.g., 4-10 years) plantations or fiber farms is expanding. In 1996, North American plantings of short rotation woody crops (SRWC) equaled approximately 52,000 ha, and estimates project that this acreage may reach 90,000-130,000 ha by 2006 (Wright and Berg 1996). *Eucalyptus* species constitute approximately 38% of all short rotation plantations, while in temperate regions, poplar, willow and black locust predominate (Perlack et al. 1995). Their primary use is for pulp and paper; however, it is likely that energy will be an increasingly important product. Biomass plantations are currently supplying energy on a commercial basis in Sweden and Brazil; Shell, the world largest oil company is making a large investment in biomass plantations and associated technologies. In the U.S., the Department of Energy's (DOE) Biofuels Feedstock Development Program (BFDP) continues to lead a collaborative effort to develop methods for increasing production of wood for energy (Wright and Berg 1996).

Though plantations offer both direct and indirect environmental benefits, they may have negative impacts as well. However, most appear to be avoidable if appropriate management practices are followed (Sedjo and Botkin 1997, Ranney and Mann 1994, Perlack et al. 1995). Key are the sites chosen for plantations. Obviously, they should not replace unmanaged forests, low-intensity managed forests or habitats crucial to biological conservation. As plantation acreage increases, genetic pollution effects may become a more significant concern. The seriousness of this issue will partly depend on the genetic diversity within plantations and whether plantations are located in environments where interfertile natural populations exist or where plantation trees can establish feral populations and displace native species. When tree plantations replace annual crops, heavily grazed pastures or degraded lands, they can have locally

significant benefits. These include reduced erosion, improved water quality, reduced use of agricultural chemicals and improved wildlife habitat (Ranney and Mann 1994, Perlack et al. 1995). In sum, plantations can reduce the pressures on as well as promote the restoration of natural forests if managed well. In addition, considerable reductions in the amount of greenhouse gases released per unit energy are expected if plantations are a source of biomass energy.

Poplars and Their Uses

Populus (poplars, cottonwoods, aspens) and *Salix* (willows) comprise the family Salicaceae, which have traditionally been considered taxonomically isolated from other families. They are placed in a separate order, Salicales, near the Violales, particularly the tropical family Flacourtiaceae (Cronquist 1988). A recent study considers the Salicales and Violales as taxonomic synonyms in the order Malpighiales (R. Price pers. comm.). Poplars are deciduous, dioecious, wind-pollinated, widely distributed across the Northern Hemisphere, and among the fastest-growing temperate trees. The genus is genetically diverse, with approximately 29 poplar species constituting six sections (Table 1.1). The earliest fossil record of *Populus* dates to about 58 million years ago (mya); sections *Aigeiros*, *Tacamahaca* and *Populus* are the most advanced (Eckenwalder 1996). Natural hybridization is common, occurring freely within sections and between species of different sections in some cases. *P. trichocarpa*, the experimental system for this study, is a member of section *Tacamahaca*, and like all poplars is a pioneer species. Though usually found in riparian areas, *P. trichocarpa* also inhabits upland sites with adequate moisture. It's geographic distribution ranges from southern Alaska to northern Baja California, and eastward to the Rocky Mountains in Idaho, Montana and Canada. *P. trichocarpa* is one of the largest *Populus* species and the tallest, fastest-growing hardwood in the West; mature trees typically attain heights of 125-150 ft. and live 200 years (Niemiec et al. 1995).

Table 1.1 The Genus *Populus* (adapted from Eckenwalder 1996)

Section	Species[1]	Distribution
Abaso	*P.mexicana*	N. America
Turanga	*P. euphratica*	E. Eurasia
	P. ilicifolia	E. Africa
	P. pruinosa	E. Eurasia
Leucoides	*P. lasiocarpa*	E. Eurasia
(swamp poplars)	*P. glauca*	E. Eurasia
	P. heterophylla	N. America
Aigeiros	***P. nigra***	Eurasia, N. Africa
(cottonwoods, black poplar)	***P. deltoides***	N. America
	P. fremontii	N. America
Tacamahaca	*P. angustifolia*	N. America
(balsam poplars)	***P. balsamifera***	N. America
	P. ciliata	E. Eurasia
	P. laurifolia	Eurasia, N. Africa
	P. maximowiczii[2]	E. Eurasia
	P. simonii	E. Eurasia
	P. suaveolens	E. Eurasia
	P. szechuanica	E. Eurasia
	P. trichocarpa	N. America
	P. yunnanensis	E. Eurasia
Populus (formerly *Leuce*)	*P. adenopoda*	E. Eurasia
(aspens, white poplars)	***P. alba***	Eurasia, N. Africa
	P. grandidentata	N. America
	P. guzmanantlensis	N. America
	P. sieboldii	E. Eurasia
	P. simaroa	N. America
	P. tremula	Eurasia, N. Africa
	P. tremuloides	N. America

[1]Species commonly used in cultivation are shown in bold.
[2]This classification is generally used by breeders; Eckenwalder (1996) classifies it as *P. suaveolens.*

Interspecific hybridization followed by clonal selection characterizes most poplar breeding programs (reviewed in Stettler et al. 1996, Stanton and Villar 1996, Bisoffi and Gullberg 1996). Intrasectional hybrids within *Tacamahaca*, *Aigeiros* and *Populus* as well as intersectional hybrids between *Tacamahaca* and *Aigeiros* are commonly employed. (Dickmann and Stuart 1983, Heilman et al. 1995). Reproductively active trees can be crossed by forcing flower buds in a greenhouse, with abundant seed set occurring in two to eight weeks, depending on the species; southern provenances of *P. deltoides*, however, require three to five months for maturation (Stanton and Villar 1996). Clonal selection criteria vary somewhat with the species and environment. In addition to growth rate, disease resistance is often a primary selection criterion. Heterotic F_1 hybrids are not uncommon, and ease of vegetative propagation via shoot or root cuttings enables rapid commercial deployment of select clones in several species. In many cases, commercial planting is routinely done with dormant, unrooted cuttings. Long-term, recurrrent breeding programs are also under development (Dinus and Tuskan 1997, Bisoffi and Gullberg 1996). Ultimately, quantitative trait locus (QTL) maps may play an important role in refining poplar breeding strategies and enabling marker-aided selection (Bradshaw 1996).

Although poplars have been cultivated since historical times, they are undergoing a renaissance. This is perhaps most evident in the Pacific Northwest. In 1978, with support from the DOE, the University of Washington and Washington State University initiated a joint program to develop and evaluate the production potential of poplar hybrids under intensive culture (Heilman et al. 1995). Industry began trials of these hybrids in 1981. As of 1996, almost 27,000 ha were in production in the Pacific Northwest, with eight to ten companies adding acreage (Wright and Berg, 1996). Intersectional hybrids of *P. trichocarpa x P. deltoides* (T X D) are by far the most prevalent, but additional hybrids are being tested, including P. trichocarpa x P. maximowiczii, P. trichocarpa x P. nigra and P. deltoides x P. nigra. Annual height growth increments range from 3-5 m under operational conditions. Seven to eight year rotations of T X D hybrids grown in the lower Columbia River Valley produced 17-21

Mg/ha/yr of oven dry woody biomass; approximately 70% of the biomass was captured as wood chips, equivalent to a production rate of 35-42 m3/ha/yr (Schuette 1995). Growth rates of the fertigated plantations east of the Cascades are 30-50% higher (Zsuffa et al. 1996). Yields from commercial poplar plantations are five to ten times greater than those from native conifer forests (Bradshaw 1998).

The success of poplar culture in the Pacific Northwest has also stimulated development of hybrids and elite clones for commercial deployment in other regions of the U.S., particularly the North Central states and the South (Wright and Berg 1996). U.S. acreage in poplar could increase dramatically if energy crop production systems become a reality. The amount of land suitable for growing energy crops exceeds 20 million hectares, and a significant portion of this is suitable for poplar cultivation (Hohenstein and Wright 1994). Although the current emphasis is on pulp and energy production, longer rotation poplar plantations may also become significant as a source of solid wood, due to growing timber scarcity and associated price premiums. Technological developments, such as laminated beams and boards, are also opening new markets for poplar wood.

Poplar cultivation is ongoing in many regions of the world (reviewed in Zsuffa et al. 1996). For example, farmers in India established an estimated 26,000 ha of *P. deltoides* in the early 1990's; harvests will be sold to industry for the manufacture of products such as matches and plywood. In 1995, there were more than 1 million ha of poplar plantations in Europe and 1.3 million ha in China (FAO 1997). Intensive poplar research and breeding programs are in progress in Europe, and an expansion of poplar plantations is anticipated. The potential of the many poplar species and hybrid combinations is still poorly understood. Further research is likely to result in better adapted clones for certain regions as well as to extend poplar cultivation to new regions.

In addition to supplying wood and biomass, poplars have long been planted for windbreaks, shelterbelts and landscaping. An emerging use for poplars is in phytoremediation. In addition to protecting stream banks from erosion, poplar's extensive root system makes it well-suited to absorb or detoxify contaminants, such as

agrochemicals from farm runoff (Stomp et al. 1994, Moffat 1995, Dix et al. 1997). Poplars are also being planted for safe disposal of wastewater and sewage sludge and for bioremediation of polluted sites. Several genotypes absorb and tolerate heavy metals, aluminum, nitrates, and herbicides (reviewed in Dix et al. 1997).

Poplar Genetics and Breeding for Commercial Use

Many of the same reasons that make *Populus* so amenable to intensive, short-rotation cultivation support its status as a model system, particularly for genetic analysis and manipulation of woody plants (for a detailed discussion see Bradshaw 1998). In addition to the features already described, all poplar species contain the same number of chromosomes (2n = 38), and the nuclear genome is relatively small (2C = 1.1 pg; equivalent to a haploid genome of approximately 500Mb), about 5 times that of *Arabidopsis* (Bradshaw and Stettler 1993; Dean and Schmidt 1995). Using a three generation hybrid poplar pedigree and DNA-based markers, a detailed genome map has been constructed (Bradshaw et al. 1994). A number of QTLs have been located, including those for stem growth and form (Bradshaw and Stettler 1995) and disease resistance (e.g., Newcombe et al. 1996, Cervera et al. 1996). Using different families, five additional maps are being generated, and markers previously used by Bradshaw and co-workers (1994) are being incorporated, which will aide in determing QTL stability across genetic backgrounds (Cervera et al. 1997). Furthermore, the ratio between physical and genetic length is approximately 200 kb/cM (Bradshaw et al. 1994), making the cloning of genes via map-based approaches feasible. In contrast, the large, repetitive genomes of conifers put this beyond the reach of current technology.

Complementing the poplar genome studies are the well-developed transgenic systems for *Populus* (reviewed in Han et al. 1996). Numerous genotypes have been transformed using *Agrobacterium*. Genetic engineering of poplars is feasible and could make important contributions to the efficiency of plantation systems. Because of the ease of vegetative propagation in most species, valuable engineered genotypes could be

amplified rapidly for testing and commercial deployment. Effective genes for resistance to insects and herbicides are available (e.g., *Bacillus thuringiensis* (*Bt*) endotoxins, proteinase inhibitors, and glyphosate resistance genes), and could aid plantation culture. Poplar is susceptible to several very damaging insect pests and even low levels of competition from weeds (Dickmann & Stuart 1983, Ostry et al. 1989, Heilman et al. 1995). Weeds and grasses also provide cover for voles, which can girdle and kill trees. Defoliating insects are the most numerous category of pests, though boring insects are also common.

As the acreage of poplar plantations increases, insects may become a more serious problem. In eastern Washington and Oregon, the cottonwood leaf beetle can cause enough damage to require multiple pesticide applications; *Bt* toxin genes effective against the beetle have been identified and introduced into hybrid poplar (Strauss et al. 1997). In general, results from greenhouse and field tests with insect- and herbicide-resistance transgenic poplars have been very positive (e.g., McCowan et al. 1991, Miranda-Brasiliero et al. 1992, Donahue et al. 1994, Leple 1995, Meilan et al. 1997, Strauss et al. 1997a, Ellis and Raffa 1997). Barker et al. (1997) estimated that use of herbicide-resistance hardwoods could reduce production costs by as much as 25%, while insect-resistant cottonwoods could result in a 10% cost reduction. These transgenic poplars may enhance environmental as well as production aspects of plantation culture by decreasing the use of insecticide sprays (Strauss et. al. 1991, Raffa 1989) and allowing preferential use of the safest herbicides as well as reducing tillage (Duke et al. 1991, Strauss et al. 1997b).

Also of considerable interest to the pulp and paper industry are trees with reduced lignin content or altered lignin composition to facilitate its extraction. Such traits are projected to have both economic and environmental benefits due to lower pulping costs, higher pulp yields and reduced mill waste. Several groups are pursing research in this area and have successfully downregulated key enzymes in the lignin biosynthesis pathway via antisense and cosuppression methods. Analyses of transgenic poplars, including field tests, are ongoing (reviewed in Boerjan et al. 1997). Genetic

engineering strategies are also being employed to improve poplar for phytoremediation. These include increasing root mass to enhance microbial activity via transformation with *Agrobacterium rhizogenes* and introducing genes for degradation of trichloroethylene (Stomp et al. 1994). In addition, aspen transformed with an *E. coli* glutathione reductase gene displayed resistance to oxidative stress induced by air pollutant (sulfur dioxide) or the herbicide paraquat (Endo et al. 1997).

Another target trait for genetic modification is disease resistance. Leaf and stem diseases are the major limitations to plantation yield in most regions of the world, making disease resistance a major selection criterion in most poplar tree improvement programs. Poplars are susceptible to a diverse range of fungi; *Melampsora* leaf rust, *Venturia* leaf and shoot blight, and *Septoria* stem canker are among the most damaging (reviewed in Newcombe 1996). Despite screening for disease resistance, few clones have both high levels of resistance and desirable growth characteristics. Hence, considerable effort is now focusing on identifying molecular markers for resistance to various diseases (reviewed in Bradshaw 1996, Cervera et al. 1997). Amplified Fragment Length Polymorphism (AFLP) markers are proving especially useful. Researchers have identified AFLP markers tightly linked to the locus conferring resistance to one of the most damaging fungi (*M. larici populina*) to poplar in Europe as well as to a pathogen (*M. medusae* f.sp. *deltoidae*) problematic in North America (Cervera et al. 1996, Stirling et al. 1998). Consequently, the first poplar gene isolated via map-based cloning may be a disease resistance gene. Additional approaches to isolating poplar disease resistance genes are also possible. Disease resistance genes from different species against different pathogens have many features in common; thus, conserved domains or motifs could be exploited to isolate homologous genes from poplar (reviewed in Hammond-Kosack and Jones 1997).

In summary, considerable evidence is accumulating that molecular techniques can enhance tree breeding programs along with plantation maintenance and production efficiencies. Compared to annual crops, trees present unique problems that may benefit from application of these techniques. Long-generation intervals, poor juvenile-mature

trait relationships and difficulties in identifying and combining valuable genes by sexually crosses limit tree improvement. However, effectively merging molecular approaches with classical breeding programs is not a simple process (discussed in Dinus and Tuskan 1997). The features that make poplar a model tree for molecular genetic studies and genetic modification, also make poplar culture a model system for the integration of molecular methods into tree improvement programs. Additionally, ecological and political concerns associated with commercial use of transgenic trees require the development of novel management strategies.

Biosafety of Transgenic Trees

Containment of transgenes inserted into poplars may be desirable or necessary before broad commercial uses are possible. In the U.S., the environmental release of transgenic plants is regulated by the U.S. Department of Agriculture's (USDA) Animal and Plant Health Inspection Service (APHIS) and the Environmental Protection Agency (EPA). When a transgenic crop has been approved for commercial use by APHIS, it is "deregulated", though additional environmental monitoring may be required during commercial cultivation. Regulations continue to evolve (recently reviewed in James et al. 1998). Concurrently, the risks and benefits of transgenic crops, approaches to safety assessment, and the regulation of transgenics continue to be discussed in both scientific circles and the general media (e.g., James et al. 1998, Kappeli and Auberson 1997, James 1997, Snow and Moran Palma 1997, Steinbrecher 1996). Nonetheless, over 3,400 field trials of transgenic plants are in progress or completed, and 30 transgenic crops have been commercially released in the U.S. or approved for release as of May 1998 (USDA-APHIS, 1998). These crops include species, such as *Brassica napus* and *Cucurbita pepo*, that are capable of genetic exchange with wild relatives.

Furthermore, among biologists who study transgenic issues, a general consensus exists on some points (Snow and Moran Palma 1997). Scientists first formally expressed most of these points about ten years ago when field tests of transgenic plants

had just begun (NAS 1987, NRC 1989, Tiedje et al, 1989). Most scientists agree that the process (i.e., genetic engineering) used to produce a transgenic crop poses little added risk of ecological impact. Rather, the focus of safety assessment should be the product---the phenotypic traits conferred by the transgene and the biology of the crop. Moreover, the need for concern varies widely depending not only on the product, but also on its ability to hybridize with wild relatives and the environment in which the crop is cultivated. For example, the level of risk is strongly correlated with the ability of the transgenic plant to persist in free-living populations and of the transgene to enhance invasiveness in the crop or a wild relative. Many also contend that current small-scale field tests are insufficient to assess risks associated with widespread cultivation (e.g., James et al. 1998, Snow and Moran Palma 1997). However, large-scale, multiple year field tests may be prohibitively expensive for trees. Thus, it is likely that ecological risks will need to be studied during the initial phases of commercial cultivation.

For transgenic poplar, genetic containment is desirable because commercial clones have undergone little domestication, several characteristics of poplar make extensive, long-distance gene flow likely, and undesirable ecological effects could result if certain transgenes become widely dispersed (reviewed in Strauss et al. 1995, James et al. 1998). Poplars produce abundant pollen and seed; large, mature *P. deltoides* has been estimated to produce greater than 25 million seeds/tree/year (Bessey 1904). Long distance movement of pollen is promoted by wind dispersal combined with tree height (typically 10-50m during reproductive maturity). Seeds, surrounded by cottony hairs, are dispersed long distances by wind and water. In addition, poplar plantations are commonly located near natural or feral stands of interfertile species.

Although it is unclear whether containment will be required for all genes and situations, the possible biological effects of escaped transgenes for resistance traits have been widely discussed. For example, the introduction of genes for insect resistance into wild populations could accelerate the evolution and spread of insects resistant to their effects (Snow and Moran Palma 1997, McGaughey and Whalon 1992, Strauss et al. 1991). The introduction of herbicide-resistance genes could impair control of cultivated

species or interfertile populations and promote greater use of herbicides (Duke et al. 1991, Goldburg 1992). In some forest stands managed for conifers as well as in other sites, such as drainage ditches and perennial crop fields, poplars are considered weeds. Thus, release of herbicide-resistant poplars may be a concern in certain areas.

Because it is difficult to predict when and where transgenic traits may create problems, engineering of reproductive sterility would be ecologically prudent, simplify regulatory approval, and promote public acceptance of transgenic trees. Further, sterility is desirable for non-transgenic, exotic trees in at least some environments. A number of species used in commercial forestry have invaded and severely affected natural and semi-natural ecosystems (Hughes 1994, Richardson 1998). For example, 19 *Pinus* species are invasive in the Southern Hemisphere, and several have established feral populations over large areas, causing major impacts. That engineering sterility will preclude further breeding is not currently a major concern in poplar culture because elite hybrid clones are vegetatively propagated for commercial plantings, and breeding efforts typically focus on selection of parents for crossing to make F_1 hybrids. Finally, even sterile transgenic poplar will require special management, at least in some cases. For example, management strategies to slow the evolution of resistant insects will be necessary when poplars containing a *Bt* transgene are deployed. Measures to prevent the spread of transgenic poplar via vegetative propagation may also be required.

In addition to regulatory and ecological concerns, society's perception of transgenic crops is also an issue. The debate has focused on annual agricultural crops, because applications of biotechnology are much more advanced in crops than they are in trees. However, the public's interest in forests and their preservation portends that transgenic trees will likely move to the forefront. Much of the public unease towards agricultural biotechnology is due to that it is unfamiliar and that the long-term effects of cultivating transgenic plants cannot be determined on an absolute basis (Kappeli and Auberson 1997). At the least, scientists conducting research in this field should be aware of these concerns, and some have suggested that scientists should take a more active role in communicating with the public. Assuredly, the degree to which the public

trusts scientists will affect their attitude towards agricultural biotechnology. Sleigh (1991) proposed that scientists have much to gain by more openly discussing not only the positive and negative aspects of their work, but also social and ethical issues--- avoidance creates the impression that scientists feel no responsibility in such areas and abandons these concerns to their critics.

Certainly, a number of individuals and groups voice very strong views against commercial use of transgenic crops (e.g., Steinbrecher 1996). Attitudes about agricultural biotechnology vary among countries, reflecting differences in culture, history and government response to this issue (Hoban 1997). While a majority of US consumers have consistently supported the use of biotechnology, German and Austrian consumers have clearly opposed genetic engineering. Hobson (1997) suggests that the lack of consumer acceptance of biotechnology in Germany and Austria may be most strongly correlated with the efforts of activist groups opposed to biotechnology. Furthermore, opposition to agricultural biotechnology is not only science-based, but also centers on economic and social issues. For instance, some perceive cultivation of transgenic crops as simply a move by large corporations to make a profit, a hindrance to the development of sustainable agriculture and forestry systems, and unfavorable to less-developed countries. Herbicide resistant crops have been a prime target because they promote use of at least some herbicides, which are the products of large corporations. However, there are also cogent arguments for their production and environmental benefits (Duke et al. 1991, Strauss et al. 1997b).

Thus, issues are complex and indicate that agricultural and forestry biotechnology has considerable potential to change human lives and the environment in positive and negative ways. Improving the clarity, accuracy and content of information communicated to the public to increase understanding and allow the public to make informed choices is a challenging but critical task. As James (1997) cautions, "an uninformed public is likely to form a dichotomy between 'productionists' and those fearful of the new technology." Similarly, determining the appropriate amount of regulation that adequately considers the risks without unduly inhibiting research,

innovation and implementation will continue to be a challenge. Kappeli and Auberson (1997) suggest that "science-based safety assessment be kept distinct from value-based risk analysis in environmental safety evaluations" in order to facilitate constructive discussion. They argue that safety needs to be addressed before risk analysis, because reliable risk analysis of environmental biotechnology applications requires more scientific knowledge and experience than is currently available. According to their definitions, safety is a relative term defined by tolerability levels. Safety assessment involves identifying potential impacts, recognizing sequences of events that lead to harmful outcomes, and appraising the safety of a transgenic organism based on comparisons to similar organisms and processes for which quantitative data or experience are available. In contrast, risk is defined as a quantitative measure of the probability of harm, and cost-benefit analyses are part of risk assessment.

Genetically Engineered Reproductive Sterility

Although several methods for engineering sterility are available, particularly for male sterility, two strategies have been successfully employed with many combinations of genes and species to engineer both male and female sterility (reviewed in Strauss et al. 1995).

Floral Tissue Ablation

This approach relies on the use of a floral-specific promoter to express a cell-autonomous cytotoxic gene product, resulting in ablation of only specific floral cells. In most cases, either a ribonuclease(RNase) or the diphtheria toxin A (*DTA*) served as the cytotoxic gene; they exhibit strong effects and are likely to be useful in any tissue. *DTA* inhibits translation through ADP-ribosylation of elongation factor 2, and is considered

highly safe because without the diphtheria toxin B chain, the A chain cannot be transported across the plasma membrane (Pappenheimer 1977).

Most of the promoters that have been used are active only at the last stages of flower development and are usually gender-specific. A construct containing the tobacco tapetum-specific *TA29* promoter fused to a cDNA encoding barnase, a RNase from *Bacillus amyloliquefaciens*, has been used to engineer male-sterility in a variety of agronomic species (Mariani et al. 1990 and 1992, Reynaerts et al. 1993). By causing the selective destruction of the tapetum, anthers were shriveled and devoid of pollen. Similarly, a stigma-specific promoter fused to barnase was introduced into tobacco, resulting in female sterile plants that had normal vegetative morphology and complete male fertility (Goldman et al. 1994). A *TA29-DTA* construct was used to engineer male-sterile tobacco plants (Koltunow et al. 1990), and an S-locus glycoprotein (*SLG*) promoter from *Brassica* fused to *DTA* was used to engineer female-sterile tobacco plants with reduced male fertility (Thorsness et al. 1991). Expression of the *SLG-DTA* construct ablated the transmitting tract of the style and caused over half the pollen grains to develop abnormally.

In contrast to the above examples, Day et al. (1995) were able to completely prevent the formation of particular floral organs. The promoter of the *Arabidopsis* floral homeotic gene *APETALA3* (*AP3*), which is expressed in petal and stamen primordia, was fused to the *DTA* gene. When introduced into tobacco or *Arabidopsis*, petals and stamens failed to develop. Sepals and carpels developed normally, although aberrant ovule development was observed. Other floral homeotic genes (discussed in following sections) are expressed in both male and female reproductive primordia and/or as the floral meristem begins to form. Therefore, the promoter of one of these genes could conceivably be used to engineer complete sterility, either by ablation of both stamens and carpels or of all floral organs.

In addition to engineering sterility, the ability to prevent formation of floral organs or entire flowers may potentially offer an additional advantage to poplar cultivation---increases in wood production. An encouraging precedent is provided by

agriculture, where the dramatic gains in yield from breeding of agronomic-groups have come primarily from selection for increased allocation to more desirable plant organs, rather than from increased photosynthesis (Evans 1980). Substantial energy and nutrients are committed to reproductive development in trees (Ledig and Linzer 1978). For example, the growth increment of Douglas fir was an average of 16% lower in cone-crop years than in non-crop years (Eis 1965). No comparable studies have been reported for poplar; however, most species reproduce heavily and regularly once past the juvenile stage (Schreiner 1971). If the initiation of reproductive morphogenesis is disrupted, considerable resources may be available for allocation to photosynthetic tissues and wood.

Conversely, using this approach could result in detrimental vegetative effects due to problems with promoter specificity. A floral-specific promoter may direct a basal level of expression in vegetative tissues that is sufficient to kill cells. Even if this is sporadic and does not severely damage the tree, significant cumulative growth effects could occur over several growing seasons. Also of particular concern for trees is that changes in environmental conditions may activate a floral-specific promoter in other tissues. Transgenic aspen containing *TA29-DTA* or *SLG-DTA* constructs are in their third growing season, and exhibit significant growth impairment compared to nontransformed trees (R. Meilan pers. comm.). The problem of promoter specificity might be avoided if barnase is used in combination with its inhibitor, barstar. A weak, constitutive promoter could be used to direct expression of barstar so that nonspecific low-level barnase expression from a "leaky" floral promoter does not result in cell death (Day and Irish 1997).

The barnase-barstar system may also be used to control fertility (i.e., reversible sterility). Though restoring fertility is not critical for poplar culture because commercially deployed F_1 hybrids are usually not used for further breeding (discussed above), it may be important in the future if transgenes are incorporated into parental genotypes. The gene encoding barstar could be placed under the control of a constitutive promoter and transformed into a sterile tree to convert it to fertility.

Alternatively, both the barnase and barstar genes could be introduced at the same time, with the barstar gene under the control of an inducible promoter.

Inhibition of genes essential for fertility

Strategies for inhibiting gene expression act at one of three levels. Either transcription of the gene is blocked, the mRNA is not translated into protein, or the activity of the encoded protein is inhibited. Antisense and cosuppression are proven methods for impairing the expression of genes required for fertility. They depend on transforming a plant with sequences that match at least part of the target gene. However, promoters need not function exclusively in floral tissues as long as the target gene is floral-specific or functionally non-essential in vegetative tissues. Typically, a strong constitutive promoter such as the cauliflower mosaic virus (CaMV) 35S promoter is used to direct expression.

Antisense RNA acts by either reducing mRNA translation or by increasing mRNA degradation. It can cause strong inhibition of gene expression, though some fraction of expression usually remains (reviewed in Mol et al. 1994). For example, when an antisense version of the floral homeotic gene *AGAMOUS* (*AG*) was introduced into *Arabidopsis*, plants with aberrant male- and female-sterile flowers were obtained (Mizukami & Ma 1995). Among the different transformed plants a range of floral phenotypes was observed that was correlated with the level of *AG* mRNA. Progeny of fertile transformants generally had the same floral phenotypes as their parents, suggesting that the phenotypic variation was inherited stably. Introduction of antisense versions of the tomato floral homeotic genes *TM5* and *TAG1* (the tomato homolog of *AG*) into tomato also produced male- and female-sterile plants (Pnueli et al. 1994a, b).

Homology dependent gene silencing is associated with the introduction of duplicate copies of either a native gene or transgene. It can involve post-transcriptional RNA turnover or transcriptional inactivation, which is associated with increased methylation (reviewed in Jorgensen 1995, Matzke & Matzke 1995). Post-

transcriptional silencing is referred to as cosuppression; this was used to inhibit the petunia floral homeotic genes *FPB1* and *FPB2*, resulting in plants in which no *FPB1* or *FPB2* mRNA was detectable (Angenent et al. 1993, 1994). Depending on which gene was suppressed, flowers were either male-sterile or male- and female-sterile. Because of the specificity of function of these homeotic genes, no pleiotropic effects were observed in vegetative organs. Gene silencing may be a result of activation of natural systems for cellular defense against aberrant genes and viruses (Ratcliff et. al., 1997) or represent a genomic response to transposable elements (Henikoff and Matzke 1997); however, the mechanisms are not fully understood. Reversion to a non-suppressed state has been observed in some cases (e.g., Jorgensen 1995), and suppression can be incomplete.

The final strategy employs transgenes with dominant negative mutations (DNMs). A gene with a DNM encodes a mutant protein that is not only non-functional, but also inhibits the activity of the coexisting, wild type protein (Herskowitz, 1987). Though not as extensively studied in plants, many DNMs are potent inhibitors of wild-type function in other eukaryotic organisms (e.g., Ince et al. 1993). The modular structure of regulatory proteins, such as those encoded by floral homeotic genes (described below), makes them amenable to generating DNMs. For example, mutating or deleting transcriptional activation domains, protein interaction domains and DNA binding regions has produced DNMs. Transgenic *Arabidopsis* expressing an apparent DNM *AG* gene phenocopied the *ag* mutant ---plants were both male and female sterile (Mizukami et al. 1996).

The effectiveness of engineering sterility via inhibition strategies is limited by genetic redundancy and because total inhibition of a gene is difficult to achieve. Suppressing two or more genes may alleviate these problems. Secondly, absolute sterility may not be necessary for effective mitigation of ecological risks (e.g., a 99% reduction in fertility may be sufficient). An additional strategy for generating sterility may also become available. Rather than inhibiting a gene essential for fertility, a gene that inhibits reproductive development could be constitutively expressed. Candidate

genes include those that repress floral induction (discussed below). Ultimately, intellectual property issues may also have a role in determining the method and genes used to engineer sterility for commercial application. Public perception may also be a factor ---DTA may not be used, simply because it is derived from a human pathogen and could, therefore, be an easy tool for those opposed to biotechnology.

Flower Development in Model Plant Species

Angiosperm shoot morphogenesis can be divided into separate phases----a juvenile vegetative phase, an adult vegetative phase in which the meristem is competent to respond to a floral stimulus, and a reproductive phase (Poethig 1990, McDaniel et al. 1992). The transition to the reproductive phase is known as floral induction. The vegetative meristem transforms into an inflorescence meristem and floral meristems form from the inflorescence meristem. In a determinate inflorescence, a terminal flower is produced, often after a number of floral meristems have formed on the flanks of the inflorescence. An indeterminate inflorescence meristem retains its identity and continues to produce floral meristems on its periphery until senescence. Furthermore, variations in the number and type of lateral shoots among species result in a diversity of inflorescence morphologies. In the typical angiosperm, the floral meristem consists of four concentric whorls, with each whorl giving rise to a different floral organ. The first (outermost) whorl gives rise to sepals, the second to petals, the third to stamens, and the fourth to carpels---forming a perfect flower. Genetic and molecular analyses of flower development (i.e., floral induction and floral morphogenesis) have focused on *Arabidopsis* and for study of floral morphogenesis, also on *Antirrhinum* (reviewed in Haughn et al. 1995, Weigel 1995, Yanofsky 1995). My discussion is limited primarily to these two species, both of which have indeterminate inflorescences and perfect flowers.

Floral Induction

In most plants, this process is influenced by environmental cues, such as photoperiod, light quality and vernalization, as well as by endogenous signals that are related to the age of the plant. Physiological and genetic studies indicate that alternate pathways to flowering exist in plants, and floral induction is the end result of the interplay between pathways that monitor both the developmental state of the plant and its surrounding environment (reviewed in Bernier et al. 1993, Amasino 1996). *Arabidopsis* is a facultative long day plant, flowering earlier and producing fewer leaves in long days than when grown in short day photoperiods. Genetic analyses of flowering time mutants in *Arabidopsis* have identified more than 20 genes that appear to be involved in the promotion (late-flowering mutants) or repression (early-flowering mutants) of this phase transition (reviewed in Haughn et al. 1995, Weigel 1995, Amasino 1996) . A genetic model (Figure 1.1A) for the control of flowering time proposes that flowering is a default state, negatively regulated by a floral repressor whose activity is under both negative and positive controls (e.g., Haughn et al. 1995).

The most extreme examples of early-flowering mutants are *embryonic flower 1 (emf1)* and *emf2*. These mutants are insensitive to photoperiod, flowering without the formation of vegetative rosette leaves under both long and short day conditions (Yang et al. 1995). In addition, the *emf* mutants are epistatic to early- and late-flowering mutants, suggesting that the *EMF* genes are components of the floral repressor or at least necessary for its activity. Haughn et al. (1995) hypothesized that *EMF* activity progressively decreases during a plant's life, and when its activity falls below a certain threshold, floral induction occurs. Mutations in the *Arabidopsis* gene *TERMINAL FLOWER 1 (TFL1)* result in early flowering under both long and short days, though short day photoperiods delay this transition (Shannon and Meeks-Wagner 1991). Moreover, the normally indeterminate inflorescence is converted to a determinate, floral meristem, leading to the production of a terminal flower. Interestingly, *TFL's Antirrhinum* ortholog *CENTRORADIALIS (CEN)* also functions to maintain the

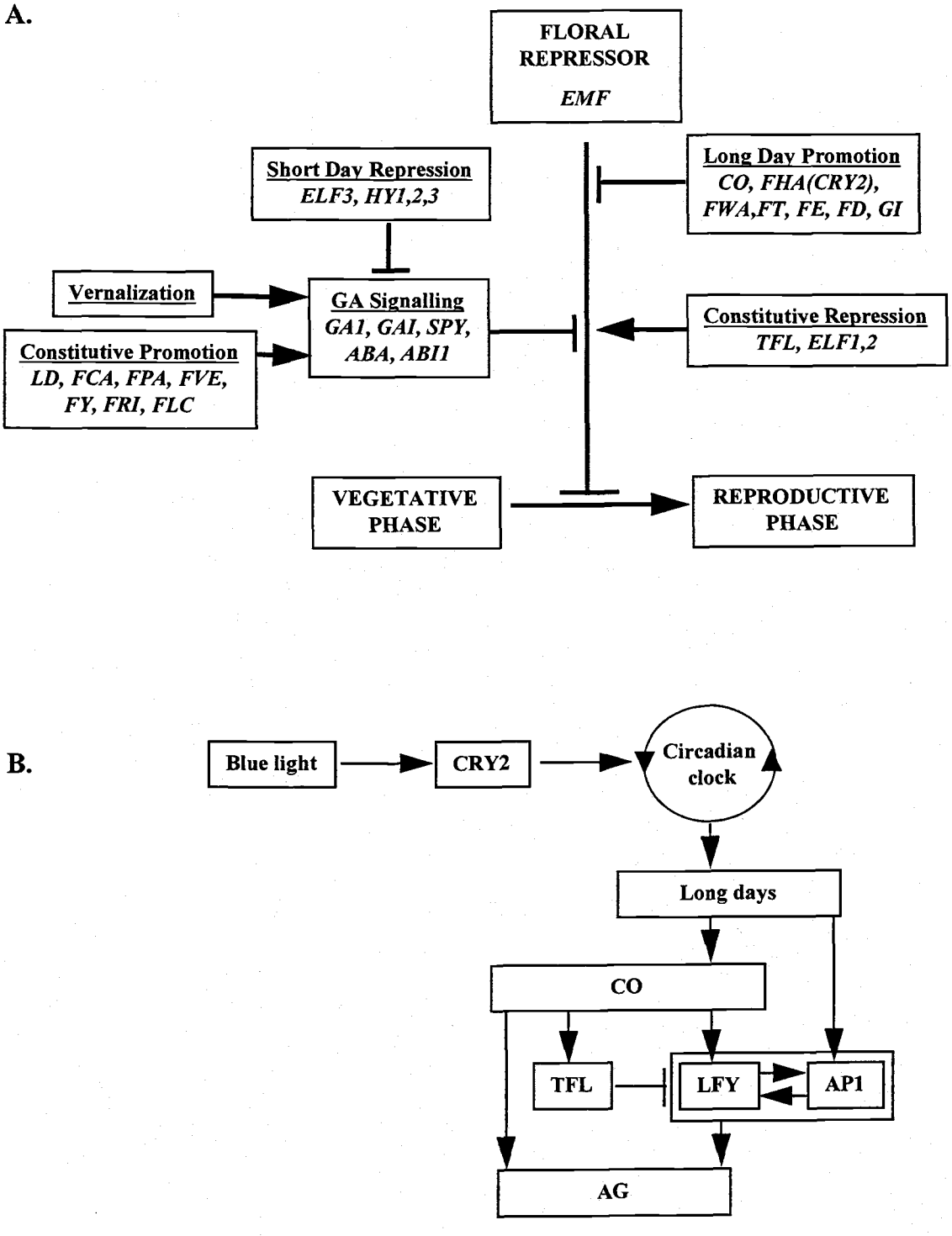


Figure 1.1. *Arabidopsis* floral induction models. **A**, model showing genes involved in flowering pathways (redrawn from Haughn et al. 1995). **B**, possible interactions among genes in the flowering pathway promoted by long days (adapted from Suarez-Lopez and Coupland 1998, Blazquez 1997, Simon et al. 1996). See text for further details.

indeterminate inflorescence meristem, but does not affect flowering time. Both *TFL1* and *CEN* have been cloned and are similar to animal phosphatidylethanolamine-binding proteins (Bradley et al. 1996 and 1997). Although the precise functions of this type of protein are unknown, they are considered to play role in intracellular signaling.

Late-flowering mutants can be divided into at least two distinct groups based on their response to environmental factors. Mutants in the first group are delayed in flowering under both long and short days; thus, these genes appear to function in an autonomous or constitutive pathway. Vernalization is able to at least partially overcome mutations in these genes. Mutants in the second group show delayed flowering only under long days and are insensitive to vernalization. Hence, these genes are involved in promoting flowering via a pathway that is activated by inductive photoperiods.

Two genes that promote flowering via the autonomous pathway have been cloned. *LUMINIDEPENDENS (LD)* encodes a nuclear protein containing a glutamine-rich region similar to domains in other transcription factors; it is expressed throughout the plant and photoperiod does not affect transcript levels (Lee et al. 1994). *FCA* encodes a protein containing two RNA-binding domains and a protein interaction domain, suggesting that it functions in the post-transcriptional regulation of transcripts involved in floral induction (MacKnight et al. 1997). In addition, the *FCA* transcript is alternatively spliced with only one form encoding the entire protein; transcripts were present at similar levels in different tissues, at different developmental stages and under different photoperiods. Interestingly, constitutive expression of the *FCA* gene in transgenic *Arabidopsis* altered not only the abundance but also the ratio of the different transcripts; the level of the shortest transcript increased dramatically, while level of the full-length transcript increased only slightly. Correspondingly, transformants flowered only slightly earlier than controls.

CONSTANS (CO), a gene involved in the photoperiod pathway, has also been cloned and encodes a protein with two zinc finger domains, similar to those found in the GATA-1 family of transcription factors (Putterill et al. 1995). *CO* is expressed in vegetative tissue and is more abundant under long days than short days.

Moreover,increasing *CO* dosage in transgenic plants accelerates flowering, and forced induction of *CO* promotes flowering irrespective of photoperiod (Putterill et al. 1995, Simon et al. 1996). Induction of *CO* activity rapidly initiated transcription of *TFL* and the floral meristem identity gene *LEAFY (LFY),* while expression of the floral meristem identity gene *APETALA1 (AP1)* was induced more slowly. This suggests that *CO* activates *LFY* and *TFL* in response to long days.

Mutations in genes involved in light perception or light-stimulated signal transduction influence flowering time in both positive and negative ways. For example, a *phytochromeB* mutant flowers early in all photoperiods, whereas a *phytochromeA* mutant shows a delay in flowering under certain light regimes (reviewed in Chory 1997). Recently, *FHA,* a gene identified genetically to promote flowering under long days, was shown to encode the blue light receptor cryptochrome 2 (*CRY2*) (Guo et al. 1998). Moreover, *cry2* mutants grown under long days showed about a three-fold reduction in *CO* mRNA compared to wild type plants. Conversely, transgenic plantsoverexpressing *CRY2* exhibited accelerated flowering and increased levels of *CO* mRNA under short days. Therefore, *CRY2* appears to act upstream of *CO* in the photoperiod pathway that promotes floral induction (Figure 1.1B).

The promotion of flowering under long days also requires a mechanism to measure day length; the circadian clock is considered to be the internal oscillator that processes inputs from the environment and whose output regulates the timing of metabolic and developmental events (reviewed in Krebs and Kay 1997). Suarez-Lopez and Coupland (1998) suggest that blue-light activation of *CRY2* sends a signal to the clock, enabling long days to be recognized and results in an output pathway that upregulates *CO* expression. Interestingly, the photoperiod insensitive *early-flowering 3* (*elf3*) mutant exhibits conditional defects in circadian rhythms (Hicks et al. 1996). Further, double mutant analyses with different alleles of the blue light receptor *CRY1* indicated that the early-flowering phenotype of *elf3* mutants may be associated with a disruption in blue-light responsive signal transduction pathways (Zagotta et al. 1996).

In *Arabidopsis*, the timing of floral induction is also affected by the plant hormone gibberellin (GA) (reviewed in Weigel 1995). Exogenous application of GA to plants grown in short days accelerates flowering; the effect of GA in long days is much less. Correspondingly, mutants of genes involved in GA biosynthesis or GA signaling have late-flowering phenotypes. Conversely, mutant alleles of *SPINDLY (SPY),* a negative regulator of GA responses, are early-flowering. An additional gene, identified by interaction with the promoter region of floral meristem identity genes, was subsequently shown to induce early flowering irrespective of photoperiod when constitutively expressed. *SQUAMOSA* promoter binding protein-like 3 (*SPL3*) from *Arabidopsis* belongs to a family of putative transcription factors, is developmentally regulated, and interacts with a motif in the *AP1* promoter (Cardon et al. 1997). However, the ability of constitutive *SPL3* expression to promote early flowering was not dependent on *AP1* activity.

A role for methylation in floral induction as well as in other plant developmental process has been proposed (reviewed in Richards 1997). Studies have used lines of *Arabidopsis* with reduced levels of overall cytosine methylation; these were either selected mutant lines (*ddm1*) or transgenic lines (AMT) expressing an antisense DNA methyltransferase gene. The *ddm1* lines and some AMT lines are late-flowering. In four independent *ddm1* lines, the late-flowering trait was found to be dominant and maps close to a previously identified late-flowering locus, *FWA* (Kakutani 1997). This gene acts in the photoperiod pathway and only dominant mutants have been identified (reviewed in Haughn et al. 1995). Thus, it is possible that the *ddm1* lines have a gain of function mutation due to hypomethylation. Others have suggested that late flowering is a result of a delay in the juvenile to adult phase transition caused by a dampening of the increasing methylation gradient normally seen in progressively older tissues (e.g., Richards 1997). However, some AMT lines are early-flowering (Finnegan et al. 1996). Additionally, recent studies have revealed that, in least in AMT lines, more complex changes in methylation patterns occur rather than simply a reduction in methylation

(Jacobsen and Meyerowitz 1997; discussed in next section). Whether methylation has an important role in normal flower development is still unclear.

Floral Morphogenesis

Floral homeotic genes encode putative transcription factors and constitute a regulatory network controlling floral organogenesis (reviewed in Yanofsky 1995, Haughn et al. 1995, Weigel 1995). Mutations in these genes cause a transformation of identity among normal floral and inflorescence organs. Molecular and genetic studies have identified two major classes of floral homeotic genes, meristem identity and organ identity genes. Most floral homeotic genes exhibit tissue- and organ-specific expression that corresponds to their functional domains. Additional types of genes appear to mediate between the meristem and organ identity genes and to spatially regulate organ identity gene expression. Meristem-structure genes are involved in the control of the number and pattern of cell divisions, and include genes that act in both flower and shoot meristems as well as genes specific for the floral meristem (reviewed in Meyerowitz 1997, Weigel and Clark 1996). Moreover, these categories of genes are not mutually exclusive; several genes fall into more than one class.

Floral meristem identity genes mediate the transition from an inflorescence meristem to a floral meristem, and are positive regulators of the organ identity genes (e.g., Weigel & Meyerowitz 1993, Shannon & Meeks-Wagner 1993). Whether the induction of inflorescence and floral meristems occurs via a linear pathway or by parallel pathways is not clear; both appear to operate in *Antirrhinum* (Bradley et al. 1996, Blazquez et al. 1997). The *Arabidopsis/Antirrhinum* genes *LFY/FLORICAULA* (*FLO*) *AP1/SQUAMOSA* (*SQUA*) initially are expressed throughout the floral meristem, and mutations in these genes cause a transformation of flowers towards inflorescence shoots (reviewed in Yanofsky 1995). *LFY/FLO* encodes a unique protein, with domains that are reminiscent of transcription factors, while *AP1/SQUA* belongs to a large family of transcription factors (described below). *LFY* is also expressed during the vegetative

phase in lateral primordia, however, it is expressed most strongly in floral meristems (Blazquez et al. 1997). *LFY/FLO* is expressed as the floral meristem begins to form, while *AP1/SQUA* xpression commences slightly later (Weigel et al. 1992, Mandel et al. 1992, Carpenter et al. 1995). Considerable functional redundancy exists among the meristem identity genes. Single *AP1* and *LFY* mutants exhibit partial conversions, while in double mutants, virtually all floral primordia develop as inflorescence shoots. Although plants homozygous for mutations in *CAULIFLOWER* (*CAL*) are phenotypically wild-type, *cal* enhances the *ap1* mutant phenotype and the two genes are very similar in sequence (Bowman et al. 1993, Kempin et al. 1995).

Long day photoperiods promote *LFY/FLO* expression and commitment to flowering is associated with the level of *LFY/FLO* expression (Bradley et al. 1996, Blasquez et al. 1997). This is consistent with the finding that long days upregulate *CO* expression, and that *CO* in turn rapidly upregulates *LFY* expression (discussed above). In addition, constitutive expression of *LFY* or *API* results in early-flowering (Weigel and Nilsson 1995, Mandel and Yanofsky 1995). Fewer rosette leaves are produced, lateral shoots are converted to solitary flowers and the normally indeterminate shoot apex produces a terminal flower, similar to *tfl* mutants (discussed above). This correlates with previous genetic analyses and RNA expression studies, which demonstrated that *TFL/CEN* is a negative regulator of *LFY/FLO* and *AP1*. *TFL/CEN* is expressed in the inflorescence apex, and in *tfl* mutants the floral meristem identity genes are ectopically expressed in the inflorescence (Weigel et al. 1992, Shannon and Meeks-Wagner 1993, Gustafson-Brown 1994, Bradley et al. 1996, 1997). Additional genes play more minor roles in specifying floral meristem identity.

Genetic and molecular analyses support a simple combinatorial model for the specification of organ identity (Coen and Meyerowitz 1991). Floral organ identity genes are necessary for three different homeotic functions, designated A, B and C, which specify the four different organ types present in most angiosperms. Each of these activities function in two adjacent whorls: A activity specifies sepals in whorl 1, combined AB activities specify petals in whorl 2, BC activities specify stamens in

whorl 3, and C activity specifies carpels in whorl 4 (Figure 1.1). *AP1/SQUA* and the *Arabidopsis* gene *APETALA2* are A function genes, *AP3/DEFICIENS* (*DEF*) and *PISTILLATA* (*PI*)/*GLOBOSA* (*GLO*) are B genes, and *AG/PLENA* (*PLE*) is the only known C gene (reviewed in Weigel & Meyerowitz 1994, Yanofsky 1995). In a triple mutant lacking all three organ identity functions, all floral organs resemble leaves (Bowman et al. 1991).

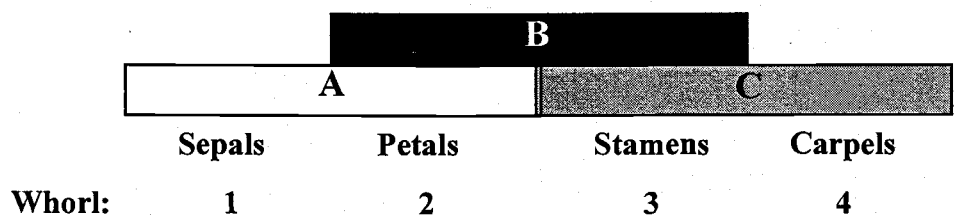


Figure 1.2. The ABC model for floral organ identity gene function.

Except for *AP2*, all belong to the MADS-box family of transcription factors (described below). *AP2* belongs to a large family of putative plant transcription factors that are characterized by a novel DNA binding domain (Okamuro et al. 1997). These genes are expressed before the primordia of the organs they specify emerge from the floral meristem. Expression continues in the developing primordia, becoming localized to specific tissues as the organs differentiate. Analysis of cell lineage patterns during *Antirrhinum* flower development showed that the stage at which floral meristem cells become fated to form particular organs correlates with the onset of organ identity gene expression (Vincent et al. 1995). Organs in two adjacent whorls are transformed if one of these genes is nonfunctional. For example, sepals and carpels develop normally in *ap3* mutants, but petals are replaced by sepals and stamens by carpels (Bowman et al. 1991)

The intermediate gene *UNUSUAL FLORAL ORGANS (UFO)/FIMBRIATA (FIM)* is expressed after the meristem identity genes, but before the organ identity genes (Simon et al. 1994, Ingram et al. 1995, 1997, Wilkinson and Haughn 1995). *UFO/FIM* affects both the identities an arrangement of organs within the flower and encodes a protein with a F-box motif. *FIM* appears to have roles in both the activation of organ identity genes, and in the positioning and maintaining of boundaries between organs. *FIM* associates with a family of F-box proteins (*FAP*s) related to yeast Skp1 proteins, which form complexes that promote protein degradation and cell cycle progression. Ingram et al. (1997) proposed that *FIM-FAP* complexes may affect cell division at organ boundaries by promoting selective degradation of regulatory proteins.

The *Arabidopsis* gene *SUPERMAN* (*SUP*) is another gene that acts to maintain a boundary in the floral meristem (Bowman et al. 1992, Sakai et al. 1995). *SUP* encodes a zinc-finger protein, suggesting a function in transcriptional regulation. In the developing meristem, *SUP* is expressed shortly after the onset of the B and C function organ identity genes in the region of the third whorl that borders the fourth whorl. Loss of function mutants have extra stamens and fewer and smaller carpels, and the expression domain of the B function gene *AP3* expands into part of the fourth whorl. Thus, *SUP* acts to maintain the boundary between the third and fourth whorls, perhaps by preventing the spread of *AP3* function or by repressing third whorl cell division and enhancing fourth whorl cell proliferation. Interestingly, a series of heritable but unstable *sup* alleles (*clk*) were found to have approximately the same pattern of excess cytosine methylation (Jacobsen and Meyerowitz 1997). Curiously, *Arabidopsis* lines (AMT; discussed in previous section) with a general reduction in cytosine methylation produced abnormal flowers, including flowers resembling those of *sup* mutants (e.g., Finnegan et al. 1996). Further, these were correlated with ectopic *AP3* expression; the pattern of *AP3* expression in the floral meristem was not examined, but *AP3* was expressed in vegetative tissues. Jacobsen and Meyerowitz (1997) found that the *SUP* gene was hypermethylated in this AMT line, correlating with the observed *sup* mutant

phenotype, but suggesting that disruption of methylation systems in the AMT lines is more complex (i.e., changes are not limited to hypomethylation).

MADS-box genes

The MADS-box is a highly conserved 57-amino-acid DNA binding domain found in numerous eukaryotic regulatory proteins (reviewed in Shore and Sharrocks 1995). The term MADS-box comes from the four genes in which this domain was first recognized: the minichromosomal maintenance gene (*MCM1*) of yeast, *AG* from *Arabidopsis*, *DEF* from *Antirrhinum* and the mammalian serum response factor (*SRF*). Almost all identified plant MADS-box proteins share the same modular structure (MIKC-type) (Figure 1.2). They contain a second, moderately conserved domain named the K-box because of its similarity to a domain known to form coiled-coils in the intermediate filament protein keratin (Ma et al. 1991). This domain of roughly 70 amino acids is predicted to form two amphipathic helices, which are thought to mediate protein-protein interactions. The MADS-box and K-box are separated by an approximately 35 amino acid Intervening (I)-region which is necessary for dimerization. The C-terminal region is the most variable in sequence and length.

Figure 1.3. Domain structure of plant MADS-box proteins. **M**ADS-box, **I**ntervening region, **K**-box, **C**-terminal region.

Evolutionary History

Twenty nine MADS-box genes have been isolated from *Arabidopsis* (M. Yanofsky pers. comm.), while about 30 MADS-box genes are estimated to be present in *Antirrhinum* and at least 60 in maize (Theissen and Saedler 1995). In contrast, only four MADS-box genes have been identified in yeast, two from *Drosophila* and five from mammals (Theissen et al. 1996). The expression patterns of at least 17 *Arabidopsis* genes have been examined; most are expressed in floral tissues and are often floral-specific. However, genes expressed specifically in root tissue as well as a gene expressed during embryogenesis have been identified (Rounsley et al 1995). In addition to numerous angiosperm species, multiple MADS-box genes have been isolated from gymnosperms (e.g., Tandre et al. 1995) and a fern (Munster et al. 1997).

Several phylogenetic studies have revealed that most plant MADS-box genes are organized into monophyletic clades (Doyle 1994, Purugganan et al. 1995, Theissen et al. 1996, Munster et al. 1997). Further, these subfamilies generally correspond to groups of genes that share similar expression patterns and related functions (i.e., orthologs and relatively recent paralogs). Thus, the four different *Arabidopsis/Antirrhinum* organ identity genes are in different clades, and the major subfamilies are commonly referred to by the name of the organ identity gene. Molecular clock estimates derived from dicot, monocot and conifer gene comparisons suggest that the different lineages began to diverge from one another around the time of the origin of land plants---450 to 500 mya (Purugganan 1997). Though multiple MADS-box genes of the MIKC-type are present in the fern *Ceratopteris*, the isolated genes are not members of the subfamilies known from seed plants (Munster et al. 1997). These results suggest that few MIKC genes were present in the last common ancestor of vascular plants (~ 395 mya), and that many of the gene duplications that led to the high number of present day MIKC-type genes occurred independently in the lineages leading to ferns and seed plants. Isolation of additional MIKC genes from ferns and other basal

vascular plant species should help to clarify the early evolutionary history of these regulatory genes.

Analyses of the angiosperm genes have revealed that the different domains are evolving at different rates (Purugganan et al. 1995). Other regulatory genes also contain both rapid and slow evolving regions, and this has been suggested to play a role in the evolution of novel developmental functions within a gene family (Purugganan and Wessler 1994). The MADS-box exhibits a high level of sequence constraint---approximately 3×10^{-10} nonsynonymous substitutions/site/year. However, several amino acid replacements occur between subfamilies, allowing that changes in this domain could contribute to the diverse functions of the plant genes. The substitution rate for the K-box is more than three times the rate for the MADS-box, while the rate estimate for the C-terminal region is ten times greater, suggesting that changes in these domains play a major role in the functional divergence of this gene family (Purugganan et al. 1995). Rates for the I-region were not calculated; however, protein studies suggest that this domain along with the K-box contribute to the functional specificities of the different genes (described below).

Protein Interactions

Both homodimers and heterodimers are formed by the MADS-box proteins. The crystal structure of the SRF core bound to DNA has been deduced (Pellegrini et al. 1995). The N-terminal region of the MADS-box forms an α helix that contacts DNA, while an adjacent hydrophobic region forms a β-sheet that mediates dimerization along with a short α helix located in the region following the MADS-box. Studies indicate that the plant MADS-box proteins exhibit partner-specificity for the formation of DNA binding dimers, suggesting that selective dimerization contributes to the functional specificity of these proteins (Riechmann et al. 1996a). While products of the A class gene *AP1* and the C class gene *AG* were able to interact with each other as well as with both B function proteins *in vitro*, only AP1-AP1 and AG-AG homodimers were capable

of binding DNA (Riechmann et al. 1996a). This indicates that the different classes of organ identity genes do not act in a combinatorial manner by forming DNA-binding heterodimers.

In both *Arabidopsis* and *Antirrhinum*, the B class organ identity proteins appear to only form a functional unit as a heterodimer with each other (i.e., AP3-PI and DEF-GLO). Though expression is initiated independently, the heterodimer is apparently necessary to maintain adequate expression levels of both AP3/DEF and PI/GLO. Research suggests that this autoregulatory mechanism operates at the level of transcription; the heterodimer binds to sites in the promoters of both B genes (Trobner et al. 1992, Schwarz-Sommer et al. 1992, Jack et al. 1994,Goto and Meyerowitz 1994). Furthermore, AP3 and PI proteins are only able to enter the nucleus when both are present, suggesting that a functional nuclear localization signal requires the formation of a heterodimer (McGonigle et al. 1995).

The MADS-box and I-region form the core or minimal DNA-binding region of AP1, AGAMOUS-like 2 (AGL2) and AG, whereas these domains plus the first several amino acids of the K-box are necessary for AP3/DEF and PI/GLO to form a DNA-binding dimer (Zachgo et al. 1995, Huang et al. 1996, Davies et al. 1996, Riechmann et al. 1996b). Plant MADS-box protein dimers bind similar sequences that are variants of the CArG-box ($CC(A/T)_6GG$) bound by SRF (reviewed in Shore and Sharrocks 1995). *In vitro* studies of the DNA binding characteristics of AP1-AP1, AP3-PI and AG-AG dimers showed that all recognized the same sites, but with somewhat different affinities (Riechmann et al. 1996b). The similarity of the binding sites suggests that domains other than the MADS-box have a major role in determining the functional specificity of these proteins.

Krizek and Meyerowitz (1996) generated chimeric genes by exchanging domains among the four *Arabidopsis* MADS-box organ identity genes. Because ectopic expression of each of these four genes produces distinctive gain of function phenotypes, the phenotypes caused by ectopic expression of the chimeric genes indicated what regions are important for specificity of function. The regions most responsible for the

functional specificities of AP3 and PI were the I-region and the K-box, while the MADS-box and I-region appeared most important for the functional specificities of AP1 and AG. However, the K-box may also have a role in determining the functional specificities of AG and AP1. Constitutive expression of a chimera containing an *AG* MADS-box and I-region and an *AP1* K-box and C-terminal region produced an *AG* gain of function phenotype. When the K-box and C-terminal region were from *AP3* rather than from *AP1*, ectopic expression produced a wild-type phenotype. Furthermore, studies indicate that the K-box mediates specific interactions between AG and other MADS-box proteins (Fan et al. 1997) (discussed in next section). Additional studies revealed that these proteins still retain their ability to specify organ identity after their DNA-binding specificity had been altered by replacing the N-terminal half of the MADS-box with the corresponding regions of SRF or MEF2A (Riechmann and Meyerowitz 1997).

The plant MADS-box protein studies indicate that dimer formation and DNA-binding specificities are insufficient to fully explain their specific functions. Although the results from *in vitro* studies may not completely or accurately reflect *in vivo* molecular mechanisms, it seems likely that interactions with accessory proteins play a major role in determining the specific functions of the different MADS-box genes. In addition, modularity of gene regulation in which multiple independent *cis* regulatory regions are required to control the expression of a gene during a developmental process is likely. In this case, a particular MADS-box protein dimer would be only one of the transcription factors required, and direct protein interactions between transcription factors bound to different sites may occur. Such mechanisms have already been shown for mammalian and yeast MADS-box proteins (reviewed in Firulli and Olson 1997, Johnson 1995). For example, the MCM1 homodimer interacts with different combinations of proteins to activate or repress transcription; the specific protein combinations are dependent on yeast cell type as well as the *cis* elements present at a particular locus. Homeodomain transcription factors are among the proteins MCM1 interacts with, while mammalian MADS-box proteins, the myocyte enhancer factor 2

(MEF2) family, interact with heterodimers of basic helix-loop-helix proteins to synergistically activate transcription and myogenesis.

The AGAMOUS Subfamily

Function, Expression and Evolution

Sepals and petals develop normally in *ag* mutants, but petals develop in place of stamens in the third whorl and carpels do not develop (Yanofsky et al. 1990). The sepal-petal-petal pattern is repeated a variable number of times. Similarly, the inner two whorls of reproductive organs are replaced by numerous petal-like and/or sepal-like organs in *ple* mutant flowers (Bradley et al. 1993). Thus, in addition to specifying stamen and carpel identity, *AG/PLE* is necessary to specify determinacy of the floral meristem. When sepal primordia are just visible on the flanks of the floral meristem, *AG* is expressed in the center of the meristem where the stamen and carpel primordia will form (i.e., whorls 3 and 4). *AG* is uniformly expressed throughout the developing stamen and carpel primordia, becoming localized to specific tissues as these organs develop (Bowman et al. 1991). After stamen differentiate, *AG* RNA is present in the connective of the anther and at lower levels in the anther walls and filament. In carpels, *AG* is expressed most strongly in the stigmatic papillae and the developing ovules. *AG* RNA is present in the integuments, becoming restricted to the endothelium (cell layer surrounding the embryo sac) in mature ovules.

Gain of function mutants resulting from introduction of a *35S-AG* construct into *Arabidopsis* or from a transposon insertion into an intron of *PLE* demonstrated that *AG/PLE* is sufficient to specify reproductive organs (Mizukami and Ma 1992, Bradley et al. 1993). In these mutants, flowers consist of four whorls of reproductive organs---carpel-stamen-stamen-carpel. Additional studies have shown that *AG's* floral determinacy function can be separated from its organ identity functions. Transgenic

Arabidopsis plants expressing an antisense *AG* construct produced a range of abnormal floral phenotypes (Mizukami and Ma 1995). The most severe were phenocopies of an *ag* loss of function mutant, while the least severe produced normal stamen and carpels, but the floral meristem was indeterminate. The severity of the phenotype generally correlated with the level of *AG* expression, suggesting that specification of floral determinacy requires higher levels of *AG* activity. Similarly, two *ag* alleles were identified that retained partial *AG* activity; further, these showed that third and fourth whorl functions are separable (Sieburth et al. 1995). Flowers of *ag-4* mutants contain stamens in the third whorl, but sepals replace carpels in the fourth whorl and the floral meristem continues to proliferate. *AG-Met205* plants produced indeterminate flowers with stamens and carpels. Both these mutations affect the K-box, suggesting that they may disrupt interactions with co-factors; they are discussed further in the next section.

These studies were extended by Jack et al. (1997) who demonstrated that *AG* is required in the fourth whorl to make the flower determinate by producing plants in which *AG* was expressed only in the second and third whorls. Due to the misexpression of *AG* in the second whorl, the flowers consist of a repetition of a sepal-stamen-stamen pattern. In addition to specifying floral meristem determinacy, studies indicate that *AG* has an important role in maintaining floral meristem identity. When grown in short day photoperiod, homozygous *ag* mutants undergo floral reversion----inflorescence shoots arise from the center of flowers (Okamuro et al. 1996). Cells that initiate floral reversion are temporally and spatially separate from those that initiate flower development, suggesting that establishment and maintenance of floral meristem identity are distinct processes. GA signal transduction also appears to be an important component of floral meristem identity maintenance. Further, when *ag* plants grown under continuous light were shifted to complete darkness for 24-48 hours, the indeterminate center of *ag* flowers produced lateral secondary flowers which also had characteristics of inflorescence shoots (Mizukami and Ma 1997).

While these studies with *ag* mutants indicated that *AG* activity is required to fully define floral meristem identity throughout reproductive growth, *35S-AG* transgenics revealed

that *AG* is sufficient to promote floral fate. Ectopic *AG* expression caused early-flowering under both long and short day photoperiods, apparently by accelerating the change from a juvenile vegetative phase to a reproductively competent adult vegetative phase and by shortening the adult phase (Mizukami and Ma 1997). In addition, terminal flowers formed at the inflorescence apex as well as at lateral shoot apices. The *35S-AG* transgene also partially suppressed the indeterminate shoot defects of *lfy* and *ap1* single and double mutants. Because *AG* is neither expressed in vegetative tissues nor in the initial stages of the floral meristem and *ag* mutants flower at the same time as wild-type plants, *AG* is not likely to have a role in floral induction. Mizukami and Ma (1997) proposed that *AG* may produce a floral promotion activity within floral meristems to maintain flower development. According to this model, commitment to floral fate is acquired gradually, owing in part to the sequential action of *LFY*, then *AP1*, and finally *AG* (Ma 1998).

Several genes isolated from various dicots and monocots exhibit expression patterns similar to *AG* and phylogenetic analysis places them in the *AG* clade (Table 1.2). Other genes belonging to the *AG* subfamily are only expressed in carpels or ovules and expression begins at a later stage of floral development. For a number of genes displaying an *AG*-like expression pattern (i.e., in both stamens and carpels), loss of function and/or gain of function mutants have been produced via antisense RNA, cosuppression and constitutive transgene expression either in homologous or heterologous plants. These include genes from tomato (Pnueli et al. 1994), tobacco (Kempin et al. 1993), brassica (Mandel et al. 1992), and rice (Kang et al. 1995). Though organ transformations were sometimes partial, the phenotypes were generally consistent with a loss or gain of C-function.

Some species contain two genes belonging to the *AG* subfamily that are also expressed in an *AG*-like pattern. Ectopic expression of the petunia homolog *PMADS3* resulted in flowers having sepals with carpelloid features and petals partially converted to antheroid structures (Tsuchimoto et al. 1993). In contrast, constitutive expression of a second petunia *AG* homolog, *Floral Binding Protein 6 (FBP6)* only affected petals;

Table 1.2. Genes belonging to the *AG* subfamily

		RNA Expression[2]			
Gene[1]	**Species**	**Stamen**	**Carpel**	**Ovule**	**Reference**
AG	*Arabidopsis*	yes	yes	yes	Bowman et al. 1991
AGL1	*Arabidopsis*	no	yes-later	yes	Flanagan & Ma 1996
AGL5	*Arabidopsis*	no	yes-later	yes	Savidge et al. 1995
AGL11	*Arabidopsis*	no	no	yes	Rounsley et al. 1995
PLE[3]	*Antirrhinum*	yes	yes	yes	Bradley et al. 1993
BAG1	*Brassica*	yes	yes	nd[4]	Mandel et al. 1992
NAG1	*Nicotiana*	yes	yes	nd	Kempin et al. 1993
TAG1	*Lycopersicon*	yes	yes	yes	Pnueli et al. 1994
GAG2	*Panax*	yes	yes	nd	Kim et al. 1995
PMADS3	*Petunia*	yes	yes	yes	Kater et al. 1998
FBP6	*Petunia*	yes	yes	yes	Kater et al. 1998
FBP7	*Petunia*	no	no	yes	Angenent et al. 1995
FBP11	*Petunia*	no	no	yes	Angenent et al. 1995
PAGL1	*Petunia*	nd	nd	nd	GenBank submission
CUM1	*Cucumis*	yes	yes	nd	Kater et al. 1998
CUM10	*Cucumis*	yes	yes	nd	Kater et al. 1998
CUS1[5]	*Cucumis*	no	yes	yes	Filipecki et al. 1997
RAP1	*Rumex*	yes	yes	nd	Ainsworth et al. 1995
SLM1	*Silene*	yes	yes	nd	Hardenack et al. 1994
OSMADS3	*Oryza*	yes	yes	nd	Kang et al. 1995
ZAG1	*Zea*	yes-weaker	yes-stronger	yes	Schmidt et al. 1993
ZAG2	*Zea*	no	yes-later	yes	Schmidt et al. 1993
ZMM1	*Zea*	nd	nd	nd	Theissen et al. 1995
ZMM2	*Zea*	yes-stronger	yes-weaker	nd	Mena et al. 1996
DAL2[6]	*Picea*	yes	yes	yes	Tandre 1997

[1]Genes expressed in female, but not in male reproductive tissues are shaded.
[2]Expression is considered ovule-specific if it is limited to ovules and the tissues from which ovules arise (i.e., placenta).
[3]An additional gene from *Antirrhinum*, *DEFH1*, has been reported to be very similar to *PLE*, but its sequence and expression pattern have not been published (Davies et al. 1996)
[4]Not determined.
[5]*CUS1* is also expressed in embryogenic callus and fruit.
[6]Tissues are male cones, female cones and the ovuliferous scales rather than stamens, carpels and ovules.

they were reduced in size and altered in shape, but no antheroid tissues were produced (Kateret al. 1998). Both genes are ectopically expressed in the *blind* mutant, which has a floral phenotype very similar to *35S-PMADS3* transgenic plants, suggesting that the two genes do not act additively. Producing single and double loss of function mutants via gene suppression strategies may elucidate a role for *FBP6* in providing C or other functions. Two *AG* homologs from cucumber, *CUM1* and *CUM10*, are also expressed in the whorls giving rise to stamens and carpels (Kater et al. 1998). Introduction of a*35S-CUM1*transgene into petunia caused a complete transformation of petals into stamens, and stigma- and style-like structures developed at the tips of abnormal sepals. Constitutive expression of *CUM10* produced petals with some antheroid tissue; however, no carpelloid features were evident on sepals, which were larger than wild-type.

Studies have also revealed a diversification of C-function in maize, which contains two genes, *ZAG1* and *ZMM2*, that appear to have distinct but partially redundant functions (Mena et al. 1996). Both genes are expressed in carpels and stamens; however, *ZAG1* RNA accumulates more in carpels than in stamen, while *ZMM2* RNA accumulates more in stamens than in carpels. A loss of function *zag1* mutant caused by transposon insertion exhibited loss of determinacy, but normal pollen-producing stamens developed and the innermost set of organs had carpelloid features. This suggests that an additional gene functions to specify stamens and perhaps, functions together with *ZAG1* to specify carpels; that this gene is *ZMM2* is suggested by its expression pattern and by the extensive sequence homology it shares with *ZAG1*.

The female-specific petunia genes *FBP7* and *FBP11* are 90% identical in amino acid sequence (Angenent et al. 1995). Both are expressed in the center of the gynoecium before ovule primordia are visible and expression later becomes restricted to the ovule primordia and finally to specific regions of the ovule. Cosuppression of both *FBP7* and *FBP11* resulted in the transformation of ovules into spaghetti-like masses of carpelloid structures. Conversely, constitutive expression of *FBP11* caused ovule-like

structures to develop on sepals and to a lesser extent on petals, suggesting that it is an ovule identity gene (Colombo et al. 1995). In both sequence and expression pattern, the *Arabidopsis* gene *AGL11* (Rounsley et al 1995) is very similar to *FBP7* and *FBP11*, but functional information is not available for *AGL11* nor for the other female-specific genes (listed in Table 1.2).

Finally, the most basal member of the *AG* subfamily is *DEF-AG-like 2* (*DAL2*) from *Picea abies* (Norway spruce)(Tandre et al. 1995). *DAL2* is expressed in both male and female cones but its transcript was not detected in vegetative tissues. In the female cone, *DAL2* is expressed in the developing ovuliferous scale, but not in the subtending bracts, the primary cone axis or cone apical meristem (Tandre 1997). Furthermore, constitutive expression of *DAL2* in *Arabidopsis* resulted in partial transformations of sepals and petals; ovule- and stigma-like structures formed on the sepal margins and filamentous, stamen-like structures replaced petals. Taken together the *DAL2* analyses support that it is involved in specifying reproductive organs and that features of an ancestral gene were conserved in both the conifer and angiosperm lineages. Tandre (1997) suggests that duplications within the angiosperm lineage gave rise to additional members of the *AG* subfamily late in angiosperm evolution, and that divergence in sequence and function of paralogs is associated with novel features important to the development of the angiosperm flower. Although the characteristics of *DAL2* might suggest that *AG* and angiosperm orthologs are more similar to the ancestral gene, it is not certain that *DAL2* is the only *Picea* gene belonging to the *AG* clade, and phylogenetic analyses have not clearly revealed the relationships among the angiosperm members of this subfamily (e.g., Purugganan 1997).

Regulatory Interactions

The research described above demonstrates that *AG* has multiple roles in flower development---it specifies reproductive organ identity, it maintains and finally defines

the floral meristem, and it functions as a meristem structure gene by preventing continued cell division in the center of the meristem to make the flower determinate. This complexity of function suggests that regulation of *AG* expression and activity and the molecular mechanisms by which *AG* achieves it different functions are also complex. The meristem identity genes *LFY* and *AP1* are partially redundant positive regulators of *AG* (Weigel and Meyerowitz 1993). However, even in *lfy ap1* double mutants, *AG* is expressed, albeit in an abnormal pattern. In addition, the effects of *lfy ap1* mutations are attenuated acropetally such that late arising lateral structures are less shoot-like and have some carpelloid features. The severity of other floral mutants, such as *ap2*, are also reduced in flowers that arise late in the life cycle. Weigel (1995, 1997) proposed that expression of downstream floral genes depends not only on activation by genes such as *LFY* and *AP1*, but also on progressive derepression during the plant's life cycle. Furthermore, floral induction signals may act directly on the late acting floral genes rather than solely via a linear hierarchy in which floral induction acts on the early expressed meristem identity genes, which then activate the organ identity and other floral genes.

At least a few of the flowering time genes appear to affect *AG* expression, though the interaction may be indirect. In the severe *emf1-2* mutant, *AG* is expressed precociously and ectopically (Chen et al. 1997). Upon germination, this mutant produces sessile cotyledons, which develop carpelloid features at their base, and then terminates in an abnormal flower consisting only of carpel-like organs. *AG* RNA accumulates at the base of the cotyledons, where ovule-like structures or stigmatic tissue will form, as well as in the developing carpelloid structures. Ectopic *AG* expression was not detected in weak *emf* mutants. Transgenic *Arabidopsis* plants in which high levels of *CO* activity were rapidly induced often terminated shoot development prematurely with the formation of a carpelloid structure (Simon et al. 1996). As described previously, forced induction of *CO* activity rapidly upregulated *LFY* and *TFL* expression; however, their expression patterns were similar to wild-type--*TFL* was expressed in the inflorescence apex, but *LFY* was not. This suggests that *CO*

ectopically activated genes required for carpel development (e.g., *AG*) in the shoot meristem via a *LFY*-independent pathway (Figure 1.1B). Furthermore, introduction of the *35S-AG* transgene suppressed the late-flowering phenotype of both a *co* mutant as well as a *fca* mutant, indicating that *AG* is downstream of two different floral promotion pathways (Mizukami and Ma 1997).

One model for floral induction proposes that *EMF* is the major constituent of a floral repressor activity that at high levels suppresses reproductive growth; floral promoters, such as *CO* and *FCA*, act to decrease this activity (e.g., Haughn et al. 1995). In addition, double mutant analyses reveled that *emf1-2* is epistatic to floral meristem identity mutants (e.g., *lfy, ap1*), indicating that *EMF* acts downstream of these genes in mediating the inflorescence to floral transition. Chen and co-workers (1997) suggested that a reciprocal negative interaction between *EMF* and the floral genes controls this transition, because high *EMF* activity appears to suppress *AP1*---precocious *AP1* expression was detected in both weak and strong *emf* mutants.

They proposed that strong *EMF* activity during vegetative growth suppresses floral genes. As flowering time genes mediate a decline in *EMF* activity, floral genes are activated; the increasing activities of floral meristem identity genes promote the further decline in *EMF* activity, allowing the late floral program to be initiated. This further decrease in *EMF* activity may correlate with derepression of *AG*, because precocious *AG* expression was only detected in the severe *emf1-2* mutant. If this were the case, *AG* derepression may coincide temporally and spatially with the activation of *AG* by *LFY* and *AP1*. This model does not preclude that *CO* or other flowering time genes activate *LFY*, *AP1* and other floral genes directly or by separate pathways rather than only via a decrease in floral repressor activity. Perhaps, in the absence of *LFY* and *AP1* activities, floral induction signals continue to decrease *EMF* activity and/or act by independent pathways to eventually derepress *AG* in lateral shoots, resulting in carpelloid features.

While *EMF* and other flowering time genes affect many genes, *CURLY LEAF* (*CLF*) appears mostly to be a regulator of *AG* (Goodrich et al. 1997). The *clf*

phenotype, leaf curling and partial transformations of sepals to carpels and petals to stamens, is very similar to the phenotype of *35S-AG* transgenics. Consequently, Goodrich et al (1997) not only verified that *AG* is ectopically expressed in *clf* mutants, but also showed that ectopic expression of *AG* is responsible for the *clf* phenotype. A severe *ag* mutation is almost completely epistatic to *clf*, and the *clf* phenotype is sensitive to *AG* dosage. In *clf* mutants, *AG* is ectopically expressed in both emerging and adult leaves during the vegetative phase, but not in the shoot apical meristem. During the early stages of flower development, the pattern of *AG* expression appears normal, but at later stages, ectopic expression was detected in petals and the inflorescence stem. These observations suggest that *CLF* is required to maintain the proper *AG* expression pattern, but not for the initial specification of the *AG* expression pattern.

Interestingly, *CLF* encodes a protein homologous to the product of the *Drosophila* gene *Enhancer of zeste* (*E(z)*), a member of the *Polycomb group* (*PcG*) (Goodrich et al. 1997). The *PcG* genes act to maintain transcriptional repression of homeotic genes and act in a somewhat antagonistic manner with the *trithorax group* (*trxG*) genes, which maintain active transcriptional states (reviewed in Pirrotta 1997, Gould 1997). The *PcG* genes encode a diverse group of proteins; however, some share motifs that are also found in heterochromatin proteins. For example, E(z), CLF and mammalian homologs contain a SET domain that is also present in two trxG proteins. Several lines of evidence indicate that these motifs mediate protein-protein interactions between members of the PcG group, leading to the formation of large, multiprotein complexes. These complexes can silence gene expression over large distances, and appear to act by modifying chromatin structure. Further, the silenced state is stable through many rounds of cell division, and PcG protein concentration affects the stability of this inactive state. Gene silencing may involve recognition of a few high-affinity *Polycomb* DNA response elements followed by recruitment of a larger number of dispersed lower-affinity sites by cooperative interactions between PcG proteins.

That *CLF* and *E(z)* both apparently maintain an inactive transcriptional state is particularly interesting because their target homeotic genes are unrelated; the *Drosophila* and mammalian homeotic genes encode homeobox proteins rather than MADS-box proteins. Further, the PcG proteins control a number of genes, and mutation in a PcG gene results in ectopic homeotic gene expression at the developmental stage when homeotic genes are normally expressed, though in a more restricted pattern. Neither seems to be the case for *CLF*; however, its role in suppressing other genes may be masked by redundant factors. It remains to be determined whether additional PcG gene homologs exist in plants and whether mechanisms of transcriptional repression are similar between plants and *Drosophila*. The observation that some transgenic AMT lines (discussed earlier) have phenotypes similar to *clf* mutants and also exhibit ectopic *AG* expression led to the suggestion that repression by *CLF* might involve methylation of the *AG* gene (Finnegan et al. 1996, Goodrich et al. 1997). Also of interest is whether *CLF* has a role in the derepression of floral genes postulated to occur over the life of a plant and whether the flowering time and meristem identity genes affect *CLF* activity.

In addition, *CLF* may interact with floral-specific repressors of *AG*. The presence of *CLF* RNA in all four floral whorls throughout development suggests that it is insufficient to repress *AG* within the flower or that post-transcriptional controls confer whorl-specific activity to *CLF*. The A function organ identity gene *AP2* is also a negative spatial regulator of *AG* RNA accumulation in whorls 1 and 2, and like *CLF*, is expressed in all floral whorls.(Drews et al. 1991, Jofku et al. 1994). In a genetic screen designed to isolate mutations that enhanced a weak *ap2* allele, an additional gene, *LEUNIG (LUG)*, was identified that also acts to prevent *AG* expression in the outer two whorls (Liu and Meyerowitz 1995). In contrast to *clf* mutants, ectopic *AG* floral expression occurs very early in flower development in *ap2* and *lug* mutants and the resulting floral phenotypes are more pronounced. While *clf* enhanced the mutant floral phenotype of a weak *ap2* allele, strong *ap2* alleles were epistatic to *clf*, suggesting that

CLF and *AP2* act in the same pathway (Goodrich et al. 1997). *LUG* and *CLF* may act independently, because *lug clf* double mutants displayed an additive phenotype.

Interestingly, *cis* regulatory elements necessary for the negative regulation of *AG* and for the activation of *AG* in the floral meristem are located in the large (~3 kb) second intron (Sieburth and Meyerowitz 1997). The expression patterns conferred by two overlapping regions of the *AG* gene were examined by introducing *AG*-β glucuronidase (*GUS*) reporter gene fusions into *Arabidopsis*. The construct lacking the *AG* intron directed a *GUS* staining pattern that differed markedly from normal *AG* expression. *GUS* staining was evident in leaves, stem and the shoot apical meristem. During early flower development, *GUS* staining was weak and not restricted to the two inner whorls, indicating that activation by *LFY* and *AP1* requires intron sequences. Further, *GUS* staining did not appear in carpels until late in development, and strong *GUS* staining in stamens also occurred at a later stage. In contrast, the construct containing intragenic sequences conferred a *GUS* staining pattern that was nearly indistinguishable from the pattern of *AG* RNA expression. The constructs were also analyzed in *ap2*, *lug* and *clf* mutants; in all cases, the expression pattern conferred by the construct containing the intragenic sequences showed the expected changes, while significant changes were not observed for the intronless construct.

Comparisons between the *AG-GUS* constructs late in carpel development indicated that regulatory interactions are complex, and that upstream sequences as well as intragenic sequences are required. Although a number of genes involved in ovule development have been identified, only one has been suggested to have a role in regulating *AG*. Mutations in *BEL1* mainly affect ovule integument morphogenesis and identity (Modrusan et al. 1994, Ray et al. 1994). Transformation of integuments into carpel-like structures was correlated with ectopic *AG* expression late in ovule development, suggesting that *BEL1* negatively regulates *AG* expression in the ovules. *BEL1* encodes a homeodomain protein and the expression patterns of *BEL1* and *AG* overlap in the ovule (Reiser et al. 1995). Therefore, additional factors may be necessary to regulate *AG* expression in the ovules.

Less is known about the genes *AG* regulates. *AP1*, which is involved in the activation of *AG*, is initially expressed throughout the floral meristem. Subsequently, *AP1* expression becomes restricted to the outer two whorls; accumulation of *AG* RNA in the inner two whorls correlates with the abatement of *AP1* expression in these whorls. In *ag* mutants, *AP1* RNA is present in the inner two whorls throughout floral morphogenesis, indicating that *AG* is a negative regulator of *AP1* expression (Gustafson-Brown 1994). Conversely, *AG* positively regulates *AGL5*, a member of the *AG* subfamily (Table 1.2). *AGL5* RNA is only detected in carpels and it begins to accumulate shortly after the onset of *AG* expression (Savidge et al. 1995). *AGL5 RNA* was not detected in flowers of *ag* or *ag ap1* mutants and ectopic *AG* expression activated *AGL5* in cauline leaves. Further, this regulation may be direct; *AG* specifically binds to a consensus binding site located in the *AGL5* promoter.

The molecular mechanisms of *AG* action are also beginning to be elucidated. As previously discussed, the AG MADS-box and I-region are sufficient for dimerization and DNA-binding *in vitro,* and these domains are also involved in determining the functional specificity of AG (Riechmann et al. 1996, Mizukami et al. 1996, Krizek and Meyerowitz 1996). Analyses of constructs encoding truncated AG proteins in *Arabidopsis* revealed that the N-terminal region is not required to produce an *AG* gain of function phenotype, whereas the K-box and C-terminal domain are necessary for *AG* function (Mizukami et al. 1996). Furthermore, transformants with an *AG* transgene lacking the C-terminal region exhibited an *ag* mutant phenotype, suggesting that this deletion generated a dominant negative mutation. When both the K-box and the C-terminal region were deleted, flowers exhibited a slight increase in the number of stamens and carpels. Because *AG* antisense plants with only a slightly reduced level of *AG* sense RNA had a similar phenotype, this truncated *AG* protein may minimally inhibit *AG* function (Mizukami et al. 1996, Mizukami and Ma 1995).

One possible explanation for these observations is that an AG protein lacking the C-terminal region competes with wild-type protein for binding to accessory proteins, and that this interaction is mediated by the K-box (Mizukami et al. 1996).

Subsequently, Fan et al. (1997) used a fusion protein containing the K-box and C-terminal region of AG as bait in a yeast two-hybrid system. Four MADS-box proteins belonging to the AP1 subfamily were found to interact: AGL2, AGL4, AGL6 and AGL9. Similarly, an earlier study using the full-length PLE protein as bait identified four interacting *Antirrhinum* proteins (DEFH200, DEFH72, DEFH49 and SQUA) that also belong to the *AP1* subfamily (Davies et al. 1996). The RNA expression domains of *AG/PLE* and these *AP1* subfamily genes overlap, suggesting that the encoded proteins may form complexes *in vivo* to regulate gene expression.

Characterization of two *ag* alleles with partial activity also support that the K-box mediates interactions with co-factors required for *AG* functions, specifically determinacy and carpel identity (Sieburth et al. 1995). The *ag-4* mutation results in a partial loss of the second amphipathic helix in the proposed K-box coiled-coil this region, whereas the *AG-Met205* mutation causes a single amino acid change near the C-terminus of the K-box. Flowers of both mutants are indeterminate, but unlike strong *ag* mutants, stamens are present in the third whorl. While *AG-Met205* flowers contain carpels, sepals replace carpels in *ag-4* mutants. Additional analyses showed that the K-box is sufficient for the binding of AG to the four AGL proteins, but interactions are stronger when the C-terminal region is present in one or both of the partner proteins (Fan et al. 1997). The AG K-box did not bind to itself nor to the AGL1 K-box, supporting that the K-box interactions are specific. In contrast, the AG MADS-box and I-region can form DNA-binding homodimers or heterodimers with AGL1 (Huang et al. 1996, Riechmann et al. 1996). Interestingly, AG did not enter the nucleus when expressed in an onion cell transient assay, suggesting that other factors interact with AG to facilitate entry (McGonigle et al. 1996).

Taken together, these studies suggest that *AG* acts via molecular mechanisms similar to those demonstrated for yeast and mammalian MADS-box genes (discussed earlier). For example, AG homodimers and heterodimers may interact with different combinations of accessory proteins via the K-box to regulate transcription. At least some of these co-factors appear to be other MADS-box proteins. Different

combinations of factors may control different functions. *AG* expression overlaps with expression of the four putative *AGL* co-factors at different floral developmental stages and in different areas within the flower. Further, all *AG* functions are not disrupted by particular mutations in the K-box, and the *cis* regulatory elements present at a particular locus may require distinct combinations of proteins. Finally, studies also indicate that the level of *AG* RNA is important and that different *AG* functions require different amounts of gene product (Mizukami and Ma 1992, 1995, Jack et al. 1997).

Poplar Floral Development

The vegetative phase of poplars lasts several years, with flowering first occurring at 5 to 10 years of age in sections *Tacamahaca* and *Aigeiros* (Braatne et al. 1996). Abundant flowering may not occur until after age 10; however, substantial flowering typically begins earlier for intensively managed plantation poplars----during their fifth growing season in Pacific Northwest plantations (Stanton and Villar 1996). Floral development takes place over a year (Figure 1.4A). Poplar flowers are borne on pedunculate, pendulous racemes (i.e., catkins, aments) that appear early in spring before foliage develops. During or soon after anthesis, the inflorescences for next year are initiated in the axils of leaves on shoots of the current year's growth. The inflorescences develop as axillary accessory buds on short shoots and axillary lateral buds below the apical bud of main shoots; however, some axillary buds are vegetative. Each inflorescence differentiates acropetally. Lateral appendages arise helically and become thin, lacerate bracts. Floral meristems arise in the axil of each bract, with their differentiation continuing through autumn. After dormancy, megasporogenesis and microsporagenesis occur, and the inflorescences rapidly elongate and emerge from the bud scales.

Poplar flowers are considered highly evolved, with their apparent simplicity due to extreme reduction rather than an expression of archaic features (Fischer 1928, Eckenwalder 1996). Flowers consist of an outer whorl organ called the perianth cup

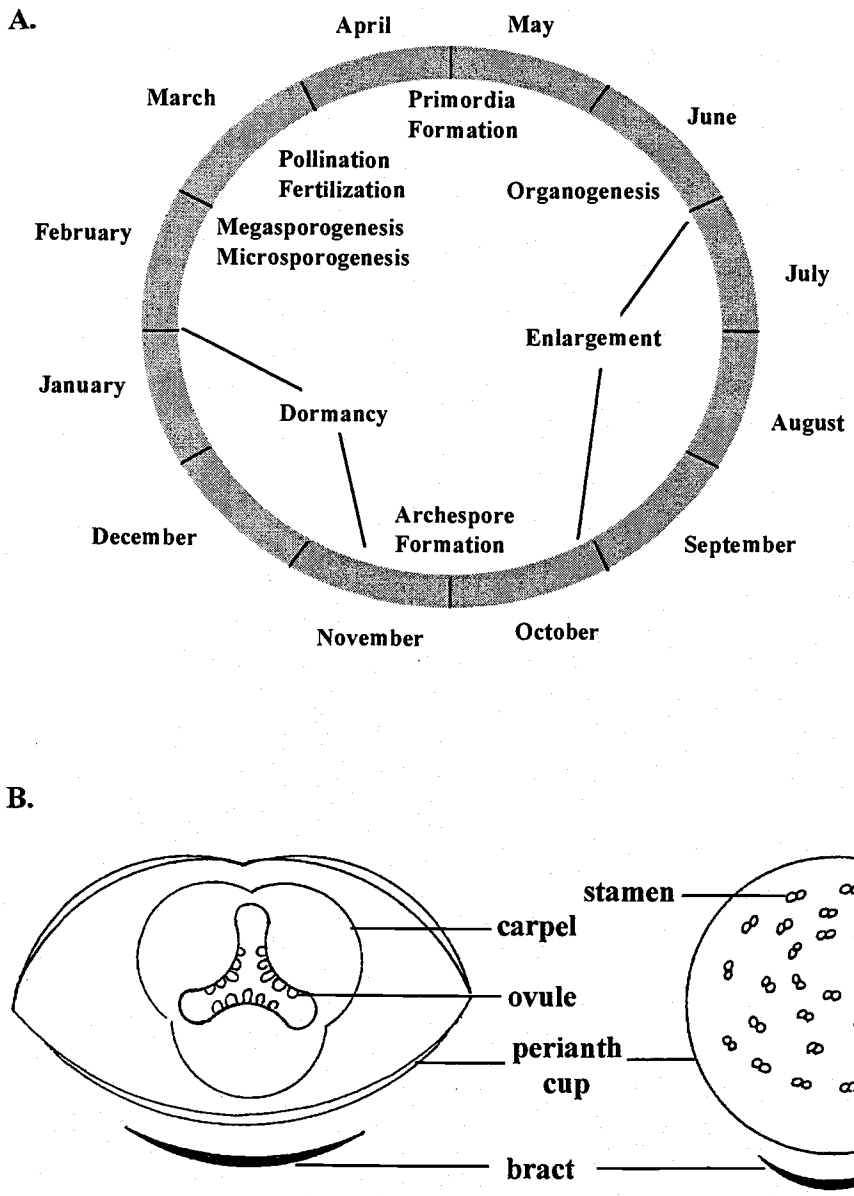


Figure 1.4. *Populus trichocarpa* floral development. **A**, A typical flowering cycle for trees in the vicinity of Corvallis, OR (redrawn from Boes and Strauss 1994). **B**, Floral diagrams for a female and male flower (S. Vollmer, unpublished data). b, bract; c, carpel; s, stamens; o, ovule; p, perianth cup.

and an inner whorl of either stamens or carpels (Figure 1.4B). Whether the perianth cup is derived from sepals, petals or both is uncertain. Female and male floral meristems are indistinguishable until the perianth cup primordia emerge from the outer whorl (Lester 1963, Boes and Strauss 1994, Kaul 1995). At this time, the central whorl of a female meristem becomes a slightly rounded, convex structure, while the center of the male meristem is concave. In a *P. trichocarpa* female, usually three, sometimes four carpel primordia emerge, eventually uniting to form a unilocular, superior ovary, style and stigmas (Boes and Strauss 1994). Thirty to fifty anatropous ovules develop from a parietal placenta. In the male flower, stamen primordia arise centrifugally; forty to fifty primordia differentiate short filaments and anthers consisting of two pollen sacs that are divided into two microsporangium locules.

Dioecy has evolved independently many times, with about 43 % of all families containing at least one dioecious species; yet only an estimated 4% of all angiosperms are dioecious (reviewed in Grant et al. 1994, Ainsworth et al. 1998). Most dioecious and monoecious species initiate both male and female reproductive primordia and then selectively degenerate or arrest development of one. The timing of the sex determination process varies among species, occurring very late in some so that male and female flowers are indistinguishable from perfect flowers. In contrast, poplars do not visibly initiate primordia of the alternate sex (e.g., Boes and Strauss 1994, Kaul 1995). However, hermaphroditic flowers as well as flowers of the opposite sex have been observed in most poplar species and in willows (e.g., Kaul 1995, Lester 1963, Meyer 1966, Stettler 1971). Studies of *P. trichocarpa* and other *Populus* species suggest that females and hybrids are more likely to exhibit variation in sex expression than male trees (Stettler 1971). Further, *P. trichocarpa* hermaphroditic flowers exhibited several forms, suggesting that alterations in the sex determination process can occur at different stages of flower development and/or in different locations within the flower. Stettler (1971) observed flowers with normal gynoecia that had antheroid structures attached to the inside of the perianth cup or to the carpel wall; other flowers had one to seven normal stamens with the filaments attached to the carpel base.

Little is known about the sex determination process in the Salicaceae. Studies support that gender is genetically controlled, though environmental factors may affect the determination process (reviewed in McLetchie and Tuskan 1994). Recent analyses indicate that gender is not controlled by sex chromosomes, ratios of autosome to sex-determining loci or simple Mendelian loci. A two-locus, multiple allele, epistatic gene model provided the simplest fit to data from studies of several full-sib families of *Salix viminalis* (basket willow). Using this model as a guide, bulk-segregant analysis identified a female-associated RAPD marker (Tuskan et al. 1996).

Ainsworth et al. (1998) proposed that dioecious species may be divided into two groups. One group includes species whose close relatives are mainly monoecious or dioecious and in which the differences between male and female flowers are programmed early in floral development. Dioecy in this group is suggested to have evolved from monoecy as an environmentally unstable system controlled by the levels of plant growth substances. The second group of species has mainly hermaphroditic relatives and the male and female flowers possess rudimentary organs of the opposite sex. Dioecy is very stable and there is no evidence for the involvement of plant growth factors; it is proposed to have arisen from hermaphroditism in these species. Although poplars and willows are clearly most like the first group of species, applications of plant growth factors have not been shown to affect sex expression. However, Stettler (1971) observed a correlation between maleness and catkin curvature in predominately female trees and suggested that auxin levels may have a role in sex determination.

Molecular studies of poplar flower development have recently begun. *PTFL*, the *P. trichocarpa* homolog of *FLO/LFY*, is weakly expressed in lateral primordia of vegetative meristems and strongly expressed in emerging male and female floral meristems (W. Rottmann unpublished data). Although the expression patterns of *PTFL* and *LFY* are very similar, differences in how these homologs interact with other genes to regulate flowering appear to exist. While transgenic hybrid aspen constitutively expressing *LFY* flowered within months (Weigel and Nilsson 1995), overexpression of *PTFL* has not induced flowering in transgenic aspen currently in their third growing

season. Characterization of the *P. trichocarpa* homolog (*PTD*) of the B function organ identity gene *AP3/DEF* revealed that it is expressed in the central whorl of male floral meristems and in developing stamens (Sheppard 1997). The absence of *PTD* expression in the outer whorl is consistent with the perianth cup being more sepal-like than petal-like.

Interestingly, *PTD* is also initially expressed in the central whorl of female meristems before carpel primordia emerge. *PTD* RNA was not detected in the carpels as they begin to form. Similarly, the *Antirrhinum* homolog, *DEF*, is transiently expressed in the fourth whorl, but not in the developing carpel primordia; *GLO*, the partner B-function gene, is not detected in the fourth whorl (Schwarz-Sommer et al. 1992, Trobner et al. 1992). Conversely, *AP3* is not expressed in the fourth whorl of *Arabidopsis* flowers, but *PI*, the *GLO* homolog, is transiently expressed in the fourth whorl (Goto and Meyerowitz 1994). Constitutive expression of both *AP3* and *PI* demonstrated that these two genes are sufficient to specify B function in *Arabidopsis* (Krizek and Meyerowitz 1996). Therefore, it would of interest to determine the expression pattern of the poplar *PI/GLO* homolog and whether *PTD* and the poplar *PI/GLO* homolog are sufficient to specify B function in a female tree (i.e., convert carpels to stamens). Alteration of the gender of poplar clones would facilitate production of interspecific hybrids, which often show non-reciprocal fertility barriers.

Chapter II: Structure and Expression of *PTAG1* and *PTAG2*: Two *AGAMOUS* homologs from Black Cottonwood (*Populus trichocarpa*)

ABSTRACT

To investigate the homeotic systems underlying floral development in a dioecious tree and to provide tools for the manipulation of floral development, we have isolated two *P. trichocarpa* genes, *PTAG1* and *PTAG2*, homologous to the *Arabidopsis* floral homeotic gene *AGAMOUS (AG)*. Intron/exon structure is conserved between *PTAG1, PTAG2* and *AG*, and intragenic regions of *PTAG1* and *PTAG2* share significant homology. *PTAG1* and *PTAG2* are the most similar paralogs with an *AG*-like floral expression pattern reported to date, and phylogenetic analysis of the *AG* subfamily supports that they are putative C-class organ identity genes. The high degree of similarity shared by *PTAG1* and *PTAG2* in both sequence (89 % amino acid identity) and expression indicate that they are unlikely to exhibit a diversification of function correlated with tree gender. Unexpectedly, *PTAG* transcripts were detected in vegetative tissue, suggesting that regulatory interactions involving C class organ identity genes may differ between *Populus* and annual herbaceous species.

INTRODUCTION

Genetic studies in the model plant genera *Arabidopsis* and *Antirrhinum* led to the formulation of the ABC model of floral organ identity (Coen and Meyerowitz 1991). According to this model, the combinatorial action of three classes of homeotic gene activities (A, B, and C) specify the four types of floral organs. Genes encoding these functions have been cloned from both species, and homologous genes have been isolated from a diversity of plants. Most organ identity genes belong to a family of transcription factors named after its highly conserved DNA binding and dimerization domain, the MADS-box (reviewed in Shore and Sharrocks 1995). Most plant members

of this family consist of four domains, including a second conserved domain, the K-box, which is thought to mediate protein-protein interactions (e.g., Fan et al. 1997). The two additional domains are the Intervening region (I-region), which lies between the MADS- and K-boxes and is necessary for dimerization (Riechmann et al. 1996), and the highly variable C-terminal region. A large number of additional genes that control various aspects of flowering have also been cloned and studied, revealing a complexity of regulatory interactions. Among these are genes that affect flowering time, enabling the complicated relationships between flowering time and floral meristem identity genes to begin to be defined (e.g., Simon et al. 1996).

Although analyses of floral homeotic genes in a diversity of species have revealed remarkable levels of conservation in sequence, expression pattern and function, these studies have also revealed significant variances. Compared to a herbaceous annual, a number of key differences are likely in the genetic control of both floral induction and floral morphogenesis in trees. To begin to address this, we are studying floral regulatory genes in poplar (genus *Populus,* includes aspens and cottonwoods). Our studies focus on *P. trichocarpa* (black cottonwood), which is native to the Pacific Northwest and widely used in poplar breeding programs in both North America and Europe. Several factors have resulted in *Populus* becoming a model system for genetic and molecular analyses of long-lived, woody plants (reviewed in Bradshaw 1998). Among these are fast growth, small genome size, ease of controlled breeding and vegetative propagation, and facile transformation/regeneration systems. An ultimate goal of our research is to enable the manipulation of flowering (promotion, inhibition, and gender determination) in order to accelerate the rate of genetic improvement and to mitigate unintended ecological effects that could result from cultivation of transgenic trees (reviewed in Strauss et al. 1995).

The reproductive phase of poplars typically begins at 5 to 10 years of age (Braatne et al. 1996). Floral development is initiated in late spring, nearly a year before anthesis occurs (Boes and Strauss 1994). Flowers are borne on pendulous racemes (i.e., catkins) that appear before flushing of vegetative buds. After anthesis, the

inflorescences for next year are initiated in the axils of leaves on shoots of the current year's growth. The inflorescences develop as axillary buds and differentiate acropetally. Floral meristems arise spirally in the axils of bracts, with their differentiation continuing through autumn. After dormancy, megasporogenesis and microsporagenesis occur, and the inflorescences rapidly elongate and emerge from the bud scales.

The distinctive development of poplar flowers makes them an attractive genus in which to study floral homeotic function. Poplars are dioecious and their flowers are considered highly evolved, having undergone extreme reduction (Fischer 1928, Eckenwalder 1996). Flowers consist of an outer whorl organ called the perianth cup and an inner whorl of either stamens or carpels. Whether the perianth cup is derived from sepals, petals or both is uncertain. Unlike most dioecious and monoecious species, which selectively arrest development of stamen or carpels, poplars do not initiate primordia of the alternate sex (e.g., Boes & Strauss 1994, Kaul 1995, Sheppard 1997). However, hermaphroditic flowers as well as flowers of the opposite gender have been observed in most poplar species and in the other genus of the family Salicaceae, *Salix* (willows) (e.g., Lester 1963, Meyer 1966, Stettler 1971). Little is known about the mechanism of sex determination in the Salicaceae. Gender is genetically controlled, though environmental factors may affect the determination process; gender is not controlled by sex chromosomes or ratios of autosome to sex-determining loci (reviewed in McLetchie and Tuskan 1994).

In the perfect flowers of *Arabidopsis* and *Antirrhinum*, class B and C genes act together to specify stamens in whorl 3, while C activity alone specifies carpels in whorl 4. In *Arabidopsis*, *AG* specifies C function, while the corresponding homolog in *Antirrhinum* is *PLENA (PLE)* (Yanofsky et al. 1990, Bradley et al. 1993). The expression of the C function genes corresponds to their domain of activity---whorls 3 and 4. *AG/PLE* also specifies determinacy of the floral meristem, and *AG* has an important role in maintaining floral meristem identity (Mizukami and Ma 1997, Okamuro et al. 1996). In flowers of *ag/ple* mutants, whorls 3 and 4 are transformed, resulting in the repetition of a sepal-petal-petal pattern. Conversely, ectopic expression

transforms whorls 1 and 2, resulting in a carpel-stamen-stamen-carpel flower. Genes from a number of species have been isolated that exhibit a similar expression pattern (e.g., Pnueli et al. 1994, Hardenack et al. 1994), and phylogenetic analysis places these genes in the same monophyletic clade as *AG* and *PLE* (Purugganan 1997).

While a single gene is sufficient to define C function in *Arabidopsis* and *Antirrhinum*, this is apparently not the case in all species. In maize, two genes, *ZAG1* and *ZMM2*, appear to have distinct but partially redundant functions (Mena et al. 1996). Flowers of *zag1* mutants are indeterminate, but normal pollen-producing stamens develop and the innermost set of organs has carpelloid features. Both genes are expressed in carpels and stamens; however, *ZAG1* RNA accumulates more in carpels than in stamens, while *ZMM2* RNA accumulates more in stamens than in carpels. In petunia, ectopic expression of the *AG* homolog *PMADS3* resulted in only partial homeotic transformations of first and second whorl organs; sepals developed carpelloid features and petals were partially converted to antheroid structures (Tsuchimoto et al. 1993). Constitutive expression of a second petunia *AG* homolog, *Floral Binding Protein 6 (FBP6)*, only affected petals; they were reduced in size and altered in shape, but no antheroid tissues were produced (Kater et al. 1998).

Floral homeotic genes have been studied in the annual dioecious species white campion (Hardenack et al. 1994) and sorrel (Ainsworth et al. 1995). Unlike poplars, both species have flowers with 4 whorls and initiate primordia of the alternate sex. Expression of a white campion *AG* homolog was not correlated with the sex determination process. In contrast, the putative sorrel C function gene, *RAP1*, is initially expressed in whorls 3 and 4 of both female and male flowers, but expression is not maintained in the whorl that undergoes developmental arrest.

We have isolated cDNA and genomic clones for two *AG* homologs from *P. trichocarpa,* a dioecious forest tree. We show that the genes are very similar to each other in sequence and expression, and are clear phylogenetic homologs to *AG*. *PTAG1* and *PTAG2* are expressed in the inner whorl of developing floral meristems from female

and male trees. They are the only close *AG* homologs yet reported that also show significant vegetative expression.

MATERIALS AND METHODS

Plant Materials and Nucleic Acid Extraction

Tissues were gathered from wild *P. trichocarpa* trees growing in the vicinity of Corvallis, Oregon, U.S.A. Immature inflorescence tissue was collected in late May to early June. At this time, floral meristems had initiated and floral organ primordia had formed in some of the meristems. Mature inflorescences were collected in late February as they were emerging from the buds. After removing the bud scales, entire inflorescences were collected; tissue samples from separate trees were not combined. Vegetative buds were gathered from flowering branches in late February as dormancy ended. Stems (current year's growth), leaves and vegetative buds were collected in late May from non-flowering basal branches and root suckers. Genomic DNA was isolated from vegetative buds using a modified CTAB method (Wagner et al. 1987). Total RNA was extracted according to the method of Hughes and Galau (1988).

Isolation of cDNA and Genomic Clones

Immature and mature female floral cDNA libraries were prepared using the λ ZAP cloning kit (Strategene). Approximately 10^6 clones were obtained per preparation, with an average insert size of 1 kb. To construct the genomic library, DNA from a single male tree was partially digested with Sau3A1, ligated into λ GEM12 (Promega) at partially filled-in Xho1 sites, and packaged using GigaPack Gold II (Stratagene).

Approximately 6.5 x 10^5 recombinant clones were recovered and the library was amplified.

All probes were labeled using the Boehringer Mannheim random primer labeling kit. To avoid cross-hybridization with other MADS-box genes, all probes lacked most or all of the MADS-box. A 737 bp HindIII-EcoRI fragment of the *AGAMOUS* cDNA (Yanofsky et al. 1990) was isolated from pCIT565 (provided by E. Meyerowitz, California Institute of Technology) and used to screen 6 x 10^5 genomic clones. Hybridization was performed at 65 °C in 5X SSPE/5X Denhardt's solution/ 0.2% SDS/100 μg/ml denatured salmon sperm DNA. Filters were washed at 42°C in 2X SSPE/0.2 % SDS and in 0.2X SSPE/0.2% SDS. A *PTAG1* cDNA clone was obtained by probing 1.3 x 10^5 clones of an immature female floral cDNA library with the HincII - XbaI fragment of pCIT565. To isolate a *PTAG2* cDNA clone, gene-specific primers 5'-CGACAGCACATGAATTTG-3' and 5'-TTACACTAACTGAAGAGCTGG-3' were designed based on the *PTAG2* genomic sequence. A 147 bp fragment corresponding to the 3' end of the *PTAG2* coding region was obtained by PCR using a mature female floral cDNA library as the template. The fragment was then used to probe 9 x 10^5 clones of a mature female floral library at high stringency. Hybridization conditions were as described above, but washes were performed at 65°C.

Fragments of the genomic clones and the *PTAG1* cDNA clone were subcloned into pBluescript KS (Stratagene) to facilitate sequencing (Appendix D). Sequencing was done at the Oregon State University Central Services Laboratory using an Applied Biosystems model 377 sequencer. Universal primers as well as specific primers designed using Generunner 3.04 (Hastings Software, Inc.) were used in sequencing reactions.

Sequence Analyses

Pairwise sequence alignments were performed using the ALIGN program of the FASTA version 2.0 sequence analysis package (Pearson 1990). Gaps were not counted

in determining percent identity. Maximum likelihood estimates of synonymous and nonsynonymous nucleotide substitution rates (Muse and Gaunt 1994) and the corresponding estimates of Nei and Gojobori (1986) were calculated using the SYNDIST program (Muse 1996). Multiple alignments of protein sequences were done using Clustal W version 1.7 (Thompson et al. 1994) and refined visually for phylogenetic analyses. Figure 2.6 was prepared using the GeneDoc program (Nicholas and Nicholas 1997). Groups of amino acids considered to be similar were L, I, V, M (hydrophobic), D, E (acidic), N, Q (amide), F, Y, W (aromatic), H, K, R (hydrophilic, basic), and P, A, G, S, T (small, neutral or weakly hydrophobic). All sequences were obtained from GenBank.

The neighbor-joining tree method (Saitou and Nei, 1987) of the MEGA computer program (Kumar et al. 1993) was used for distance-based phylogenetic analysis. Poisson-correction distance (d) was used to estimate the number of amino acid substitutions per site assuming a Poisson distribution: $d = -\ln (1 - p)$, where $p = Nd/N$ is the proportion of different amino acids between two sequences compared, and Nd and N are the number of amino acid differences and the total number of amino acids compared, respectively. Consensus trees and estimates of statistical confidence were inferred from 1,000 bootstrapped data sets. Parsimony analysis was performed using the PROTPARS program in the PHYLIP software package (Felsenstein, 1989). The SEQBOOT program was used to generate 1,000 data sets, and the majority-rule and strict consensus trees were generated from bootstrap parsimony trees using the program CONSENSE. The phylogenetic trees were viewed and drawn using the TREEVIEW program (Page 1996). For the analysis displayed in Figure 2.3, the mammalian (*Mus musculus*) MADS-box gene *MEF2C* served as the outgroup. For analysis of the *AG* subfamily (Figures 2.4 and 2.5), *Arabidopsis* and *Antirrhinum* orthologs [*APETALA1 (AP1)/SQUAMOSA (SQUA), APETALA3 (AP3)/ DEFICIENS (DEF),* and *PISTILLATA (PI)/ GLOBOSA (GLO)*] from different floral homeotic gene clades were used as outgroups.

Northern and Southern Analyses

Analyses were done using standard procedures (Sambrook e al. 1989). For the Southern blot, 5 μg DNA samples from a male and a female tree were digested with restriction enzymes and blotted onto a nylon membrane. Gene-specific probes were obtained by PCR and labeled with a random primer labeling kit (Boehringer Mannheim). A 274 bp *PTAG1* PCR fragment, corresponding to the last 147 bp of the coding region and part of the 3' UTR was synthesized using 5'-CGACAGAGCATGAATTTG-3' and 5'-CCAGACAAATATGATTTACG-3'. The *PTAG2* probe was the 147 bp PCR fragment used to isolate the cDNA clone (described above). Hybridization was performed at 68°C in 5X SSPE/5X Denhardt's solution/ 0.2%SDS/0.05X Blotto/100 μg/ml denatured salmon sperm DNA and washes were done at 68°C in 2X SSPE/0.2% SDS followed by 0.1X SSPE/0.5%SDS.

Total RNA samples (20 μg) were electrophoresed in formaldehyde agarose gels and blotted onto Hybond (Amersham). Short PCR fragments corresponding to portions of the 3' UTR regions were synthesized for use as probes to eliminate any cross-reaction between *PTAG1* and *PTAG2*. A 115 bp *PTAG1* fragment was obtained using primers 5'-CCTGGGTTTCCATTGAGC-3' and 5'-GGATAGTTAATACATAGAGGAAGAG-3' and a 118 bp *PTAG2* fragment was generated with primers 5'-GTACCTACTATTTC-ACTGAGCG-3' and 5'-AAAGCAATACATGGAGGAAGAG-3'. Fragments were subcloned into pBluescript KS (Stratagene), linearized with EcoRI, and ^{32}P-labelled antisense RNA probes were synthesized using T7 polymerase and Ambion's Maxiscript kit. Blots were hybridized at 65°C in 50% formamide/5X SSC/5X Denhardt's solution/ 0.5% SDS/100 μg/ml denatured salmon sperm DNA and final washes were at 65°C in 0.1X SSC/0.1% SDS. The specificity of the probes was tested under identical conditions using *in vitro* transcribed *PTAG1* and *PTAG2* RNAs (data not shown). Blots were stripped and re-probed with a SstI- XbaI fragment of the *P. deltoides* 18S rDNA isolated from pPD5 (D'Ovidio et al. 1991). Blots were exposed to Kodak BioMax Film. Quantitation of RNA bands was done using a Molecular Dynamics phosphorimager and

ImageQuaNT software version 4.2. Expression levels of *PTAG1/2* (percents of total signal among all bands) were adjusted for background (signal in empty lane) and equal loading, based on the assumption that the 18S RNA is expressed equally in all tissues.

RT-PCR

PTAG1 transcripts from floral and vegetative tissues were analyzed using the Superscript One-Step RT-PCR system (Life Technologies). cDNAs were synthesized using 200 ηg of total RNA according to the manufacturer's protocol. Primers 5'-AAGATCCTCACTTTCTCTACAC-3' and 5'-CCCAGACAAATATGATTTAC-3' were used to synthesize and amplify a product containing all intron-exon junctions.

In situ Hybridization Studies

Tissue samples were fixed, embedded, sectioned, and hybridized according to the procedure described by Kelly et al. (1995) with some modifications. Antisense and sense *35S*-probes were generated from a plasmid consisting of a 800 bp Pst1-Xho1 fragment from the *PTAG1* cDNA clone inserted in pBluescript KS (Stratagene), and were not alkaline hydrolyzed. These probes lacked the MADS-box and included part of the I-region and the remaining 3' sequences of the cDNA. Short, gene-specific probes derived from the *PTAG1* and *PTAG2* 3' UTR regions (described above for northern blots) were also used (data not shown). In some cases, probes were purified through Nunctrap columns (Stratagene) to reduce background. Probes were applied to 10 μm sections at a final concentration of 2×10^7 cpm/ml (approximately 30-40 ηg/ml). Slides were photographed using a Contax camera mounted on a Zeiss Axioskop microscope equipped with a darkfield illuminator. Photographic slides were scanned, digitized and adjusted for contrast, brightness and color balance using Adobe Photoshop version 4.0.

RESULTS

Structure of *PTAG1* and *PTAG2* cDNAs and Genes

A genomic library prepared using DNA isolated from a single male tree was screened with a fragment of the *AG* cDNA that lacked most of the MADS-box region. Based on restriction mapping, two classes of genomic clones were identified. Sequencing of the genomic clones revealed that the two classes corresponded to two closely related genes. Using a fragment of the *AG* cDNA 3' to the MADS-box as a probe, a near full-length *PTAG1* cDNA clone was isolated from an immature female floral library. Based on *PTAG2* genomic sequence, a partial *PTAG2* cDNA clone was isolated from a mature female library by PCR, and this sequence was then used to isolate a near full-length cDNA clone.

Figure 2.1 shows the *PTAG1* and *PTAG2* cDNA sequences and the predicted translation of the coding sequences. The cDNA sequences are identical to the corresponding genomic sequences except for two nucleotide changes in the 5' UTR and one synonymous base transition in the I-region of *PTAG2*. *PTAG1* encodes a 241 amino acid protein, while *PTAG2* encodes a slightly shorter (238 amino acid) protein, apparently due to a frameshift caused by a four bp deletion near the C-terminal end of coding sequence. A portion of the 5' flanking sequence is also displayed, and regions matching the eukaryotic TATA box consensus sequence TATA(T/A)A(T/A) are indicated (Breathnach and Chambon 1981). A total of 11,485 bp of the *PTAG1* gene locus contained in a single genomic clone, and 10,007 bp of the *PTAG2* locus contained in two overlapping clones, were sequenced and are available as GenBank accession numbers AF052570 and AF052571 (Appendix A). Like *AG* (Yanofsky et al. 1990) and *PLE* (Bradley et al. 1993), both *PTAG1* and *PTAG2* contain 8 introns at conserved positions (Figure 2.1). All introns have canonical donor (GT) and acceptor (AG) sites.

Figure 2.1. *PTAG1* and *PTAG2* cDNA and upstream sequences. The complete *PTAG1* nucleotide and deduced amino acid sequences are shown. The MADS-box is underlined and in boldface type; the K-box is underlined. Only nucleotides and residues in *PTAG2* that are different from *PTAG1* are displayed. Nucleotides present in the cDNAs are in capital letters. Dashes indicate gaps introduced in the alignment; dots denote *PTAG2* nucleotides identical to *PTAG1*. Intron positions are indicated by triangles. TATA motifs are boxed. Three nucleotides that differ between the *PTAG2* cDNA and genomic sequence are indicated by bold type.

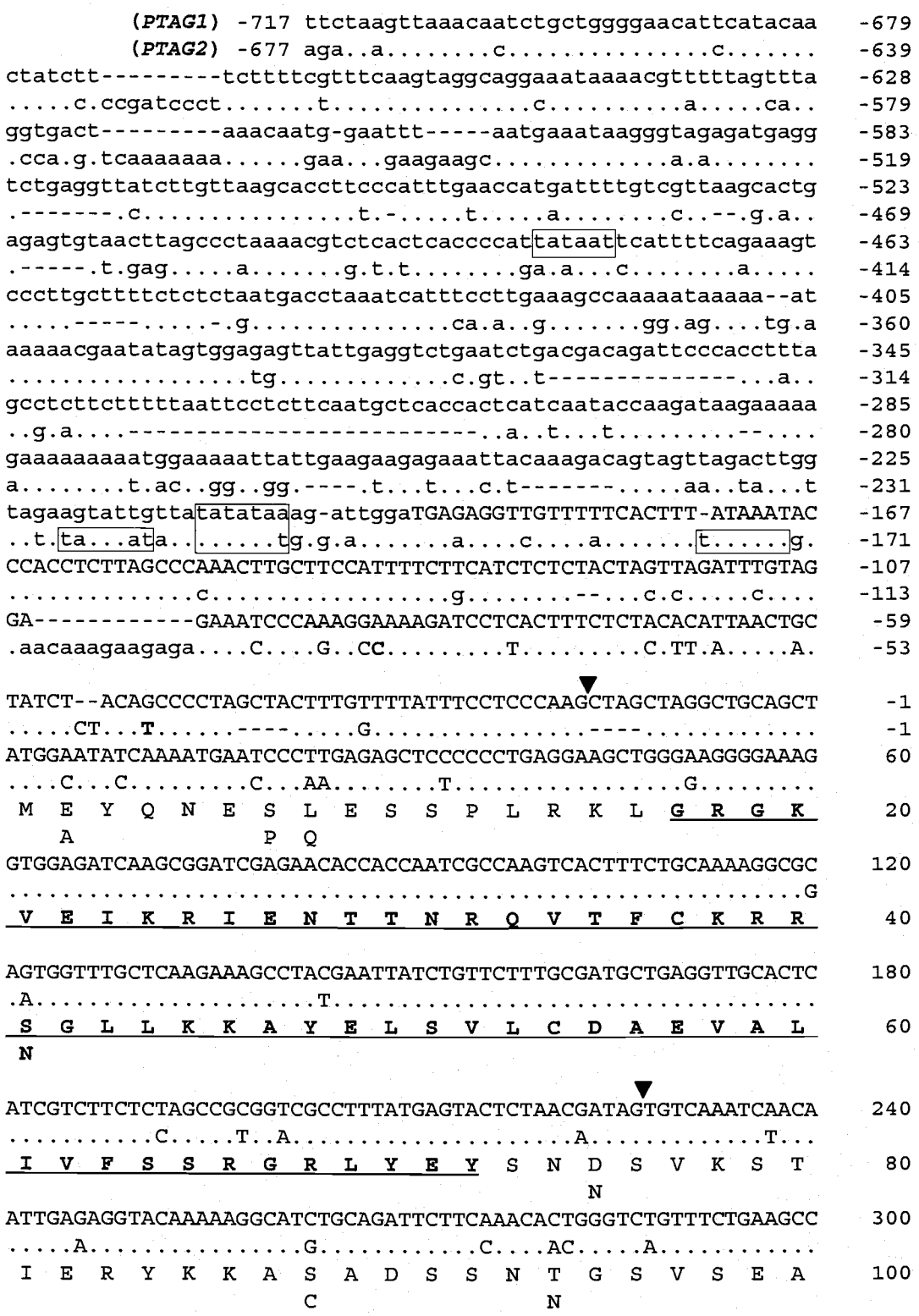


Figure 2.1, continued

```
AATGCTCAGTACTACCAGCAAGAAGCTGCCAAGCTGCGTTCCCAAATTGGTAATTTGCAG     360
.........T...T.....................C..G.................
 N  A  Q  Y  Y  Q  Q  E  A  A  K  L  R  S  Q  I  G  N  L  Q     120
          F

AATTCAAACAGGCATATGCTGGGTGAAGCGCTTAGTTCATTGAGTGTGAAGGAACTTAAG     420
...........A.............T.A......G.....................
 N  S  N  R  H  M  L  G  E  A  L  S  S  L  S  V  K  E  L  K     140
             N              S        A

AGTTTGGAAATACGACTTGAGAAAGGAATAAGCAGAATTCGTTCCAAAAAGAATGAGCTG     480
..C.....G...AA.............TG.T..........G.............
 S  L  E  I  R  L  E  K  G  I  S  R  I  R  S  K  K  N  E  L     160
             K                 G

TTGTTTGCAGAAATCGAGTATATGCAGAAGAGGGAGGTTGACTTGCACAACAATAACCAG     540
........T.....T.....................A.....................
 L  F  A  E  I  E  Y  M  Q  K  R  E  V  D  L  H  N  N  N  Q     180
                                     I

CTTCTCCGAGCAAAGATTTCAGAGAATGAAAGAAAGCGACAGAGCATGAATTTGATGCCA     600
.................G.......................CA..............G
 L  L  R  A  K  I  S  E  N  E  R  K  R  Q  S  M  N  L  M  P     200
                   A                       H

GGAGGAGCAGACTTTGAGATCGTGCAGTCTCAACCATATGACTCTCGGAACTATTCTCAA     660
.....T.TCA....C......A..............T.....................
 G  G  A  D  F  E  I  V  Q  S  Q  P  Y  D  S  R  N  Y  S  Q     220
       V  N           M              F
GTGAATGGATTACAGCCTGCAAGTCATTACTCACATCAAGATCAGATGGCCCTTCAGTTA     720
..T........G.C......C.A.......C.T...G....C...C.----.........     716
 V  N  G  L  Q  P  A  S  H  Y  S  H  Q  D  Q  M  A  L  Q  L     240
             P        N        P     E        L      F  S  *     238

GTTTAATAATCTCCAAGTGCAGCAGTTTCTCGCATTTCCATATTCCATGGAGAGTACCTG     780
......A..G.A............AC.C.............................A     776
 V  *                                                          241

GGTTTCCATTGAGCGCAAAAGCTACATGTATGCTAAAA----AACCTGAAGTAGCGTAAA     836
CTA..T..C.............G..A...C........CAAA..T..........A...C     836
TCATATTTGTCTGGGTGGGAGGGCCTAGTACTCTTCCTCTATGTATTAACTATCCTGTCC     896
...A......GCC.....-..A................C.......GCT.T...A....     895
CAGTTAAGACATAAGAAATGTCAGAGAAGGATTTCTTTTCTGTATGTTTCATGAAGGCAT     956
.............C..........T..............C.................C     955
TAAGATGCTGTTACAGT---TGTGACTAACTTATTATATAT---GTCTTACTGCTTC(A)n  1007
..........G.....ACT.....................TTT......TATTTCTT(A)n 1013
```

Figure 2.1, continued

The *PTAG1* and *PTAG2* cDNAs are highly homologous in both the coding and untranslated regions, suggesting that they may be the result of a relatively recent duplication event (Figure 2.1, Table 2.1). Interestingly, the coding regions share slightly greater nucleotide identity than amino acid identity. The different coding domains are evolving at different rates. Ratios of nonsynonymous substitution rate to synonymous substitution rate revealed that the MADS-box is under the greatest level of constraint, while the C-terminal region is evolving at the fastest rate, with a ratio seven times greater than the MADS-box ratio. Further evidence of a recent duplication event is seen in a comparison of the genomic sequences (Table 2.2). Homologous introns share significant sequence identity, with most of the differences due to insertions/deletions (Appendix A). Although differing by almost 1 kb in length, the large second introns contain several stretches of around 100 to 400 bp that are approximately 85% identical. The 5' flanking regions also exhibit strong homology, particularly in the 500 bp proximal to the start of the cDNAs. Sequence homology decreases markedly in more distal upstream sequences and appears largely insignificant, though short (~10 bp) homologous motifs may have some regulatory significance.

Initial analyses of SSR markers derived from the *PTAG1* and *PTAG2* genomic sequences in a mapping pedigree (Bradshaw et al. 1994) indicate that the genes are located in different linkage groups (S.P. Difazio, S. Leonardi, and H.D. Bradshaw, pers. comm.). Southern analysis using gene-specific probes demonstrated that single copies of *PTAG1* and *PTAG2* are present in both male and female genomes (Figure 2.2). DNA restriction fragment patterns were identical for *PTAG1* between a female and male tree. A polymorphism was detected for *PTAG2*; two bands were seen in the XbaI digest of DNA from a female tree versus only one band in the digest of DNA from a male tree.

Table 2.1. Comparison of *PTAG1* and *PTAG2* by domain.

Region (length)	Percent Identity (nucleotide)	Percent Identity[1] (amino acid)	K_a/K_s[2]
N-terminal (48 bp)	**87.5**	**81.3** (81.3)	**0.295**
MADS-box (168 bp)	**95.8**	**98.2** (100)	**0.050**
I-region (99 bp)	**89.9**	**87.9** (90.9)	**0.176**
K-box (201 bp)	**92.0**	**92.5** (98.5)	**0.118**
C-terminal (207/198 bp)	**88.9**	**78.8** (87.9)	**0.344**
Entire coding (723/714 bp)	**91.5**	**88.7** (93.7)	**0.196**
5'UTR[3] (195/200 bp)	**88.1**		
3'UTR (284/299 bp)	**87.2**		

[1]Percent similarity is shown in parentheses.

[2]Maximum likelihood estimates of non-synonymous substitution rate(K_a)/synonymous substitution rate (K_s) ratio (Muse and Gaunt 1994). The corresponding estimates of Nei and Gojobori (1986) were not significantly different.

[3]Start corresponds to the first nucleotide of the *PTAG1* cDNA; the first 102 bp of the *PTAG2* 5' UTR comes from the genomic sequence.

Table 2.2. Comparison of *PTAG1* and *PTAG2* genomic sequences

Region	Percent Identity	Length *PTAG1/PTAG2*
Upstream region	61.7	2410/2234
Upstream region (distal)	57.0	1890/1761
Upstream region (proximal)	78.8	520/473
Exon 1[1]	**88.4**	**178/187**
Intron 1	80.4	464/492
Exon 2	**93.8**	**244/240**
Intron 2	78.8	4865/3882
Exon 3	**91.5**	**82/82**
Intron 3	70.7	651/829
Exon 4	**91.9**	**62/62**
Intron 4	80.7	85/86
Exon 5	**88.0**	**100/100**
Intron 5	80.0	143/137
Exon 6	**95.2**	**42/42**
Intron 6	76.3	203/156
Exon 7	**97.6**	**42/42**
Intron 7	82.2	140/129
Exon 8	**85.0**	**167/163**
Intron 8	76.7	1052/533
Exon 9	**87.3**	**285/295**
All introns	77.9	7603/6244
All exons	**90.0**	**1202/1213**

[1]Start of exon 1 corresponds to the first nucleotide of the *PTAG1* cDNA clone.

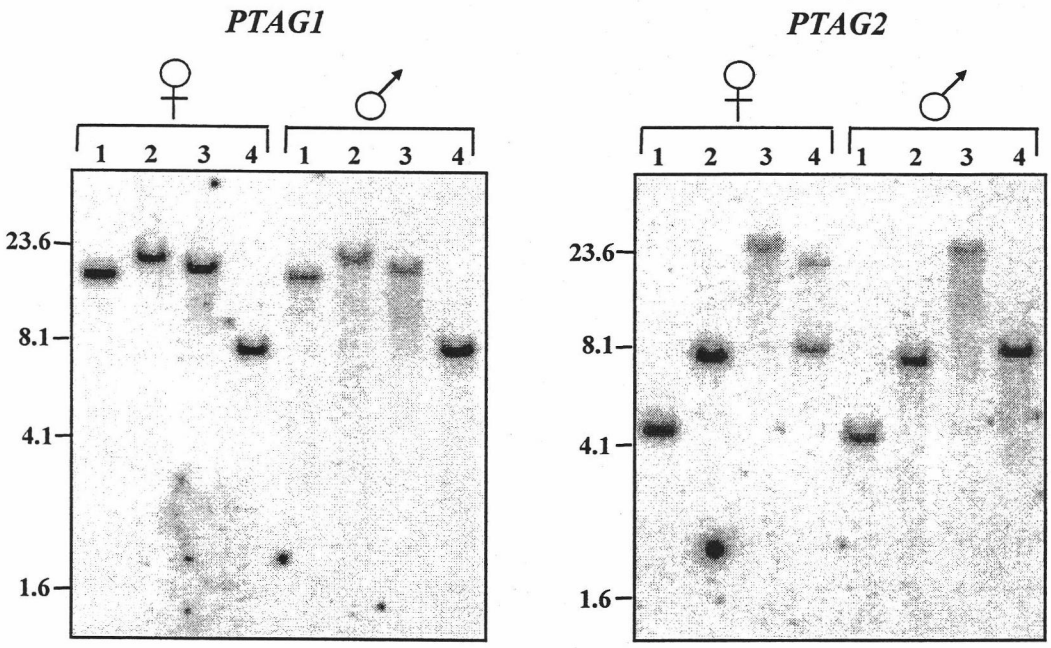


Figure 2.2 Southern blot analysis of *PTAG1* and *PTAG2*. Genomic DNA from a female and a male *P. trichocarpa* tree was digested with the same four restriction enzymes: 1, EcoRI; 2, PstI; 3, SstI; 4, XbaI. Molecular weight markers (kb) are shown.

Relationships among Members of the *AG* Subfamily

Previous phylogenetic analyses using the MADS-box, I-region and K-box (MIK region) revealed that most plant MADS-box genes are organized into monophyletic clades; these clusters generally correspond to groups of genes that share related functions (e.g., Purugganan et al. 1995). Based on the these studies, we selected a representative subset of dicot MADS-box genes to show that *PTAG1* and *PTAG2* are clearly members (95% bootstrap support) of the *AG* clade (Figure 2.3; Appendix C). In addition to dicot genes, several monocot genes as well as a conifer gene belong to the *AG* group. In order to study the relationships within the *AG* subfamily, we performed phylogenetic analyses using all members of this group and members of separate floral homeotic gene clades as outgroups. By limiting the analyses to this group, entire coding regions could be aligned reasonably well (Appendix B).

Figure 2.4 depicts a tree derived using the neighbor-joining procedure. A phylogenetic tree estimated by maximum parsimony has a similar topology, though a few differences were apparent (Figure 2.5). In both trees, *PTAG1* and *PTAG2* cluster with all but one of the dicot genes expressed in both male and female reproductive primordia, which we refer to as the dicot C-class group. The exception is the bisexually expressed cucumber gene *CUM10* (Kater et al. 1998), which forms a strongly supported group (91% and 87% bootstrap support) with three dicot genes expressed in ovules (Colombo et al. 1995, Angenent et al. 1995, Rounsley et al. 1995). The female-specific *Arabidopsis* genes *AGL1* and *AGL5* (Ma et al. 1991) are most closely related to the dicot C-group, and maximum parsimony analysis places them in the same cluster (Figure 2.5). Similar to the dicots, the monocot members of the *AG* subfamily organize into groups that correspond to their expression patterns. The gymnosperm gene *DAL2* (Tandre et al. 1995) is separate from all other members of the *AG* subfamily in both trees.

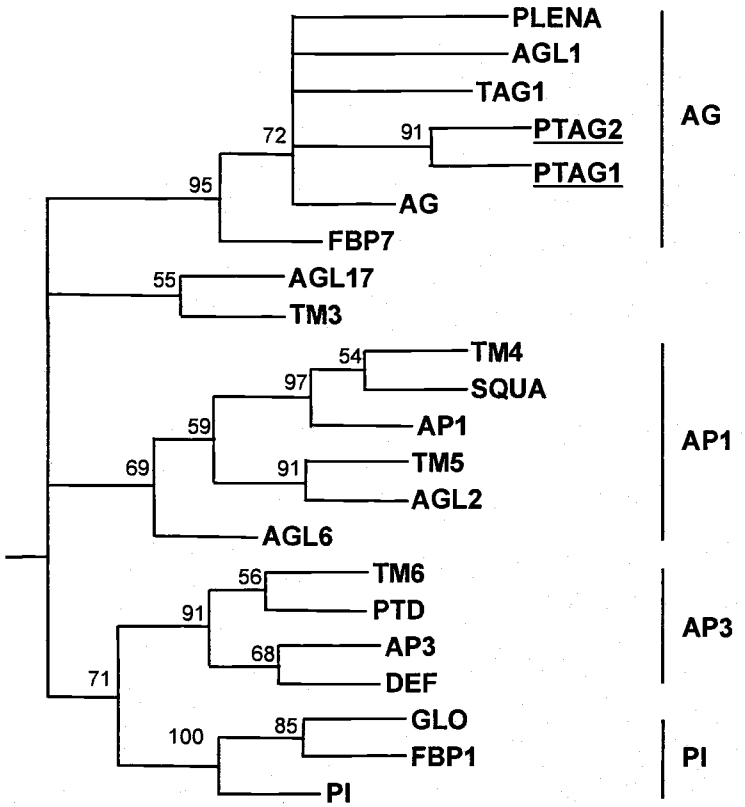


Figure 2.3 Phylogeny of plant MADS-box genes estimated by maximum parsimony. This analysis was conducted on deduced amino acid sequences of the MIK region. Numbers at nodes are bootstrap values. All nodes with less than 50% bootstrap support are collapsed on the tree. The mammalian gene *MEF2C* was the outgroup. Genes are from the following species: *Arabidopsis thaliana* (*AG, AGL1, AGL2, AGL6, AGL17, AP1, AP3, PI*); *Antirrhinum majus* (*DEF, GLO, PLENA, SQUA)*; *Petunia hybrida (FBP1, FPB7)*; *Lycopersicon esculentum (TAG1, TM3-6); Populus trichocarpa (PTAG1, PTAG2, PTD).*

Figure 2.4 Phylogeny of the *AG* subfamily estimated by the neighbor-joining procedure. Bootstrap values are indicated and nodes with values less than 50% are collapsed. Predicted amino acid sequences of the entire coding region were used in this analysis. Genes expressed in both female and male tissues are in gray boxes, while those expressed only in female tissues are in black boxes. Genes are from the following species: *Arabidopsis thaliana* (*AG, AGL1, AGL5, AGL11, AP1, AP3, DEF); Antirrhinum majus (PLE, SQUA, DEF, GLO); Brassica napus (BAG); Cucumis sativus (CUM1, CUM10, CUS1); Panax ginseng (GAG2); Lycopersicon esculentum (TAG1); Nicotiana tabacum (NAG1)Petunia hybrida (FBP6, FBP7, FBP11, PMADS3); Petunia inflata (PAGL1), Picea abies (DAL2); Populus trichocarpa (PTAG1, PTAG2); Oryza sativa (OSMADS3); Rumex acetosa (RAP1); Silene latifolia (SLM1); Zea mays (ZAG1, ZAG2, ZMM1, ZMM2).*

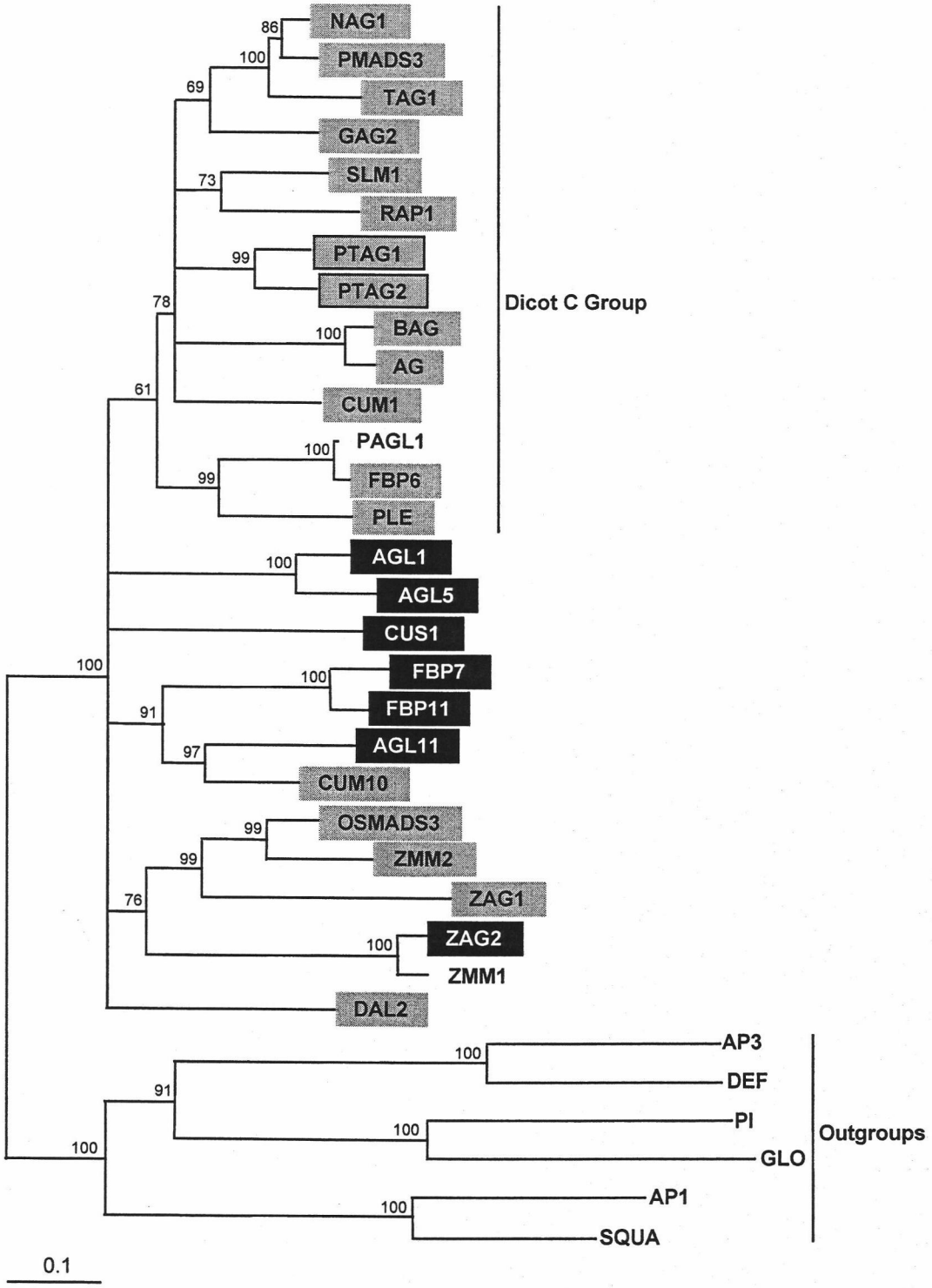


Figure 2.4, continued

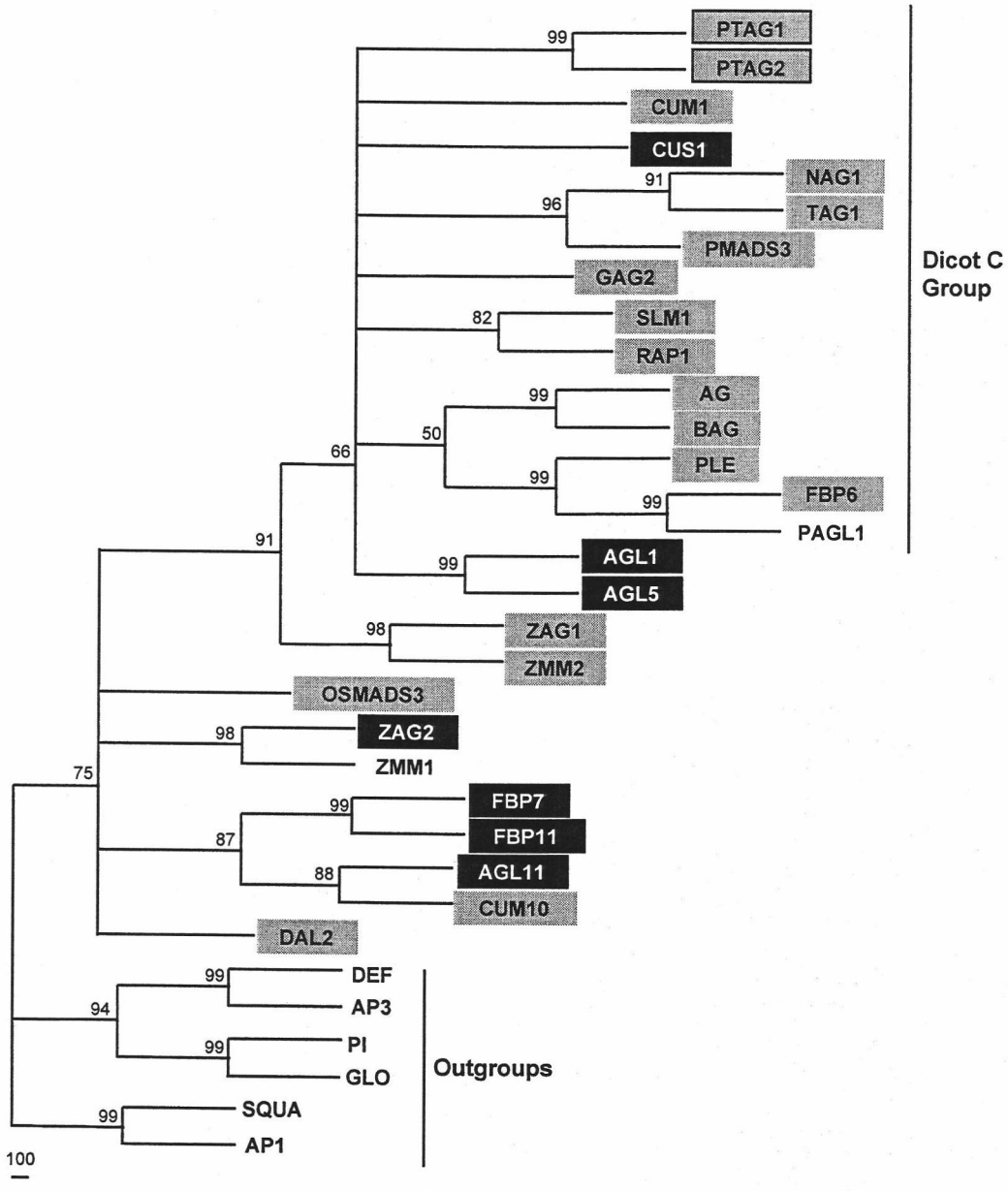
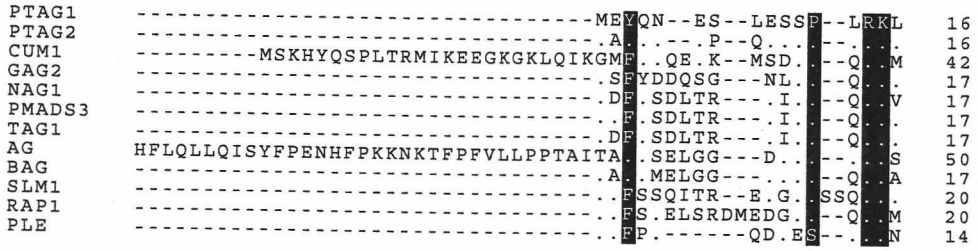


Figure 2.5 Phylogeny of the *AG* subfamily estimated by maximum parsimony analysis. See legend to Figure 2.4 for details.

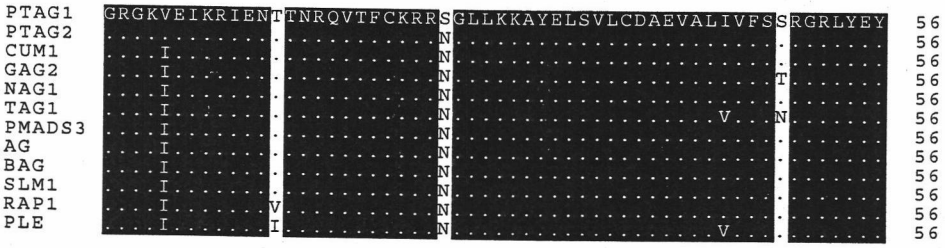
The PTAG1 and PTAG2 proteins are 76% and 72% identical to AG and share a similar level of identity with other members of the dicot C-group. Figure 2.6 is an alignment of PTAG1 and PTAG2 with other members of the dicot C-group. The two α helices of a proposed coiled-coil motif in the K-box and the a and d positions of the helix heptad repeat sequence are indicated (Ma et al. 1991). These positions are typically occupied by hydrophobic and apolar residues in coiled-coils (Kohn et al. 1997). The second α helix and the adjacent C-terminus of the K-box are particularly well conserved among these proteins. Although the C-terminal domain is highly variable, blocks of amino acids are conserved among all members of this group, particularly near the N-terminal end of this domain.

Of the genes belonging to the dicot C-group, only the gene structures of *AG* (Yanofsky et al. 1990), *PLE* (Bradley et al. 1993), *PTAG1* and *PTAG2* have been reported---- all have eight introns at conserved positions (Figure 2.7). While intron 1 is within the predicted translation of *AG*, it is located in the 5' UTR of *PLE*, *PTAG1* and *PTAG2*. *AG* is unusual in that it does not contain an initiating methionine and its N-terminal domain is 3 times the length of the *PTAG1/2* and *PLE* N-terminal domains (Figure 2.6). The initiating methionine of *BAG* (Mandel et al. 1992), the *AG* homolog from the close *Arabidopsis* relative *Brassica napus*, clearly aligns (Figure 2.6) with an *AG* residue 3' of the site of the first intron. Therefore, the position of intron 1 appears essentially homologous in all four genes. Unlike the other three genes, intron 8 is located in the 3' UTR of *PTAG2* rather than in the last amino acid codon. However, alignment of the *PTAG1* and *PTAG2* cDNAs (Figure 2.1) clearly show that this is due to a frameshift introducing a slightly premature stop codon rather than a shift in intron position. In contrast, the two genes most closely related to the dicot C-group have somewhat different structures. *AGL1* and *AGL5* lack intron 8 and *AGL5* also lacks intron 4 (Ma et al. 1991, Savidge et al. 1995).

N-terminal region



MADS-box



I-region

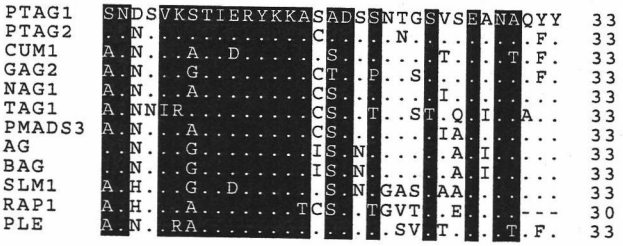
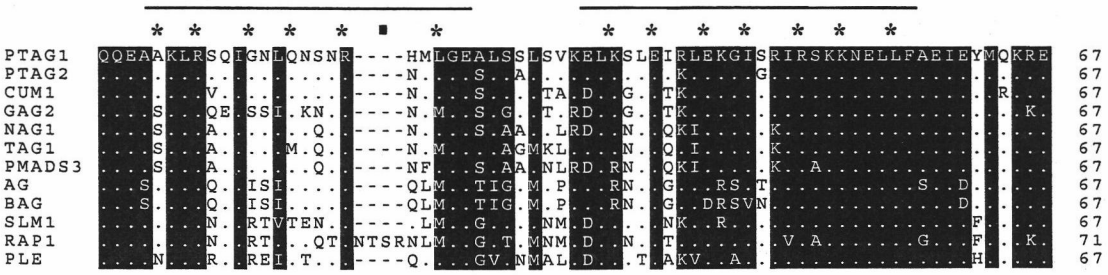


Figure 2.6. Alignment of PTAG1 with other members of the dicot C-group of the AG subfamily. Amino acids identical to PTAG1 residues are indicated by dots. Residues that are identical or similar in all sequences are shaded black. Two potential α helices of a coiled coil structure are indicated; the a and d positions of the heptad repeat sequence are denoted by asterisks. The square indicates one d position unique to RAP1.

K-box



C-terminal region

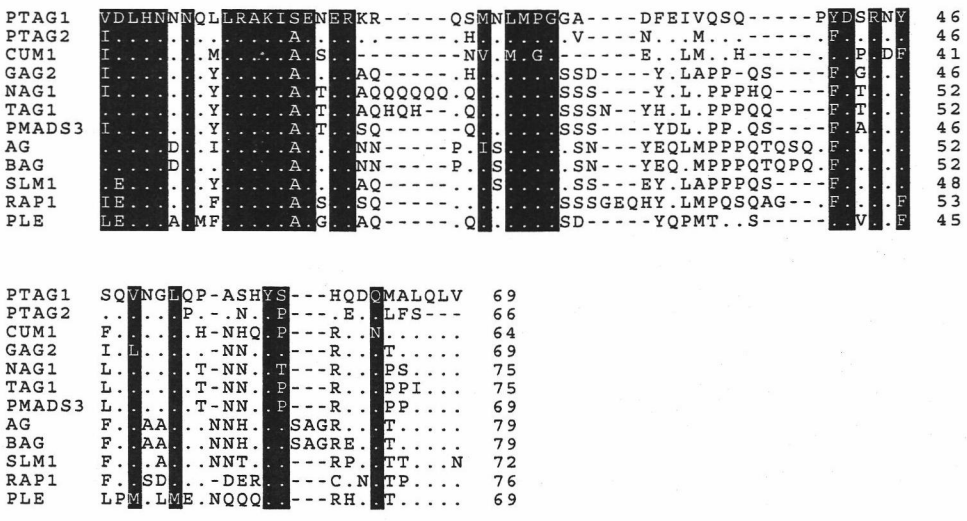


Figure 2.6, continued

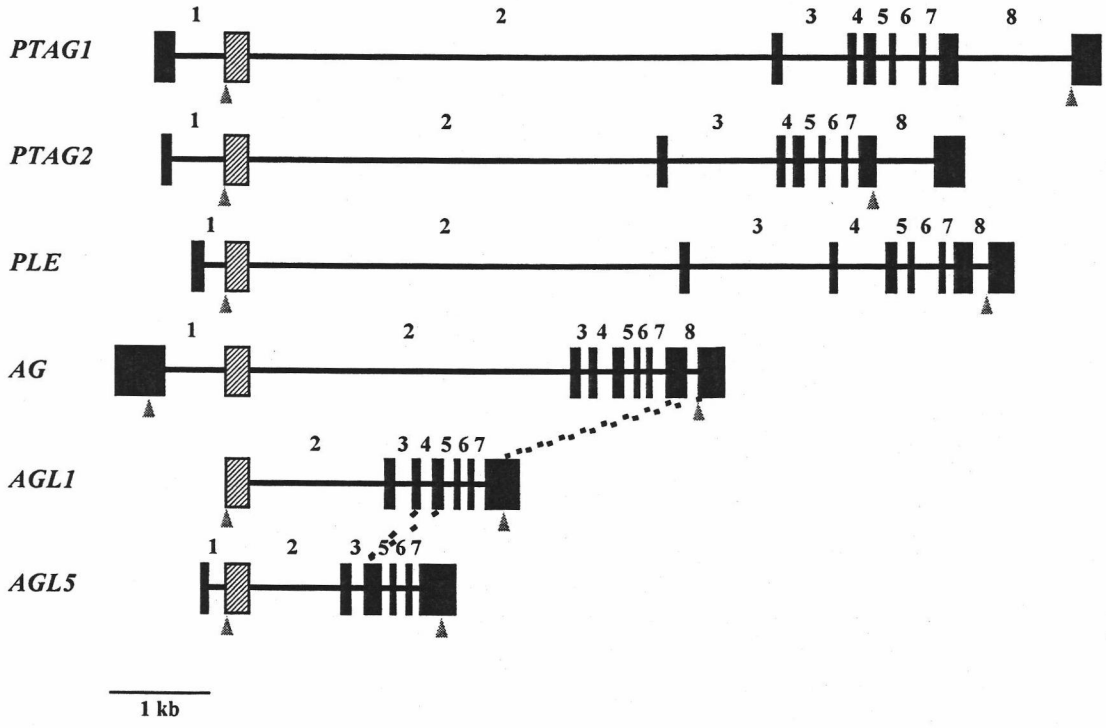


Figure 2.7. Gene Structures of *AG* subfamily members. Exons are depicted as boxes and introns by lines. The first nucleotides of the reported cDNA sequences were considered to be the start of first exon. The exon containing the MADS-box region is hatched. Homologous introns are numbered the same for all genes. The positions of translational start and stop codons on indicated by triangles. Dotted lines indicate intron differences between adjacent genes.

PTAG1 and *PTAG2* Expression

Gene-specific probes corresponding to portions of the *PTAG1* and *PTAG2* 3' UTRs were used to probe gel blots of RNA isolated from developing and mature female and male inflorescences and vegetative tissues (Figure 2.8). Both probes gave a very similar pattern of expression; transcripts were detected in immature and mature flowers from both male and female trees. Surprisingly, weak bands were also detected in all vegetative tissues tested. Vegetative buds, stems and leaves from the current year's growth collected from basal, non-flowering branches or root sprouts of mature trees contained the lowest levels of both transcripts. Compared to these vegetative tissues, transcript levels were approximately two-fold greater in vegetative buds that had undergone dormancy and were from flowering branches. *PTAG1* transcripts were about 1.2 to 1.4 kb, while *PTAG2* transcripts were in the 1.1 to 1.3 kb range. The size of both transcripts appeared to vary among the different tissues in a consistent manner for the two genes. In both cases, transcripts from vegetative tissues were shorter than the floral transcripts, and this size difference was consistently observed across separate RNA gel blot experiments. To investigate whether this size difference is due to alternate splicing, *PTAG1* transcripts were analyzed via RT-PCR with gene-specific primers encompassing all intron-exon junctions. Products of the expected size (940 bp) were synthesized from both vegetative and floral tissues (data not shown).

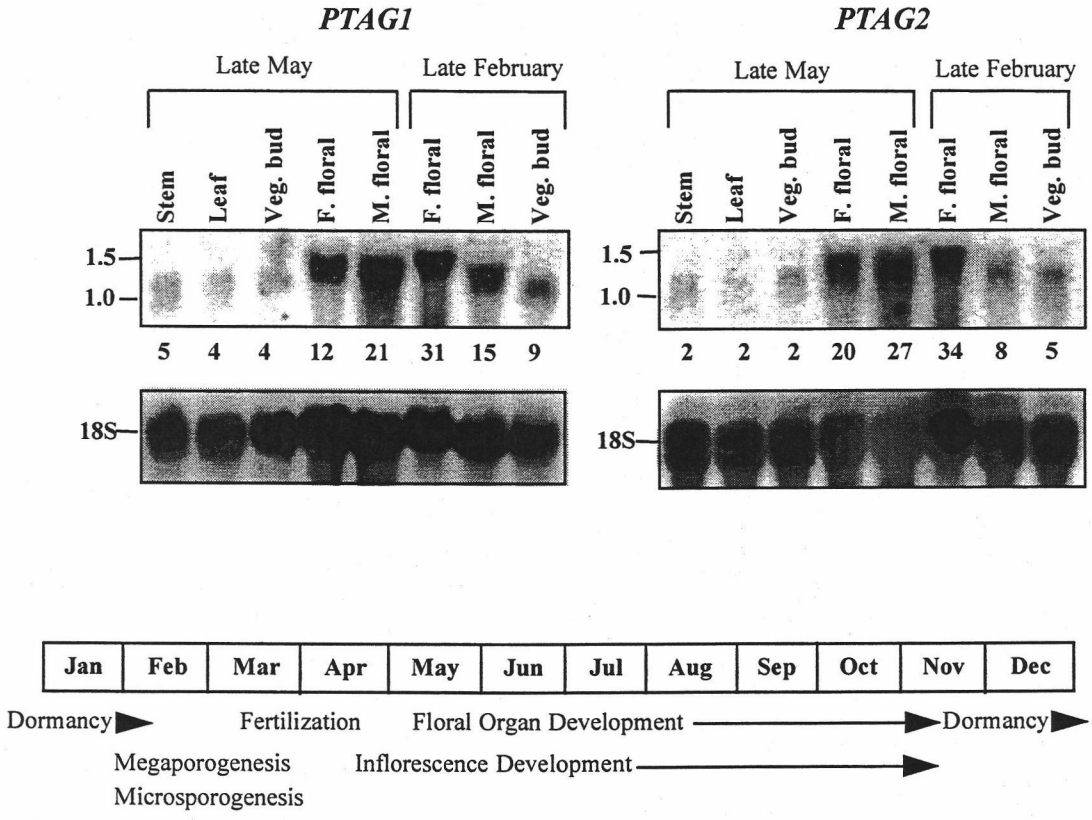


Figure 2.8. Expression of *PTAG1* and *PTAG2* in floral and vegetative tissues. Each lane contains 20 μg of total RNA. Molecular weight markers (kb) are indicated. RNA was prepared from female (F) and male (M) inflorescences at both immature (early spring) and mature (late spring) developmental stages. Late spring vegetative tissues were collected from non-flowering basal branches, while early spring overwintered buds were from flowering branches. Blots were stripped and re-probed with an 18S rDNA fragment. Numbers below the blots indicate relative expression levels of *PTAG1/2* after adjustment for equal loading, based on the 18S RNA controls. A typical annual flowering cycle for *P. trichocarpa* in the vicinity of Corvallis, OR is depicted below.

The expression patterns of *PTAG1/2* in various floral tissues is shown in Figure 2.9. Sections were hybridized to antisense or sense probes corresponding to a fragment of the *PTAG1* cDNA that lacked the MADS-box, but contained part of the I-region and the remaining 3' sequence. The antisense probe cross-reacts with *PTAG2*; however, studies using the very short, gene-specific probes (described above) produced hybridization patterns that were indistinguishable from the patterns produced using the longer probe, though the short probes did produce weaker signals (data not shown).

Floral meristems develop acropetally in poplar catkins in the axils of bracts (Figure 2.9 a). In male flowers, 30 to 50 stamen primordia arise centrifugally. Three to four carpel primordia form in female flowers; thirty to fifty anatropous ovules develop from a parietal placenta. *PTAG* expression was first detected in the central whorl of male and female meristems when the perianth cup primordia have barely begun to emerge, but before stamen and carpel primordia were visible (Figure 2.9 a,b,e,f). Expression continued in developing stamen and carpels (Figure 2.9 d,g,k). Transcripts were not detected in the outer whorl where the perianth cup forms at any developmental stage (e.g., Figure 2.9 h). As the stamen primordia differentiate, *PTAG* expression appeared to decrease and was confined to the filament, connective, and anther walls; expression levels were near the detection limit in mature male flowers, though a weak signal appeared to be associated with the connective and filament (data not shown). In mature female flowers, *PTAG* transcripts were evident in the placenta, but were not detected in the stigma (Figure 2.9 i). *PTAG* expression was seen in the ovule integuments, but not in the nucellus or embryo sac. A band of expression was also detected in the funiculus, directly below the ovule (Figure 2.9 m-p).

Figure 2.9. Expression of *PTAG* in developing male and female flowers. Longitudinal sections were hybridized to a probe synthesized from the *PTAG1* cDNA that cross-hybridized with *PTAG2*. **(a)** Immature male inflorescence; 10X objective (obj.). **(b)** Developing male floral meristems; stamen primordia have not initiated in the top meristem, but are forming in the lower two; 25X obj. **(c)** Negative control (sense probe), developing male flowers; 25X obj. **(d)** Male flower with stamen primordia; 50X obj. **(e)** Brightfield photo of developing female floral meristems; carpel primordia have just initiated in the bottom flower; 25X obj. **(f)** Darkfield photo of **(e)**. **(g)** and **(k)** Female flowers with developing carpels; 25X and 50X obj. **(h)** Female flower with carpel primordia just forming; 50X obj. **(i)** Negative control, developing female flower; 50X obj. **(I)** Mature female flower; 10X obj. **(j)** Negative control, mature female flower; 10X obj. **(m)** and **(n)** brightfield and darkfield photos of mature female flower, showing ovules; 10X obj.; rectangles frame region shown at higher magnification (50X obj.) in **(o)** and **(p)**. **b**, bract; **c**, carpel; **es**, embryo sac; **f**, funiculus; **fm**, floral meristem; **i**, integument; **o**, ovule; **p**, perianth cup; **pl**, placenta; **s**, stamen, **st**, stigma.

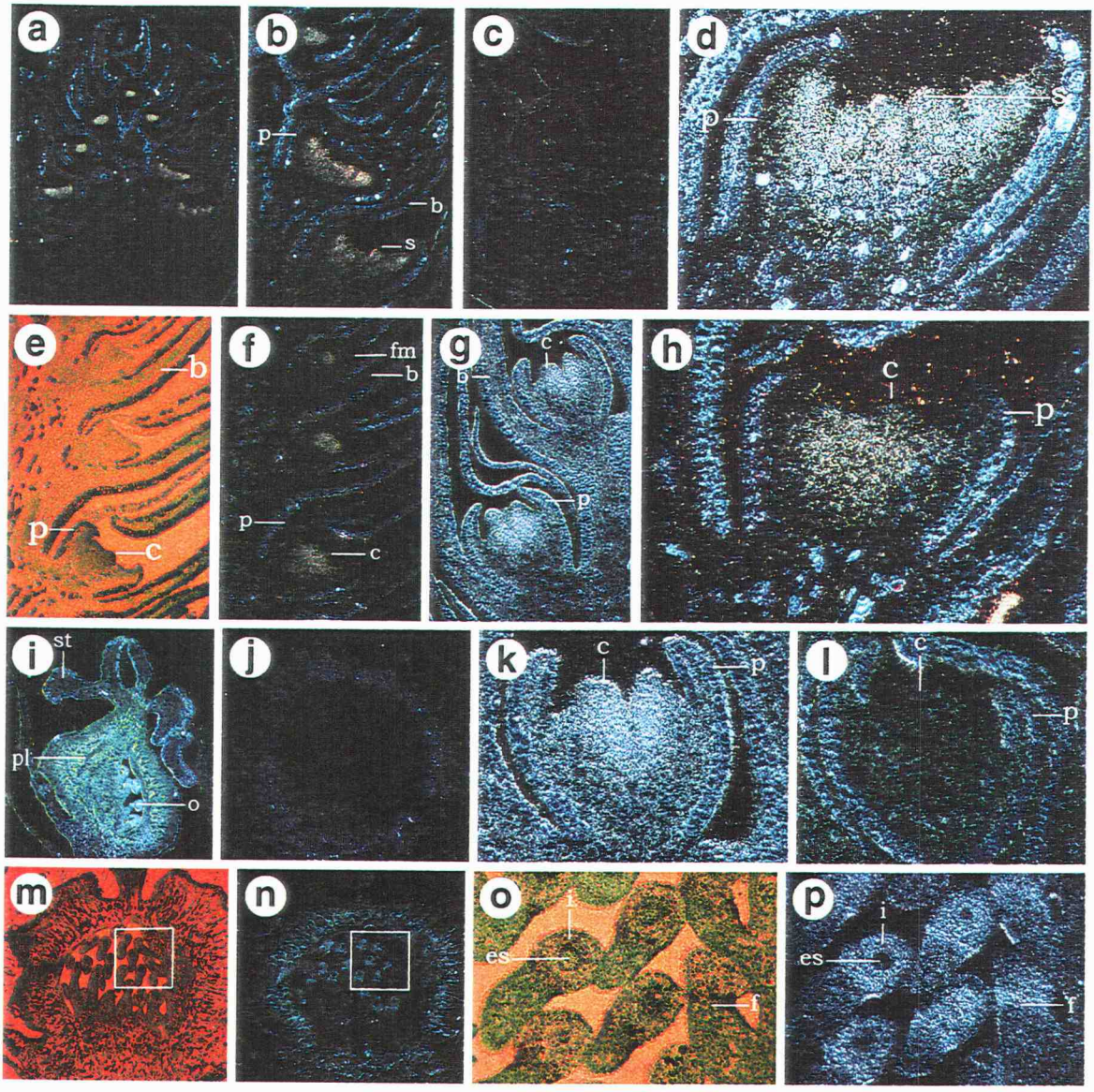


Figure 2.9, continued

DISCUSSION

Evolution of the *AG* Subfamily

Molecular clock estimates derived from dicot, monocot, and conifer gene comparisons suggest that the different floral homeotic lineages (Figure 2.3) began to diverge from one another around the time of the origin of land plants---450-500 mya (Purugganan 1997). Within the *AG* clade, duplications have led to a diversification of expression and functions, but most still have a role in development of reproductive structures. The subfamily contains genes expressed in both developing female and male reproductive primordia, as well as genes expressed exclusively in female tissues. Further, features present in an ancestral gene of this subfamily appear to have been conserved in both angiosperm and gymnosperm lineages. The *Picea* gene *DAL2* is a member of this clade, and expression and functional studies indicate a role in specifying reproductive structures (Tandre et al. 1995, 1997). Although analyses using the MIK region identified closely related pairs of genes within this subfamily, they have not strong statistical support concerning relationships within the group (e.g., Purugganan 1997).

Our phylogenetic analyses using the entire coding region discerned some larger clusters within the *AG* clade that show a general correspondence with spatial expression pattern (Figures 2.4 and 2.5). In particular, the dicot C-group includes all but one of the bisexually expressed dicot genes. The deduced proteins of the dicot C-group members are very similar, including regions of the highly variable C-terminal domain (Figure 2.6). The C-terminal half of the K-box, which includes the second predicted amphipathic helix, is strongly conserved among the dicot C-group. Study of *ag* alleles indicate an important role for this region in specifying a subset of *AG's* functions (Sieburth et al. 1995). The *ag-4* mutation results in a partial loss of this region of the K-box; flowers are indeterminate with carpels replaced by sepals, but unlike strong *ag* alleles, stamen develop in the third whorl. In the *AG-Met205* allele, the second to last

residue of the K-box, which is normally an arginine, is changed to a methionine. This allele specifies both stamens and carpels, but flowers exhibit a loss of determinacy.

In addition to *AG* and *PLE*, other members of the dicot C-group have also been shown to cause homeotic organ transformations indicative of C function (e.g., Pnueli et al. 1994, Kempin et al. 1993), further supporting that this group corresponds to C-function genes. However, a few exceptions indicate that evolutionary variances in function are not uncommon. In petunia, ectopic expression of *FBP6* (*P. hybrida*) affected petal size and shape but did not alter organ identity (Kater et al. 1998). In contrast, ectopic expression of *PMADS3* (*P. hybrida*) does result in partial C function organ transformations (Tsuchimoto et al. 1993). An additional petunia gene, *PAGL1* (*P. inflata*), is present in the dicot C-group; however, its expression pattern and function has not been reported. *PAGL1* and *FBP6* encode proteins that are 99% identical, indicating that they are likely to be orthologs.

An exception to the general correlation between phylogenetic relationship and expression pattern is the bisexually expressed cucumber gene *CUM10*. It forms a well-supported cluster with the petunia ovule-identity gene *FBP11* (Colombo et al. 1995), and two other dicot genes expressed in ovules, *FBP7* (Angenent et al. 1995) and *AGL11* (Rounsley et al. 1995). In contrast, the cucumber gene *CUM1* clearly belongs to the dicot C-group. *CUM1's* placement also correlates with its function; ectopic expression in transgenic petunia transformed petals into stamens and sepals into carpelloid structures, whereas ectopic *CUM10* expression resulted in only partial transformations of petals into antheroid structures (Kater et al. 1998). The placement of a third cucumber gene *CUS1* is problematic. The neighbor-joining analysis separates it from all other genes, and parsimony places it in the dicot C-group with only weak bootstrap support. The expression pattern of *CUS1* is also unusual. It is weakly expressed in female, but not male flowers, and is most strongly expressed in fruits and embryogenic callus (Filipecki et al. 1997).

PTAG1 and *PTAG2* appear to be a divergent pair of the dicot C-group (Figures 2.4, 2.5). Their coding regions share high nucleotide identity with each other (92%),

and intronic and near upstream sequences also exhibit significant homology (Table 2.2), consistent with a relatively recent duplication event. Whether the *PTAG* duplication coincides with the putatively recent origin of *Populus* (earliest fossil evidence is 58 mya) or Salicaceae, and is functionally related to their highly reduced floral morphology can only be determined by further evolutionary and functional studies within the genus and family (Eckenwalder 1996).

Although information is limited, intron-exon structure provides some possible insights into evolutionary relationships. In a previous analysis, Doyle (1994) proposed that an ancestral gene with 6 exons (numbers 2-7 in Figure 2.7) gives the most parsimonious explanation for the origin of introns in the plant MADS-box family. This requires both intron loss as well as intron gain to explain the known gene structures. Within the *AG* subfamily, complete or nearly complete gene structures for 8 genes have now been reported. The 4 genes of the dicot C-group, *AG, PLE, PTAG1* and *PTAG2*, share the same gene structure---all have 8 introns (Figure 2.7; Yanofsky et al 1990, Bradley et al. 1993). All other plant MADS-box genes for which gene structures have been reported have fewer than 8 introns. The female-specific *Arabidopsis* genes, *AGL1* and *AGL5*, group most closely to the dicot C-group and both lack intron 8 (Ma et al. 1991). In addition, intron 4 is absent in *AGL5*. The maize genes, *ZAG2* and *ZMM1*, which are well separated from the dicot-C group (Figures 2.4, 2.5), also lack intron 8 (Theissen et al. 1995). The presence or absence of intron 1 in *AGL1* and the maize genes is unclear. Thus, the phylogenetic analysis of the *AG* subfamily, together with what is know about intron-exon structures, do not contradict Doyle's hypothesis. In addition, intron 8 could possibly be uniquely characteristic of the dicot C-group. Genes belonging to the *AP1* clade (Figure 2.3) contain an intron 3' of intron 7; however, it does not appear to be at a position homologous to the position of intron 8 in the dicot C-group (Doyle 1994).

Also of note is the size of intron 2, which is particularly large (~3-5 kb) in the dicot C-group genes *AG, PLE, PTAG1* and *PTAG2*. Studies by Sieburth and Meyerowitz (1997) have shown that sequences in this intron are essential for normal *AG*

expression. As described above, *PTAG1* and *PTAG2* share regions of high homology within this intron and these genes are expressed in a nearly identical pattern (see next section). Sequence comparisons with the *AG* intron 2 (GenBank accession number AL021711), revealed numerous short stretches (~10 bp) of sequence identity and a few stretches of 30-40 bp with greater than 85% identity. Further dissection of the *AG* second intron may indicate whether any of these sequences correspond to conserved regulatory elements. Also of evolutionary interest is whether the presence of regulatory sequences in this intron is characteristic of the entire *AG* subfamily or only a subset.

PTAG1 and *PTAG2* appear to be Redundant C Function Genes

Various levels of genetic redundancy are common among regulatory genes of plants and animals (reviewed in Pickett and Meeks-Wagner 1995, Cooke et al. 1997). The high degree of amino acid identity (89%) as well as their very similar expression patterns suggests that *PTAG1* and *PTAG2* have largely overlapping functions. However, other close paralogs belonging to the plant MADS-box gene family exhibit a degree of functional divergence. For example, the proteins encoded by the *Arabidopsis* genes *AP1* and *CAULIFLOWER* (*CAL*) are 81% identical (Mandel et al. 1992, Kempin et al. 1995). Based on single and double mutant analyses, *CAL* activity appears to be completely redundant with *AP1* activity, but *AP1* has additional functions that are not redundant with *CAL* (Bowman et al. 1993). A restriction fragment polymorphism was detected for *PTAG2* between a female and male tree (Figure 2.2). Because we studied wild trees, heterozygosity is a likely explanation for this difference; however, a DNA modification that prevents restriction is also possible. Analysis of additional trees is needed to determine whether this polymorphism is associated with tree gender.

That *PTAG1* and *PTAG2* are involved in specifying C function is suggested by their floral expression patterns and phylogenetic analyses. Like *AG*, *PLE* and other genes shown to specify C functions (e.g., Yanofsky et al. 1990, Bradley et al. 1993, Pnueli et al. 1994), *PTAG1/2* are not expressed in the whorl giving rise to perianth

structure, but are expressed in the whorls giving rise to stamen and carpels, both before these primordia begin to form as well as in the developing stamen and carpels. Furthermore, their expression pattern in mature female flowers is similar to the patterns reported for *AG* (Bowman et al. 1991, Reiser et al. 1995). *PTAG1/2* transcripts are detected in the placenta, funiculus, and ovule integuments, but not in the nucellus or embryo sac. In contrast to *AG*, *PTAG* expression was not seen in the stigma.

The differentiation between members of the *AG* subfamily provided by our phylogenetic analyses using entire coding regions generally correlates with expression patterns and with what is known about gene functions (Figures 2.4 and 2.5). Thus, the grouping of *PTAG1/2* with similarly expressed dicot C-class genes is consistent with *PTAG1/2* specifying C function in poplars. However, the genes in this group may nonetheless not be the functional equivalents of *AG*. As examples in maize and petunia show, the presence of a close paralog can be indicative of functional diversification. The petunia genes *FBP6* and *PMADS3* both belong to the dicot C-group and are expressed similarly; yet, ectopic expression studies reveal that their activities are not identical (Kater et al. 1998). The monocot *AG* homologs, *ZAG1* and *ZMM2*, exhibit quantitative differences in expression---*ZAG1* accumulates more in developing carpels than stamens, while *ZMM2* exhibits the reverse pattern. Further, *ZAG1* is necessary to specify only a subset of the C functions, suggesting that these genes have nonidentical, but overlapping functions (Mena et al. 1996).

A similar degree of functional divergence appears unlikely for the poplar paralogs. The PTAG proteins are more similar (89% identity) than the paralogs from petunia (70%) and maize (63%). Although we can not rule out the possibility that there are additional poplar genes which belong to the dicot C-group, this does not appear to be the case. When Southern blots were probed at high stringency with a fragment of the *PTAG1* cDNA that lacked the MADS-box, but contained most of the remaining 3' sequences, both *PTAG1* and *PTAG2* gene fragments were detected (data not shown). When stringency was reduced (42°C), no other bands appeared, suggesting that additional, closely related paralogous genes are not present. Despite their high

similarity, it is still possible that *PTAG1* and *PTAG2* have evolved or are evolving some differences in function. The different domains of both genes are diverging at the rates (Table 2.1) expected for plant MADS-box genes (Purugganan et al. 1995), suggesting that both are under similar forms of functional selection. Functional differences among these genes will only be discerned by ectopic expression or gene-specific suppression via poplar transformation.

Significance of Vegetative Expression?

Unlike *AG, PLE* and putative orthologs from other species, we detected *PTAG* transcripts in vegetative tissues, though at lower levels than in floral tissues (Figure 2.8). Recently, weak vegetative expression of the tomato homolog *TAG1* was also reported, and increased levels of the *TAG1* transcript were associated with the development of fruit-like features on sepals cultured *in vitro* (Ishida et al. 1998). Though *PTAG* transcripts were detected, it is possible that they were not translated. Even if protein was produced, factors that suppress PTAG function may be present or necessary accessory factors may be absent in vegetative tissues. In addition, the level of *AG* RNA is important to function. The severity of the phenotype in *35S-AG*, *35S-antisense AG*, and *AP3-AG* transgenic plants correlates with the level of *AG* RNA expression, and the organ identity and indeterminacy functions of *AG* require different amounts of gene product (Mizukami and Ma 1992, 1995, Jack et al. 1997). Thus, the amount of *PTAG* RNA present in vegetative tissues may not be sufficient to specify floral organ identity or indeterminacy. While the observed vegetative expression may not have direct functional significance, it may be significant from the viewpoint of understanding regulatory interactions and how these may vary between trees and herbaceous annuals.

Regulation of *AG* is complex. The floral meristem identity genes *LFY* and *AP1* are positive regulators of *AG*, while *AP2* and *LUG* act to spatially regulate *AG* expression within the developing flower (reviewed in Yanofsky 1995). Interestingly,

CLF is necessary to prevent *AG* expression in vegetative tissues not only during the reproductive phase of the plant, but also during the vegetative phase (Goodrich et al. 1997). *CLF* is homologous to the *Drosophila* gene *Enhancer of zeste* (*E(z)*), a member of the *Polycomb group* (*PcG*). The *PcG* genes act to maintain transcriptional repression by modifying chromatin structure. The observation that the effects of floral mutations are attenuated acropetally led to the proposal that expression of downstream floral genes may rely on both activation by genes such as *LFY* and *AP1,* and on progressive derepression during the plant's life cycle (Weigel 1995). For example, late arising lateral shoots of *lfy ap1* double mutants are less shootlike and have some carpelloid features, and *AG* is expressed in these mutant shoots. There are also some indications that the flowering time genes, *CO* and *EMF1*, affect *AG* expression, though the interaction may be indirect (Simon et al. 1996, Chen et al. 1997).

In addition, *35S-AG* transgenics demonstrated that *AG* is sufficient to cause early-flowering, apparently both by accelerating the change from a juvenile vegetative phase to a reproductively competent adult vegetative phase, and by shortening the adult phase (Mizukami and Ma 1997). Because *AG* expression is not detected in vegetative tissues nor in the initial stages of the floral meristem, and *ag* mutants flower at the same time as wild-type plants, *AG* is not likely to have a role in floral induction. Mizukami and Ma (1997) proposed that *AG* may provide a floral promotion activity within floral meristems to maintain flower development. According to this model, meristem commitment to floral fate is acquired gradually, owing in part to the sequential action of *LFY*, then *AP1*, and finally *AG* (Ma 1998).

While the mechanisms underlying floral induction and development may be largely conserved between annuals and trees, substantial differences are also expected. Among the characteristics of *Populus* that suggest this are an extended vegetative phase, the year-long floral development cycle, and the persistence of apical vegetative meristems (i.e., they never convert to inflorescence meristems). *PTAG* vegetative expression may simply be the result of less stringent repression controls. Alternatively, it might be correlated with derepression that occurs as a tree approaches reproductive

competency and flowering. In addition, branches within a tree might exhibit varying degrees of derepression; lower branches typically do not produce flowers. It may also be possible that at low levels *PTAG* does contribute to floral induction, is part of an induction pathway, or maintains reproductive competency rather than only maintain floral fate within floral meristems as was suggested for *AG* (Mizukami and Ma 1997)

We observed that *PTAG* expression levels in vegetative buds change with season and/or location on the tree. Vegetative buds that had overwintered were collected from flowering branches in early spring as inflorescences initiated in the previous year were nearing anthesis. Expression levels in these buds were two times the levels observed in newly initiated vegetative buds from non-flowering branches collected in late spring. Although the buds that had undergone dormancy would eventually give rise to shoots with inflorescence buds as well as vegetative buds, it is unlikely that inflorescence meristems had initiated in the overwintered buds at the time of collection. Certainly, floral meristems were not present. Further analysis of vegetative expression at different locations within the tree's crown and at various times during a year, analysis of both juvenile and mature trees, and ectopic expression in transgenic trees should identify whether vegetative expression of *PTAG* is associated with reproductive competence.

In addition, vegetative transcripts are shorter (~150-200 bp) than floral transcripts (Figure 2.8), indicating differences in transcription or RNA processing. RT-PCR experiments suggest that this size difference is not due to alternate splicing (data not shown). Differences in the site of transcriptional initiation or the in polyA site could account for the shorter vegetative transcripts, and both could affect translational efficiency and mRNA stability. Thus, determining the cause of the different transcript sizes should provide insight into the mechanisms regulating *PTAG* expression. Also of interest is whether *AG* homologs from other species exhibit this type of regulation. As discussed above, factors that repress *AG* expression in vegetative tissues as well as positive regulators of *AG* expression in floral meristems have been identified, but the molecular mechanisms underlying their regulatory functions are unknown.

Chapter III: Summary

Conclusions

The main objective of this project was to isolate the *P. trichocarpa* homolog of the *Arabidopsis* C function organ identity gene *AG* and the genomic sequences necessary to direct its proper expression. Although our analyses support that this was accomplished, there were a few suprising results. First, *Populus* contains two putative C function genes. The coding regions of *PTAG1* and *PTAG2* share 92% nucleotide identity and 89% amino acid identity. Further, homologous introns and upstream sequences share significant identity. Southern analysis showed that both genes are present as single copies in female and male trees. Although the high similarity of these gene loci indicates a relatively recent duplication event, preliminary analyses of gene-specific SSR markers in a mapping pedigree places these genes on separate chromosomes.

Phylogenetic analysis of a representative subset of the known plant MADS-box genes clearly places the *PTAG* genes in the *AG* clade. Further analysis of the *AG* subfamily differentiated groups that were generally correlated with expression pattern and function. The *PTAG* genes formed a group with other dicot genes exhibiting a similar expression pattern, and several of these have also been shown to specify C function. Duplications and diversification in the *AG* lineage appears to have resulted in a group of genes involved in different aspects of reproductive organ development; the known genes are either expressed in both female and male reproductive organs or only in female tissues. An interesting question is whether the ancestral gene of this lineage was female-specific or expressed bisexually. Our analyses only weakly (less than 50% bootstrap support) suggest that the bisexually expressed conifer gene *DAL2* is basal to all the angiosperm genes of the *AG* family. This raises the possibility that additional conifer genes belonging to this group may exist, and that elaboration of this lineage into

bisexual and female-specific genes may have begun before the divergence of angiosperms and gymnosperms.

The expression profiles of *PTAG1* and *PTAG2* are very similar and their cell-specific floral expression patterns strongly resemble those of *AG*, *PLE* and C function genes from other species. *In situ* hybridization studies showed that *PTAG1/2* were expressed in the central whorl of both female and male floral meristems before the appearance of reproductive primordia. No transcripts were detected in the outer whorl at any developmental stage. Expression continued in both developing stamens and carpels. Transcripts levels were lower in mature male flowers compared to developing male flowers, while expression levels were highest in mature female flowers. In mature male flowers, only very weak signals associated with the filament and connective were detected. In mature female flowers, *PTAG1/2* were expressed in the placenta, the funiculi and the ovule integuments. Expression was not detected in the stigma, nucellus or embryo sac. Surprisingly, we also detected weak vegetative expression of both genes. Further, a two-fold difference was observed in expression between overwintered buds collected from flowering branches in early spring and newly initiated buds from non-flowering branches collected in late spring. This hints that all vegetative expression may not be simply the result of a general, basal level of transcription. In addition, vegetative transcripts are shorter than floral transcripts, indicating that the site of transcriptional initiation or mRNA processing differs between tissues.

These results suggest that inhibiting the *PTAG* genes is likely to be an effective way to engineer reproductive sterility in poplars. Because the functions of *PTAG1* and *PTAG2* may be largely redundant, inhibiting both will probably be necessary. However, this should not be difficult due to the high levels of both nucleotide and amino acid identity. Thus, if the *PTAG1* cDNA is used in antisense, cosuppression or DNM constructs, introduction of any of these constructs into poplar may inhibit both *PTAG1* and *PTAG2* in at least some transgenic lines. At the onset of this project, we had predicted that the promoter from the poplar *AG* homolog fused to a cytotoxin gene could also be used to engineer sterility. However, the unexpected weak vegetative

expression makes use of the *PTAG* promoters more problematic. Although use of the barnase-barstar system, or a less potent cytotoxin, may provide a means to mitigate the potential detrimental effects caused by weak vegetative expression of a cytotoxin, this strategy is largely untested. Alternatively, the *cis* regulatory elements necessary for the vegetative expression of *PTAG1/2* may be separable from the elements required for floral expression. Thus, vegetative expression could possibly be eliminated by modifying the promoters.

SUGGESTIONS FOR FUTURE RESEARCH

The *PTAG* sequences were isolated with the ultimate goal of using them to engineer reproductive sterility. Thus, I recommend pursuing experiments specifically directed to this goal first, though I also suggest additional experiments. For engineering sterility using *PTAG* sequences, I advise pursuing two methods. One is the DNM approach for inhibition of *PTAG*. I recommend this method over antisense and cosuppression for a few reasons. Due to the long juvenile phase of poplars, verification of a sterility construct's effectiveness requires several years. Further, the effectiveness of a particular construct will vary between transgenic lines. Once a particular DNA alteration has been shown to result in a DNM, the effectiveness of the construct depends on whether or not it is expressed at an adequate level. Thus, the transgene expression level may be a good indicator of whether or not a particular transformant will be sterile. In contrast, there is currently no proven method for indicating whether an antisense or cosuppression construct is likely to be working until the endogenous gene is expressed (i.e. at the time of flowering). The absence of transgene expression during the juvenile phase could be due to chromosomal position effects and thus, is not necessarily a reliable indicator that the endogenous gene will be cosuppressed.

In addition, a higher proportion of transformants typically exhibit transgene expression rather than cosuppression, and genes are often only partially suppressed by antisense RNA. A high frequency of cosuppression was reported in tobacco plants

transformed with a replicating potato virus X cDNA, but whether this provides a means to consistently silence any transgene is unclear (Angell and Baulcombe 1997). Cosuppression can also exhibit somatic instability (reviewed in Baulcombe 1996). Similarly, transgene expression can be unstable---cosuppression of transgene and homologous endogenes occurred in some transgenic tobacco lines following transfer from the greenhouse to the field (Brandle et al. 1995). However, in the case of a DNM transgene, gene silencing may achieve the same, desired result as transgene expression. Silencing of a previously highly expressed DNM transgene is most likely to be caused by cosuppression rather than position effects; thus, the homologous endogene may also be silenced.

I also suggest pursuing the use of a *PTAG* promoter for directing expression of a cytotoxin. Although vegetative expression makes this more difficult, the floral expression pattern of *PTAG* is ideally suited for engineering both female and male sterility. Moreover, the ablation approach offers some advantages over inhibition approaches. In particular, ablation circumvents the problem of genetic redundancy. The difference in vegetative and floral transcript size hints that floral and vegetative expression may be under different regulatory controls. Determing the cause of this difference in transcript size will inform on *PTAG* regulatory mechanisms and may suggest a way to modify the *PTAG* sequence in order to eliminate vegetative expression. For example, a difference in the transcriptional start site may account for the shorter vegetative transcripts. If this were the case, alternating or deleting the vegetative initiation site could eliminate vegetative expression without affecting floral expression.

Additional *PTAG* promoter analysis could identify other separable regulatory motifs and inform on the conservation of motifs and regulatory interactions between species. Knowing the sequence from two close paralogs that are expressed in a very similar pattern would certainly facilitate this analysis. In addition, the genomic sequence of the *AG* locus is now available. The majority of regulatory regions are located in the second intron of *AG* (Sieburth and Meyerowitz 1997) and further dissection of the regulatory elements of this intron is being pursued, providing a guide

for the analysis of *PTAG* promoter regions. The extended juvenile phase of poplars hinders analysis; however, sequences necessary to inhibit vegetative expression may be identified in a more reasonable amount of time. In *clf* mutants, *AG* is expressed in leaves during the vegetative phase as well as the reproductive phase of the plant, and sequences necessary for *CLF* regulation are located in the second intron (Goodrich et al. 1996, Sieburth and Meyerowitz 1997). Regulation of *AG* is complex, making it especially interesting to determine the degree to which regulatory networks are conserved between a tree and a herbaceous annual. *PTAG* promoters could be analyzed in both *Populus* and *Arabidopsis*. In addition, introducing the *PTAG1/2* gene locus into an *ag* mutant may be an interesting way to determine both functional and regulatory equivalency.

Inhibition of the *PTAG* genes via DNM transgenes will also inform on *PTAG* function. As mentioned above, it may not be simple to inhibit just one *PTAG*, potentially making it difficult to determine if they exhibit functional differences. However, they can be ectopically expressed independently. Because constitutive expression of *AG* in *Arabidopsis* causes not only floral organ transformation, but also early-flowering (Mizukami and Ma 1997), it would be particularly interesting to test the effects of constitutive *PTAG1/2* expression in poplar. Overexpression of *PTFL* was not sufficient to induce flowering in poplar, whereas its *Arabidopsis* homolog *LFY* induced early-flowering in *Arabidopsis* and in at least some poplar genotypes (Rottmann et al. unpublished data, Weigel and Nilsson 1995). In addition, *PTAG1/2* could also be constitutively expressed in a model herbaceous plant such as *Arabidopsis* or tobacco. Their functional equivalency with *AG* could be further addressed by introducing *PTAG1/2* into *ag* mutants.

Further analysis of the vegetative expression patterns of *PTAG1/2* are needed to determine if expression levels exhibit any significant correlations with tree maturity, location within the tree, or season. Samples collected from seedlings, juvenile trees, mature trees and young vegetative propagules could be compared. In addition, vegetative tissues should be collected from different locations within the tree's crown

(e.g., flowering and non-flowering branches) and at various times during a year. Because vegetative expression is weak and differences may be small, a careful quantitative analysis will be the most informative; RT-PCR or RNase protection are likely to be the best strategies to use for this expression analysis.

STUDYING AND MANIPULATING FLOWERING IN TREES

Molecular genetic analysis of developmental processes in trees has barely begun. Certainly, many molecular mechanisms are common to all plants and are much easier to study in model herbaceous plants such as *Arabidopsis*. But the differences are of interest and can be crucial to the successful manipulation of a molecular pathway for commercial purposes. In addition, herbaceous plants are not suitable for the study of some processes, such as the genetic control of wood formation. Assuredly, studies in trees will continue to be based largely on research in model annuals in most cases. However, the emerging status of *Populus* and perhaps a few other trees, such as *Eucalyptus*, as model systems may lessen this dependency on model annuals as guides. The advances that come from a critical mass of scientists studying the same tree combined with technological developments will likely provide new opportunities to address questions directly in trees.

Flowering is an example of a developmental process that is largely conserved between all angiosperms, but important differences are likely between trees and herbaceous annuals. The cloning of flowering time genes in *Arabidopsis* is allowing investigation of the molecular mechanisms underlying floral induction. One approach to studying the genetic control of this process in *Populus* is to isolate homologs of the *Arabidopsis* genes as this study has done. However, the genes identified as important regulators in *Arabidopsis* may not represent the genes with the most significant roles in the transition to reproductive competency and flowering in trees. The demonstration that constitutive expression of *LFY* in an aspen hybrid induced precocious flowering

(Weigel and Nilsson 1995) caused much excitement in the forestry community; however, it became apparent that manipulation of flowering in a usable way in trees will not be as straightforward as initially indicated. The main point is that manipulation of flowering in trees requires a better understanding of this process in trees.

Such investigations in trees may be greatly facilitated by the development of genomic methods. By providing a means to rapidly quantitate the expression level of many genes in parallel, including weakly expressed ones, cDNA microarray technology provides a powerful method for identifying differentially expressed genes (reviewed in Schena 1996, Ramsey 1998). Microarrays of *Arabidopsis* expressed sequence tags (ESTs) have been constructed and more are being developed. A heterologous approach, whereby poplar mRNAs are hybridized to *Arabidopsis* microarrays may be an effective way to identify key genes. Further, the development of poplar ESTs is underway. In particular, an EST bank from tissues relevant to wood formation is expected to be available in the near future. Combined with developments in transformation/ regeneration systems in trees, genomic technologies may allow major advances in both our understanding of developmental and physiological processes in trees as well as our ability to manipulate these processes in trees.

Bibliography

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

Ainsworth, C., Parker, J., and Buchanan-Wollaston, V. 1998. Sex determination in plants. In, Current Topics in Developmental Biology, Volume 38. R.A. Pederson and G.P. Schatten, eds. Academic Press, San Diego, CA. pp. 169-223.

Amasino, R.M. 1996. Control of flowering time in plants. Curr. Opin. Gen. Dev. 6: 480-487.

Angell, S.D., and Baulcombe, D.C. 1997. Consistent gene silencing in transgenic plants expressin a replicating potato virus X RNA. EMBO J. 16: 3675-3684.

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

Barker, R.G., Wann, S.R., and Kaphammer, B.J. 1997. Biological sciences in the pulp and paper industry---the industry's viewpoint. Tappi Proceedings, 1997 Biological Sciences Symposium, San Francisco, CA, Oct. 19-23.

Baulcombe, D.C. 1996. RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. Plant Mol. Bio. 32: 79-88.

Bessy, C.E. 1904. The number and weight of cottonwood seeds. Science 20: 118-119.

Blazquez, M.A. 1997. Illuminating flowers: *CONSTANS* induces *LEAFY* expression. BioEssays 19: 277-279.

Boerjan, W., Baucher, M., Chabbert, B., Petit-Conil, M., Leple, J.-C., Pilate, G., Cornu, D., Monties, B., Inze, D., Van Doorsselaere, J., Jouanin, L., Van Montagu, M., Tsai, C.-j, Podila, G.K., Joshi, C.P., and Chiang, V.L. 1997. Genetic modification of lignin biosynthesis in quaking aspen and poplar. Chapter 26. In, Micropropagation, genetic engineering, and molecular biology of *Populus*. N.B. Klopfenstein, Y.W. Chun, M.-S. Kim, M.R. Ahuja, eds. M.C. Dillon, R.C. Carman, L.G. Eskew, tech. eds. Gen Tech. Rep. RM-GTR-297. Fort Collins, CO: USDA, Forest Service, Rocky Mountain Research Station. pp. 193-205

Boes, T.K. and Strauss, S.H.. 1994. Floral phenology and morphology of black cottonwood, *Populus trichocarpa* (Salicaceae). Am. J. Bot. 81: 5 62-567.

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

Bowman, J.L., Sakai, H., Jack, T., Weigel, D., Mayer, U., and Meyerowitz, E.M. 1992. *SUPERMAN*, a regulator of floral homeotic genes in *Arabidopsis*. Development 114: 599-615.

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

Braatne, J.H., Rood, S.B., and Heilman, P.E. 1996. Life history, ecology, and conservation of riparian cottonwoods in North America. In, Biology of *Populus* and its implications for management and conservation. Part I, Chapter 3. R.F. Stettler, H.D. Bradshaw, Jr., P.E. Heilman, and T.M. Hinckley, eds. NRC Research Press, National Research Council of Canada, Ottawa, ON. pp. 57-85.

Bradley, D., Carpenter, R., Copsey, L., Vincent, C., Rothstein, S., and Coen, E. 1996. Control of inflorescence architecture in *Antirrhinum*. Nature 379: 791-797.

Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R., and Coen, E. 1997. Inflorescence commitment and architecture in *Arabidopsis*. Science 275: 80-82.

Bradley, D., Vincent, C., Carpenter, R., and Coen, E. 1996. Pathways for inflorescence and floral induction in *Antirrhinum*. Development 122: 1535-1544.

Bradshaw, Jr., H.D. 1996. Molecular genetics of *Populus*. In, Biology of *Populus* and its implications for management and conservation. Part I, Chapter 8. R.F. Stettler, H.D. Bradshaw, Jr., P.E. Heilman, and T.M. Hinckley, eds. NRC Research Press, National Research Council of Canada, Ottawa, ON. pp. 183-199.

Bradshaw, H.D. 1998. Case history in genetics of long-lived plants: molecular approaches to domestication of a fast-growing forest tree: *Populus*. In, Molecular Dissection of Complex Traits. A.H. Paterson, ed. CRC Press, NY. pp. 219-228.

Bradshaw, Jr., H.D. and R.F. Stettler. 1995. Molecular genetics of growth and development in *Populus* IV. Mapping QTLs with large effects on growth, form and phenology traits in a forest tree. Genetics 139: 963-973.

Bradshaw Jr., H.D., Villar, M., Watson, B.D., Otto, K.G., Stewart, S., and Stettler, R.F. 1994. Molecular genetics of growth and development in *Populus* III. A genetic

linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. Theor. Appl. Genet 89: 167-178.

Brandle, J.E., McHugh, S.G., James, L., Labbe, H., and Miki, B.L. 1995. Instability of transgene expression in field grown tobacco carrying the csr1-1 gene for sulfonylurea herbicide resistance. Bio/Technology 13: 994-998.

Breathnach, R., and Chambon, P. 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. 50: 349-383.

Cervera, M.T., Gusmao, J., Steenackers, M., Peleman, J., Storme, V., Vanden Broeck, A., Van Montagu, M. and Boerjan, W. 1996. Identification of AFLP molecular markers for resistance against *Melampsora larci-populina* in *Populus*. Theor. Appl. Genet. 93: 733-737.

Cervera, M.T., Villar, M., Faivre-Rampant, P. Goue, M.-C., Van Montagu, M., and Bojeran, W. 1997. Applications of molecular marker technologies in *Populus* breeding. Chapter 14. In, Micropropagation, genetic engineering, and molecular biology of *Populus*. N.B. Klopfenstein, Y.W. Chun, M.-S. Kim, M.R. Ahuja, eds. M.C. Dillon, R.C. Carman, L.G. Eskew, tech. eds. Gen Tech. Rep. RM-GTR-297. Fort Collins, CO: USDA, Forest Service, Rocky Mountain Research Station. pp. 101-115.

Chen, L., Cheng, J.-C., Castle, L., and Sung, Z.R. 1997. *EMF* genes regulate *Arabidopsis* inflorescence development. Plant Cell 9: 2011-2024.

Chung, Y.-Y., Kim, S.-R., Finkel, D., Yanofsky, M.F., and An, G. 1994. Early flowering and reduced apical dominance result from ectopic expression of a rice MADS-box gene. Plant Mol. Bio. 26: 657-665.

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

Cooke, J., Nowak, M.A., Boerlijst, M. and Maynard-Smith, J. 1997. Evolutionary origins and maintenance of redundant gene expression during metazoan development. Trends Genet. 13: 360-364.

Cronquist, A. 1988. The Evolution and Classification of Flowering Plants, Second Edition. The New York Botanical Garden, Bronx, NY.

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

Day, C.D., Galgoci, B.F.C., and Irish, V.F. 1995. Genetic ablation of petal and stamen primordia to elucidate cell interactions during floral development. Development 121: 2887-2895.

Day, C.D. and Irish, V.F. 1997. Cell ablation and the analysis of plant development. Trends Plant Sci. 2: 106-111.

Dean, C. and Schmidt, R. 1995. Plant genomes: a current molecular description. Annu. Rev. Plant Physiol. Plant Mol. Biol. 45: 395-418.

Dickmann, D.I. and Stuart, K.W. 1983. The Culture of Poplars in Eastern North America. University Publications, Michigan State University, East Lansing.

Dinus, R.J. and Tuskan, G.A.. 1997. Integration of molecular and classical genetics: a synergistic approach to tree improvement. Chapter 29. In, Micropropagation, genetic engineering, and molecular biology of *Populus*. N.B. Klopfenstein, Y.W. Chun, M.-S. Kim, M.R. Ahuja, eds. M.C. Dillon, R.C. Carman, L.G. Eskew, tech. eds. Gen Tech. Rep. RM-GTR-297. Fort Collins, CO: USDA, Forest Service, Rocky Mountain Research Station. pp. 220-235.

Dix, M.E., Klopfenstein, N.B., Zhang, J.-W., Workman, S.W., and Kim, M.-S. 1997. Potential use of *Populus* for phytoremidiation of environmental pollution in riparian zones. Chapter 27. In, Micropropagation, genetic engineering, and molecular biology of *Populus*. N.B. Klopfenstein, Y.W. Chun, M.-S. Kim, M.R. Ahuja, eds. M.C. Dillon, R.C. Carman, L.G. Eskew, tech. eds. Gen Tech. Rep. RM-GTR-297. Fort Collins, CO: USDA, Forest Service, Rocky Mountain Research Station. pp. 206-211.

Donahue, R.A., T.D. Davis, C.H. Michler, D.E. Reimenschneider, D.R. Carter, P.E. Marquardt, N. Sankhla, B.E. Haissig and J.G. Isebrands. 1994. Growth, photsynthesis and herbicide tolerance of genetically modified hybrid poplar. Can. J. For. Res. 24: 2377-2383.

D'Ovidio, R., Mugnozza, G.S., and Tansarella, O.A. 1991. rDNA cloning and rapid hybrid identification in the *Populus* spp. (Salicaceae) Plant Syst. Evol. 177: 165-174.

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

Drews, G.N., J.L. Bowman and E.M. Meyerowitz. 1991. Negative regulation of the *Arabidopsis* homeotic gene *AGAMOUS* by the *APETALA2* product. Cell 65: 991-1002.

Duke, S.O., A.L. Christy, F.D. Hess, and J.S. Holt. 1991. Herbicide-resistant crops. Council for Agricultural Science and Technology: Ames, Iowa.

Eckenwalder, J.E. 1996. Systematics and evolution of *Populus*. In, Biology of *Populus* and its implications for management and conservation. Part I, Chapter 1. R.F. Stettler, H.D. Bradshaw, Jr., P.E. Heilman, and T.M. Hinckley, eds. NRC Research Press, National Research Council of Canada, Ottawa, ON. pp. 7-32.

Evans, L.T. 1980. The natural history of crop yield. Am. Sci. 68: 388- 397.

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

Felsenstein, J. 1989. PHYLIP -- Phylogeny Inference Package (Version 3.2). Cladistics 5: 164-166.

Filipecki, M.K., Sommer, H., and Malepszy, S. 1997. The MADS-box gene *CUS1* is expressed during cucumber somatic embryogenesis. Plant Sci. 125: 63-74.

Finnegan, E.J., Peacock, W.J., and Dennis, E.S. 1996. Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. Proc. Natl. Acad. Sci. 93: 8449-8454.

Firulli, A.B. and Olson, E.N. 1997. Modular regulation of muscle gene transcription: a mechanism for muscle cell diversity. Trends Genet. 13: 364-369.

Fisher, M.J. 1928. The morphology and anatomy of the flowers of the Salicaceae II. J. Bot. 15: 372-394.

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

Food and Agricultural Organization of the United Nations (FAO). 1997. State of the World's Forests, 1997. Rome, Italy.

Gladstone, W.T. and Ledig, F.T. 1990. Reducing pressure on natural forests through high-yield forestry. For. Ecol. Manage. 35: 69-78.

Goldburg, R., J. 1992. Environmental concerns with the development of herbicide tolerant plants. Weed Techol. 6: 647-652.

Goldman, M. H., Goldburg, R.B., and Mariani, C. 1994. Female sterile tobacco plants are produced by stigma-specific cell ablation. EMBO J. 13:2976-2984.

Goodrich, J., Puangsomlee, Martin, M., Long, D., Meyerowitz, E.M., and Coupland, G. 1997. A polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. Nature 386: 44-51.

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

Grant, S., Houben, A., Vyskot, B., Siroky, J., Pan, W.-H., Macas, J., and Saedler, H. 1994. Genetics of sex determination in flowering plants. Devel. Genet. 15: 214-230.

Guo, H., Yang, H., Mockler, T., and Lin, C. 1998. Regulation of flowering time by *Arabidopsis* photoreceptors. Science 279: 1360-1363.

Gustafson-Brown, C., Savidge, B., and Yanofsky, M. 1994. Regulation of the *Arabidopsis* floral homeotic gene *APETALA1*. Cell 76: 131-143.

Haines, R. 1994. Biotechnology in forest tree improvement. FAO Forestry Paper 118, FAO, Rome, Italy.

Hammond-Kosack, K.E. and Jones, J.D.G. 1997. Plant disease resistance genes. Annu. Rev. Plant Physiol. Plant Mol. Biol. 48: 575-607.

Han, K.-H., Gordon, M.P., and Strauss, S.H. 1996. Cellular and molecular biology of Agrobacterium-mediated transformation of plants and its application to genetic transformation of *Populus*. In, Biology of *Populus* and its implications for management and conservation. Part I, Chapter 9. R.F. Stettler, H.D. Bradshaw, Jr., P.E. Heilman, and T.M. Hinckley, eds. NRC Research Press, National Research Council of Canada, Ottawa, ON. pp. 201-222.

Hardenack, S., Ye, D., Saedler, H., and Grant, S. 1994. Comparison of MADS-box gene expression in developing male and female flowers of the dioecious plant white campion. Plant Cell 6: 1775-1787.

Haughn, G.W., Schultz, E.A., and Martinez-Zapater, J.M. 1995. The regulation of flowering in *Arabidopsis thaliana*: meristems, morphogenesis, and mutants. Can. J. Bot. 73: 959-981.

Heilman, P.E., Stettler, R.F., Hanley, D.P., Carker, R.W. 1995. High yield hybrid poplar plantations in the Pacific Northwest. PNW Extension Bullention No. 356, revised edition. Pacific Northwest Regional Extension Publications, WA, OR, ID.

Henikoff, S. and Matzke, M.A. 1997. Exploring and explaining epigenetic effects. Trends Genet. 13: 293-295.

Hohenstein, W.G. and Wright, L.L. 1994. Biomass energy production in the United States: an overview. Biomass and Bioenergy 6: 161-173.

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

Hughes, C.E. 1994. Risks of species introductions in tropical forestry. Commonwealth Forestry Rev. 73: 243-252.

Hughes. D.W. amd Galau, G. 1988. Preparation of RNA from cotton leaves and pollen. Plant Mol. Biol. Rep. 6: 253-257.

Ingram, G.C., Goodrich, J., Wilkinson, M.D., Simon, R., Haughn, G.W., and Coen, E.S.. 1995. Parallels between *UNUSUAL FLORAL ORGANS* and *FIMBRIATA*, genes controlling flower cevelopment in *Arabidopsis* and *Antirrhinum*. Plant Cell 7: 1501-1510.

Ince, B.A., Zhuang, Y., Wrenn, C.K., Shapiro, D.J., and Katzenellenbogen, B.S. 1993. Powerful dominant negative mutants of the human estrogen receptor. J. Biol. Chem. 268: 14026-14032.

Ishida, B.K., Jenkins, S.M., and Say, B. 1998. Induction of *AGAMOUS* gene expression plays a key role in ripening of tomato sepals *in vitro*. Plant Mol. Biol. 36: 733-739.

Jack, T., Sieburth, L., and Meyerowitz, E. 1997. Targeted misexpression of *AGAMOUS* in whorl 2 of *Arabidopsis* flowers. Plant J. 11: 825-839.

Jacobsen, S.E. and Meyerowitz, E.M. 1997. Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*. Science 277: 1100-1103.

James, R.R. 1997. Utilizing a social ethic toward the environment in assessing genetically engineered insect-resistance in trees. Agricultural and Human Values 14: 237-249.

James, R.R., DiFazio, S.P., Brunner, A.M., and Strauss, S.H. 1998. Environmental effects of genetically engineered woody biomass crops. Biomass and Bioenergy 14: 403-414.

Jofuku, K.D., den Boer, B.G.W., Van Montagu, M., and Okamura, J.K. 1994. Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. Plant Cell 6: 1211-25.

Johnson, A.D. 1995. Molecular mechanisms of cell-type determination in budding yeast. Curr. Opin. Genet. Dev. 5: 552-558.

Kakutani, T. 1997. Genetic characterization of late-flowering induced by DNA hypomethlation mutation in *Arabidopsis thaliana*. Plant J. 12: 1447-1451.

Kang, H.-G., Noh, Y.-S., Chung, Y.-Y., Costa, M.A., An, K., and An, G. 1995. Phenotypic alterations of petal and sepal by ectopic expression of a MADS box gene in tobacco. Plant Mol. Biol. 29: 1-10.

Kappeli, O. and Auberson, L. 1997. The science and intracacy of environmental safety evaluations. Trends Biotech. 15: 342-349.

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

Kaul, R.B. 1995. Reproductive structure and organogenesis in a cottonwood, *Populus deltoides* (Salicaceae). Int. J. Plant Sci. 156: 172-180.

Kelly, A.J., Bonnlander, M., and Meeks-Wagner, D.R.. 1995. *NFL*, the tobacco homolog of *FLORICAULA* and *LEAFY*, is transcriptionally expressed in both vegetative and floral meristems. Plant Cell 7: 225-234.

Kempin, S.A., Mandel, M.A., and Yanofsky, M.F. 1993. Conversion of perianth into reproductive organs by ectopic expression of the tobacco floral homeotic gene *NAG1*. Plant Physiol. 103: 1041-1046.

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

Koch, P. 1992. Wood versus nonwood materials in U.S. residential construction: some energy-related global implications. For. Prod. J. 42: 31-42.

Kohn, W.D., Mant, C.T., and Hodges, R.S. 1997. α-helical protein assembly motifs. J. Biol. Chem. 272: 2583-2586.

Koltunow, A.M., Truettner, J., Cox, K.H., Wallroth, M., and Goldberg, R.B. 1990. Different temporal and spatial gene expression patterns occur during anther development. Plant Cell 2: 1201-1224.

Kumar, S, Tamura, K and Nei, M. 1993. MEGA: Molecular Evolutionary Genetic Analysis, version 1.0. The Pennsylvania State University, University Park, PA

Lee, I., Aukerman, M.J., Gore, S.L., Lohman, K.N., Michaels, S.D., Weaver, L.M., John, M.C., Feldmann, K.A., and Amasino, R.M. 1994. Isolation of *LUMINIDEPENDENS*: a gene involved in the control of flowering time in *Arabidopsis*. Plant Cell 6: 75-83.

Leple, J.-C., Bonade-Bottino, M., Augustin, S., Pilate, G., Dumanois Le Tan, V., Delplanque, A., Cornu, D. and Jouanin, L. 1995. Toxicity to *Chrysomela tremulae* (Coleoptera: Chrysomelidae) of transgenic Poplars expressing a cysteine proteinase inhibitor. Mol. Breed. 1: 319-328.

Lester, D.T. 1963. Floral initiation and development in quaking aspen. For. Sci. 9: 323-329.

Libby, W.J. 1993. Mitigating some consequences of *in-situ* genetic conservation. In, Proc. INRA/IUFRO International Symposium on Population Genetics and Gene Conservation of Forest Trees, Caracans Maubuisson, 1992, P. Bardat, ed.

Liu, Z. and Meyerowitz, E.M.. 1995. *LEUNIG* regulates *AGAMOUS* expression in *Arabidopsis* flowers. Development 121: 979-991.

Ma, H. 1998. To be, or not to be, a flower---control of floral meristem identity. Trends Genet. 14: 26-32.

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

Macknight, R., Bancroft, I., Page, T., Lister, C., Schmidt, R., Love, K., Westphal, L., Murphy, G., Sherson, S., Cobbett, C., and Dean, C. 1997. *FCA*, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. Cell 89: 737-745.

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

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

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

Matzke, M.A. and Matzke, A.J.M. 1995. Homology-dependent gene silencing in transgenic plants: what does it really tell us? Trends Genet. 11:1-3.

Mariani, C., De Beuckeleer, M., Truettner, J., Leemans, J., and Goldberg, R.B. 1990. Induction of male sterility in plants by a chimaeric ribonuclease gene. Nature 347:737-741.

Mariani, C., Gossele, V., De Beuckeleer, M., De Block, M., Goldberg, R.B., De Greef, W., and Leemans, J. 1992. A chimaeric ribonuclease-inhibitor gene restores fertility to male-sterile plants. Nature 357: 384-387.

McGaughey, W.H. and Whalon, M.E.. 1992. Managing insect resistance to *Bacillus thuringiensis* toxins. Science 258: 1451-1455.

McGonigle, B., Bouhidel, K., and Irish, V.F. 1996. Nuclear localization of the *Arabidopsis APETALA3* and *PISTILLATA* homeotic gene products depends on their simulataneous expression. Genes Dev. 10: 1812-1821.

McLetchie, D.N. and G.A. Tuskan. 1994. Gender determination in *Populus*. Nor. J. Agric. Sci. Suppl. 18: 57-66

Meilan, R., Ma, C., James, R.R., Strauss, S.S., Eaton, J., Taylor, M., Crockett, R.P., Rogan, G., Holden, L., Hoien, E., Stanton, B. and Han, K.-H. 1997. Development of glyphosate-resistant hybrid cottonwoods. Tappi Proceedings, 1997 Biological Sciences Symposium, San Francisco, CA, Oct. 19-23.

Mena, M., Ambrose, B.A., Meely, R.B., Briggs, S.P., Yanofsky, M.F. and Schmidt, R.J. 1996. Diversification of C-function activity in maize flower development. Science 274: 1537-1540.

Meyer, V.G. 1966. Flower abnormalities. Bot. Rev. 32: 165-217.

Meyerowitz, E.M. 1997. Genetic control of cell division patterns in developing plants. Cell 88: 299-308.

Miranda-Brasileiro, A.C., Tourneur, C., Leple, J.C., Combes, V., and Jouanin, L. 1992. Expression of the mutant *Arabidopsis thaliana* acetolactate synthase gene confers chlorsulfron resistance to transgenic poplar plants. Transgenic Res. 1: 133-141.

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

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

Mizukami, Y. and Ma, H. 1995. Separation of *AG* function in floral meristem determinacy from that in reproductive organ identity by expresing antisense *AG* RNA. Plant Mol. Biol. 28: 767-784.

Mizukami, Y. and Ma, H.. 1997. Determination of *Arabidopsis* floral meristem identity by *AGAMOUS*. Plant Cell 9: 393-408.

Modrusan, Z., Reiser, L., Feldmann, K.A., Fischer, R.L., and Haughn, G.W. 1994. Homeotic transformation of ovules into carpel-like structures in *Arabidopsis*. Plant Cell 6: 333-349.

Moffat, A.S. 1995. Plants providing their worth in toxic metal cleanup. Science 269: 302-303.

Moffat, A.S. 1997. Resurgent forests can be greenhouse sponges. Science 277: 315-316.

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

National Academy of Science. 1987. Introduction of Recombinant DNA-engineered Organisms into the Environment: Key Issues. National Academy Press, Washington D.C.

National Research Council. 1989. Field Testing Genetically Modified Organisms: Framework for Decision. National Academy Press, Washington D.C.

Newcombe, G., H.D. Bradshaw, Jr., G.A. Chastagner and R.F. Stettler. 1996. A major gene for resistance to *Melampsora Medusae f. sp. deltoidae* in a hybrid poplar pedigree. Phytopathology 86: 87-94.

Niemiec, S.S., Ahrens, G.R., Willits, S., and Hibbs, D.E. 1995. Hardwoods of the Pacific Northwest. Forest Research Laboratory, Oregon State University.

Okamuro, J.K., Caster, B., Villarroel, Van Montagu, M., and Jofuku, K.D. 1997. The AP2 domain of *APETALA2* defines a large new family of DNA binding proteins in *Arabidopsis*. Proc. Natl. Acad. Sci. 94: 7076-7071.

Okamuro, J.K., den Boer, B.G.W., Lotys-Prass, C., Szeto, W., and Jofuku, K.D. 1996. Flowers into shoots: photo and hormonal control of a meristem identity switch in *Arabidopsis*. Proc. Natl. Acad. Sci. 93: 13831-13836.

Ostry, M.E., L.F. Wilson, H.S. McNabb, Jr. and L.M. Moore. 1989. A guide to Insect, Disease, and Animals Pests of Poplars. USDA For. Serv. Agric. Handbook 677, Washington, D.C. 118pp.

Pappenheimer, A.M., Jr. 1977. Diphtheria toxin. Annu. Rev. Biochem. 46: 69-94.

Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences 12: 357-358.

Perlack, R.D., Wright, L.L., Huston, M.A., and Schramm, W.E. 1995. Biomass fuel fom woody crops for electric power generation. ORNL-6871 Report, Oak Ridge National Laboratory, Oak Ridge, TN.

Perez-Garcia, J.M. 1995. Global economic and land use consequences of North American timberland withdrawls. J. Forestry 93 (7): 34-38.

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

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

Pnueli, L., D. Hareven, A.D. Rounsley and M.F. Yanofsky. 1994b. Isolation of the tomato *AGAMOUS* gene *TAG1* and analysis of its homeotic role in transgenic plants. Plant Cell 6: 163-173.

Polak, J. 1997. Sustaining profits and forests. Sustainable Forestry Working Group, The John D. and Catherine T. MacArthur Foundation, Chicago, IL., 32pp.

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

Purugganan, M.D., Rounsley, S.D., Schmidt, R.J., and Yanofsky, M.F. 1995. Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. Genetics 140: 345-356.

Putterill, J., Robson, F., Lee, K., Simon, R., and Coupland, G. 1995. The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80:847-857.

Raffa, K.F. 1989. Genetic engineering of trees to enhance resistance to insects. BioScience 39: 524-534.

Ramsey, G. 1998. DNA chips: state-of-the art. Nature Biotech. 16: 40-44.

Ranney, J.W. and Mann, L.K. 1994. Environmental considerations in energy crop production. Biomass and Bioenergy 6: 211-228.

Ray, A. Robinson-Beers, K., Ray, S., Baker, S.C., Lang, J.D., Preuss, D., Milligan, A.B., and Gasser, C.S. 1994. *Arabidopsis* floral homeotic gene *BELL* (*BEL1*) controls ovule development through negative regulation of *AGAMOUS* gene (*AG*). Proc. Natl. Acad. Sci. 91: 5761-5765.

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

Riechmann, J.L. and Meyerowitz, E.M. 1997. Determination of floral organ identity by *Arabidopsis* MADS domain homeotic proteins AP1, AP3, PI, and AG is independent of their DNA-binding specificity. Mol. Biol. Cell 8: 1243-1259.

Riechmann, J.L., Wang, M., and Meyerowitz, E.M. 1996b. DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucleic Acids Res. 24: 3134-3141.

Reiser, L., Modrusan, Z., Margossian, L., Samach, A., Ohad, N., Haughn, G.W., and Fischer, R.L. 1995. The *BEL1* gene encodes a homeodomain protein involved in pattern formation in the *Arabidopsis* ovule primordium. Cell 83: 735-742.

Reynaerts, A., Van de Wiele, H., De Sutter, G., and Janssens, J. 1993. Engineered genes for fertility control and their application in hybrid seed production. Sci. Hortic. 55:125-139.

Richards, E.J. 1997. DNA methylation and plant development. Trends Genet. 13: 319-323

Richardson, D.M. 1998. Forestry trees as invasive aliens. Conserv. Biol. 12: 18-26.

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

Saitou, N and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425

Sakai, H., Medrano, L.J., and Meyerowitz, E.M. 1995. Role of *SUPERMAN* in maintaining *Arabidopsis* floral whorl boundaries. Nature 378: 199-203.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual, 2nd ed., cold spring Harbor Laboratory Press, N.Y.

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

Schena, M. 1996. Genome analysis with gene expression microarrays. Bioessays 18: 427-431.

Schmidt, R.J., Veit, B., Mandel, M.A., Mena, M., Hake, S., and Yanofsky, M.F. 1993. Identification and molecular characterization of *ZAG1*, the maize homolog of the *Arabidopsis* floral homeotic gene *AGAMOUS*. Plant Cell 5: 729-737.

Schreiner, E.J. 1971. Genetics of eastern cottonwood. USDA-For. Ser. Res. Paper WO-11. 24 pp.

Schuette, W.R. 1995. Growth and yield of hybrid poplars at the lower Columbia River fiber farm (abstract). International Poplar Symposium, Seatttle, WA, Aug. 20-25, 1995.

Schwarz-Sommer, Z., Hue, I., Huijser, P., Flor, P.J., Hansen, R., Tetens, F., Lonnig, W.-E., Saedler, H., and Sommer, H. 1992. Characterization of the *Antirrhinum* floral homeotic MADS-box gene *DEFICIENS*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. EMBO J. 11: 251-263.

Sedjo, R.A., and Botkin, D. 1997. Using forest plantations to spare natural forests. Environment 39(10): 14-20.

Shannon, S. and Meeks-Wagner, D.R. 1991. A mutation in the *Arabidopsis TFL1* gene affects inflorescence meristem development. Plant Cell 3: 877-892.

Shannon, S. and Meeks-Wagner, D.R. 1993. Genetic interactions that regulate inflorescence development in *Arabidopsis*. Plant Cell 5: 639-655.

Sheppard, L.A. 1997. *PTD*: a *Populus trichocarpa* gene with homology to floral homeotic transcription factors Ph.D. Dissertation, Genetics Program, Oregon State University, Corvallis.

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

Sieburth, L.E. and Meyerowitz, E.M. 1997. Molecular dissection of the *AGAMOUS* control region shows that cis elements for spatial regulation are located intragenically. Plant Cell 9: 355-365.

Sieburth, L.E., Running, M.P., and Meyerowitz, E.M. 1995. Genetic separation of third and fourth whorl functions of *AGAMOUS*. Plant Cell 7: 1249-1258.

Simon, R., Carpenter, R., Doyle, S. and Coen, E. 1994. *Fimbriata* controls flower development by mediating between meristem and organ identity genes. Cell 78: 99-107.

Simon, R., Igeno, M.I., and Coupland, G.. 1996. Activation of floral meristem identity genes in *Arabidopsis*. Nature 384: 59-62.

Sleigh, M.J. 1991. Perceptions and misconceptions of genetic engineering. Search 22: 127-128.

Snow, A.A. and Moran Palma, P. 1997. Commercialization of transgenic plants: potential ecological risks. BioScience 47: 86-96.

Stanton, B.J. and Villar, M. 1996. Controlled reproduction of *Populus*. In, Biology of *Populus* and its implications for management and conservation. Part I, Chapter 5. R.F. Stettler, H.D. Bradshaw, Jr., P.E. Heilman, and T.M. Hinckley, eds. NRC Research Press, National Research Council of Canada, Ottawa, ON. pp. 113-138.

Steinbrecher, R.A. 1996. From green to gene revolution: the environmental risks of genetically engineered crops. Ecologist 26: 273-281.

Stettler, R.F. 1971. Variation in sex expression of black cottonwood and related hybrids. Silvae Genet. 20: 42-46.

Stettler, R.F., Zsuffa, L. and Wu, R. 1996. The role of hybridization in the genetic manipulation of *Populus*. In, Biology of *Populus* and its implications for management and conservation. Part I, Chapter 4. R.F. Stettler, H.D. Bradshaw, Jr., P.E. Heilman, and T.M. Hinckley, eds. NRC Research Press, National Research Council of Canada, Ottawa, ON. pp. 87-112.

Stirling, B.V., Necombe, G., and Bradshaw, H.D. 1998. Development of a fine stucture genetic map around a disease resistance locus in hybrid poplar (abstract). Plant and Animal GenomeVI, San Diego, CA, Jan 18-22, 1998.

Stomp, A.-M., K.-H. Han, S. Wilbert, M.P. Gordon and S.D. Cunningham. 1994. Genetic strategies for enhancing phytoremediation. *In*, Recombinant DNA Technology II Vol. 721 of the Annals of the New York Academy of Sciences.

Strauss, S.H., Howe, G.T., and Goldfarb, B. 1991. Prospects for genetic engineering of insect resistance in forest trees. For. Ecol. Manage. 43: 181-209.

Strauss, S.H., James, R., Brunner, A., DiFazio, S. and Meilan, R. 1997a. Tree Genetic Engineering Research Cooperative (TGERC) Annual Report: 1996-1997. Forest Research Laboratory, Oregon State University.

Strauss, S.H., Knowe, S.A., and Jenkins, J. 1997b. Benefits and risks of genetically engineered, glyphosate-resistant cottonwoods. J. Forestry 95: 12-19.

Strauss, S.H., Rottmann, W.H., Brunner, A.M., and Sheppard, L.A. 1995. Genetic engineering of reproductive sterility in forest trees. Mol. Breed. 1: 5-26.

Suarez-Lopez, P. and Coupland, G. 1998. Plants see the blue light. Science 279: 1323-1324.

Tandre, K., Albert, V.A., Sundas, A., and Engsrom, P. 1995. Conifer homologues to genes that control floral development in angiosperms. Plant Mol. Bio. 27: 69-78.

Tandre, K. 1997. Molecular approaches to the developmental biology of Norway Spruce, *Picea abies*. Ph.D. Dissertation, Department of Physiological Botany, Uppsala University, Uppsala, Sweden.

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

Theissen, G. and H. Saedler. 1995. MADS-box gene in plant ontogeny and phylogeny: Haeckel's 'biogenetic law' revisted. Curr. Opin. Gen. Dev. 5:628-639.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nuc. Acids Res. 22:4673-4680.

Thorsness, M.K., Kandasamy, M.K., Nasrallah, M.E., and Nasarallah, J.B. 1991. A *Brassica* S-locus gene promoter targets toxic gene expression and cell death to the pistil and pollen of transgenic *Nicotiana*. Devel. Biol. 143: 173-184.

Tiedje, J.M., Colwell, R.K., Grossman, Y.L., Hodson, R.E., Lenski, R.E., Mack, R.N., and Regal, P.J. 1989. The planned introduction of genetically engineered organisms: ecological considerations and recommendations. Ecology 70: 298-315.

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

Tsuchimoto, S., van der Krol, A.R., and Chua, N.-H. 1993. Ectopic expression of *PMADS3* in transgenic petunia phenocopies the petunia blind mutant. Plant Cell 5: 843-853.

Tuskan, G.A., Alstrom-Rapaport, C., Wang, Y.C., Roberts, G., and Gullberg U. 1996. The detection of female-associated DNA markers in Salicaceae. SREIG Workshop: Novel applications of molecular markers in forest trees.

USDA-APHIS, Biotechnology and Scientific Services. 1998. http:/www.aphis.usda.gov/biotech/, May 7, 1998.

Vincent, C.A., Carpenter, R., and Coen, E. 1995. Cell lineage patterns and homeotic gene activity during *Antirrhinum* flower development. Curr. Biol. 5:1449-1458.

Wagner, D.B., Furnier, G.R., Saghai-Maroof, M.A., Williams, S.M., Dancik, B.P., and Allard, R.W. 1987. Chloroplast DNA polymorphisma in lodgepole pines and their hybrids. Proc. Natl. Acad. Sci. USA 84: 2097-2100.

Weigel, D. 1995. The genetics of flower development: from floral induction to ovule morphogenesis. Annu. Rev. Genetics 29:19-39.

Weigel, D. 1997. Flower development: repressing reproduction. Curr. Biol. 7:R373-R375.

Weigel, D. and Clark, S.E. 1996. Sizing up the floral meristem. Plant Physiol. 112:5-10.

Weigel, D. and E.M. Meyerowitz. 1993. Activation of floral homeotic genes in *Arabidopsis*. Science 261: 1723-1726.

Weigel, D. and E.M. Meyerowitz. 1994. The ABCs of floral homeotic genes. Cell 78: 203-209.

Weigel, D. and O. Nilsson. 1995. A developmental switch sufficient for flower initiation in diverse plants. Nature 377: 495-500.

Wright, L.L. 1994. Production technology status of woody and herbaceous crops. Biomass and Bioenergy 6: 191-209.

Wright, L.L., and Berg, S. 1996. Industry/government collaborations on short-rotation woody crops for energy, fiber and wood products. Proc., Bioenergy '96 - The Seventh National Bioenergy Conference: Partnerships to Develop and Apply Biomass Technologies, Nashville, TN, September 15-20, 1996.

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

Yanofsky, M.F. 1995. Floral meristems to floral organs: genes controlling early events in *Arabidopsis* flower development. Annu. Rev. Plant Physiol. Plant Mol. Biol. 46: 167-188.

Zachgo, S., E. de Andrade Silva, P. Motte, W. Trobner, H. Saedler and Z. Schwarz-Sommer. 1995. Functional analysis of the *Antirrhinum* floral homeotic *DEFICIENS* gene *in vivo* and *in vitro* by using a temperature-sensitive mutant. Development 121: 2861-2875.

Zhang, Y., Buongiorno, J., and Zhang, D. 1997. China's economic and demographic growth, forest products consumption, and wood requirements: 1949 to 2010. For. Prod. J. 47: 27-35.

Zsuffa, L., Giordano, E., Pryor, L.D., and Stettler, R.F. 1996. Trends in poplar culture: some global and regional perspectives. In, Biology of *Populus* and its implications for management and conservation. Part II, Chapter 20. R.F. Stettler, H.D. Bradshaw, Jr., P.E. Heilman, and T.M. Hinckley, eds. NRC Research Press, National Research Council of Canada, Ottawa, ON. pp. 515-539.

Appendices

APPENDIX A: ALIGNMENT OF *PTAG* GENOMIC SEQUENCES

Exons are underlined, and noted in the left margin. Translational start and stop codons are in bold type.

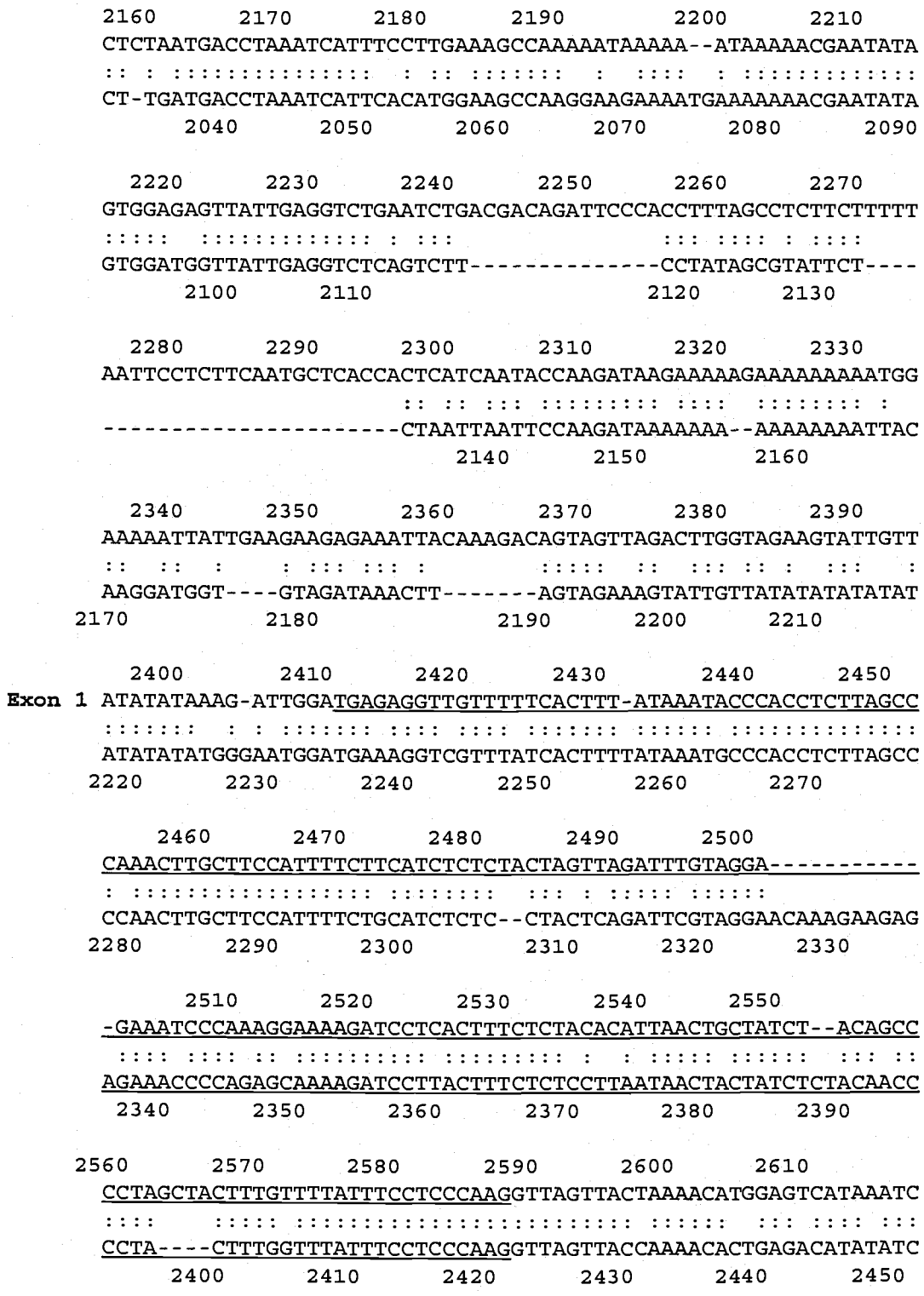
```
                10         20         30         40         50
PTAG1 GGATCCACCTCCACGTCAG-TCCATCCGCATTCGTAAG--TCCATAAAACTACCAGATTT
       :  :: ::   :: : :::   :: : : : :  : : : ::   :  :::  :
PTAG2 TCCATTATTTCAACAATAGATTCATTTACACTAGCATGGATACTTCAATGAATAAGAAGT
                10         20         30         40         50         60

       60         70         80          90          100         110
      TGCTTATTCTTGTTATTCTTCATCA--TTTACTTC-----CTTTTTAGCTTCTATTCATT
         ::::: ::   : :  :   ::  : :: :::     : :::::  :: : :: :::
      GTGTTATTGTTTGGAGTAATAGACACCTATAATTCTCAAACCTTTTACTTTATTTTTATT
                70         80         90        100        110        120

              120        130        140        150         160
      GCCTTTTTGAGCCCTCTTCCTATAAAGAGGCAATTCTTGATCCGCTT--CGGCAACAAGC
       ::::: : :    : :: :  :       :: ::::: ::  : ::  :    ::: :
      TCCTTTGTTA----TATTACATTTT----TCATTTCTTTATTGGGTTTTCATTGACAGGA
              130            140           150        160        170

           170        180        190        200        210        220
      T------ATGAATGAAGAATTTTCTGCTTTGCATAAGACAGATACTTGGGATCTGGTTCC
      :      :: :::  ::  :::  ::  ::   ::: : : : :     ::::  :   :
      TGGCTAGATTAATATAG--TTTCTTGACTT---TAATAAATAAAAAAAAGATCAAGACTC
           180          190        200          210        220

              230        240        250        260        270
      TCT--ACCTCCCGGTAAGAGTGTTGTTGGTTGTCATTGGGTG-TATAAGATCAAGACTAA
      :::  ::   ::  ::  :              :::::: : :::   ::: : :::::
      TCTTCACAAACCTTTACAAA-------------ATTGGGCGCTAT---ATCTAAACTAA
      230        240                    250          260        270

    280        290        300        310        320        330
      TTCTGATGGGTCTATTGAGCAATACAAAGCTAGGCTGGTTGCAAAAGGATACTCTCAACA
           :  :  :::       :::: :    :::   :: ::: :     : :  ::  :
      AAAACTTAAGATTAT------ATACTATCTAAGG--AGTAGCACACTATAAATAACATTA
              280            290         300        310        320
```

```
340        350         360         370         380             390
  TTATGGTATGGACTATGAG-GAAACATTTGCCCCGGTTGCAAAAATGACT----ACTATT
  : : :::: :      :::: : :::  : : :    :::::::: : :::     : :::
  TAAAGGTA-GTTTGTTGAGCGGAAC--TAGACT---TTGCAAAA-TAACTTTCCAATATA
       330         340          350          360         370

     400         410          420         430         440         450
  CGTACTCTTATTGTCGTAG--CTTCGATTCGTCAGTGGCATATTTCTCAGCTTGATGTTA
    :  :::: :::  :: :  :::  : :       :   : : :: : :  :  :  :
  GCTTTTCTTGTTGATGTTGACCTTTTAATTTAGGATCAAACACTTGTAAATTACA--ATT
    380         390         400         410         420         430

       460         470         480         490         500         510
  AAAATGCCT-TCTTGAATGGAGATCTTCAAGAAGAAGTTTATGTGGCACTCCCTCCTGGT
  :::: :: : : ::   : :  :: :: :  ::: : : :  :     ::       : ::
  AAAAGGCTTATTTTTGTTTGCCATTTTTACCAAGCAATGTTAGGATTGCTAGAGATTAGT
      440         450         460         470         480         490

        520         530         540           550         560
  ATTTCATATGACTCTGGATATGTTTGTAA---GCTTAAGAAAGCATTAATTATATGGTCT
   ::::    :: :  : :  : : : :::   :::::: : : :    : ::  : :
  TTTTCCATGGAATAAGAAGTTATCTTTAAAGGGCTTAAAAGACCT---AGTAGCTTG---
      500         510         520         530           540

 570          580         590         600          610         620
  CAAACAAGC--ACCCCGTGCTTGGTTTGAGAAATTCTC--TATTGTGATCTCGTCTCTTG
     :::::   :   : ::  : ::::  ::   : :  ::::::  ::   :     ::
  ---ACAAGGCTATGACTTGTGTTGTTTTGGATCGTATGGTTATTGTTATAGAG----GTG
     550         560         570         580         590           600

    630         640         650         660          670         680
  GCATTGTTTCTAG-CAGTCATGATTCTGCTCTTTTTA--TTAAGTGCACT-GATGCAGGT
    : :: ::  :: ::  ::: ::   : :  :  :   ::::  :   : :::: :  :
  CTAGTGGTTAAAGACATCCATCATGGAGGTGGTGATGACTTAAAAGAGTTAGATGTAAAT
         610         620         630         640         650         660

         690                 700         710           720
  CGTATCATTCTGTCTTTAT-------ATGTTGATAACATG----ATTATTATTGGTGATG
   : :   :: :::  :: :        :   :: :: : :     ::: :: ::::   ::
  TGGAGACTTATGTTATTCTCACATAAAAAATGTTAGCCTCCGACATTGTTTTTGGATGTG
         670         680         690         700         710         720

  730        740          750         760             770         780
  -ACATTGATGGTA---TTTCAGTCTTGAAGACAAAGTTGGCTA-----GACGATTTGAAA
   : : : :  :::   ::: : :::: :        ::  ::       :::  ::  :::
  TAAAATCAATGTACCATTTTATTCTTCATTGTTTGTTTTCCTTATTATGACTTTTACAAA
         730         740         750         760         770         780
```

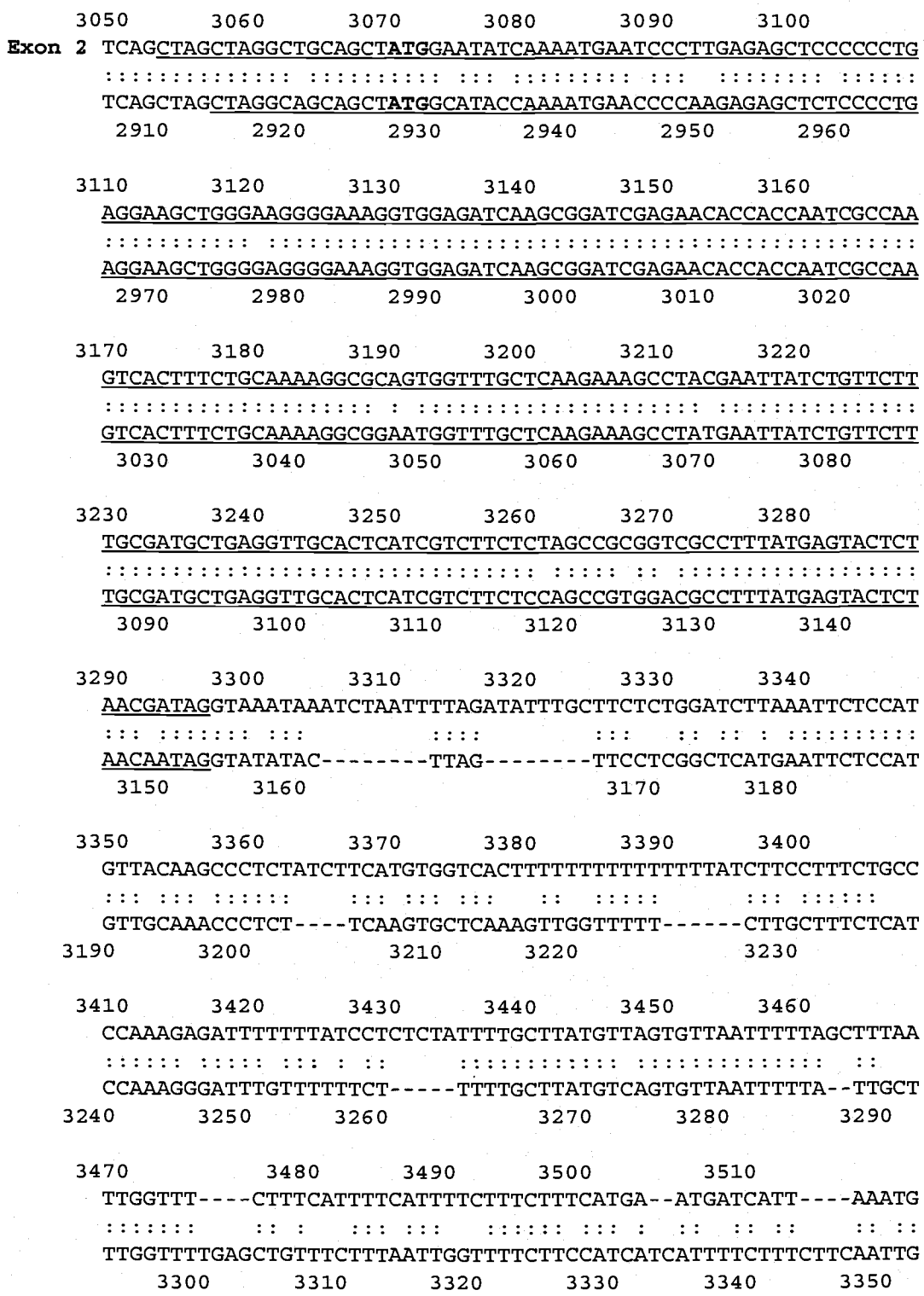
```
       790               800         810       820       830
TGAAAGATTTGGGTTAT-------CTTCAATATT-TCCTGGGTATTGAGGTAGCATACTC
:  :   ::: ::: ::        ::::::: :: : ::   : ::    ::  :: ::
TTTATCCTTTAGGTGATGAAATTCCTTCAATCTTGTTCTATTTTTTTTTTTA--ATTCTT
       790       800       810       820       830

     840       850       860       870          880       890
ACCTAGAGGTTACCTTCTTTCTCAGTCGAAATATGTTGCAGATAT--TCTTGAGCAGACT
     :  :::   : :::  :::  : : :::   :   :::    :::: : ::  ::
GGTACGTAGTTCTGTACTTAATCAAGCAACATAAAATAGTGATGCCATCTTCATCACTCT
840       850       860       870       880       890

       900       910       920       930       940       950
AGACTTACTGATAACAAAACTGTAGATACTCCTATTGAGGTCAACGTGAGGTACTCTTCT
: :              ::: :: :: :: :  ::           :  ::  :: :: :
ATA-------------AAC-GTGGAAACCCAAATC----------TCTGG--CTTTTAT
900                      910                  920         930

       960         970       980          990      1000
TCTGATGGTTTACCT--TTGATAGATCTTACTTTATAC---CACACTATTGTTAGGAGTT
::  ::: :: :  :  ::  ::::: ::  :::: ::   ::    :   ::::: :::
TC--ATGATTAAAGTCATTTCTAGATTTT--TTTAGACGTTCAAGTGAGATTTAGG-GTT
        940       950         960       970       980

 1010      1020      1030      1040      1050      1060
TGGTATATCTCACCATTACTCGTCCAGATATTGCATATGCTGTTCATGTTGTTAGTCAGT
    :::    :  :: : : ::  : :::    : :    :::  ::: ::::
C--AATAAGAGAGGATCAATGGTGAAAATAGAAGA-ACAAAGTTGTTGTGGTTAA-----
   990      1000      1010       1020      1030

 1070      1080      1090      1100      1110      1120
TTGTTGCTTCTCTTACTACTGTTCACTGGGCAGCTGTTATTCGTATTTTGCGATATCTTC
                    ::: ::: ::  : ::::         ::: :::::
--------------------GTTGACTCGGTGGTTGTTG------AGTTGGGATATGA--
                   1040      1050             1060      1070

 1130      1140      1150      1160      1170      1180
GGGGTACAGTTTTTCAGAGTCTTTTACTTTCATCCACCTCTTTCTTGGAGTTGCGTGCAT
  :: : :: :  : ::: :  : :: : :          :::  :  : ::: :: : :
-AGGAATAGATGGT-AGACTAATCTAGTGT----------TTTTGTCCACTTGAGTTCTT
       1080      1090                1100      1110

 1190      1200      1210      1220      1230      1240
ACTCTGATGCTGATCATGGTAGTGATCCCACAGATCGCAAGTCTGTTACCGGGTTCTGTA
: :               :: :  :::  :   :: :: : :: :: ::   ::::
AAT---------------TATTATTCCATC---TC-CATGACTATTTCCATCTTCT---
1120                    1130          1140      1150
```

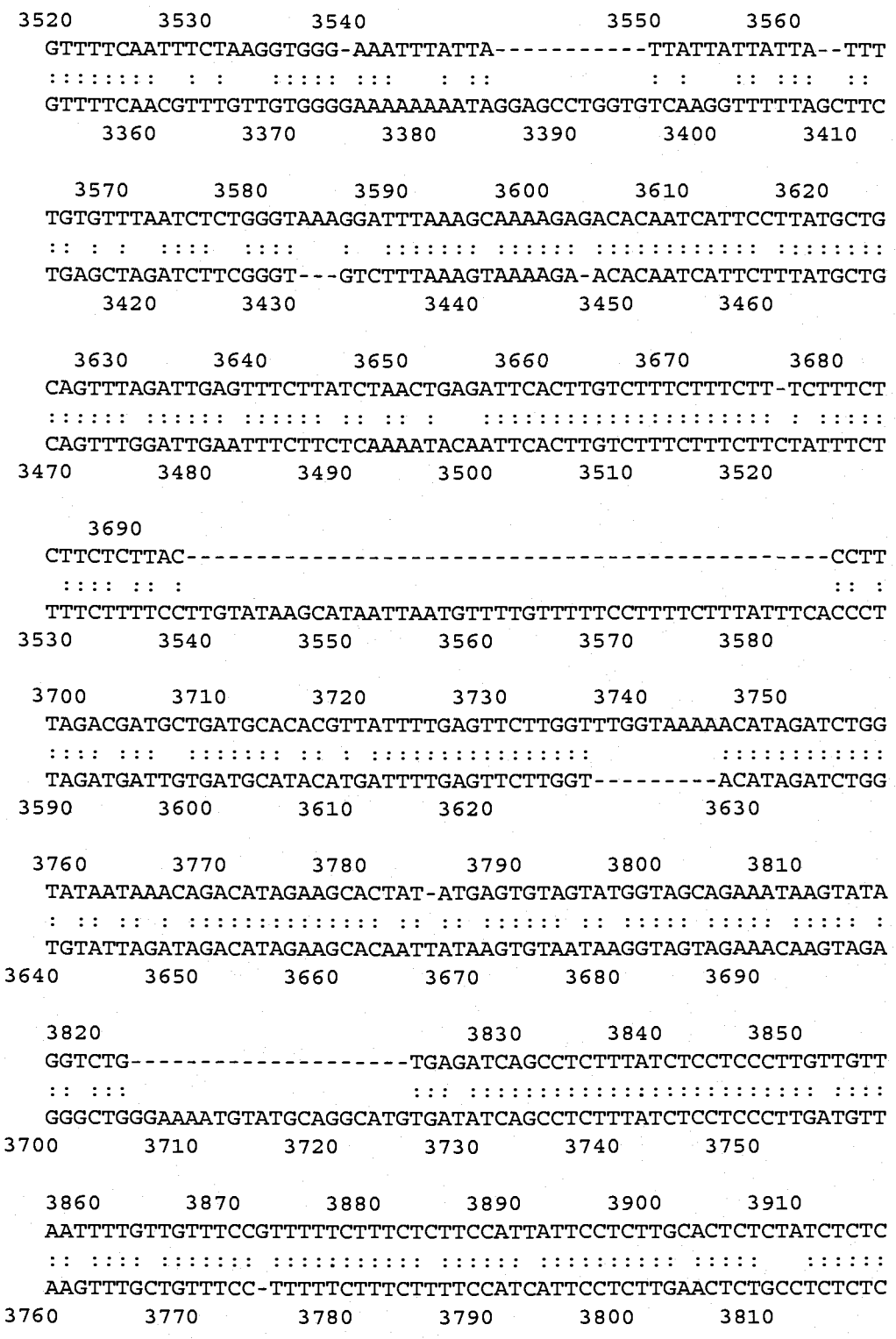
```
        1250       1260       1270       1280       1290
TCTT------TTTAGGTGATTCTCTTATTTCTTGGAAGAGCAAGAAACAATCTATTGTTT
::::      : : : : :: ::::    : : ::   :    : :::    :::::
TCTTCAGTGATATTGTTTATACTCTGTGATTTGGGTTTATT--GGAACTTATTATTGAGG
 1160       1170       1180       1190         1200       1210

1300       1310       1320       1330       1340          1350
CTCAATCATCCATCGAAGCAGAATATCGTGCCATGACATCTACTA-CCA--AAGAGATTG
:    :::::::: :::     :: : :: :  :  : :: :: : :::  :: :  :::
CAGC-TCATCCATAGAA-----ATTTGGTAC--TTGCTTCAACAAACCACTAAAATGTTG
    1220            1230        1240       1250       1260

1360       1370       1380       1390       1400       1410
TTTGGTTA-TGTTGGTTACTTGCTGATATGAGAGTTTCATTTTCTCATCCTACTCCTATG
: :::::: : :: :  : :    :: :  : ::  :: :  :    :     :::
TGTGGTTAATATTTGAGAATGCGCGAAA--AAAGCATCGTACTAAATTTGGGTTCCCG--
 1270       1280       1290         1300       1310       1320

 1420       1430       1440       1450         1460       1470
TATTGTGACAACCAGAGTTCTATTCAGATTGCTCACA--ACTCGGTTTTTCATGAGCGAA
 : :: ::  :  :::: : :     :::: :: :    : : :: :::::  : :   :
-ACTG-GATGAAGAGAGATGT-----GATTACTTAATTTATTTGGATTTTCG-GGGTTTA
       1330       1340            1350       1360       1370

   1480       1490       1500       1510       1520       1530
CTAAGCACATTGAGATCGATTGTCATCTTACTCATCATCATCTCAAGCATGGCACCATTG
 ::       : :: :   : :   :: : : :  :  : :    ::  ::   : :::::
TTAGATTTTTGGAAAGGTAATACGATATCATTGGTTTTGAGAGGAA--AT---AACATTG
   1380       1390       1400       1410         1420

   1540       1550       1560        1570        1580       1590
CTTTACCTTTTGTTCCTTCTTCCTTAC-AGATTGCAGATTT--CTTTATCAAGGCGCATT
   :    :::: :  :: ::   ::  : : :  :: :::  ::: ::
GGAT----TTTGATGATTTTTGAATAATAAAATTAAGTTTTTTCTTGAT-----------
1430          1440       1450       1460       1470

      1600       1610        1620       1630       1640
CCATCTCTCGTTTTTGTTTTCAG--GTTGGCAAACTCTCGATGCTTGTAGCTGCCGCAT-
   ::  : :::  :       ::  :  ::  : :::::  :  :: :::: :  : :
---TCATTTGTTAATAGAAAGAGAAGAGGGATAGCTCTC--TTATTCTAGCAGAAGTACG
      1480       1490       1500         1510       1520

 1650       1660       1670       1680        1690       1700
--TGTGAGTTTGAGGGGAGATGTTAAATAATATTTATGTA-GTCTTATTTATTAAGGGTA
  : :::: :  : ::::    :::: :  :: :: :::: :  ::::: :: :: :  :
TATATGAGCT--ATGGGA---TTTAATTCTTAATTTTGTATGAGTTATTGATCAAAGAAA
1530        1540         1550       1560       1570       1580
```

```
  1710       1720       1730       1740          1750       1760
GAATAGTACTTTCAGTTTAACCTATATATACTTTATTT---GTATTTAGGTTAAGACTAA
 :   : :   : : ::         :::::::: ::: :      ::: :   : : ::    :
AAGCAATGATGTGAGAAGTCTATATATATAATTTCTCCTACGTACTCCGTTGAACCTTTT
   1590       1600       1610       1620       1630       1640

     1770       1780       1790       1800       1810       1820
GCATTCATAATAAATGTATCATTAAGAATTCTAGCCTCCTTTTCGTGTTTCATTTTAATT
   ::: :::::::    ::    ::  :: ::::  :  : : :   :     :: : :: :
---TTCCTAATAAAA--ATTGATAGAAAATCTA--CAACATATACAGAGAAATGTGAAGT
      1650         1660       1670         1680       1690

     1830       1840       1850       1860        1870       1880
ATTTTTAACAATCTTGTATTAATATATGTGAATTGATTGAATTAAATC-TTGTAATACAA
  ::     ::::    : :: :::    :    ::  ::    :    :  :: ::::: :  :
TCTT----CAATTGAGAATAAATCGTTTCAAAAGGACGTAGGAATCTCCTTGTAGTG-AG
1700           1710       1720       1730       1740        1750

      1890       1900       1910       1920       1930
TTTAATTGATTCTAAGTTAAACAATCTGCTGGGGAACATTCATACAACTATCTT------
:   :: :       :: :::::::: :::::::::::::: :::::::::::: :
TGAAACTCCAAGAAAATTAAACAACCTGCTGGGGAACATCCATACAACTATCCTCCGATC
     1760       1770       1780       1790       1800       1810

     1940       1950       1960       1970       1980
---TCTTTTCGTTTCAAGTAGGCAGGAAATAAAACGTTTTTAGTTTAGGTGACT------
   ::::::: :::::::::::::::: :::::::::: ::::  :::    : :
CCTTCTTTTCTTTTCAAGTAGGCAGGCAATAAAACGTATTTAGCATAGCCAAGTTCAAAA
     1820       1830       1840       1850       1860       1870

    1990             2000       2010       2020       2030
---AAACAATG-GAATTT-----AATGAAATAAGGGTAGAGATGAGGTCTGAGGTTATCT
   ::::::   :::         :::::::::::: : ::::::        :::: :::
AAAAAACAAGAAGAAGAAGAAGCAATGAAATAAGGGAAAAGATGA-------GGTTCTCT
     1880       1890       1900       1910             1920

2040       2050       2060       2070       2080       2090
TGTTAAGCACCTTCCCATTTGAACCATGATTTTGTCGTTAAGCACTGAGAGTGTAACTTA
:::::::::::: : ::::: ::::: :::::::: ::   : : :::      : :    :
TGTTAAGCACCTTTC-ATTTGTACCATAATTTTGTCCTT--GGAATGA-----TTAGAGA
  1930       1940       1950       1960             1970

2100       2110       2120       2130       2140       2150
GCCCTAAAACGTCTCACTCACCCCATTATAATTCATTTTCAGAAAGTCCCTTGCTTTTCT
:::: ::::::: : : ::::::::  : ::: :::::::: :::::::::     ::::
GCCCAAAAACGTGTTATTCACCCCAGAAAAATCCATTTTCAAAAAGTCCCT-----TTCT
1980       1990       2000       2010       2020            2030
```


2160 2170 2180 2190 2200 2210
CTCTAATGACCTAAATCATTTCCTTGAAAGCCAAAAATAAAAA--ATAAAAACGAATATA
CT-TGATGACCTAAATCATTCACATGGAAGCCAAGGAAGAAAATGAAAAAAACGAATATA
2040 2050 2060 2070 2080 2090
2220 2230 2240 2250 2260 2270
GTGGAGAGTTATTGAGGTCTGAATCTGACGACAGATTCCCACCTTTAGCCTCTTCTTTTT
GTGGATGGTTATTGAGGTCTCAGTCTT----------CCTATAGCGTATTCT----
2100 2110 2120 2130
2280 2290 2300 2310 2320 2330
AATTCCTCTTCAATGCTCACCACTCATCAATACCAAGATAAGAAAAAGAAAAAAAAATGG
----------CTAATTAATTCCAAGATAAAAAAA--AAAAAAAATTAC
2140 2150 2160
2340 2350 2360 2370 2380 2390
AAAAATTATTGAAGAAGAGAAATTACAAAGACAGTAGTTAGACTTGGTAGAAGTATTGTT
AAGGATGGT----GTAGATAAACTT-------AGTAGAAAGTATTGTTATATATATATAT
2170 2180 2190 2200 2210
2400 2410 2420 2430 2440 2450
Exon 1 ATATATAAAG-ATTGGATGAGAGGTTGTTTTTCACTTT-ATAAATACCCACCTCTTAGCC
ATATATATGGGAATGGATGAAAGGTCGTTTATCACTTTTATAAATGCCCACCTCTTAGCC
2220 2230 2240 2250 2260 2270
2460 2470 2480 2490 2500
CAAACTTGCTTCCATTTTCTTCATCTCTCTACTAGTTAGATTTGTAGGA----------
CCAACTTGCTTCCATTTTCTGCATCTCTC--CTACTCAGATTCGTAGGAACAAAGAAGAG
2280 2290 2300 2310 2320 2330
2510 2520 2530 2540 2550
-GAAATCCCAAAGGAAAAGATCCTCACTTTCTCTACACATTAACTGCTATCT--ACAGCC
AGAAACCCCAGAGCAAAAGATCCTTACTTTCTCTCCTTAATAACTACTATCTCTACAACC
2340 2350 2360 2370 2380 2390
2560 2570 2580 2590 2600 2610
CCTAGCTACTTTGTTTTATTTCCTCCCAAGGTTAGTTACTAAAACATGGAGTCATAAATC
CCTA----CTTTGGTTTATTTCCTCCCAAGGTTAGTTACCAAAACACTGAGACATATATC
2400 2410 2420 2430 2440 2450

```
2620       2630       2640       2650       2660       2670
  TCGTTGTATTCTTCAGTGCTTCATCACTTGTTTTGGGCTAATTAATCAATCTTTTCACGT
  :::::::::::::: :::::::   ::::::::: :::::     ::::::::: : :  :
  TCGTTGTATTCTTGAGTGCTT---CACTTGTTTGGGGCT----TATCAATCTTCTGATCT
       2460       2470          2480           2490       2500

2680       2690       2700       2710       2720       2730
  TTCAAAACCCACCTCTTCTTTTTCTGTTTTGATCACTCAGAAACCCCAAAAAATACAACT
  :       :  : ::::::::  ::      :  : ::: :: :::::::  : ::  :::::
  T------CTTATCTCTTCTTCATC----ATAGTGACTGAGGAACCCCATCAGATGAAACT
          2510       2520          2530       2540       2550

2740                                  2750       2760       2770
  TT-------------------------CAAACATTTCTGTCTCCCTTTCCCATTTCAATC
  ::                          ::::: :::::::: : :: : :: ::::::::
  TTTAATTTTCTAAAAAAGATTTACTTACAAACGTTTCTGTCACTCTCTGCCGTTTCAATC
    2560       2570       2580       2590       2600       2610

     2780       2790       2800         2810       2820       2830
  TCCAGATTGAAGCACCAGTGATTTATT--TTTGTTTTGTTGATTGATTATTTTGACCATA
  ::::::::::::::  : :  :: ::   :::::::::::  :  ::::::::   ::::
  TCCAGATTGAAGCATTACTAGTTCATCCCTTTGTTTTGTTTCTCAATTATTTT---CATA
    2620       2630       2640       2650       2660          2670

       2840       2850         2860          2870       2880
  ACCAATAAACCATAACAATCGC-AATTCA-GA---AGCTCCAGACGTTCATCGACCCCTT
   :::  ::::::::::::  :: :::::: ::   :::: ::: :::::::: :: ::: :
  TCCATGAAACCATAACAAGGGCTAATTCAAGAGCTAGCTGCAGGCGTTCATGGAACCCCT
       2680       2690       2700       2710       2720       2730

  2890       2900       2910       2920       2930       2940
  TTTCTTATGTTTATTTTATATTACTTCCATCCTGGACTACTCATTTGGACAAAAAAAGTA
  :: ::: :::::::::: :    :::::::: ::  ::: ::: : :  ::: :  : :
  TT-CTTCTGTTTATTTTGT----CTTCCATCATGAGCTATTCAGT-GCTCAAGAGTATTC
        2740          2750       2760       2770       2780

  2950       2960       2970             2980       2990
  TTGCTAAATATGCTATGAGTTGTGCATATATT------ATTCTTGAATTAGT--------
   ::::::::::::::::: :: : : :::::       ::::::::::: :
  CTGCTAAATATGCTATGAATTATCCTTATATATAAATCATTCTTGAATTAATTACTAGCT
   2790       2800       2810       2820       2830       2840

     3000       3010       3020          3030       3040
  AGTATTTTTTTCATTTTATTACATT-TTTTGTGTTGTC---ACTCAGTTTGTGTTTTGGA
  :::: ::    : ::::::::::  : :::: :: ::::   :: ::::::::::::::::
  AGTAGTTCAGTAATTTTATTACTCTCTTTTCTGCTGTCTTCACCCAGTTTGTGTTTTGGA
   2850       2860       2870       2880       2890       2900
```

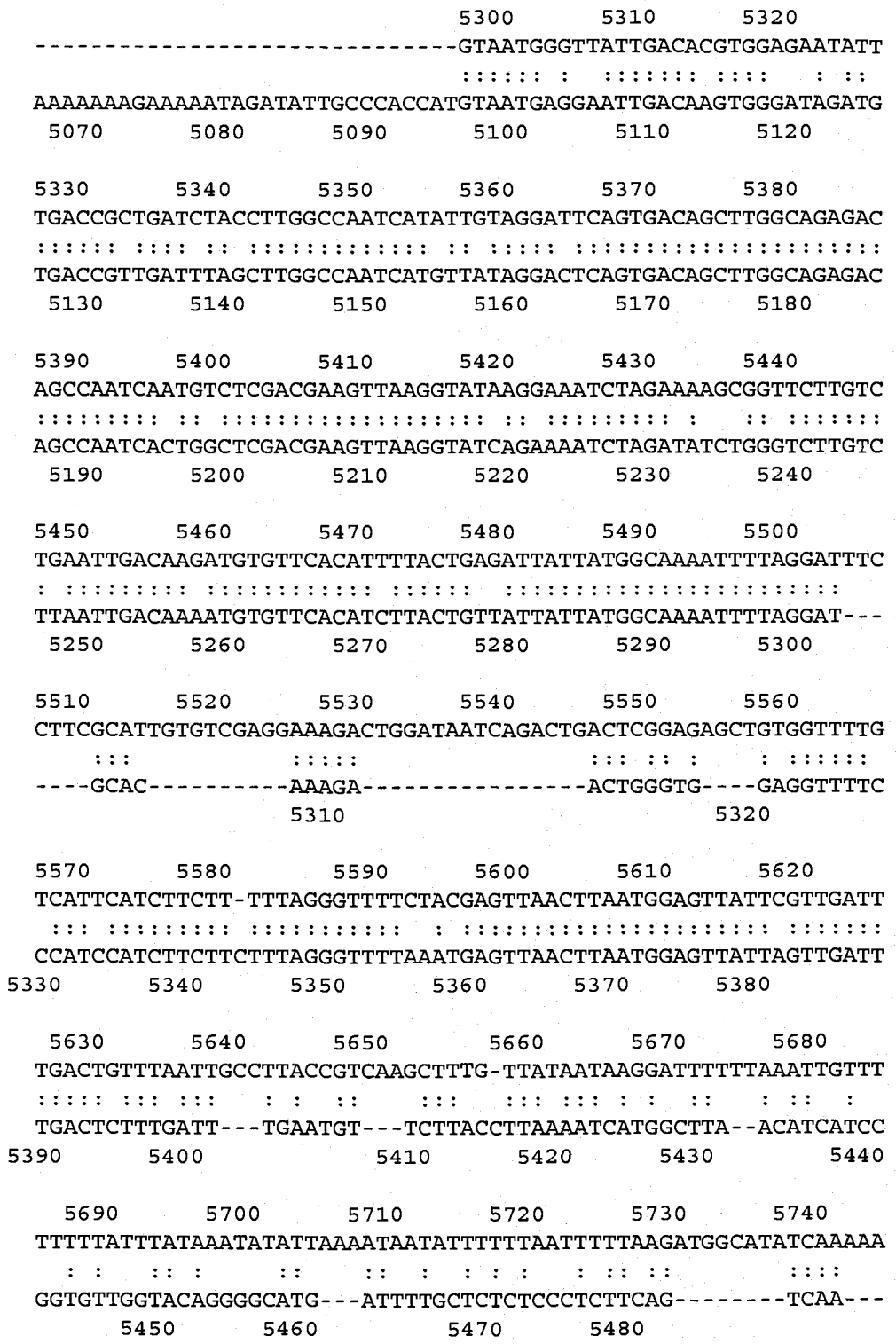

3050 3060 3070 3080 3090 3100
Exon 2 TCAGCTAGCTAGGCTGCAGCTATGGAATATCAAAATGAATCCCTTGAGAGCTCCCCCCTG
TCAGCTAGCTAGGCAGCAGCTATGGCATACCAAAATGAACCCCAAGAGAGCTCTCCCCTG
2910 2920 2930 2940 2950 2960
3110 3120 3130 3140 3150 3160
AGGAAGCTGGGAAGGGGAAAGGTGGAGATCAAGCGGATCGAGAACACCACCAATCGCCAA
AGGAAGCTGGGGAGGGGAAAGGTGGAGATCAAGCGGATCGAGAACACCACCAATCGCCAA
2970 2980 2990 3000 3010 3020
3170 3180 3190 3200 3210 3220
GTCACTTTCTGCAAAAGGCGCAGTGGTTTGCTCAAGAAAGCCTACGAATTATCTGTTCTT
GTCACTTTCTGCAAAAGGCGGAATGGTTTGCTCAAGAAAGCCTATGAATTATCTGTTCTT
3030 3040 3050 3060 3070 3080
3230 3240 3250 3260 3270 3280
TGCGATGCTGAGGTTGCACTCATCGTCTTCTCTAGCCGCGGTCGCCTTTATGAGTACTCT
TGCGATGCTGAGGTTGCACTCATCGTCTTCTCCAGCCGTGGACGCCTTTATGAGTACTCT
3090 3100 3110 3120 3130 3140
3290 3300 3310 3320 3330 3340
AACGATAGGTAAATAAATCTAATTTTAGATATTTGCTTCTCTGGATCTTAAATTCTCCAT
AACAATAGGTATATAC--------TTAG--------TTCCTCGGCTCATGAATTCTCCAT
3150 3160 3170 3180
3350 3360 3370 3380 3390 3400
GTTACAAGCCCTCTATCTTCATGTGGTCACTTTTTTTTTTTTTTTATCTTCCTTTCTGCC
GTTGCAAACCCTCT----TCAAGTGCTCAAAGTTGGTTTTT------CTTGCTTTCTCAT
3190 3200 3210 3220 3230
3410 3420 3430 3440 3450 3460
CCAAAGAGATTTTTTTATCCTCTCTATTTTGCTTATGTTAGTGTTAATTTTTAGCTTTAA
CCAAAGGGATTTGTTTTTTCT-----TTTTGCTTATGTCAGTGTTAATTTTTA--TTGCT
3240 3250 3260 3270 3280 3290
3470 3480 3490 3500 3510
TTGGTTT----CTTTCATTTTCATTTTCTTTCTTTCATGA--ATGATCATT----AAATG
TTGGTTTTGAGCTGTTTCTTTAATTGGTTTTCTTCCATCATCATTTTCTTTCTTCAATTG
3300 3310 3320 3330 3340 3350


3520 3530 3540 3550 3560
GTTTTCAATTTCTAAGGTGGG-AAATTTATTA----------TTATTATTATTA--TTT
GTTTTCAACGTTTGTTGTGGGGAAAAAAAATAGGAGCCTGGTGTCAAGGTTTTTAGCTTC
3360 3370 3380 3390 3400 3410
3570 3580 3590 3600 3610 3620
TGTGTTTAATCTCTGGGTAAAGGATTTAAAGCAAAAGAGACACAATCATTCCTTATGCTG
TGAGCTAGATCTTCGGGT---GTCTTTAAAGTAAAAGA-ACACAATCATTCTTTATGCTG
3420 3430 3440 3450 3460
3630 3640 3650 3660 3670 3680
CAGTTTAGATTGAGTTTCTTATCTAACTGAGATTCACTTGTCTTTCTTTCTT-TCTTTCT
CAGTTTGGATTGAATTTCTTCTCAAAATACAATTCACTTGTCTTTCTTTCTTCTATTTCT
3470 3480 3490 3500 3510 3520
3690
CTTCTCTTAC----------CCTT
TTTCTTTTCCTTGTATAAGCATAATTAATGTTTTGTTTTTCCTTTTCTTTATTTCACCCT
3530 3540 3550 3560 3570 3580
3700 3710 3720 3730 3740 3750
TAGACGATGCTGATGCACACGTTATTTTGAGTTCTTGGTTTGGTAAAAACATAGATCTGG
TAGATGATTGTGATGCATACATGATTTTGAGTTCTTGGT---------ACATAGATCTGG
3590 3600 3610 3620 3630
3760 3770 3780 3790 3800 3810
TATAATAAACAGACATAGAAGCACTAT-ATGAGTGTAGTATGGTAGCAGAAATAAGTATA
TGTATTAGATAGACATAGAAGCACAATTATAAGTGTAATAAGGTAGTAGAAACAAGTAGA
3640 3650 3660 3670 3680 3690
3820 3830 3840 3850
GGTCTG----------TGAGATCAGCCTCTTTATCTCCTCCCTTGTTGTT
GGGCTGGGAAAATGTATGCAGGCATGTGATATCAGCCTCTTTATCTCCTCCCTTGATGTT
3700 3710 3720 3730 3740 3750
3860 3870 3880 3890 3900 3910
AATTTTGTTGTTTCCGTTTTTCTTTCTCTTCCATTATTCCTCTTGCACTCTCTATCTCTC
AAGTTTGCTGTTTCC-TTTTTCTTTCTTTTCCATCATTCCTCTTGAACTCTGCCTCTCTC
3760 3770 3780 3790 3800 3810

```
  3920          3930       3940       3950       3960       3970
  GCTT----TTTTTTTGCACATACTTGTTTGTTTGTGTCATCTACGAGGCTAAAGAGATTG
    ::     :::: :::::::::: ::  :::::: :::::::     ::::::::::::
  CTTTACTCTTTTCTTGCACATACATGCATGTTTGAGTCATCTCTAGGGCTAAAGAGATTA
3820       3830       3840       3850       3860       3870

     3980       3990       4000       4010       4020       4030
  CCTATAGCCAAAGCTGTCATCTTCTCATTAGTCCAAACCCTCC-ATCTCTTTTCACTTCC
  :::::::: :::::::::::::::::::::::::::::::::: ::::::: ::::::::
  CCTATAGCTAAAGCTGTCATCTTCTCATTAGTCCAAACCCTCCCATCTCTTCTCACTTCC
3880       3890       3900       3910       3920       3930

       4040       4050       4060       4070       4080       4090
  TAGTTAAATAGCACGTCAATTAGACATCAAGAAAGCAAAAGTACCATGTCAAAT-AACCG
       :::::::::::: :  ::::: :::::   :: ::::: : ::::::: ::  :
  ----AAAATAGCACGTCAGTCGGACAT-AAGAAGAAAAGAGTACAAAGTCAAATTAAATG
    3940       3950       3960       3970       3980       3990

       4100       4110       4120       4130              4140
  TGAAAAAGAAGAAGAACAAAGAAAGGTTTTTTTAATTTGTCATGTCACT-------CAAA
  ::::::: ::          ::::::: :::::: :: ::::::::::         ::::
  TGAAAAAAAA---------AGAAAGGGTTTTTTTATATGTCATGTCACCAAACACACAAA
     4000               4010       4020       4030       4040

     4150       4160       4170       4180       4190       4200
  CATATATTATTAGGGTTTCAAATCCCAAATCCCCAGATGGGTTT-TTCATCTTATTTTAT
  ::::::::: :::::::::::  :::::::::: :::::::: ::::::::::::::
  CATATATTACTAGGGTTTCAAAATCCAAATCCCCAAATGGGTTTCTTCATCTTATTTTAT
    4050       4060       4070       4080       4090       4100

       4210       4220                          4230       4240
  TTTTCCAAACCAATC-CAGGG------------------TTTTTCCCCTAAT-CACACG
  ::::::::: ::::: :  ::                   ::::::: ::::: ::::::
  TTTTCCAAAACAATCACTAGGATCTCTCAATTTAGGATTCTTTTTCCTCTAATTCACACG
    4110       4120       4130       4140       4150       4160

       4250       4260       4270       4280       4290       4300
  AAATTTCCCAAAATCTCAGTTTGAACCCACGAGGGGATAGTGAAAACCTTTCTGTTAGTC
  :: :::: ::::::::: ::: ::::::::: ::::: :::::::: :::: ::::  ::
  AA-TTTCACAAAATCTCTGTTCGAACCCACGTGGGGAAAGTGAAAAGCTTTTTGTTTTTC
     4170       4180       4190       4200       4210       4220

       4310       4320       4330       4340       4350       4360
  AATGCATAACCCCAGTTAGGGTTCATAGTTAGGGTTCATATTCAAGTAACCACATGAAAT
  :: ::::: :::              ::::::::::::::: :::::::::: :::: :
  AA-GCATAGCCC-------------TAGTTAGGGTTCATATTTAAGTAACCACTTGAAGT
      4230                  4240       4250       4260
```

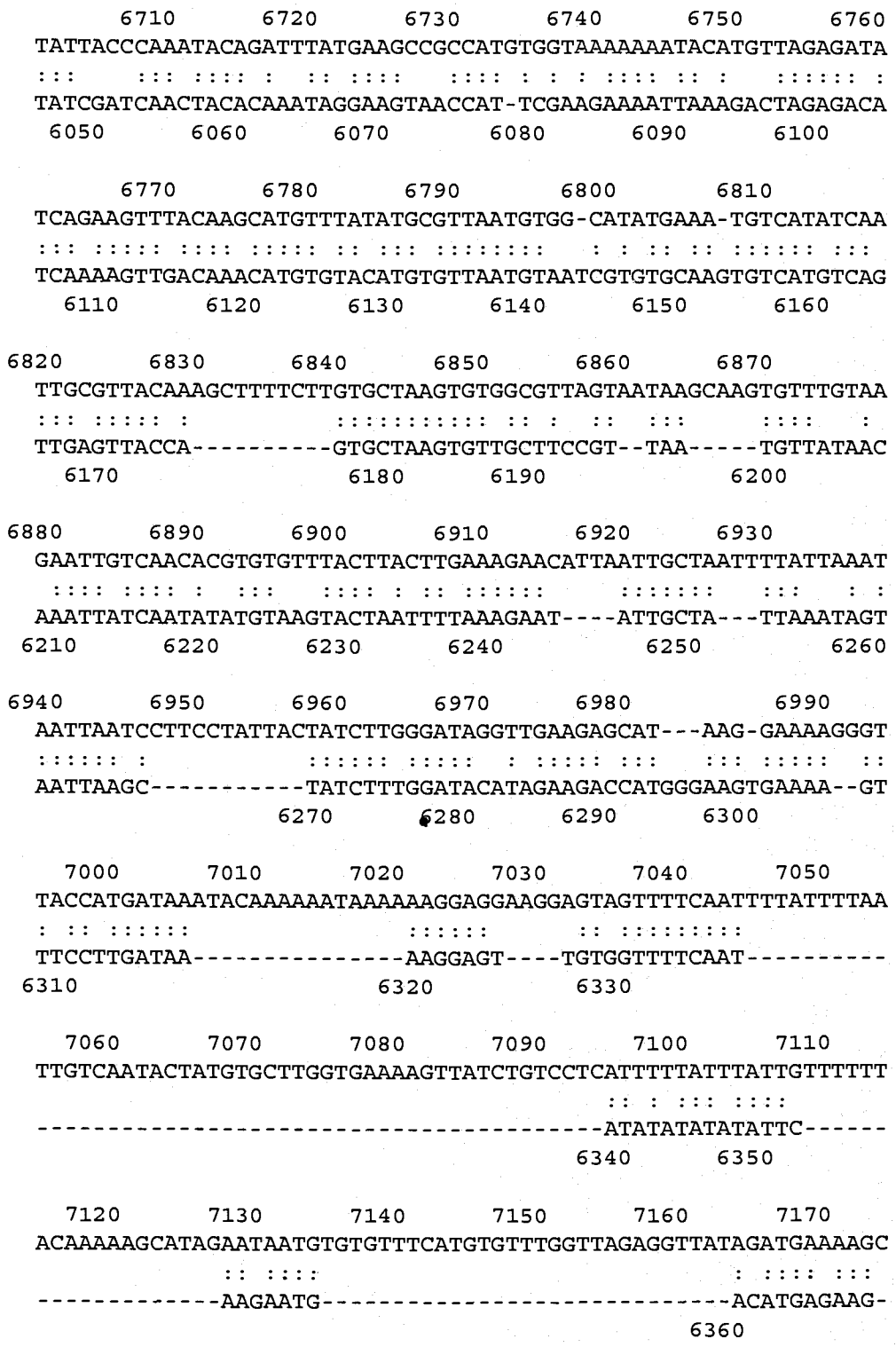
```
      4370       4380       4390       4400       4410       4420
   CATCGAAATCGTACATTAACATTCAAGGAAAACTGTTAAATCAAGCAAGTGGACCCTTCC
   :::: :::: : ::   ::: :: :  : ::::: :: :::::: :: :::::  :::::
   CATCAAAATTGAACCGAAAC-TTTAGTGCAAACTATTCAATCAACCATGTGGATTCTTCC
4270       4280       4290       4300       4310       4320

      4430       4440       4450       4460       4470       4480
   ACAACCAATCAAAACTCA-GTTAGATTTCACCTAGATTTTTACCCCTTTTTTAACCTGGG
   : ::::: :::::: : : :::::::::::::::::::::::::::  :::::::::: ::
   ATAACCAGTCAAAAATTAAGTTAGATTTCACCTAGATTTTTACCC--TTTTTAACCTCGG
4330       4340       4350       4360       4370       4380

      4490       4500       4510       4520       4530       4540
   TAAGTATGGTACAGTAATCGGTTAGGGTTTAGTAGCCAGTCAAATAGATCAGATTGTTGT
   :::: : ::::::::::  :::::::::::: :::::::  :::: ::::::::::::
   TAAGAAGGGTACAGTAACTGGTTAGGGTTTAATAGCCAGTTCAATATATCAGATTGTTGT
   4390       4400       4410       4420       4430       4440

      4550       4560       4570       4580       4590       4600
   TCGGGTTTATGAACAGAATCTTTGGTAACGTCACACACGATTTTTCAGTTCTTGCCTACT
   :  :::::::::: :::::::::::: :::::::::::::::::::::::::::: :::::
   TTTGGTTTATGAAAAGAATCTTTGGTCACGTCACACACGATTTTTCAGTTCTTGACTACT
   4450       4460       4470       4480       4490       4500

      4610       4620       4630       4640       4650       4660
   GACAAAAGGCTTTATGTCATGATTCCTTAAACTGAACCCAAGATTTTTAACTTCCGATCC
   ::::::::: :: : :::::::::: : ::: ::::::  : ::::: :::: :: :::
   GACAAAAGGGTTCAAGTCATGATTCATGAAAATGAACCATAAATTTTGAACTCCCAATC-
   4510       4520       4530       4540       4550       4560

      4670                  4680            4690       4700
   CCCTGGAAAAAATATGAA---------ATTCCA-----AAAATTGTCCATTTCTTCTCCT
      :: :::::: : :::          :: :::      :: ::::::::::::::::::
   ---TGCAAAAAAAAAGAAGAAGAAGCAATACCACACAGAATATTGTCCATTTCTTCTCCT
       4570       4580       4590       4600       4610       4620

   4710       4720       4730       4740       4750       4760
   TAGATCTCTCTCTATCTCTCTCCCGGTTAAATTGTTTCC-ATGGTGAAAGCAGAGAGATG
   :::::: ::: :: :::::       ::::  :: ::::: :: :::::::   :::::::
   TAGATCCCTCCCTCTCTCT------GTTATTTTCTTTCCCATAGTGAAAG---AGAGATG
       4630       4640             4650       4660          4670

    4770       4780       4790       4800       4810       4820
   GATCAATGAGAATGGGTTAACCAAGGCCATAATGATGGCACTGTTTAAGATC-TTGTATA
   :: ::: ::::: :::::: : :::: ::: :::::: :: ::::   : :: ::::
   GAACAACGAGAAAGGGTTAGCTAAGGTCATGATGATGCCATTGTT---GGTCATTGT---
      4680       4690       4700       4710          4720
```

```
  4830       4840       4850       4860       4870       4880
GATATATTTATATAAGTTTTTTTTTTTTTTAATTTAAAGAGAGATTTAGCCCCATTTGTA
            : ::: :  :::    :::    :: ::::                ::: :
------------TGAGTGTGGTTTGCGTTTG--TTCAAGATC-------------TTGAA
           4730       4740        4750                    4760

  4890       4900       4910       4920       4930       4940
TTTTTACGGTGAGAAAACACTTTTATAAAAAATTGATATTTTTTTAAAAATTATTTTTTA
:                                    :::: : : ::
T-----------------------------------ATATGTATGTA--------------
                                          4770

  4950       4960       4970       4980       4990       5000
TATTTTTTAGATTATTTTTATGTGTTAATATTAAAAATAAATTTTTTAAAATATAAAAAA
                  : :::::::   :::                        ::   ::
----------------TGTATGTGTATGTATG--------------------TACGCAAGT
                     4780                          4790

  5010       5020       5030       5040       5050       5060
TATTATATTAATATATTTTAAATAAAAAATTAACCGTTGATGACAATATTGAGAGAAAGA
: :: ::  :: : :   : :::: : :                          ::::::::::
TCTTTTAGGAAGAGAGAGTTAATACAGA---------------------GAGAGAAAGA
4800       4810       4820                          4830

  5070       5080       5090       5100       5110       5120
GAGTCGTGAAGAGAGAATGAACGACAACTGTTAACCAGTGGAAGAGTTCTGTCAATTTTG
::        :::::  ::: :::::::: :: :: ::: ::: ::: ::::::::::::::
GA-------AGAGACGATGTACGACAAGTGCTAGCCAATGGGAGACTTCTGTCAATTTTG
         4840       4850       4860       4870       4880

  5130       5140       5150       5160       5170       5180
GTTTCTTCTATGTAA-TAGAAAGCCTACAACTCTAGCTGGTATTGTACGGCTCTGCTTCT
: :: ::  :::::: ::::: :: :: ::::::::: : :  ::  : :::   : :::
GCTTTTTTAATGTAAATAGAA-GCGTAAAACTCTAGCAGCTGCTG--CTGCTGCTCCTCT
4890       4900       4910       4920       4930       4940

   5190       5200       5210       5220       5230       5240
CTCAGAGTTTCAGTCTGAGACTAATAAA-ATGTCCGATTAGTACAATATTTTATTACAA-
:: ::::::::::::  :  : :  ::: :      : :: :    ::: :::::::::
CTGAGAGTTTCAGTCACATTCAAGAAAACAAAAAAAAAATAATTTTTTATCTTATTACAAA
 4950       4960       4970       4980       4990       5000

     5250       5260       5270       5280       5290
TGAAATAGAATATCGAGGTGGGTAATAGAGTGAGTTTAAGGAGATTATCCACTAT-----
:::::::::::::::::::::::::::::  : :  : : :::::::::::: ::
TGAAATAGAATATCGAGGTGGGTAATAGATGGTGGCTCAAGAGATTATCCACCATAGAAG
 5010       5020       5030       5040       5050       5060
```


5300 5310 5320
----------GTAATGGGTTATTGACACGTGGAGAATATT
AAAAAAAGAAAAATAGATATTGCCCACCATGTAATGAGGAATTGACAAGTGGGATAGATG
5070 5080 5090 5100 5110 5120
5330 5340 5350 5360 5370 5380
TGACCGCTGATCTACCTTGGCCAATCATATTGTAGGATTCAGTGACAGCTTGGCAGAGAC
TGACCGTTGATTTAGCTTGGCCAATCATGTTATAGGACTCAGTGACAGCTTGGCAGAGAC
5130 5140 5150 5160 5170 5180
5390 5400 5410 5420 5430 5440
AGCCAATCAATGTCTCGACGAAGTTAAGGTATAAGGAAATCTAGAAAAGCGGTTCTTGTC
AGCCAATCACTGGCTCGACGAAGTTAAGGTATCAGAAAATCTAGATATCTGGGTCTTGTC
5190 5200 5210 5220 5230 5240
5450 5460 5470 5480 5490 5500
TGAATTGACAAGATGTGTTCACATTTTACTGAGATTATTATGGCAAAATTTTAGGATTTC
TTAATTGACAAAATGTGTTCACATCTTACTGTTATTATTATGGCAAAATTTTAGGAT---
5250 5260 5270 5280 5290 5300
5510 5520 5530 5540 5550 5560
CTTCGCATTGTGTCGAGGAAAGACTGGATAATCAGACTGACTCGGAGAGCTGTGGTTTTG
----GCAC----------AAAGA----------ACTGGGTG----GAGGTTTTC
5310 5320
5570 5580 5590 5600 5610 5620
TCATTCATCTTCTT-TTTAGGGTTTTCTACGAGTTAACTTAATGGAGTTATTCGTTGATT
CCATCCATCTTCTTCTTTAGGGTTTTAAATGAGTTAACTTAATGGAGTTATTAGTTGATT
5330 5340 5350 5360 5370 5380
5630 5640 5650 5660 5670 5680
TGACTGTTTAATTGCCTTACCGTCAAGCTTTG-TTATAATAAGGATTTTTTAAATTGTTT
TGACTCTTTGATT---TGAATGT---TCTTACCTTAAAATCATGGCTTA--ACATCATCC
5390 5400 5410 5420 5430 5440
5690 5700 5710 5720 5730 5740
TTTTTATTTATAAATATATTAAAATAATATTTTTTAATTTTTAAGATGGCATATCAAAAA
GGTGTTGGTACAGGGGCATG---ATTTTGCTCTCTCCCTCTTCAG--------TCAA---
5450 5460 5470 5480

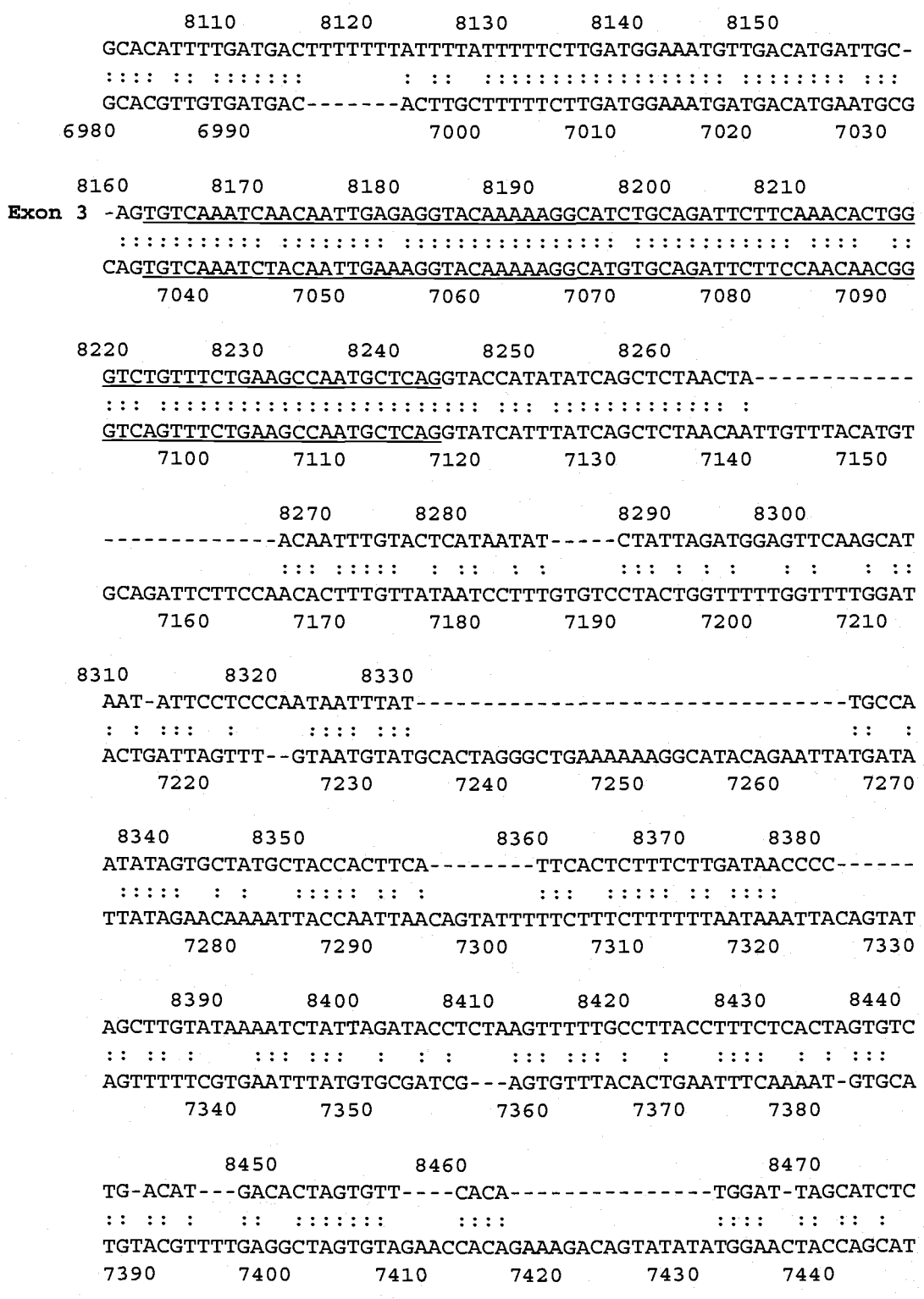
```
  5750       5760       5770       5780       5790       5800
TATTTTAAAAAATAAAAAAATAATTTGAAATAAAACAAAAATTAATTTTTTTAAAACAAT
   ::: :   : :        :: ::::: :::           ::::::: ::
---TTTCATCATT-------TATTTTGATATA----------TAATTTTCTT--------
   5490               5500                 5510

  5810       5820       5830       5840       5850       5860
ATTTTTAACGCAATAACAAATTCTTAATCTTTTACTCATATATCTTAAATTTACGAGAGT
  :::      :  :::      : :::: : : ::::   : ::::: :: :::: ::::
--TTT-----CCCTAA-----TGTTAAGCCTCTACTGTCACATCTTTAAATTACTAGAGG
5520                 5530       5540       5550       5560

  5870       5880       5890       5900       5910       5920
TTTTTCCAAAAAGATAAAGAGATATATGTAAGCGATAAAGTATTAGTAACCTCACATAAA
                          ::::::::  :::::     ::::::: :::::::::
G---------------------ATATGTAACTGATAAC----TAGTAACTTCACATAAA
                         5570       5580          5590       5600

  5930       5940         5950       5960       5970         5980
ATAATGTACAATAATAGAT--AAAAACTAAATTTTATATAAAAATTGAATT-TCAATCCA
::: :::::::::: :::::  :::::   :::::::::::::: ::::: : :::::
ATACTGTACAATA-TAGATTTAAAAAAGGAATTTTATATAAAATTTGAACTCTCAAT---
      5610        5620       5630       5640       5650

     5990       6000       6010       6020       6030       6040
CTTTCTTTTTTCGTGGATCATAAGGAGTTGGACTTGCTTTTTTCACGGTAATTTGACCAA
 ::: :::: :  : : :        : :::::  ::            ::: :::: : ::
-TTTATTTTATTTTTGTT-------ATTTGGA--TGA----------GTAGTTTGTCTAA
 5660       5670              5680                  5690

     6050       6060       6070       6080       6090       6100
AGAAAGAGTTAATACAAATAATATTAATTAAGATATTATCTCTTGTTGTTTGTTCTTGTT
:: :: :: :::::   :   :::::    ::: : :  :::: ::::::
AGCAATAGCTAATATGGAGGGTATTA----AGAAACTGCCTCTAGTTGTT----------
5700       5710       5720           5730       5740

     6110       6120       6130       6140       6150       6160
TTGAAATAATTTAGTTTTTTTTTTAAGAAAAAAAAGTTTTTCCAATACATAAGCAATACA
                          :: : :::::  ::          :::::
-------------------------GACACAAAAGCCTTG---------AAGCAC----
                                5750               5760

     6170       6180       6190       6200       6210       6220
AAAGTGTTTGAACATGGTAATTCTTCTTCTTCTTAGTTGACCAAATTACATTTGGTAGAC
   :: ::: : :  : ::::::     :::::: : ::  :   :  ::: : : : :
---GTATTTTA-CTCGCTAATTCACA--CTTCTTGGCTGTGCTCTTACCATCTTGGAAAA
      5770       5780       5790       5800       5810
```

```
      6230       6240       6250       6260       6270       6280
TAAA-GTTGTTCATATATATGCTACCATTGATAGAGTCATTGGCCAATTATATGTTTTTA
:::: :    :::: : :                                : :: ::: :::
TAAATGGATTTCAAAAA----------------------------AGTACATGGTTTC-
5820       5830                                    5840

      6290       6300       6310       6320       6330       6340
CGTCATTATATTTGAATTCTTTTGTTAATAGTAATTATTAATCACTGAAGTTATTGCATT
  : : : :::::::: :  :::          ::: :: ::       ::
--TTAATCTATTTGAAATGATTT---------AATAATGAA------AA-----------
 5850       5860                5870

      6350       6360       6370       6380       6390       6400
CTTGTCAGCTGATAAACTCCAAGTTGTAATTTTATGTTTGATCTTGTAATTAAGAGCAAG
           ::::: :: ::  :::::
-----------ATAAAATCAAA--TGTAA------------------------------
         5880       5890

      6410       6420       6430       6440       6450       6460
CCAGGAGGACATCTCTAGTGTTCGAGGAAATTGACAAAATTTGCTTCCTCAAATATATTT
                 ::: ::               :::: :          ::::
----------------AGTCTT--------------AAATGT---------AATA-----
                 5900                              5910

      6470       6480       6490       6500       6510       6520
TTGTTTTTCATTGGACAAAAATACATGTTATATATATATATATATATATATATATATATA
                 ::::::: :: ::
----------------AAAAATA-ATTTT------------------------------
                      5920

      6530       6540       6550       6560       6570       6580
TATATATATATAATGCCTATATTTTGTGAGTAGTTCCATAAGTTTAGGATATGTTTGAGG
                 : :: :::: :::::           ::::       : :::
----------------CGATGTTTTTTGAGT----------GTTT------TTTTT----
                   5930                5940

      6590       6600       6610       6620       6630       6640
TAGTTTAACATAAGCATTTGATTTTTTTTTTCAATCCTTATATCAAAATTATCATAAAAC
   ::::  ::      :::::  : ::       ::  ::::::    ::    :::::::
---TTTATAAT-----TTTGAAATCTT-----CATATTTATATGTCCAT----ATAAAAC
  5950            5960            5970       5980           5990

      6650       6660       6670       6680       6690       6700
AATTAAAAAATCATTAATTTATTTTATTTTTTTAATTAAAAAAAACACTTATAAACACAG
:    ::::::::: ::::       ::: ::        : ::::::: :: : :  ::
ATAGGAAAAATCATCAATTCTCAAAATTATTGG-----AGAAAAACACCTACATATGCAT
      6000       6010       6020            6030       6040
```

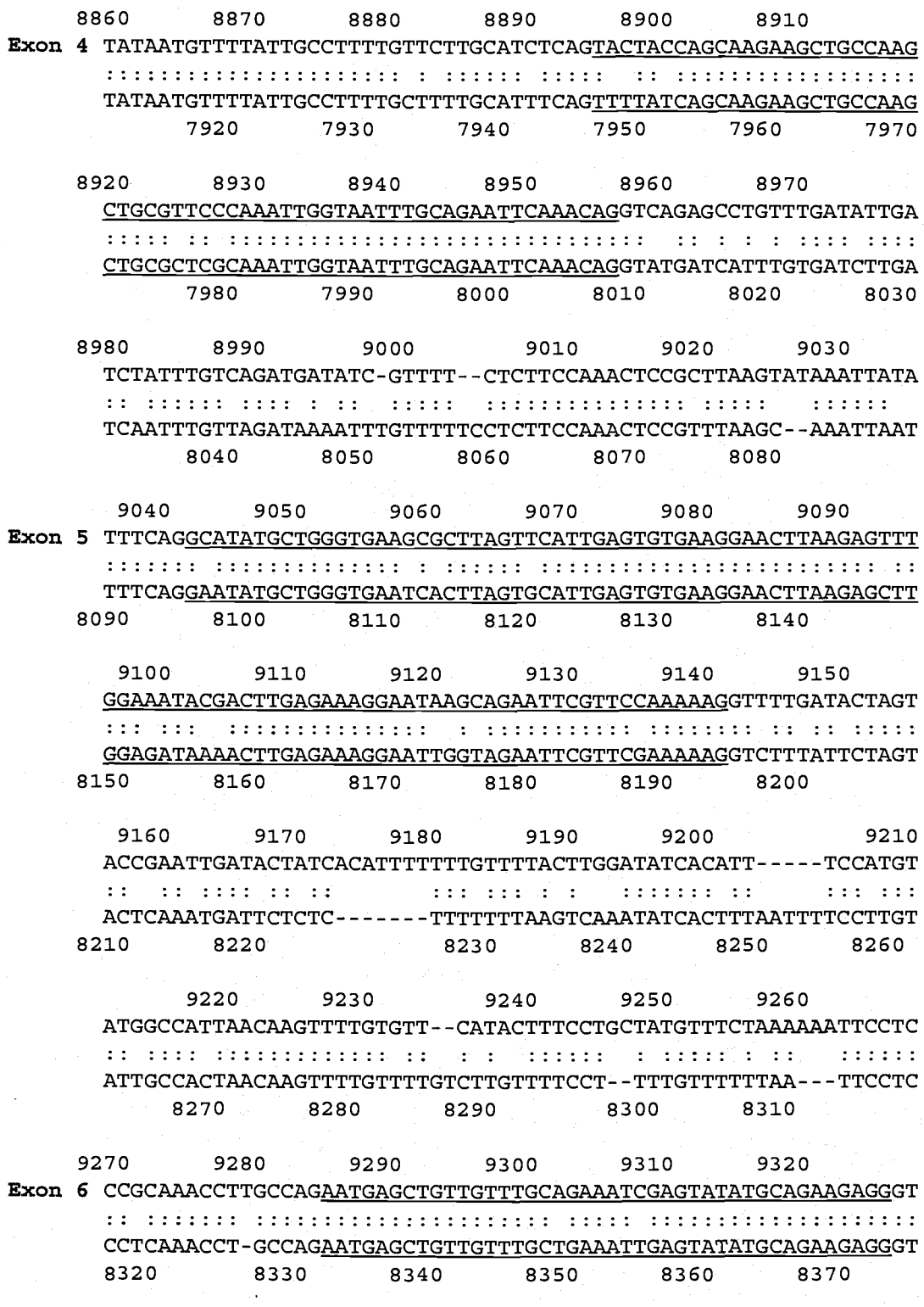

6710 6720 6730 6740 6750 6760
TATTACCCAAATACAGATTTATGAAGCCGCCATGTGGTAAAAAAATACATGTTAGAGATA
TATCGATCAACTACACAAATAGGAAGTAACCAT-TCGAAGAAAATTAAAGACTAGAGACA
6050 6060 6070 6080 6090 6100
6770 6780 6790 6800 6810
TCAGAAGTTTACAAGCATGTTTATATGCGTTAATGTGG-CATATGAAA-TGTCATATCAA
TCAAAAGTTGACAAACATGTGTACATGTGTTAATGTAATCGTGTGCAAGTGTCATGTCAG
6110 6120 6130 6140 6150 6160
6820 6830 6840 6850 6860 6870
TTGCGTTACAAAGCTTTTCTTGTGCTAAGTGTGGCGTTAGTAATAAGCAAGTGTTTGTAA
TTGAGTTACCA----------GTGCTAAGTGTTGCTTCCGT--TAA-----TGTTATAAC
6170 6180 6190 6200
6880 6890 6900 6910 6920 6930
GAATTGTCAACACGTGTGTTTACTTACTTGAAAGAACATTAATTGCTAATTTTATTAAAT
AAATTATCAATATATGTAAGTACTAATTTTAAAGAAT----ATTGCTA---TTAAATAGT
6210 6220 6230 6240 6250 6260
6940 6950 6960 6970 6980 6990
AATTAATCCTTCCTATTACTATCTTGGGATAGGTTGAAGAGCAT---AAG-GAAAAGGGT
AATTAAGC-----------TATCTTTGGATACATAGAAGACCATGGGAAGTGAAAA--GT
6270 6280 6290 6300
7000 7010 7020 7030 7040 7050
TACCATGATAAATACAAAAAATAAAAAAGGAGGAAGGAGTAGTTTTCAATTTTATTTTAA
TTCCTTGATAA---------------AAGGAGT----TGTGGTTTTCAAT----------
6310 6320 6330
7060 7070 7080 7090 7100 7110
TTGTCAATACTATGTGCTTGGTGAAAAGTTATCTGTCCTCATTTTTATTTATTGTTTTTT
----------ATATATATATATTC------
6340 6350
7120 7130 7140 7150 7160 7170
ACAAAAAGCATAGAATAATGTGTGTTTCATGTGTTTGGTTAGAGGTTATAGATGAAAAGC
------------AAGAATG----------ACATGAGAAG-
6360

```
   7180       7190       7200       7210       7220       7230
TTTAATAATAAATAGTAGCTAAATATACTTCATTGTTTGAGTGGTAGAGGAGATTTTTAA
: :  ::::: :     ::         :: :: ::           : :::: ::
TCTCTTAATACAG----GC--------CTCCAATGA---------AAAGGAAAT------
6370      6380                  6390                6400

   7240       7250       7260       7270       7280       7290
AATTTATGAAGACTACAATTCTCTTTCATTTCAAATAACATCCCTATTTTAGTGGTGAGA
       ::::: ::  :::          ::: ::: ::         ::: : :   : :
-------GAAGAATA--ATT---------TTCCAATTAC--------TTTCGAGTAAAAA
           6410                  6420               6430

   7300       7310       7320       7330       7340       7350
TTAATGTATTTGTTTCTCTTTTTCTATTTTCTTTTATCAATATTATATATAAAACTAAAA
 : :: ::: :          :   :::::::::: :             :::::  ::::
GTTATCTATGT---------TCAATATTTTCTTTTCTT-----------TAAAA--AAAA
6440               6450      6460                  6470

   7360       7370       7380       7390       7400       7410
TGCATCAGTGTTTTACTATGGATTGATCATAATGCAATTCACTATAAAATAATTGATGCT
                     ::  :: :: :::             : :: :::  ::
--------------------GAGAGAACAGAAT-----------AGAAGAATGTATAGA
                       6480                   6490       6500

   7420       7430       7440       7450       7460       7470
TCCCTTAAAAAACCAAATAATTAAACAAACACTCAGGGTTAATTTTGTATTTTCATATCT
::  :                            ::    :::  ::::::  ::: ::: ::
TCTGT-------------------------CT----GTT--TTTTGTTCTTTGATAACT
                                  6510        6520       6530

   7480       7490       7500       7510       7520       7530
TTATTGCATAGTGTAATTATTTCTATGTCCTTGAAAAAAGAAAAAAAACACTAGGGTTTT
     : ::::::::  : ::::  : :: : : :: : :  : :       :  :   ::
A----GAATAGTGTA--TGTTTCAGTTTCTTGGTAAGAGGTTAGA------TGTGAAGTT
          6540        6550       6560            6570

   7540       7550       7560       7570       7580       7590
TTTAAAAAAGTTTCATATTTTTTTGTATAGTGTAATTATCCCACTTTTGGGGCCAACTTT
 ::: :: : : : : :                 :: :::   ::::       :: : ::
CTTATAATACTATAACA---------------AAATAT---ACTT-------CATCATT
6580      6590                    6600                 6610

   7600       7610       7620       7630       7640       7650
TTTTTACCTAAGGTAAAGGGGTATTTTTGGTTTTTTTATGTTTGTTTTTTTGCAATTATT
:                 ::::::
TG-------------AGGGGT-------------------------------------
```

```
  7660       7670       7680       7690       7700       7710
ATATGGGATCAAGAGTGTTATGATCTTTTTATATAAAAAAAAAATGGTTGACACGTGATC
     :::   :           ::::::        : ::: ::: :      ::    ::: :
-----GGAGGA----------GATCTT-------ACAAACAAACT-----AC---TGAAC
                      6630              6640              6650

  7720       7730       7740       7750       7760       7770
TACAATTCCCCCTCCCTTTTCATTCCTAACCTTGAAAGTCTTAGTGAAACATATAGTTAT
  ::::       ::::::::   ::                   :::::  :::::
C-CAAT------TCCCTTTTACTT----------------TTAGT--AACATCCC-----
              6660                        6670        6680

  7780       7790       7800       7810       7820       7830
AATAAAGAAATATTATCTCTAGTTTTGCAAATTAATTTCATAACATCAATTAAATATTCT
         ::: ::        :: :::                          : ::::: :
--------AATTTT------GGTGTTG-----------------------AGATATTGT
                     6690                                6700

  7840       7850            7860       7870       7880       7890
GATAAGGTAAAGTTATTTAGGATGG----AGAAATTTACATAATGAAGCCTCCTTCTGCC
:::::::::::::::: ::    :::     :: :::::::  ::::::::: ::::: ::
GATAAGGTAAAGTTATATACTTTGGCCAAAGTAATTTACAGGATGAAGCCTTCTTCTCCC
    6710       6720       6730       6740       6750       6760

      7900       7910         7920
TGAGTAGTGCATTTCTATGGTATTTA--TGAG-----------------------CATC
::: :::::  :::::  :::::: :  : ::                         ::::
TGATTAGTGACTTTCTGAGGTATTAACATCAGAGTTCTGAAGATTATCCAAAAAAACATC
    6770       6780       6790       6800       6810       6820

    7930       7940       7950       7960       7970         7980
AA--TTCTACAATCCATTGAAGCAAAAGAACTAACCTTCTTGAAACCCTC--TTGCAGAT
:   ::::  :::  :::   :: : :::  :::  ::::::::: ::::  ::::::::
AGAGTTCTGAAATTTATTAGGGCCAGAGATTTAATATTCTTGAAATCCTCAATTGCAGAT
    6830       6840       6850       6860       6870       6880

      7990       8000       8010       8020       8030       8040
AATTGTGAGTGAATGTAAGTCCACTACGAAATATTCACACGATTACGCACTTAGTTATCA
:::: :::   ::: ::::  ::  : :::::: :::: :::    :::::: :::: :
AATTATGA---AATTTAAGATCAAAATGAAATATCCACAGGAT----CACTTATTTATGA
    6890         6900       6910       6920         6930

      8050       8060       8070       8080       8090       8100
TTAAACTTTGTTTTTGGTGCTTTGCATTTTCTTAATTAGATTCTTCCACAGCTTTCCAAT
 :::: ::::::::::: :                 : : ::::::::::   ::  :::
CTAAAATTTGTTTTTGGGG---------------TGACATTCTTCCACATTATTAGAAT
 6940       6950                    6960       6970
```


8110 8120 8130 8140 8150
GCACATTTTGATGACTTTTTTTATTTTATTTTTCTTGATGGAAATGTTGACATGATTGC-
GCACGTTGTGATGAC-------ACTTGCTTTTTCTTGATGGAAATGATGACATGAATGCG
6980 6990 7000 7010 7020 7030
8160 8170 8180 8190 8200 8210
Exon 3 -AGTGTCAAATCAACAATTGAGAGGTACAAAAAGGCATCTGCAGATTCTTCAAACACTGG
CAGTGTCAAATCTACAATTGAAAGGTACAAAAAGGCATGTGCAGATTCTTCCAACAACGG
7040 7050 7060 7070 7080 7090
8220 8230 8240 8250 8260
GTCTGTTTCTGAAGCCAATGCTCAGGTACCATATATCAGCTCTAACTA------------
GTCAGTTTCTGAAGCCAATGCTCAGGTATCATTTATCAGCTCTAACAATTGTTTACATGT
7100 7110 7120 7130 7140 7150
8270 8280 8290 8300
-------------ACAATTTGTACTCATAATAT-----CTATTAGATGGAGTTCAAGCAT
GCAGATTCTTCCAACACTTTGTTATAATCCTTTGTGTCCTACTGGTTTTTGGTTTTGGAT
7160 7170 7180 7190 7200 7210
8310 8320 8330
AAT-ATTCCTCCCAATAATTTAT--------------------------------TGCCA
ACTGATTAGTTT--GTAATGTATGCACTAGGGCTGAAAAAAGGCATACAGAATTATGATA
7220 7230 7240 7250 7260 7270
8340 8350 8360 8370 8380
ATATAGTGCTATGCTACCACTTCA--------TTCACTCTTTCTTGATAACCCC------
TTATAGAACAAAATTACCAATTAACAGTATTTTTCTTTCTTTTTTAATAAATTACAGTAT
7280 7290 7300 7310 7320 7330
8390 8400 8410 8420 8430 8440
AGCTTGTATAAAATCTATTAGATACCTCTAAGTTTTTGCCTTACCTTTCTCACTAGTGTC
AGTTTTTCGTGAATTTATGTGCGATCG---AGTGTTTACACTGAATTTCAAAAT-GTGCA
7340 7350 7360 7370 7380
8450 8460 8470
TG-ACAT---GACACTAGTGTT----CACA---------------TGGAT-TAGCATCTC
TGTACGTTTTGAGGCTAGTGTAGAACCACAGAAAGACAGTATATATGGAACTACCAGCAT
7390 7400 7410 7420 7430 7440

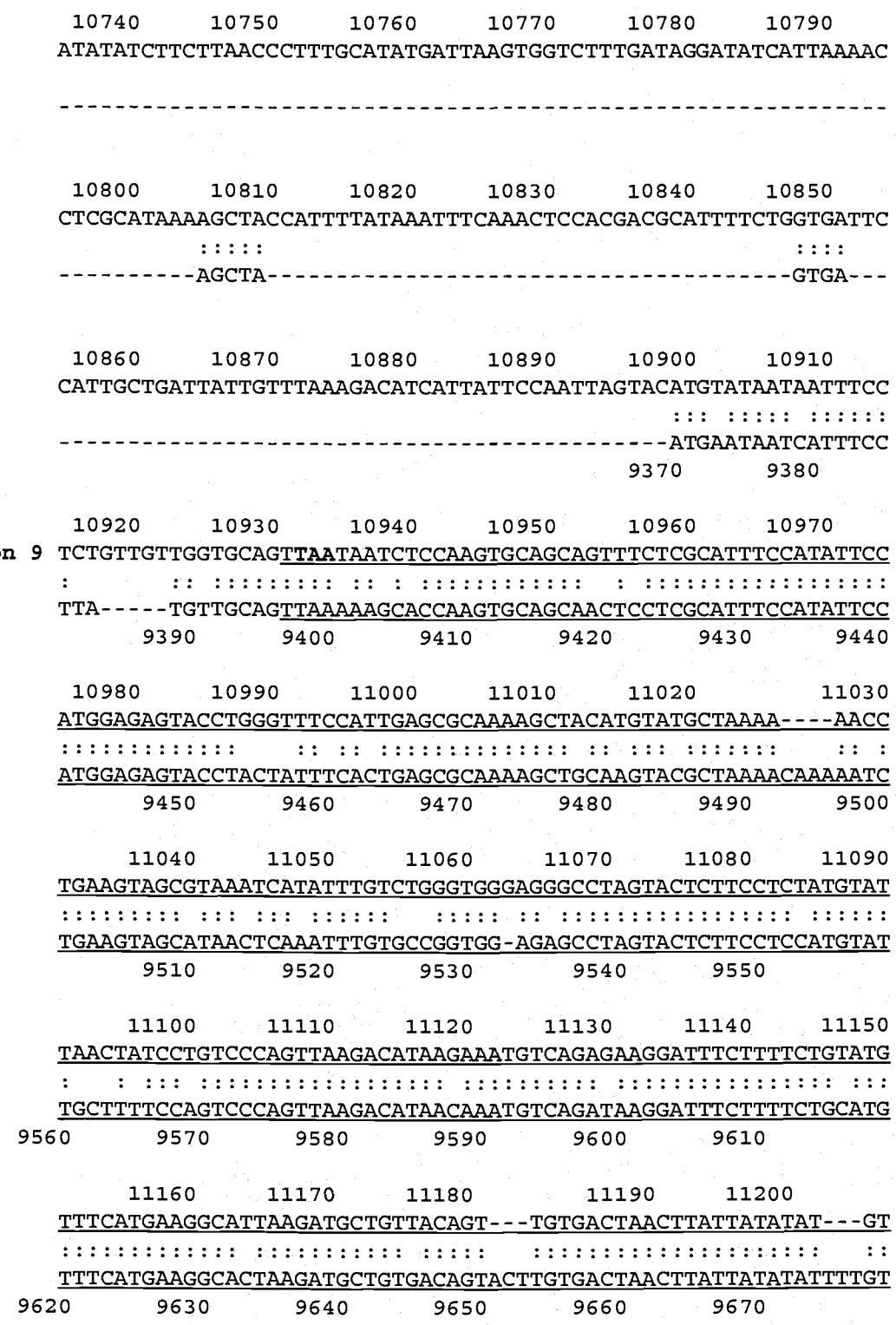
```
8480      8490      8500      8510      8520      8530
GGAGTTGAAGGTTGTCTGGCTTCTTCGAANATCCAGGGTTTTCAAGAAGGTTTGTA----
  :    ::   : : :      ::  : . :: : : ::: ::  ::  : : :
ATAACAAAATCCTTTTTATGAAATTTTATCGTCGATGTTTTACACTAAATTCTCTCACTA
7450      7460      7470      7480      7490      7500

     8540      8550      8560      8570                8580
--CATTGGGAGGCCCGTGGTTATAAACCTACTGTGTAAATG---------GTTTGATAAA
  ::::   ::   :::  :::  ::: : :::: :::::          :::  :  ::
TTCATTAACAG---CGTAATTAACAACATGCTGT-TAAATTATAGAAGGAGTTCAAGCAA
7510         7520      7530       7540      7550      7560

             8590          8600      8610          8620
TA---------ATGATTCATC----AGATTTGAGTAATAGTCTTT----TAATTTC-TTT
::         :: ::: ::     : :::::   ::::  ::::    :::  :: :::
TATTCCTAACAATCATTTATTGCCAATATTTGTCAAATACCCTTTGTGATAACCTCATTT
    7570      7580      7590      7600      7610      7620

                                 8630                8640
GTA--------------------------AATGTTG--TCT--------ATGTTTTTTC
::                               ::: :::  :::        : ::::::::
GTGTAAAATCGATTAAATACATACCTATTTAATTTTGCTTCTCAGAGTGGAGGTTTTTTC
    7630      7640      7650      7660      7670      7680

           8650      8660      8670      8680      8690
C------------AGTCCTCCCTACACACACTCTGATAATTATAACCAATTTTGT-----
:            :::: :   :  : ::   :::   :::::: ::: ::::::
CTACTGCATTGGGAGTCATGAGTGTAAAC---CTGC--ATTATAGCCAGTTTTGTGTACA
    7690      7700      7710           7720      7730

              8700      8710      8720          8730
--------TTCGCTTCCTCCTTTCGCTATGCTCCTACTG----AATTTATTTCCAGTTTG
        ::  ::::::: : : ::: ::  :::: ::    ::::::::::::::::
GAAACCCTTTTCCTTCCTC-TGTTGCTGTGGCCCTATTGTATCAATTTATTTCCAGTTTG
7740      7750       7760      7770      7780      7790

8740      8750      8760      8770      8780      8790
ATTCAGTATTATATGCATGTTTACAAGAAAATAGAAGGGGGGAATCTACATCACTGAGAT
:::: ::::::::: ::::::: ::::::  :: ::: : : ::: :::::::::: ::
ATTCGGTATTATATACATGTTTCCAAGAAG-TATAAGAGAGAAATGTACATCACTGATAT
7800      7810      7820       7830      7840      7850

8800      8810      8820      8830      8840      8850
TTTCTACCTGTATTTTATCAACTGATCTAATATGAACTTGAGGCTCTTAATTTTGTTATA
::::::: : ::::::       : :::::: :::::: :::: ::::::: : :::::
TTTCTACTTATATTTTG-----AGTTCTAATCTGAACTCGAGGATCTTAATCTAGTTATT
 7860      7870           7880      7890      7900      7910
```


8860 8870 8880 8890 8900 8910
Exon 4 TATAATGTTTTATTGCCTTTTGTTCTTGCATCTCAGTACTACCAGCAAGAAGCTGCCAAG
TATAATGTTTTATTGCCTTTTGCTTTTGCATTTCAGTTTTATCAGCAAGAAGCTGCCAAG
7920 7930 7940 7950 7960 7970
8920 8930 8940 8950 8960 8970
CTGCGTTCCCAAATTGGTAATTTGCAGAATTCAAACAGGTCAGAGCCTGTTTGATATTGA
CTGCGCTCGCAAATTGGTAATTTGCAGAATTCAAACAGGTATGATCATTTGTGATCTTGA
7980 7990 8000 8010 8020 8030
8980 8990 9000 9010 9020 9030
TCTATTTGTCAGATGATATC-GTTTT--CTCTTCCAAACTCCGCTTAAGTATAAATTATA
TCAATTTGTTAGATAAAATTTGTTTTTCCTCTTCCAAACTCCGTTTAAGC--AAATTAAT
8040 8050 8060 8070 8080
9040 9050 9060 9070 9080 9090
Exon 5 TTTCAGGCATATGCTGGGTGAAGCGCTTAGTTCATTGAGTGTGAAGGAACTTAAGAGTTT
TTTCAGGAATATGCTGGGTGAATCACTTAGTGCATTGAGTGTGAAGGAACTTAAGAGCTT
8090 8100 8110 8120 8130 8140
9100 9110 9120 9130 9140 9150
GGAAATACGACTTGAGAAAGGAATAAGCAGAATTCGTTCCAAAAAGGTTTTGATACTAGT
GGAGATAAAACTTGAGAAAGGAATTGGTAGAATTCGTTCGAAAAAGGTCTTTATTCTAGT
8150 8160 8170 8180 8190 8200
9160 9170 9180 9190 9200 9210
ACCGAATTGATACTATCACATTTTTTTGTTTTACTTGGATATCACATT-----TCCATGT
ACTCAAATGATTCTCTC-------TTTTTTTAAGTCAAATATCACTTTAATTTTCCTTGT
8210 8220 8230 8240 8250 8260
9220 9230 9240 9250 9260
ATGGCCATTAACAAGTTTTGTGTT--CATACTTTCCTGCTATGTTTCTAAAAAATTCCTC
ATTGCCACTAACAAGTTTTGTTTTGTCTTGTTTTCCT--TTTGTTTTTTAA---TTCCTC
8270 8280 8290 8300 8310
9270 9280 9290 9300 9310 9320
Exon 6 CCGCAAACCTTGCCAGAATGAGCTGTTGTTTGCAGAAATCGAGTATATGCAGAAGAGGGT
CCTCAAACCT-GCCAGAATGAGCTGTTGTTTGCTGAAATTGAGTATATGCAGAAGAGGGT
8320 8330 8340 8350 8360 8370

```
9330      9340      9350      9360      9370      9380
AATGCTTCTTATGTTATCACATTTCCCATTTATTTAATATTTATTGTTTTCTGGTGGAGT
::  :  ::: :::  ::: :::  ::::                 :  ::::
AA--CAACTTTTGTGCTCATATTCACCAT----------------GACTTCT--------
   8380      8390      8400

9390      9400      9410      9420      9430      9440
ATATTCTATATGATTGTTATATATTCTGAGGTAAAAGTCATCTAGTGTTTATTAACATAA
    :::::  :::               : :::::   ::: ::: :::   ::  :::
----TCTATTTGA----------------GATAAAAAA-ATCAAGTTTTTGCCAATTTAA
 8410                       8420      8430      8440

9450      9460      9470      9480      9490      9500
TGATTCTATGGTCAACTTATTCCTTCCTGTTTTCACTCCGAGATTTTCCTTTGATTCCTT
:::: ::::::: ::: : :::  :  : ::::::::::: : : ::: ::::::::: ::
TGATCCTATGGTGAACCTCTTCTATTGTATTTTCACTCCAAAAATTTTCTTTGATTCATT
8450      8460      8470      8480      8490      8500

9510      9520      9530      9540      9550      9560
Exon 7 GAATGAAAATGCACATTACAGGAGGTTGACTTGCACAACAATAACCAGCTTCTCCGAGCA
:::::::::::::: ::: :::::: ::::::::::::::::::::::::::::::::::::
GAATGAAAATGCAAATTGCAGGAGATTGACTTGCACAACAATAACCAGCTTCTCCGAGCA
8510      8520      8530      8540      8550      8560

9570      9580      9590      9600      9610      9620
AAGGTCTTTCTTCTATCTATCTATTTATCCATCTCGAGTGAGGGCAAGGATGCGTGCGTG
:::::::::::  ::   ::::::::::::  ::           ::       : :: :::
AAGGTCTTTCTACT---TATCTATTTATCAAT----------GC-------CTTGTGTG
8570      8580        8590                        8600

9630      9640      9650      9660      9670      9680
TGCATGAATGAAGATCTCTATGTCTTATATCGTTAGTGAGCTGTTTATAATTTAG-AAAT
::  ::::    :::::  :: ::::: :::::: ::: : ::::: : :::::: ::::
TGTCTGAACTTGGATCTTAATATCTTAGATCGTTGGTGGGTTGTTT-TTATTTAGTAAAT
8610      8620      8630      8640      8650       8660

9690               9700      9710      9720      9730
Exon 8 ATGA----------GGCTTATCTTGATAGTGCAGATTTCAGAGAATGAAAGAAAGCGACA
::::          ::::::: :::::  :::::::: ::::::::::::::::::::::
ATGACACTACGTGGGGCTTATGTTGATGTTGCAGATTGCAGAGAATGAAAGAAAGCGACA
8670      8680      8690      8700      8710      8720

9740      9750      9760      9770      9780      9790
GAGCATGAATTTGATGCCAGGAGGAGCAGACTTTGAGATCGTGCAGTCTCAACCATATGA
:  ::::::::::::::: ::::: :   :::: :::::: ::::::::::::::: :::
GCACATGAATTTGATGCCGGGAGGTGTCAACTTCGAGATCATGCAGTCTCAACCATTTGA
8730      8740      8750      8760      8770      8780
```

```
9800      9810      9820      9830      9840      9850
CTCTCGGAACTATTCTCAAGTGAATGGATTACAGCCTGCAAGTCATTACTCACATCAAGA
:::::::::::::::::::: :::::::: : :::::: : ::::::: : ::: ::::
CTCTCGGAACTATTCTCAAGTTAATGGATTGCCGCCTGCCAATCATTACCCTCATGAAGA
8790      8800      8810      8820      8830      8840

9860      9870      9880      9890      9900      9910
TCAGATGGCCCTTCAGTTAGTGTAAGTATCTCCTTTGTAACGAATAATAGGTTTTCATTA
 ::: :    :::::::::::::::::: ::::::: :: ::    ::: ::::::: :
CCAGCT----CTTCAGTTAGTGTAAGTATTTCCTTTGCAATGAGCTGTAGTTTTTCATCA
8850          8860      8870      8880      8890      8900

9920      9930      9940      9950      9960      9970
ACCGGACAACCAGATTTAGTGTTGTGCATTCATAAAATACAATTAATTACTTTAATTTGG
:           : ::: :: :: ::::         : :::::: :::::: :: :::
ATT----------AATTACTGATGAGCAT---------ATAATTAACTACTTTGATCTGG
              8910      8920              8930      8940

9980      9990     10000        10010        10020     10030
AGATGTTCCAAAAGTTGCAACTGC---ATGGTTCATGGGCTCTAA---TTTCTTGGAAGT
:   ::: ::  ::  ::: : ::   ::::::: ::: :: :::   ::: :::::::
ATGGGTTTCAGTAGCAGCAGCGGCTGAATGGTTCGTGGTCTGTAAAAATTTATTGGAAGG
   8950      8960      8970      8980      8990      9000

       10040     10050     10060     10070      10080
ATATAA---CCGATGCTATGTCTTTTCATTCTCATAATTACT-GATCAGTCCCTTA--TA
::::::   : :::::: :: ::: : :::::::::::: : : ::::  ::  :::  ::
ATATAATAACTGATGCTGTGCCTTCTAATTCTCATAATCATTTGATCTTTCAATTAGTTA
   9010      9020      9030      9040      9050      9060

10090     10100      10110     10120     10130
GATGATTATTTGCAGATTCTTAT-GACCATTTTCCCATTGAGATTATAAGA-----TTTT
:::::: :::: :  :::::::: ::   :::: :::::: ::: ::::::      ::
GATGATGATTTACGCATTCTTATTGAGATTTTTACCATTG-GATGATAAGAGGGAATTGC
   9070      9080      9090      9100       9110      9120

10140     10150     10160     10170     10180     10190
GACATCGAATAGTTGGACTAGGAGTAAAGAGCTGTTGCTGTTATTTAGCACCCCAAAGGA
 : ::  :   :::: ::::  :::: :       :::::::::  :::::::::
AATATTTAGCTGTTGTACTAAAAGTAGAC------TGCTGTTATC-AGCACCCCA-----
    9130      9140      9150             9160       9170

10200     10210     10220     10230     10240          10250
AATATTATATACCTCTGAACCAATTGAATGGCCGACCTAGGTTTACT-----GAAATGTT
     :    ::    :::: :   :: :  :: :::::: ::::::     :::  :::
-----TGCTCACTGAAGAACTAGAAGATTACCCAACCTAGTTTTACTTCACTGAACCGTT
           9180      9190      9200      9210      9220
```

```
 10260     10270     10280     10290     10300     10310
TAGCTGTAAGAAGGTTAAGTGTTATCAGATTCCCCAAGTGAGAAGTACATGTTTCTTAGC
:   :: :::::  : ::: :: ::: :::: ::::::::: ::::: :::::::: :
TGCATGCAAGAACTTAAAGCGTAATCTGATTTCCCAAGTGACAAGTATATGTTTCTAA--
 9230      9240      9250      9260      9270      9280

 10320     10330     10340     10350     10360     10370
ATACTTTATGTTTCACGCACCTTGATTTTTCAAACTTTGTTTATCGATTTCTGAACTAAA
                 : ::::::                   ::: : :::
-----------------CTCCTTGA------------------CGAAT-CTG-------
                    9290

 10380     10390     10400     10410     10420     10430
GTGACTACATTATAGAACTTGAACCTAAAATTACTCTCCTCACTATAGGTGAAATCAGAT
                                    :: ::                 :::
-----------------------------------CTGCT----------------GAT
                                   9300

 10440     10450     10460     10470     10480     10490
TACTTGAAAATACTACTAAAAAAAATTATGGCGTTTGCTGGTATTTCTAACATCTTTTCT
: :::                               :::::::
TCCTT-------------------------------TGCTGGT-----------------
9310

 10500     10510     10520     10530     10540     10550
GCTAATCTTGTATTAATTTTCTCCTAGATGAACTTGTTATTATGTAAAAAGGTTTCATTA
       :: :::::                      :: :::::            : ::
-------TTATATTA--------------------TTCTTATG------------ACTA
    9320                           9330

 10560     10570     10580     10590     10600     10610
CTCATGCAATGGTGCACTAATGCTTGAGGAGTTCCAAGTAACTTTGCTGTCTCATGTAAA
:  :  ::::                                :::::  :
CAAACACAAT-------------------------------ACTTTTC-------------
9340                                    9350

 10620     10630     10640     10650     10660     10670
GAAGAGTGCTGAAGTTCACTATGGTTTAACTTCTACTGCACTGCTTGATATTGCCATGAA
                             ::::
-----------------------------AACT-----------------------------
                            9360

 10680     10690     10700     10710     10720     10730
CTCTGACATCATTTGGCTTGATCTTGTTCTAAAATCTAAATGAAATAATTCTCTCTTACT

-------------------------------------------------------------
```


10740 10750 10760 10770 10780 10790
ATATATCTTCTTAACCCTTTGCATATGATTAAGTGGTCTTTGATAGGATATCATTAAAAC
10800 10810 10820 10830 10840 10850
CTCGCATAAAAGCTACCATTTTATAAATTTCAAACTCCACGACGCATTTTCTGGTGATTC
AGCTA
GTGA
10860 10870 10880 10890 10900 10910
CATTGCTGATTATTGTTTAAAGACATCATTATTCCAATTAGTACATGTATAATAATTTCC
ATGAATAATCATTTCC
9370 9380
10920 10930 10940 10950 10960 10970
Exon 9 TCTGTTGTTGGTGCAGTTAATAATCTCCAAGTGCAGCAGTTTCTCGCATTTCCATATTCC
TTA-----TGTTGCAGTTAAAAAGCACCAAGTGCAGCAACTCCTCGCATTTCCATATTCC
9390 9400 9410 9420 9430 9440
10980 10990 11000 11010 11020 11030
ATGGAGAGTACCTGGGTTTCCATTGAGCGCAAAAGCTACATGTATGCTAAAA----AACC
ATGGAGAGTACCTACTATTTCACTGAGCGCAAAAGCTGCAAGTACGCTAAAACAAAAATC
9450 9460 9470 9480 9490 9500
11040 11050 11060 11070 11080 11090
TGAAGTAGCGTAAATCATATTTGTCTGGGTGGGAGGGCCTAGTACTCTTCCTCTATGTAT
TGAAGTAGCATAACTCAAATTTGTGCCGGTGG-AGAGCCTAGTACTCTTCCTCCATGTAT
9510 9520 9530 9540 9550
11100 11110 11120 11130 11140 11150
TAACTATCCTGTCCCAGTTAAGACATAAGAAATGTCAGAGAAGGATTTCTTTTCTGTATG
TGCTTTTCCAGTCCCAGTTAAGACATAACAAATGTCAGATAAGGATTTCTTTTCTGCATG
9560 9570 9580 9590 9600 9610
11160 11170 11180 11190 11200
TTTCATGAAGGCATTAAGATGCTGTTACAGT---TGTGACTAACTTATTATATAT---GT
TTTCATGAAGGCACTAAGATGCTGTGACAGTACTTGTGACTAACTTATTATATATTTTGT
9620 9630 9640 9650 9660 9670

```
          11210     11220     11230     11240     11250
   CTTAC--------TGCTTCATCTTGTGATATTTTCTTGCATGTTAATCTGATTAAAGTGT
   ::::         :  :::::::::: ::::::  : :: :    ::::       ::: :
   CTTATATTTCTTATCTTTCATCTTGTAATATTTCTTCGCGT----ATCT------AGTAT
9680      9690      9700      9710      9720

  11260           11270          11280              11290
   AGCTT-------AGACCATTCA-----CCATGTTAATGGTGACT---------TGTTGGT
    ::::       : ::: :::       :::   : : :   :::          :: :: :
   TGCTTTTCATTCAAACCCTTCCGTGACCCAGAATCAGGACCACTGCCTTAGCATGCTGTT
9730      9740      9750      9760      9770      9780

   11300     11310           11320     11330     11340
   GACTACTAGTAGCTGTAG-------CTCTCCGTAGTACTGCTATGCCTTCAAAAAATGAT
    :  :   :::  ::::        ::::   :  : :::: : :: :   : : :   :
   CATCAGCGGTACATGTAATAGAGGCCTCTATATTTTGCTGCCA-GCTTAATATACAGTTT
9790      9800      9810      9820      9830       9840

11350     11360       11370        11380     11390     11400
   GGGTCGGAAATTACT-AGCT-AGCT--AGTA-TTGCTGTTTCATTCAATCTCTGCTTTAA
      ::    ::    : :: : :::   :::: ::  : ::    : : : :::  : :::
   ACATCTTTCATGTGTGAGTTCAGCACGAGTAATTAATTTTATGGTTATTTTCTT-TGTAA
 9850      9860      9870      9880      9890      9900

    11410       11420     11430         11440     11450
   CCCAAAAA--TCAGGACTAGTGGATTAGCATACC----TCTCACCAGGACAATGCACTAG
   :  :       : : : ::: : : : :::::  :    : :    ::   ::   : ::
   CAGAGCCTCTTGATGTCTATTTG-TAAGCATTGCGAGGTTTTTAAAGATTAAATTAATAC
  9910      9920      9930       9940      9950      9960

    11460                   11470     11480
   ---AGCACATTTTC-------------ATCTTCTTCTCATATTT
      :::  : : ::              :  : ::::    :  :
   GTAAGCTGAATGTCTCGCAAAAGGTACAAATTGCTTC----AGCT
   9970      9980      9990     10000
```

APPENDIX B: AMINO ACID SEQUENCE ALIGNMENTS USED FOR PHYLOGENETIC ANALYSES

Gaps are represented by dashes and amino acids identical to PTAG1 residues are shown as dots.

MIK regions of plant MADS-box sequences representing different subfamilies.

MADS-box

```
PTAG1 GRGKVEIKRIENTTNRQVTFCKRRSGLLKKAYELSVLCDAEVALIVFSSRGRLYEY  56
PTAG2 ......................N.............................  56
TAG1  ....I..................N...................V...N.....  56
PLENA ....I.......I...........N...................V.........  56
AGL1  ....I..................N...................VI..T.....  56
AG    ....I..................N.............................  56
FBP7  ....I.......N...........N.............E..I......T...V...  56
AP1   ...R.QL.....KI......S...A......H.I..........V...HK.K.F..  56
SQUA  .....QL.....KI......S...G......H.................NK.K.F..  56
TM4   ...R.QL.....KI......S..........H.I.........G.....TK.K.F..  56
TM5   ...R..L....GKI......A...N...................I..N..K...F  56
AGL2  ...R..L.....KI......A...N...................I..N..K...F  56
AGL6  ...R..M.....KI......S...N...................I.....K...F  56
GLO   ....I.......SS.....YS...N.IM...K.I......H.SV.I.A.S.KMH.F  56
FBP1  ....I.......SS.....YS...N.I....K.I......R.SV.I.A.S.KMH.F  56
PI    ....I.......AN..V...S...N..V...K.IT.....K....I.A.N.KMID.  56
DEF   A...IQ......Q......YS...N..F...H........K.SI.MI..TQK.H..  56
TM6   --..I...K...S......YS...N.IF..RK..T.....KIS..ML..TRKYH..  54
PTD   ....I...K...P......YS...N.IF...Q..T.....K.S..IVPNTNK.N..  56
AP3   A...IQ......Q......YS...N..F...H..T.....R.SI.M...SNK.H..  56
AGL17 ....IV.QK.DDS.S.....S...K..I...K..AI......C..I..NTDK..DF  56
TM3   V...TQMR....A.S.....S...N......F..........G..I..P..K...F  56
MEF2C ..K.IQ.T..MDER......T..KF..M...........C.I...I.N.TNK.FQ.  56
```

I-region

```
PTAG1 --NDSVKSTIERYKKASADSSN-TGSVSEANAQYY  32
PTAG2 --.N............C.....-N.........F.  32
TAG1  --.N...A........CS....-............  32
PLENA --.N..RA..............-SV.T....T.F.  32
AGL1  --.N..RG........CS.AV.-PP..T...T...  32
AG    --.N...G........IS.N..-....A.I.....  32
FBP7  --.NNIRAI.D.....TVET..-AFTTQ.L...F.  32
AP1   T-DSCMEKIL...ERY.YAERQ-LIAPESDVNTNW  33
SQUA  T-DSCMDRIL.K.ERY.FAERQ-LV.NEPQSPANW  33
TM4   N-DSCMERIL...ERY.FAEKQ-LVPTDHTSPVSW  33
TM5   S-SS.MLK.L...Q.CNYGAPEPNI.TR..LEISS  34
AGL2  S-SSNMLK.LD..Q.C.YG.IEVNNKPAKELEN-S  33
AGL6  --SVGIE......NRCYNC.LS-NNKPE.TTQS-W  31
GLO   SPSTTLVDMLDH.H.L.------GKRLWDPKHEHL  29
FBP1  S--T.LVDILDQ.H.LT------GRRLLD.KHENL  27
PI    CPSMDLGAMLDQ.Q.L.------GKKLWD.KHENL  29
DEF   SPTTAT.QLFDQ.Q..VG-----VDLW.SHYEK-M  29
TM6   SP.TTT.KM.DQ.QS.LG-----VDIW.IHYEK-M  29
PTD   SPST.T.KIYDQ.QN.LG-----IDLWGTQYEK-M  29
AP3   SP.TTT.EIVDL.QTI.D-----VDVWATQYER-M  29
AGL17 --SS.......--RFNT.KMEEQELMNPASEVKFW  31
TM3   --SS.TQEI.RGN.RHTK.RVQPENQAGPQYL..M  33
MEF2C ----------------------------------   -
```

K-box

```
PTAG1 QQEAAKLRSQIGNLQNSNRHMLGEALSSLSVKELKSLEIRLEKGISRIRSKKNELLFAEIEYMQKRE  67
PTAG2 ....................N....S..A...........K.....G.....................  67
TAG1  ....S...A.....M.Q..N.M....AGMKL....N..Q.I.....K.....................  67
PLENA ....N...R..REI.T...Q....GV.NMAL.D...T.AKV..A.................H.....  67
AGL1  ....S...R..RDI......IV..S.G..NF....N..G........V........V..........  67
AG    ...S....Q..ISI.....QLM..TIG.M.P...RN..G...RS.T...........S..D......  67
FBP7  ...SK...Q..QLI......LV..G....N.R...Q..N...R..A......H.MIL..S.DL....  67
AP1   SM.YNR.KAK.EL.ERNQ..Y...D.QAM.P...QN..QQ.DTALKH..TR..Q.MYES.NEL..K.  67
SQUA  TL.YS..KAR.EL..RNH..YM..D.D.M.L..IQ...QQ.DTALKN..TR..Q..YDS.SEL.HK.  67
TM4   TL.HR..KARLEV..RNQK.YV..D.E...M...QN..HQ.DSALKH...R..Q.MHES.SVL..KD  67
TM5   ...YL..KGRYEA..R.Q.NL...D.GP.NS...E...RQ.DMSLKQ...TRTQ.MLDQLTDY.RK.  67
AGL2  YR.YL..KGRYE...RQQ.NL...D.GP.NS...EQ..RQ.DGSLKQV..I.TQYMLDQLSDL.NK.  67
AGL6  C..VT..K.KYES.VRT..NL...D.GEMG....QA..RQ..AALTAT.QR.TQVMME.M.DLR.K.  67
GLO   DN.INRVKKENDSM.IEL..LK..DITT.NY...MV..DA..N.T.ALKN.QM.FVRMMRKHNEMV.  67
FBP1  DN.IN.VKKDND.M.IEL..LK..DIT..NHR..MI..DA..N.LTS..N.Q..V.RMMRKKT.SM.  67
PI    SN.IDRIKKENDS..LEL..LK..DIQ..NL.N.MAV.HAI.H.LDKV.DHQM.I.ISKRRNEKMMA  67
DEF   .EHLK..NEVNR..RREI.QRM..S.ND.GYEQIVN.IEDMDNSLKL..ER.YKVISNQ.DTSK.KV  67
TM6   .ENLKR.KEINNK.RREI.QRT..DM.G.NLQ..CH.QENITESVAE..ER.YHVIKNQTDTCK.KA  67
PTD   .EHLR..NDINHK.RQEI.QRR..G.ND..IDH.RG..QHMTEALNGV.GR.YHVIKTQN.TYR.KV  67
AP3   .ETKR..LETNR..RTQIKQR...C.DE.DIQ..RR..DEM.NTFKLV.ER.FKS.GNQ..TTK.KN  67
AGL17 .R..ET..QELHS..ENY.QLT.VE.NG......QNI.SQ..MSLRG..M.REQI.TN..KELTRKR  67
TM3   .H...N.MKK.EL.ETAK.KF...G.Q.CTLQ.VQQI.KQ..RSVGT..AR.LQVFKEQV.RLK.KK  67
MEF2C ----------------------------------------------------------------  -
```

Alignment of AG family (entire coding regions)

N-terminal

```
PTAG1     ---------------------------------MEYQNESL------ESSP------------LRKL-  16
PTAG2     ---------------------------------.A....PQ------....------------....-  16
NAG1      ---------------------------------.DF.SDLTR------.I..------------Q..V-  17
PMADS3    ---------------------------------..F.SDLTR------.I..------------Q....  17
TAG1      ---------------------------------.DF.SDLTR------.I..------------Q...-  17
BAG       ---------------------------------.A..M.LGG------....------------Q..A-  17
AG        HFLQLLQISYFPENHFPKKNKTFPFVLLPPTAITA..S.LGG-----D...------------...S-  50
GAG2      ---------------------------------.SFYDDQSG-----NL..------------Q....  17
CUM1      --------MSKHYQSPLTRMIKEEGKGKLQIKG.FQNQ.EKM-----SD..------------Q..M-  42
SLM1      ---------------------------------..FSSQITR----E.G..SS----------Q....  20
RAP1      ---------------------------------..FS..LSRD--MEDG..------------Q..M-  20
FBP6      ---------------------------------.VFP.QEFE-----S..S------------Q..S-  17
PAGL1     ---------------------------------.VFP.QEFE-----S..S------------Q..S-  17
PLE       ---------------------------------..FP.QDSE-----SLRK------------NGRG-  17
CUS1      -----------------------------MSCY.EED.ESGVVGLRR..S------------SSRT-  25
AGL1      ---------------------------------..EGGS--------HDAES-----------SK.--  14
AGL5      ---------------------------------..GGASN-------.VAES-----------SK.I-  16
FBP7      ------------------------------------------------------------------M   1
FBP11     ------------------------------------------------------------------M   1
AGL11     ------------------------------------------------------------------M   1
CUM10     ------------------------------------------------------------------M   1
OSMADS3   ------------------------------------------------------------------M   1
ZAG1      -----------------MHIREEEATPSTVTGI.STLTSAGQQKLKEPI..GGGSASVAGSAAE.NNG  51
ZMM2      ----------------------------MLNMMTDLSCGP.SKVKEQVAAA.TG-----SG-D.QGQ  33
ZAG2      ------------------------------------------------------------------M   1
ZMM1      ------------------------------------------------------------------M   1
DAL2      ------------------------------------------------------------------M   1
AP1       ------------------------------------------------------------------M   1
SQUA      ------------------------------------------------------------------M   1
AP3       ------------------------------------------------------------------M   1
DEF       ------------------------------------------------------------------M   1
PI        ------------------------------------------------------------------M   1
GLO       ------------------------------------------------------------------M   1
```

MADS-box

```
PTAG1     GRGKVEIKRIENTTNRQVTFCKRRSGLLKKAYELSVLCDAEVALIVFSSRGRLYEY  56
PTAG2     ........................N.............................  56
NAG1      ....I...................N.............................  56
PMADS3    ....I...................N.............................  56
TAG1      ....I...................N.................V...N.......  56
BAG       ....I...................N.............................  56
AG        ....I...................N.............................  56
GAG2      ....I...................N.......................T.......  56
CUM1      ....I...................N.............................  56
SLM1      ....I...................N.............................  56
RAP1      ....I.......V...........N.............................  56
FBP6      ....I...................N.............................  56
PAGL1     ....I...................N.............................  56
PLENA     ....I.......I...........N.................V...........  56
CUS1      ....I...................N.............................  56
AGL1      ....I...................N.................VI..T.......  56
AGL5      ....I...................N.................VI..T.......  56
FBP7      ....I.......N...........N.............E..I......T...V...  56
FBP11     ....I.......N...........N.................I......T...V...  56
AGL11     ....I.......S...........N.......................T.......  56
CUM10     ....I...................N.............................  56
OSMADS3   ....I...................N.............................  56
ZAG1      .K..T...................N.............................  56
ZMM2      ....I...................N.................V...........  56
ZAG2      ...RI.......N.S.........N.............................  56
ZMM1      ...RI.......N.S.........N.................V...........  56
DAL2      ....I...................N.............................F  56
AP1       ...R.QL.....KI......S...A......H.I..........V...HK.K.F..  56
SQUA      .....QL.....KI......S...G......H..............NK.K.F..  56
AP3       A...IQ......Q......YS...N..F...H..T.....R.SI.M...SNK.H..  56
DEF       A...IQ......Q......YS...N..F...H........K.SI.MI..TQK.H..  56
PI        ....I.......AN..V...S...N..V...K.IT.....K....I.A.N.KMID.  56
GLO       ....I.......SS.....YS...N.IM...K.I......H.SV.I.A.S.KMH.F  56
```

I-region

```
PTAG1    --SNDSVKSTIERYKKASAD-SSNTGSVSEANAQ-YY  33
PTAG2    --..N..........C..-...N.........-F.  33
NAG1     --A.N...A........CS.-......I......-..  33
PMADS3   --A.N...A........CS.-......IA.....-..  33
TAG1     --A.NNIR.........CS.-.T..ST.Q.I..A-..  33
BAG      --..N...G........IS.-N......A.I...-..  33
AG       --..N...G........IS.-N......A.I...-..  33
GAG2     --A.N...G........CT.-.P..S.........-F.  33
CUM1     --A.N...A..D......S.-......T....T.F.-  33
SLM1     --A.H...G..D......S.-N.GAS.AA.....-..  33
RAP1     --A.H...A........TCS.-.TGVT..E.....---  30
FBP6     --A.N..RA..D....HH..-.TS.........T.-..  33
PAGL1    --A.N..RA..D....HH..-.TS.........T.-..  33
PLE      --A.N..RA...........-...SV.T....T.-F.  33
CUS1     --A.N..RA..S.....YS.-P.TAMT.....T.-F.  33
AGL1     --A.N..RG........CS.-AV.PP..T...T.-..  33
AGL5     --A.N..RG........CS.-AV.PPTIT...T.-..  33
FBP7     --..NNIRAI.D.....TVE-T..AFTTQ.L...-F.  33
FBP11    --A.NNI.G........T.E-T..ACTTQ.L...-F.  33
AGL11    --A.NNIR.........CS.-.T..ST.Q.I..A-..  33
CUM10    ---.N.I.T........CS.-..A.S..T.L.T.-..  32
OSMADS3  --A.N.....V......NS.-T..S.T.A.V...-H.  33
ZAG1     --A.N...G........TS.N..AA.TIA.VTI.-H.  34
ZMM2     --A.N............NS.-...S.T.A.V...-..  33
ZAG2     --A.N...A.V......HTVG..SGPPLL.H...QF.  35
ZMM1     --A.N...A........H.VG..SGPPLL.H...QF.  35
DAL2     --A.H...R........TCV.-NNHG.VI..S.S.YW-  33
AP1      -STDSCMEKIL...ERY.YAERQLIAPE.DV.TNW--  34
SQUA     -STDSCMDRIL.K.ERY.FAERQLVSNEPQSP.NW--  34
AP3      ISP.TTT.EIVDL.QTI.DV--DVWATQY.RM-----  30
DEF      ISPTTAT.QLFDQ.Q..VGV--DLWS.HY.KM-----  30
PI       CCPSMDLGAMLDQ.Q.L.GK--KLWDAKH.NL-----  30
GLO      CSPSTTLVDMLDH.H.L.GK--RLWDPKH.HL-----  30
```

K-box

```
PTAG1     QQEAAKLRSQIGNLQNSN----RHMLGEALSSLSVKELKSLEIRLEKGISRIRSKKNELLFAEIEYMQKRE  67
PTAG2     ....................----.N....S..A...........K.....G.....................  67
NAG1      ....S...A.......Q.----.N....S.AA..LRD..N..QKI.....K.....................  67
PMADS3    ....S...A.......Q.----.NF...S.AA.NLRD.RN..QKI.....K..A...................  67
TAG1      ....S...A.....M.Q.----.N.M....AGMKL....N..Q.I.....K.....................  67
BAG       ...S....Q..ISI.........QLM..TIG.M.P...RN..G..DRSVN.............D.......  67
AG        ...S....Q..ISI.........QLM..TIG.M.P...RN..G...RS.T...........S..D.......  67
GAG2      ....S...QE.SSI.KN.----.N.M..S.G..T.RD..G..TK.........................K.  67
CUM1      ........V.............----.N....S....TA.D..G..TK.....................R...  67
SLM1      ........N..RTVTEN.----..LM..G....NM.D.....NK..R...................F.....  67
RAP1      ........N..RT...QTRNTS.NLM..G.T.MNM.D..N..T........V.A.......G...F...K.  71
FBP6      ........R..RDI.TY.----.QIV........PRG..N..GK...A.G.V.........S...L.....  67
PAGL1     ........R..RDI.TY.----.QIV........PRD..N..GK...A.G.V.........S...L.....  67
PLE       ....N...R..REI.T..----.Q....GV.NMAL.D...T.AKV..A.................H.....  67
CUS1      ...S....A.......L.----..L...SI......D.....VK..........R......S.........  67
AGL1      ....S...R..RDI.........IV..S.G..NF....N..G........V........V...........  67
AGL5      ....S...R..RDI..L.----..I...S.G..NF....N..S........V....H.M.V..........  67
FBP7      ...SK...Q..QLI.........LV..G....N.R...Q..N...R..A......H.MIL..S.DL....  67
FBP11     ...SK...Q..QL...T.----..LV..G..A.N.R...Q..N...R..T......H.MIL..T.NL....  67
AGL11     ...S....Q..QTI.....----.NLM.DS..........QV.N....A........H...LV...NA....  67
CUM10     ...S....Q..QM.....SNLV..LM.DS..A.T.....Q..N...R..T......H.M.L.....L....  71
OSMADS3   ...SS...Q..SS...A.S---.TIV.DSINTM.LRD..QV.N......AK..AR.....Y..V.......  68
ZAG1      K..S.R..Q..V..........----.ALI.DSITTM.H....H..T..D.ALGK..A...DV.CS.V....R..  67
ZMM2      ...SS...QM.HS...A.T---.NIV.DSIHTMGLRD..QM.GK...A.IK..AR.....Y..VD......  68
ZAG2      ...S....N..QM...T.----..LV.DSVGN..L....Q..S.......K..AR.S...A...S..A...  67
ZMM1      ...SV...N..QM...T.----..LV.DSVGN..L....Q..S.......K..AR.S...A...N..A...  67
DAL2      ....G...Q..EI...A.----..LM.DG.TA.NI....Q..V......G.V......M.LE..DI..R..  67
AP1       SM.YNR.KAK.EL.ERNQ----..Y...D.QAM.P...QN..QQ.DTALKH..TR..Q.MYES.NEL..K.  67
SQUA      TL.YS..KAR.EL..RNH----..YM..D.D.M.L..IQ...QQ.DTALKN..TR..Q..YDS.SEL.HK.  67
AP3       .ETKR..LETNR..RTQI----KQR...C.DE.DIQ..RR..DEM.NTFKLV.ER.FKS.GNQ..TTK.KN  67
DEF       .EHLK..NEVNR..RREI----.QRM..S.ND.GYEQIVN.IEDMDNSLKL..ER.YKVISNQ.DTSK.KV  67
PI        SN.IDRIKKENDS..LEL----..LK..DIQ..NL.N.MAV.HAI.H.LDKV.DHQM.I.ISKRRNEKMMA  67
GLO       DN.INRVKKENDSM.IEL----..LK..DITT.NY...MV..DA..N.T.ALKN.QM.FVRMMRKHNEMV.  67
```

C-terminal

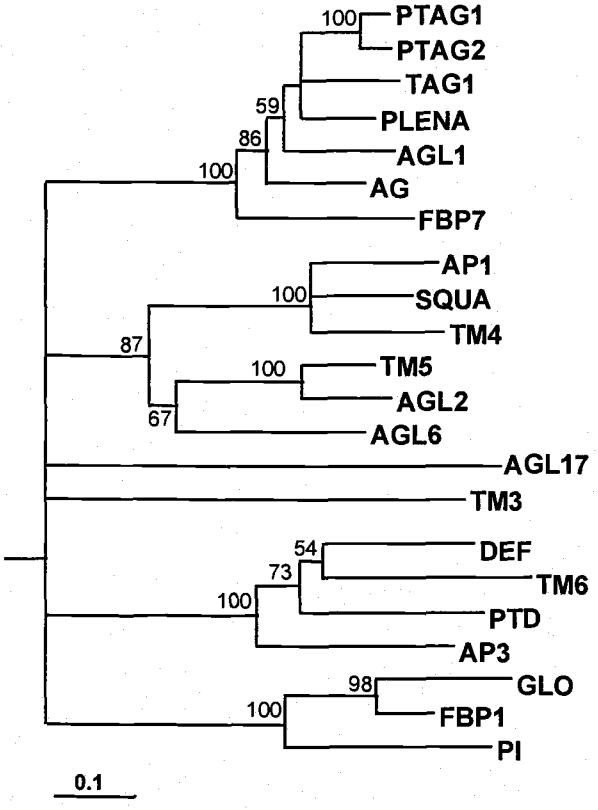
```
PTAG1    VDLHNNNQLLRAKISE---NERKR-----QSMNLMPGGAD------FEIVQS-----QPY   41
PTAG2    I............A.---.....-----.H.......VN-------...M..------..F   41
NAG1     I.......Y.....A.---T..AQQQQQQ.Q......SSS------Y.L.PPPHQ----F   47
PMADS3   I.......Y.....A.---T..SQ------Q......SSS------YDL.PPQ.S----F   41
TAG1     ........Y.....A.---T..AQHQH--.Q......SSSN----YH.L.PPP.Q----F   47
BAG      .....D........A.---...NN-----P..S.....SN-----YEQ.MPPP.TQPQPF   47
AG       .....D..I.....A.---...NN-----P.IS.....SN-----YEQLMPPP.TQSQPF   47
GAG2     I.......Y.....A.---...AQ-----.H......SS.------Y.LAPP-.S----F   41
CUM1     I.......M.....A.---S..--------NV.M.G.--------E..LM..-H.-----.   36
SLM1     .E......Y.....A.---...AQ-----...S.....SS-----EY.LAPPP.S----F   43
RAP1     IE......F.....A.---S..SQ-----.........SSSGEQ--HY.LMPQS.AG--.F   48
FBP6     IEMQ.A.MY.....A.---V..AT-----.Q....H..G------SEYQQ.PMSSTS...   46
PAGL1    IEMQ.A.MY.....A.---V..AT-----.Q.......G------SEYQQ.PMSSTS...   46
PLE      LE...A.MF.....A.---G..AQ-----.Q.....--------SDY--.PM--TS.S.   40
CUS1     IE..T....I....A.---T..SQ-----.NT.ASNNNG-------IATRRGEEG-----   40
AGL1     ME.QH..MY.....A.---GA.LNPD--Q.ESSVIQ.TTVY----ESGVSSHD.SQ---H   48
AGL5     IE.Q.D.MY..S..T.---RTGLQ----Q.ESSVIHQ.TVY----ESGVTS.H.SG---Q   46
FBP7     IQ.EQE.AF..S..A.---...------L.ELSM..A.GQ-----EYNAI.QYLA-----   41
FBP11    IQ.EQE.TF..S..A.---...------L.ELSM..ATGQ-----EYNAF.QYFA-----   41
AGL11    IE.D.E.IY..T.VA.---V..------Y.QHHHQMVSGS-----EINAIEALAS-----   41
CUM10    IE.E.E.VCI.T..A.---V..------V.QA.--MVSGQ-----ELNAI.ALANS----   40
OSMADS3  .E.Q.D.MY..S.VV.---...GQ-----.PL.M-M.A.S-----TS.YDHMVNN---..   43
ZAG1     ME.Q.D.LY..SRVD.---...AQ-----.TA.M-M.APS-----TS.YQ.HGFT---..   43
ZMM2     M..QTD.MY..S..A.S---.TGQ-----PA.HMTM.APP-----TS.YDHMA-----.F   43
ZAG2     TE.Q.DHMT..T..E.---G.Q-Q----L.QVTVARSV.AAAAATNL.LNPFLEMDTKCF   52
ZMM1     TE.Q.DHMN..T..E.---G.Q-Q----L.QVTVAQSV.AAAA-TDV.LNPFLEMDTKCF   51
DAL2     HI.IQE.EI..S..A.C--.SH--------NT.MLSAP-------EYDALPAFDS-----   38
AP1      KAIQEQ.SM.SKQ.K.REKIL.AQ----QEQWDQQNQ.HNMPPPLPPQQH.IQH.YMLSH   56
SQUA     KAIQEQ.TM.AK..K.KE-K.IAQ----QPQWEHHRHHTN------AS.MPPPPQ----.   45
AP3      ---KSQQDIQKNL.H.L--EL.AE----DPHYG.VDN.G.------YDS.LGY.IE--GS   43
DEF      ---R.VEEIH.NLVL.F--DA.RE----DPHFG.VDNEG.------YNS.LGFPNG--GP   43
PI       EEQRQLTFQ.QQQEMAIAS.A.G------MM.RDHD.QFG------YRVQPIQPN-----   43
GLO      EENQSLQFK..QMHLDPMNDNVMES----.AVYDHHHHQN------IADYEAQM.----F   46
```

C-terminal, continued

```
PTAG1    D-SRNYSQVNGLQP----ASHYS-----HQDQM---ALQLV-------------------   69
PTAG2    .-..........P.-----.N..P-----.E..L---FS----------------------   66
NAG1     .-T...L......T----NN..T-----R...P---S.....------------------   75
PMADS3   .-A...L......T----NN..P-----R...P---P.....------------------   69
TAG1     .-T...L......T----NN..P-----R...P---PI...------------------   75
BAG      .-....F..AA...N---NH...--SAGRE..T---.....------------------   79
AG       .-....F..AA...N---NH...--SAGR...T---.....------------------   79
GAG2     .-G...I.L.....----NN...-----R...T---.....------------------   69
CUM1     .-P.DFF......H----NHQ.P-----R..N.---.....------------------   64
SLM1     .-....F...A...N---NT...-----RP..T---T...N------------------   72
RAP1     .-...FF..SD...----DER..-----C.N.T---P.....------------------   76
FBP6     .-A..FLP..L.E.N---P-...-----R...T---.....------------------   74
PAGL1    .-A..FLP..L.E.N---P-...-----R...T---.....------------------   74
PLE      .-V..FLPM.LME.N---QQQ..-----RH..T---.....------------------   69
CUS1     SMGT.LEDN.HH.YDS--TNYFDP----.HNHP--IS.....------------------   73
AGL1     Y-N...IP..L.E.----NQQF.-----G...P---P.....------------------   76
AGL5     Y-N...IA..L.E.----NQNS.-----N...P---P.....------------------   74
FBP7     ----..ML.L.MMEG----VPS.---PLPSD----KKS.D.E------------------   68
FBP11    ----..ML.L.MMEGG---VPS.D--PLPAH.---KKS...E------------------   71
AGL11    ------.FAHSIMTAG---SGSGN--GGSYS.PD-KKI.H.G------------------   73
CUM10    ----..FFSP.IME.A---G------PVSYSH.D-KKM.H.G------------------   68
OSMADS3  .-....FL...IM.Q----PQ..A-----..L.P--TT...GQQPAFN-------------   78
ZAG1     .PI.SFL.F.IV.Q----PQF..-----Q.EDRKDFND.GGR------------------   76
ZMM2     .-....FL..---SM----PQ...------..L.P--TT...G------------------   69
ZAG2     FTGGPFATLDMKCFL---PGSLQQMLEAQ.R..LATE.N.GYQLAPPGSDAANNNPHHQF  109
ZMM1     FPGGPFATLDMKCFF---PGSLQ-MLEAQ.R..LATE.N.GYQLAPPDTDVANNNPQ-QF  106
DAL2     ----..FLHA.LIDA----.H..A-----..E.T---T...G------------------   64
AP1      Q-PSPFLNMG..YQE---DDPMAMRNDLELTLEPVYNCN.GCFAA---------------   97
SQUA     SMAPQFPCI.VGNTYEGEGANEDRRNELDLTLDSLYSCH.GCFAA---------------   90
AP3      RAYALRFHQ.HHHYYPNHGL.AP----SAS.II---TFH.LE------------------   78
DEF      RIIALRLPT.HHP-----TL.SG----GGS.LT---TFA.LE------------------   73
PI       --LQEKIMSLVID-----------------------------------------------   54
GLO      AFRVQPM.P.LQERF---------------------------------------------   61
```

APPENDIX C: ADDITIONAL *AG* FAMILY COMPARISONS

Phylogeny estimated by neighbor-joining procedure for MIK regions of a representative subset of plant MADS-box genes. *MEF2C* was the outgroup. Nodes less than 50% are collapsed.

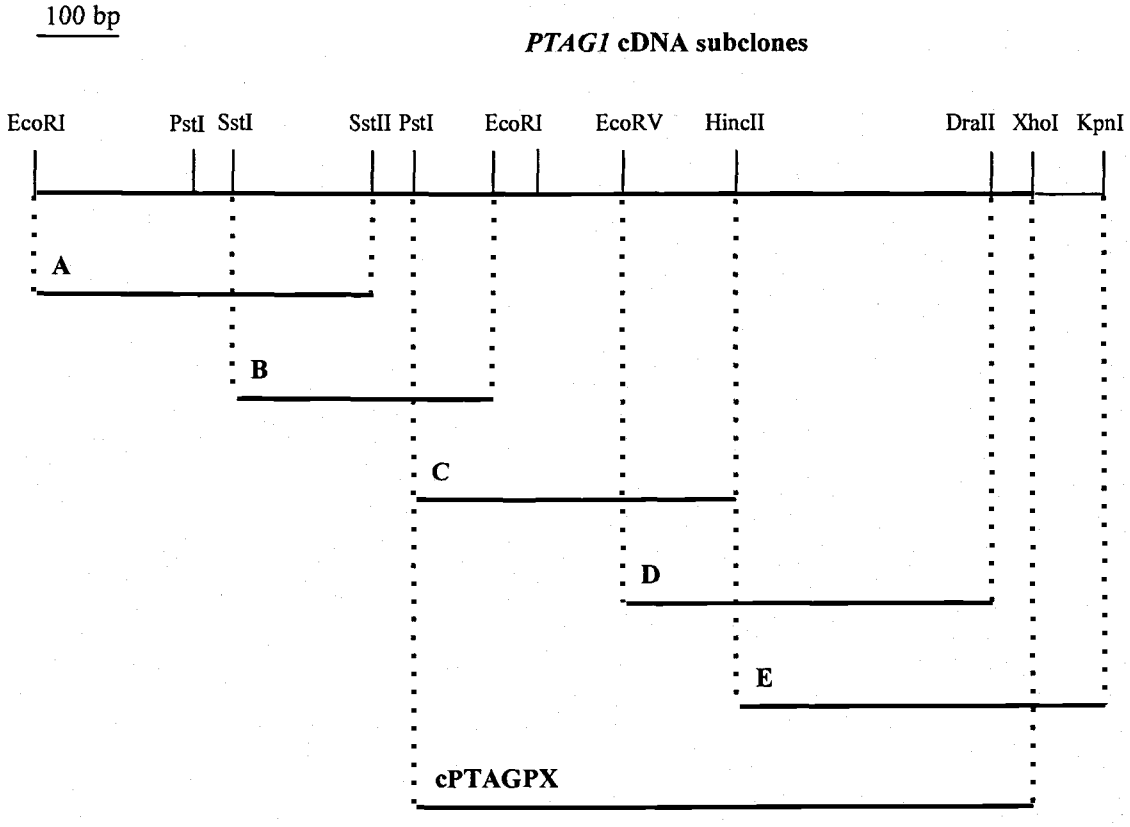


Percent identity by domain between the deduced amino acid sequences of *PTAG1/2* and additional members of the *AG* subfamily[1]

Gene	*PTAG1*				*PTAG2*			
	M	**I**	**K**	**C**	**M**	**I**	**K**	**C**
TAG1 *(Lycopersion esculatum)*	93	85	78	64	95	85	75	62
NAG1 *(Nicotiana tobaccum)*	96	82	78	62	98	82	82	61
PMADS3 *(Petunia hybridia)*	96	79	72	63	98	79	76	63
GAG2 *(Panax ginseng)*	95	76	72	68	96	82	73	65
CUM1 *(Cucumis sativus)*	96	76	85	64	98	76	87	58
BAG *(Brassica napus)*	96	79	67	70	98	76	64	68
AG *(Arabidopsis thaliana)*	96	79	69	70	98	76	66	62
SLM1 *(Silene latifolia)*	96	67	75	65	98	61	72	62
RAP1 *(Rumex acetosa)*	95	63	69	64	96	67	66	61
PLE *(Anirrhinum majus)*	93	73	70	54	95	76	70	48
AGL1 *(Arabidopsis thaliana)*	91	64	78	47	93	67	73	40
AGL5 *(Arabidopsis thaliana)*	91	58	75	43	93	61	70	35
AGL11 *(Arabidopsis thaliana)*	93	61	72	33	95	61	70	38
DAL2 *(Picea abies)*	95	55	66	36	96	55	66	34
ZMM2 *(Zea mays)*	95	75	52	46	96	76	57	44
ZAG2 *(Zea mays)*	91	55	67	28	93	61	64	38

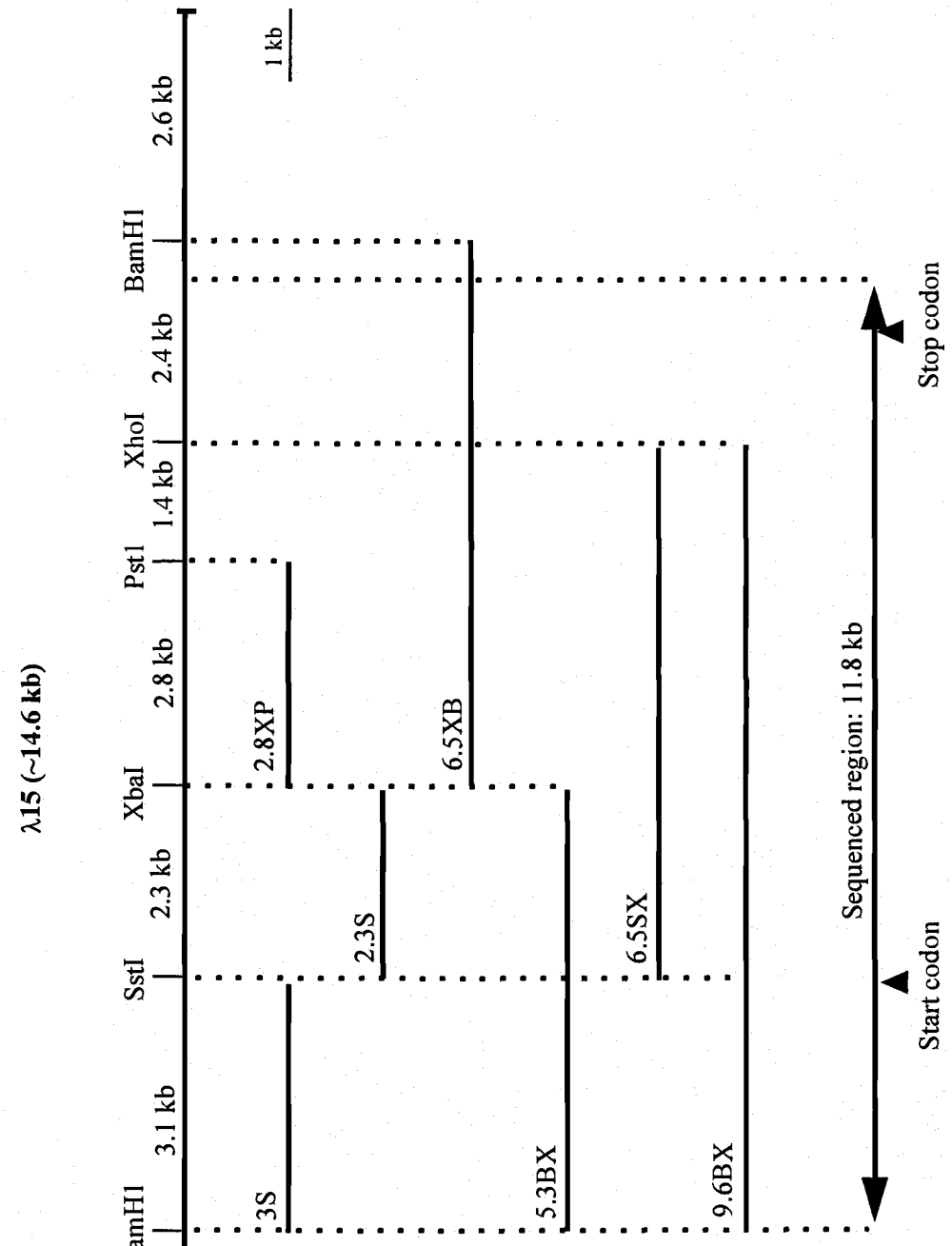
[1]The solid lines separates genes belonging to the dicot C-group from other members of the *AG* subfamily.

APPENDIX D: MAPS OF CDNA AND GENOMIC SUBCLONES

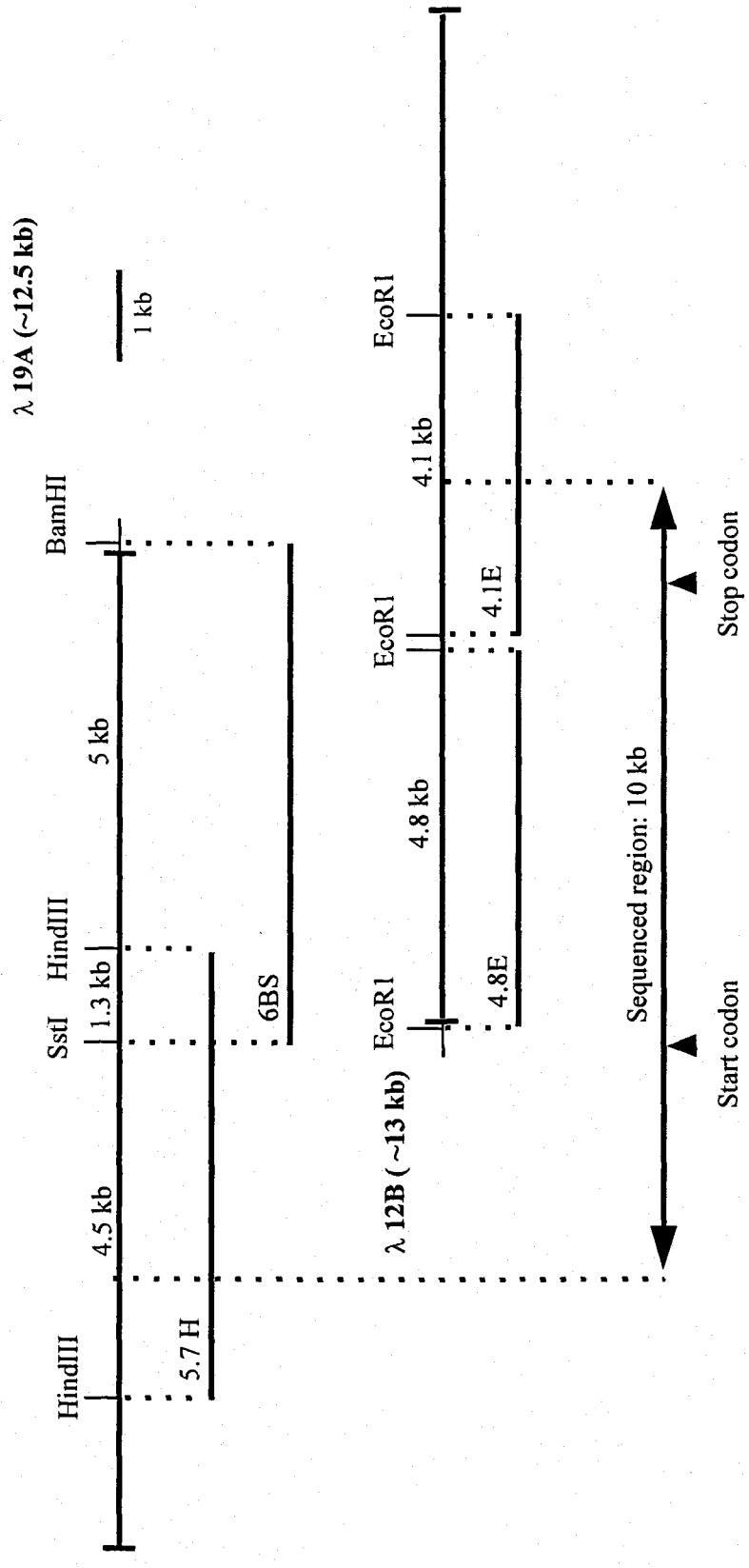


Note: The *PTAG1* cDNA clone contains one unspliced intron (intron 5, 143 bp)

PTAG1 Genomic Subclones



PTAG2 Genomic Subclones



APPENDIX E: SEQUENCING AND PCR PRIMERS

The position of the 5' nucleotide of each primer in the final *PTAG1/2* genomic and cDNA sequences (Appendix A, Figure 2.1) is shown.

PTAG1 Genomic Sequencing Primers

Primers used for sequencing 5.3BX subclone, which includes all 5' flanking sequence and extends 3' into Intron 2 (Appendix D). FB: forward primer series starting from near the BamH1 site. FS: forward primers starting from the near the SstI site. RS: reverse primers starting near the SstI site. RX: reverse primers starting near the XbaI site. 65SX forward and reverse primers sequenced across the XbaI site using the 6.5SX subclone as a template.

Name	Position		Sequence
53FB1	5' flanking	294	TTG AGC AAT ACA AAG CTA G
53FB2	5' flanking	675	GCA GGT CGT ATC ATT CTG
53FB3	5' flanking	1,091	CTG GGC AGC TGT TAT TCG
53FB4	5' flanking	1,508	TCA TCA TCT CAA GCA TGG
53FB5	5' flanking	1,912	GGG AAC ATT CAT ACA ACT ATC
53FB6	5' flanking	2,360	CAA AGA CAG TAG TTA GAC TTG G
53FS	Intron 1	2,952	AAA TAT GCT ATG AGT TGT GC
53FS1	Intron 2	3,355	AGC CCT CTA TCT TCA TGT GG
53FS2	Intron 2	3,723	TTT GAG TTC TTG GTT TGG
53FS3	Intron 2	4,073	TAC CAT GTC AAA TAA CCG
53FS4	Intron 2	4,474	ACC TGG GTA AGT ATG GTA C
53FS5	Intron 2	4,763	ATG GAT CAA TGA GAA TGG
53FS6	Intron 2	5,086	CGA CAA CTG TTA ACC AGT G
53RS7	5' flanking	262	CAC CCA ATG ACA ACC AAC
53RS6	5' flanking	684	TAC GAC CTG CAT CAG TGC
53RS5	5' flanking	1,106	AAT AAC AGC TGC CCA GTG
53RS4	5' flanking	1,602	AAC GAG AGA TGG AAT GCG
53RS3	5' flanking	2,033	TAA CCT CAG ACC TCA TCT C
53RS2	Exon 1/5'flank	2,421	ACA ACC TCT CAT CCA ATC
53RS1	Intron 1	2,769	GAA ATG GGA AAG GGA GAC AG
53RX6	Exon 2	3,215	TCG TAG GCT TTC TTG AGC
53RX5	Intron 2	3,620	ATA AGG AAT GAT TGT GTC TC
53RX4A	Intron 2	3,989	ACA GCT TTG GCT ATA GGC
53RX4	Intron 2	3,979	CTA TAG GCA ATC TCT TTA GC
53RX3	Intron 2	4,412	CAC TTG CTT GAT TTA ACA G
53RX2	Intron 2	4,809	ACA GTG CCA TCA TTA TGG
53RX1	Intron 2	5,168	CAA TAC CAG CTA GAG TTG TAG
65SXF	Intron 2	5,076	AGA GAA TGA ACG ACA ACT G
65SXR	Intron 2	5,631	CAG TCA AAT CAA CGA ATA AC

PTAG1 Genomic Sequencing Primers, continued

Primers used for sequencing the 6.5XB sublone, which contains part of Intron 2 and all downstream regions. FX: forward primer series starting near the XbaI site. FP: forward primers starting near the Pst1 site. RP: reverse primers starting near the Pst1 site. RC: reverse primers starting at the 3' end of sequence.

Name	Position		Sequence
65FX1	Intron 2	5,896	GCG ATA AAG TAT TAG TAA CCT
65FX2	Intron 2	6,329	AGT TAT TGC ATT CTT GTC AG
65FX3	Intron 2	6,720	TAT GAA GCC GCC ATG TGG
65FX4	Intron 2	7,056	TGT CAA TAC TAT GTG CTT GG
65FX5	Intron 2	7,370	CTA TGG ATT GAT CAT AAT GC
65FX6	Intron 2	7,769	TGG TTG ACA CGT GAT CTA C
65FP1	Intron 2	8,087	CAC AGC TTT CCA ATG CAC
65FP2	Intron 3	8,417	TTG CCT TAC CTT TCT CAC
65FP3	Intron 3	8,778	GGG AAT CTA CAT CAC TGA G
1E5F	Exon 5	9,118	ATA AGC AGA ATT CGT TCC
1E7F	Exon 7	9,536	GAC TTG CAC AAC AAT AAC C
1E8F	Exon 8	9,795	TGA CTC TCG GAA CTA TTC TC
65FP4	Intron 8	10,069	TGA TCA GTC CCT TAT AGA TG
65FP5	Intron 8	10,469	TTG CTG GTA TTT CTA ACA TC
65FP6	Intron 8	10,820	AAT TTC AAA CTC CAC GAC
65FP7	Exon 9	11,120	AAA TGT CAG AGA AGG ATT TC
65RP7	Intron 2	6,012	CCA ACT CCT TAT GAT CCA C
65RP6	Intron 2	6,342	AGA ATG CAA TAA CTT CAG TG
65RP5	Intron 2	6,727	GCT TCA TAA ATC TGT ATT TGG G
65RP4	Intron 2	7,076	ACC AAG CAC ATA GTA TTG AC
65RP3	Intron 2	7,458	AAT TAA CCC TGA GTG TTT G
65RP2	Intron 2	7,718	TGT AGA TCA CGT GTC AAC C
65RP1	Intron 2	7,938	TGG ATT GTA GAA TTG ATG C
65RC8	Intron 3	8,718	TAG GAG CAT AGC GAA AGG
65RC7	Exon 5	9,121	TTA TTC CTT TCT CAA GTC G
65RC6	Intron 6	9,473	AAG GAA TAA GTT GAC CAT AG
65RC5	Intron 8	9,908	CCT ATT ATT CGT TAC AAA GG
65RC4	Intron 8	10,179	CAG CAA CAG CTC TTT ACT C
65RC3	Intron 8	10,578	AGC ATT AGT GCA CCA TTG
65RC2	Intron 8	10,869	CAA TAA TCA GCA ATG GAA TC
65RC1	Exon 9	11,179	TGT AAC AGC ATC TTA ATG C
65RC	3' flanking	----	CTA GCA CCA AAG ATC AAT AG

PTAG2 Genomic Sequencing Primers

The series of reverse primers for the 5.7H subclone, which contains the 5' flanking region and extends 3' into Intron 2 (Appendix D), start near the Sst1 site (located in Exon 2).

Name	Position		Sequence
57FH1	5' flanking	----	CAC AAA GCA CTA GCT TGG
57FH2	5' flanking	339	AGC GGA ACT AGA CTT TGC
57FH3	5' flanking	691	TGT TAG CCT CCG ACA TTG
57FH4	5' flanking	1,223	CAT AGA AAT TTG GTA CTT GC
57FH5	5' flanking	1,724	AAG GAC GTA GGA ATC TCC
57FH6	5' flanking	2,106	GGT CTC AGT CTT CCT ATA GC
57FH7	Intron 1	2,466	AGT GCT TCA CTT GTT TGG
57RH7	5' flanking	492	CTA ATC TCT AGC AAT CCT AAC
57RH6	5' flanking	865	GCT TGA TTA AGT ACA GAA CTA C
57RH5	5' flanking	1,241	CAA GTA CCA AAT TTC TAT GG
57RH4	5' flanking	1,744	CAA GGA GAT TCC TAC GTC
57RH3	5' flanking	2,064	TTG GCT TCC ATG TGA ATG
57RH2	Exon 1	2,372	AGG AGA GAA AGT TAA GGA TC
57RH1	Intron 1	2,694	AGC CCT TGT TAT GGT TTC ATG
57RH	Exon 2	3,027	ACT TGG CGA TTG GTG GTG TTC

PTAG2 Genomic Sequencing Primers, continued

Forward and reverse primers for sequencing the 6BS subclone, which extends from Exon 2 to Intron 8, and for sequencing the remaining 3' region, which is contained in subclone 4.1E (Appendix D).

Name	Position		Sequence
PTAGF1	Intron 2	3,232	TTT CTC ATC CAA AGG GAT TTG
PTAGF2	Intron 2	3,583	CAC CCT TAG ATG ATT GTG ATG
PTAGF3	Intron 2	3,949	GTC AGT CGG ACA TAA GAA G
6BSF4	Intron 2	4,296	GCA AAC TAT TCA ATC AAC CAT G
6BSF5	Intron 2	4,671	ATG GAA CAA CGA GAA AGG
6BSF6	Intron 2	5,170	ACA GCT TGG CAG AGA CAG C
6BSF6A	Intron 2	5,160	GGA CTC AGT GAC AGC TTG
6BSF7	Intron 2	5,474	CCT CTT CAG TCA ATT TCA TC
6BSF8	Intron 2	5,796	GTG CTC TTA CCA TCT TGG
6BSF9	Intron 2	6,161	TCA GTT GAG TTA CCA GTG C
6BSF10	Intron 2	6,542	TGT TTC AGT TTC TTG GTA AG
6BSF11	Intron 2	6,768	AGT GAC TTT CTG AGG TAT
6BSF12	Intron 3	7,222	TTT GTA ATG TAT GCA CTA GG
6BSF13	Intron 3	7,685	ACT GCA TTG GGA GTC ATG
6BSF14	Exon 5	8,118	GTG CAT TGA GTG TGA AGG
6BSF15	Intron 7	8, 626	AAT ATC TTA GAT CGT TGG TG
6BSR14	Intron 2	3,426	GAA GAT CTA GCT CAG AAG C
6BSR13	Intron 2	3,705	CAG CCC TCT ACT TGT TTC
6BSR12	Intron 2	4,094	ATG AAG AAA CCC ATT TGG
6BSR11	Intron 2	4,611	TGG ACA ATA TTC TGT GTG G
6BSR10	Intron 2	4,969	CTT GAA TGT GAC TGA AAC TC
6BSR9	Intron 2	5,316	CCA GTT CTT TGT GCA TCC
6BSR8	Intron 2	5,698	CTT TAG ACA AAC TAC TCA
6BSR7	Intron 2	6,050	CGA TAA TGC ATA TGT AGG TG
PTAGR6	Intron 2	6,302	TCA CTT CCC ATG GTC TTC
PTAGR5	Intron 2	6,662	AAA GGG AAT TGG GTT CAG TAG
PTAGR4	Intron 2	6,973	ATA ATG TGG AAG AAT GTC ACC
PTAGR3	Intron 3	7,520	ACG CTG TTA ATG AAT AGT GAG
PTAGR2A	Exon 4/Intron 4	8,014	TCA TAC CTG TTT GAA TTC TG
PTAGR2	Intron 5/Exon 6	8,338	GCT CAT TCT GGC AGG TTT GAG
PTAGR1	Exon 8	8,823	AGG CGG CAA TCC ATT ACC TTG
PTAGFE	Intron 8	8,943	GAT GGG TTT CAG TAG CAG CAG
PTAGFE1	Intron 8	9,282	AAC TCC TTG ACG AAT CTG CTG
41FE2	Exon 9	9,566	TCC AGT CCC AGT TAA GAC
41RE2	Intron 8	9,320	AAC CAG CAA AGG AAT CAG
41RE1	Exon 9	9,641	GCA TCT TAG TGC CTT CAT G
41RE	3' flanking	----	TAG GCA TGC ACT GGA TTG

PTAG2 cDNA sequencing primers

Name	Position		Sequence
AG2F	MADS-box	265	CTG AGG TTG CAC TCA TCG
AG2F2	C-terminal	679	AGC ACA TGA ATT TGA TGC
AG2R2	K-box	415	TGA TAG AAC TGA GCA TTG G
AG2R	3' UTR	857	ATA TGG AAA TGC GAG GAG

PTAG1 PCR Primers

Primer pairs are listed consecutively.

Name	Position		Sequence	Product Size (bp)
PTAG1F1	C-terminal	772	CGA CAG AGC ATG AAT TTG	274
PTAG1R1	3' UTR	1,045	CCA GAC AAA TAT GAT TTA CG	
PTAG1F1	C-terminal	772	CGA CAG AGC ATG AAT TTG	146
PTAG1R2	C-terminal	917	ACT AAC TGA AGG GCC ATC TG	
AG1F3	3' UTR	972	CCT GGG TTT CCA TTG AGC	115
AG1R3	3' UTR	1,086	GGA TAG TTA ATA CAT AGA GGA AGA G	
AG1RTF	5' UTR	107	AAG ATC CTC ACT TTC TCT ACA C	940
AG1RTR	3' UTR	1,046	CCC AGA CAA ATA TGA TTT AC	

PTAG2 PCR Primers

Name	Position		Sequence	Product Size (bp)
PTAG2F1	C-terminal	675	CGA CAG CAC ATG AAT TTG	146
PTAG2R2	C-terminal	*816	TTA CAC TAA CTG AAG AGC TGG	
AG2F3	3' UTR	868	GTA CCT ACT ATT TCA CTG AGC G	118
AG2R3	3' UTR	985	AAA GCA ATA CAT GGA GGA AGA G	
AG2RTF	5' UTR	16	AAG ATC CTT ACT TTC TCT CC	877
AG2RTR	3' UTR	892	TTG CGC TCA GTG AAA TAG	

* First bp that is located in exon 8; first 4 bp of primer are in intron 8.

APPENDIX F: DEPOSITION OF MATERIALS

-80°C Glycerol Stocks

Original *PTAG* cDNA clones are in pBluescript SK (Stratagene). All *PTAG* subclones are in pBluescript KS. Relevant paperwork is deposited in the laboratory inventory notebook.

Name	Description	Location
pCPTAG1	*PTAG1* cDNA clone	B1, B2
pCPTAG1A	*PTAG1* cDNA subclone (see Appendix D)	B3, B4
pCPTAG1B	"	B5, B6
pCPTAG1C	"	B7, B8
pCPTAG1D	"	B9, B10
pCPTAG1E	"	B11, B12
pCPTAGPX	"	B13, B14
pPTAG1P	274 bp *PTAG1* 3' PCR fragment (primers: PTAG1F1/R1)	B15, B16
p3UTRAG1	115 bp *PTAG1* 3' UTR PCR frag. (primers: AG1F3/R3)	B17, B18
pCPTAG2	*PTAG2* cDNA clone	B19, B20
pPTAG2P	146 bp *PTAG2* C-term. PCR frag. (primers: PTAG2F1/R2)	B21, B22
p3UTRAG2	118 bp *PTAG2* 3' UTR PCR frag. (primers: AG2F3/R3)	B23, B24
p3S	*PTAG1* genomic subclone (see Appendix D)	B25, B26
p2.3S	"	B27, B28
p5.3BX	"	B29, B30
p2.8XP	"	B31, B32
p6.5XB	"	B33, B34
p6.5SX	"	B35, B36
p9.6BX	"	B37, B38
p5.7H	*PTAG2* genomic subclone (see Appendix D)	B39, B40
p6BS	"	B41, B42
p4.8E	"	B43, B44
p4.1E	"	B45, B46
pCIT565	*AG* cDNA in pGEM7Zf (Promega)	B47, B48
pAM116	*AP1* cDNA in pGEM7Zf	B49, B50
pPD4	*P.deltoides* 25S rDNA EcoRV/EcoRI frag. in pSP72 (Promega)	R34, R35
pPD5	*P.deltoides* 18S rDNA SstI/XbaI frag. in pBS (Stratagene)	R36, R37
pPD6	*P. deltoides* 18S rDNA EcoRV/SstI frag. in pSP72 (Promega)	R38, R39

Miscellaneous

PTAG primers are stored in the FSL 075 -20°C freezer in Nalgene CyroBoxes labeled "PTAG1 primers" and "PTAG2 primers".

The amplified male *P. trichocarpa* genomic library, prepared 7/28/94, is cloned into λGEM12 (Promega). The titer of the library was 1.7×10^{10} pfu/ml. It is stored in 50 ml Corning tubes in the glass-front refrigerator in FSL 075. Additional aliquots are stored on the second shelf of the -80°C freezer in FSL 077 in a cardboard freezer box labeled "poplar genomic library".

An electronic copy of this thesis is located at p:\tgerc\thesis\amy\ambthes.doc