Studies of Genes Involved in Regulating Flowering Time in *Arabidopsis thaliana*

Maria Svensson
Faculty of Natural Resources and Agricultural Sciences
Department of Plant Biology and Forest Genetics
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2006

Acta Universitatis Agriculturae Sueciae
2006: 2

ISSN 1652-6880
ISBN 91-576-7051-X

Tryck: SLU Service/Repro, Uppsala 2006

Abstract

Svensson, M. 2006. Studies of genes involved in regulating flowering time in *Arabidopsis thaliana*.
Doctor's dissertation.
ISSN 1652-6880, ISBN 91-576-7051-X.

Transition from a vegetative growth phase to flowering in plants occurs in response to both environmental conditions and endogenous signals. Identification of genes that are involved in regulating the time of flowering is of great importance in agri- and horticulture. Flowering-time genes can be used for crop improvement by, for instance, engineering plants to flower earlier. This shortening of the time to flowering could result in an extended growing season that could enable farmers to grow more than one crop each year.

In this work, a gene knockout approach using T-DNA tagging and *in vivo* gene fusion has been employed to identify and characterise genes that are involved in regulating flowering time in the model plant *Arabidopsis thaliana*. This approach resulted in the identification of two genes, *At4g20010* and its homologue *At1g31010*. Expression studies and GUS histochemical analysis of a reporter gene revealed that *At4g20010* is mainly expressed in rapid growing tissues such as root tips, shoot apex, flowers and stem nodes. T-DNA insertional mutants of *At4g20010* and *At1g31010* exhibit a late-flowering phenotype that can largely be repressed by application of gibberellin. Plants with an insertional mutation in *At4g20010* contain a reduced amount of the bioactive gibberellin GA_4 compared to wild-type plants. The decreased level of GA_4 is not due to a transcriptional repression of the GA-biosynthetic genes *AtGA3ox1* or *AtGA20ox1*, since their expressions were increased in the mutant plants. *In silico* analyses revealed that the C-terminal protein sequences encoded by *At4g20010* and *At1g31010* contain RNA-binding motifs, whereas the N-terminal sequences have three-dimensional structures similar to single stranded nucleic acid-binding proteins. To conclude, *At4g20010* and *At1g31010* may encode two RNA-binding proteins that are involved in regulating flowering time in *A. thaliana* by affecting the metabolism of GA. This can be possible either by a positive regulation of GA3ox at the post-transcriptional level or by a negative regulation of GA2ox.

Keywords: *Arabidopsis thaliana*, flowering time, fold recognition, GA biosynthesis, promoter trapping, RNA-binding, T-DNA tagging

Author's address: Maria Svensson, Department of Life Sciences, University of Skövde, Box 408, SE-541 28 Skövde, Sweden. E-mail: Maria.Svensson@his.se

Appendix

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

I. Svensson, M., Lundh, D., Ejdebäck, M. & Mandal, A. 2004. Functional prediction of a T-DNA tagged gene of *Arabidopsis thaliana* by *in silico* analysis. *Journal of Molecular Modeling* 10, 130-138.

II. Svensson, M., Lundh, D., Bergman, P. & Mandal, A. 2005. Characterisation of a T-DNA-tagged gene of *Arabidopsis thaliana* that regulates gibberellin metabolism and flowering time. *Functional Plant Biology* 32, 923-932.

III. Svensson, M., Lundh, D., Bergman, P. & Mandal, A. 2005. *At4g20010* and its homologue *At1g31010* encode two putative nucleic acid-binding proteins involved in regulating flowering time in *Arabidopsis thaliana*. (Manuscript).

Contents

Abbreviations

CDP	*ent*-copalyl diphosphate
CPS	*ent*-copalyl diphosphate synthase
FMI	Floral meristem identity
GA	Gibberellin
GA13ox	GA 13-hydroxylase
GA20ox	GA 20-oxidase
GA2ox	GA 2-oxidase
GA3ox	GA 3β-hydroxylase
GA-3P	Glyceraldehyde-3-phosphate
GGDP	Geranylgeranyl diphosphate
I-PCR	Inverse-PCR
IPP	Isopentenyl diphosphate
KAO	*ent*-kaurenoic acid oxidase
KO	*ent*-kaurene oxidase
KS	*ent*-kaurene synthase
MVA	Mevalonic acid (Mevalonate)
PDB	Protein Data Bank
RMSD	Root mean square deviation
RT-PCR	Reverse transcriptase PCR
UTR	Untranslated region

Introduction

Arabidopsis thaliana

During the last decades *Arabidopsis thaliana* has become one of the most widely used model plant in biological research. Its relatively small genome (~125Mbp) with low amounts of repetitive sequences, short generation time, and its close relationship to the organisms it is meant to model, are some of the reasons why Arabidopsis has been adopted as a model system for molecular and genetic studies (Meyerowitz, 1987; Meinke *et al.*, 1998). Arabidopsis's value as a model plant increased even more at the end of year 2000 when it was the first plant to have its genome sequenced (The Arabidopsis Genome Initiative, 2000). At that time it was the third genome of a higher eukaryote, after *Caenorhabditis elegans* and *Drosophila melanogaster*, that was completely sequenced (The *C. elegans* sequencing consortium, 1998; Adams *et al.*, 2000). The Arabidopsis ecotype Columbia was sequenced by a public consortium (The Arabidopsis Genome Initiative, 2000), whereas the private company Cereon sequenced the ecotype Landsberg *erecta* (Jander *et al.*, 2002). After completion of the sequencing process, the genes and other features of the entire Arabidopsis genome were annotated (The Arabidopsis Genome Initiative, 2000). The annotation process is iterative and on-going, and since the end of year 2000 the genome has been reannotated by TIGR, The Institute for Genomic Research (Wortman *et al.*, 2003). The total number of Arabidopsis genes was initially estimated to 25,498; however, the latest released version of the TIGR ATH1 genome (Version 5) includes an estimated number of 30,700 genes.

Approximately one third of the initially predicted genes could not be assigned any biological function (based on homology searches), and only about 10% of the annotated genes have yet been thoroughly established to have a definitive function (The Arabidopsis Genome Initiative, 2000; Ostergaard & Yanofsky, 2004). These numbers will probably increase drastically in the next few years as the genetic resources in Arabidopsis research have recently been boosted (Ostergaard & Yanofsky, 2004). For example, the development of different public collections of T-DNA tagged lines enables researchers to search a sequence database and find a mutant line with an insertion in their gene of interest. The largest insertion collection for *Arabidopsis thaliana* was created by Ecker and co-workers at the Salk Institute and it contains more than 225,000 T-DNA tagged lines. For approximately 90,000 of these lines the location of the T-DNA has been determined by sequencing, and it revealed that about 22,000 of the Arabidopsis genes contain T-DNA insertions (Alonso *et al.*, 2003). The SALK T-DNA collection together with several other insertion collections that have been developed for *Arabidopsis thaliana*, including the SAIL collection, very much contributes to the process of determining gene function through reverse genetics (Balzergue *et al.*, 2001; Sessions *et al.*, 2002; Till *et al.*, 2003).

Another resource that will be helpful in assigning biological function to the Arabidopsis genes is the growing number of publicly available gene expression profiles (http://www.ag.arizona.edu/microarray; 5-Dec-2005; https://www.genevestigator.ethz.ch; 5-Dec-2005; http://www.affymetrix.com/products/arrays/specific/arab.affx; 5-Dec-2005). Through microarray analysis the expression level of genes can be analysed in different tissues during different developmental stages and stress conditions, and in this way potential functions of the genes can be revealed.

T-DNA tagging

The insertion of foreign DNA into a plant genome is a powerful approach for identifying new genes and determining gene function. A knockout mutation can be generated by inserting a DNA segment with a known sequence into a plant gene, for instance by disrupting the expression of the gene. The knockout of the plant gene may consequently result in plants with a recognisable mutant phenotype. The insertion of the T-DNA does not need to be occurred in the exon of a gene to result a mutant phenotype. Several researchers have shown that insertions can also occur in introns and in 5´ or 3´ non-coding regions as well as resulting plants with mutant phenotypes (reviewed by Azpiroz-Leehan & Feldmann, 1997). The insertional mutagen does not create only a mutation, it also ´tags´ the affected gene. This enables researchers to identify the gene in question. The tagged gene can be identified and isolated by amplifying and sequencing the plant DNA flanking the known insert.

One of the most commonly used methods for transferring foreign DNA into plants genome is *Agrobacterium*-mediated transformation (Topping *et al.*, 1995; Tinland, 1996; Zupan *et al.*, 2000). In this method researchers take advantage of the soil bacterium *Agrobacterium tumefaciens*'s natural ability to transfer a fragment of its own DNA into plant genomes. The DNA that is being transferred (T-DNA) is flanked by 25 bp imperfect direct repeat border sequences, named right and left border. In theory, only the sequence within these borders (the T-DNA sequence) is transferred to the plant genome by *Agrobacterium* in a random manner. However, in practice sequences outside the T-DNA borders that belong to the transformation vector can also be transferred to the plants (Ramanathan & Veluthambi, 1995; Kononov, Bassuner & Gelvin, 1997; De Buck *et al.*, 2000). Other rearrangements of the T-DNA and the plant DNA sequence at the site of insertion have also been observed by several researchers (Mayerhofer *et al.*, 1991; Ohba *et al.*, 1995; Forsbach *et al.*, 2003). Despite these rearrangements that can occur, the *Agrobacterium* mediated T-DNA transfer system is usually the method of choice since, in comparison with other transformation methods, it usually results in stable transgenes that are intact, non-rearranged and that exist in a low copy number (Gelvin, 1998). Feldmann (1991) showed that the average number of independent inserts was 1.5 per diploid genome, where 57% of the transformed plants contained a single insert and 25% of the plants contained two inserts. Similar results have also been obtained for other T-DNA insertion collections (McElver *et al.*, 2001; Alonso *et al.*, 2003).

An additional feature of the T-DNA tagging approach is the use of *in vivo* gene fusion technology. In this technique, a T-DNA vector containing a promoterless or enhancerless reporter gene placed at the right or left end of the T-DNA is employed for gene tagging. Following random insertion of the T-DNA into plant genomes a transcriptional or translational gene fusion between the plant gene and the promoterless reporter gene can be achieved and identified by screening the plants for the activity of the reporter gene (Topping & Lindsey, 1995). In a promoter trap approach the promoterless reporter gene will be activated when inserted downstream of a native plant gene promoter (Fig. 1). An advantage of the promoter trap approach compared to regular T-DNA insertion mutagenesis is that it relies not only on the ability of generating a mutant phenotype but also reveals information about the expression pattern of the tagged gene. This is because of the fact that the expression pattern of the reporter gene usually reflects the expression of the tagged gene (Topping *et al.*, 1995). The first vector designed for promoter trapping was developed by Koncz and co-workers (1989) and contained aminoglycoside (kanamycin) phosphotransferase as a reporter gene. Other reporter genes used frequently in promoter traps are the *uidA* (β-glucuronidase; GUS), green fluorescent protein (GFP) and the luciferase genes (Riggs & Chrispeels, 1987; Kertbundit *et al.*, 1991; Topping & Lindsey, 1995; Stewart, 2001; Ryu *et al.*, 2004).

Although several different genes have been identified by the promoter trap approach it is not always easy to demonstrate an exact correlation between the expression pattern of the reporter gene and the tagged gene (Pereira, 2000). For example, Stangeland and co-workers (2005) observed GUS activity even when the promoterless *gus* reporter gene was inserted in intergenic regions and in inverted orientation in respect of the direction of the promoter of the tagged gene. This activation can be explained by the presence of cryptic promoters or by promoters of still unannotated genes (Stangeland *et al.*, 2005). However, when the promoterless reporter gene is inserted in the same orientation as the tagged gene in the 5´UTR or in the intron region, the expression pattern of the reporter gene greatly reflects the pattern of the tagged gene (Stangeland *et al.*, 2005).

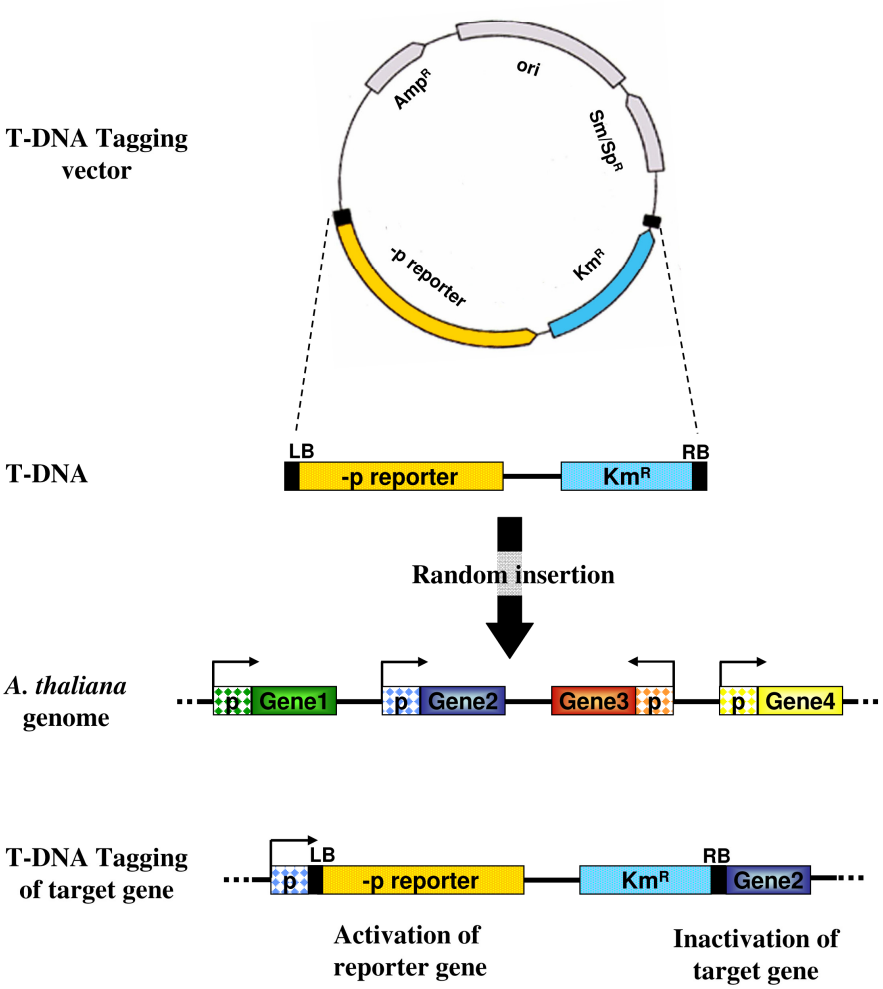

Fig. 1. Schematic presentation of T-DNA mediated gene tagging using a promoter trap and *in vivo* gene fusion. LB, left border; RB, right border; p, promoter; -p reporter, promoterless reporter gene; Km^R, kanamycin resistant selectable marker gene.

Protein structure prediction

The sequencing and annotation of the Arabidopsis genome have revealed thousands of genes and still today, five years after the release of the full sequence, most of the proteins encoded by theses genes have not been assigned any function, despite structural genomics initiatives and biochemical efforts. Today, the fastest way to achieve information about a protein's function is through computational methods. By sequence comparison and other analysing tools a protein's structure and function can be predicted by establishing the relationship to other structurally and functionally determined proteins. To achieve a deeper understanding of the biological function of a protein it can be essential to know its three-dimensional

structure (Pawlowski *et al.*, 2001). The three-dimensional structure can reveal details of binding, catalysis and signalling, events that happen at the molecular level and that control the function of a protein (Thornton *et al.*, 1999). Another reason for also analysing a protein's three-dimensional structure and not only its sequence is that the three-dimensional structures of proteins are better conserved than their sequences during evolution. As a consequence, two proteins can share a similar structural fold even though they are not similar at the sequence level. This is why a structure-structure comparison of proteins can reveal novel and complementary information about the relationship between proteins than sequence-sequence comparison alone (Xu, Xu & Uberbacher, 2000).

The three-dimensional structure of a protein can be predicted either by a template-based method, which uses a known protein structure as a template, or by an *ab initio* method. The *ab initio* method predicts a protein structure by optimising the energy function that describes the physical properties of the amino acids. As a consequence, and in contrast to the template-based method, it predicts the three-dimensional protein structures without any reference to existing protein structures. The *ab initio* method requires long computing time and powerful computers, and the prediction programmes are normally not readily available (Xu, Xu & Uberbacher, 2000). This makes the process of structure prediction difficult and time consuming and explains why the method of choice for structure prediction is often the template-based method. Template-based prediction includes the methods of comparative (homology) modelling and fold recognition (threading). A schematic summary of these methods are presented in Fig. 2. The first step in predicting a protein's three-dimensional structure by the template-based method is to find out if the protein sequence has any similarity to any sequence with an already known structure. This can be performed by searching a structural database, e.g. the Protein Data Bank, PDB (Berman *et al.*, 2000). If a search against the PDB database reveals a match to a protein with a sequence identity of above 30%, the comparative modelling approach can be performed. The comparative modelling approach uses the theory that similar sequences have similar structures, and following this it assembles the coordinates of all the atoms in the protein to be predicted based on a sequence alignment between the protein and a similar protein with a known structure. The success of the outcome from comparative modelling is highly dependent on high sequence identity between the query protein and the template with a known structure, this in order to be able to select the correct template and to create a good alignment. As mentioned previously, a 30% sequence identity between the query sequence and the template is needed in order to perform comparative modelling successfully. However, this percentage is usually applied when using multiple sequence alignments. If pairwise sequence alignments are used then a sequence identity of around 50% is recommended in order to obtain a satisfactory result. Although the models built by using the comparative modelling approach are reasonably accurate, the drawback of this method is that it is very dependent on the existence of a good sequence homologue and on the quality of the produced alignment. In other words, it might not be applicable to a large fraction of protein sequences.

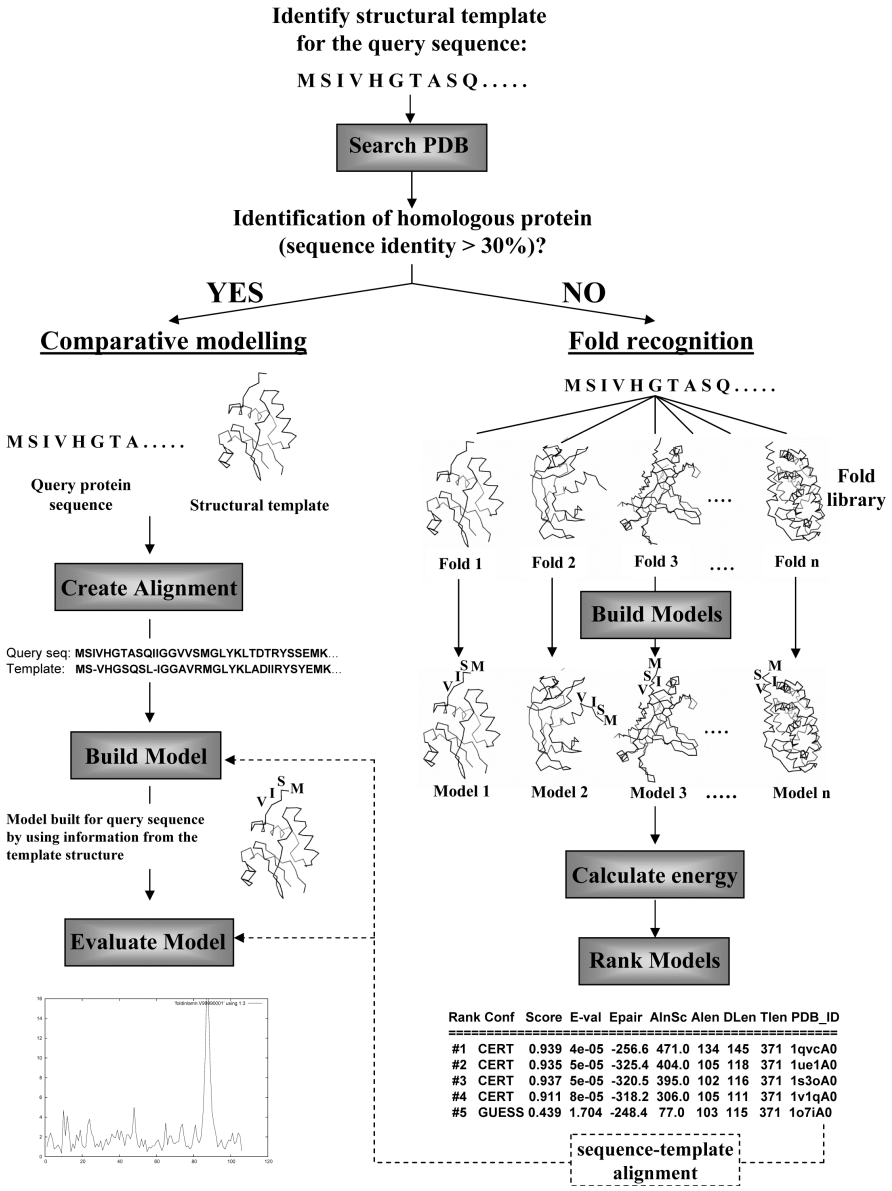

Fig. 2. Schematic view of the template-based approach for prediction of protein structure. Adapted from Jones & Hadley (2000) and Marti-Renom *et al.* (2003).

As mentioned previously, it is not only proteins with similar sequences that adopt similar folds; also proteins with no detectable sequence similarity can have structural similarities. It has often been shown that proteins which appear to be unrelated at first, when comparing the protein sequences, still adopt similar folds. In fact, more than 50% of proteins that were believed to be unrelated to any known protein were later shown to have a well known fold (Godzik, 2003). Studies have also shown that sequences with a sequence identity of less than 5% still adopt the

same fold (Orengo *et al.*, 1993). Because of these facts a new method for predicting protein structure was developed in the 1990's, the fold recognition or threading method (Bowie, Luthy & Eisenberg, 1991; Godzik & Skolnick, 1992; Jones, Taylor & Thornton, 1992). The fold recognition method is based on the suggestion that there are a limited number of possible folds and that a large percentage of the proteins adopt one of these folds (Chothia, 1992). Rather than trying to find the accurate structure for a protein by considering all the possible conformations available, the fold recognition method uses the fact that the correct structure probably has been observed previously and that the structure already is stored in a database (Jones & Hadley, 2000). The fold recognition method analyses and finds folds that are compatible with a query sequence, *i.e*. instead of predicting how the sequence will fold it predicts how well different folds will fit the sequence. The basic idea behind the fold recognition method is outlined in Fig. 2 and described as follows. At first, a query sequence is searched through a library of representative protein structures in order to find the best fit. In this search the query sequence is optimally fitted to all the folds in the library by creating sequence-structure alignments (Jones & Hadley, 2000). Later, an energy function is used to describe the fitness of the alignments between the query sequence and the template fold. In order to find the most optimal alignments a threading algorithm is used to search for the possible alignments with lowest energy (Xu, Xu & Uberbacher, 2000). The final output of the fold recognition method is a ranking of the folds in the library, where the top-ranked fold is the one used to create the alignment with lowest energy and as a consequence the most probable fold for the query sequence (Jones & Hadley, 2000).

Once the template structure has been identified and an alignment built, either by a comparative method or by a threading process, several different methods can be used to build a three-dimensional structure model for the query protein. Some methods are based on the assembly of rigid fragments where equivalent fragments are extrapolated from the known protein structure to the query sequence. Other methods use restraints such as interatomic distances in order to build models of the query sequence that has the best agreement with the template structure (Srinivasan, Guruprasad & Blundell, 2002). Usually several different models are built, and then the quality of the models are evaluated. This can, for example, be performed by comparing the built model with the template structure in terms of visual inspection and root mean square deviation (RMSD), and by evaluating the stereochemistry of the model (Edwards & Cottage, 2003).

Flowering time

The control of flowering time is of great importance in agriculture and also of great scientific interest for understanding the mechanism underlying plant development. Genes that regulate flowering time of plants can be used for crop improvement by, for instance, engineering plants to flower earlier. This could, for example, extend the growing season for grains and fruits and perhaps enable farmers to grow more than one crop each year (Moffat, 2000). Shortening the time to flowering would also contribute to early maturity and permit a more northerly cultivation. The

transition from a vegetative growth phase to flowering in plants has in the past been studied almost entirely by plant physiologists. However, during the last decade molecular genetics have provided powerful new tools for studies of developmental processes. During the end of the 1990's, several genes involved in regulating the flowering time were described (Koornneef, 1997). Today, more than 80 different flowering-time genes have been discovered in *Arabidopsis thaliana* (Blazquez, Koornneef & Putterill, 2001), and the genetic analysis of these genes has led to identification of four major pathways controling flowering time (reviewed by Araki, 2001; Mouradov, Cremer & Coupland, 2002; Simpson & Dean, 2002; Bastow & Dean, 2003; Komeda, 2004; Parcy, 2005). The four major pathways are the photoperiod-, the vernalisation-, the autonomous- and the gibberellin pathway. The photoperiod and the vernalisation pathways mediate flowering in response to environmental factors such as day length and low temperature, whereas the autonomous and the gibberellin pathways mainly act independently of these external signals (Mouradov, Cremer & Coupland, 2002). Two of the most prominent genes of these flowering pathways are the *CONSTANS* (*CO*) and the *FLOWERING LOCUS C* (*FLC*) genes (Parcy, 2005). The *CO* gene promotes flowering and plays a key role in the photoperiod pathway. The expression and the protein accumulation of *CO* are regulated by light and the circadian rhythm (Valverde *et al.*, 2004). The *co* mutants flower late in long-days but similarly or identically to wild-type plants in short-days (Koornneef, Hanhart & van der Veen, 1991). Another prominent flowering-time gene, *FLC,* represses flowering and is the convergence point for the vernalisation and the autonomous pathways (Michaels & Amasino, 1999). Both the vernalisation and the autonomous pathways promote flowering by repressing *FLC* (Amasino, 2005).

A plant switches from a vegetative to a reproductive growth phase, as environmental conditions and endogenous signals influence a change in the identity of cells of the shoot apical meristem (Coupland, 1995). Leaves, stems and flowers are all derived from cells within the shoot apical meristem, but during the onset of flowering these cells change their identity so that they give rise to floral structures and stems rather than leaves (Coupland, 1995). More specifically, the signals following the flowering pathways will result in an induction of the floral meristem identity (FMI) genes *LEAFY* (*LFY*), *APETALA1* (*AP1*) and *CAULIFLOWER* (*CAL*). The induction of these FMI genes changes the fate of the meristem and initiate the development of floral primordia (Parcy, 2005). However, most of the flowering-time genes within the different pathways do not directly induce the FMI genes, instead a few regulators known as Floral Pathway Integrators are able to integrate the cascade of inputs from the different flowering pathways and convert it into an induction of the FMI genes (Simpson & Dean, 2002; Parcy, 2005). Genes that are considered to be Floral Pathway Integrators are *FLOWERING LOCUS T* (*FT*), *SUPPRESSOR OF CO OVEREXPRESSION* (*SOC1*) and *LFY* (Simpson & Dean, 2002; Parcy, 2005). The gibberellin pathway promotes flowering by inducing at least two of these three Floral Pathway Integrators, namely *SOC1* and *LFY*, and possibly also *FT* (Parcy, 2005).

Gibberellins

Gibberellins (GAs) are a group of tetracyclic diterpenes that regulate growth and especially control developmental processes such as seed germination, stem elongation, flowering and fruit development of plants (Davies, 1995; Ross, Murfet & Reid, 1997). There are two main types of GAs, the C20-GAs which contain 20 carbon atoms, and the C19-GAs, which have lost the twentieth carbon through metabolism (Cleland, 1999). Gibberellins were first isolated in 1926 from the fungus *Gibberella fujikuroi*, and since then several additional GAs have been identified from various species. At the time of writing of this thesis, 136 different GA structures have been identified in plants, fungi and bacteria. The nomenclature of each different GA, which is found to be naturally occurring and whose structure has been chemically characterised, is numbered in an approximate order of discovery. Not all of the GAs have high biological activity, in fact only a few of the identified compounds have been shown to give rise to biological responses in plants (Hedden & Phillips, 2000). Most of the GAs within a plant are precursors or catabolites of the biologically-active GAs. Examples of biological active compounds in higher plants are GA_1, GA_3, GA_4 and GA_7 (Hedden & Phillips, 2000).

GA mutants

The most conspicuous phenotype of plants with mutations concerning gibberellin is dwarfism. Plants with this dwarfed phenotype can be divided into two different groups: those that respond to applied GA and those that do not. Mutant plants that respond to exogenous GA are usually called biosynthesis- or GA-sensitive mutants (Hedden, 1999). These mutants have a reduced GA biosynthesis and their mutant phenotype can be restored to wild-type by the application of exogenous GA. Mutant plants that are not rescued by the application of GA are called response- or GA-insensitive mutants (Hedden, 1999). The use of genetic mutants that are affected in their biosynthesis of GA or in their response to GA has resulted in a deep understanding of the general pathways of gibberellin metabolism and an increasing knowledge of the GA signalling process.

GA biosynthesis

Today, most of the genes encoding GA biosynthetic enzymes in *A. thaliana* have been identified (Hedden & Phillips, 2000). This knowledge has been obtained by studying plants that contain gene mutations disrupting the GA biosynthesis. The GA biosynthetic pathway in higher plants can be classified into three different stages: (1) the first stage includes the synthesis of *ent*-kaurene in the plastids; (2) the second stage includes the conversion of *ent*-kaurene to GA_{12} and occurs on membranes outside the plastid; (3) the third and last stage includes the synthesis of C_{19}- and C_{20}-GAs in the cytoplasm (Hedden & Phillips, 2000; Olszewski, Sun & Gubler, 2002). The first and the major part of the second stage are general for all plants, whereas after the formation of GA_{12}-aldehyde there can be variations between species (Srivastava, 2002). As a consequence, several alternative pathways of GA biosynthesis appear after the formation of GA_{12}-aldehyde. In this

thesis I have focused only on describing the main pathways. The steps involved in the GA biosynthetic pathway are outlined in Fig. 3. The GA biosynthetic pathway has also been reviewed previously by Hedden & Phillips (2000), Yamaguchi & Kamiya (2000) and in Olszewski, Sun & Gubler (2002).

(1) The formation of GAs begins with the cyclisation of geranylgeranyl diphosphate (GGDP). GGDP is synthesised from isopentenyl diphosphate (IPP), where four five-carbon isoprene units are joined to build up the C_{20} precursor GGDP (Srivastava, 2002). IPP is synthesised either through the mevalonate-dependent pathway in the cytosol or through the non-mevalonate pathway (utilizes pyruvate and glyceraldehydes-3-phosphate) in the plastid (Lichtenthaler *et al.*, 1997). The latter pathway is believed to be the main provider of IPP for GA biosynthesis since the following conversion of GGDP to *ent*-kaurene via *ent*-copalyl diphosphate (CDP) also occurs in the plastid. The conversion of GGDP to CDP and CDP to *ent*-kaurene is catalysed by *ent*-copalyl diphosphate synthase (CPS) and *ent*-kaurene synthase (KS), respectively (Hedden & Phillips, 2000).

Genes coding for both the enzymes, CPS and KS, together with related mutants have been identified in various plants (Hedden & Phillips, 2000). The gene encoding CPS was first cloned from the *GA1* locus of *A. thaliana* by genomic subtraction (Sun, Goodman & Ausubel, 1992; Sun & Kamiya, 1994). Mutations in the *GA1* locus result in plants with a GA-deficient phenotype, where the most severe mutant phenotype can be observed in *ga1-3* plants. *ga1-3* mutants cannot germinate without GA treatment, they are severely dwarfed, flower later under long-days and are unable to flower under short days (Koornneef & Van der Veen, 1980). The gene encoding KS was first cloned from pumpkin (*Cucurbita maxima*) and later also from *A. thaliana*, locus *GA2* (Yamaguchi *et al.*, 1996; Yamaguchi *et al.*, 1998a). The phenotype of *ga2-1* mutants is similar to that of *ga1-3*, showing a nongerminating and extreme dwarfed phenotype (Koornneef & Van der Veen, 1980).

(2) After formation of *ent*-kaurene, the substrate is transported from the plastid to membranes outside the plastid by a mechanism that today is not fully understood (Olszewski, Sun & Gubler, 2002). *ent*-Kaurene is oxidised to *ent*-kaurenoic acid by *ent*-kaurene oxidase (KO), and e*nt*-kaurenoic acid is later oxidised in a three step procedure to GA_{12} by the endoplasmic reticulum associated *ent*-kaurenoic acid oxidase (KAO). All above mentioned reactions in stage two are catalysed by cytochrome P450 monooxygenases (Helliwell *et al.*, 1998; Helliwell *et al.*, 2001). GA_{12} can further be converted into GA_{53} by GA 13-hydroxylase (GA13ox).

ent-Kaurene oxidase was first cloned from *A. thaliana* locus *GA3* by Helliwell *et al*. (1998), and its corresponding mutant, the GA-responsive dwarf *ga3*, was characterised by Koornneef and van der Veen (1980). All of the enzymes that appear early in the gibberellin biosynthetic pathway (CPS, KS and KO) are, in most plant species, encoded by a single gene, which explains why loss-of-function mutations at these loci result in plants with a severe dwarfed phenotype (Hedden & Phillips, 2000).

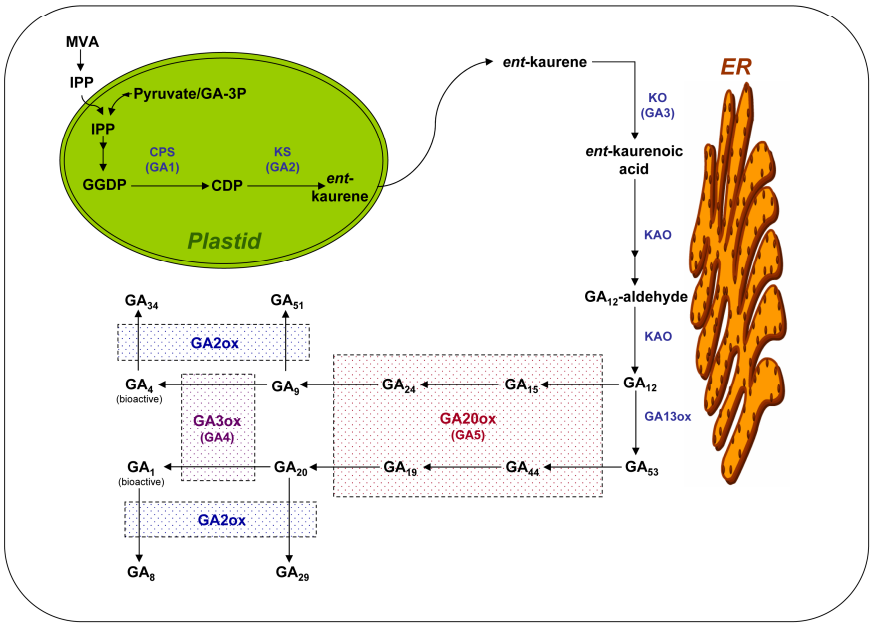

Fig. 3. Major steps of the GA biosynthesis pathway in higher plants. Mevalonic acid (MVA), Isopentenyl diphosphate (IPP), Glyceraldehyde-3-phosphate (GA-3P), Geranylgeranyl diphosphate (GGDP), ent-copalyl diphosphate (CDP), ent-copalyl diphosphate synthase (CPS), ent-kaurene synthase (KS), ent-kaurene oxidase (KO), ent-kaurenoic acid oxidase (KAO), GA 13-hydroxylase (GA13ox), GA 20-oxidase (GA20ox), GA 2-oxidase (GA2ox), GA 3β-hydroxylase (GA3ox). Genetic loci encoding GA biosynthesis enzymes in *Arabidopsis thaliana* are written in parenthesis.

(3) In the last stage of GA biosynthesis, GA_{12} and GA_{53} are, in parallel pathways, converted into various C_{19}- and C_{20}-GAs by a series of oxidative reactions involving 2-oxoglutarate-dependent dioxygenases. The pathways that convert GA_{12} and GA_{53} are called early non-hydroxylation pathway and early 13-hydroxylation pathway, respectively. The preference of pathway varies between species, for example in cowpea, rice and lettuce the early 13-hydroxylation pathway is the most dominant, and as a consequence these plants produce more of the bioactive GA_1 rather than GA_4. In Arabidopsis and cucumber the predominant pathway is the early non-hydroxylation pathway, and GA_4, instead of GA_1, plays the role of the main active hormone (Davies, 1995; Kamiya & Garcia-Martinez, 1999). The first reactions in this final stage of GA biosynthesis involve GA 20-oxidase (GA20ox) which converts GA_{12} and GA_{53} in a stepwise procedure to GA_9 and GA_{20}, respectively (Hedden & Phillips, 2000). GA_9 and GA_{20} can then be converted into the bioactive GAs, GA_1 and GA_4, by a 3β-hydroxylation, catalysed by GA 3β-hydroxylase (GA3ox). Another dioxygenase, GA 2-oxidase (GA2ox), can deactivate the biological active GAs, GA_1 and GA_4, by converting them into the biological inactive GA_8 and GA_{34} by a 2β-hydroxylation (Hedden & Phillips, 2000). The level of bioactive GAs can be regulated in an additional way since GA2ox also converts GA_9 and GA_{20} to GA_{51} and GA_{29}, respectively, and in this way it diverts the precursor GAs away from the route of becoming bioactive GAs.

Unlike CPS, KS and KO (enzymes that appear early in the pathway and that are encoded by a single gene), GA20ox, GA3ox and GA2ox are each encoded by a small gene family. At present, five, four and eight different genes that encode GA20ox, GA3ox and GA2ox, respectively, have been identified in *A. thaliana* (Hedden *et al.*, 2001; Schomburg *et al.*, 2003). Due to this redundancy of the GA oxidase genes, plants containing a null mutation in any of the genes within the gene family do not exhibit such severe GA-deficient phenotype as plants with mutations in the *CPS*, *KS* and *KO* genes (Talon, Koornneef & Zeevaart, 1990; Hedden & Phillips, 2000).

The first cloned gene to encode GA20-oxidase was isolated from pumpkin (*Cucurbita maxima*) by Lange, Hedden & Graebe (1994). Shortly afterwards, several genes encoding GA20-oxidases were isolated from *A. thaliana*, where *GA20ox1* corresponds to the Arabidopsis locus *GA5* (Phillips *et al.*, 1995; Xu *et al.*, 1995). The first GA3-oxidase encoding gene to be cloned was *GA3ox1* from the *GA4* locus of *A. thaliana* (Chiang, Hwang & Goodman, 1995; Williams *et al.*, 1998). No general expression patterns have so far been distinguished for any of the GA dioxygenase families. Instead, the genes within the same family are expressed differently in the plant tissues. On the other hand, orthologues of *GA20ox* in closely related species have been shown to have similar expression patterns (Hedden & Phillips, 2000). This can for example be seen when comparing the expression pattern of *GA20ox* in *A. thaliana* and rice (*Oryza sativa*). In *Arabidopsis thaliana* the *AtGA20ox1* is mainly expressed in growing vegetative tissues but it is also expressed in flowers (Hedden & Phillips, 2000). This expression pattern is similar to that observed in rice by Sakamoto *et al*. (2004). They could see the *OsGA20ox1* expression in both reproductive and vegetative organs. A similar expression pattern between Arabidopsis and rice can also be observed when studying *GA20ox3*. No expression of *GA20ox3* has been observed in the vegetative organs in either Arabidopsis or rice. Instead, its expression occurs only in the siliques and in the reproductive organs, respectively (Hedden & Phillips, 2000; Sakamoto *et al.*, 2004). In *Arabidopsis thaliana GA3ox1* is expressed mainly in siliques and germinating seeds but it is also active in seedlings, leaves, stems and flowers. On the other hand, *GA3ox2* (*GA4H*) is predominantly expressed in germinating seeds and young seedlings (Yamaguchi *et al.*, 1998b; Hedden & Phillips, 2000).

The gene encoding one of the dioxygenases that appears last in the GA biosynthetic pathway, *GA2ox1*, was first cloned from runner bean (*Phaseolus coccineus*) by Thomas, Phillips & Hedden (1999). *GA2ox1* has also now been cloned from several other species, including *A. thaliana*, together with the other GA2-oxidase genes within the gene family (Lester *et al.*, 1999; Martin, Proebsting & Hedden, 1999; Sakamoto *et al.*, 2001; Sakai *et al.*, 2003; Schomburg *et al.*, 2003). Expression studies in *Arabidopsis thaliana* have shown that transcript levels of *AtGA2ox1* and *AtGA2ox2* were highest in flowers, siliques and in upper stems, but some expression could also be observed in lower stems and leaves, and in the case of *AtGA2ox2* also in roots (Thomas, Phillips & Hedden, 1999). In rice, Sakamoto *et al*. (2001) observed a ring-shaped expression pattern of *OsGA2ox1*

around the vegetative shoot apex. This expression decreased drastically when the plants changed from a vegetative to a reproductive growth phase.

Feedback and feed-forward regulation of GA biosynthesis

Plants with mutations in genes encoding GA biosynthesis enzymes that appear early in the pathway have been shown to contain reduced levels of the bioactive GAs. They can also exhibit elevated transcript levels of the *GA20ox* and *GA3ox* genes (Thomas, Phillips & Hedden, 1999; Hedden & Phillips, 2000). These elevated transcription levels of the *GA20ox* and *GA3ox* genes are in most cases reduced by the application of bioactive GAs (Olszewski, Sun & Gubler, 2002). This was noted for instance in the *ga1-2* mutant by Thomas, Phillips & Hedden (1999). Thomas, Phillips & Hedden (1999) showed that the elevated expression levels of *AtGA20ox2* and *AtGA3ox1* observed in the *ga1-2* mutant were markedly reduced by the application of GA_3. This negative feedback regulation achieved by the application of bioactive GA mainly inhibits the enzymes that appear in the last steps of GA biosynthesis. In other words, it has not been shown to have any effect on the mRNA levels of the genes encoding CPS, KS and KO (Helliwell *et al.*, 1998; Hedden & Phillips, 2000). Thomas, Phillips & Hedden (1999) also observed that the expression of *AtGA2ox1* and *AtGA2ox2* in the *ga1-2* mutant was upregulated by the application of GA_3. Similar type of feed-forward regulation of the GA 2-oxidase genes has also been observed in rice and pea (Elliot *et al.*, 2001; Sakai *et al.*, 2003). Generally, when a plant contains low levels of bioactive gibberellins there is an upregulation of the GA biosynthesis genes *GA20ox* and *GA3ox*, whereas high levels of bioactive GAs cause a downregulation of these genes (feedback regulation) and an upregulation of the deactivating GA 2-oxidase genes (feed-forward regulation). These feedback and feed-forward regulations help the plant to keep the concentrations of bioactive GAs within a limited range. These regulations of the GA-dioxygenase genes are illustrated in Fig. 4.

Some exceptions to the feedback and feed-forward regulations of the GA dioxygenase genes, such as the seed germination in Arabidopsis where only *AtGA3ox*1 but not *AtGA3ox2* was regulated by a negative feedback response, have previously been observed (Yamaguchi *et al.*, 1998b). It has been suggested that *AtGA3ox2* is involved in maintaining active GA levels to promote seed germination (a process that needs high levels of active GAs during a short period), which explains why an inhibition of its expression through feedback regulation would not seem necessary or appropriate (Hedden & Phillips, 2000). Another example that describes the exception to the rule of GA dioxygenase feed-forward regulation was presented by Ogawa *et al*. (2003). They did not find any evidence for a feed-forward upregulation of the *AtGA2ox* genes when analyzing GA_4 imbibed seeds. This is in agreement with the idea that high levels of bioactive GAs are needed during seed germination and that the synthesis, rather than the deactivation, is the most important process that regulates bioactive GA levels during the germination of *A. thaliana* seeds (Ogawa *et al.*, 2003).

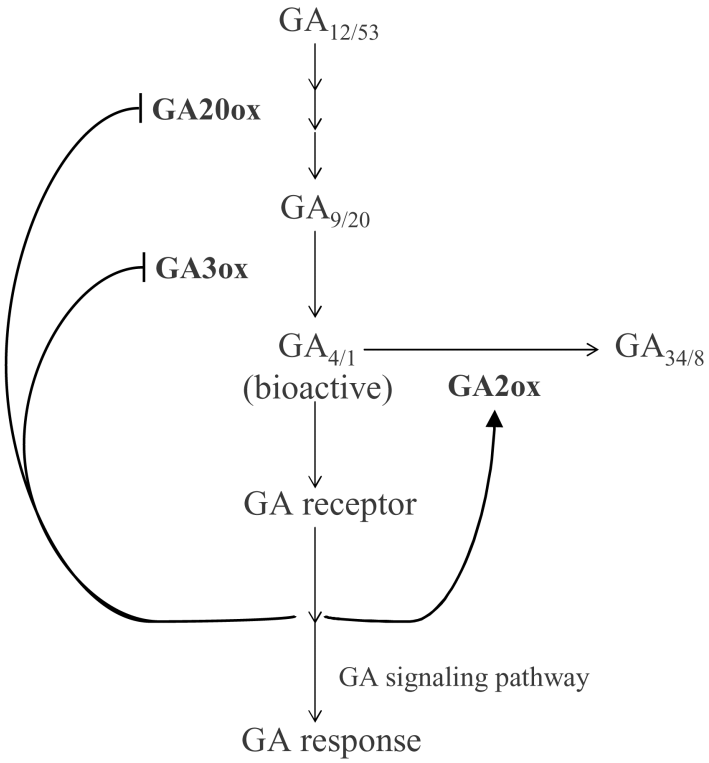

Fig. 4. Model showing the feedback (indicated by T-bar) and feed-forward (indicated by closed triangular arrowhead) regulation of GA biosynthesis.

Aims of this study

Identification of genes that are involved in regulating the time of flowering is of great importance in agri- and horticulture. Genes involved in regulating flowering time can be used for crop improvement by, for instance, engineering plants to flower earlier. In an attempt to identify and isolate genes that are involved in regulating flowering time I have employed a gene knockout approach using promoter probe T-DNA tagging and *in vivo* gene fusion in the model plant *Arabidopsis thaliana*. The thesis is mainly based on the identification and characterisation of a T-DNA tagged gene and a mutant of *A. thaliana* exhibiting delayed flowering.

Specific aims were as follows:

- Screening of T-DNA tagged lines for identification of mutants showing delayed flowering.
- Physiological and molecular characterisation of the T-DNA tagged mutant.
- Identification, cloning and characterisation of the T-DNA tagged gene.
- *In silico* analyses for predicting function of the T-DNA tagged gene.

Results and discussion

Functional prediction of a T-DNA tagged gene of *Arabidopsis thaliana* by *in silico* analysis (I)

By screening a library of T-DNA tagged *Arabidopsis thaliana* lines, I identified a mutant line (197) exhibiting a late-flowering phenotype and a tissue specific expression of the promoterless *gus* reporter gene. The T-DNA tagged plants had been transformed with the promoter-trap vector pMHA2, which contained a promoterless *gus* gene placed adjacent to the right end of the T-DNA (Mandal *et al.*, 1995). In order to identify the T-DNA tagged gene an inverse-PCR (I-PCR) cloning method was performed. The results showed that the T-DNA had been inserted in chromosome 2, about 4 kb upstream of *At2g36400* and 2 kb downstream of *At2g36410*. The function of both of these genes was unknown. When searching for similar sequences in the *A. thaliana* GenBank, it was observed that the protein sequence encoded by *At2g36400* had a high sequence similarity to a growth-regulating factor (Os-GRF1) of *Oryza sativa*. According to van der Knapp, Kim & Kende (2000), Os-GRF1 is encoded by a gibberellin-induced gene and has a regulatory role in stem growth.

The focus of our main investigation turned to *At2g36400*, rather than *At2g36410*, for several reasons. First of all, the mutant phenotype observed in our T-DNA tagged line was a late-flowering phenotype that could be largely reversed by the

application of exogenous gibberellin. The protein sequence encoded by *At2g36400* showed a high sequence similarity to a growth-regulating factor (Os-GRF1) that is encoded by a gibberellin induced gene. A second reason for focusing on *At2g36400*, which was located downstream of the T-DNA insert, was that the T-DNA tagged plants showed GUS activity. Activation of the *gus* reporter gene indicated that the promoter of the target gene might be located upstream of the right junction, whereas the coding sequences might be downstream of the integrated T-DNA. Another reason for selecting *At2g36400* for further studies was that reverse transcriptase PCR (RT-PCR) and Northern blot analyses showed that the expression of *At2g36400* was clearly reduced in the T-DNA tagged line compared to wild-type plants (data not shown). The results also indicated that *At2g36400* was mainly expressed in the shoot apex of wild-type plants (data not shown). The expression pattern observed for *At2g36400* was in agreement with that observed for the promoterless *gus* reporter gene, since GUS analysis revealed a predominant GUS activity in the shoot apex of the tagged plants. The result from the expression study of *At2g36400* in the T-DNA tagged line and in the wild-type plants was also a great factor that contributed to make our decision for selection of *At2g36400* for further studies.

Because of the long distance between the T-DNA insert and *At2g36400* it was very difficult to explain a connection among the T-DNA insert, the mutant phenotype and the GUS activity. Since several T-DNA lines with mutant phenotype have shown to contain the T-DNA insert far up in the 5´-UTR (Klucher *et al.*, 1996; Azpiroz-Leehan & Feldmann, 1997) and the fact that the transcription level of *At2g36400* was reduced in line 197, we believed that the T-DNA insert in the upstream region of *At2g36400* caused the mutant phenotype. However, the GUS activity was harder to explain. A possible explanation to the observed GUS activity would be that *At2g36400* contained additional upstream exons that were not yet annotated in the *A. thaliana* GenBank. The gene prediction programme GenScan (Burge & Karlin, 1997) revealed the possible existence of two exons upstream of *At2g36400*. However, when analysing the presence of these exons by RT-PCR, it was revealed that additional upstream exons did not exist. By confirming these results, the transcriptional start of *At2g36400* was analysed by a rapid amplification of the 5´ cDNA end, 5´-RACE, (Schaefer, 1995). The results revealed that the transcription start of *At2g36400* was in the vicinity of what was annotated, confirming that *At2g36400* did not contain any upstream exons (data not shown).

In order to investigate if the GUS expression and the mutant phenotype seen in plants of line 197 were directly associated with the T-DNA insertion upstream of *At2g36400*, we backcrossed the mutant line with wild-type *A. thaliana*. However, segregation analysis of the F_2 hybrid offsprings revealed that kanamycin resistant plants of some lines did not exhibit GUS activity, whereas other lines did. These results showed that the GUS activity and the kanamycin resistance segregated, indicating that plants of line 197 contained more than one T-DNA insertion. Since several T-DNA insertions were integrated into the plants of line 197 it was impossible to conclude that the late-flowering phenotype and the GUS activity

observed in the mutant plants was a result of the T-DNA insert in *At2g36400* or whether it was an affect of another T-DNA insertion.

To conclude whether the mutant phenotype was linked to the gene *At2g36400*, I analysed several T-DNA insertion lines from the Salk Institute (Alonso *et al.*, 2003). However, the results could not confirm that the late-flowering phenotype observed in the plants of line 197 was due to the T-DNA insertion upstream of *At2g36400*. Because of this, we studied another line (197/4) with the same ancestors as line 197. This line also exhibited a late-flowering phenotype together with a tissue-specific GUS expression. Based on PCR results it was confirmed that the plants of line 197/4 did not contain a T-DNA insert upstream of *At2g36400*. Our further investigations were focused on line 197/4. The results derived from characterisation of this mutant are described in paper II.

Although it was later revealed that the insertion in *At2g36400* did not cause the late-flowering phenotype and the GUS activity seen in plants of line 197, we reported the prediction of the three-dimensional protein structure of At2g36400 in paper I. The results from the structure prediction showed that At2g36400 may have a structure similar to a ligand-binding domain of the human retinoic acid receptor gamma-2 protein (RRG2). RRG2 belongs to the superfamily of nuclear hormone receptors and more specifically functions as a receptor for retinoic acid (Krust *et al.*, 1989). Nuclear hormone receptors are ligand-activated transcription regulators that are involved in diverse physiological functions such as the control of embryonic development, cell division and differentiation (Escriva, Bertrand & Laudet, 2004). Further studies of the gene *At2g36400*, and other members of its gene family (AtGRF), have been performed by Kim and co-workers (2003; 2004). When analysing single T-DNA insertion mutants for three of the *AtGRF* genes, including *At2g36400*, Kim, Choi & Kende (2003) found that none of these lines exhibited a visible mutant phenotype. However, triple insertional null mutants had smaller leaves and cotyledons compared to wild-type plants (Kim, Choi & Kende, 2003). It was later also concluded that two members of the GRF family act as transcriptional activators and that they interact with the coactivator GIF1, GRF-interacting factor (Kim & Kende, 2004). According to Kim & Kende (2004), GIF1 is a functional homolog of the human synovial sarcoma translocation (SYT) protein. A recent study showed that SYT functions as a transcriptional coactivator of nuclear hormone receptors (Iwasaki, Koibuchi & Chin, 2005).

Characterisation of a T-DNA-tagged gene of *Arabidopsis thaliana* that regulates gibberellin metabolism and flowering time (II)

As mentioned previously, plants of line 197/4 exhibited a significant delay in flowering time compared to control plants. When analysing the mutant plants by histochemical assay for GUS activity it was shown that the activity of the promoterless *gus* reporter gene was expressed predominantly in rapid growing tissues such as root tips, stem nodes and shoot apex. To confirm that the plants of line 197/4 contained only one T-DNA insertion, a Southern blot analysis was performed. Although the results indicated that two T-DNA copies had been

integrated into the plant genome, it was verified that these copies were inserted in the same position. The integration of several T-DNAs at a common site is a fairly frequent event that has been observed by many researchers (Krizkova & Hrouda, 1998; De Buck *et al.*, 1999; Kumar & Fladung, 2000; Meza *et al.*, 2002). It has been proposed that the phenomenon of multiple T-DNA integration can either be due to the existence of transient "integration hotspots", or to the fact that the T-DNAs travel together to the insertion point (Tinland, 1996). After the conclusion that plants of line 197/4 harbour only one insertion containing two T-DNA copies, the plant DNA sequence flanking the T-DNA was identified by an I-PCR cloning method. The results revealed that the T-DNA had been inserted in the 3´-UTR of *At4g20010*, 15 bp downstream of the translational termination codon. To analyse whether this T-DNA integration actually caused a reduction of the transcript steady-state level of *At4g20010*, a RT-PCR analysis was performed. The results showed that the RNA level of *At4g20010* was severely reduced in the mutant plants compared to wild-type plants. When analysing the steady-state transcript level of *At4g20010* in different wild-type tissues, it was shown that *At4g20010* was highly expressed in the shoot apex and in flowers. The expression of *At4g20010* was also fairly high in stems, whereas in roots and leaves its expression was the lowest. The expression pattern of the T-DNA tagged gene *At4g20010* in wild-type plants is similar to that of the *gus* reporter gene observed in the plants of line 197/4. Both the promoterless *gus* reporter gene and *At4g20010* were highly expressed in the shoot apex. *At4g20010* was also highly expressed in flowers, this was however not compatible with the results obtained in GUS analysis. The GUS histochemical assay revealed no *gus* activity in the flowers of the plants of line 197/4, although its activity was observed in the abscission zone of developing siliques. Activity of the *gus* gene was also observed in parts of the stems, leaves and roots, *i.e.* in stem nodes, leaf veins and in root tips. These results are in agreement with results obtained from the expression study of *At4g20010*. The expression of *At4g20010* was observed in all of these tissues, but its level of expression was lower than that observed in the shoot apex. Although the promoterless *gus* reporter gene was inserted in the 3´-UTR of *At4g20010* the results indicate that the expression pattern of the *gus* gene reflects the expression pattern of *At4g20010*.

When considering the delayed flowering time in plants of line 197/4 and the fact that gibberellin is a plant hormone that regulates growth and controls developmental processes such as flowering (Davies, 1995; Ross, Murfet & Reid, 1997), we hypothesised that the delayed flowering observed in mutant plants was due to a disturbance in the GA metabolism. To verify this hypothesis, the plants of line 197/4 were treated with exogenous gibberellin and the endogenous levels of different GAs were measured. The results indicated that the late flowering phenotype was largely repressed by the application of GA_3. Similar results have been obtained in other GA-deficient mutants (Magome *et al.*, 2004; Alcazar *et al.*, 2005). The measurement of endogenous GAs showed that the amount of the bioactive gibberellin, GA_4, was reduced in the mutant plants compared to wild-type plants. GA quantification also revealed that GA_9 and GA_{20}, the immediate precursors of GA_4 and GA_1, were increased, possibly because of a feedback regulation. Low levels of bioactive GAs may trigger an upregulation of the GA

biosynthetic genes, which consequently may result in elevated levels of the immediate precursors (Hedden & Phillips, 2000). The pattern of GA levels observed in plants of line 197/4 is similar to that observed in other GA-deficient plants where the decreased levels of bioactive gibberellins are caused by either an inhibition of *GA3ox* or by an overexpression of *GA2ox* (Talon, Koornneef & Zeevaart, 1990; Sakamoto *et al.*, 2001). In order to verify whether the plants of line 197/4 contained elevated or repressed RNA levels of these genes, I analysed the expression of different GA-oxidase genes. The results of this analysis are presented in paper III and described below.

As described previously, plants of line 197/4 contain a T-DNA insertion in the 3´-UTR of *At4g20010*. *At4g20010* encodes a protein that belongs to a family of several uncharacterised plant proteins (mainly from *A. thaliana*) called DUF371 (Marchler-Bauer *et al.*, 2003). In addition, the protein sequence of At4g20010 is similar to a RNA-binding protein (RB38) from *Chlamydomonas reinhardtii* that associates with the 5´-UTR of the chloroplast *psbA* mRNA (Barnes *et al.*, 2004). RB38 contains four repeats that, according to Barnes *et al*. (2004), might be involved in RNA-binding. Some of these repeats were also identified in At4g20010 and in two of its homologues, At1g31010 and At5g44785.

To verify whether or not the late-flowering phenotype observed in the plants of line 197/4 was actually caused by the mutation in the target gene *At4g20010*, I analysed two SALK T-DNA mutants containing insertions in *At4g20010* (SALK_145209) and *At1g31010* (SALK_018261). The results showed that the flowering time was significantly delayed in the SALK lines compared to wild-type plants when grown under short-day conditions. Unlike plants of line 197/4, the difference was not that prominent when the plants were grown in long-days. This phenomenon might be explained by the difference in ecotype (197/4 are generated from the ecotype C24, whereas the SALK lines are derived from Columbia) or by the difference in T-DNA localisation. Further characterisation of the SALK T-DNA lines and the tagged genes are described in paper III.

At4g20010 and its homologue *At1g31010* encode two putative nucleic acid-binding proteins involved in regulating flowering time in *Arabidopsis thaliana* (III)

In our third paper, we report the results of our further studies on the mutant 197/4 and the SALK T-DNA insertion lines 145209 and 018261. SALK T-DNA lines that were homozygous for their respective insertion were identified. RT-PCR and Southern blot analysis were performed to confirm the knockout of the T-DNA tagged genes and the existence of one T-DNA insertion, respectively. As described in paper II, the SALK T-DNA lines flowered later than the wild-type plants when grown under short-day conditions. This significant delay in flowering time was shown to be completely reversed by the application of GA_4. This result is in agreement with that observed for the plants of line 197/4. This because, the late-flowering phenotype observed in the plants of line 197/4 was also largely repressed by the application of exogenous gibberellin. In addition to the late flowering

phenotype, the SALK T-DNA lines also exhibit other mutant phenotypes. The plants of SALK_145209 and especially SALK_018261 showed a wavy leaf phenotype. The plants of line SALK_018261 were also relatively smaller than the wild-type plants, where some plants were very small.

To determine if the reduced amount of the bioactive GA_4 in the plants of line 197/4 was due to a downregulation of the *GA3ox* or *GA20ox* genes, I performed RT-PCR. The results showed that there was a significant increase in the steady-state transcript levels of *GA3ox1* and *GA20ox1* in plants of line 197/4 compared to wild-type plants. This indicates that the reduced amount of GA_4 is not due to a transcriptional repression of *GA3ox1* or *GA20ox1*. Instead, the increased expression of *GA3ox1* and *GA20ox1* is probably due to a positive feedback regulation, where the low level of bioactive GA triggers the upregulation of these genes. Increased expression levels of *GA3ox1* and *GA20ox1* have been reported previously in several GA-deficient and GA-insensitive mutants (Xu *et al.*, 1995; Cowling *et al.*, 1998; Thomas, Phillips & Hedden, 1999). In the SALK line 018261 there was also an observed upregulation of *GA3ox1* and *GA20ox1*.

Since it was established that the lower amount of bioactive GA_4 in the plants of line 197/4 was not due to a transcriptional downregulation of *GA3ox1* or *GA20ox1*, we proposed that the decreased level of GA_4 could be a result of an increased expression of *GA2ox*. However, when analysing the expression of *GA2ox2* there was no significant difference in the steady-state transcript level between the mutant and the wild-type plants. In plants of line 197/4 there was a slight downregulation of *GA2ox2*, whereas in the SALK line 018261 there was a small upregulation. The analysis of the transcriptional steady-state level of *GA2ox1* revealed that there was a significant increase of *GA2ox1* in plants of line 197/4 compared to wild-type plants. However, these results were obtained when analysing RNA isolated from flowers, and a similar expression result was not obtained from the SALK line 018261.

To further investigate the functions of *At4g20010* and *At1g31010,* the three-dimensional structures of the proteins encoded by these genes were predicted by a fold-recognition method. The template fold identified for the N-terminal sequence of both At4g20010 and At1g31010 was a single stranded DNA-binding protein. The single stranded-binding proteins belong to the superfamily of nucleic acid-binding proteins and this group also contains the RNA-binding proteins (all contain an OB-fold). Considering the results obtained from the structure prediction and the fact that both At4g20010 and At1g31010 contain putative RNA-binding motifs in their C-terminal sequences, we believe that the proteins encoded by the *At4g20010* and *At1g31010* genes are putative RNA-binding proteins. Several genes within the flowering pathway have been shown to encode RNA-binding proteins, including *FCA*, *FLK* and *FPA*, all of which promote flowering by repressing *FLC* (Macknight *et al.*, 1997; Schomburg *et al.*, 2001; Lim *et al.*, 2004).

If *At4g20010* and *At1g31010* encode RNA-binding proteins, they may regulate GA metabolism at the post-transcriptional level. Based on our results obtained from the GA measurement and the expression studies of the GA-oxidase genes we

postulate that At4g20010 function as a positive regulator of GA3ox at the post-transcriptional level or act as a negative regulator of GA2ox.

Conclusions

By using T-DNA tagging and *in vivo* gene fusion I have identified two genes, *At4g20010* and its homologue *At1g31010*, involved in regulating flowering time in *A. thaliana*. Expression studies and GUS histochemical analysis revealed that *At4g20010* is mainly expressed in rapid growing tissues such as root tips, shoot apex, flowers and stem nodes. Plants with reduced steady-state transcript levels of either *At4g20010* or *At1g31010* exhibit a late-flowering phenotype that can be largely repressed by the application of gibberellin. In addition, plants with an insertional mutation in *At4g20010* contain reduced amount of the bioactive gibberellin GA_4 compared to wild-type plants. This suggests that the late-flowering phenotype observed in the mutant plants is a result of a decreased level of GA_4. However, this reduction is not due to a transcriptional repression of the GA-biosynthetic genes *GA3ox1* or *GA20ox1*, since their expressions were increased in the mutant plants.

By using bioinformatic tools, it was revealed that the protein sequences encoded by *At4g20010* and *At1g31010* contain RNA-binding motifs in their C-terminal sequences. A protein structure prediction by a fold recognition method indicated that the N-terminal sequences of both At4g20010 and At1g31010 have a three-dimensional structure resembling a nucleic acid-binding protein. Based on these results obtained *in silico* in combination with the results from the experimental studies, we postulate that *At4g20010* and *At1g31010* encode two RNA-binding proteins that are involved in regulating flowering time in *A. thaliana* by affecting the metabolism of GA. This can be possible either by a positive regulation of GA3ox at the post-transcriptional level or by a negative regulation of GA2ox.

A three-dimensional protein structure prediction by fold recognition revealed that the *At2g36400* encoded protein might have a structure similar to a ligand-binding domain of the nuclear hormone receptor RRG2, and it acts as a transcriptional regulator.

Future perspectives

In my thesis, I present the identification and characterisation of two genes, *At4g20010* and *At1g31010*, which are believed to encode RNA-binding proteins involved in regulating GA-metabolism. However, experimental studies will be needed in order to conclude that these proteins actually bind RNA. An example of a method that could be applied to address this is the gel mobility shift assay. This method would tell us if the proteins bind a specific RNA. However, another method like the yeast three-hybrid system (Sengupta, Wickens & Fields, 1999)

could be performed in order to identify which RNA, if any, the proteins bind. Results from this type of experiments may show whether or not the proteins of At4g20010 and At1g31010 interact with RNA encoding the GA-biosynthesis enzymes GA3ox or GA2ox, or if the possible regulation of these proteins is indirect. Part of this work has been initiated; I have expressed, isolated and purified the protein encoded by *At4g20010*.

To further confirm that *At4g20010* and *At1g31010* are involved in regulating the flowering time in *A. thaliana*, the genes could be transferred back in *sense* orientation to the late-flowering mutants in order to analyse whether or not flowering time is restored. The genes could also be overexpressed in Arabidopsis or some economically or ornamentally important plants to analyse if an increased transcription of *At4g20010* or *At1g31010* enhances the flowering time. Plants containing null mutations in both *At4g20010* and *At1g31010* could also be analysed to see if a more severe mutant phenotype arises. Plants containing the double mutations have been produced, but homozygous plants for both of these mutations, at the time of writing this thesis, have not yet been analysed. To further investigate the connection between the late-flowering phenotype and the lowered amount of bioactive GA_4, the endogenous GA levels could be measured in all of the above described transgenic plants.

To analyse whether the lowered amount of the bioactive GA_4 in the T-DNA tagged mutant is a result of a decreased activity of GA3ox or an increased activity of GA2ox, a GA-supplementation experiment could be performed. The mutant phenotype could be studied after the application of the immediate precursor of GA_4, namely GA_9, in the mutant plants. If the late-flowering phenotype would not be restored after the treatment, it could be concluded that the reduction of GA_4 in the mutant plants is probably due to a decreased activity of GA3ox. This would strengthen the hypothesis that At4g20010 acts as a positive regulator of GA3ox and in that way regulates GA biosynthesis. Another way to analyse whether it is the reduced activity of GA3ox or the increased activity of GA2ox that causes the reduced amount of GA_4, would be to compare the protein levels of GA3ox and GA2ox between wild-type and mutant plants.

In a future perspective it would also be of interest to analyse whether the activity of other flowering-time genes such as *FLC, FT*, *LFY* and *SOC1* are affected in the mutant plants. I believe that the best way to obtain an overall picture of which genes are affected in the mutant plants would be to perform global gene expression analysis by means of microarrays.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., WoodageT, Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. & Venter, J.C. 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185-2195.

Affymetrix. Genechip arrays. Arabidopsis ATH1 genome array. http://www.affymetrix.com/products/arrays/specific/arab.affx (accessed 5-Dec-2005).

Alcazar, R., Garcia-Martinez, J.L., Cuevas, J.C., Tiburcio, A.F. & Altabella, T. 2005. Overexpression of ADC2 in Arabidopsis induces dwarfism and late-flowering through GA deficiency. *Plant Journal* 43, 425-436.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C. & Ecker, J.R. 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653-657.

Amasino, R.M. 2005. Vernalization and flowering time. *Current Opinion in Biotechnology* 16, 154-158.

Araki, T. 2001. Transition from vegetative to reproductive phase. *Current Opinion in Plant Biology* 4, 63-68.

Azpiroz-Leehan, R. & Feldmann, K.A. 1997. T-DNA insertion mutagenesis in Arabidopsis: going back and forth. *Trends in Genetics* 13, 152-156.

Balzergue, S., Dubreucq, B., Chauvin, S., Le-Clainche, I., Le Boulaire, F., de Rose, R., Samson, F., Biaudet, V., Lecharny, A., Cruaud, C., Weissenbach, J., Caboche, M. & Lepiniec, L. 2001. Improved PCR-walking for large-scale isolation of plant T-DNA borders. *Biotechniques* 30, 496-498, 502, 504.

Barnes, D., Cohen, A., Bruick, R.K., Kantardjieff, K., Fowler, S., Efuet, E. & Mayfield, S.P. 2004. Identification and characterization of a novel RNA binding protein that associates with the 5'-untranslated region of the chloroplast *psbA* mRNA. *Biochemistry* 43, 8541-8550.

Bastow, R. & Dean, C. 2003. Plant sciences. Deciding when to flower. *Science* 302, 1695-1696.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Research* 28, 235-242.

Blazquez, M., Koornneef, M. & Putterill, J. 2001. Flowering on time: genes that regulate the floral transition. Workshop on the molecular basis of flowering time control. *EMBO Reports* 2, 1078-1082.

Bowie, J.U., Luthy, R. & Eisenberg, D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253, 164-170.

Burge, C. & Karlin, S. 1997. Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* 268, 78-94.

Chiang, H.H., Hwang, I. & Goodman, H.M. 1995. Isolation of the Arabidopsis *GA4* locus. *Plant Cell* 7, 195-201.

Chothia, C. 1992. Proteins. One thousand families for the molecular biologist. *Nature* 357, 543-544.

Cleland, R. 1999. *Biochemistry and molecular biology of plant hormones*. Elsevier Science, Amsterdam.

Coupland, G. 1995. Genetic and environmental control of flowering time in Arabidopsis. *Trends in Genetics* 11, 393-397.

Cowling, R.J., Kamiya, Y., Seto, H. & Harberd, N.P. 1998. Gibberellin dose-response regulation of GA4 gene transcript levels in Arabidopsis. *Plant Physiology* 117, 1195-1203.

Davies, P.J. 1995. *Plant hormones: physiology, biochemistry and molecular biology*. Kluwer Academic Publishers, Dordrecht.

De Buck, S., De Wilde, C., Van Montagu, M. & Depicker, A. 2000. T-DNA vector backbone sequences are frequently integrated into the genome of transgenic plants obtained by Agrobacterium-mediated transformation. *Molecular Breeding* 6, 459-468.

De Buck, S., Jacobs, A., Van Montagu, M. & Depicker, A. 1999. The DNA sequences of T-DNA junctions suggest that complex T-DNA loci are formed by a recombination process resembling T-DNA integration. *Plant Journal* 20, 295-304.

Edwards, Y.J. & Cottage, A. 2003. Bioinformatics methods to predict protein structure and function. A practical approach. *Molecular Biotechnology* 23, 139-166.

Elliot, R., Ross, J., JJ, S., DR, L. & JB, R. 2001. Feed-forward regulation of gibberellin deactivation in pea. *Journal of Plant Growth Regulation* 20, 87-94.

Escriva, H., Bertrand, S. & Laudet, V. 2004. The evolution of the nuclear receptor superfamily. *Essays in Biochemistry* 40, 11-26.

Feldmann, K.A. 1991. T-DNA insertion mutagenesis in Arabidopsis: mutational spectrum. *Plant Journal* 1, 71-82.

Forsbach, A., Schubert, D., Lechtenberg, B., Gils, M. & Schmidt, R. 2003. A comprehensive characterization of single-copy T-DNA insertions in the *Arabidopsis thaliana* genome. *Plant Molecular Biology* 52, 161-176.

Gelvin, S.B. 1998. The introduction and expression of transgenes in plants. *Current Opinion in Biotechnology* 9, 227-232.

Godzik, A. 2003. Structural Bioinformatics. John Wiley and Sons Ltd, New York.

Godzik, A. & Skolnick, J. 1992. Sequence-structure matching in globular proteins: application to supersecondary and tertiary structure determination. *Proceedings of the National Academy of Sciences of the USA* 89, 12098-12102.

Hedden, P. 1999. *Biochemistry and molecular biology of plant hormones*. Elsevier Science, Amsterdam.

Hedden, P. & Phillips, A.L. 2000. Gibberellin metabolism: new insights revealed by the genes. *Trends in Plant Science* 5, 523-530.

Hedden, P., Phillips, A.L., Rojas, M.C., Carrera, E. & Tudzynski, B. 2001. Gibberellin Biosynthesis in Plants and Fungi: A Case of Convergent Evolution? *Journal of Plant Growth Regulation* 20, 319-331.

Helliwell, C.A., Chandler, P.M., Poole, A., Dennis, E.S. & Peacock, W.J. 2001. The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. *Proceedings of the National Academy of Sciences of the USA* 98, 2065-2070.

Helliwell, C.A., Sheldon, C.C., Olive, M.R., Walker, A.R., Zeevaart, J.A., Peacock, W.J. & Dennis, E.S. 1998. Cloning of the Arabidopsis ent-kaurene oxidase gene *GA3*. *Proceedings of the National Academy of Sciences of the USA* 95, 9019-9024.

Iwasaki, T., Koibuchi, N. & Chin, W.W. 2005. Synovial sarcoma translocation (SYT) encodes a nuclear receptor coactivator. *Endocrinology* 146, 3892-3899.

Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M. & Last, R.L. 2002. Arabidopsis map-based cloning in the post-genome era. *Plant Physiology* 129, 440-450.

Jones, D. & Hadley, C. 2000. *Bioinformatics: sequence, structure and databanks: a practical approach*. Oxford University Press, Oxford.

Jones, D.T., Taylor, W.R. & Thornton, J.M. 1992. A new approach to protein fold recognition. *Nature* 358, 86-89.

Kamiya, Y. & Garcia-Martinez, J.L. 1999. Regulation of gibberellin biosynthesis by light. *Current Opinion in Plant Biology* 2, 398-403.

Kertbundit, S., De Greve, H., Deboeck, F., Van Montagu, M. & Hernalsteens, J.P. 1991. In vivo random beta-glucuronidase gene fusions in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the USA* 88, 5212-5216.

Kim, J.H., Choi, D. & Kende, H. 2003. The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis. *Plant Journal* 36, 94-104.

Kim, J.H. & Kende, H. 2004. A transcriptional coactivator, AtGIF1, is involved in regulating leaf growth and morphology in Arabidopsis. *Proceedings of the National Academy of Sciences of the USA* 101, 13374-13379.

Klucher, K.M., Chow, H., Reiser, L. & Fischer, R.L. 1996. The *AINTEGUMENTA* gene of Arabidopsis required for ovule and female gametophyte development is related to the floral homeotic gene *APETALA2*. *Plant Cell* 8, 137-153.

Komeda, Y. 2004. Genetic regulation of time to flower in *Arabidopsis thaliana*. *Annual Review of Plant Biology* 55, 521-535.

Koncz, C., Martini, N., Mayerhofer, R., Koncz-Kalman, Z., Korber, H., Redei, G.P. & Schell, J. 1989. High-frequency T-DNA-mediated gene tagging in plants. *Proceedings of the National Academy of Sciences of the USA* 86, 8467-8471.

Kononov, M.E., Bassuner, B. & Gelvin, S.B. 1997. Integration of T-DNA binary vector 'backbone' sequences into the tobacco genome: evidence for multiple complex patterns of integration. *Plant Journal* 11, 945-957.

Koornneef, M. 1997. Plant development: timing when to flower. *Current Biology* 7, R651-652.

Koornneef, M., Hanhart, C.J. & van der Veen, J.H. 1991. A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Molecular and General Genetics* 229, 57-66.

Koornneef, M. & Van der Veen, J.H. 1980. Induction and analysis of gibberellin sensitive mutants in *Arabidopsis thaliana* (L.) Heynh. *Theoretical and Applied Genetics* 58, 257-263.

Krizkova, L. & Hrouda, M. 1998. Direct repeats of T-DNA integrated in tobacco chromosome: characterization of junction regions. *Plant Journal* 16, 673-680.

Krust, A., Kastner, P., Petkovich, M., Zelent, A. & Chambon, P. 1989. A third human retinoic acid receptor, hRAR-gamma. *Proceedings of the National Academy of Sciences of the USA* 86, 5310-5314.

Kumar, S. & Fladung, M. 2000. Transgene repeats in aspen: molecular characterisation suggests simultaneous integration of independent T-DNAs into receptive hotspots in the host genome. *Molecular and General Genetics* 264, 20-28.

Lange, T., Hedden, P. & Graebe, J.E. 1994. Expression cloning of a gibberellin 20-oxidase, a multifunctional enzyme involved in gibberellin biosynthesis. *Proceedings of the National Academy of Sciences of the USA* 91, 8552-8556.

Lester, D.R., Ross, J.J., Smith, J.J., Elliott, R.C. & Reid, J.B. 1999. Gibberellin 2-oxidation and the *SLN* gene of *Pisum sativum*. *Plant Journal* 19, 65-73.

Lichtenthaler, H.K., Schwender, J., Disch, A. & Rohmer, M. 1997. Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway. *FEBS Letters* 400, 271-274.

Lim, M.H., Kim, J., Kim, Y.S., Chung, K.S., Seo, Y.H., Lee, I., Kim, J., Hong, C.B., Kim, H.J. & Park, C.M. 2004. A new Arabidopsis gene, *FLK*, encodes an RNA binding protein with K homology motifs and regulates flowering time via *FLOWERING LOCUS C*. *Plant Cell* 16, 731-740.

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

Magome, H., Yamaguchi, S., Hanada, A., Kamiya, Y. & Oda, K. 2004. dwarf and delayed-flowering 1, a novel Arabidopsis mutant deficient in gibberellin biosynthesis because of overexpression of a putative AP2 transcription factor. *Plant Journal* 37, 720-729.

Mandal, A., Sandgren, M., Holmström, K.-O., Gallois, P. & Palva, E.T. 1995. Identification of *Arabidopsis thaliana* sequences responsive to low temperature and abscisic acid by T-DNA tagging and *in vivo* gene fusion. *Plant Molecular Biology Reporter* 13, 243-254.

Marchler-Bauer, A., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S., Hurwitz, D.I., Jackson, J.D., Jacobs, A.R., Lanczycki, C.J., Liebert, C.A., Liu, C., Madej, T., Marchler, G.H., Mazumder, R., Nikolskaya, A.N., Panchenko, A.R., Rao, B.S., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Vasudevan, S., Wang, Y., Yamashita, R.A., Yin, J.J. & Bryant, S.H. 2003. CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Research* 31, 383-387.

Marti-Renom, M., Fiser, A., Madhusudhan, M., John, B., Stuart, A., Eswar, N., Pieper, P., Shen, M.-Y. & Sali, A. 2003. *Current protocols in bioinformatics*. John Wiley & Sons, New York.

Martin, D.N., Proebsting, W.M. & Hedden, P. 1999. The *SLENDER* gene of pea encodes a gibberellin 2-oxidase. *Plant Physiology* 121, 775-781.

Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Bakkeren, G., Crameri, A., Angelis, K., Redei, G.P., Schell, J., Hohn, B. & Koncz, C. 1991. T-DNA integration: a mode of illegitimate recombination in plants. *Embo Journal* 10, 697-704.

McElver, J., Tzafrir, I., Aux, G., Rogers, R., Ashby, C., Smith, K., Thomas, C., Schetter, A., Zhou, Q., Cushman, M.A., Tossberg, J., Nickle, T., Levin, J.Z., Law, M., Meinke, D. & Patton, D. 2001. Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. *Genetics* 159, 1751-1763.

Meinke, D.W., Cherry, J.M., Dean, C., Rounsley, S.D. & Koornneef, M. 1998. *Arabidopsis thaliana*: a model plant for genome analysis. *Science* 282, 662, 679-682.

Meyerowitz, E.M. 1987. *Arabidopsis thaliana*. *Annual Review of Genetics* 21, 93-111.

Meza, T.J., Stangeland, B., Mercy, I.S., Skarn, M., Nymoen, D.A., Berg, A., Butenko, M.A., Hakelien, A.M., Haslekas, C., Meza-Zepeda, L.A. & Aalen, R.B. 2002. Analyses of single-copy Arabidopsis T-DNA-transformed lines show that the presence of vector backbone sequences, short inverted repeats and DNA methylation is not sufficient or necessary for the induction of transgene silencing. *Nucleic Acids Research* 30, 4556-4566.

Michaels, S.D. & Amasino, R.M. 1999. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11, 949-956.

Moffat, A.S. 2000. Plant research. Can genetically modified crops go 'greener'? *Science* 290, 253-254 P.

Mouradov, A., Cremer, F. & Coupland, G. 2002. Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell* 14 Suppl, S111-130.

Ogawa, M., Hanada, A., Yamauchi, Y., Kuwahara, A., Kamiya, Y. & Yamaguchi, S. 2003. Gibberellin biosynthesis and response during Arabidopsis seed germination. *Plant Cell* 15, 1591-1604.
Ohba, T., Yoshioka, Y., Machida, C. & Machida, Y. 1995. DNA rearrangement associated with the integration of T-DNA in tobacco: an example for multiple duplications of DNA around the integration target. *Plant Journal* 7, 157-164.
Olszewski, N., Sun, T.P. & Gubler, F. 2002. Gibberellin signaling: biosynthesis, catabolism, and response pathways. *Plant Cell* 14 Suppl, S61-80.
Orengo, C.A., Flores, T.P., Jones, D.T., Taylor, W.R. & Thornton, J.M. 1993. Recurring structural motifs in proteins with different functions. *Current Biology* 3, 131-139.
Ostergaard, L. & Yanofsky, M.F. 2004. Establishing gene function by mutagenesis in *Arabidopsis thaliana*. *Plant Journal* 39, 682-696.
Parcy, F. 2005. Flowering: a time for integration. *International Journal of Developmental Biology* 49, 585-593.
Pawlowski, K., Rychlewski, L., Zhang, B. & Godzik, A. 2001. Fold predictions for bacterial genomes. *Journal of Structural Biology* 134, 219-231.
Pereira, A. 2000. A transgenic perspective on plant functional genomics. *Transgenic Research* 9, 245-260; discussion 243.
Phillips, A.L., Ward, D.A., Uknes, S., Appleford, N.E., Lange, T., Huttly, A.K., Gaskin, P., Graebe, J.E. & Hedden, P. 1995. Isolation and expression of three gibberellin 20-oxidase cDNA clones from Arabidopsis. *Plant Physiology* 108, 1049-1057.
Ramanathan, V. & Veluthambi, K. 1995. Transfer of non-T-DNA portions of the *Agrobacterium tumefaciens* Ti plasmid pTiA6 from the left terminus of TL-DNA. *Plant Molecular Biology* 28, 1149-1154.
Riggs, C.D. & Chrispeels, M.J. 1987. Luciferase reporter gene cassettes for plant gene expression studies. *Nucleic Acids Research* 15, 8115.
Ross, J.J., Murfet, I.C. & Reid, J.B. 1997. Gibberellin mutants. *Physiologia Plantarum* 100, 550-560.
Ryu, C.H., You, J.H., Kang, H.G., Hur, J., Kim, Y.H., Han, M.J., An, K., Chung, B.C., Lee, C.H. & An, G. 2004. Generation of T-DNA tagging lines with a bidirectional gene trap vector and the establishment of an insertion-site database. *Plant Molecular Biology* 54, 489-502.
Sakai, M., Sakamoto, T., Saito, T., Matsuoka, M., Tanaka, H. & Kobayashi, M. 2003. Expression of novel rice gibberellin 2-oxidase gene is under homeostatic regulation by biologically active gibberellins. *Journal of Plant Research* 116, 161-164.
Sakamoto, T., Kobayashi, M., Itoh, H., Tagiri, A., Kayano, T., Tanaka, H., Iwahori, S. & Matsuoka, M. 2001. Expression of a gibberellin 2-oxidase gene around the shoot apex is related to phase transition in rice. *Plant Physiology* 125, 1508-1516.
Sakamoto, T., Miura, K., Itoh, H., Tatsumi, T., Ueguchi-Tanaka, M., Ishiyama, K., Kobayashi, M., Agrawal, G.K., Takeda, S., Abe, K., Miyao, A., Hirochika, H., Kitano, H., Ashikari, M. & Matsuoka, M. 2004. An overview of gibberellin metabolism enzyme genes and their related mutants in rice. *Plant Physiology* 134, 1642-1653.
Schaefer, B.C. 1995. Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. *Analytical Biochemistry* 227, 255-273.
Schomburg, F.M., Bizzell, C.M., Lee, D.J., Zeevaart, J.A. & Amasino, R.M. 2003. Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. *Plant Cell* 15, 151-163.
Schomburg, F.M., Patton, D.A., Meinke, D.W. & Amasino, R.M. 2001. FPA, a gene involved in floral induction in Arabidopsis, encodes a protein containing RNA-recognition motifs. *Plant Cell* 13, 1427-1436.
Sengupta, D.J., Wickens, M. & Fields, S. 1999. Identification of RNAs that bind to a specific protein using the yeast three-hybrid system. *RNA* 5, 596-601.
Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., Clarke, J.D., Cotton, D., Bullis, D., Snell, J., Miguel, T., Hutchison, D., Kimmerly, B., Mitzel, T., Katagiri, F., Glazebrook, J., Law, M. & Goff,

S.A. 2002. A high-throughput Arabidopsis reverse genetics system. *Plant Cell* 14, 2985-2994.
Simpson, G.G. & Dean, C. 2002. Arabidopsis, the Rosetta stone of flowering time? *Science* 296, 285-289.
Srinivasan, N., Guruprasad, K. & Blundell, T. 2002. *Proteins structure prediction*. Oxford University Press, New York.
Srivastava, L. 2002. *Plant growth and development: Hormones and environment*. Academic Press, San Diego.
Stangeland, B., Nestestog, R., Grini, P.E., Skrbo, N., Berg, A., Salehian, Z., Mandal, A. & Aalen, R.B. 2005. Molecular analysis of Arabidopsis endosperm and embryo promoter trap lines: reporter-gene expression can result from T-DNA insertions in antisense orientation, in introns and in intergenic regions, in addition to sense insertion at the 5' end of genes. *Journal of Experimental Botany* 56, 2495-2505.
Stewart, C.N., Jr. 2001. The utility of green fluorescent protein in transgenic plants. *Plant Cell Reports* 20, 376-382.
Sun, T., Goodman, H.M. & Ausubel, F.M. 1992. Cloning the Arabidopsis *GA1* Locus by Genomic Subtraction. *Plant Cell* 4, 119-128.
Sun, T.P. & Kamiya, Y. 1994. The Arabidopsis *GA1* locus encodes the cyclase ent-kaurene synthetase A of gibberellin biosynthesis. *Plant Cell* 6, 1509-1518.
Swiss Federal Institute of Technology Zurich. Genevestigator. Arabidopsis thaliana. https://www.genevestigator.ethz.ch (accessed 5-Dec-2005).
Talon, M., Koornneef, M. & Zeevaart, J.A. 1990. Endogenous gibberellins in *Arabidopsis thaliana* and possible steps blocked in the biosynthetic pathways of the semidwarf *ga4* and *ga5* mutants. *Proceedings of the National Academy of Sciences of the USA* 87, 7983-7987.
The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796-815.
The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012-2018.
Thomas, S.G., Phillips, A.L. & Hedden, P. 1999. Molecular cloning and functional expression of gibberellin 2- oxidases, multifunctional enzymes involved in gibberellin deactivation. *Proceedings of the National Academy of Sciences of the USA* 96, 4698-4703.
Thornton, J.M., Orengo, C.A., Todd, A.E. & Pearl, F.M. 1999. Protein folds, functions and evolution. *Journal of Molecular Biology* 293, 333-342.
Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C., Odden, A.R., Young, K., Taylor, N.E., Henikoff, J.G., Comai, L. & Henikoff, S. 2003. Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Research* 13, 524-530.
Tinland, B. 1996. The integration of T-DNA into plant genomes. *Trends in plant science* 1, 178-184.
Topping, J. & Lindsey, K. 1995. Insertional mutagenesis and promoter trapping in plants for the isolation of genes and the study of development. *Transgenic research* 4, 291-305.
Topping, J., Wei, W., Clarke, M., Muskett, P. & Lindsey, K. 1995. *Plant gene transfer and expression protocols*. Humana Press, Totowa, New Jersey.
University of Arizona. Department of Plant Sciences. Arabidopsis oligonucleotide microarrays. http://www.ag.arizona.edu/microarray (accessed 5-Dec-2005).
Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A. & Coupland, G. 2004. Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science* 303, 1003-1006.
van der Knaap, E., Kim, J.H. & Kende, H. 2000. A novel gibberellin-induced gene from rice and its potential regulatory role in stem growth. *Plant Physiology* 122, 695-704.
Williams, J., Phillips, A.L., Gaskin, P. & Hedden, P. 1998. Function and substrate specificity of the gibberellin 3beta-hydroxylase encoded by the Arabidopsis *GA4* gene. *Plant Physiology* 117, 559-563.

Wortman, J.R., Haas, B.J., Hannick, L.I., Smith, R.K., Jr., Maiti, R., Ronning, C.M., Chan, A.P., Yu, C., Ayele, M., Whitelaw, C.A., White, O.R. & Town, C.D. 2003. Annotation of the Arabidopsis genome. *Plant Physiology* 132, 461-468.

Xu, D., Xu, Y. & Uberbacher, E.C. 2000. Computational tools for protein modeling. *Current Protein and Peptide Science* 1, 1-21.

Xu, Y.L., Li, L., Wu, K., Peeters, A.J., Gage, D.A. & Zeevaart, J.A. 1995. The *GA5* locus of Arabidopsis thaliana encodes a multifunctional gibberellin 20-oxidase: molecular cloning and functional expression. *Proceedings of the National Academy of Sciences of the USA* 92, 6640-6644.

Yamaguchi, S. & Kamiya, Y. 2000. Gibberellin biosynthesis: its regulation by endogenous and environmental signals. *Plant Cell Physiology* 41, 251-257.

Yamaguchi, S., Saito, T., Abe, H., Yamane, H., Murofushi, N. & Kamiya, Y. 1996. Molecular cloning and characterization of a cDNA encoding the gibberellin biosynthetic enzyme ent-kaurene synthase B from pumpkin (*Cucurbita maxima* L.). *Plant Journal* 10, 203-213.

Yamaguchi, S., Smith, M.W., Brown, R.G., Kamiya, Y. & Sun, T. 1998b. Phytochrome regulation and differential expression of gibberellin 3beta-hydroxylase genes in germinating Arabidopsis seeds. *Plant Cell* 10, 2115-2126.

Yamaguchi, S., Sun, T., Kawaide, H. & Kamiya, Y. 1998a. The *GA2* locus of *Arabidopsis thaliana* encodes ent-kaurene synthase of gibberellin biosynthesis. *Plant Physiology* 116, 1271-1278.

Zupan, J., Muth, T.R., Draper, O. & Zambryski, P. 2000. The transfer of DNA from agrobacterium tumefaciens into plants: a feast of fundamental insights. *Plant Journal* 23, 11-28.

Acknowledgments

I would like to express my sincere gratitude to all people who has helped, supported and encouraged me during my PhD period. In particular, I would like to thank:

My supervisors, **Abul Mandal** and **Per Bergman**, for giving me the opportunity to work with this project and for introducing me to the intriguing field of molecular biology. Thank you for your encouragement, support, and guidance during these years.

Dan Lundh, for enjoyable collaboration and your friendly and positive attitude. Your enthusiastic way of explaining and answering all my questions regarding bioinformatics has truly been stimulating and helpful.

All people at the old Department of Natural Sciences in Skövde for creating such a warm and friendly atmosphere, my heartiest thanks goes to:

The Molbiol group (past and present members): **Heléne**, **Tjowe**, **Åsa**, **Linda**, **Anna-Karin**, **Henrik A**, **Henrik E**, **Ida**, **Helena**, **Cristofer**, **Patric** and especially **Mikael** for handling my TGS's and for guiding me in the lab.

The PhD students: **Elie**, **Erik**, **Magnus**, **Andreas**, **Diana**, **Peter**, **Sazzad, Sandra**, **Fredrik**, **Emma**, **Patrik**, **Simon**, **Niclas**, and **Jenny**. I have really enjoyed our parties, movie nights, lunches and discussions and I hope that I also in the future will be invited to the yearly Christmas party.
My special thanks goes to the two "SKs". **Sazzad**, you welcomed and introduced me to the lab on my first day, and ever since you have been one of my closest friends. You are the only one that truly has understood my situation and you have always taken the time to help me. I wish you all the best in the world, because nobody deserves it more. **Sandra**, you have been my training, shopping and party companion and your encouragement and friendship has meant so much to me. You have supported and helped me (even without me asking) and I will really miss our daily chats and discussions. In conclusion, I couldn't have asked for a better friend and colleague.

The administrative and technical staff for all your help concerning money, courses, labs, logistics and employment: **Amila**, **Pernilla**, **Siv**, **Ann-Louise**, **Pia**, **Mikael**, **Noel**, **Kajsa, Jonas, Bengt**, and especially **Mats** and **Jesper** for helping me every time my computer crashed.

Håkan, my personal tennis coach, for trying to teach me how to play tennis.

All the past and present **students** (both Swedish and foreign master students) that I have tried to teach. I never thought that it would be so fun and stimulating!

Dr Sibdas Ghosh (Dominican University of California), **Dr Thomas Moritz** (SLU, Umeå) and **Dr Reidunn Aalen** (Oslo University) for your collaboration and valuable scientific advice.

The people at the Department of Plant Biology and Forest Genetics, SLU, Uppsala that has helped me with administrative and scientific matters: **Dr Folke Sitbon**, **Professor Christer Jansson**, **Kristin-Sophie Mellsjö**, **Birgitta Eriksson** and **Ingrid Eriksson**.

I would also like to acknowledge the Nilsson-Ehle foundation for financial support.

Last but absolutely not least, I would like to thank my wonderful and supportive family: my love **Henrik**, my parents **Elisabeth** and **Conny**, my sisters **Linda** and **Karin**, **Peder**, **Urban**, **Annika**, **Åke**, **Karin**, **Erika**, **Rickard**, **Hanna**, **Joel**, **Anton**, **Anna**, **Sara**, **Kerstin**, **Stefan**, **Margaretha, Eje** and other family members and friends. Your love and endurance has really helped me and without you *nothing else matters*.