ACTA UNIVERSITATIS AGRICULTURAE SUECIAE

SILVESTRIA 173

A Functional Genomics Approach to Wood Development

Magnus Hertzberg

SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

A Functional Genomics Approach to Wood Development

Magnus Hertzberg.

Akademisk avhandling som för vinnande av filosofie doktorsexamen kommer att offentligt försvaras i hörsal Björken, SLU Umeå, fredagen den 9 feruari 2001, kl 10.00

Abstract

Wood is widely used for various purposes, e.g. for pulp and paper, construction material and heating. Thus it is surprising that the cellular and molecular regulation of wood formation remains poorly characterised. This study therefore addresses the process of wood formation in order to elucidate factors controlling this process at the molecular level. Two different approaches were taken, using novel technologies that have become available in recent years.

First, a new group of homeobox transcription factors were identified, denoted the PALE class due to a five amino acid insertion between helix1 and helix 2 in the homeodomain. All identified members of this class in hybrid aspen have an expression pattern suggesting a role in wood formation. This group of transcription factors were also identified in *Arabidopsis thaliana*, the model plant for plant molecular biology. Function of the PALE-class of homeobox genes was investigated using, transgenic hybrid aspen and transgenic Arabidopsis plants, expressing different PALE class members in sense and anti-sense direction under the control of the CaMV 35S promotor. Transactivation properties of the PttHB1 protein was analysed in yeast, and it was demonstrated that a region in PttHB1, which show homology to the strong transactivation domain of VP16, can function as an activator of transcription in yeast.

In the second approach, an EST sequencing project was initiated where 4809 EST's, originating from the cambial region of hybrid aspen, were sequenced. These EST's corresponded to 2988 genes, based on cluster analysis. 745 different proteins were identified and annotated (not counting iso-enzymes). Using these EST's as a base, a cDNA microarray was constructed representing the 2988 genes. The microarray was used for detailed transcriptprofiling of the wood-differentiation process, initiating in the meristematic cambium cells and terminating in mature xylem cells. In order to enable this experiment, a high fidelity target amplification method was developed, which is based on a 3'-prime tagged, PCR amplification protocol. This method allows transcript profiling with cDNA microarrays using minute amounts of starting material (~0.1mg of plant tissue). In practical use, two-fold expression changes observed with this technology is significant with 99% confidence. Using this technology, a transcriptional roadmap to xylem formation was produced. This data set describes unique, hitherto, unknown expression patterns for thousands of genes during xylem formation. We also demonstrate the utility of this expression data as a platform to rapidly perform functional analysis of genes involved in xylem formation, using *Arabidopsis thaliana* mutant collections.

Key words: Homeobox genes, PALE genes, wood development, transcript profiling, target amplification, genomics, EST sequencing.

Distribution:
Swedish University of Agricultural Sciences
Department of Forest Genetics and Plant Physiology
SE-901 83 Umeå, Sweden

Umeå 2001
ISSN: 1401-6230
ISBN:97-576-6057-3

A Functional Genomics Approach to Wood Development

Magnus Hertzberg
Department of Forest Genetics and Plant Physiology
Umeå

Doctoral thesis
Swedish University of Agricultural Sciences
Umeå 2001

Acta Universitatis Agriculturae Sueciae
Silvestria 173

ISSN 1401-6230
ISBN 91-576-6057-3

Printed by: SLU, Grafiska Enheten, Umeå, Sweden, 2001

Abstract

Magnus Hertzberg. 2001. A Functional Genomics Approach to Wood Development. Doctor's thesis. Silvestria 173. ISSN 1401-6230, ISBN 97-576-6057-3

Wood is widely used for various purposes, e.g. for pulp and paper, construction material and heating. Thus it is surprising that the cellular and molecular regulation of wood formation remains poorly characterised. This study therefore addresses the process of wood formation in order to elucidate factors controlling this process at the molecular level. Two different approaches were taken, using novel technologies that have become available in recent years.

First, a new group of homeobox transcription factors were identified, denoted the PALE class due to a five amino acid insertion between helix1 and helix 2 in the homeodomain. All identified members of this class in hybrid aspen have an expression pattern suggesting a role in wood formation. This group of transcription factors were also identified in *Arabidopsis thaliana*, the model plant for plant molecular biology. Function of the PALE-class of homeobox genes was investigated using, transgenic hybrid aspen and transgenic *Arabidopsis* plants, expressing different PALE class members in sense and anti-sense direction under the control of the CaMV 35S promotor. Transactivation properties of the PttHB1 protein was analysed in yeast, and it was demonstrated that a region in PttHB1, which show homology to the strong transactivation domain of VP16, can function as an activator of transcription in yeast.

In the second approach, an EST sequencing project was initiated where 4809 EST's, originating from the cambial region of hybrid aspen, were sequenced. These EST's corresponded to 2988 genes, based on cluster analysis. 745 different proteins were identified and annotated (not counting iso-enzymes). Using these EST's as a base, a cDNA microarray was constructed representing the 2988 genes. The microarray was used for detailed transcriptprofiling of the wood-differentiation process, initiating in the meristematic cambium cells and terminating in mature xylem cells. In order to enable this experiment, a high fidelity target amplification method was developed, which is based on a 3'-prime tagged, PCR amplification protocol. This method allows transcript profiling with cDNA microarrays using minute amounts of starting material (~0.1mg of plant tissue). In practical use, two-fold expression changes observed with this technology is significant with 99% confidence. Using this technology, a transcriptional roadmap to xylem formation was produced. This data set describes unique, hitherto, unknown expression patterns for thousands of genes during xylem formation. We also demonstrate the utility of this expression data as a platform to rapidly perform functional analysis of genes involved in xylem formation, using *Arabidopsis thaliana* mutant collections.

Key words: Homeobox genes, PALE genes, wood development, transcript profiling, target amplification, genomics, EST sequencing.

Author's address:Umeå Plant Science center. Department of Forest Genetics and Plant Physiology, The Swedish University of Agricultural Sciences, SE 901 83 Umeå, Sweden.

Contents

Appendix

List of papers

The present thesis is based upon the following papers, which will be referred to by the corresponding Roman numerals.

I. **Hertzberg, M.** and Olsson, O, (1998), Molecular characterization of a novel plant homeobox gene expressed in the maturing xylem zone of Populus tremula x tremuloides, Plant Journal **16**. 285-295.

II. Sterky, F., Regan, S., Karlsson, J., **Hertzberg, M.**, Rhode, A., Holmberg, A., Amini, A., Bhalerao, R., Larsson, M., Villarroel, R., Montagu, M., Sandberg, G., Olsson, O., Terri, T. T., Boerjan, W., Gustafsson, P., Uhlen, M., Sundberg, B. And Lundeberg, J. (1998) Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags. Proc. Natl. Acad. Sci. USA **95**,13330-13335.

III. **Hertzberg, M.**, Sivertzon, M., Aspeborg, H., Nilsson, P., Sandberg, G. and Lundeberg, J, (2001). cDNA microarray analysis of small plant tissue samples using a cDNA tag target amplification protocol (accepted for publication in Plant Journal).

IV. **Hertzberg, M.**, Aspeborg, H., Schrader, J., Andersson, A., Bhalerao, R., Blomqvist, K., Rahman, D., Marchant, A., Bennett, M., Uhlén, M., Teeri, T., Lundeberg, J., Sundberg, B., Nilsson, P. and Sandberg, G. (2001) A transcriptional roadmap to xylogenesis revealed by cDNA microarray analysis. (manuscript)

V. **Hertzberg, M.,** Martinsson, J. and Olsson, O, (2001). PALE homeobox genes in *Arabidopsis thaliana*: Cloning and initial characterisation of the *AtPALE2* gene. (Manuscript).

Introduction

Wood formation

Xylem formation (xylogenesis) is the process whereby meristematic cells differentiate into xylem cells. The xylem serves as a mechanical support for the stem and has an essential role in transporting water to all the different plant organs. Along with the water, mineral nutrients and other substances such as plant hormones are co-transported. In some species xylogenesis continues even after elongation growth has ceased, leading to the production of secondary xylem. In woody species this results in wood. Wood formation is initiated by a secondary meristem, called the vascular cambium. This meristem originates from the pro-cambium, which in turn develops from the apical meristem (Esau, 1977; Mauseth, 1988). In woody species, the cambium occurs as a continuous ring of cells between the xylem and the phloem throughout the entire length of fully expanded shoots and roots. The cambium is composed of a few layers of cells, one layer of which is believed to comprise the cambial initial cells (Larson, 1994). The dividing cells on each side of the initial cells are called phloem- and xylem- mother cells, respectively. The initials plus the mother cells collectively comprise the cambium or cambial zone, (Larson, 1994; Wilson et al., 1966). The characteristic cell files seen in a cross-section of a stem are the result of cambial cell divisions, in tightly controlled frequency, places and planes, and the subsequent, regulated expansion and differentiation of various cambial derivatives into tracheary elements, vessels, fibres, parenchyma, and sieve elements (Mauseth, 1988). The vascular cambium increases the diameter of an axis by periclinal divisions and its circumference by anticlinal divisions. On the xylem side of the cambium the entire process, from the first division of a cambial initial to the final development of the different forms of mature xylem cell types, occurs in a number of defined phases. Anatomical observations have shown that these phases can be visualised in several different zones. Closest to the cambium there is a dividing zone, where the xylem mother cells continue to divide. Immediately inside the dividing zone an expansion zone can be distinguished, where the derivative cells expand in all directions, both radially and longitudinally (forming tracheids and fibre tracheids) and radially and tangentially (forming vessel elements), to their final sizes. Next, a maturation zone develops, where secondary cell wall thickening and lignification of the xylem cells occur. Finally, programmed cell death is initiated, leading to the termination of all cellular processes (Pennell and Lamb, 1997; Wilson *et al.,* 1966). The ray cells continue to live and contribute to the chemical structure of the fibres and vessels (Barnett, 1981; Pickett-Heaps, 1968) through continued lignification, and in some species in the production of heartwood that occurs several years later (Magel, 2000).

Ways to study xylem development

Classical work

To gain insight into the developmental regulation of xylogenesis, a number of approaches have been adopted. Early work focused on anatomical and developmental description of wood, and the influence of environmental stimuli on the developmental patterns. This yielded knowledge that still provides a basis for all studies of xylogenesis (Larson, 1994; Barnett, 1981). Another area that has been extensively studied is the involvement of phytohormones in xylogenesis and wood formation (Little and Pharis, 1995). Auxin, cytokinins, abscisic acid, ethylene, gibberellins and brassinosteroids have all been implicated in the control of xylem formation (Savidge, 1996; Higuchi, 1997). Auxin, more specifically indole-3-acetic-acid, has been the most intensively studied and is probably the most important hormone in xylem formation. According to the canalisation theory auxin is a key regulator of the initial vascular strand formation and patterning (see Berleth et al., 2000 for a review). Alternative theories have been presented, but in these too Auxin is still perceived as having a role. (Koizumi et al, 2000). Auxin is also required in the regulation of secondary xylem formation, both for maintaining the cambial meristem and cambial activity, and for the subsequent differentiation process (Little and Pharis, 1995; Sundberg et al 2000). In order to elucidate the function of auxin in xylogenesis, exogenous applications of the hormone and inhibitors of auxin transport have been widely used, in conjunction with auxin measurements. *In planta* overproduction and down regulation of auxin-levels via transgene technology has also been a useful approach (Klee et al., 1987; Romano et al., 1991; Sitbon et al., 1992; Tuominen et al., 1996).

Plant model systems used for studies of xylogenesis

Studies on xylem formation have been conducted in several different experimental systems, from cell suspension cultures of *Zinnia* (Fukuda, 1997), *in vitro* systems using callus or explants, and intact plants. Today, three different model plant systems are much used in molecular and genetic studies of xylogenesis. These are *Arabidopsis thaliana*, *Zinnia* and *Populus*, but other plants are also used (e.g. *Pinus taeda, Pinus radiata*, birch, tobacco).

Arabidopsis, the main model system for plant research (Meinke et al., 1998), is a small plant with a life cycle of just 6-8 weeks under optimal growth conditions.

This, in combination with its capacity for self-pollination makes it an excellent system to use in controlled crosses for purposes such as mutant mapping and genetic complementation experiments, and for the production of double and triple mutants. *Arabidopsis* is also easy to transform, and large collections of mutants are widely available, produced by both chemical mutagenesis and T-DNA or transposon insertion mutagenesis. Furthermore, *Arabidopsis* has a relatively small genome, of 125Mbp, for which the almost complete DNA sequence was reported in December 2000 (The Arabidopsis Genome Initiative, 2000). The availability of the genomic sequence of *Arabidopsis* gives the opportunity to approach plant biology in new ways, and in the near future it will probably be possible to monitor the whole transcriptome of *Arabidopsis thaliana* using microarrays. Already, cDNA-microarrays and oligonucleotide-based microarrays containing ~10,000 clones are available (see Weisman and Ohlrogge, 2000; Zhu and Wang, 2000).

The *Zinnia* system is based on cultures of *Zinnia* leaf mesophyll cells that are induced to re-differentiate, by appropriate applications of auxin and cytokinins, into tracheary elements with an efficiency of up to 80% (Roberts and McCann, 2000). This system has been used for gene finding and for studying the effects of applying plant hormones and various other substances (Fukuda, 1997; Roberts and McCann, 2000). One of the advantages of the *Zinnia* system is that it is possible to synchronise the entire re-differentiation process, enabling easy sampling at specific stages and the possibility of adding or withdrawing substances quickly and efficiently at any time during the re-differentiation process.

The third commonly-used model system for xylem differentiation is *Populus*. Being a tree, *Populus* produces large amounts of secondary xylem (wood), in contrast to *Arabidopsis*, which only produces secondary xylem under special conditions (Lev-Yadun, 1994; Regan et al., 1999; Little et al., 2000). *Populus* is fast growing, it has a relatively small genome of ~450Mbp, and is easily transformable, enabling reverse genetic approaches to be used, such as over-expression and down regulation of specific genes. *Populus spp.* are widely used for commercial growth of fibres, facilitating their use in practical biotechnology. Different *Populus* species have been used in a large number of investigations related to xylem formation including studies on lignification (Baucher et al., 1996; Hu et al., 1999) and the roles of auxin and gibberellins in xylem formation (Tuominen et al., 1996; Eriksson et al., 2000). The initiation of a large EST sequencing project in *Populus tremula x tremuloides* (Sterky et al., 1998; this thesis) has further strengthened the position of *Populus* as the model system of choice for investigating wood formation by identifying a large number of new genes potentially involved in the process.

Genetic approaches

Two fundamentally different approaches are used to elucidate molecular/ genetic factors involved in xylem formation. One is forward genetics, where a mutated population of individuals is screened for a specific trait, such as a thicker inflorescence stem. The other approach is reverse genetics, which is used to identify the function of a gene by introducing specific mutations into it (for example by changing the promotor) *in vitro* and re-inserting the mutated gene back into the plant by transgene technology.

Arabidopsis thaliana is an excellent model system to use in forward genetic experiments to study xylem formation. The most difficult step associated with this approach is to set up an easy and robust assay that enables a large population of mutated plants to be screened for the phenotype of interest. After a primary mutant has been found much of the subsequent work involves classical genetics (Foster and Twell, 1996). During the final stage, in which the phenotype of the mutatant is fully characterized to elucidate the function of the mutated gene, further transgenic strategies can be exploited. A number of screens for vascular mutants have been performed in *Arabidopsis*, (reviewed in Roberts and McCann, 2000; Fagard et al., 2000). These screens have identified mutations in various classes of genes related to xylogenesis. The *irx3* (*irregular xylem 3*) mutant, for example, has a mutation in a cellulose synthase gene involved in secondary cell wall synthesis, giving the plant a collapsed xylem phenotype (Turner and Sommerville, 1997; Taylor et al., 1999). Another example of such a mutation alters the gene encoding Ferulate-5-hydroxylase (F5H, an enzyme involved in the lignin pathway), thus generating the *fah1/sin1* mutant, which has a reduced proportion of syringyl units in its lignin (Meyer et al., 1996). Mutants have also been identified with mutations in regulatory genes, like the *ifl1* or *interfascicular fiberless 1* mutant (Zhong et al., 1997), used to demonstrate that the *IFL1* gene encodes a homeodomain (HD) protein (Zhong and Ye, 1999). The *IFL1* gene is part of a class of homeobox (HB) genes that encode proteins containing a HD-leucine zipper domain (Ruberti et al., 1991; Schena and Davids, 1992; Zhong and Ye, 1999). *IFL1* is most similar to genes in the HD-zip 3 sub group, to which *ATHB-8* belongs. *ATHB-8* has previously been implicated in vascular development due to its expression in procambial cells and during early stages of re-vascularisation after wounding (Baima et al., 1995). Other mutants that were not initially identified when screening for phenotypes indicative of vascular malfunction can also sometimes give insights into xylem development. A number of *Arabidopsis* mutants have been described, for instance, that are impaired in different aspects of auxin transport and auxin perception. Many of these mutants show pleiotropic phenotypes, and at least six examples (*monopteros, bodenlos, pin1, gnom, lop1* and *axr6*) display vascular abnormalities (see Hardtke et al., 1998; Hamann et

al., 1999; Galweiler et al., 1998; Mayer et al., 1993; Carland and McHale, 1996 and Hobbie et al., 2000, respectively). This provides further genetic evidence that auxin is fundamentally involved in xylem formation. It also shows that previously-described mutants should be checked for vascular phenotypes if present knowledge gives any reason for thinking that the mutated function could be involved in xylogenesis, as is the case with auxin related mutants.

Homeobox genes

The homeodomain

Homeobox genes encode transcription factors and contain a semi-conserved DNA region of ~180bp, denoted the homeobox (Bürglin, 1994). The homeobox encodes the homeodomain (HD) part of these factors, a DNA binding domain with a number of highly conserved amino acids that are essential for the ordered helix-loop-helix-turn-helix structure of the HD and its DNA binding properties (Billeter *et al.,* 1990; Otting *et al.,* 1990; Gehring *et al.,* 1994).

Homeobox genes are part of a large gene class

Homeobox (HB) genes have been found in all eukaryotic organisms for which large amounts of DNA sequence data are available. Most likely, they are present in all eukaryotes. Use of the "Simple Modular Architecture Research Tool", SMART (Schultz et al., 2000), to search protein databases (http://smart.embl-heidelberg.de/) shows that humans and mice contain at least 246 and 217 HD-encoding genes, respectively (these numbers will probably increase until the complete genomes have been sequenced and analysed). In insects such as *Drosophila melanogaster* 176 different HD-coding genes can be identified and in the *Caenorhabditis elegans* 113 HD-coding genes can be found using the SMART tool. In *Arabidopsis thaliana* 92 HD-coding genes are known to be present and, finally, in the single cell organism *Saccharomyces cerevisiae* 10 different HB genes have been found. Thus, multicellular organisms such as plants and animals contain large numbers of homeobox genes while yeast has only a few. Bacteria do not contain any HD proteins. However, *Arabidopsis* has more genes encoding transcription factors than the above-mentioned animal species (The Arabidopsis genome initiative, 2000) so the class of HD-coding genes, has expanded relatively less in *Arabidopsis* than in animal species.

Function of HD transcription factors

In total, 2248 HB genes are present in protein databases, as identified by the SMART-tool. HD proteins have been shown to be crucial in the control of many, diverse cellular and developmental processes, such as spatial patterning, positional information, cell fate determination and cell differentiation, in virtually all eukaryotic organisms studied, from yeast to man (Gehring *et al.,* 1994; Lawrence and Morata 1994; Scott et al, 1989). A large number of HB genes have been studied in plants (see the *Introductions* and *Discussions* of Papers I and V included in this thesis for a number of examples). Many of these genes have been shown to be involved in developmental processes at different levels. For example, the *SHOOT MERISTEMLESS* (*SM*) gene plays a role in the formation and maintenance of the shoot apical meristem, showing that homeobox genes help determine stem cell function (Barton and Poethig, 1993; Endrizzi *et al.,* 1996; Long *et al.,* 1996). Similarly, the *GLABRA2* gene has been shown to control trichome differentiation (Rerie *et al.,* 1994) and the hairless cell character in the root epidermis (Masucci *et al.,* 1996; Schiefelbein *et al.,* 1997). These observations show that some HB genes can be crucial in determining the identity of single cells.

The role of HB genes (like that of any other gene) can be identified by both reverse and forward genetics. The putative function of two HB genes involved in xylem formation, for instance, was elucidated in two different ways. The *Arabidopsis* gene *IFL1* was identified by screening an EMS-mutated *Arabidopsis* collection for mutants with altered fibre formation through histological analysis of the inflorescence stem (Zhong et al., 1997). The mutated gene in identified mutants was cloned, using a positional cloning strategy (Zhong and Ye, 1999). The *ifl1* mutant totally lacks inter-fascicular fibres, so the inflorescence stem is unable to stand erect. The mutation also reduces the level of vascularisation in the vascular bundles. The expression pattern of this gene in the interfascicular regions and in vascular bundles further supports the hypothesis that *IFL1* is involved in xylem differentiation. Another HB gene, the *Oshox1* gene from rice (Meijer et al., 1997), was cloned by homologue screening of a rice cDNA library with the hetorologous probe *HAT4* from *Arabidopsis* (Schena and Davids, 1992). Initial reverse genetics experiments in *Arabidopsis* indicated that it has a role in general growth. In later reverse genetics experiments the *Oshox1* gene was over-expressed in rice and the results (in combination with expression data from *Oshox1* promoter-GUS fusions in the same species) indicate that it is involved in the step that commits cells to a pro-vascular fate (Scarpella et al., 2000). The *IFL1* and *Oshox1* genes both belong to the plant specific HD-ZIP class of homeobox genes (see Paper I in this thesis for a short description of the different homeobox gene classes in plants). If one assumes that phylogenetically related genes are likely to be functionally related, this indicates that other HD-ZIP genes could be

involved in vascular differentiation. Indeed, two other HD-ZIP genes have been proposed to be involved in regulating different stages of xylem formation, namely *ATHB-2* and *ATHB-8* (Steindler et al., 1999; Baima et al., 1995).

Functional genomics

Now that the genome of *Arabidopsis* has been completely sequenced (The Arabidopsis Genome Initiative, 2000) and genomic sequencing efforts are underway in other species such as rice and maize, together with numerous EST sequence projects on different plant species, new methods can be adopted to investigate plant biology and gene function. Microarray technology has also recently made it possible to monitor the expression of thousands of genes simultaneously (extensively reviewed in the supplement to volume 22 of Nature, 1998). The two kinds of microarray approach that are currently most commonly used are the Stanford or spotted cDNA, and the Affymetrix techniques. In the spotted cDNA approach, cDNA clones are spotted in high density on microscope-type glass slides using a robotic arrayer (Schena et al., 1995, 1996). With this technique, it is possible to array all 25,000 gene sequences from *Arabidopsis* on a single slide. The other method, from Affymetrix, exploits a technique developed in the electronic industry to synthesise short (20-mers) complementary to the 3'-end of the corresponding mRNA, this method can used to synthesise as many as ~0.7 million oligonucleotides per square cm (Lipshutz et al., 1998) making it possible to monitor up to 60,000 genes per hybridisation (Affymetrix uses multiple oligos per gene analysed), these numbers could be increased. Which of these methods is best with respect to factors like specificity, linear range and and other analytical parameters is debatable. However, the current method of choice, for experiments on species that are not main model species (like *Arabidopsis*), is the spotted cDNA technique, because it is relatively cheap and simple. There are also large-scale methods being developed to monitor the proteome and metabolome by measuring the levels and status of thousands of proteins and metabolites in parallel (Pandey and Mann, 2000; Fiehn, et al., 2000).

Combining data from transcript profiling experiments with previous knowledge of the genes and other large data sets, such as those obtained from proteomic and metabolomic screens as well as data from large scale 2-hybrid screens, will facilitate the elucidation of gene function. The proposed function should of course then be tested using the standard techniques that are currently in use, a key technique being reverse genetics using transgenic technology (e.g. the production and analysis of over-, under-, miss- and null- expressers of the gene being considered).

Present Study

The PALE class of Homeobox genes

Cloning of hybrid aspen PALE genes

In order to gain insight into the molecular regulation of xylogenesis we initially started to clone genes that were potential regulators of the process, using differential display (Liang and Pardee, 1992). Different tissues from the cambial region were compared and tissue-specific bands were cloned (data not shown). However, only the 3' ends of the cDNA's are obtained using this technique, which makes it difficult to identify possible homologues by searching DNA and protein databases. Thus, the identity of the cloned differentially expressed genes in most cases remained unknown. Because of this, it was decided to adopt an alternative strategy.

As homeobox genes are known to be involved in the regulation of many different processes we decided to clone genes of this class expressed in the cambial region of *Populus tremula x tremuloides* (hereafter denoted hybrid aspen). Our strategy was to investigate the expression pattern of the cloned HB genes and select candidate genes with an expression pattern indicating a specific role in xylem-formation. A cDNA library was constructed with mRNA from the cambial region of hybrid aspen (see Papers I and II). This cDNA library was used in two ways. In the first approach, the library was screened with probes specific to HB genes (Paper I), and in the second the library was used for EST sequencing (see Paper II, and below).

Two different probes were used to screen the cDNA library for HB genes: a PCR-amplified fragment covering the HD of the Knotted (KN) type of HD and a degenerate primer mix complementary to the third helix region of the HB (originally described by Bürglin et al., 1989 and later modified by Baima et al., 1995). From the screen with the KN probe, one KN-like gene was cloned. However, this gene was predominantly expressed in the bark region of the stem, which doesn't correlate with a function in xylem development (data not shown). A screen conducted with the Helix 3-specific oligo mix, on the other hand, revealed two different homeobox genes, named *PttHB1* and *PttHB2*, belonging to a type of homeobox gene that had not been described previously. We assigned these genes to a new class, designated "PALE" since they have a five amino acid insertion between Helix 1 and Helix 2 of the homeodomain (Paper I). Expression pattern analysis revealed that both these PALE genes had a distinctive and differential expression during the xylem formation process. The *PttHB1* gene was mainly expressed in the maturation zone, during which the secondary cell wall is laid down and lignified. *PttHB2* was expressed in

earlier developmental phases on both sides of the cambium, as well as in the cambium itself (Paper I). Both of these expression patterns are indicative of a possible role in xylem/vascular development. The *PttHB1* gene could be involved in regulating the developmental switch that changes cells undergoing expansion into cells undergoing secondary cell wall formation. It could also be involved in regulating a specific process that is induced, or up-regulated, in this zone (e.g. cellulose synthesis). The *PttHB2* gene is expressed in dividing and expanding cells during vascular development and might thus be involved in early stages of the process by, for instance, keeping the cells competent for vascular formation, controlling cell proliferation or repressing secondary cell wall formation.

Cloning and identification of *Arabidopsis* PALE genes

In order to investigate whether members of the PALE class of homeobox genes are present in other organisms, protein databases were regularly searched for the presence of genes homologous to *PttHB1* and *PttHB2*. Approximately a year after the two *PttHB* genes were cloned, an *Arabidopsis* homologue, with high similarity to *PttHB2*, not only in the HD region but over the whole protein, was found and named *AtPALE2* (Paper V). This sequence was elucidated during the Arabidopsis Genome Initiative's (AGI) sequencing of the *Arabidopsis* genome. An *AtPALE2* cDNA covering the whole coding region was RT-PCR amplified from mRNA isolated from young bolting *Arabidopsis* plants. Initial expression analysis of *AtPALE2* revealed that it was expressed in most tissue types, with the highest expression in hypocotyls (see Paper V). This expression pattern is consistent with the possibility that *AtPALE2*, like *PttHB2*, is expressed in developing vascular tissues. Thus, since they have high sequence similarity, a close phylogenetic relationship (Paper V), and a possible convergence in expression pattern, *AtPALE2* and *PttHB2* may be orthologues to each other.

A search of available protein databases at the end of the year 2000 (2000-12-01) revealed a small family of PALE homeobox genes in *Arabidopsis* consisting of 14 members. At least four of these are expressed, since corresponding EST's have been found, indicating that they are real genes (Paper V). Possible redundancy amongst different members could be anticipated due to high identity between pairs of the PALE genes. For example *AtPALE3* and *AtPALE4* display 78% similarity over the whole protein sequence and 86% identity in the HD.

Functional analysis of PALE genes

In order to gain information concerning the function of PALE genes, especially with respect to their possible function in regulating xylem development, a number of transgenic plants were constructed. In a first set of experiments the *PttHB1* gene was ligated into the pPCV702.km binary plant transformation vector (Koncz and Schell, 1985) in sense and antisense direction and transformed into hybrid aspen and *Arabidopsis* (essentially as described in Nilsson et al., 1992 and in Bechtold et al., 1993). For hybrid aspen, primary transformants had to be used for analysis. Therefore, in these experiments, wt plants that had been raised from tissue culture and cultivated in the same way as the transgenics were always included as controls. In *Arabidopsis,* independent antibiotic resistant transgenic plants (T1) were selected and self crossed. Plants from the resulting seeds (T2) were again tested for antibiotic resistance, and lines showing a Mendelian ~3:1 inheritance ratio of antibiotic resistant to sensitive plants, were identified for further study. Homozygous plants were isolated in the following generation (T3), by scoring their selfed seeds, and picking plants that gave an all:0 ratio for antibiotic resistance. Homozygous T4 plants were then used for phenotypic analyses.

Seven hybrid aspen 35S:*PttHB1* sense, three 35S:*PttHB1* antisense plants and two wt plants were grown in parallel. In an initial analysis, no difference could be detected in growth rate or morphology between these plants and wt. Total RNA was prepared from each plant and used for northern blot analysis using the *PttHB1* cDNA as a probe (Figure 1). This showed that the several chosen plants over-expressed the *PttHB1* gene to differing degrees and that the *PttHB1* gene may have been down-regulated in two anti-sense plants, AS2 and AS4. Four additional *PttHB1* anti-sense plants were later produced, but none of them displayed any obvious phenotypic deviation from wt.

The hybrid aspen plants were sampled at a height of ~1.5m. Samples were taken from internodes at the top and bottom of the plant for further sectioning. The sections were stained with toludine blue as a general stain and with phloroglucinol to stain lignin. This analysis did not reveal any differences between the control and transgenic hybrid aspen plants. A sample from the basal region of the stem was used for measuring fibre length and width, as well as cell wall thickness. For this purpose, stem pieces were cut out and macerated using acetic acid and hydrogen peroxide for ~4 hours in a boiling water bath (Franklin, 1945). The macerated fibres were then analysed using a Kajaani FiberLab instrument (Valmet Automation, Finland). Again, no differences were detected between wt, *PttHB1* anti-sense and *PttHB1* over-expressers (Figure 2).

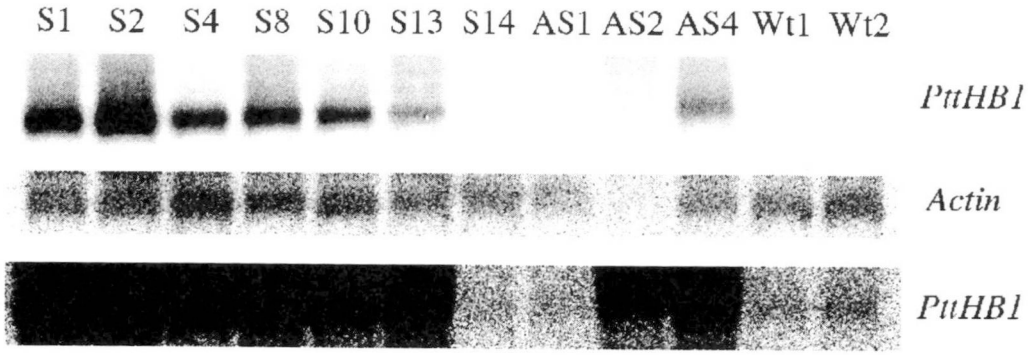

Figure 1
Northern blot analysis of transgenic 35S:*PttHB1* sense and anti sense hybrid aspen plants. Probes used indicated on the right. The 2 *PttHB1* sub figures represent different exposures of the same hybridisation.

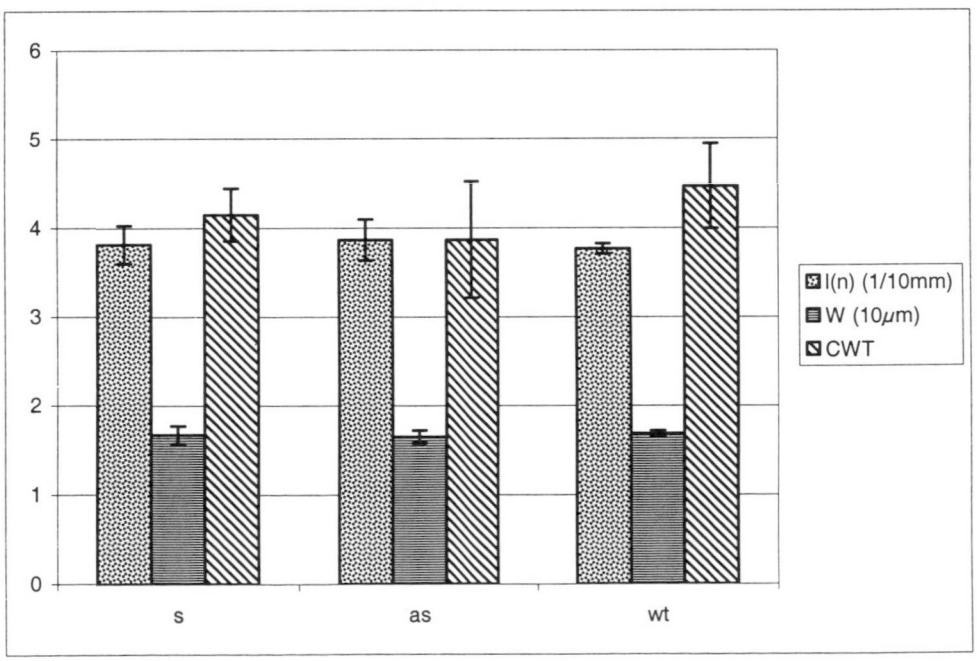

Figure 2
Fiber analysis of wt and 35S:*PttHB1* sense and anti-sense hybrid aspen plants. 3 wild type plants were poled, 7 sense plants were poled and 3 anti-sense plants were poled in this analysis. l(n)= average fiber length (x 0.1mm); W= average fiber width (x 10μm); CWT= average Kajaani cell wall thickness. Error bars indicate standard deviation.

Arabidopsis PttHB1 over-expressers were analysed by examining several different features: growth rate, lateral root formation, anatomy of inflorescence stem, toludine blue and phloroglucinol stained hypocotyls, roots and inflorescence stems. No differences were seen in any of these experiments. However, an interesting difference was observed when germinating and growing the *PttHB1* transgenic *Arabidopsis* seedlings in darkness on 1% sucrose + MS + 1% agar plates (Figure 3), where a stochastic occurrence of a light-grown phenotype was observed in all four lines analysed. Photomorphogenesis is a complex biological process involving several distinct developmental stages, transcription factors and regulatory complexes. The relevance of this observation, if any, is therefore difficult to discern at this stage but it is definitely worth further study. Based on the *Arabidopsis* genome sequence, we know that there are no obvious orthologues to the hybrid aspen *PttHB1* gene in *Arabidopsis* (Paper V). Interestingly, *ATHB-2* (*HAT4*), which is proposed to be involved in light regulated processes (Steindler et al., 1999), is related to *PttHB1* (see paper I, Figure 2d). However, the phenotypic characteristics observed in experiments in which *ATHB-2* transcript levels in transgenic *Arabidopsis* are altered are not phenocopied or reversed in the *PttHB2* transgenics. Thus, the functional relationship between the two genes, if any, is not straightforward. However it is still possible that the PttHB1 protein in the transgenic plants interacts in the same pathway as ATHB-2 by partially binding to the same cis-elements. Such an interaction could be of physiological relevance, but could also be an artefact of over-expressing the PttHB1 protein.

In order to elucidate the function of *AtPALE2*, sense 35S:*AtPALE2* and two types of anti-sense (full length and exon 2) *Arabidopsis* plants have been produced recently see paper V. However, analysis of these plants has not yet given any insight into the function of *AtPALE2* (Paper V).

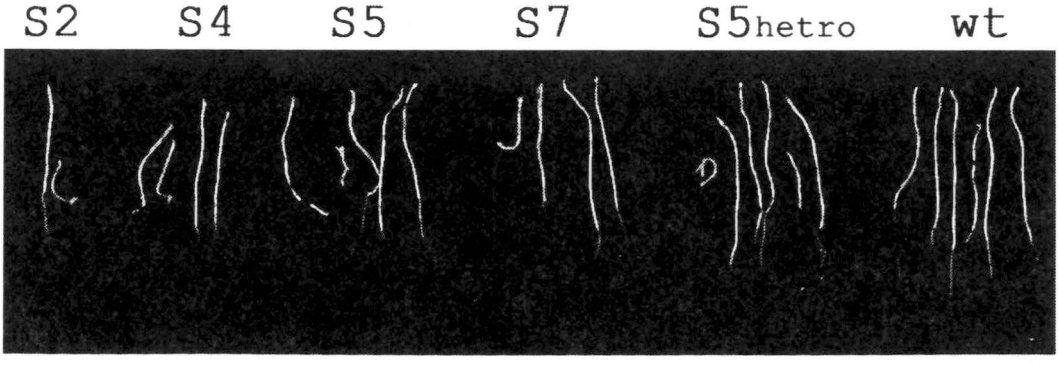

Figure 3
Phenotype of 4 different 35S:*PttHB1* homozygot Arabidopsis transgenenic lines. Plants from one heterozygot parent, for one of the lines, are indicated in the figure. Six seeds of each line were germinated and grown for 5 days in the dark.

PttHB1 contains a domain that is homologous to the transactivating domain of VP16 (see Paper I). In order to investigate if this domain could function as a transactivating domain, serial deletions of the *PttHB1* cDNA were fused in frame with the GAL4-DNA binding domain (Figure 4), in the pAS2 vector (Durfee et al., 1993). These constructs were introduced into yeast strain Y190, which carries a *lacZ* reporter gene under the control of a yeast minimal promoter behind four copies of the DNA binding site for GAL4. Colonies with the different constructions were grown on Hybond-N membranes placed on synthetic solidified growth media. After three days of growth, the colonies were stained with x-gal (Durfee et al., 1993), for 16 hours, and a strong signal was fully developed on the filter after three hours of incubation (figure 5). This experiment demonstrates that PttHB1 contains a transactivating domain, which lies in the VP16 transactivation-like domain (Figures 4 and 5), indicating that PttHB1 is an activator of transcription.

Future strategies for functional analysis of PALE genes, will be presented in the section on future perspectives.

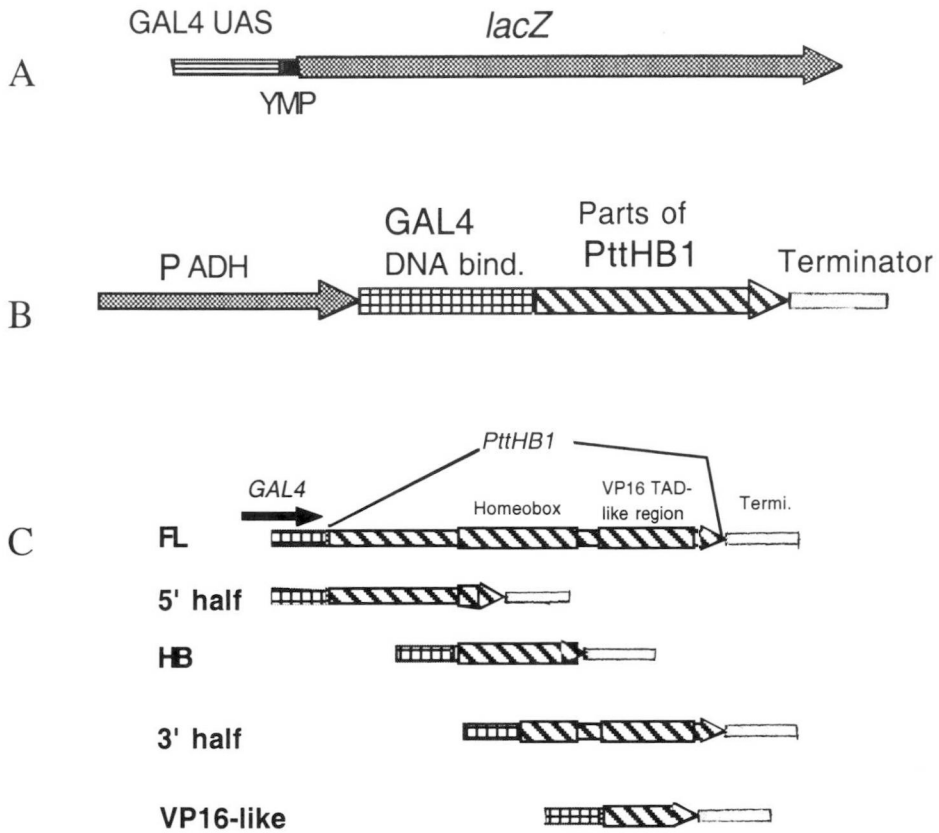

Figure 4
Schematic presentation of a transactivation study of PttHB1. A: Reporter-gene construct in Y190. B: Overview of the expression cassette of pAS2 vector. C: The different deletions used.

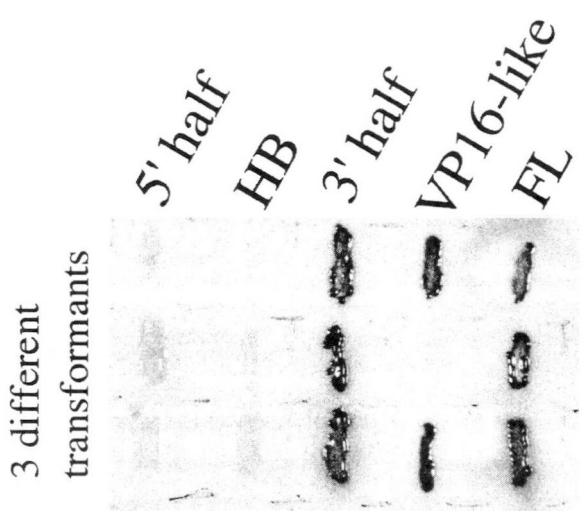

Figure 5.
X-gal staining of 3 individual colonies of each construct introduced into Y190. Constructs indicated in the figure (se figure 4C).

Identifying genes by EST sequencing

The cambial region specific cDNA library from hybrid aspen was used as the basis for an EST sequencing project. From 4809 cambial region EST sequences, 2988 different genes were identified by cluster analysis, and EST's (singletons and clusters) encoding 745 different proteins (counting sets of iso-enzymes as single proteins) with known functions were identified and annotated from the cambial region library. In total, including EST sequences produced from a xylem specific cDNA library (from *Populus trichocarpa*), 820 different proteins with known function were identified and annotated (Paper II). For example, all the identified genes involved in lignin biosynthesis were found among the cambial region EST's that were sequenced (Paper II). Genes involved in diverse aspects of cell biology were identified (Paper II, Figure 2), including many with putative regulatory functions, such as MADS-box genes, Homeobox genes, and genes encoding trimeric G-proteins, protein kinases, protein phosphatases and others. An interesting group of genes is represented by EST's that showed no significant homology to any sequence in the databases (12% at that time, Paper II), and it is tempting to speculate that some of these genes could be specific, previously unknown, regulators of wood formation.

Transcript profiling during secondary xylem formation

Construction of an hybrid aspen wood cDNA microarray

Using a unigen set based on the produced EST's (Paper II), a transcript-profiling project was initiated (Papers III and IV). Singletons and one member from each cluster (Paper II) from the cambial region EST's were arrayed into a unigen set consisting of 2,995 bacterial plasmid lysates. Each of these was re-transformed into *Escherichia coli* and plasmids were subsequently prepared (Papers III and IV). The cDNA inserts were PCR amplified, purified and spotted on glass slides using a GMS 417 Arrayer (Papers III and IV).

Method development

In order to facilitate the analysis of specific cell types during wood formation, a technique was developed that allowed small samples to be used as targets for transcript profiling (Paper III). We showed that the method can be used for samples with as little as 0.1 μg of total RNA, corresponding roughly to 0.05 - 0.1 mg of plant tissue. Comparing transcript profiles obtained using this method with profiles ob-

tained using standard techniques and 1 μg mRNA for labelling, it was estimated that a 2-fold difference in expression could be distinguished with 99 % confidence using the new method and 0.1 μg of total RNA (Paper III). The method is based on fragmentation of the transcriptome prior to PCR amplification of the 3' part of the transcripts (Paper III). The amplified cDNA population is subsequently labelled with Cy3- or Cy5- labelled nucleotides, using asymmetric PCR (paper III). This approach circumvents the bias previously seen for shorter transcripts when amplifying the full-length transcriptome (Van Gelder et al., 1990). Another advantage of this amplification method is that a single amplification is enough for the several repetitive hybridisations that are needed in order to gain reliable expression data. Furthermore, specific amplification of the 3' end should avoid some of the problematic cross-hybridisation that can occur between closely related gene-family members when hybridising cDNA microarrays (DeRisi et al., 1997), by omitting most of the coding regions in the target mix.

In an attempt to simplify the amplification procedure presented in Paper III, the 5' template-switching effect was used (Clonetech, USA). In the SMART kits, produced by Clonetech, single stranded cDNA with adaptor sequences on each end (5' and 3') is produced from RNA in one reverse transcriptase reaction. This is achieved by the use of a template-switching adapter that anneals to the short string of the extra dC-nucleotides that certain reverse trancriptases add to the cDNA when reaching the end of the template RNA molecule. The reverse transcriptase then switches template from the RNA molecule to the adapter and synthesises the complementary sequence to the adapter. This technique was developed in order to create full-length cDNA populations. By fragmenting the RNA sample before cDNA synthesis and subsequent PCR amplification, a 3'-specific cDNA population, similar to the one described in paper III, can be produced in six hours. In order to test the reliability of this simplified amplification method, the same approach was taken as earlier described (see Paper III). 0.1μg of total RNA from the same xylem and phloem RNA samples as used in Paper III was fragmented by passing it through a 10 μl Hamilton syringe 50 times. cDNA was synthesised and amplified using the SMART cDNA synthesis kit as described by the manufacturer. The amplified cDNAs were labelled, hybridised, washed and analysed as described in Paper III. The results suggested that this method too could be used to differentiate between a 2-fold change in expression with 99% confidence (Figure 6). However the dynamic range of this experiment was lower than for the experiments in Paper III, using the same amount of starting material. To conclude, this simplified amplification method looks promising, but requires further optimisation.

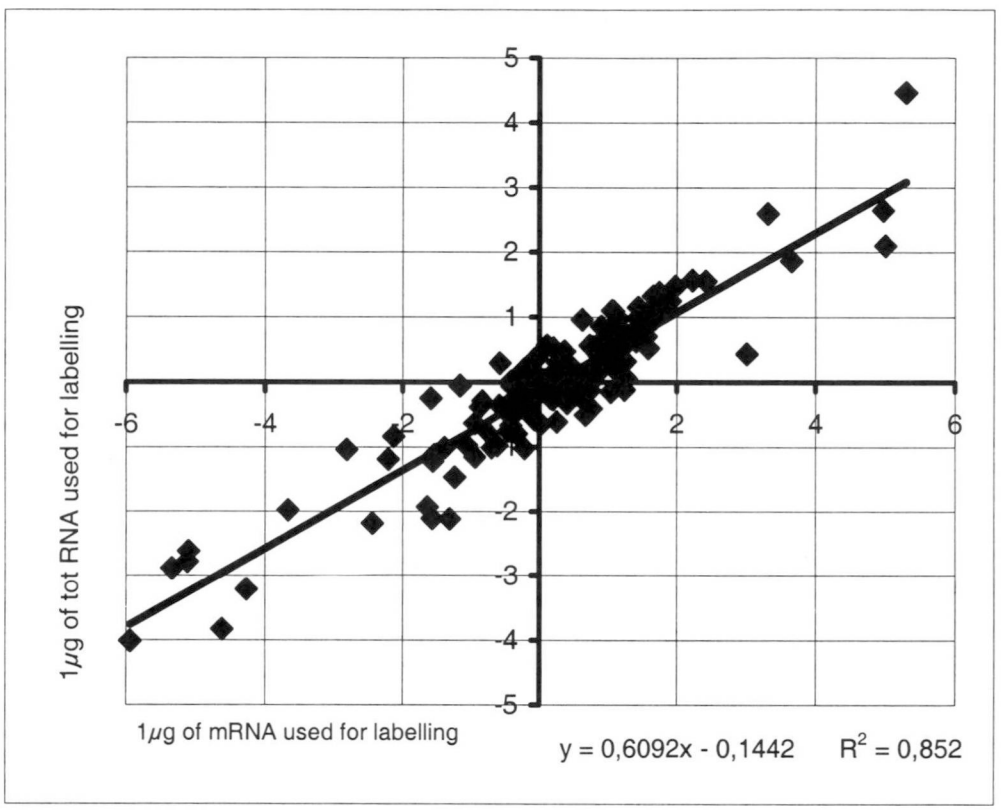

Figure 6
Scatter plots of log2 transformed expression-ratios from a hybridisation using 1μg of mRNA for labelling (X-axis), plotted against a hybridisation using 0.1μg total RNA as starting material for labelling (Y-axis). The solid line represents the regression line. The equation of the regression line is shown.

Creating a gene expression map for xylem development

As part of the EST sequencing project, a large number of genes potentially involved in xylem formation were identified. In order to learn more about these genes, their expression patterns were analysed in the vascular cambium and during subsequent developmental phases in the differentiation process leading to mature xylem cells (Paper IV). This analysis was conducted using cDNA microarrays in combination with the newly developed target amplification protocol (Paper III). Six different types of sample were used representing (progressing inwards from the bark): phloem, cambium cells, early phase of expansion, late phase of expansion, secondary cell-wall formation and finally the late maturation phase (see Figure 1 in Paper IV). These specific differentiation stages were sampled using a tangential cryo-sectioning technique developed by Uggla et al., (1996), which produces cell layer specific sections through the cambial region.

cDNA microarray analysis of a wide range of samples, spanning meristematic vascular cambium cells to cells late in terminal differentiation processes (up to and including autolysis), poses considerable problems. Two issues that have to be dealt with are normalisation and selection of a reliable reference. The presented array contains only ~3,000 different genes, which probably mostly originate from the early stages of xylem differentiation (due to RNA amount being present in tissues sampled for the cDNA library used in the EST sequencing). This causes a considerable normalisation problem since the ratio of the theoretical signal for all spots from the sample, divided by the theoretical signal for all spots from the control, will vary when different samples are analysed, thus a normalisation based on that the total sample signal divided with the total reference signal should equal 1 can not be used (or similar methods). To solve this problem we used spiked human cDNA for normalisation (see paper IV). Secondly, different samples will probably have large differences in gene expression; some genes being completely turned off in one sample and highly expressed in another. If the control sample lacks expression, or has low expression, of a significant amount of genes, expression-data for these genes would be lost or result in unreliable estimates. To solve this problem, we used a mixture of the samples as the control (Paper IV). Control experiments showed that the expression data from this experiment were reliable, according to standard filter hybridisations in which the target is placed on a filter and incubated in a solution containing a radioactively labelled probe (see Figure 2a and b in paper IV). After analysing expression profiles from two different hybridisation schemes it was concluded that 2-fold changes in expression reflect a significant change in expression (Paper IV).

From this transcript profiling experiment, one can conclude that a large proportion of the genes on the array have specific functions during the differentiation process that leads to mature xylem. For example, a significant group of genes were specifically induced in the zone of secondary cell wall formation, strongly indicating that they have a function associated with this process (Figure 3b:VIII in Paper IV). These data will substantially increase our knowledge of xylogenesis, providing both patterns of expression of genes with known function, and a basis for further investigation of the genes with unknown functions (see Paper IV). The data will also help us produce a toolbox with promoters expressed during specific stages of xylem development. Such promoters will be of great importance in functional analysis of different genes, and for biotechnological purposes.

Another class of genes of special interest are those with homology to *Arabidopsis* genes, either of unknown function, or with proposed functions in other stages of plant development and physiology. We have demonstrated the utility of the transcript profiling data from hybrid aspen using a gene that is a member of the GRAS family of plant transcription factors. This class of genes had not earlier been proposed to be involved in xylogenesis. A hybrid aspen homologue to the GRAS family is present on the array and shows a specific expression pattern (Figure 8A, Paper IV). Thus, one can now combine results from the two model systems, *Arabidopsis* and *Populus*. With knowledge of the specific expression pattern in the *Populus* stem we analysed *Arabidopsis* mutants with mutations in GRAS family members, observing a vascular phenotype in *short-root* (Paper IV). This indicates the potential of using hybrid aspen for transcript-profiling during xylem formation, and *Arabidopsis* for the subsequent functional analysis. Several other candidate genes for this kind of analysis are presented in Paper IV, figure 8.

Future perspectives

A new class of Homeobox transcription factors was identified (Papers I and V) with a possible function in xylogenesis. Initial functional analysis was performed in hybrid aspen and *Arabidopis* (Papers V and the *Introduction* to this thesis). We have established that PttHB1 probably can activate transcription, but we do not yet have conclusive information on which processes *PttHB1* regulates. The lack of an altered phenotype in *PttHB1* transgenic plants could be due to *PttHB1* acting via combinatory effects and/or redundancy in the regulatory pathways. In the future, further functional analysis of PALE genes should primarily be done in *Arabidopsis*, since this will allow insertional T-DNA and transposon knock out mutants to be obtained, and the possibility of creating multiple (e.g. double and triple) mutants, thus accelerating the analysis. The possibility of rapidly making specific crosses between series of mutants and over-expressors is especially useful when studying developmental and regulatory networks where redundancy and cross regulation could be present. If a putative function is identified in *Arabidopsis* one should, if relevant, go back to hybrid aspen plants and study its function in a tree species. Analysis of DNA binding preferences for the PALE genes would probably also give novel insights into the biological function of some of these proteins.

Secondly, EST sequencing in combination with transcript profiling revealed the expression pattern of almost 3,000 genes during the xylogenesis process. These data will be useful for identifying genes. They could also be used as the basis for a large-scale functional genomics analysis of factors involved in xylem formation, by combining the model systems *Populus* and *Arabidopsis*, exploiting the transcript profiling advantages of the former and the genetic tools available in the latter. This data set also gives expression profiles at high spatial resolution for genes that are already being studied (for example, see Figure 5D, Paper IV).

With the recent completion of the sequencing of the *Arabidopsis* genome, a number of possibly synergistic approaches could be envisioned for xylogenesis research combining hybrid aspen and *Arabidopsis*. We initiated a large-scale EST sequencing project, where the goal is to produce over 70,000 EST's, current status 36 000 EST's (The Swedish Tree Functional Genomics Program). However, although a full-scale genomic sequencing project for a *Populus spp.* would be a breakthrough for the field, the cost is considerable and not a realistic option for the present time. There are however a number of different ways in which the *Arabidopsis* genome sequence could be exploited. For example a full suite of PALE genes could easily be identi-

fied in *Arabidopsis*, a task that would be much more time consuming in hybrid aspen. Searching all the hybrid aspen EST's against the *Arabidopsis* genome DNA-sequence and deduced protein complement will identify which hybrid aspen EST's have possible counterparts in *Arabidopsis*, and which genes are present in hybrid aspen but not in *Arabidopsis*. Genes with expression profiles (in hybrid aspen) indicating a certain function can be studied either in *Arabidopsis* or in hybrid aspen. All hybrid aspen genes with a possible *Arabidopsis* orthologue are preferably studied in *Arabidopsis*, using the wide variety of genetic tools available to elucidate their function. Hybrid aspen genes that do not have an *Arabidopsis* orthologue obviously have to be further studied in hybrid aspen, preferentially using techniques that could potentially be applied on a large scale like, for example, virus-induced gene silencing (Burton et al., 2000), although more standard transgene techniques could also be adopted. There is also a possibility of combining the detailed spatial expression profiles obtained from wood formation in hybrid aspen with the genomic sequence from *Arabidopsis*, in order to identify new *cis* acting DNA elements regulating developmental stage specific xylem expression.

References

Baima, S., Nobili, F., Sessa, G., Lucchetti, S., Ruberti, I. and Morelli, G. (1995) The expression of the *Athb-8* homeobox gene is restricted to provascular cells in *Arabidopsis thaliana*. *Development,* **121**, 4171-4182.

Barnett, J.R. (1981) Secondary xylem development. In *Xylem cell development*. (Barnett, J.R., ed.). Tunbridge Wells, Kent, UK: Castle house publications Ltd, pp. 47-95.

Barton, M.K. and Poethig, R.S. (1993) Formation of the shoot apical meristem in *Arabidopsis thaliana* - An analysis of development in wild type and in the *SHOOT MERISTEMLESS* mutant. *Development,* **119**, 823-831.

Baucher, M., Chabbert, B., Pilate, G., VanDoorsselaere, J., Tollier, M.T., PetitConil, M., Cornu, D., Monties, B., VanMontagu, M., Inze, D., Jouanin, L. and Boerjan, W. (1996). Red xylem and higher lignin extractability by down-regulating a cinnamyl alcohol dehydrogenase in poplar. *Plant Physiol.* **112**, 1479-1490.

Bechtold, N., Ellis, J. and Pelletier, G. (1993) In-planta *Agrobacterium*-mediated gene-transfer by infiltration of adult *Arabidopsis thaliana* plants. *G.C.R. Acad. Sci.* **316**, 1194-1199.

Berleth, T., Mattsson, J. and Hardtke, C.S. (2000). Vascular continuity and auxin signals. *Trends Plant Sci.* **5**, 387-393.

Billeter, M., Qian, Y., Otting, G., Muller, M., Gehring, W.J. and Wuthrich, K. (1990) Determination of the three-dimensional structure of the *Antennapedia* homeodomain from *Drosophila* in solution by 1H nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **5**, 183-197.

Bürglin, T.R. (1994) A comprehensive classification of homeobox genes. In *Guidebook to the homeobox genes* (Duboise, D. ed.). Oxford: Oxford University Press, pp. 27-71.

Bürglin, T.R., Finney, M., Coulson, A. and Ruvkun, G. (1989) *Caenorhabditis elegans* has scores of homoeobox-containing genes. *Nature,* **341,** 239-243.

Carland F.M. and McHale, N.A. (1996) *LOP1*: A gene involved in auxin transport and vascular patterning in *Arabidopsis*. *Development* **122**, 1811-1819.

Chory, J. (1993) Out of darkness: Mutants reveal pathways controlling light regulated development in plants. *Trends Genet.* **9**, 167-172

DeRisi, J.L., Iyer, V.R. and Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680-686.

Durfee, T., Becherer, K., Chen, P.L., Yeh, S.H., Yang, Y., Kilburn, A.E., Lee, W.H. and Elledge, S.J. (1993) The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. *Gene. Dev.* **7**, 555-69

Endrizzi, K., Moussian, B., Haecker, A., Levin, J.Z. and Laux, T. (1996) The *SHOOT MERISTEMLESS* gene is required for maintenance of undifferentiated cells in *Arabidopsis* shoot and floral meristems and acts at a different regulatory level than the meristem genes *WUSCHEL* and *ZWILLE*. *Plant J.* **10**, 967-979.

Eriksson, M.E., Israelsson, M., Olsson, O. and Moritz, T. (2000) Increased gibberellin biosynthesis in transgenic hybrid aspen trees promotes growth, biomass production and xylem fibre length. *Nat. Biotechnol.* **18**, 784-788.

Esau, K. (1977) *Anatomy of seed plants*,2nd edn. New York: John Wiley and Sons Ltd.

Fagard, M., Höfte, H., and Vernhettes, S. (2000) Cell wall mutants. *Plant Physiol. Biochem.* **38**, 15-25.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N. and Willitmitzer, L. (2000) Metabolitic profiling for plant functional genomics. Nat. Biotechnol. **18**, 1157-1161.

Foster, G. and Twell, D. (1996) Plant Gene Isolation: Principles and Practise. (Foster, G. and Twell, D., eds.). Chichester, UK: John Wiley and Sons Ltd.

Franklin, (1945). *Nature*.

Fukuda, H. (1997) Tracheary element differentiation. *Plant Cell* **9**, 1147-1156.

Galweiler, L., Guan, C.H., Muller, A., Wisman, E., Mendgen, K., Yephremov, A. and Palme, K. (1998) Regulation of polar auxin transport by AtPIN1 in *Arabidopsis* vascular tissue. *Science* **282**, 2226-2230.

Gehring, W., Affolter, M. and Bürglin, T. (1994) Homeodomain proteins. *Annu. Rev. Biochem.* **63**, 487-526.

Hamann, T., Mayer, U. and Jurgens, G. (1999) The auxin-insensitive *bodenlos* mutation affects primary root formation and apical-basal patterning in the *Arabidopsis* embryo. *Development* **126**, 1387-1395.

Hardtke, C.S. and Berleth, T. (1998) The *Arabidopsis* gene *MONOPTEROS* encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J.* **17**, 1405-1411.

Higushi, T. (1997) Biochemistry and molecular biology of wood. New York: Springer Verlag,.

Hobbie, L., McGovern, M., Hurwitz, L.R., Pierro, A., Liu, N.Y., Bandyopadhyay, A. and Estelle, M. (2000) The *axr6* mutants of *Arabidopsis thaliana* define a gene involved in auxin response and early development. *Development* **127**, 23-32.

Hu, W.J., Harding, S.A., Lung, J., Popko, J.L., Ralph, J., Stokke, D.D., Tsai, C.J., and Chiang, V.L. (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nat. Biotechnol.* **17**, 808-812.

Klee, H.J., Horsch, R.B., Hinchee, M.A., Hein, M.B. and Hoffman, N.L. (1987) The effects of overproduction of two *Agrobacterium tumefaciens* T-DNA biosynthetic gene products in transgenic petunia plants. *Gene. Dev.* **1**, 86-96.

Koizumi, K., Sugiyama, M. and Fukuda, H. (2000) A series of novel mutants of *Arabidopsis thaliana* that are defective in the formation of continous vascular network: calling the auxin signal flow canalisation hypothesis into question. *Development* **127**, 3197-3204.

Koncz, C. and Schell, J. (1986) The promoter of T_L-DNA gene 5 controls the tissue-specific expression of chimeric genes carried by a new type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* **204**, 383-396.

Larson, P.R. (1994) The vascular cambium: Development and structure. Berlin: Springer Verlag.

Lawrence, P.A. and Morata, G. (1994) Homeobox genes: Their function in *Drosophila* segmentation and pattern formation. *Cell,* **78**, 181-189.

Lev Yadun, S. Induction of sclerid differentiation in the pith of *Arabidopsis thaliana* (L.) Heynh. *J. Exp. Bot.* **45**, 1845-1449.

Liang, P. and Pardee, A.B. (1992) Differential display of eukaryotic messenger-RNA by means of the polymerase chain reaction. *Science* **157**, 967-971.

Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R. and Lockhart, D.J. (1999) High density synthetic oligonucleotide arrays. *Nat. Genet.* **21**, 21-24. Suppl.

Little, C.H.A., MacDonald, J.E. and Olsson, O. (2000) Involvement of indole-3-acetic acid in fascicular and interfascicular cambial growth and interfascicular extraxylary fiber differentiation in *Arabidopsis thaliana* inflorescence stems. Submitted.

Little, C.H.A. and Pharis, R.P. (1995) Hormonal control of radial and longitudinal growth in tree stems. In *Plant stems: Physiology and functional morphology* (Gartner, B.L., ed.). San Diego, CA, USA: Academic Press, pp. 281-319.

Long, J.A., Moan, E.I., Medford, J.I., and Barton, M.K. (1996) A member of the *KNOTTED* class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. *Nature,* **379,** 66-69.

Magel E.A. (2000) Biochemistry and physiology of heartwood formation. In *Cell and Molecular Biology of Wood Formation*. (Savidge, R.A., Barnett, J.R. and Napier, R., eds.). Oxford, UK: BIOS Scientific Publishers Ltd,

Masucci, J.D., Rerie, W.G., Foreman, D.R., Zhang, M., Galway, M.E., Marks, M.D., and Schiefelbein, J.W. (1996) The homeobox gene *GLABRA2* is required for position-dependent cell differentiation in the root epidermis of *Arabidopsis thaliana*. *Development* **122**, 1253-1260.

Mauseth, J.D. (1988) *Plant anatomy*. Menlo Park, California, USA: The Benjamin/Cummings Publishing Company.

Mayer, U., Buttner, G. and Jurgens, G. (1993) Apical-basal pattern-formation in the *Arabidopsis* embryo- Studies on the role of the *GNOM* gene. *Development* **117**, 149-162.

Meijer, A.H., Scarpella, E., vanDijk, E.L., Qin, L., Taal, A.J.C., Rueb, S., Harrington, S.E., McCouch, S.R., Schilperoort, R.A. and Hoge, J.H.C. (1997) Transcriptional repression by Oshox1, a novel homeodomain leucine zipper protein from rice. *Plant J.* **11**, 263-276.

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

Meyer, K., Cusumano, J.C., Somerville, C. and Chapple, C.C.S. (1996) Ferulate-5-hydroxylase from *Arabidopsis thaliana* defines a new family of cytochrome P450-dependent monooxygenases. *Proc. Natl. Acad. Sci. USA* **93**, 6869-6874.

Nilsson, O., Aldén, T., Sitbon, F., Little, C.H.A., Chalupa, V., Sandberg, G. and Olsson, O. (1992) Spatial pattern of cauliflower mosaic virus 35S promoter-luciferase expression in transgenic hybrid aspen trees monitored by enzymatic assay and non-destructive imaging. *Transgenic Res.* **1**, 209-220.

Otting, G., Qian, Y.Q., Billeter, M., Muller, M., Affolter, M., Gehring, W.J. and Wuthrich, K. (1990) Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* **9**, 3085-3092.

Pandey, A. and Mann, M. (2000) Proteomics to study genes and genomes. *Nature* **404**, 837-845.

Pennel, R.I. and Lamb, C. (1997) Programmed cell death in plants. *Plant Cell,* **9**, 1157-1168.

Pickett-Heaps, J. D. (1968). Xylem wall deposition: Radioautographic investigations using lignin precursors. *Protoplasma* **65**, 181-205.

Regan, S., Chaffey, N. and Sundberg, B. (1999) Exploring cambial growth with *Arabidopsis* and *Populus*. *J. Exp. Bot.* **50**, 33. Suppl.

Rerie, W.G., Feldmann, K.A. and Marks, M.D. (1994) The *GLABRA2* gene encodes a homeodomain protein required for normal trichome development in *Arabidopsis*. *Genes Dev.* **8**, 1388-1399.

Roberts, K., and McCann, M.C. (2000) Xylogenesis: The birth of a corpse. *Curr. Opin. Plant Biol.* **3**, 517-522.

Romano, C.P., Hein, M.B. and Klee, H.J. (1991) Inactivation of auxin in tobacco transformed with the indoleacatic acid-lysine synthase gene of *Psudomonas savastanoi*. *Gene. Dev.* **5**, 438-446.

Ruberti, I., Sessa, G., Lucchetti, S. and Morelli G (1991) A novel class of plant proteins containing a homeodomain with a closely linked leucine zipper motif. *EMBO J.* **10**, 1787-1791.

Savidge, A.S. (1996) Xylogenesis, generic and environmental regulation- A review. *IAWA Journal*, **17**, 269-310.

Scarpella, E., Rueb, S., Boot, K.J.M., Hoge, J.H.C. and Meijer, A.H. (2000) A role for the rice homeobox gene *Oshox1* in provascular cell fate commitment. *Development* **127**, 3655-3669.

Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray. *Science* **270**, 467-470.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O. and Davis R.W. (1996) Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci USA* **93**, 10614-10619.

Schena, M. and Davis, R.W. (1992) HD-Zip proteins: members of an *Arabidopsis* homeodomain protein superfamily *Proc. Natl. Acad. Sci. USA* **89**, 3894-3898.

Schiefelbein, J.D., Masucci, J.D. and Wang, H. (1997) Building a root: The control of patterning and morphogenesis during root development. *Plant Cell,* **9,** 1089-1098.

Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P. and Bork, P. (2000) SMART: a web-based tool for the study of genetically mobile domains. *Proc. Natl. Acad. Sci. USA* **28**, 231-234

Scott, M.P., Tamkun J.W. and Hartzell G.W. (1989) The structure and function of the homeodomain. *Biochim. Biophys. Acta* **989**, 25-48.

Sitbon, F., Hennion, S., Sundberg, B., Little, C.A.H., Olsson, O. and Sandberg, G. (1992) Transgenic Tobacco Plants Coexpressing the Agrobacterium tumefaciens iaaM and iaaH genes display altered growth and Indoleacetic acid metabolism. Plant Physiol. **99**, 1062-1069.

Steindler, C., Matteucci, A., Sessa, G., Weimar, T., Ohgishi, M., Aoyama, T., Morelli, G. and Ruberti, I. (1999) Shade avoidance responses are mediated by the ATHB-2 HD-Zip protein, a negative regulator of gene expression. *Development* **126**, 4235-4245.

Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rhode, A., Holmberg, A., Amini, A., Bhalerao, R., Larsson, M., Villarroel, R., Montagu, M., Sandberg, G., Olsson, O., Terri, T. T., Boerjan, W., Gustafsson, P., Uhlen, M., Sundberg, B. and Lundeberg, J. (1998) Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags. *Proc. Natl. Acad. Sci. USA* **95**,13330-13335.

Sundberg, B., Uggla, C. and Tuominen, H. (2000). Cambial growth and auxin gradients. In *Cell and Molecular Biology of Wood Formation* (Savidge, R.A., Barnett, J.R. and Napier, R., eds.) Oxford, UK: BIOS Scientific Publishers Ltd.

Supplement to *Nat. Genet.* 22 no1. (1999).

Taylor, N.G., Scheible, W.R., Cutler, S., Somerville, C.R. and Turner, SR. (1999). The *irregular xylem3* locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. *Plant Cell* **11**, 769-779.

The Arabidopsis genome initiative: Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**, 796-815 (2000).

Tuominen, H., Sitbon, F., Jacobsson, C., Sandberg, G., Olsson, O. and Sundberg, B. (1995) Altered growth and wood characteristics in transgenic hybrid aspen expressing *Agrobacterium tumefaciens* T-DNA indoleacetic acid-biosynthetic genes. *Plant Physiol* **109**,1179-1189.

Turner, S.R. and Somerville, C.R. (1997) Collapsed xylem phenotype of Arabidopsis identifies mutants deficient in cellulose deposition in the secondary cell wall. *Plant Cell* **9**, 869-701.

Uggla, C., Moritz, T., Sandberg, G. and Sundberg, B. (1996) Auxin as a positional signal in pattern formation in plants. *Proc. Natl. Acad. Sci. USA,* **93**, 9282-9286.

Wilson, B.F., Wodzicki, T.J. and Zahner, R. (1966) Differentiation of cambial derivates: Proposed terminology. *Forest Science,* **12**, 438-440.

Wisman, E. and Ohlrogge, J. (2000). *Arabidopsis* microarray service facilities. *Plant Physiol.* **127**, 1468-1471.

Zhong, R.Q., Taylor, J.J. and Ye, Z.H. (1997) Disruption of interfascicular fiber differentiation in an *Arabidopsis* mutant. *Plant Cell* **9**, 2159-2170.

Zhong, R.Q. and Ye, Z.H. (1999) *IFL1*, a gene regulating interfascicular fiber differentiation in Arabidopsis, encodes a homeodomain-leucine zipper protein. *Plant Cell* **11**, 2139-2152.

Zhu, T. and Wang, X. (2000). Large-scale profiling of the *Arabidopsis* transcriptome. *Plant Physiol.* **124**, 1472-1476.

Acknowledgements

First I want to thank my supervisors, Olof Olsson and Göran Sandberg for support, freedom and opportunities given under these years. I am sincerely grateful. Although, we of course have had the innebandy situation!

I also want to thank all co-authors, that are too many, to all be mentioned by name. However I will put a special thanks to Björn Sundberg, Peter Nilsson and Joakim Lundeberg, it has been nice working with you.

I want to thank all present and former members of the lab, it has been a nice time. I also thank all the non-lab-working members of the department for help with various things and for being nice work companions. It has not only been work, so I give an extra thank-you to all party-animals that have made a number of parties end really late.

Anneli S, I thank for the long time I known her, will we ever be separated?

I thank Rishi B and Ove N for valuable discussions and comments on science and other matters.

Maria puts an extra thanks to Crister Å., David M. and Stefan N, that from time to time have taking me out training.

On the other hand I want to thank Hannele T and later Ove N for almost daily joining me for lunch. Hannele, I hope they have PLOPP in Helsinki.

A special thanks, to all making the out-of-lab-experience present. As an old university friend said: working is not life. That's a god one. You should all remember that.

A thanks to my parents and sister for always being my family.

And finally I thank Maria for all the years that have been shared and all the years to come.