Out of the woods: forest biotechnology enters the genomic era
Rishikesh Bhalerao, Ove Nilsson and Goran Sandberg*

Trees represent a unique life form of utmost importance for mankind, as these organisms have developed a perennial lifestyle that produces the majority of terrestrial biomass. The difference between trees and annual plants is one of the main arguments behind the effort to sequence the entire genome of the poplar tree. This initiative is being backed up with a full-scale functional genomics effort on trees that will set a completely new agenda for forest research.

Introduction
Trees constitute the majority of terrestrial biomass production and the forestry and wood-processing industries that depend upon them are major contributors to the economies of the industrialised world. The human population is still growing dramatically, increasing the demand for products derived from trees. In addition, trends in developed countries to reduce their use of non-renewable resources and to conserve the environment are likely to increase requirements for products generated from renewable resources such as wood. Meeting the increased demands for forestry products is likely to require increased forest productivity and more intensive research to create novel products from wood. To date, research related to the forest industry has mainly concentrated on forest management and timber, pulp and paper technology. By contrast, there has been little emphasis on exploiting advances in biotechnology, either to enhance productivity or to generate novel products from renewable resources such as wood. Therefore, a challenge for the future is to ensure that the forest industry benefits from rapidly developing biotechnological advances. In particular, forest biotechnology could derive enormous advantages from information generated through functional genomics approaches. The aims of this review are to summarise the current status of genomics in forest trees, to consider potential uses of genomics in novel areas of biotechnological research, and to identify concerns likely to arise from the rapid anticipated growth of forest biotechnology industries.

Gene discovery
Forest genomics began when expressed sequence tag (EST) projects were initiated in pine and poplar, after the pioneering work of Venter and colleagues [1] had proven the value of EST sequencing as a cheap but efficient method of finding genes. Given the unique biology of trees, it was felt that gene discovery should be prioritised to identify putative tree-specific genes. These initial efforts based on poplar and pine were modest in scope: the first publications reported about 5000 [2] and 1000 EST sequences for the respective species [3]. Since then, over 80 000 ESTs have been sequenced from pine (www.pinetree.ccgb.umn.edu) and over 130 000 from poplar (www.poppel.fysbot.umu.se). Furthermore, nearly 80 000 ESTs have been sequenced from birch (T Palva, personal communication) and a spruce sequencing project has recently been initiated. In addition to these academic efforts, impressive EST projects based on radiata pine and eucalyptus have been undertaken and reported by industrial laboratories. Unlike earlier activities, where the objective was simply to identify the main sequences expressed in the species being considered, more recent efforts have focused on the creation and comparison of multiple cDNA libraries. These libraries were made from RNA isolated from a variety of tissues and from plants either in various developmental stages or subjected to different treatments. In each case, single pass 5’ sequencing has resulted in a large number of sequences with no apparent homologues in databases derived from other types of plant. However, subsequent analysis has indicated that the high number of such sequences is a gross overestimate of the genes unique to trees, as the number decreases when the sequence length of ESTs is extended beyond the usual 400–500 base pairs (bp) to ~800 bp [4]. Nevertheless, as a result of these sequencing efforts we now know considerably more about the genetic composition of trees than we did previously. In addition, using these EST resources we may be able to elucidate the genetic basis for the great differences in wood quality observed between gymnosperms and angiosperms.

Transcript profiling
It was recognised from an early stage that mere gene discovery from EST analysis would not allow the full
potential of the genomics approach to forestry research to be realized. It was also necessary to obtain a better understanding of the expression patterns of different genes, as EST frequency gives rather rudimentary data on gene expression. Furthermore, it is often far more interesting to obtain temporal and spatial descriptions of the expression patterns, and this type of information is especially important for biotechnological purposes. For instance, it may be desirable to modify a metabolic pathway in a small subset of cells. Thus, a platform for global transcript profiling was established in poplar using the first set of EST sequences. The first microarray slide contained about 2500 features and was used to investigate the molecular basis of xylem development [5**]. In addition, a new amplification technique was developed, allowing RNA to be isolated from submilligram amounts of tissues to generate probes for microarray analysis [6**]. Since then, a further generation of microarray slides has been produced with over 13 500 features representing 33 000 ESTs; a further generation is planned to accommodate over 20 000 features derived from 100 000 sequenced ESTs.

**Genomic sequencing**

Although EST sequencing is a cheap and quick way to identify expressed genes, we also need to know the complete genomic sequence of one or more tree species. The genomic sequence is necessary for several reasons. Firstly, it is highly unlikely that all the genes of any tree will be identified by EST sequencing alone. Secondly, even if there are several hundred genes unique to trees, it would be extremely useful to identify their individual contributions to the observed architectural and other differences between simple annual weeds like Arabidopsis and trees. Thirdly, acquisition of a full tree genome sequence would be very valuable for quantitative trait locus (QTL) analysis, marker-assisted breeding and, importantly, the genome sequence from one tree could be used as a platform for identifying synteny between tree species, as has been done for Arabidopsis and Brassica and other species. Therefore, the news that the United States Department of Energy (DOE) has decided to sequence the genome of poplar is welcome to all tree biologists (www.ornl.gov/ipgc). The DOE’s Joint Genome Institute (www.jgi.doe.gov) is expected to produce six times coverage of the entire genome sequence during 2003. However, without further advances in sequencing and bioinformatics, it seems unlikely that we will obtain genomic information from any gymnosperm in the near future. This is because gymnosperms have a massive genome with haploid DNA contents of, on average, 15 500 Mbp [7], as compared with 125 Mbp for Arabidopsis [8] and 550 Mbp for poplar [9].

**Genetic mapping**

Genetic maps of varying quality have been generated for several forest tree species, using a variety of approaches. QTLs have been identified for a range of traits, such as wood density, fibre length and resistance [10–13,14**,15]. However, the gains from QTL mapping for breeding purposes have been severely restricted by the great difficulties in identifying the gene (or genes) located at the QTL. There are two reasons for these difficulties. Firstly, the long time to flowering hampers repeated...
crossings to generate isogenic lines and, secondly, without access to large numbers of gene sequences and microarrays, candidate gene selection is not really possible. However, recent advances in sequencing, especially genomic sequencing in poplar, and transcript profiling should considerably enhance QTL analyses.

The use of model systems in forest biotechnology
It is apparent from work in Caenorhabditis elegans, Arabidopsis and yeast that development of a suitable model system provides a great tool for understanding complex biological processes. Similarly, poplar has been developed as an important model system for tree molecular genetics. The major advantages of poplar are its ease of transformation, relatively small genome size and rapid growth. Moreover, in addition to the EST and genomic sequencing efforts described above, several other resources are already, or soon will be, available for this species. These include a BAC (bacterial artificial chromosome) physical map (currently with 46,000 BAC clones fingerprinted, 80% assembled into contigs and 92,000 BAC end sequences), a completed chloroplast DNA sequence, and several hundred mapped simple sequence repeats (SSR) markers [16]. Global transcript profiling and other genomics technologies are also being used in poplar studies and effective forward and reverse genetic screens are being designed. Collectively, these resources will make poplar not only the model for tree biology, but will also provide an excellent model for many unique aspects of plant biology associated with perenniality and developmental phase changes, adaptation to harsh climates, herbivory, secondary growth and secondary metabolism (Figure 1).

A general strategy for combining poplar and Arabidopsis
While great strides have been taken towards developing poplar as a model system for tree biology, the value of Arabidopsis as a test bed for studying ‘tree’ genes and their functions should not be underestimated. For example, Arabidopsis undergoes secondary growth (Figure 2) [17,18] and can therefore be used to identify important regulatory genes involved in wood formation, potentially the most important area from the perspective of forest biotechnology. To illustrate the opportunities presented by this approach, preliminary phenotypic characterisation of an Arabidopsis line with a T-DNA (transfer DNA) tagged gene homologous to a cambially expressed poplar gene was carried out. This work enabled the gene’s function in poplar to be rapidly elucidated, as shown in Figures 3 and 4. This example highlights the potential use of Arabidopsis to rapidly identify the function of poplar genes. As more and more information is derived from exploratory experiments such as transcript profiling in poplar and other tree species, the scope for genetic analysis will be limited in trees. However, we believe that the close link between the two models will make it possible to identify target genes in Populus and then isolate the corresponding knockouts in Arabidopsis for genetic and functional analysis.

From genomics to biotechnology
From the above discussion it is clear that forestry has already entered the genomics era, but what benefits will
these genomics initiatives bring to forest biotechnology? To date, the major focus of research in forest biotechnology has been to modify lignin and/or cellulose contents. Numerous reviews have already been written on the advances made in these areas [19]. For example, ways in which lignin levels can be modified and the consequences for wood and tree growth have been intensively studied [20–22]. Thus, as a large body of data covering these aspects has already been amassed and discussed, we will focus here on areas of forestry research that have been neglected so far, but which are expected to benefit from recent advances in forest genomics.

Wood formation

As the demand for products generated from wood increases, it will be necessary to increase production of this raw material [23]. To date, the forest industry has sought to boost productivity mainly at production sites in tropical countries where high growth rates can be achieved. However, genomics research related to the control of cambial activity, which underlies wood production, may provide alternative approaches to enhance productivity. For example, it has been shown that overexpressing phytochrome can prevent growth cessation [2]. This phenomenon is well known and is exploited in practical breeding programmes, but the findings presented by Olsen et al. [24] demonstrate the potential for altering tree growth in trees of temperate regions by biotechnological means. Similarly, several reports have indicated that growth control and wood formation in trees can be biotechnologically modified by altering their hormone content (e.g. auxins and gibberellins), and/or by modifying hormonal signalling systems, which regulate primary plant development [25–28]. For example, overexpression of a gene encoding a protein involved in the biosynthesis of gibberellins, GA20 oxidase, increases the growth rate and enhances cell division in hybrid aspen trees [27]. In the cited examples, genes that play important roles in the regulation of diverse pathways were targeted, but for practical purposes it may be essential to identify genes that alter growth in a very specific manner. This is the kind of application for which genomic techniques such as transcript profiling and EST sequencing of tissue-specific libraries will be highly useful, notably for identifying candidate genes to alter cambial activity.

Figure 3

Example of using transcript expression data from poplar combined with studies in Arabidopsis. (a) Vascular expression pattern of Scarecrow-family homologue (SCR; Scarecrow encodes a putative transcription factor that has been implicated in cell fate specification in Arabidopsis root) from poplar (see Figure 4 for specifications of tissues A–E). (b) Vascular expression pattern of SCR:GUS in primary leaves of wild-type Arabidopsis. (c) Expression of RGA-like promoter trap insertion in primary leaves of Arabidopsis. (d) Cleared Arabidopsis primary leaf illustrating wild-type venation pattern. (e) Nomarski image of wild-type primary leaf venation. (f) Cleared shr (shortroot) primary leaf showing loss of higher order venation. (g) Nomarski image of shr primary leaf venation showing breaks in the vascular strands. (Data courtesy of Alan Marchant and Malcolm Bennet.)
Engineering wood quality

Wood formation is a process that can be divided into a series of well-defined developmental events that are initiated in the vascular cambium [29]. Cambial derivatives develop into xylem cells through the processes of division, expansion, secondary wall formation, lignification and, finally, programmed cell death (Figure 4). These processes are strongly interlinked; thus, modulation of any one aspect of wood formation may affect many other aspects, and any strategy designed to engineer wood quality will have to take full account of such complications. Therefore, wood engineering almost necessitates the use of genomics, as genomic approaches can provide information on the regulation of not just one gene or enzyme, but an entire pathway or several pathways at the same time. Recent microarray experiments by Hertzberg et al. [6] demonstrated this point by providing expression profiles of over 2300 genes across the developmental gradient during wood formation (Figure 5). These experiments not only clearly indicate the complexity that wood engineers will have to deal with, but also provide tempting glimpses into how regulators of specific aspects of wood development may be identified for modulation of wood properties.

Meristem identity and reproductive control

All above-ground portions of a plant are ultimately derived from the activity of the shoot apical meristem (SAM). This meristem produces shoot, leaf and flower primordia on its flanks, and also produces the primary vascular meristems. Later, the vascular cambium, responsible for the secondary growth of the stem, is formed from these primary meristems. Genetic analysis, mainly in Arabidopsis, has identified a large array of genes that appear to help determine the identity of the primordia and the SAM itself. However, owing to the limited extent of secondary growth in Arabidopsis we still know very little about the genes regulating the growth of the vascular meristems. We know even less about the regulation of meristem function in trees, which show alternating cycles of growth and dormancy, as well as alternating periods of vegetative and reproductive growth. Coupled to our very poor knowledge of the regulation of the juvenility-to-maturity transition in trees, which also affects the competence of different meristems to respond to developmental signals, we clearly need much more research on these 'tree-specific' aspects of meristem function. It is equally clear that these tree meristems ultimately control the growth and production of the tree, and are therefore of obvious interest to the forest industry.

An important and apparently unique aspect of tree development is the late juvenility-to-maturity transition, which makes trees flower later than any other known plants. In Populus, the time to first flowering can take up to 20 years. During the juvenile phase the tree is not competent to induce flowering. The reason for this is completely unknown, but is presumably due to a complex interaction
between plant growth regulators (e.g. gibberellins), floral repressors and epigenetic factors such as alterations in methylation status and/or chromatin structure. It is not clear whether this particular aspect of the regulation of tree flowering has a counterpart in annual plants such as *Arabidopsis*. Although genes regulating flower meristem identity appear to have conserved functions between *Arabidopsis* and trees [30,31] we still lack evidence demonstrating that any genes normally involved in regulating *Arabidopsis* flowering time also have a function in trees. From a forest biotechnology perspective, there is considerable interest in the development of technologies to control flowering, for two fundamentally different reasons. The possibility to induce early flowering in trees would dramatically enhance the speed of breeding programmes, especially in combination with marker-assisted breeding. This could, for the first time, enable tree breeders to achieve similar, dramatic gains in productivity and quality to those already achieved by breeders of agricultural crops. However, there is also a general consensus within the forest biotechnology community that public and regulatory acceptance to widespread plantations of genetically modified trees will require the development of techniques that prevent the flowering of such trees [32]. One way of achieving this is to express cytotoxic genes from flower-specific promoters. For this purpose it is important to characterise flower-specific promoters from trees and isolate the corresponding promoters [33–40]. One major problem, however, is the subsequent characterisation of the isolated promoters and sterility constructs. This cannot be done until the transgenic plants expressing the constructs are flowering, which can take decades. One possible way of circumventing this is to test the sterility constructs in transgenic trees that have been engineered to flower early using the techniques described above. Alternatively, they could be tested in naturally early-flowering variants of otherwise late-flowering trees [41].

Transcript profiling of the wood-forming tissue presented as a hierarchical cluster analysis of 1791 selected genes with differential expression in the sampled tissues. The colour scale at the bottom depicts fold change between samples. I–VIII: groups of genes with different differential expression patterns, the expression ratios are given in log2 scale. The different samples represent the tissues marked with black bars in Figure 3. (The figure was reproduced from [5]. Copyright 2001 National Academy of Sciences, USA.)
This represents another important challenge for forest biotechnologists.

Conclusions

To summarise, the era of functional genomics has come to the forest sector with several EST sequencing projects being initiated in a range of forest trees. Even more exciting is the complete sequencing of the genome of the model tree poplar. Thus, the stage is set for the next level of analysis, namely elucidation of gene function in forest trees. This analysis will be accelerated by using several approaches in parallel. In this respect, global transcript profiling is already being used in poplar. Another exciting strategy to rapidly assess gene function is suggested by promising experiments that used data from transcript profiling in poplar. The data were used to identify candidate genes whose function is now being investigated at a rapid pace by analyzing the phenotypes of mutants in the corresponding Arabidopsis orthologues. One area that needs to be targeted, and which will be important in the future, is the integration of information obtained from the use of different profiling technologies.

In conclusion, in the coming years, both academic research and the forest biotechnology industry are set to benefit from the advent of functional genomics in forest trees.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

• of special interest
** of outstanding interest


A landmark paper describing the application of microarrays to tree species to chart the development of xylem formation.


This paper describes a breakthrough technology that allows RNA to be isolated from submilligram amounts of tissue for the purpose of probe preparation in microarray experiments.


One of the most detailed map descriptions of poplar derived using two different markers. Interestingly, mapping was performed on the species of poplar that will be sequenced in the near future.


