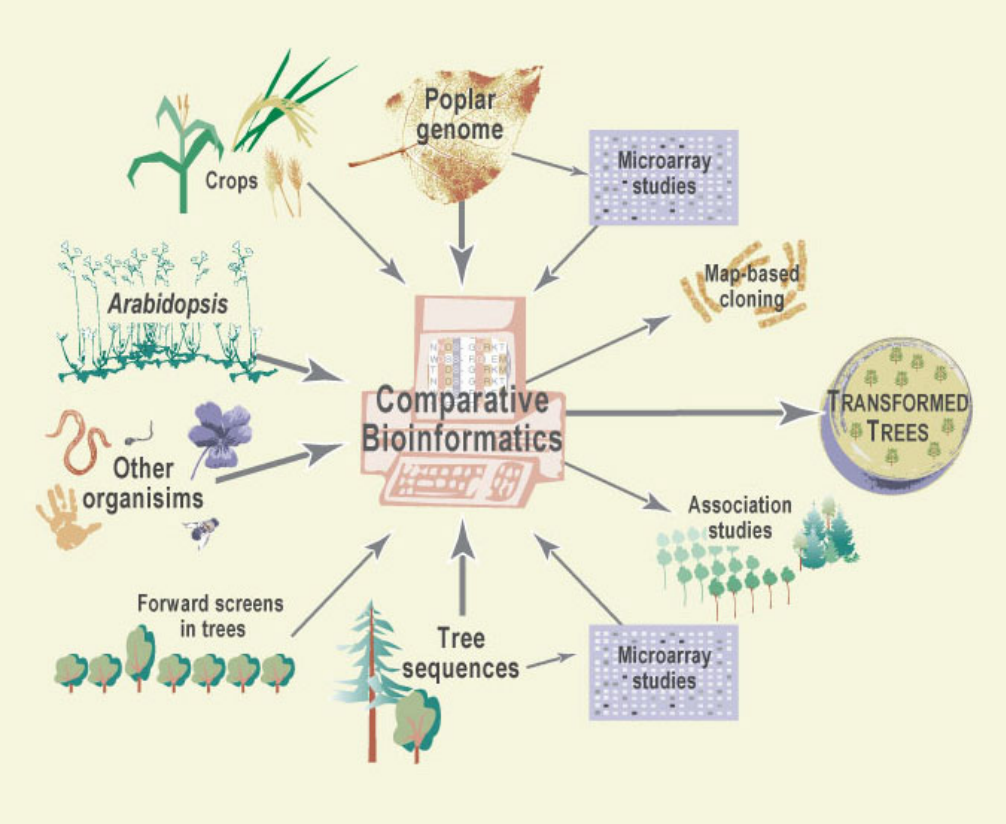
Tree biotechnology in the 21st century: Transforming trees in the light of comparative genomics

Steven H. Strauss and Amy M. Brunner

Department of Forest Science, Oregon State University, Corvallis 97331-5752

Corresponding author: Steve.Strauss@orst.edu, phone 541 737 6578, fax 541 737 1393



SUMMARY

Genetic engineering (GE) is the physical isolation, modification, and asexual transfer of genes. It allows new, qualitative changes to trees to improve productivity and pest management traits. It complements, but does not replace, the quantitative, polygenic, and incremental breeding for basic productivity and adaptation to environment of traditional breeding programs. As a new method for improving the productivity of planted forests, it could be important to meeting future challenges for preserving biodiversity via protection of wild forests from exploitation; sequestering carbon; replacing fossil energy and materials with renewably generated biomaterials; providing new means for bioremediation of polluted lands and water; and producing trees that are better able to cope with a variety of biotic and abiotic anthropogenic stresses.

Because of the conservation of the genetic code—and of basic developmental programs within plants—genes can be isolated, modified, and re-inserted into the same species, or moved across species, to produce effective trait modifications. Except where entirely new traits are desired [e.g., production of new chemicals (Rishi et al. 2001) or novel functions such as enhanced bioremediation capacity, and herbicide resistance (Merkle and Dean 2001)], it will usually be most effective to use knowledge gained from comparative genomic studies to identify and modify native genes or those obtained from within closely related species. In the near term, GE will require a clonal propagation system for delivering modified trees to plantations (Griffin et al. 1996). However, genes that speed breeding and increase clonal propagation efficiency may result in transgenes being deployed widely in breeding programs and sexual propagules in the future.

GE and gene transfer enable the production and wide deployment of novel, dominant "domestication" alleles that may have large benefits for productivity or product quality under intensive plantation management systems (Bradshaw and Strauss 2001a). These genes—which reduce competitiveness and reproductive success of single trees—are virtually inaccessible to traditional breeding due to their adverse effects on fitness in wild populations and consequent rarity and recessive gene action. They should therefore pose extremely little threat of "invasion" or disruption of intact wild populations via introgression.

The main technical obstacles to GE are the difficulty and costs of transformation and clonal propagation for most commercially important species, and the very limited knowledge of structural and functional comparative genomics in trees. Biosocial constraints include the lack of effective demonstration of a genetic confinement system that may be needed to convince the public of the environmental safety and legality of transgenic plantations, and strong social resistance to GE in forestry by some sectors of the public (Strauss et al. 2001). As widely agreed at a recent international conference (Bradshaw and Strauss 2001b), research is the most important need for development of forest biotechnology—particularly improvements in the efficiency of GE and clonal propagation to speed research and testing; new functional genomic knowledge; and field demonstrations of value and biosafety. A serious research commitment will be required in the near-term if GE products are to be available for commercial uses in one to two decades.

TREE BREEDING AND DOMESTICATION

Wherever humans have chosen to cultivate trees for specific products—be they fruit, wood, pulp, or energy—breeding has played a large part in boosting yields and adapting trees to the new environments and economics of production systems (e.g., Zobel and Talbert 1984). Some of the most dramatic examples come from the use of trees as exotics, where major changes in form, pest resistance, and yield often accompany crop development in novel environments and social contexts (Figure 1). The extent of genetic diversity in trees, and thus of the opportunity for genetic change, is most readily seen where trees have been deployed as clonal, vegetative propagules, as in some poplars (genus *Populus*) and eucalypts (genus *Eucalyptus*). This allows the genetic variation present, including the non-additive forms such as heterosis, to be readily observed and a large proportion of it captured. However, most tree species are not deployed as clonal propagules, and have undergone very little domestication of any sort. In contrast to most domesticated crops, only specialists would be able to distinguish bred trees from their wild relatives.

Figure 1 Change in tree form and growth during domestication. *Pinus radiata* in New Zealand after a single generation of selection for form and productivity (left), compared to trees from wild progenitors in California (right). The consequences of the increased growth and straightness for wood yield and product quality are obviously very large. This rapid change was enabled by the existence of highly polymorphic alleles for these traits. Other desirable changes in form and growth that may have large productivity benefits, but would be opposed by natural selection in the wild (e.g., absence of sexual reproduction), will be extremely difficult to achieve via conventional breeding. GE will be a far more effective means for domestication for these kinds of traits. Photo used with permission: Pauline Newman, Forest Research - New Zealand.

Difficulties imposed by trees

The limited domestication of trees is a consequence of the short time that serious breeding efforts have been made, and result from constraints to breeding inherent in their life history and management characteristics (Bradshaw and Strauss 2001b). The main obstacles to domestication of trees are:

1. The multiple-year delay until the onset of flowering, therefore slowing breeding cycles;
2. Intolerance of inbreeding that prevents the fixation of desirable alleles, and genotypic configurations, as inbred lines;
3. Large size and slow development that makes establishment and measurement of large breeding populations costly;
4. Low uniformity of the growth and testing environment when species are planted on non-agricultural lands, reducing trait heritability;
5. Long time frames needed to assess tree health and productivity;
6. Limited alteration of the production environment toward agronomic conditions, reducing the ability to modify genotypes to adapt to altered, productivity enhancing conditions; and

7. The perennial life cycle, and the consequent need for adaptation to unpredictable annual climatic cycles, requiring conservativeness in alteration of adaptive characteristics during domestication.

CHANGES IN BREEDING ENABLED BY GE

The constraints imposed by the environment, perennial growth, and the need to assess tree health over several years and sites, cannot be changed by GE. These are fundamental constraints of the biology of the organism and production system. However, for genes with little effect on organism physiology and adaptation (e.g., herbicide resistance), it should be possible to shorten the selection cycle once the extent of somaclonal variation—the background mutation imparted by the gene transfer process—is understood. This situation should be the case with poplars in the near future, where very low levels of somaclonal variation appear to be present, and highly stable, productive transgenic lines can be identified in the first few years of testing (e.g., Meilan et al. 2002; Figure 2). However, for novel transgenic changes that are expected to have multiple pleiotropic effects on productivity and physiology (e.g., significant changes in lignin chemistry), the testing cycle may need to be longer than is currently the case—at least during initial stages of use—to ensure that unexpected changes in adaptation to abiotic and biotic stresses have not been imparted. As with other forms of novel breeding, the extent of testing needed will be determined empirically—via adaptive management—during early commercial applications.

Figure 2 Example of healthy transgenic trees growing in a research trial in the United States. The trees in the left row comprise a number of different transgenic versions of a single hybrid cottonwood clone (*Populus deltoides* x *nigra*), and those on the right

Speeding flowering and clonal propagation

Flowering

GE may be able to accelerate breeding of trees by speeding the onset of sexual reproduction, and by facilitating clonal propagation from a diversity of genotypes and maturation states. The Arabidopsis genes *LEAFY* and *APETALA1* can speed flowering and provide normal seed set in citrus trees, potentially accelerating breeding markedly (Pena et al. 2001). *LEAFY* also speeds flowering in poplars, but is ineffective in many genotypes and does not appear to result in production of normal pollen or seed production (Rottmann et al. 2000). Nonetheless, these results inspire confidence that genes will ultimately be identified that can serve as genetic switches to allow flowering to be induced precociously in trees. Because of the genetic complexity and evolutionary diversity of mechanisms of flowering onset (reviewed in Battey and Tooke 2002), these may need to be different genes in different tree taxa.

Clonal propagation

Vegetative propagation is the most direct means for making large genetic gains because it allows a large proportion of the extensive genetic diversity in trees to be captured in a single cycle of selection. However, the extent to which genotypes submit to various propagation methods, such as rooting of cuttings, itself shows very high genetic variation, and as most trees age they lose the ability for the "rejuvenation" required for effective clonal propagation and derivation of healthy, fast growing clonal "seedlings" (reviewed in Brunner et al. 2002). The capability for rejuvenation, and the related trait "cellular regeneration capacity," also tend to show very high genetic variation, and may be controlled by major genes (e.g., Han et al. 1995). This suggests that it is possible to identify genes whose selective induction could promote the efficiency of clonal propagation, and perhaps enable lower cost methods to be employed (e.g., by avoiding the need for cryostorage to maintain juvenility). Fundamental studies of the genes that control epigenetic states, such as DNA methylation and histone acetylation (reviewed in Verbsky and Richards 2001), may also provide new means for reversing epigenetic programs, and studies of genes that control apomixis (Grimanelli et al. 2001) may someday provide new ways to induce the formation of seed-like clonal propagules. Finally, and most important for GE, methods for enhancing rejuvenability/totipotency of plant cells should enable increased rates of tranformation and regeneration of transgenic plants—currently the largest bottleneck to the use of transformation as a breeding and research tool.

Inserting novel alleles

GE derives its greatest novelty from the ability to isolate and selectively modify genes. All other breeding methods work from phenotypes or statistical tools for inferring abstract genetic properties from related individuals. Due to the conservation of the genetic code, this enables genes such as those for herbicide resistance to be utilized from other organisms, as well as desirable alleles that are rare in the gene pool, or present only as recessive forms, to be effectively used in breeding via conversion to dominant forms and insertion into a variety of genotypes. This is the case with traits such as reproductive sterility. Male-sterile genotypes, though rare, can be readily identified in wild populations of many species; however, they often involve organelle genomes or recessive loss-of-function nuclear genes (e.g., Schnable and Wise 1998). Using methods such as RNA interference (RNAi: e.g., Matzke et al. 2001; Wesley et al. 2001) or promoter:cell toxin fusions (ablation: e.g., Nilsson et al. 1998), dominant loss-of-function alleles can be created anew and inserted (or crossed) into a diversity of genotypes with dominant trait inheritance.

Because plant domestication often involves alterations in plant phenotypes that make them more productive under high density, high fertility, monoclonal plantings, it often involves the loss of functions important to promoting fitness in biotically and environmentally diverse wild populations. For example, dwarf varieties are commonly used in grains and highly desirable in tree orchards, and several genes for dwarfism have been identified (Silverstone and Sun 2000). Similar changes may be desirable in forest plantation trees as short, narrow-crowned trees may be more productive in monoclonal wood plantations if the population invests less in roots, branches, and reproductive tissues relative to stems, and more stems can be packed per unit area (Bradshaw and Strauss 2001b). Alleles that cause such deleterious changes in form would be

very difficult to find in wild populations, but should be readily produced via GE. Likewise, genes that change wood structural or chemical qualities to improve its value as an industrial feedstock will tend to move trees away from the normal range produced by natural selection. Such alleles will therefore be difficult to identify at a high frequency in the wild, and thus very difficult to increase in frequency via traditional population breeding in trees. However, they could also be readily produced via GE. Genes that affect lignin quantity and chemistry are examples of such genes; they could be of tremendous economic and ecological value by reducing the great cost, and byproducts, of lignin removal during pulping ({Sedjo, this volume}), and possibly also increase fiber productivity and quality (Hu 1999, Dinus et al. 2001).

TRANSLATING GENOMIC KNOWLEDGE INTO TREE BIOTECHNOLOGY

Comparative genomics

In addition to the ability to produce and mobilize alleles for desired traits that would not be accessible to traditional breeding, another major advantage to GE is the ability to capitalize on knowledge from comparative genomics and molecular biology (Figure 3). This is enabled by the extensive conservation of gene and derived protein sequences, the availability of powerful computer methods for searching and comparing large sequence databases, and the extensive functional and microsyntenic (map position) conservation within broad taxonomic classes of plants. Thus, large scale mutagenesis projects in model organisms (e.g., the *Arabidopsis* 2010 project: Chory et al. 2000) should produce a great deal of information useful to molecular

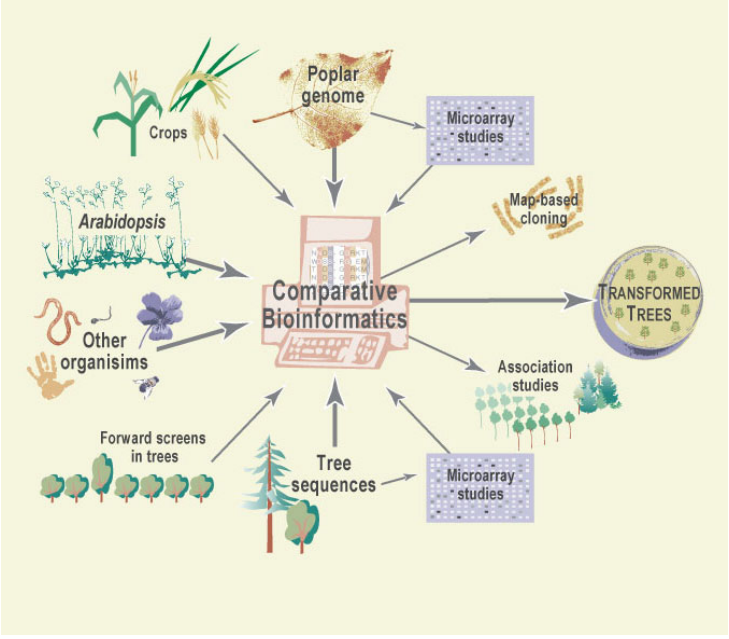


Figure 3 Genomics information important to molecular domestication of forest trees. Bold items and arrows show information and analysis expected to be of most important to genomic analysis or commercial deployment. Because of the high degree of conservation of the genetic content of organisms, especially within plants, genomic information about any species can be used to suggest routes for study, marker-aided breeding, or transgenic modification of trees. Information from intensively studied species such as Arabidopsis, and fundamental studies of molecular genetic biology in other model organisms, are expected to be most important because they will contain a large amount of information on gene function as well as sequence. This information, plus functional genomic studies from other crop and plant species, can be used to identify candidate genes in databases of tree genomes that may be useful for modifying economically important traits. Poplars will be of particular value because of its soon-to-be-completed genome sequence, and its amenability to transformation—the latter allowing hypotheses of gene function and trait modification in trees to be directly and rigorously tested. Forward screens in trees, where transformation is used to create tagged mutants, is also expected to be useful for species like poplar where transformation can be conducted on a large scale. High-throughput studies of gene expression through microarray hybridization allows new genes from trees to be recognized based on expression patterns, and the expression characteristics of large numbers of homologous genes to be verified. Recognition of gene homologies with other organisms via sequence or synteny will aid identification of genes in map-based cloning experiments, and in selection of genes for studying associations of natural variants in traits with sequence polymorphisms. Because of the costs and statistical difficulties of fine mapping for quantitative trait loci (QTL) in most tree species, transformed trees will be needed for demonstrating the functional roles of suspected allelic variants for important traits, as well as for deploying dominant alleles that can impart domestication phenotypes in the variety of genetic backgrounds needed for commercial applications.

breeders of any plant species. By contrast, traditional breeding is necessarily confined to knowledge from the species or closely related taxa.

For example, for genes that affect wood chemistry it will often be profitable to look to studies of model organisms—particularly where there are complete genome, extensive EST sequence banks, or large scale expression/function assays—to try and identify key transcription factors that efficiently control wood quality in trees. Likewise, because of the complexity of flowering it will nearly always be preferable to study candidate genes identified by extensive studies of flowering genes in rapid-cycling model organisms than to try and isolate genes for this process *de novo* from trees. However, large sequence databases, either from genome or expressed genes (cDNA), must be available in trees for comparative genomic methods to proceed. Therefore, one of the great benefits of structural genomics projects in trees is the *in silico* "functional" access it provides to studies in model organisms. Alternatively, though far less powerful, if there is sufficient gene conservation—and the ortholgous gene family members can be targeted (no small task for most genes)—specific genes can be chosen for isolation in tree species based on their sequence and putative functional conservation.

Even where large sequence banks exist for tree species, the ability to select the best genes from among the numerous (often hundreds) of candidate genes and gene families for control of specific traits is challenging. Gene deletion, duplication, mutation, rearrangement, and functional divergence often make recognition of functional evolutionary orthologs difficult. The same gene or its functional ortholog may perform differently depending on its organismal context. Moreover, evolutionary changes in key genes can alter entire regulatory networks. For example, changes in the level of an encoded protein due to gene duplication/deletion, or novel interactions among proteins, can modify a signaling pathway as well as the extent of cross-talk and redundancy among pathways. Thus, some kind of functional assay in trees is usually required. The three basic experimental options are association studies, microarray expression analysis, and production of transgenic trees.

Association studies: Searching for statistical associations between traits and natural genetic variants of candidate genes in wild or experimental populations

Association studies will help to define effective targets for GE, as well to provide options for marker-aided breeding. However, there are a number of serious obstacles to their application. The main constraints are the limited knowledge of potential candidate genes underlying specific traits, and the difficulty and cost of screening large numbers of candidate loci and progeny. Large numbers need to be screened because of low linkage disequilibrium in outbred tree populations and high background environmental and genetic variability. Phenotyping of progeny can also be a major cost, especially for complex traits such as wood chemistry, or delayed expression traits such as onset of flowering.

Array expression studies: Scanning large numbers of genes from trees for patterns of expression that suggest an important role in specific processes

This requires a large sequence bank of genes, derived via either cDNA or genomic sequencing, and thus is also a route for novel gene discovery (reviewed in Aharoni and Vorst 2002). Key

constraints are the costs of generating the sequence data, and that even large EST banks usually fail to include a large proportion of genes. EST databases are usually biased toward highly expressed "structural" genes, whereas weakly expressed transcription and signal transduction factors can play key roles in generating phenotypic variation (Doebley and Lukens 1998). Conducting and statistically analyzing large arrays from a diversity of experiments is also problematic (e.g., Quackenbush 2001), and generally requires verification studies with more precise, single gene or small-array methods. A key advantage of array studies is that this method allows many genes, including many "novel" genes (i.e., without clear homologs/orthologs in other species), to be identified. It is therefore also a method of *de novo* gene discovery. By allowing global expression changes to be monitored under different environmental conditions, at different time points following an environmental stimulus, at different ages and stages of development, and in different tissues, microarray studies can point to the most appropriate gene(s) to manipulate for a desired phenotype. For example, such studies may distinguish genes which are likely to have pleiotropic effects or undesired additional effects under a particular environmental stress from a gene which acts downstream in a single biochemical pathway (e.g., Chen et al. 2002). An in depth microarray study of xylem differentiation in poplar successfully clustered large numbers of known and novel genes (Hertzberg et al. 2001).

Transgenic studies: The production of directed mutants—via impairment/loss of gene expression or hyperexpression—in an isogenic background

Transformation provides the most informative and precise information on the basic physiological function of genes, and thus of what might be accomplished with them via GE. The main constraint is that high throughput GE is not possible in most tree species, with the notable exception of some poplar genotypes. However, even here the cost of generating the tens to hundreds of transgenics desired for investigating the roles of a number of candidate genes for a specific trait is considerable. And because of the variation in transgene expression with each gene transfer event it is necessary to generate, and both phenotypically and molecularly evaluate, at least a dozen independent lines per gene. Finally, because of the strong canalization and redundancy of developmental mechanisms—and the presence of many gene family members with similar roles—single gene knock-outs often fail to display obvious phenotypes. Seeing the effects of specific genes therefore often requires multiple-gene suppression, or gain-of-function transgenes, such as those with ectopic overexpression. However, gain-of-function mutations are much harder to interpret in terms of native function. Inducible or tissue-specific overexpression is preferable, but is rarely employed, for large scale screens. The recent demonstration (Abbott et al. 2002) that three unrelated endogenes were coordinately suppressed by a single chimeric transgene suggests that multiple gene suppression can become a feasible approach. Adapting this approach to large-scale studies will likely require use of the more efficient RNAi method (rather than cosuppression) for inducing gene-silencing (Wesley et al. 2001), and development of a high-throughput system for vector construction.

The main obstacle to comparative genomics is therefore cost for conducting the large-scale sequencing, and/or functional studies, needed. Like with other organisms, this is best addressed by large, multi-investigator, internationally coordinated efforts to establish basic sequence/informatics databases, and national or international centers for efficient functional

assays. Unfortunately, private databases, such as the large EST databases for *Pinus* and *Eucalyptus* produced by Genesis/Arborgen, are generally not available to others, and public efforts, while somewhat coordinated, have often been duplicative (e.g., multiple studies of poplar cambial tissues). There has also been very limited provision of printed microarrays to the international community from expression centers, impeding standardization of array experiments. However, the upcoming sequencing of the poplar genome by the Department of Energy has been well planned internationally, and its informatic analyses and resulting database composition is expected benefit from provision of data from several countries.

Novel gene discovery in trees

In addition to use of genome knowledge from model organisms to select genes for investigation in trees, due to their different phylogenetic and adaptive histories there are likely to be many genes that control important "tree-specific" traits that will be missed by this method alone. It is therefore highly desirable to independently identify developmental regulatory genes in trees. The options for gene discovery in trees are map-based cloning, large scale DNA sequencing, and gene tagging.

Map-based cloning: Use of genetic linkage mapping of natural genetic variants to identify the underlying genes

Map-based cloning depends on the availability of natural or synthesized populations that have substantial linkage disequilibrium (such as a single full-sib family), and that segregates for genes that affect a target trait. Wild populations could also be employed, however, they would require a density of markers that is currently unavailable and unaffordable for any tree species. For example, in landraces of maize disequilibrium declines rapidly beyond 200 bp (Tenaillon et al. 2001); a similar level of disequilibrium in trees would therefore require hundreds of thousands of markers to identify unknown genes. In pedigree populations, the most significant obstacles are the need for very large populations to be screened to enable sufficient genetic precision for physical isolation, and the difficulty of fine-mapping with quantitative trait loci (QTL; Flint and Mott 2001). For example, even a highly heritable trait like vegetative growth phenology—when studied intensively in specific environments and clonally replicated genotypes—appears to be controlled by numerous genes of small effect (Jermstad et al. 2001). More complex and economically important traits such as yield or wood quality are likely to prove far more intractable. This will make QTL localization extremely difficult. Moreover, in most cases it would still be necessary to transform with a number of candidate genes from a mapped region to identify the causative one, assuming its effects are at least partially dominant and statistically detectable (Mackay 2001). Map-based cloning therefore appears to be technically infeasible for the large majority of tree species and genes in the foreseeable future.

Identifying novel genes through genome and EST sequencing

Either array hybridization or high-throughput transformation can be used to determine the effects of novel (or highly dissimilar) genes that have been revealed in sequencing projects. For example, genes important to onset of flowering might be identified by their associated changes in expression with tree age/reproduction (cf. Chen et al. 2002). Gene selection could be aided by

searching for motifs such as DNA binding or kinase domains that give indications of broad function, and transformation studies could be conducted after informatic and array studies suggest which genes are most likely to affect target traits and processes. The costs and complications of these methods are similar to those discussed above for array and transformation methods in comparative genomics.

Transformation-based gene tagging: Use of random gene insertion as a means to identify novel genes based on developmental phenotypes or reporter-gene expression

The vast majority of genes that have been identified in genetics are the result of analysis of mutants. Mutations are generally caused by recessive, loss-of-function alleles that require inbreeding and consequent homozygosity to reveal. Despite the vast abundance of natural mutations in tree populations, because of the long generation times, intolerance of inbreeding, and poor genetic maps, for trees large mutant populations have neither been created nor screened for the causative genes. However, two recent techniques for creating tagged, dominant mutants in trees facilitates direct mutation-based gene isolation.

Activation tagging

Activation tagging is a method whereby a strong enhancer of gene expression—active up to several kb from a promoter—is randomly inserted into the genome, disturbing (elevating) expression of a nearby gene (Weigel et al. 2000). This gain-of-function causes dominant phenotypes and the presence of the known transgene sequence—generally with an antibiotic resistance marker to facilitate plasmid rescue—enables the affected portion of the genome to be isolated and sequenced. Ultimately, a candidate gene nearest to the enhancer is identified and then re-transformed to verify that it is the cause of the phenotype. This method yields tagged phenotypes in about 1% of independent transgenic lines, including in poplars (Ma et al. 2001). It therefore requires that very large populations of transgenics, and rapid screening methods for specific traits, are available. Moreover, the complex rearrangements that can occur during transformation require that putatively tagged genes be re-transformed to verify their association with phenotypes (cf. Tax and Vernon 2001). A large-scale tagging project in a tree—even poplar—is therefore not a small task.

Gene/promoter trapping

Gene/promoter trapping is a method whereby a reporter gene, such as GUS, is randomly inserted into the genome either with a basal promoter (allowing it to be expressed whenever it lands by an enhancer), or without a basal promoter, requiring direct incorporation into a gene for expression (reviewed in Springer 2000). The population of transgenics is then screened for developmental-, tissue-, cell-, or environment-induced expression patterns of interest. This method yields a considerably higher number of tagged mutants than does activation tagging, and rapidly identifies promoters that may have desired expression properties, but does not itself give developmental mutations. A large number of vascular expressed genes have been identified in this manner in poplar (A. Groover, U.S. Forest Service, pers. comm.)

REFINING EXPERIMENTAL TRANSGENICS INTO COMMERCIAL PRODUCTS

Creating new alleles for target traits

If the social controversies over GE do not curtail funding for research, the next decade or two could be productively spent both scanning genomes using comparative and *de novo* gene discovery methods, and testing proof-of-concept transgenic trees for the ability to modify important traits. These traits will certainly include mainstays such as yield, wood structure, and wood chemistry, but could also include more novel and specific traits such as for crown structure, flowering, vegetative propagation capacity, transformability, growth rhythms, and biotic and abiotic stress resistances. With the upcoming complete genome sequence of poplar—the model forest tree—tackling it vigorously in an internationally coordinated manner via both array and transgenic studies would seem highly desirable. It would also be desirable to conduct similar studies in taxonomically distinct forestry species, such as in a conifer and eucalyptus. However, studies in these species would be likely to rely less primarily on array studies, rather than on functionally more informative transformation tests. In a decade, it might be possible to have identified a large number of new genes that are likely to, or have been demonstrated in transgenics, to affect important traits in trees. However, much additional work would then be required to develop actual commercial varieties.

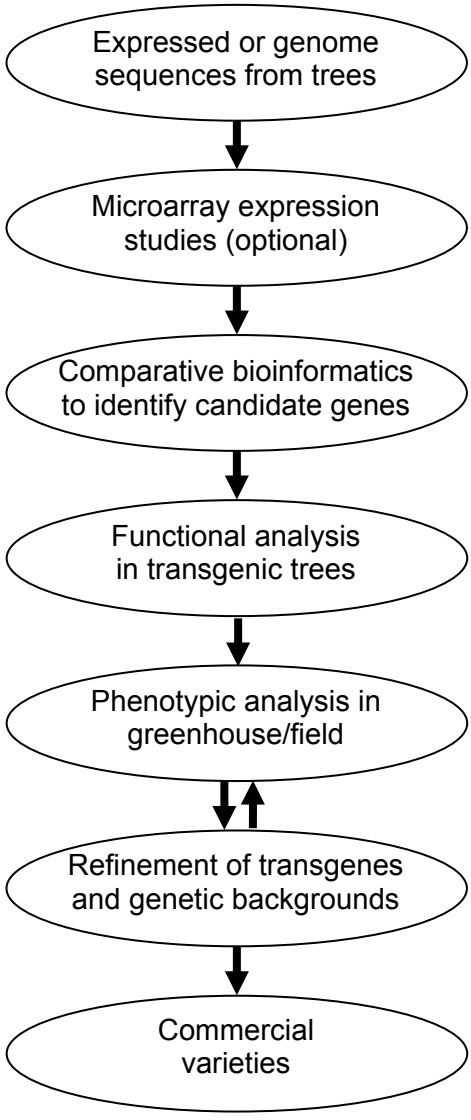


Figure 4 Summary of steps in tree domestication via comparative bioinformatics and transformation.

In many cases simple gene knock-outs or overexpression phenotypes will be inadequate for delivering the precise phenotypic changes desired in a domesticated tree. For example, excessive lignin reduction is likely to be deleterious, but an intermediate level may be both ecologically safe and economically valuable. It will therefore be necessary to identify the best genes of several that can affect a given trait, and then to learn how to engineer the best alleles for those genes (Figure 4). RNAi may be adequate where reductions in gene activity are required, however, too little is known at present about its stability and quantitative consistency during development and environmental variation. Alternative means for modulating gene expression, such as by using designed zinc-finger transcription factors for transcription repression or activation (Beerli and Barbas 2002), may be needed for more refined control of gene expression. In addition, facile methods for coordinate suppression or overexpression of multiple genes, such as use of a single

polyprotein transgene to express multiple proteins, will be needed to generate many desired phenotypes (reviewed in Halpin et al. 2001).

Likewise, there are too few native promoters and reliable means for inducing, or elevating, gene expression in tree species. This will be important where high levels of expression of foreign genes are desirable in specific tissues to enhance activity while avoiding unintended ecological effects (e.g., for production of industrial or bioremediation proteins). Thus, if we are to be successful in generating genomic information that leads to useful transgenic varieties, it will be important not only to conduct large screens of gene function, but also to learn to engineer the most promising genes so their effects can be carefully controlled. Toward this goal system-level genomic tools, such as arrays, that allow the monitoring of changes in metabolism imparted by transgenes, may be useful for detecting adverse pleiotropic effects as early as possible. For example, it might be helpful in verifying that genetic circuitries for responding to herbivory are not unduly impaired by lignin modification genes. The analysis of array data to infer the function, and responses to perturbations, of complex regulatory networks are currently the subject of intensive bioinformatics research (e.g., Gifford 2001, Wyrick and Young 2002).

Ecological and social issues

Hubris and skepticism

Despite the abundant genetic diversity present in forest tree species, because of their outcrossing system of mating, long generation times, very large effective populations sizes, strong constraints from natural selection, and limited efforts at domestication, there have been only modest changes to tree form, productivity, or wood properties. Even the poplars and eucalypts, where marked improvements of productivity have resulted from interspecific hybridization combined with clonal propagation, have still retained their basic form and properties. The tools of genetic engineering can clearly move species beyond—into new developmental zones—compared to all prior breeding technologies. This requires caution and skepticism, as well as technological enthusiasm, for the possibilities and new knowledge it will bring.

Although we know that one or a few major genes can result in domestication traits in crops (e.g., Paterson et al. 1995; Bradshaw and Strauss 2001b), it is unclear whether adapted varieties can be produced with those major gene changes if there are not also the many polygenic modifications that also accompanied annual crop breeding during domestication. In forestry, such fine tuning can be done to a much more limited extent, mainly by careful selection of transgenic events and genetic backgrounds, and extensive field testing prior to widespread use. The need to assess a diversity of transgenic lines and genetic backgrounds further underlines the importance of gene transfer efficiency. If there is a single kind of trait that is most important to molecular domestication via transformation—and thus should be screened for vigorously in functional genomic screens—perhaps it is the ability to regenerate from single transformed cells. Although great advances have been made in transformation methods to utilize positive phenotypic markers of selection and to enable the elimination of unwanted transgenes such as those for antibiotic resistance (e.g., Ebinuma et al. 2001), there are still no transgenes whose expression can increase transformation/regeneration frequency substantially in a variety of genotypes and species.

The main ecological risks of using novel, domesticated trees are to the growers and local economies, rather than to wild ecosystems, as these highly-altered trees are very unlikely to be competitive in wild environments. However, the public may insist on a high degree of precaution until a record of safety is established. The novelty of such trees may also force a social requirement that all feasible steps be taken to effectively confine them to plantations, requiring some form of engineered infertility. Thus, in addition to genomic scans it will be important during the upcoming years to continue developing and testing the efficiency of infertility transgenes. The best means for doing this might be through release of some commercial horticultural products whose only trait is male and/or female sterility (to reduce allergens and nuisance tissues), combined with monitoring of gene flow or other effects (cf. NRC 2002). This would allow long-term, commercial scale "testing," under highly diverse environments, of the reliability of engineered sterility, with little ecological consequence should it fail. Rigorous testing of infertility systems is also likely to be required before genes that might increase fitness in wild populations—such as novel pest toxin genes like those from *Bacillus thuringiensis*—could be deployed (e.g., {Raffa, this volume}).

Patents and cooperation

The scientific and technological possibilities for forest biotechnology are vast. However, the costs, social concerns, and ecological unknowns are also considerable. It seems clear that progress will require collaboration on many fronts, as no one organization—academic, government, corporate, or consortium—can do all that is needed. In this respect we suggest that broad patents on gene sequences—unless clearly tied to utility in specific commercial tree taxa—be avoided whenever possible (Williamson 2001). The prospect of large license costs and legal fees to negotiate the intellectual property maze, combined with the already very considerable regulatory and public relations costs of biotechnology, clearly discourage the broad interest and support from the forest industry, small landowners, and the professional natural resource community that is needed for progress. The associated secrecy also provides an environment in which "Frankensteinian paranoia" seems to grow and prosper (cf. Bobrow and Thomas 2001), as do the risks of liability for genetic pollution from one company's patented genes. The forest biotechnology community can hardly afford much more of this social turbulence. We therefore encourage companies and researchers to carefully consider the costs of a highly patent and secrecy oriented approach to research. We believe that the presumed benefits to single companies it provides in the short-term may be overshadowed by the enormous, cumulative burden it places on the entire forest biotechnology enterprise.

REFERENCES

Abbot, J. C., A. Barrakate, G. Pincon, M. Legrand, C. Lapierre, I. Mila, W. Schuch, and C. Halpin. 2002. Simultaneous Suppression of Multiple Genes by Single Transgenes: Down-Regulation of Three Unrelated Lignin Biosynthetic Genes in Tobacco. *Plant Physiology* 128:844-853.

Aharoni, A., and O. Vorst. 2002. DNA Microarrays for Functional Plant Genomics. *Plant Molecular Biology* 48:99-118.

Battey, N.H., and F. Tooke. 2002. Molecular Control and Variation in the Floral Transition. *Current Opinions Plant Biology* 5:62-68.

Beerli, R.R., and C.F. Barbas III. 2002. Engineering Polydactyl Zinc-Finger Transcription Factors. *Nature Biotechnology* 20:135-141.

Bobrow, M., and S. Thomas. 2001. Patents in a Genetic Age. *Nature* 409:763-764.

Bradshaw, H.D., and S.H. Strauss. 2001a. Plotting a Course for GM Forestry. *Nature Biotechnology* 19:1104

Bradshaw, H.D., Jr., and S.H. Strauss. 2001b. Breeding Strategies for the 21st Century: Domestication of Poplar. In *Poplar Culture in North America,* edited by D.I. Dickmann, J.GT. Isebrands, J.E. Eckenwalder, and J. Richardson. Ottawa, Canada: NRC Research Press, National Research Council of Canada, 383-394.

Brunner, A.M., B. Goldfarb, V. Busov, and S.H. Strauss. 2002. Controlling Maturation and Flowering for Forest Tree Domestication. In *Transgenic Plants: Current Innovations and Future Trends*, edited by C.N. Stewart, Wymondham, UK. Horizon Scientific Press, in press.

Chen, W., N.J. Provart, J. Glazebrook, F. Katagiri, H-S. Chang, T. Eulgem, F. Mauch, S. Luan, G. Zou, and S.A. Whitham. 2002. Expression Profile Matrix of *Arabidopsis* Transcription Factor Genes Suggests their Putative Functions in Response to Environmental Stresses. *The Plant Cell* 14:559-574.

Chory, J., J.R. Ecker, S. Briggs, M. Caboche, G.M. Coruzzi, D. Cook, J. Dangl, S. Grant, M.L. Guerinot, A. Henikoff, R. Martienssen, K. Okada, N.V. Raikhel, C.R. Somerville, and D. Weigel. 2000. National Science Foundation-Sponsored Workshop Report: "The 2010 Project". *Plant Physiology* 123:423-425.

Dinus, R.J., P. Payne, M.M. Sewell, V.L. Chiang, and G.A. Tuskan. 2001. Genetic Modification of Short Rotation Poplar Wood Properties for Energy and Fiber Production. *Critical Reviews Plant Science* 20:51-69.

Doebley, J. and L. Lukens. 1998. Transcriptional Regulators and the Evolution of Plant Form. *Plant Cell* 10: 1075-1082.

Ebinuma, H., K. Sugita, E. Matsunaga, S. Endo, K. Yamada, and A. Komamine. 2001. Systems for the Removal of a Selection Marker and their Combination with a Positive Marker. *Plant Cell Reports* 20:383-392.

Flint, J., and R. Mott. 2001. Finding the Molecular Basis of Quantitative Traits: Successes and Pitfalls. *Nature Reviews: Genetics* 2:437-445.

Gifford, D.K. 2001. Blazing Pathways Through Genetic Mountains. *Science* 293:2049-2051.

Griffin, R. 1996. Genetically Modified Trees: The Plantations of the Future or an Expensive Distraction? *Commonwealth Forestry Review* 75:169-175.

Grimanelli, D., O. Leblanc, E. Perotti, and U. Grossniklaus. 2001. Developmental Genetics of Gametophytic Apomixis. *Trends Genetics* 17:597-604.

Halpin, C., A. Barakate, B.M. Askari, J.C. Abbot, and M.D. Ryan. 2001. Enabling Technologies for Manipulating Multiple Genes on Complex Pathways. *Plant Molecular Biology* 47:295-310.

Han, K.-H., H.D. Jr. Bradshaw, M.P. Gordon. 1995. Adventitious Root and Shoot Regeneration *In Vitro* is Under Major Gene Control in an F_2 Family of Hybrid Poplar (*Populus trichocarpa* x *P. deltoides*). *Forest Genetics* 1:139-146.

Hertzberg, M., H. Aspeborg, J. Schrader, A. Andersson, R. Erlandsson, K. Blomqvist, R. Bhalerao, M. Uhlen, T. Teeri, J. Lundeberg, B. Sundberg, P. Nilsson, and G. Sandberg, 2001. A Transcriptional Roadmap to Wood Formation. *Proceedings of the National Academy of Science USA* 98:14732-14737.

Hu, W.-J., S.A. Harding, J. Lung, J.L. Popko, J. Ralph, D.D. Stokke, C.-J. Tsai, and V.L. Chiang. 1999. Repression of Lignin Biosynthesis Promotes Cellulose Accumulation and Growth in Transgenic Trees. *Nature Biotechnology* 17:808-812.

Jermstad, K.D., D.L. Bassoni, K.S. Jech, N.C. Wheeler, and D.B. Neale. 2001. Mapping of Quantitative Trait Loci Controlling Adaptive Traits in Coastal Douglas-fir. I. Timing of Vegetative Bud Flush. *Theoretical Applied Genetics* 102:1142-1151.

Ma, C., R. Meilan, A.M. Brunner, J.A. Carson, J. Li, and S.H. Strauss. 2001. Activation Tagging in Poplar: Frequency of Morphological Mutants. Poster presented at Tree Biotechnology in the New Millennium, July 2001, Stevenson, WA. http://www.fsl.orst.edu/tgerc/iufro2001/poster_abstracts.pdf (accessed April 4, 2002)

Mackay, T.F.C. 2001. The Genetic Architecture of Quantitative Traits. *Annual Reviews Genetics* 35:303-339.

Matzke, M.A., A.J.M. Matzke, G.J. Pruss, and V.B. Vance. 2001. RNA-Based Silencing Strategies in Plants. *Current Opinions Genetics Development* 11:221-227.

Meilan, R., D.J. Auerbach, C. Ma, S.P. DiFazio, and S.H. Strauss. 2001. Stability of Herbicide Resistance and *GUS* Expression in Transgenic Hybrid Poplars (*Populus* sp.) During Several Years of Field Trials and Vegetative Propagation. *HortScience* 37 (2):1-4.

NRC (National Research Council). 2002. *Environmental Effects of Transgenic Plants: The Scope and Adequacy of Regulation.* Washington, DC: National Academy Press.

Merkle, S.A., and J.F.D. Dean. 2001. Forest Tree Biotechnology. *Current Opinion Biotechnology* 11:298-302.

Nilsson, O., E. Wu, D.S. Wolfe, and D. Weigel. 1998. Genetic Ablation of Flowers in Transgenic *Arabidopsis* 15:799-804.

Paterson, A.H., Y.-R. Lin, Z. Li, K.F. Schertz, J.F. Doebley, S.R.M. Pinson, S.-C. Liu, J.W. Stansel, and J.E. Irvine. 1995. Convergent Domestication of Cereal Crops by Independent Mutations at Corresponding Genetic Loci. *Science* 269:1714-1717.

Pena, L., M. Martin-Trillo, J. Juarez, J. Pina, L. Navarro, and J.M. Martinez-Zapater. 2001. Constitutive Expression of *Arabidopsis LEAFY* or *APETALA1* Genes in Citrus Reduces their Generation Time. *Nature Biotechnology* 19:263-267.

Quackenbush, J. 2001. Computational Analysis of Microarray Data. *Nature Reviews: Genetics* 2:418-427.

Rishi, A.S., N.D. Nelson and A. Goyal. 2001. Molecular Farming in Plants: A Current Perspective. *Journal of Biochemistry and Biotechnology* 10:1-12.

Rottmann, W.H., R. Meilan, L.A. Sheppard, A.M. Brunner, J.S. Skinner, C. Ma, S. Cheng, L. Jouanin, G. Pilate, and S.H. Strauss. 2000. Diverse Effects of Overexpression of *LEAFY* and *PTLF*, a Poplar (*Populus*) Homolog of *LEAFY/FLORICAULA*, in Transgenic Poplar and *Arabidopsis*. *Plant Journal* 22:235-246.

Schnable, P., and R.P. Wise. 1998. The Molecular Basis of Cytoplasmic Male Sterility and Fertility Restoration. *Trends Plant Science* 3:175–180.

Silverstone, A.L. and T.-P. Sun. 2000. Gibberellins and the Green Revolution. *Trends Plant Science* 5 (1):1-2.

Springer, P.S. 2000. Gene Traps: Tools for Plant Development and Genomics. *Plant Cell* 12:1007-1020.

Strauss, S.H., M.M. Campbell, S.N. Pryor, P. Coventry, and J. Burley. 2001. Plantation Certification and Genetic Engineering Research: Banning Research is Counterproductive. *Journal of Forestry* 99 (12):4-7.

Tax, F.E., and D.M. Vernon. 2001. T-DNA-Associated Duplication/Translocations in *Arabidopsis*. Implications for Mutant Analysis and Functional Genomics. *Plant Physiology* 126:1527-1538.

Tenaillon, M.I., M.C. Sawkins, A.D. Long, R.L. Gaut, J.F. Doebley, and B.S. Gaut. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *Mays* L.). *Proceedings of the National Academy of Science* 98:9161-9166.

Verbsky, M.L. and E.J. Richards. 2001. Chromatin Remodeling in Plants. *Current Opinion in Plant Biology* 4:494-500.

Weigel, D., J.H. Ahn, A. Blazquez, J.O. Borevitz, S.K. Christensen, C. Fankhauser, C. Ferrandiz, I. Kardailsky, E.J. Malancharuvil, and M.M. Neff. 2000. Activation Tagging in *Arabidopsis*. *Plant Physiology* 122:1003-1013.

Wesley, S.V., C.A. Helliwell, N.A. Smith, M. Wang, D.T. Rouse, Q. Liu, P.S. Gooding, S.P. Singh, D. Abbott, P.A. Stoutjesdijk, S.P. Robinson, A.P. Gleave, A.G. Green, and P.M. Waterhouse. 2001. Construct Design for Efficient, Effective and High-Throughput Gene Silencing in Plants. *Plant Journal* 27:581-590.

Williamson, A.R. 2001. Gene Patents: Socially Acceptable Monopolies or an Unnecessary Hindrance to Research? *Trends Genetics* 17:670-673.

Wyrick, J.J., and R.A. Young. 2002. Deciphering Gene Expression Regulatory Networks. *Current Opinion in Genetics and Development* 12:130-136.

Zobel, B., and J. Talbert. 1984. *Applied Forest Tree Improvement.* New York: John Wiley and Sons.