Biotechnology techniques

Richard Meilan, Zhonglian Huang, Gilles G. Pilate

2. Biotechnology techniques

R. Meilan, Z. Huang and G. Pilate

Biotechnology can be divided into two broad areas: conventional breeding and molecular genetics. The former has been used for centuries to improve plant and animal species to satisfy human needs. Advances in molecular genetics have been rapidly adopted by the scientific community over the last two decades, and they complement tools already available to conventional breeders.

Molecular genetics can itself be subdivided into two distinct categories. In the first, which could be called 'non-controversial technologies', the plant genome is not altered. This category comprises molecular markers, which are used for DNA fingerprinting and MAS (e.g. QTL mapping and association genetics); sequence analysis (genomic DNA, cDNA libraries [ESTs], and bacterial artificial chromosome [BAC] clones), which aid in gene discovery; and *in vitro* propagation (e.g. somatic embryogenesis). The benefits of research using these technologies are increased genetic gain per generation through improved selection in conventional breeding programmes, faster deployment of genetically improved material to plantations, and a deeper understanding of the genes controlling commercially important traits.

The second major subdivision of molecular genetics, termed 'controversial technologies', includes recombinant DNA and gene-transfer techniques. These are the basis for genetic engineering, which is defined as the stable, usually heritable, modification of an organism's genetic makeup via asexual gene transfer, regardless of the origin and nature of the introduced gene. The product of this process is generally referred to as a genetically modified organism (GMO). Genetic engineering offers the opportunity to add new genes to existing, elite genotypes. Although much progress has been made, genetically engineered forest species are not likely to be deployed commercially in much of the world for several more years. One reason for this delay is our limited understanding of the key genes that contribute to the control of commercially important traits, such as wood properties, flowering control and pest resistance. Research in these areas will broaden our knowledge of the genetic and physiological mechanisms that govern tree growth and development. In addition, it will allow the assessment of risks associated with these controversial technologies–assessments that will be required if we are to produce genetically improved material for meeting the growing societal demands for high-quality wood and fibre (Farnum, Lucier and Meilan, 2007).

To make more rapid progress with tree biotechnology, certain innovations are needed, including improved regeneration protocols, alternative *in vitro* selection strategies, dependable excision mechanisms and reliable confinement strategies. One limitation is in our understanding of the roles played by genes controlling key aspects of tree development. Poplar is widely accepted as the model tree for

forest biology owing to its small genome, expanding molecular resources, fast growth, and the relative ease with which it can be clonally propagated *ex vitro* and transformed and regenerated *in vitro* (Bradshaw *et al.*, 2000; Wullschleger, Jansson and Taylor, 2002). The recently released *Populus trichocarpa* genome sequence (Tuskan *et al.*, 2006) and newly developed genomics approaches have already and will continue to expedite gene discovery. The knowledge gained through our work with poplar can then be applied to other tree species.

TECHNIQUES

Recombinant DNA

The application of a variety of techniques collectively referred to as 'recombinant DNA technology' permits the study of gene structure and function, gene transfer to various species, and the efficient expression of their products. Using microbiological methods, it is possible to combine genetic material from various organisms in novel ways. Through these techniques it has been possible to expand our knowledge concerning the way in which genes are regulated, eukaryotes synthesize proteins, and eukaryotic genomes are organized. With regard to genetic engineering, recombinant DNA techniques are essential for:

- identifying genes responsible for specific traits;
- isolating these genes;
- creating genetic constructs harbouring both these genes and flanking regulatory sequences needed for expression in the host organism (in our case, a tree);
- selecting transgenic cells (generally by using an antibiotic or herbicide resistance gene).

Once genetically modified plants have been produced, this technology also allows us to select the best individuals with preferred levels of integration and expression and to monitor, at the molecular level, whether transgene integration and expression are maintained from one growing season to the next, after sexual reproduction, and in various environments.

Transformation

The main steps required for the production of GMOs are:

- stably introducing a novel piece of DNA into the genome of a cell (i.e. transformation);
- isolating transgenic plant cells on a medium containing a selection agent (e.g. the antibiotic or herbicide against which the selectable marker gene imparts resistance);
- regenerating whole plants from the transformed cells through *in vitro* culture;
- screening various transgenic lines that result from independent transformation events on the basis of insert copy number and configuration, and expression.

To date, much of the research on genetic engineering of trees has concentrated on optimizing transformation. Three gene-transfer techniques are commonly

utilized here: protoplast transformation, biolistics and *Agrobacterium*-mediated transformation. Historically, angiosperms were transformed primarily through the use of *Agrobacterium tumefaciens*. Because of early difficulties encountered when transforming conifers with common *Agrobacterium* strains, gymnosperms were initially transformed using particle bombardment (Pena and Seguin, 2001). These problems have now largely been resolved, and several different species are being efficiently transformed via standard *Agrobacterium* strains (e.g. Pilate *et al.*, 1999; Tang, Newton and Weidner, 2007; Tereso *et al.*, 2006). However, except for larch (*Larix kaempferi* × *L. decidua*) (Levee *et al.*, 1997), much work remains to be done on the other steps leading to the production of genetically modified trees, particularly with regard to the regeneration of whole plants from transgenic cells. Plants are regenerated through one of two methods: direct organogenesis or somatic embryogenesis. The latter leads to the production of embryos from somatic tissues, whereas the former involves the generation of organs, such as shoots and roots, from various mature tissues or undifferentiated cell masses derived therefrom. No matter which approach is used, *in vitro* regeneration is often a genotype-dependent process.

Protoplast transformation

Protoplasts are derived by enzymatically digesting the walls of plant cells that are usually isolated from the leaf mesophyll, and are often grown in a liquid suspension culture. Frequently, protoplasts can be transformed either by direct DNA uptake, following polyethylene glycol pre-treatment, or by electroporation. Although many studies have resulted in successful transient expression of a transgene in cell-derived protoplasts (Bekkaoui, Tautorus and Dunstan, 1995), very few have described the regeneration of transgenic trees (e.g. Chupeau, Pautot and Chupeau, 1994). This is probably due to difficulties in regenerating whole plants from protoplasts.

Biolistics

Particle bombardment relies on the delivery of DNA-coated tungsten or gold microprojectiles, which are accelerated variously by ignited gunpowder, compressed gases (helium, nitrogen or carbon dioxide) or electrical discharge (Hansen and Wright, 1999). Although this technique was used to produce some of the first transgenic plants from recalcitrant coniferous or monocotyledonous species (Klein *et al.*, 1988; Ellis *et al.*, 1993), such transformation efficiency remains generally low, and usually results in a high number of transgene inserts in the genome. For these reasons, direct DNA transfer techniques have been avoided in favour of *Agrobacterium*-mediated protocols.

Agrobacterium-mediated transformation.

Agrobacterium tumefaciens is a soil-borne bacterium responsible for crown gall, a disease of dicotyledonous plants that causes chaotic cell proliferation at the infection site, ultimately leading to the development of a plant tumour. During

the complex infection process, bacterial DNA is stably incorporated into the plant genome. Today *A. tumefaciens* co-cultivation is the most widely used and preferred method for transforming many types of plants (reviewed by Gelvin, 2003).

A. tumefaciens harbours a large, tumour-inducing (Ti) plasmid, which encodes several products needed to transfer a piece of its DNA into the host-plant genome. This transferred sequence, called T-DNA, contains a region delimited by two borders, and carries genes that are responsible for tumour development and for the synthesis of opines (molecules that serve as a carbon and nitrogen source for the bacterium, and which result from an association between amino acids and sugars. The virulence genes (*Vir*), located outside the T-DNA region on the Ti plasmid, facilitate T-DNA transfer.

This naturally occurring mechanism for DNA transfer has been exploited by plant biotechnologists, who have demonstrated that the bacterium recognizes the DNA to be transferred to the plant cell genome by its unique borders. An *A. tumefaciens* strain is said to be disarmed when the genes within those T-DNA borders are removed. Another plasmid, a binary vector that contains the genes of interest between the border sequences, is then transformed into the disarmed strain of *A. tumefaciens*. The *Vir* genes located on the disarmed vector are able to act in trans.

The transfer of T-DNA into the host-plant genome takes place following the co-cultivation of explants (generally leaf disks, petioles, stem internodes or root segments) with the bacterium. The explants are then extensively washed to remove excess bacterium before being maintained on media containing bacteriostatins (e.g. cefotaxime or timentin) and the appropriate selection agent. Transgenic cells are multiplied then transferred to a series of media that have been optimized to contain the proper amounts of nutrients and plant growth regulators so that the various phases of plant regeneration are induced through either somatic embryogenesis or organogenesis.

The first genetically modified tree, a poplar, was produced 20 years ago (Fillatti *et al.*, 1987). Today, the number of forest tree species for which transformation and regeneration techniques have been optimized remains low; they include aspen, cottonwood, eucalyptus and walnut. Recently, transformation and regeneration protocols have been developed for several gymnosperms, mostly species within the genera *Pinus*, *Larix* and *Picea*. Within each of these species, only a few genotypes have been amenable to the recovery of transgenic plants. In general, for a wide range of genotypes, effective plant regeneration has been more difficult to achieve through organogenesis than through somatic embryogenesis.

Transgene type and its control

A gene comprises a coding sequence that is preceded by a promoter, which controls where, when and to what extent it will be expressed in a plant. This coding sequence might originate from a different species and therefore may not be present in the host plant. For example, Bt genes, which confer resistance to insects, are derived from a bacterium, *Bacillus thuringiensis*. Alternatively, the

transgene may already exist in the host plant (i.e. an endogene). For example, ferulate-5-hydroxylase (F5H) is an enzyme specific for the synthesis of syringyl lignins; homologues of this gene are found in angiosperm trees. In general, foreign genes are relatively easy to express in the host plant. Depending on the configuration of the genetic construct (e.g. the orientation of the coding sequence or the occurrence of an inverted repeat), expression of the introduced gene may be ectopic (e.g. expressed in a tissue or at a stage not ordinarily seen in the wild-type plant), elevated or down-regulated (e.g. RNA interference (RNAi)). Moreover, a promoter could be fused to a reporter gene, such as β-glucuronidase (GUS) (Jefferson, Burgess and Hirsch, 1986) or to the green fluorescence protein (GFP) gene from jellyfish (*Aequoria victoria*) (Haseloff *et al.*, 1997), which can be used to reveal the pattern of expression conferred by a given promoter.

IDENTIFYING CANDIDATE GENES

Mutation analysis

Several experimental approaches have been taken to isolate genes that either confer a commercially useful trait or control a key aspect of plant development. The first, mutation analysis, involves screening thousands and possibly millions of seedlings for rare mutations that might aid in identifying desirable genes. This is a random, hit-or-miss approach that is slow, labour-intensive and sporadic when applied to tree species. In addition, because trees have long generation times, mate by cross-pollination and are highly heterozygous, rare recessive mutations are difficult to detect. A directed programme of inbreeding could be employed to expose recessive mutations, but inbreeding can also result in trees with poor form and low vigour owing to their high genetic loads, confounding attempts to identify valuable alleles. Tree improvement through these conventional means could require many decades, even with rapid advances in the area of plant genetics and the ease with which biotechnological tools can be applied to certain tree species (e.g. poplar; Bradshaw and Strauss, 2001).

In silico cloning

A second method for identifying candidate genes involves utilizing information from other model plants, such as the herbaceous annual *Arabidopsis thaliana*, to identify tree orthologs. An example of this approach is the identification of the *NAC1* gene, a root-specific member of a family of transcriptional regulators in plants. A mutation in *NAC1* diminishes lateral root formation and perturbs expression of *AIR3* (Xie *et al.*, 2000), a downstream gene associated with the emergence of lateral roots (Neuteboom *et al.*, 1999a, b). Furthermore, transgenic complementation with a functional *NAC1* gene restores lateral root formation, and overexpression results in a proliferation of lateral roots. Thus, the *NAC1* gene product appears to be both necessary and sufficient for lateral root formation. In this case, both sequence and functional information are being tested for functionality via transgenesis (B. Goldfarb, personal communication, North Carolina State University).

Forward genomics

A third way to facilitate gene discovery relies on the use of direct, random mutagenesis. Gene and enhancer trapping are methods for insertion-based gene discovery that both reference genome sequence data and result in a dominant phenotype (Springer, 2000). In short, gene-trap vectors carry a reporter gene lacking a functional promoter, while enhancer-trap constructs contain a minimal promoter preceding a reporter gene. In each case, the reporter gene is expressed in a fashion that imitates the normal expression pattern of the native gene at the insertion site, as has been demonstrated for *Arabidopsis* gene- and enhancer-trap lines (e.g. Springer *et al.*, 1995; Gu *et al.*, 1998; Pruitt *et al.*, 2000). The genomic region flanking the insertion site is amplified using PCR and sequenced; alignment of the flanking sequence with the genome sequence allows immediate mapping of insertions (Sundaresan *et al.*, 1995). This technique has recently been applied to identify genes likely to be involved in vascularization (Groover *et al.*, 2004). A similar strategy, using a luciferase-based promoter-trap vector, has allowed the identification of tissue- or cell-specific promoters (Johansson *et al.*, 2003).

Another forward genomics approach, namely activation tagging, utilizes a strong enhancer element that is randomly inserted into the genome and can be effective some distance from a promoter (Weigel *et al.*, 2000). Elevated expression of the nearby native gene may result in an aberrant phenotype. Lines exhibiting an obvious difference (early flowering, modifications in crown form, adventitious root development, etc.) are then analysed for the causative gene. Overexpression of some native genes (e.g. those affecting wood quality) may not give rise to a visually apparent change. In such cases, high throughput analyses are needed for screening a population of transgenics. The feasibility of this approach has already been demonstrated in poplar (Busov *et al.*, 2003). The recent release of the annotated draft of the *Populus trichocarpa* genome (www.phytozome.net/poplar.php) is facilitating the isolation and characterization of loci underpinning mutations found in similar ways.

Microarrays

A fourth approach to identifying candidate genes utilizes differential gene expression. The development of microarray technology has provided biologists with a powerful tool for studying the effects of gene expression on development and environmental responses (Brown and Botstein, 1999; Rishi, Nelson and Goyal, 2002). Expression levels of entire suites of genes, of both known and unknown function, can be measured simultaneously rather than one or a few genes at a time. This approach has already been successful in many systems. For root formation, a screen of loblolly pine shoots given a rooting treatment (auxin pulse) yielded a putative membrane transport protein that was induced by auxin treatment in juvenile (rooting) but not in mature (non-rooting) stem bases (Busov *et al.*, 2004). This gene shows homology to a large multigene family in *Arabidopsis*, members of which are similar to what was first classified as a nodulin from alfalfa.

PCR-based techniques

The fifth molecular technique to identify candidate genes is based on PCR, and includes suppression subtractive hybridization (SSH), differential display PCR (DD-PCR), and cDNA-AFLP (amplified fragment length polymorphism).

SSH is a PCR-based technique that was developed for the generation of subtracted cDNA libraries, and combines normalization and subtraction in a single procedure. Diatchenko *et al.* (1996) demonstrated that SSH could result in the enrichment of rare sequences by over 1000-fold in one round of subtractive hybridization. This technique has been a powerful tool for many molecular genetic and positional cloning studies to identify developmental, tissue-specific and differentially expressed genes (Matsumoto, 2006). For example, using SSH, bract-specific genes have been successfully identified in the ornamental tree *Davidia involucrata* (Li *et al.*, 2002), and genes responsive to benzothiadiazole (BTH; used to induce systemic acquired resistance) in the tropical fruit tree papaya (Qiu *et al.*, 2004). Genes involved in flowering have also been isolated from carnation (*Dianthus caryophyllus*; Ok *et al.*, 2003) and black wattle (*Acacia mangium*; Wang, Cao and Hong, 2005) using this method.

DD-PCR is another widely used method for detecting altered gene expression between samples, often derived from the treated and untreated individuals from the same genotype or species. An amplification is done using a primer that hybridizes to the poly(A) tail and an arbitrary 5' primer. The first application of this technology was reported by Liang and Pardee (1992), and has since been used with a wide variety of organisms, including bacteria, plants, yeast, flies and higher animals, to expedite gene discovery. A Myb transcription factor HbMyb1 associated with a physiological syndrome known as tapping panel dryness has been identified and characterized from rubber trees using differential display reverse transcriptase PCR (DDRT-PCR) (Chen *et al.*, 2002). Transcriptional profiling of gene expression from leaves of apricot (*Prunus armeniaca*) was conducted by DDRT-PCR and up- or down-regulated genes in response to European stone fruit yellows phytoplasma infection were identified (Carginale *et al.*, 2004). A significant disadvantage of this technique is its high percentage of false-positives (Zegzouti *et al.*, 1997).

cDNA-AFLP was first used by Bachem *et al.* (1996) to analyse differential gene expression during potato tuber development and was subsequently modified by Breyne *et al.* (2003). It too is a PCR-based method, which starts with cDNA synthesis, using random hexamer primers and total or mRNA as a template. Following digestion with two different restriction enzymes, adapters are ligated before amplification via PCR. This method has proven to be an efficient tool for differential quantitative transcript profiling and a useful alternative to microarrays (Breyne *et al.*, 2003). cDNA-AFLP was used to identify transcripts that accumulated in mature embryos and in *in vitro*-cultured plantlets subjected to desiccation or abscisic acid (ABA) treatment in almond (*Prunus amygdalus*; Campalans, Pages and Messeguer, 2001). Using this approach a novel gene, designated *Mal-DDNA*, was cloned and confirmed to play an important role in lowering the acidity of apple fruit (Yao *et al.*, 2007).

RNA interface

Double-stranded RNA-mediated gene suppression, also known as RNA interference (RNAi), was first reported in *Caenorhabditis elegans* a decade ago (Fire *et al.*, 1998). It is currently the most widely used method to down-regulate gene expression. It can be used to knock out all copies of a given gene, thus providing insight into its functionality. However, it does not always result in complete inhibition of a gene's expression. Recent advances in targeted gene mutagenesis and replacement using the yeast *RAD54* gene (Shaked, Melamed-Bessudo and Levy, 2005) or zinc-finger nucleases (Lloyd *et al.*, 2005; Wright *et al.*, 2005) may eventually lead to efficient methods for engineering null alleles in trees.

IMPROVEMENTS NEEDED

Regeneration

Regeneration protocols are typically optimized for a single genotype by conducting complex, labour-intensive, complete-factorial experiments. A more universal protocol has not been developed because of a lack of fundamental understanding of how plant cells acquire the competence to regenerate *in vitro*. Using rapidly advancing genomics tools, it is now possible to unravel this mystery. The research community now has access to a chip on which sequence information for all poplar genes has been spotted. Using this microarray, it is possible to identify genes that interfere with or promote regeneration by evaluating expression levels for all genes in tissues that differ in their regeneration potential, before and after being induced to regenerate. In addition, gene expression profiling that is done on tissues gathered during the juvenility-to-maturity transition could help identify genes affecting regeneration, in a similar manner to the approach described by Brunner and Nilsson (2004) to identify genes involved in flowering control.

Selection systems

As described above, a selectable marker gene is linked to the gene of interest that is being inserted. Transformed cells can then be isolated on a medium containing the appropriate selection agent. While this method is convenient, it is often problematic. First, performing subsequent rounds of transformation may not be possible because only a limited number of selectable marker genes are available. Second, various selection agents can have dramatic and negative effects on regeneration. Finally, the presence of a selectable marker gene is usually an impediment to gaining public acceptance of genetically engineered plants.

Recently, alternative selection systems have been developed. These are based on a growth medium that lacks a substance needed for metabolic activity or proper development. A particularly attractive option exploits the inability of a cell to regenerate a whole plant without the addition of a phytohormone, or its derivative, to the culture medium at a precise step in the regeneration process. For example, most regeneration protocols rely on an exogenous supply of cytokinin to induce differentiation of adventitious shoots or embryos from transgenic calli.

The *GUS* gene, a common reporter, encodes an enzyme that cleaves glucuronide residues. The glucuronide derivative of benzyladenine is biologically inactive; if it is the sole cytokinin incorporated in the induction medium, regeneration will not occur. However, upon hydrolysis by β-glucuronidase, a biologically active cytokinin is liberated to induce regeneration (Okkels, Ward and Joersbo, 1997). This supplement must necessarily be transitory because cytokinin can inhibit subsequent steps in development.

Another positive selection strategy involves inserting a gene whose product imparts a metabolic advantage to the transformed cell. Mannose is a sugar that plants are unable to metabolize; cells starve when grown on a medium containing mannose as the sole carbon source. When taken up by the cells, this sugar is phosphorylated by a native hexokinase. However, plants lack a native phosphomannose isomerase gene, which encodes an enzyme that catalyses the conversion of mannose to a usable six-carbon sugar (Joersbo *et al.*, 1998). Similarly, xylose isomerase, another enzyme that plants lack, is able to convert xylose to a sugar that can be utilized (Haldrup, Petersen and Okkels, 1998). Regeneration protocols that exploit positive-selection strategies such as these can be up to ten fold more efficient than those that rely on more traditional, negative-selection strategies.

Excision systems

The ability to delete unwanted pieces of DNA reliably is a valuable tool for both basic and applied research. Excision systems can remove selectable marker genes, thereby alleviating public concern and allowing for easy re-transformation using vectors derived from a common backbone. Moreover, some alternative regeneration methods (e.g. MAT, discussed below) depend on excision for their success. Because transposons have proven too unreliable, alternative systems, such as *Cre/lox* (Russell, Hoopes and Odell, 1992), FLP/FRT (Lyznik, Rao and Hodges, 1996) and *R/RS* (Onouchi *et al.*, 1995), have been utilized. Excision vectors typically include a recombinase gene, usually under the control of an inducible promoter, and recognition sites that flank the DNA being targeted for removal. However, these systems have not proven to be reliable in certain plants. Thus, it is necessary to determine which is the most appropriate for use with various tree species. For each system, one must ascertain the efficacy of the recombinase and how cleanly it excises the target sequence. Moreover, it is imperative to have an inducible promoter that functions reliably in the plant being transformed.

Producing marker-free plants

The recently developed multiautonomous transformation system (MAT) allows for the production of transgenic plants lacking selectable marker genes from a variety of species (e.g. tobacco, aspen, rice, snapdragon) (Ebinuma *et al.*, 1997; Ebinuma and Komamine, 2001). These vectors harbour *Agrobacterium* genes (*ipt* or *rol*) that control sensitivity to or the biosynthesis of phytohormones. Cells transformed with these vectors regenerate into plants with either a 'shooty'

or 'hairy-root' phenotype. MAT vectors also contain a site-specific, inducible recombinase for excision of both the recombinase and the oncogenes. This alternative production system is attractive because it has the potential to increase both the yield and speed with which transgenic plants can be produced, and may eliminate the need for specific selection and regeneration conditions, making it possible to transform a wider array of genotypes. Such a system will also be useful for stacking genes in forest trees, as described by Halpin and Boerjan (2003).

Mitigating transgene spread

The Coordinated Framework of the United States Animal and Plant Health Inspection Service (APHIS) now gives consideration to transgenic woody perennials. It is likely that before such trees can be deployed commercially, a method to mitigate the risk of transgene spread in the environment will be required, particularly in the cases when the introduced gene will improve the fitness of the genetically engineered tree. Many researchers are investigating ways to modify floral development to satisfy this need. The two most common approaches are to engineer trees that are either reproductively sterile or have delayed flowering. The latter may be particularly useful for short-rotation intensive culture (SRIC), where trees are harvested before the onset of maturation. Nevertheless, the main techniques being employed to modify floral development are:

- cell ablation (floral-specific expression of a cytotoxin gene);
- RNAi (silencing native genes via short, interfering RNAs);
- dominant negative mutations (DNMs), which lead to the production of a dysfunctional version of a gene product, such as a transcription factor (reviewed by Meilan *et al.*, 2001).

Because of functional redundancy, suppression of more than one floral regulatory gene is likely to be needed to achieve complete sterility. Where redundancy is obvious, RNAi constructs can be designed to silence effectively several members of a multigene family (Waterhouse and Helliwell, 2003). It is also advisable to utilize multiple techniques (e.g. cell ablation, RNAi or DNM, alone or in combination) to alter the expression of genes in more than one family to increase the likelihood of developing a durable confinement strategy. Transgene expression has been found to be unstable under various conditions (Brandle *et al.*, 1995; Köhne *et al.*, 1998; Metz, Jacobsen and Stiekema, 1997; Neumann *et al.*, 1997; Scorza *et al.*, 2001). Matrix attachment regions (MARs) have been used to enhance and stabilize transgene expression (Han, Ma and Strauss, 1997; Allen, Spiker and Thompson, 2000); however, there is some question about their utility (Li *et al.*, 2008). Given the potential for instability, it will be imperative to conduct multiyear field studies, in a variety of environments, and extending past the onset of maturity, in order to ensure the reliability of a given confinement system.

Progress in this area has been hampered by the inherent, delayed maturation of trees. Even the five- to seven-year juvenile period for poplar is a serious impediment. There is a report of a *Populus alba* genotype (6K10) that can be

induced to flower precociously, but it is of limited practical use (Meilan *et al.*, 2004). Its induction regime is lengthy and complex, and specialized equipment is required. In addition, not every plant in a population responds to induction. Moreover, the efficiency with which the genotype can be transformed and regenerated is very low. Because both male and female sterility will be needed, poplar is dioecious and 6K10 is a female, confinement systems will need to be tested in another poplar genotype. Early-flowering genotypes are rare and many trees do not respond well to treatments that induce precocious flowering (Meilan, 1997). Thus, there is a need for alternative genotypes that can be reliably and efficiently induced to flower.

BIO-INFORMATICS TECHNOLOGY

Bio-informatics is an interdisciplinary approach that utilizes computational and statistical techniques to aid in solving biological problems at the molecular level. Initially, bio-informatic tools were merely used to store, retrieve and analyse nucleic acid and protein sequence information. The field is now evolving rapidly, and being employed in newly emerging disciplines such as comparative genomics, transcriptomics, functional genomics and structural genomics. Below we briefly discuss some of the basic bio-informatics applications that are commonly used today.

Sequence analysis

One of the fundamental goals of sequence analysis is to determine the similarity of unknown or 'query' sequences to those previously identified and stored in various databases. A commonly used algorithm known as BLAST (basic local alignment search tool) provides a way to rapidly search nucleotide and protein databases. Since BLAST performs both local and global alignments, regions of similarity embedded in other, seemingly unrelated, proteins can be detected. Sequence similarity can provide important clues concerning the function of uncharacterized genes and the proteins they encode.

Other sequence-analysis tools are available to aid in determining the biological function and structure of genes and proteins, or to cluster them into related families based on their sequence information. Some software packages need to be purchased, others are available at no cost. The European Molecular Biology Open Software Suite (EMBOSS) is free, open-source software that can be downloaded from http://emboss.sourceforge.net/. It integrates many bio-informatics tools for sequence analysis into a single environment and can be used to analyse DNA and protein sequence in a variety of formats. Within EMBOSS there are hundreds of applications covering areas such as sequence alignment, rapid database searching for sequence patterns (e.g. to identify islands or repeats), protein motif identification (domain analysis), codon usage analysis for small genomes, and rapid identification of sequence patterns in large sequence sets. In addition, because extensive libraries are provided with this package, it is possible for users to develop and release software of their own. An example of another integrated

bio-informatics software can be found at http://ca.expasy.org/tools. As with EMBOSS, this package is helpful for characterizing and predicting the function of biomolecules of interest. Other commonly used sequence analysis applications include ClustalW and IMAGE.

Structure prediction

There are also software packages that can predict protein structure based on its sequence information or that of the gene encoded by it. Understanding protein structure is the key to revealing its function. Currently there are many programs for performing primary, secondary and tertiary structural analyses. ProtParam is a tool that computes physical or chemical parameters for a protein, such as molecular weight, amino acid and atomic composition, isoelectric point, extinction coefficient, estimated half-life, stability index and aliphatic index, based on user-entered sequence information. RasMol is an excellent graphics tool for visualizing macromolecular structure in order to help elucidate function. Other structure-prediction programs include Dowser, FastDNAml, LOOPP, MapMaker/QTL and PAML.

THE -OMICS

The 'omics' suffix is used to describe disciplines in which researchers analyse biological interactions on a genome-wide scale. The associated prefix indicates the object of study in each field. Examples include genomics, transcriptomics, metabolomics and proteomics. These encompass the study of the genetic make-up, the complete set of mRNA produced, the collection of metabolites, and protein function and interaction, respectively, in organisms, tissues or cells. The main focus of -omics is on gathering information at a given level and using computer-based tools to identify relationships in order to understand heterogeneous, biological networks, often with the ultimate goal of manipulating regulatory mechanisms. Omics require a multidisciplinary approach, bringing scientists together from a variety of fields to interpret the data collected.

APPLICATIONS

Rapidly emerging biotechnological tools can be used to help us better understand how biological systems function. The resulting discoveries allow us to introduce novel or alter existing traits that are useful to humans. Chapter 4 by McDonnell *et al.* in this volume provides a description of some commercially important and environmentally beneficial traits that have been incorporated into trees.

REFERENCES

Allen, G.C., Spiker, S.L. & Thompson, W.F. 2000. Use of matrix attachment regions (MARs) to minimize transgene silencing. *Plant Molecular Biology,* 43: 361–376.

Bachem, C.W.B., van der Hoeven, R.S., de Bruijn, S.M., Vreugdenhil, D., Zabeau, M. & Visser, R.G.F. 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant Journal,* 9: 745–753.

Bekkaoui, F., Tautorus, T.E. & Dunstan, D.I. 1995. Gymnosperm protoplasts. pp. 167–191, *in:* Jain, S.M., Gupta, P.K. & Newton, R.J. (editors). *Somatic embryogenesis in woody plants*, Vol. 1. Kluwer Academic Publishers, Dordrecht, Netherlands.

Bradshaw, H.D. Jr & Strauss, S.H. 2001. Breeding strategies for the 21st century: domestication of poplar. pp. 383–394, *in:* D.I. Dickmann, J.G. Isebrands, J.E. Eckenwalder & J. Richardson (editors). *Poplar culture in North America*. Part 2. NRC Research Press, National Research Council of Canada, Ottawa, Canada.

Bradshaw, H.D. Jr, Ceulemans, R., Davis, J. & Stettler, R.F. 2000. Emerging model systems: poplar (*Populus*) as a model forest tree. *Journal of Plant Growth Regulation,* 19: 306–313.

Brandle, J.E., McHugh, S.G., James. L., Labbe, H. & Miki, B.L. 1995. Instability of transgene expression in field grown tobacco carrying the *csr1-1* gene for sulfonylurea herbicide resistance. *Bio-Technology*, 13(9): 994–998.

Breyne, P., Dreese, R., Cannoot, B., Rombaut, D., Vandepoele, K., Rombauts, S., Vanderhaeghen, R., Inze, D. & Zabeau, M. 2003. Quantitative cDNA-AFLP analysis for genome-wide expression studies. *Molecular Genetics and Genomics,* 269: 173–179.

Brown, P.O. & Botstein, D. 1999. Exploring the new world of the genome with DNA microarrays. *Nature Genetics,* 21: 33–37.

Brunner, A.M. & Nilsson, O. 2004. Revisiting tree maturation and floral initiation in the poplar functional genomics era. *New Phytologist,* 164(1): 43–51.

Busov, V.B., Meilan, R., Pearce, D.W., Ma, C., Rood, S.B. & Strauss, S.H. 2003. Activation tagging of a dominant gibberellin catabolism gene (*GA 2-oxidase*) from poplar that regulates tree stature. *Plant Physiology,* 132: 1283–1291.

Busov, V.B., Johannes, E., Whetten, R.W., Sederoff, R.R., Spiker, S.L., Lanz-Garcia, C. & Goldfarb, B. 2004. An auxin-inducible gene from loblolly pine (*Pinus taeda* L.) is differentially expressed in mature and juvenile-phase shoots and encodes a putative transmembrane protein. *Planta,* 218(6): 916–927.

Campalans, A., Pages, M. & Messeguer, R. 2001. Identification of differentially expressed genes by the cDNA-AFLP technique during dehydration of almond (*Prunus amygdalus*). *Tree Physiology,* 21: 633–643.

Carginale, V., Maria, G., Capasso, C., Ionata, E., Cara, F.L., Pastore, M., Bertaccini, A. & Capasso, A. 2004. Identification of genes expressed in response to phytoplasma infection in leaves of *Prunus armeniaca* by messenger RNA differential display. *Gene*, 332: 29–34.

Chen, S., Peng, S., Huang, G., Wu, K., Fu, X. & Chen, Z. 2002. Association of decreased expression of a Myb transcription factor with the TPD (tapping panel dryness) syndrome in *Hevea brasiliensis*. *Plant Molecular Biology,* 51: 51–58.

Chupeau, M.C., Pautot, V. & Chupeau, Y. 1994. Recovery of transgenic trees after electroporation of poplar protoplasts. *Transgenic Research,* 3(1): 13–19.

Diatchenko, L., Lau, Y.F.C., Campbell, A.P., Chenchik, A., Mogadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. & Siebert, P.D. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proceedings of the National Academy of Sciences of the United States of America,* 93(12): 6025-6030.

Ebinuma, H. & Komamine, A. 2001. MAT (Multi-Auto-Transformation) vector system. The oncogenes of *Agrobacterium* as positive markers for regeneration and selection of marker-free transgenic plants. In vitro *Cellular & Developmental Biology-Plant,* 37(2): 103–113.

Ebinuma, H., Sugita, K., Matsunaga,, E. & Yamakado, M. 1997. Selection of marker-free transgenic plants using the isopentenyl transferase gene. *Proceedings of the National Academy of Sciences of the United States of America,* 94(6): 2117–2121.

Ellis, D.D., McCabe, D.E., McInnis, S., Ramachandran, R., Russell, D.R., Wallace, K.M., Martinell, B.J., Roberts, D.R., Raffa, K.F. & McCown, B.H. 1993. Stable transformation of *Picea glauca* by particle acceleration. *Bio-Technology*, 11(1): 84-89.

Farnum, P., Lucier, A. & Meilan, R. 2007. Ecological and population genetics research imperatives for transgenic trees. *Tree Genetics and Genomes*, 3(2): 119–133.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. & Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391: 806–811.

Fillatti, J.J., Sellmer, J., McCown, B., Haissig, B. & Comai, L. 1987. *Agrobacterium*-mediated transformation and regeneration of *Populus*. *Molecular Genomics and Genetics*, 206: 192–196.

Gelvin, S.B. 2003. *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiology and Molecular Biology Reviews,* 67(1): 16–37.

Groover, A., Fontana, J., Dupper, G., Ma, C., Martienssen, R., Strauss, S. & Meilan, R. 2004. Gene and enhancer trap tagging of vascular-expressed genes in poplar trees. *Plant Physiology,* 134: 1742–1751.

Gu, Q., Ferrandiz, C., Yanofsky, M. & Martienssen, R. 1998. The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development,* 125: 1509–1517.

Haldrup, A., Petersen, S.G. & Okkels, F.T. 1998. Positive selection: a plant selection principle based on xylose isomerase, an enzyme used in the food industry. *Plant Cell Reporter,* 18(1-2): 76–81.

Halpin, C. & Boerjan, W. 2003. Stacking transgenes in forest trees. *Trends in Plant Science,* 8: 363–365.

Han, K.-H., Ma, C. & Strauss, S.H. 1997. Matrix attachment regions (MARs) enhance transformation frequency and transgene expression in poplar. *Transgenic Research,* 6(6): 415–420.

Hansen, G. & Wright, M.S. 1999. Recent advances in the transformation of plants. *Trends in Plant Science,* 4: 226–231.

Haseloff, J., Siemering, K.R., Prasher, D.C. & Hodge, S. 1997. Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. *Proceedings of the National Academy of Sciences of the United States of America,* 94: 2122–2127.

Jefferson, R.A., Burgess, S.M. & Hirsch D. 1986. β-glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proceedings of the National Academy of Sciences of the United States of America,* 83: 8447–8451.

Joersbo, M., Donaldson, I., Kreiberg, J., Petersen, S.G., Brunstedt, J. & Okkels, F.T. 1998. Analysis of mannose selection used for transformation of sugar beet. *Molecular Breeding,* 4(2): 111–117.

Johansson, A.-M., Wang, C., Stenberg, A., Hertzberg, M., Little, C.H.A. & Olsson, O. 2003. Characterization of a PttRPS18 promoter active in the vascular cambium region of hybrid aspen. *Plant Molecular Biology,* 52(2): 317–329.

Klein, T.M., Fromm, M., Weissinger, A., Tomes, D., Schaaf, S., Sletten, M. & Sanford, J.C. 1988. Transfer of foreign genes into intact maize cells with high-velocity micro-projectiles. *Proceedings of the National Academy of Sciences of the United States of America,* 85: 4305–4309.

Köhne, S., Neumann, K., Pühler, A. & Broer, I. 1998. The heat-treatment-induced reduction of the *pat* gene encoded herbicide resistance in *Nicotiana tabacum* is influenced by the transgene sequence. *Plant Physiology,* 153: 631–642.

Levee, V., Lelu, M.A., Jouanin, L., Cornu, D. & Pilate, G. 1997. *Agrobacterium tumefaciens*-mediated transformation of hybrid larch (Larix kaempferi × L. decidua) and transgenic plant regeneration. *Plant Cell Reports,* 16(10): 680–685.

Li, J., Brunner, A.M., Shevchenko, O., Meilan, R., DiFazio, S.P., Ma, C., Skinner, J.S. & Strauss, S.H. 2008. Efficient and stable transgene suppression via RNAi in field-grown poplars. *Transgenic Research,* 17(4): 679–694.

Li, Y.X., Chen, L., Juan, L., Li, Y. & Chen, F. 2002. Suppression subtractive hybridization cloning of cDNAs of differentially expressed genes in Dovetree (*Davidia involucrata*) bracts. *Plant Molecular Biology Reports*, 20: 231–238.

Liang, P. 2005. A decade of differential display. *Biotechniques*, 33: 338–346.

Liang, P. & Pardee, A.B. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, 257(5072): 967–971.

Lloyd A., Plaisier, C.L., Carroll, D. & Drews, G.N. 2005. Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America,* 102: 2232–2237.

Lyznik, L.A., Rao, K.V. & Hodges, T.K. 1996. FLP-mediated recombination of FRT sites in the maize genome. *Nucleic Acids Research,* 24: 3784–3789.

Matsumoto T.K. 2006. Genes uniquely expressed in vegetative and potassium chlorate induced floral buds of *Dimocarpus longan*. *Plant Science,* 170(3): 500–510.

Meilan, R. 1997. Floral induction in woody angiosperms. *New Forests,* 14(3): 179–202.

Meilan, R., Brunner, A., Skinner, J. & Strauss, S. 2001. Modification of flowering in transgenic trees. pp. 247–256, *in:* A. Komamine & N. Morohoshi (editors). *Molecular breeding of woody plants.* (Progress in Biotechnology series.) Elsevier Science BV, Amsterdam, Netherlands.

Meilan, R., Sabatti, M., Ma, C. & Kuzminsky, E. 2004. An early-flowering genotype of *Populus*. *Journal of Plant Biology,* 47(1): 52–56.

Metz, P.L.J., Jacobsen, E. & Stiekema, W.J. 1997. Occasional loss of expression of phosphinothricin tolerance in sexual offspring of transgenic oilseed rape (*Brassica napus* L.). *Euphytica,* 98(3): 189–196.

Neumann, K., Dröge-Laser, W., Köhne, S. & Broer, I. 1997. Heat treatment results in a loss of transgene-encoded activities in several tobacco lines. *Plant Physiology,* 115: 939–947.

Neuteboom, L.W., Ng, J.M., Kuyper, M., Clijdesdale, O.R., Hooykaas, P.J. & van der Zaal, B.J. 1999a. Isolation and characterization of cDNA clones corresponding with mRNAs that accumulate during auxin-induced lateral root formation. *Plant Molecular Biology,* 39: 27–278.

Neuteboom, L.W., Veth-Tello, L.M., Clijdesdale, O.R., Hooykaas, P.J. & van der Zaal, B.J. 1999b. A novel subtilisin-like protease gene from *Arabidopsis thaliana* is expressed at sites of lateral root emergence. *DNA Research,* 6: 13–19.

Ok, S.H., Park, H.M., Kim, J.Y., Bahn, S.C., Bae, J.M., Suh, M.C., Jeung, J.U., Kim, K.N. & Shin, J.S. 2003. Identification of differentially expressed genes during flower development in carnation (*Dianthus caryophyllus*). *Plant Science,* 165: 291–295.

Okkels, F.T., Ward, J.L. & Joersbo, M. 1997. Synthesis of cytokinin glucuronides for the selection of transgenic plant cells. *Phytochemistry,* 46: 801–804.

Onouchi, H., Nishihama, R., Kudo, M., Machida, Y. & Machida, C. 1995. Visualization of site-specific recombination catalyzed by a recombinase from *Zygosaccharomyces rouxii* in *Arabidopsis thaliana*. *Molecular Genomics and Genetics,* 247: 653–660.

Pena, L. & Seguin, A. 2001. Recent advances in the genetic transformation of trees. *Trends in Biotechnol*ogy, 19(12): 500–506.

Pilate, G., Leplé, J.C., Cornu, D. & Lelu, M.A. 1999. Transgenic larch (*Larix* species). pp. 125–141, *in:* Y.P.S. Bajaj (editor). *Transgenic trees.* Vol. 44. Springer-Verlag, Berlin and Heidelberg, Germany.

Pruitt, R., Vielle-Calzada, J., Ploense, S., Grossniklaus, U. & Lolle, S. 2000. FIDDLEHEAD, a gene required to suppress epidermal cell interactions in *Arabidopsis*, encodes a putative lipid biosynthetic enzyme. *Proceedings of the National Academy of Sciences of the United States of America,* 97: 1311–1316.

Rishi, A.S., Nelson, N.D. & Goyal, A. 2002. DNA micro-arrays: gene expression profiling in plants. *Reviews in Plant Biochemistry and Biotechnology,* 1: 81–100.

Russell, S.H., Hoopes, J.L. & Odell, J.T. 1992. Directed excision of a transgene from the plant genome. *Molecular Genomics and Genetics,* 234: 49–59.

Qiu, X., Guan, P., Wang, M.L., Moore, P.M., Zhu, Y.J., Hu, J., Borth, W. & Albert, H.H. 2004. Identification and expression analysis of BTH induced genes in papaya. *Physiological and Molecular Plant Pathology,* 65(1): 21–30.

Scorza, R., Callahan, A., Levy, L., Damsteegt, V., Webb, K. & Ravelonandro, M. 2001. Post-transcriptional gene silencing in plum pox virus-resistant transgenic European plum containing the plum pox potyvirus coat protein gene. *Transgenic Research,* 10(3): 201–209.

Shaked, H., Melamed-Bessudo, C. & Levy, A.A. 2005. High-frequency gene targeting in *Arabidopsis* plants expressing the yeast *RAD54* gene. *Proceedings of the National Academy of Sciences of the United States of America,* 102: 12265–12269.

Springer, P. 2000. Gene traps: tools for plant development and genomics. *Plant Cell*, 12: 1007–1020.

Springer, P., McCombie, W., Sundaresan, V. & Martienssen, R. 1995. Gene trap tagging of Prolifera, an essential MCM2-3-5-like gene in *Arabidopsis*. *Science,* 268: 877–880.

Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J., Dean, C., Ma, H. & Martienssen, R. 1995. Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes and Development,* 9: 1797–1810.

Tang, W., Newton, R.J. & Weidner, D.A. 2007. Genetic transformation and gene silencing mediated by multiple copies of a transgene in eastern white pine. *Journal of Experimental Botany,* 58(3): 545–554.

Tereso, S., Miguel, C., Zoglauer, K., Valle-Piquera, C. & Oliveira, M.M. 2006. Stable *Agrobacterium*-mediated transformation of embryogenic tissues from *Pinus pinaster* Portuguese genotypes. *Plant Growth Regulation*, 50(1): 57–68.

Tuskan, G.A., DiFazio, S., Jansson, S. *et al.* 2006.: The genome of black cottonwood, *Populus trichocarpa (*Torr & Gray). *Science,* 313(5793): 1596–1604.

Wang, X.J., Cao, X.L. & Hong, Y. 2005. Isolation and characterization of flower-specific transcripts in *Acacia mangium*. *Tree Physiology,* 25: 167–185.

Waterhouse, P.M. & Helliwell, C.A. 2003. Exploring plant genomes by RNA-induced gene silencing. *Nature Reviews Genetics,* 4: 29–38.

Weigel, D., Ahn, J.H., Blazquez, M.A., Borevitz, J.O., Christensen, S.K., Fankhauser, C., Ferrandiz, C., Kardailsky, I., Malancharuvil, E.J., Neff, M.M., Nguyen, J.T., Sato, S., Wang, Z.Y., Xia, Y.J., Dixon, R.A., Harrison, M.J., Lamb, C.J., Yanofsky, M.F. & Chory, J. 2000. Activation tagging in *Arabidopsis*. *Plant Physiology*, 122(4): 1003–1013.

Wright, D.A., Townsend, J.A., Winfre,y R.J. Jr, Irwin, P.A., Rajagopal, J., Lonosky, P.M., Hall, B.D., Jondle, M.D. & Voytas, D.F. 2005. High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *Plant Journal,* 44: 693–705.

Wullschleger, S.D., Jansson S. & Taylor G. 2002. Genomics and forest biology: *Populus* emerges as the perennial favorite. *Plant Cell*, 14: 2651–2655.

Xie, Q., Frugis, G., Colgan, D. & Chua, N.-H. 2000. *Arabidopsis NAC1* transduces auxin signal downstream of *TIR1* to promote lateral root development. *Genes & Development,* 14(23): 3024–3036.

Yao, Y.X., Li, M., Liu, Z., Hao, Y.J. & Zhai, H. 2007. A novel gene, screened by cDNA-AFLP approach, contributes to lowering the acidity of fruit in apple. *Plant Physiology and Biochemistry*, 45(2): 139–145.

Zegzouti, H., Marty, C., Jones, B., Bouquin, T., Latché, A., Pech, J. & Bouzayen, M. 1997. Improved screening of cDNA generated by mRNA differential display enables the selection of true positives and the isolation of weakly expressed messages. *Plant Molecular Biology Reports,* 15: 238–245.