Critical Reviews in Plant Science, 28:285–308, 2009

ISSN: 0735-2689 print / 1549-7836 online
DOI: 10.1080/07352680903241014

Taylor & Francis
Taylor & Francis Group

Poplar Genomics: State of the Science

Xiaohan Yang,[1] Udaya C. Kalluri,[1] Stephen P. DiFazio,[2] Stan D. Wullschleger,[1] Timothy J. Tschaplinski,[1] (Max) Zong-Ming Cheng,[3] and Gerald A. Tuskan[1]

[1]*Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA*
[2]*Department of Biology, West Virginia University, Morgantown, West Virginia 26506, USA*
[3]*Department of Plant Sciences, University of Tennessee, Knoxville, Tennessee 37996, USA*

Table of Contents

Address correspondence to Xiaohan Yang, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA. E-mail: yangx@ornl.gov

Recent advances in *Populus* genomics have greatly expanded its popularity as a reference for fundamental as well as applied research in woody plants. In this review, we provide an overview of the state-of-the-science in *Populus* genomics research, including experimental and computational genomics. We have surveyed and summarized the following: 1) pioneering as well as more recent reports of genetics- and genomics-based investigations in *Populus*, 2) the positive impact of technological improvements, 3) findings from phylogenetic analyses of gene families, and 4) genomic databases. In the area of *Populus* experimental genomics, genetic approaches have been advanced to the genome scale with resolution to the gene and/or single nucleotide level. On the other hand, the modern "omics" approaches have been successfully applied to analysis of gene function, such as transcriptome profiling using microarrays as well as the next-generation DNA sequencing technology, and characterization of proteome and metabolome using modern instruments. In the area of *Populus* computational genomics, significant progresses have been made in sequence-based discovery of predicted gene function, comparative analysis of gene families, development of genomic databases, and studies of the evolutionary dynamics at both the gene and genome level. While significant advancements have been achieved in *Populus* genome-based science, several challenges need to be addressed, such as 1) better annotation of the *Populus* genome, 2) robust technology for large-scale molecular profiling, 3) efficient system for genome-wide mutagenesis, and 4) high-performance computational pipelines to keep up with the pace of the rapid accumulation of data and to integrate "omics" data into functional systems biology platforms.

Keywords *Populus*, genomics, bioinformatics, gene expression, protein, metabolomics, evolution, microRNA, microarray, gene family, systems biology

I. INTRODUCTION

Populus plants are fast-growing angiosperm trees. Owing to the perennial growth habit and wide-ranging habitat, *Populus* serves as a model for ecological genomics as well as a reference for fundamental scientific investigations of wood formation, secondary cell wall development, and morpho-physiological changes associated with seasonal variations. *Populus* has been a subject of intensive research for its end use in timber and paper-pulp industries, and recently it also garnered worldwide recognition as an important model bioenergy crop. Since the U.S. Department of Energy (DOE) announced plans to sequence the *Populus* genome in early 2002 (Wullschleger *et al.*, 2002), genomics research in *Populus* has been greatly accelerated, as reflected by the number of genomics-related papers published on *Populus* (Fig. 1). The most important hallmark of *Populus* genomics research was the publication of the *Populus* genome sequence (Tuskan *et al.*, 2006), which has already been cited more than 380 times. The availability of the *Populus* genome sequence enabled the application of high-throughput genomics technology and facilitated comparative and evolutionary genomics studies, solidifying the role of *Populus* as a reference organism for tree biology. The present review provides an overview of the state-of-the-science in *Populus* genomics research, including both experimental and functional genomics as well as the newly emerged field of computational genomics.

II. EXPERIMENTALLY-BASED FUNCTIONAL GENOMICS

A. Genetic Approaches

A primary goal of functional genomics is to identify the molecular and genetic bases of phenotypes. In *Populus* functional genomics research, genetic approaches such as random mutagenesis, identification of quantitative trait loci (QTL), characterization of nucleotide polymorphisms, and creation of transgenic lines with up- or down-regulated gene expression have helped link genetic loci and/or genes to phenotypes.

1. Mutagenesis

Insertional mutagenesis is complicated for organisms with long generation times like trees because of the difficulty in creating homozygous lines. Therefore, efforts have focused on

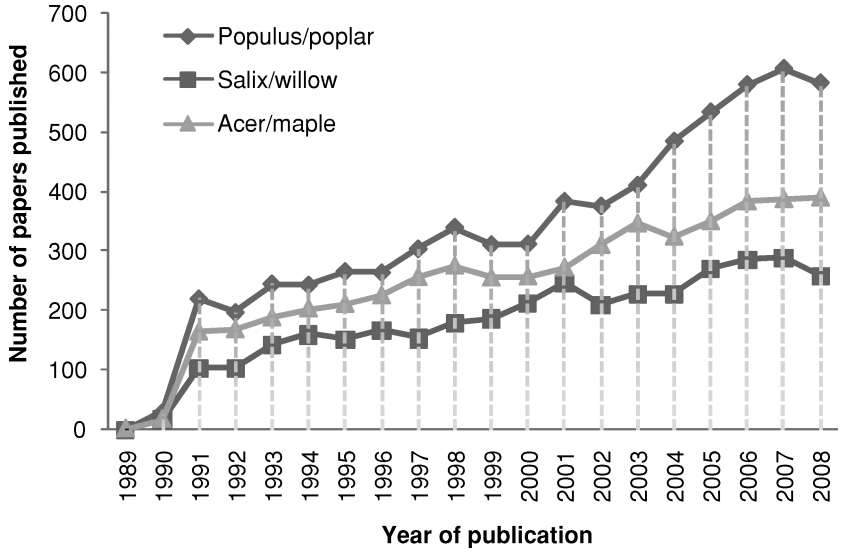

FIG. 1. Number of scientific papers related to genomics in *Populus*, *Salix*, and *Acer* published during a 20-year period from 1989 to 2008.

producing dominant mutations that can be observed in the first generation. One such approach is activation tagging, which was first developed in *Nicotiana tabacum* by Hayashi *et al.* (1992) and successfully applied to *Arabidopsis* by Weigel and colleagues (Weigel *et al.*, 2000). The general procedure of activation tagging includes: 1) introducing an enhancer cassette (typically four copies of the cauliflower mosaic virus 35S enhancer) into a plant genome by *Agrobacterium*-mediated transformation; 2) screening transgenic plants for mutants that exhibit interesting phenotypes; 3) identifying candidate gene(s) in the vicinity of the insertion site using either tail-PCR or plasmid rescue; and 4) functional characterization of the candidate gene(s) by knockdown and/or overexpression. Busov *et al.* (2003) generated 627 independent activation-tagged *Populus* lines, of which nine exhibited an obvious morphological phenotype. Recently, Harrison *et al.* (2007) produced approximately 1,800 independent activation-tagged *Populus* lines. Of the first 1,000 lines screened for developmental abnormalities, 2.4% exhibit alterations in leaf and stem structure as well as overall stature, mostly representing new phenotypes that have not previously been identified in *Populus* and/or other plants. While the activation-tagged lines provide a solid genetic basis for elucidating gene function, they are far less than required to characterize the majority of *Populus* genes, as the genome contains over 45,000 genes (Tuskan *et al.*, 2006). Therefore, a more efficient system needs to be developed to achieve saturation mutagenesis of the *Populus* genome, or substantially more resources need to be invested in activation tagging.

2. *Marker-Based Approaches*

QTLs are genomic regions associated with quantitative traits, such as yield, wood quality and resistance to abiotic and biotic stresses. QTL analyses have been invaluable in forest trees because of the insights they provide about the genetic architecture of quantitative traits (Wu and Lin, 2006). In theory, mapping of QTLs also provides the foundation for marker-assisted selection and gene discovery via positional cloning. However, several factors in forest trees work against the realization of the promise of QTLs. First, due to the aforementioned long generation times that preclude advanced generation breeding, the essential tools for positional cloning such as recombinant inbred lines and near isogenic lines are not practical in forest trees. Second, the low levels of linkage disequilibrium typical of most forest tree populations erode marker-trait associations in unstructured populations, and therefore inhibit applications of marker-assisted selection (Strauss *et al.*, 1992; Neale and Savolainen, 2004). Nevertheless, substantial insights have been gained by identifying genomic regions associated with adaptive traits, and availability of the map-anchored *Populus* genome sequence opens the possibility of linking QTLs to candidate genes, thereby linking genetic and genomic approaches to elucidating gene functions (Yin *et al.*, 2008).

Early attempts at QTL mapping in *Populus* focused on morphological and productivity-related traits in an inbred F_2 mapping population (Family 331) derived from a cross between *P. trichocarpa* 93–968 and *P. deltoides* ILL-129 (Bradshaw *et al.*, 1994), the two taxa from which most commercially-important hybrid clones have been derived. These attempts provided the first views of genetic control of growth and development, revealing a relatively small number of loci with major effects controlling growth and development (Bradshaw and Stettler, 1995; Wu, 1998). This outcome can be attributed in part to a bias against detecting loci with small effects (Beavis, 1998). QTL mapping also helped deconvolute the genetic bases of hybrid vigor in *Populus*, revealing, for example, that superior height growth was primarily conferred by the *P. trichocarpa* parent while enhanced diameter growth was provided by alleles derived from the *P. deltoids* parent (Bradshaw and Stettler,

1995). Similar observations extended to leaf morphology (Wu *et al.*, 1997) and stomatal density (Ferris *et al.*, 2002). More recent studies have largely upheld these initial findings and extended them to multiple families and environments. For example, Rae *et al.* (2008) recently performed a QTL analysis to develop new high-yield genotypes for wide-scale planting and identified 82 QTLs for eight stem and biomass traits. Directly dealing with one of the major factors limiting the deployment of QTLs in breeding, they assessed the stability of these QTLs across three contrasting environments, and found a number of "stability QTLs" that appeared at all sites (Rae *et al.*, 2008).

Important insights have also been gained from the relative positions of QTLs in *Populus*. For example, it is commonly observed that multiple related traits map to the same location in the genome, suggesting that some loci have pleiotropic effects on *Populus* development and productivity (Bradshaw and Stettler, 1995; Wullschleger *et al.*, 2005; Rae *et al.*, 2008). Co-location of QTLs for biomass production and for production of sylleptic branches (those originating from meristems produced within the same growing season) provided some evidence for the importance of sylleptic branching in heterosis of hybrid clones (Bradshaw and Stettler, 1995), a hypothesis that has since been upheld by physiological studies (Scarascia-Mugnozza *et al.*, 1999). Conversely, lack of co-location has also provided important insights. For example, Wullschleger *et al.* (2005) found separate QTLs for production of above- and below-ground biomass, suggesting separate genetic control for these components of productivity.

One of the more striking cases of map-based inferences about phenotypic traits is the discovery of a possible incipient sex chromosome in *Populus*. Although *Populus* is typically dioecious, with separate male and female sexes, morphologically distinct sex chromosomes are not apparent, leading to speculation that sex is determined by multiple autosomal loci (Alstrom-Rapaport *et al.*, 1998). However, multiple studies mapped sex determination to the same chromosomal region recently. Gaudet *et al.* (2008) applied the pseudo-test-cross strategy for linkage analysis of a F_1 pedigree obtained by crossing two genotypes of *P. nigra* from contrasting natural Italian populations, and mapped sex determination to linkage group XIX of the male parent map. Markussen *et al.* (2007) independently mapped sex determination to a similar position in a hybrid F_1 aspen (*P. tremula* x *P. tremuloides*) pedigree, demonstrating the generality of the genetic basis of this trait in the *Populus* genus. Yin *et al.* (2008) mapped sex determination to the same region in a *P. trichocarpa* x *P. deltoides* pedigree, and combined genetic, genomic, and mapping information to demonstrate that this region of linkage group XIX displays multiple characteristics of a sex chromosome, including haplotype divergence, and suppressed recombination. They suggested that *Populus* is in the process of evolving a ZW system of sex determination, with the females containing heteromorphic sex chromosomes (Yin *et al.*, 2008).

QTL studies in *Populus* have also been extended to understand plant adaptation and to pinpoint likely evolutionary responses to future climatic change. A series of experiments have examined the above- and below-ground responses of Family 331 to ambient CO_2 and elevated CO_2. Ferris *et al.* (2002) identified a total of 18 QTLs related to leaf traits, including leaf size and stomatal density. Rae *et al.* (2007) identified three QTLs on linkage groups I, IX and XII for aboveground growth and another three QTLs on linkage groups IV, XVI and XIX for root growth. To find QTLs for drought stress responses in the *Populus* genome, Street *et al.* (2006) mapped 25 QTLs in family 331 under controlled conditions, including 44 in drought. Tschaplinski *et al.* (2006) performed a large-scale drought treatment on family 331 and identified seven QTLs for osmotic potential, and two of these QTLs were later confirmed in approximately the same genomic positions by Street *et al.* (2006), who mapped 25 QTLs for physiological data in family 331 under control conditions and 44 QTLs under drought.

Progress has also been made in mapping QTLs for biotic interactions in *Populus*. In order to map the QTLs controlling resistance to leaf rust, Jorge *et al.* (2005) evaluated *P. deltoides* x *P. trichocarpa* F_1 progeny for quantitative resistance to *Melampsora larici-populina*, and detected nine QTLs, of which two had large, broad-spectrum effects. Interestingly, one of these QTLs is co-located with a QTL for formation of an ectomycorrhizal symbiosis with the fungus *Laccaria bicolor* (Tagu *et al.*, 2005). The availability of genome sequences for both of these fungi promises exciting developments in the characterization of these biotic interactions in the coming years (Martin *et al.*, 2004; Martin *et al.*, 2008; Whitham *et al.*, 2008).

One of the challenges of the 'omics era is the integration of different types of genome-scale data sets. One promising approach is the integration of metabolite profiles with QTL analysis to reveal loci that control complex metabolic pathway of closely related compounds. Tschaplinski *et al.* (2005) proposed combining metabolite profiling with QTL analysis as a novel approach to identify metabolite (m)QTL (i.e., loci that control metabolite abundance). In a preliminary assessment, the metabolite concentrations of fine roots of progeny of an interspecific backcross between *P. trichocarpa* × *deltoides* '52–225' × *P. deltoides* 'D124' were subjected to QTL analysis. mQTLs were identified for a number of secondary metabolites, including a large-effect mQTL that explained >10% of the phenotypic variation in the concentration of trichocarpinene, a secondary metabolite, and its glucoside, trichocarpin. The approach was further validated by Morreel *et al.* (2006), who identified mQTLs that control flavonoid biosynthesis in two F_1 families, *P. deltoides* cv. S9–2 x *P. nigra* cv. Ghoy and *P. deltoides* cv. S9-2 x *P. trichocarpa* cv. V24. Based on multi-trait and single-trait analyses, they identified one mQTL that explained 24% of the variation in the concentration of pinobanksin 3-acetate, one mQTL that explained 19% and 14% of the variation in the concentrations of quercetin and quercetin 3-methyl ether, respectively, and one mQTL that controlled the level of an unknown flavanone.

A similar approach relies on integration of whole-genome microarray analysis with QTL mapping. Kirst *et al.* (2005)

developed an eQTL approach to map cis- or transregulatory elements in *Eucalyptus*. Recently, Kirst and colleagues performed eQTL analysis in a *P. trichocarpa* x *P. deltoides* pseudo-backcross pedigree to identify transcriptional networks and their regulators across multiple tissues (Benedict *et al.*, 2009). While the eQTL approach looks very promising, it has not yet been widely used in *Populus*, due to a lack of cost-effective, high-performance technology for transcriptome profiling.

Due to limited recombination events in the genome, traditional QTL analysis cannot achieve a resolution at the gene and/or single nucleotide level. A promising alternative to traditional QTL analysis is association mapping based on single nucleotide polymorphisms (SNP), which takes advantage of the low linkage disequilibrium typically observed in unstructured *Populus* populations (Neale and Savolainen, 2004). In one of the first published assays of nucleotide variation in natural *Populus* populations, Gilchrist *et al.* (2006) analyzed SNPs in *Populus* using an ecotilling technique, which detects natural DNA polymorphisms using a mismatch-specific endonuclease. Specifically, they examined DNA variation in nine different genes among individuals from 41 different populations of *P. trichocarpa*, and showed that genes examined varied considerably in their level of variation, from one SNP to more than 23 SNPs per 1000-bp region. Ingvarsson (2008) demonstrated the feasibility of association studies in *P. tremula* by analyzing nucleotide polymorphism and linkage disequilibrium using multi-loci data from 77 sequence fragments, 550 bp on average. An approximate Bayesian computation was used to evaluate a number of different demographic scenarios and to estimate parameters for the best-fitting model. His analysis showed that *P. tremula* harbors substantial nucleotide polymorphism across loci and recombination rates are likely to be two to ten times higher than the mutation rate (Ingvarsson, 2008). In the first demonstration of a phenotypic association with candidate gene polymorphisms, Ingvarsson *et al.* (2008) surveyed SNPs at the phytochrome B_2 locus and identified two nonsynonymous SNPs that were independently associated with variation in the timing of bud set, explaining 1.5–5.0% of the observed phenotypic variation. These studies indicate that association studies could be a very powerful genetic approach to delineate molecular mechanisms at the single nucleotide level. An ideal resource for association studies would be a collection of genome sequencing data from each individual plant in a genetic population, which is an intimidating task due to the high cost of current genome sequencing technology. However, the future development of low-cost high-throughput $1,000 genome sequence technology (Service, 2006) would unleash the potential of association studies in *Populus* for genome-scale analysis with resolution at the single nucleotide level.

3. Transgenic Manipulation of Candidate Genes

In general, verification of gene function involves transformation to over-express and/or knockdown candidate genes in transgenic plants. One of the features that sets *Populus* apart from other tree species is the ease with which it can be manipulated in tissue culture (Taylor, 2002). Consequently, *Populus* was one of the first forest trees to be routinely transformed using *Agrobacterium*, and protocols have been developed for transformation of multiple species (Han *et al.*, 1996). Recently, Ma *et al.* (2004) and Song *et al.* (2006) independently established *Agrobacterium*-mediated transformation systems for the sequenced *P. trichocarpa* genotype Nisqually-1, in which *A. tumefaciens* C58 produced transgenic calli with regeneration efficiency of up to 13% five months after co-cultivation. More recently, Cseke *et al.* (2007) developed a transformation system for *P. tremuloides* that is suitable for high-throughput transformations using *A. tumefaciens*. This system uses *Agrobacterium*-inoculated aspen seedling hypocotyls followed by direct thidiazuron-mediated shoot regeneration on selective media, allowing fully formed transgenic trees to be generated in only three to four months.

Genetic transformation has been a primary tool for determination of gene function and genetic improvement in *Populus*, and many studies have been published since the technique was first used in the mid 1980s. One of the primary targets has been elucidation of the mechanisms of cell wall formation and lignin biosynthesis, due to the commercial importance of these traits. In the late 1990s, transgenic *P. tremuloides* plants were produced to down-regulate a caffeic acid O-methyltransferase (CAOMT) gene by homologous sense suppression (Tsai *et al.*, 1998) and a 4-coumarate:coenzyme A ligase gene by antisense inhibition (Hu *et al.*, 1999). Downregulation of cinnamoyl-CoA reductase (CCR) was achieved by Leple *et al.* (2007) in transgenic *P. tremula* x *P. alba* hybrid using antisense and sense construct via an *A. tumefaciens* procedure, with the levels of target transcript reduced down to 3 to 4% of wild-type levels. They showed that the downregulation of CCR was associated with up to 50% reduced lignin content, an orange-brown, often patchy, coloration of the outer xylem, reduced biosynthesis, and increased breakdown or remodeling of non-cellulosic cell wall polymers. Up to now, genetic engineering in *Populus* has been largely restricted to manipulation of one gene at a time. However, biological pathways generally involve multiple genes. Thus, to study gene functions at the pathway level in *Populus*, efficient and stable genetic engineering systems for manipulating multiple genes at a time need to be established in the future.

B. "Omics" Approaches

Compared to genetic approaches, "omics" approaches are relatively new in *Populus* functional genomics research. With the emergence of new life science technologies, high-throughput approaches have been utilized in *Populus* functional genomics studies, such as gene expression analysis using large-scale sequencing of expressed sequence tags (EST), microarray expression studies, protein and metabolite analysis using state-of-the-art instruments, and genome-wide identification of microRNAs using modern sequencing techniques.

1. Transcriptome Sequencing

EST profiling plays important roles in functional genomics efforts such as gene discovery, genome annotation, cDNA microarray design, and *in silico* transcript profiling. The pioneering effort in *Populus* EST sequencing was spearheaded by the Umea Plant Sciences Center in Sweden, with a primary motivation of gaining an understanding of the molecular mechanisms of cell wall formation (Sterky *et al.*, 1998). This led to the establishment of the first comprehensive public *Populus* EST source, consisting of 102,019 ESTs clustered into 11,885 clusters and 12,759 singletons, generated from 19 cDNA libraries each originating from different tissues (Sterky *et al.*, 2004). In another study aimed at understanding wood formation, Andersson-Gunneras *et al.* (2006) sequenced 5,723 ESTs from cellulose-enriched tension wood forming tissues in a *P. tremula* x *P. tremuloides* hybrid. To understand the molecular bases of the enhanced growth caused by downregulation of an enzyme in the lignin biosynthetic pathway, 4-coumarate:coenzyme A ligase in transgenic *P. tremuloides*, Ranjan *et al.* (2004) sequenced 11,308 ESTs from shoot apex, young leaf, young stem and root tip libraries, enriched by a PCR-based suppression subtractive hybridization between control and transgenic plants. The ESTs were clustered and assembled into 5,028 nonredundant transcripts, with a large number of ESTs (16%) associated with signal transduction in transgenic leaves, as well as some homologs of transposable elements upregulated in transgenic tissues. This effort was extended with 5,410 additional ESTs from two hybrids of additional *Populus* species, *P. angustifolia* and *P. fremontii*, which differ markedly in their phenylpropanoid profiles (Harding *et al.*, 2005).

EST sequencing has also been used to gain insights into responses to abiotic and biotic stresses in *Populus*. Nanjo *et al.* (2004) sequenced over 30,000 ESTs from *P. nigra* leaves treated with dehydration, chilling, high salinity, heat, ABA or H_2O_2, and discovered over 4,500 nonredundant full-length cDNAs that were responsive to these stresses. To facilitate gene discovery related to water and nutrient uptake and assimilation in roots, Kohler *et al.* (2003) sequenced 7,013 ESTs representing 4,874 unique transcripts (1,347 clusters and 3,527 singletons) in the roots of *P. trichocarpa* x *P. deltoides*, with 6% of the ESTs assumed to be associated with root functions. Brosché *et al.* (2005) sequenced 13,838 ESTs from 17 libraries derived from *P. euphratica* trees exposed to a variety of stresses. This study was particularly interesting because *P. euphratica* is a unique species in the genus, occurring in desert environments where it is exposed to drought and salinity stress, in stark contrast to the mesic environments where most *Populus* species occur. This study resulted in the identification of 7,841 unigene clusters, 26% of which were novel *Populus* transcripts (Brosche *et al.*, 2005).

To characterize inducible defenses against insect herbivory in *Populus*, Ralph *et al.* (2006) developed an EST resource as a complement of the existing *Populus* genome sequence and *Populus* ESTs by focusing on herbivore- and elicitor-treated tissues and incorporating normalization methods to capture rare transcripts. They generated 139,007 3′- or 5′-end sequenced ESTs from 15 cDNA libraries and assembled the 107,519 3'-end ESTs into 14,451 contigs and 20,560 singletons, representing 35,011 putative unique transcripts. This resource was recently expanded using full-length cDNAs derived from insect-attacked leaves of the *P. trichocarpa* x *P. deltoides* hybrid (Ralph *et al.*, 2008).

Other *Populus* EST resources have been created primarily to enhance gene prediction and annotation. Unneberg *et al.* (2005) created nine cDNA libraries from *P. tremula* and *P. trichocarpa*, and sequenced 70,747 ESTs that were clustered into 14,213 putative genes. Nanjo *et al.* (2007) sequenced female *P. nigra* var. *italica* full-length cDNA libraries, and generated about 116,000 5′- end or 3′-end ESTs that were assembled into 19,841 nonredundant full-length clones. Another 4,664 full-length cDNAs were generated recently by Ralph *et al.* (2008) using the biotinylated CAP trapper method from xylem, phloem and cambium, green shoot tips, and leaves from the *P. trichocarpa* genotype Nisqually-1.

Using public *Populus* EST resources, Moreau *et al.* (2005) performed *in silico* transcript profiling in the woody tissues in relation to programmed death of xylem fibers and identified a large number of previously uncharacterized transcripts possibly related to the death of xylem fibers. To identify expression of genes encoding carbohydrate-active enzymes in the *P. trichocarpa*, Geisler-Lee *et al.* (2006) compared EST frequencies in a collection of 100,000 ESTs from 17 different tissues, and showed that genes involved in pectin and hemicellulose metabolism were expressed in all tissues, indicating that these genes are essential for the development of cell wall matrix. The EST data also indicated that sucrose synthase genes were highly expressed in wood-forming tissues along with cellulose synthase and homologs of KORRIGAN and ELP1 whereas the expression levels of genes related to starch metabolism were low during wood formation, indicating the preferential flux of carbon to cell wall biosynthesis.

The aforementioned traditional EST sequencing approach is based on vector-cloning, which generally misses a significant portion (∼40%) of transcripts even by sequencing millions of cDNA clones from multiple tissues (Sun *et al.*, 2004; Gowda *et al.*, 2006). It is also expensive due to the use of the Sanger sequencing method. Fortunately, next-generation DNA sequencing technology such as 454 sequencing makes it feasible to directly sequence plant transcriptomes in a cost-effective and efficient manner (Emrich *et al.*, 2007). In a collaboration between Oak Ridge National Laboratory and the DOE Joint Genome Institute, transcriptome sequencing using 454 DNA sequencing technology was performed to profile gene expression associated with dehydration response in *Populus deltoides*. About 2.6 million sequencing reads were obtained from six leaf cDNA libraries. These 454 sequencing reads were mapped to the JGI *Populus* genome browser (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.info.html), providing experimental

support for approximately 50% of the *ab initio* gene models in *Populus*. A set of genes which were upregulated by dehydration were identified by bioinformatic analysis of the 454 sequencing reads using a computational pipeline developed at Oak Ridge National Laboratory. The gene ontology analysis revealed that the genes relevant to drought response were enriched in this gene set (Yang *et al.*, unpublished). The success of this 454 sequencing project demonstrated that sequencing-based profiling is an excellent approach for the quantitative analysis of *Populus* gene expression.

2. *Hybridization-Based Transcript Profiling*

Microarrays provide high-throughput transcript profiling based on hybridization of labeled transcripts to glass slides. The two major classes of arrays are cDNA arrays, based on spotting portions of cDNA clones directly onto slides, and oligonucleotide arrays, which have shorter probes that may be synthesized in situ or spotted. The first generation of *Populus* arrays was a cDNA microarray, the 13K POP1 array, developed by Andersson *et al.* (2004) based on 13,490 unigenes assembled from 36,354 ESTs. The POP1 array was used in transcript profiling across the wood-forming meristem to identify potential regulators of cambial stem cell identity (Schrader *et al.*, 2004), during tension wood formation to identify the genes responsible for the change in carbon flow into various cell wall components (Andersson-Gunneras *et al.*, 2006), to characterize drought responses (Street *et al.,* 2006), in isolated cambial meristem cells during dormancy to better understand the environmental and hormonal regulation of this process (Druart *et al.*, 2007), and in studying the auxin-regulated wood formation process (Nilsson *et al.*, 2008). The second generation *Populus* cDNA microarray, the 25K POP2 array, was developed by Moreau *et al.* (2005) based on 24,735 different cDNA fragments. It was used in transcript profiling in leaf tissue of *P. deltoides* to monitor nocturnal changes in gene expression during leaf growth (Matsubara *et al.*, 2006), in a subset of extreme genotypes exhibiting extreme sensitivity and insensitivity to drought in the mapping population Family 331 (Street *et al.*, 2006), in apical bud formation and dormancy induction (Ruttink *et al.*, 2007), in a transgenic *P. tremula* x *P. alba* hybrid with downregulated expression of cinnamoylcoenzyme A reductase to investigate how CCR downregulation impacted metabolism and the biosynthesis of other cell wall polymers (Leple *et al.*, 2007), and in leaves of free-growing *P. tremula* throughout multiple growing seasons (Sjodin *et al.*, 2008).

Brosché *et al.* (2005) constructed a 6K *P. euphratica* cDNA microarray representing 6,340 unigenes assembled from the EST resource described above. This array was used to assess gene expression in adult *P. euphratica* trees growing in the desert (Brosche *et al.*, 2005; Bogeat-Triboulot *et al.*, 2007) and young plants submitted to a gradually increasing water deficit for four weeks in a greenhouse (Bogeat-Triboulot *et al.*, 2007).

Ralph *et al.* (2006) developed a 15.5K *Populus* cDNA microarray containing 15,496 unigenes assembled from 107,519 3'-end ESTs obtained from 15 cDNA libraries, and utilized it to monitor gene expression in *Populus* leaves in response to herbivory by forest tent caterpillars (*Malacosoma disstria*). The 15.5K *Populus* microarray was also used by Miranda *et al.* (2007) to study the transcriptional response of *P. trichocarpa* x *P. deltoides* hybrid to infection by leaf rust (*Melampsora medusa*).

Multiple whole-genome oligonucleotide microarrays have been developed for *Populus* (reviewed by Tsai *et al.*, 2009). The first whole genome oligonucleotide microarray was designed by Oak Ridge National Laboratory in collaboration with NimbleGen (Madison, WI, USA). This array contained three different 60-mer probes for every predicted gene model in the *P. trichocarpa* genome, as well as nearly 10,000 divergent transcripts from the *P. tremula*, *P. tremuloides*, and *P. alba* EST resources described above (Groover *et al.*, 2006). This array has been used in transcript profiling in *P. trichocarpa* to study expression of invertase genes (Bocock *et al.*, 2008), in 14 different tissues of *P. trichocarpa* genotype Nisqually-1 to study expression of auxin response regulators genes (Kalluri *et al.*, 2007), in vegetative organs of the *P. trichocarpa* genotype Nisqually-1 for comparative analysis of the transcriptomes of *P. trichocarpa* and *Arabidopsis thaliana* (Quesada *et al.*, 2008), in five different tissues (young leaves, mature leaves, nodes, internodes and roots) of *P. trichocarpa* Nisqually-1 to study the expression pattern of the cytokinin response regulator gene family (Ramirez-Carvajal *et al.*, 2008), and in *Populus* leaves upon infection with compatible and incompatible strains of the foliar rust *Melampsora larici-populina* (Rinaldi *et al.*, 2007).

Affymetrix has also produced a *Populus* whole-genome array based on their photolithographic fabrication technique. This array targets 61,251 predicted genes, including 47,835 from the *P. trichocarpa* genome sequence, and the remainder representing divergent unigenes predicted from an assembly of over 260,000 ESTs from 13 *Populus* species. The array has eleven 25-mer probes targeting each predicted gene, preferentially selected from the 3′ end of each predicted transcript. This array has been used to characterize responses to nitrogen stress (Qin *et al.*, 2008), and to explore patterns of expression of R2R3 Myb transcription factors in *Populus* (Wilkins *et al.*, 2009).

A third *Populus* whole-genome array has recently been produced by Agilent in a four-plex format targeting 43,803 genes with a single 60-mer probe per gene. This subset of the *Populus* gene models was selected by excluding gene models that showed high homology to transposable elements or to bacterial, fungal, or mammalian sequences that were likely contaminants of the original sequencing template (Tuskan *et al.*, 2006). As of this writing, there are not yet any published studies using this array, but several are in progress (C.J. Tsai, personal communication).

The diversity of *Populus* array designs presents challenges for the integration of data across platforms. A web-based tool, PopArray (http://aspendb.uga.edu) has recently been created to facilitate cross-referencing of probes across platforms (Tsai *et al.*, 2009). However, each of these microarray platforms/systems still suffers from several limitations. First is

a lack of full-genome coverage because the probes were designed according to either limited EST information or the first version of the *Populus* genome annotation, which contains many incomplete gene models. Second, some of the probes are not gene-specific due to high sequence identity of recently duplicated genes in the genome. Finally, splice variants are not well represented. This latter problem is compensated somewhat by a new version of the *Populus* NimbleGen array that has seven independent 60-mer probes per gene target (http://www.nimblegen.com/products/exp/custom.html). Future array designs will benefit from deep transcriptome sequencing data and the release of the next *Populus* genome annotation in the near future.

3. *PCR-Based Transcript Profiling*

Traditional techniques like differential display still have a role to play in the genomics era (Liang, 2002), even in a genus like *Populus* in which a species has been fully sequenced, and large numbers of ESTs are available. These enrichment techniques allow efficient use of sequencing resources, allowing researchers to focus on transcripts that are up- or down-regulated in response to specific treatments. For example, Caruso *et al.* (2008) searched for genes differentially expressed in response to drought in young rooted cuttings under PEG 6000 treatment using the differential display technique and identified 36 differentially expressed leaf cDNAs between stressed and control conditions. Another useful technique is cDNA-amplified fragment length polymorphism (AFLP) transcript profiling. It has been used to map differential gene expression during dormancy induction, dormancy, dormancy release by chilling, and subsequent bud break in apical buds of a *P. tremula* × *P. alba* hybrid, revealing novel genes linked to a crucial transitory step in dormancy induction, and to dormancy release through chilling (Rohde *et al.*, 2007). Finally, Zhuang and Adams (2007) used a clever application of the SNaPshot SNP genotyping assay (Applied Biosystems) to reveal allelic variation in gene expression in a *P. trichocarpa* × *P. deltoides* F_1 hybrid. Using this single-base primer extension assay with extension primers designed to anneal to the amplified DNA adjacent to the SNP site, they identified cis-regulation for six genes, trans-regulation for one gene, and combined cis- and trans-regulation for nine genes, demonstrating that species-specific alleles can have variable expression patterns depending on the genetic background (Zhuang and Adams, 2007).

4. *Protein Profiling*

Protein profiling offers multiple advantages over transcript profiling for characterizing the functional state of an organism at a given point in time. First, a given transcript can give rise to different proteins due to post-translational modifications, which will only be apparent with a direct observation of the protein. Furthermore, mRNA levels may be a poor indicator of protein levels due to differential rates of turnover of mRNA compared to proteins (e.g., a long-lived protein may be in relatively high abundance, even though the mRNA might be in low abundance due to degradation of the transcript) (Pandey and Mann, 2000). Therefore, extensive efforts have been undertaken to characterize protein responses at the whole genome scale in model organisms, including *Populus*.

Two studies have examined the proteomic responses of *P. euphratica* to experimentally-induced heat and drought stress. Ferreira *et al.* (2006) studied protein accumulation profiles of leaves from young *P. euphratica* plants submitted to 42/37°C for three days using two-dimensional electrophoresis. They detected up- or down-regulation of 45% of the 1,355 spots assayed by 2-dimensional gel electrophoresis (2-DE), and identified 51 out of 62 selected spots using matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) analysis. They showed that short-term up-regulated proteins were related to membrane destabilization and cytoskeleton restructuring, sulfur assimilation, thiamine and hydrophobic amino acid biosynthesis, and protein stability; long-term upregulated proteins were involved in redox homeostasis and photosynthesis. Late downregulated proteins were involved mainly in carbon metabolism, indicating that moderate heat response involves proteins related to lipid biogenesis, cytoskeleton structure, sulfate assimilation, thiamine and hydrophobic amino acid biosynthesis, and nuclear transport. Bogeat-Triboulot *et al.* (2007) determined the expression profiles of proteins in mature leaves at four stress levels and after recovery in young, vegetatively propagated *P. euphratica* plants that were submitted to a gradually increasing water deficit for four weeks in a greenhouse and were allowed to recover for ten days after full re-irrigation. Using 2-DE, they identified 375 spots that were responsive to the treatments, and characterized 100 of these using MALDI-TOF-MS and MALDI-TOF/TOF-MS analysis, resulting in identification of 39 drought-responsive proteins, including proteins related to energy and C metabolism, and proteins involved in glycolysis. Surprisingly, protein levels were not correlated with stress levels, in contrast to transcript abundance as determined by microarrays, which were directly correlated with the level of water deficit (Bogeat-Triboulot *et al.*, 2007).

Plomion *et al.* (2006) performed a more comprehensive analysis of the *Populus* proteome. They first created highly reproducible and well-resolved 2-DE maps of proteins for eight tissues/organs of adult *P. trichocarpa* (both male and female) plants and 2-month-old rooted cuttings of a *P. trichocarpa* x *P. deltoides* hybrid. They excised 398 spots from the 2-DE gels and identified 363 proteins (~91.2%) by nanospray LC-MS/MS, based on comparison with 260,000 *Populus* ESTs. In order to assess the resolution of protein detection based on peptide mass fingerprinting (PMF) and compare the identification rate to that obtained by LC-MS/MS, they generated MALDI-TOF-MS profiles for 320 spots and obtained reliable PMFs for only 163 spots (51%), from which about half (83 spots) positively matched gene models from the *P. trichocarpa* genome sequence.

Du *et al.* (2006) analyzed proteins expressed in different wood regeneration stages in a system that can mimic the

initiation and differentiation of cambium cells for *P. tomentosa*. They obtained PMFs for 244 differentially-expressed proteins and assigned putative functions for 199 of these proteins. They showed that regulatory genes for cell cycle progression, differentiation and cell fate were expressed during formation of cambial tissue, while genes involved in secondary wall formation were predominantly found in the xylem developmental stage, indicating that changes in patterns of gene expression correspond to developmental stages of the secondary vascular system.

Sasaki *et al.* (2007) surveyed the localization of anionic peroxidase isoenzymes, which are important for lignification, in various organs of *P. alba* using 2-DE followed by PMF analysis. They showed that the expression profile of each isoenzyme was quite different, suggesting that individual anionic isoenzymes are differently regulated at transcription, translation, or posttranslational level.

Availability of a high-throughput, broad-spectrum proteome profiling method is highly desirable to improve turnaround time and comprehensiveness of a proteomics approach. A shotgun proteome profiling method has recently been reported for *Populus* (Kalluri *et al.*, 2009, in review). In this study, MudPIT technique was applied to profiling subcellular protein fractions of developing xylem. Identification of nearly 6,000 proteins using this method significantly increases the number of proteins validated from *Populus* beyond what was previously reported from 2-DE-based approaches.

5. *MicroRNA Profiling*

MicroRNAs (miRNA) are small RNAs approximately 21 nucleotides in length that negatively control gene expression by cleaving or inhibiting the translation of target gene transcripts (Barakat *et al.*, 2007). To test whether miRNAs play roles in the regulation of wood development in tree species, Lu *et al.* (2005) isolated small RNAs from the developing xylem of *P. trichocarpa* stems and cloned 22 miRNAs, which are the founding members of 21 miRNA gene families for 48 miRNA sequences. Their computational prediction revealed that a majority of these miRNAs potentially target developmental- and stress/defense-related genes. Ko *et al.* (2006) cloned the microRNA 166 (Pta-miR166) families from *P. tremula* × *P. alba* hybrid using a combination of *in silico* and PCR-based methods. They showed the expression of class III HD-Zip transcription factor (PtaHB1) was inversely correlated with the level of Pta-miR166, which directed the cleavage of PtaHB1 *in vivo*, as confirmed using modified 5′-rapid amplification of cDNA ends. They also found that the expression of Pta-miR166 was much higher in the winter than during the growing season, suggesting seasonal and developmental regulation of microRNA. Recently, a high throughput pyrosequencing technique was used by Barakat *et al.* (2007) to profile small RNAs from leaves and vegetative buds of *Populus*. After analysis of 80K small RNA reads, they identified 123 new sequences belonging to previously identified miRNA families as well as 48 new miRNA families that could be *Populus*-specific. They also identified putative targets of nonconserved miRNA including both previously identified targets as well as several new putative target genes involved in development, resistance to stress, and other cellular processes. They showed that almost half of the genes predicted to be targeted by nonconserved miRNAs appear to be *Populus*-specific.

6. *Metabolite Profiling*

The metabolome is the quantitative complement of all of the low-molecular weight molecules present in a cell in a given physiological state, representing the products of cellular biochemical processes. As such, analysis of the metabolome potentially provides an even finer snapshot of an organism's physiological state than proteomic analysis (Fiehn, 2002). Advances in metabolite profiling have proceeded rapidly in the past decade, raising the possibility of simultaneously assaying the levels of thousands of compounds in a high-throughput manner (Weckwerth, 2003). This great promise is beginning to be realized in model organisms like *Populus*, as tools and metabolite databases rapidly accumulate, and metabolomic analyses are being used to dissect phenotypes with unprecedented precision.

A major use of metabolomics is the characterization of biochemical changes that occur following experimental down- or up-regulation of candidate gene expression. For example, Busov *et al.* (2006) performed metabolic profiling to gain insight into the biochemical changes associated with the dramatic architectural changes of the dwarfed transgenic plants overexpressing *Arabidopsis gai* and *rgl1* using gas chromatography–mass spectrometry (GC-MS). They showed that transgenic plants had increased concentrations of citric acid and several amino acids, including asparagine and arginine, and two unidentified glucosides, but reduced concentrations of monosaccharides in roots. The combined responses were indicative of increased respiratory consumption of monosaccharides to generate Krebs cycle organic acids that are required for amino acid synthesis and root growth. In leaves, the concurrent decline in glutamine and other N-containing metabolites, including phenylalanine, was consistent with increased N allocation to roots via perturbations in the secondary carbon pathways. The transgenic dwarf plants displayed increased concentrations of various products and intermediates of the phenylpropanoid biosynthetic pathway, such as the accumulation of syringin (sinapyl alcohol glucoside), likely reflected the reduced shoot growth of the transgenics, leading to a buildup of a storage form of the monolignol precursor. Phenolic glucosides that are associated with defense, including salicin and tremulacin, were similarly found at much higher levels in both *gai* and *rgl1* expressing plants. The accumulation of 3-O-caffeoylquinic acid that results from the conjugation of a key phenolic acid precursor of monolignol biosynthesis with an upstream organic acid intermediate of the shikimic acid pathway, coupled with declines of the monomers including quinic acid and other phenolic acid conjugates, may have been indicative of the reduced carbon flux through the lignin biosynthetic pathway. A similar shift in carbon partitioning away from the

upstream lignin precursor conjugates was revealed by metabolite profiling of poplar transgenic plants expressing a bacterial nahG gene encoding salicylate hydroxylase that converts salicylic acid to catechol (Morse *et al.*, 2007). Expression of nahG decreased quinic acid-phenolic acid conjugates and phenolic acid-glucosides, including salicylate glucoside, but increased catechol glucoside, while exerting little effect on levels of salicylic acid and catechol, the substrate and product, respectively, of the nahG enzyme.

To investigate the effects of downregulating cinnamoyl-CoA reductase (CCR) on metabolism in transgenic *P. tremula* x *P. alba*, Leple *et al.* (2007) analyzed metabolome of young developing xylem of wild-type and CCR-downregulated lines using GC-MS followed by principal component analysis. They identified 20 known metabolites that were accumulated differentially in the CCR-downregulated lines compared with the wild type, with the largest fraction of differential metabolites being carbohydrates such as glucose, mannose, galactose, myo-inositol, raffinose, and melezitose, reflecting changes in central carbohydrate metabolism. Most notably, CCR-downregulated mutants had large accumulations in 4-O-β-D-glucopyranosyl sinapic acid and 4-O-β-D-glucopyranosyl vanillic acid. These were the same phenolic acid glucosides that accumulated in caffeoyl-CoA O-methyltransferase (CCoAOMT) mutants (Meyermans *et al.*, 2000). To identify the metabolites that control bud development in *P. tremula* x *P. alba*, Ruttink *et al.* (2007) analyzed developing buds of wild-type and transgenic plants that upregulate or downregulate the ABI3 transcription factor at weekly intervals during 6 weeks of short-day treatment with GC-MS, and quantified 8,852 m/z peaks, of which 1,702 m/z peaks have significant changes either between any two genotypes at a given time, or between any two sampling points for a given genotype. The 1,702 m/z peaks corresponded to 176 metabolites, of which 162 had more than a fourfold differential accumulation, including 110 unidentified compounds (67.9%), 13 organic acids (8.0%), 16 amino acids (9.9%), and 14 sugars or sugar alcohols (8.6%).

Metabolic profiling has also been a valuable tool for characterizing developmental processes in *Populus*. For example, Andersson-Gunneras *et al.* (2006) performed metabolite analysis in developing *P. tremula* tension wood using gas chromatography/time-of-flight mass spectrometry. They detected more than 350 peaks and revealed that 26 metabolites were significantly changed, among which sucrose, arabinose, inositol, shikimate, monolignols and gamma-butyric acid were decreased in tension wood, while xylose, xylitol and two fatty acid metabolites were more abundant in tension wood. Gou *et al.* (2008) analyzed the cell-wall acylesters of *P. trichocarpa* with liquid chromatography-mass spectrometry, Fourier transform-infrared microspectroscopy, and synchrotron infrared imaging facility. They showed that the cell wall of *Populus* contained a considerable amount of acylesters, primarily acetyl and p-hydroxycinnamoyl molecules and the "wall-bound" acetate and phenolics displayed a distinct tissue specific-, bending stress responsible- and developmental-accumulation pattern, indicating that different "wall-bound" acylesters play distinct roles in *Populus* cell wall structural construction and/or metabolism of cell wall matrix components. To characterize the environmental and hormonal regulation of dormancy in perennial plants, Druart *et al.* (2007) performed metabolite profiling of isolated cambial meristem cells during the course of their activity-dormancy cycle using GC-MS analysis. They detected more than 1,000 peaks, of which 227 changed significantly in one or more pairwise sample class comparisons. The responsive metabolites were classified into four groups: carbohydrates, organic acids, amines (amines and amino acids) and sterols.

7. *"Omics" Application in Wood Formation and Secondary Cell Wall Formation*

The occurrence of extensive secondary xylem development (wood formation) where xylem developmental phase transitions can be distinguished spatially across various radial growth layers makes/renders *Populus* as an excellent model to study molecular dynamics that underlie the processes of secondary cell wall formation and wood development. Several studies based on targeted expression profiling techniques of RT-PCR and in-situ hybridization (Kalluri and Joshi, 2004) as well as broad spectrum expression profiling based on ESTs (Sterky *et al.*, 2004) and microarray technologies (Nilsson *et al.*, 2008) have provided insights into genes that are important to secondary xylem development. Some of the relevant gene families have also been reported in greater detail elsewhere in this review, including those involved in carbohydrate biosynthesis process (Geisler-Lee *et al.*, 2006), phenyl propanoid pathway (Tsai *et al.*, 2006), transcription regulation (Arnaud *et al.*, 2007), cell cycle (Espinosa-Ruiz *et al.*, 2004), and structural proteins (Oakley *et al.*, 2007).

III. COMPUTATIONAL GENOMICS

A. Sequence-based Discovery

The availability of transcriptome and genome sequencing data opened the door for new discoveries using computational approaches, such as analysis of differential gene expression using *in silico* transcript profiling, large-scale prediction of cis-elements in genomic regions upstream of the transcription start sites, identification of lineage-specific motifs and genes by comparing gene sequences among different species.

1. *Identification of Putative Cis-regulatory Sequences*

To identify putative cis-regulatory sequences in the cellulose synthase (CesA) gene family which encodes the catalytic subunits of a large protein complex responsible for the deposition of cellulose into plant cell walls, Creux *et al.* (2008) carried out a comparative sequence analysis of orthologous CesA promoters from *Arabidopsis*, *Populus* and *Eucalyptus*, and identified 71 conserved sequence motifs, of which 66 were

significantly over-represented in either primary or secondary wall-associated promoters. Krom and Ramakrishna (2008) studied the co-expression and interspecies conservation of divergent and convergent gene pairs in the *Oryza*, *Arabidopsis* and *Populus* genomes. They showed that strongly correlated expression levels between divergent and convergent genes were quite common in all three species, suggesting that shared as well as unique mechanisms operate in shaping the organization and function of divergent and convergent gene pairs in different plant species. They also identified 56 known regulatory elements overrepresented in the intergenic regions of divergent genes (separated by 1 kb or less) with correlated expression in *Arabidopsis* (39 elements), *Oryza* (16 elements), and *Populus* (1 element) (Krom and Ramakrishna, 2008).

2. Analysis of Alternative Splicing in Transcripts

Baek *et al.* (2008) analyzed the structure of 5′-splice junctions in *Medicago truncatula*, *P. trichocarpa*, *A. thaliana*, and *O. sativa* and observed commonalities between the species. They found that for *M. truncatula*, *P. trichocarpa* and *A. thaliana*, but not in *O. sativa*, alternative splicing was most prevalent for introns with decreased UA content.

3. Identification of Lineage-specific Motifs and Genes

Although proteins lacking currently defined motifs or domains (POFs) appear similar to proteins with experimentally defined domains or motifs (PDFs) in their relative contribution to biological functions, the POFs have more predicted disordered structure than the PDFs, implying that they may exhibit preferential involvement in species-specific regulatory and signaling networks (Gollery *et al.*, 2006). Gollery *et al.* (2007) performed a comparative analysis of POFs and PDFs in the predicted proteomes derived from *Arabidopsis*, *Oryza* and *Populus*, finding that >26% of the *Populus* proteome was comprised of POFs, compared with 19% and 33% of the *Arabidopsis* and *Oryza* proteomes, respectively. In a comparison among the three plant proteomes, ~75% of the unique proteins of *Populus* were POFs. Due to higher proportion of disordered structures in POFs, the shorter length of POFs compared with PDFs, and the low level of sequence similarity between POFs from different organisms, most POFs exist in plants as singletons. It was therefore hypothesized that POFs could represent newly evolving genes or genes that are evolving much faster than the genome average, suggesting that these unique *Populus* proteins could be lynchpins of the evolutionary process in this genus (Gollery *et al.*, 2006; Gollery *et al.*, 2007).

Recently, Yang *et al.* (2009) conducted a genome-wide analysis of lineage specific genes in *Arabidopsis, Oryza* and *Populus* and identified three differential gene (DG) sets: i) 917 *Arabidopsis* genes without homologues in *Oryza* or *Populus*, ii) 2,781 *Oryza* genes without homologues in *Arabidopsis* or *Populus*, and iii) 594 *Populus* genes without homologues in *Arabidopsis* or *Oryza*. Furthermore, they used the DG sets to search against a plant transcript database, NR protein database, NCBI dbEST and six newly sequenced genomes (*Carica*, *Glycine*, *Medicago*, *Sorghum*, *Vitis* and *Zea*) and identified 165, 638 and 109 species-specific genes (SS) genes in *Arabidopsis*, *Oryza* and *Populus*, respectively. They showed that some SS genes were preferentially expressed in flowers, roots, xylem and cambium or upregulated by stress, reflecting functional and/or anatomical differences between monocots and eudicots or between herbaceous and woody plants.

B. Phylogenetic Analysis of Gene Families

Phylogenetic analyses of individual gene families have been playing a very important role in *Populus* genomics since the draft genome sequence became available. Thus far analyses of about 50 gene families containing more than 2,400 genes in *Populus* have been published (Table 1). These gene families are involved in various biological processes, as discussed in the following sections.

1. Transcription Regulation

MADS-box transcription factor genes control diverse developmental processes in flowering plants ranging from root to flower and fruit development in plants (Becker and Theissen, 2003). Leseberg *et al.* (2006) identified 105 putative functional MADS-box genes and 12 pseudogenes in the *Populus* genome, comparable to those in *Arabidopsis* (107 genes) (Parenicova *et al.*, 2003). They constructed a phylogenetic tree using all MADS-box genes from *Arabidopsis* and *Populus* and classified the *Populus* MADS-box genes using the *Arabidopsis* MADS-box gene dataset. Their phylogenetic analysis revealed that *Populus* has 64 type II MADS-box genes, implying a higher birth rate when compared with *Arabidopsis* (64 vs. 47). In contrast, there were 41 putative functional type I genes and 9 type I pseudogenes in *Populus*, suggesting that the *Populus* type I MADS-box genes have experienced a high death rate, but a relatively lower birth rate, leading to a smaller number of type I genes in *Populus* than in *Arabidopsis* (41 vs. 60). They also found that the *Populus* MADS-box gene family has become expanded through tandem gene duplication and segmental duplication events. Recently, Diaz-Riquelme *et al.* (2009) performed a phylogenetic analysis of full-length *Vitis*, *Arabidopsis* and *Populus* MIKC^C-type MADS box protein sequences, dividing the MIKC gene family into 13 subfamilies. All *Vitis* MIKC genes were grouped with their *Arabidopsis* and *Populus* counterparts, and in most cases, two *Populus* genes were found for every homolog in *Vitis* or *Arabidopsis*, consistent with the recent discovery that *Populus* experienced a recent genome-wide duplication event (Tuskan *et al.*, 2006; Tang *et al.*, 2008a; 2008b).

The two related transcription factor gene families, Auxin/Indole-3-Acetic Acid (Aux/IAA) and Auxin Response Factor (ARF), regulate auxin-induced gene expression, each with unique localized functions as well as overlapping redundant functions (Liscum and Reed, 2002). Kalluri *et al.* (2007) identified 35 Aux/IAA and 39 ARF genes in the *Populus* genome.

TABLE 1
Populus gene familes studied by phylogenetics analysis

Category	Gene family	References
Biosynthesis of structural components	alpha- and beta-tubulin	Oakley *et al.* (2007)
Biosynthesis of carbohydrates	Starch branching enzymes	Han *et al.* (2007)
Biosynthesis of carbohydrates	Invertase	Bocock *et al.* (2008)
Biosynthesis of carbohydrates	endo-beta-mannanase	Yuan *et al.* (2007)
Biosynthesis of carbohydrates	Cellulose synthase	Suzuki *et al.* (2006)
Biosynthesis of carbohydrates	Xyloglucan endo-transglycosylases	Baumann *et al.* (2007)
Biosynthesis of carbohydrates	COBRA	Ye *et al.* (2009a)
Biosynthesis of hormone	YUCCA	Ye *et al.* (2009b)
Cell cycle	NIMA-related kinase	Vigneault *et al.* (2007)
Cell cycle	D-type cyclins	Menges *et al.* (2007)
Disease resistance	NBS resistance	Kohler *et al.* (2008)
Fatty acid metabolism	acyl:coenzyme A synthetase	Souza Cde *et al.* (2008)
Flower development	FT/TFL1	Igasaki *et al.* (2008)
Photorespiration	Glycine decarboxylase complex	Rajinikanth *et al.* (2007)
Protein degradation	F-box	Yang *et al.* (2008)
Protein degradation	Kunitz trypsin inhibitor	Major and Constabel (2008)
Protein degradation	Protease	Garcia-Lorenzo *et al.* (2006)
Regulating transcription	AP2/ERF	Zhuang *et al.* (2008)
Regulating transcription	MADS-box	Leseberg *et al.* (2006); Diaz-Riquelme *et al.* (2009)
Regulating transcription	Aux/IAA	Kalluri *et al.* (2007)
Regulating transcription	ARF	Kalluri *et al.* (2007)
Regulating transcription	LIM	Arnaud *et al.* (2007)
Regulating transcription	DOF	Yang *et al.* (2006)
Regulating transcription	Cytokinin response regulator	Ramirez-Carvajal *et al.* (2008)
Regulating transcription	R2R3-MYB	Wilkins *et al.* (2008)
Signal transduction	LysM kinase	Zhang *et al.* (2007)
Signal transduction	Calcineurin B-Like	Zhang *et al.* (2008)
Phenylpropanoid metabolism	Arogenate dehydratase (ADT)	Tsai *et al.* (2006)
Phenylpropanoid metabolism	beta-Alanine N-methyltransferase (NMT)	Tsai *et al.* (2006)
Phenylpropanoid metabolism	Flavonoid 3′,5′-hydroxylase (F3′5′H)	Tsai *et al.* (2006)
Phenylpropanoid metabolism	Flavone synthase II (FNSII)	Tsai *et al.* (2006)
Phenylpropanoid metabolism	Flavanone 3-hydroxylase (F3H)	Tsai *et al.* (2006)
Phenylpropanoid metabolism	Flavonol synthase (FLS)	Tsai *et al.* (2006)
Phenylpropanoid metabolism	Dihydroflavonol 4-reductase (DFR)	Tsai *et al.* (2006)
Phenylpropanoid metabolism	Anthocyanidin synthase (ANS)	Tsai *et al.* (2006)
Phenylpropanoid metabolism	Anthocyanidin reductase (ANR/BAN)	Tsai *et al.* (2006)
Phenylpropanoid metabolism	Leucoanthocyanidin reductase (LAR)	Tsai *et al.* (2006)
Phenylpropanoid metabolism	Flavonoid O-methyltransferase (FOMT)	Tsai *et al.* (2006)
Phenylpropanoid metabolism	trans-Cinnamate 4-hydroxylase (C4H)	Tsai *et al.* (2006); Hamberger *et al.* (2007)
Phenylpropanoid metabolism	Coumarate 3-hydroxylase (C3H)	Tsai *et al.* (2006); Hamberger *et al.* (2007)
Phenylpropanoid metabolism	Caffeic acid O-methyltransferase (COMT)	Tsai *et al.* (2006); Hamberger *et al.* (2007)
Phenylpropanoid metabolism	Ferulate 5-hydroxylase (F5H)	Tsai *et al.* (2006); Hamberger *et al.* (2007)
Phenylpropanoid metabolism	Caffeoyl-CoA O-methyltransferase	Tsai *et al.* (2006); Hamberger *et al.* (2007)
Phenylpropanoid metabolism	hydroxycinnamoyltransferase (HCT)	Tsai *et al.* (2006); Hamberger *et al.* (2007)
Phenylpropanoid metabolism	Aldehyde dehydrogenase (ALDH)	Hamberger *et al.* (2007)
Phenylpropanoid metabolism	Cinnamoyl CoA reductase (CCR)	Hamberger *et al.* (2007)
Phenylpropanoid metabolism	NADPH-cytochrome P450 oxydoreductase	Hamberger *et al.* (2007)
Phenylpropanoid metabolism	Cinnamyl alcohol dehydrogenase-related	Hamberger *et al.* (2007)
Phenylpropanoid metabolism	Phenylalanine amonnia lyase (PAL)	Hamberger *et al.* (2007)
Phenylpropanoid metabolism	4-Coumarate:CoA Ligase (4CL)	Hamberger *et al.* (2007)

From the phylogenetic tree reconstructed from *Populus*, *Arabidopsis* and *Oryza* Aux/IAA amino acid sequences, they found that four groups of *Populus* Aux/IAAs (PoptrIAA3, 16, 27 and 29) expanded to contain three or more members each. Similarly, the phylogeny of *Populus*, *Arabidopsis* and *Oryza* ARF protein sequences revealed differential expansion or contraction between *Arabidopsis* and *Populus*. For example, the number of activator ARFs (defined by the Q-rich middle region) in *Populus* is 2.6 times that in *Arabidopsis* whereas the ratio of repressor and other ARFs between *Arabidopsis* and *Populus* is 1:1.4. Kalluri *et al.* (2007) also showed that the differential expansion or contraction in Aux/IAA and ARF between *Arabidopsis* and *Populus* was caused by high segmental and low tandem duplication events in *Populus*. Furthermore, their expression studies showed that genes in the expanded PoptrIAA3 subgroup display differential expression.

Genes in the AP2/Ethylene Response Factor (ERF) family encode transcriptional regulators with a variety of functions involved in developmental and physiological processes in plants (Okamuro *et al.*, 1997; Nakano *et al.*, 2006). Zhuang *et al.* (2008) identified 200 AP2/ERF genes in the *Populus* genome. According to a phylogeny created from the AP2/ERF domains of 200 AP2/ERF proteins in *P. trichocarpa* and 145 AP2/ERF proteins in *A. thaliana*, they divided the AP2/ERF family into four subfamilies (AP2, DREB, ERF, and RAV), and revealed that two subfamilies (DREB and ERF) were expanded in *P. trichocarpa* in comparison with *A. thaliana*.

The LIM domain is an evolutionarily-conserved double-zinc finger motif found in a variety of proteins exhibiting diverse biological roles that act as modular protein-binding interfaces mediating protein-protein interactions in the cytoplasm and the nucleus (Khurana *et al.*, 2002). In tobacco, one LIM protein (Ntlim1) is a transcription factor affecting gene expression of lignin biosynthesis (Kawaoka and Ebinuma, 2001). Arnaud *et al.* (2007) identified 12 LIM gene models in the *P. trichocarpa* genome and their phylogenetic analysis of 24 LIM domain proteins (12 *P. trichocarpa* + six *A. thaliana* + six *O. sativa* proteins) revealed that plant LIM proteins have undergone one or several duplication events during their evolution. They classified the plant LIM proteins into four groups: alphaLIM1, betaLIM1, gammaLIM2 and deltaLIM2.

Cytokinin is one key hormone regulating many aspects of plant growth and development. Response Regulators (RRs) are transcription factors that function in the final step of the two-component cytokinin signaling system, which involves histidine kinase receptors that perceive cytokinin and transmit the signal via a multistep phosphorelay, with Type-A and Type-B RRs as negative and positive regulators of cytokinin, respectively (Mason *et al.*, 2005; To *et al.*, 2007). Ramirez-Carvajal *et al.* (2008) identified 33 cytokinin RR genes in *Populus*: 11 type As, 11 type Bs, and 11 pseudo-RRs. Their phylogenetic analysis using the conserved receiver domains of RR gene family members in *Populus*, *Arabidopsis* and *Oryza* revealed that a significant number of the RRs (26 in *Populus*, 20 in *Arabidopsis* and 14 in *Oryza*) grouped in species-specific pairs. Their expression analysis showed that *Populus* RR type As and type Bs appear to be preferentially expressed in nodes, while the pseudo-RRs are preferentially expressed in mature leaves.

Plant-specific R2R3-type MYB transcription factors control plant secondary metabolism as well as the identity and fate of plant cells (Stracke *et al.*, 2001). Wilkins *et al.* (2009) performed phylogenetic analysis of the predicted R2R3-MYB protein sequences in *Populus* (192 sequences), *Vitis* (119 sequences), *Arabidopsis* (126 sequences) and other plant species (54 sequences) and divided the R2R3-MYB family into 49 clades (C1 – C49). The phylogeny revealed unequal representation of *Populus*, *Vitis* and *Arabidopsis* R2R3-MYB proteins within individual clades. For example, clades C3 and C25 contained more *Populus* genes, whereas clades C13 and C37 did not include any *Populus* genes. This differential expansion/contraction indicates lineage-specific gene duplication and gene loss. Furthermore, they showed that the expanded *Populus* R2R3-MYB members were associated with wood formation and reproductive development.

2. *Flowering*

The FLOWERING LOCUS T (FT) and TERMINAL FLOWER1 (TFL1) genes function, respectively, as a promoter and a repressor of the floral transition in *Arabidopsis* (Danilevskaya *et al.*, 2008). There is increasing evidence that the FT protein is a major component of flower signal transduction which causes changes in the gene expression that reprograms the shoot apical meristem (SAM) to form flowers instead of leaves (Turck *et al.*, 2008). Igasaki *et al.* (2008) performed phylogenetic analysis of 32 FT/TFL1 proteins in *P. nigra* var. *italica* (nine sequences), *A. thaliana* (seven sequences), tomato (six sequences), *Vitis* (five sequences), apple (three sequences) and citrus (two sequences). Their phylogenetic tree revealed that the genes fall into four different clades: the TFL1 clade, the FT clade, the MOTHER OF FT AND TFL1 clade, and the BROTHER OF FT AND TFL1 clade. Their gene expression and transgenic studies suggest that one *Populus* gene in the TFL1 clade, PnTFL1, represses flowering and two *Populus* genes in the FT clade, PnFT1 and PnFT2, promote flowering.

3. *Signal Transduction*

Calcineurin B-like (CBL) proteins have been implicated as important sensors in signaling of calcium which plays a crucial role as a second messenger in mediating various defense responses under environmental stresses (Gu *et al.*, 2008). Zhang *et al.* (2008) identified 10 CBL candidate genes (PtCBLs) in the *P. trichocarpa* genome. Their phylogenetic analysis of *Arabidopsis*, *Oryza* and *Populus* CBL genes divided the CBL gene family into four groups. Their comparative analyses indicate that the duplication events in *Populus* might have contributed to the expansion of the CBL family. Furthermore, they cloned nine CBL genes (PeCBLs) from *P. euphratica*, a mostly salt- and drought-tolerant *Populus* species, and performed gene

specific RT-PCR analysis which suggests that seven CBL gene members may play an important role in responding to specific external stimuli.

The combination of the lysin motif (LysM) and receptor kinase domains is present exclusively in plants (Zhang *et al.*, 2007). LysM domain-containing receptor-like kinases (LYK) family members are critical for both nod factor and chitin signaling (Wan *et al.*, 2008). Zhang *et al.* (2007) identified a total of 48 LYK genes in *Arabidopsis* (five genes), *Oryza* (six genes), *M. truncatula* (eight genes), *L. japonicus* (six genes), *P. trichocarpa* (eleven genes) and *Glycine max* (twelve genes), and created two congruent plant LYK phylogenies using LysM domain sequences (all LysM motifs sequences + spacer sequences) and the full-protein sequences (LysM + kinase domain), respectively. The phylogenies revealed that the plant LYK proteins fall into three major clades and two minor clades. They identified six distinct types of LysM motifs in plant LYK proteins and five additional types of LysM motifs in non-kinase plant LysM proteins. Their genomic analysis revealed that the plant LYK gene family has evolved through local and segmental duplications. Their expression data showed that most plant LysM kinase genes were expressed predominantly in the root (Zhang *et al.*, 2007).

4. Protein Degradation

Proteases play key roles in the regulation of biological processes in plants, maintaining strict protein quality control and degrading specific sets of proteins in response to diverse environmental (i.e., defense responses to pathogens and pests) and developmental stimuli (van der Hoorn and Jones, 2004; Garcia-Lorenzo *et al.*, 2006). Garcia-Lorenzo *et al.* (2006) performed a comparative analysis of protease genes in *A. thaliana* and *P. trichocarpa*. They showed that most protease families were larger in *Populus* than in *Arabidopsis*, reflecting the recent genome duplication. They also showed that different *Populus* tissues expressed unique suites of protease genes and that the mRNA levels of different classes of proteases changed along a developmental gradient.

The F-box gene family is involved in posttranslational regulation of gene expression by selective degradation of proteins. In plants, F-box genes influence a variety of biological processes such as long-distance signaling, floral development, shoot branching, leaf senescence, root proliferation, cell cycle, and responses to biotic and abiotic stresses (Gagne *et al.*, 2002; Jain *et al.*, 2007; Yang *et al.*, 2008). Yang *et al.* (2008) recently performed a comparative genomics analysis of F-box gene sequences from *Arabidopsis*, *Oryza*, *Populus*, *Vitis* and papaya. The data revealed that the F-box gene family is expanded in herbaceous annuals (*Arabidopsis* and *Oryza*) relative to woody perennials (*Populus*, *Vitis* and papaya), supporting the hypothesis that compared to long-lived plants like trees, short-lived herbaceous annuals require a more diverse set of protein degradation mechanisms to successfully complete development over a short life cycle.

Programmed cell death, a central regulatory process in both plant development and in plant responses to pathogens, involves various caspase-like cysteine proteases as well as serine proteases such as Kunitz trypsin inhibitors (KTI) that have specific inhibitory activity solely against trypsin proteases (Li *et al.*, 2008). Major and Constabel (2008) performed phylogenetic analysis of 22 Kunitz trypsin inhibitor (KTI) protein sequences of *P. trichocarpa* x *P. deltoides*, and divided them into three clades A (eight sequences), B (four sequences), and C (eight sequences), with the clade A divided further into three subclades A1, A2 and A3. They selected five wound- and herbivore-induced genes representing the phylogenetic clades of the KTI gene family for functional analysis and cloned them into *Escherichia coli* to produce active KTI proteins. The recombinant KTI proteins were all biochemically distinct and showed clear differences in efficacy against trypsin-, chymotrypsin- and elastase-type proteases, suggesting that functional specialization of different members of this gene family is consistent with phylogenetic diversity (Major and Constabel, 2008).

5. Cell Cycle

In both animals and plants, it appears that D-type cyclins (CYCD) play an important role in cell cycle responses to external signals, by forming the regulatory subunit of cyclin-dependent kinase complexes (Meijer and Murray, 2000). Menges *et al.* (2007) constructed a phylogeny using 46 CYCD protein sequences from *Arabidopsis* (10 sequences), *Oryza* (14 sequences), and *Populus* (22 sequences), along with six CYCD sequences from moss (*P. patens*) and algae (*C. reinhardtii* and *O. tauri*) as outgroups. They divided the CYCDs into six clades according to the phylogenetic tree and identified remarkable conservation in intron/exon boundaries and in the location of potential cyclin-dependent kinase phosphorylation sites within CYCD proteins. A promoter sequence analysis and global expression correlation analysis revealed that the phylogenetic clade structure was supported by conserved regulatory elements and by the distinct expression patterns.

The NIMA-related family of serine/threonine kinases (Neks), defined by similarity in their N-terminal catalytic domains to the founding member, Never In Mitosis A from the fungus *Aspergillus*, play essential roles in cell cycle regulation and/or localize to centrosomes in fungi and mammals (Parker *et al.*, 2007). Vigneault *et al.* (2007) identified 22 Neks in *Arabidopsis* (seven sequences), *Populus* (nine sequences) and *Oryza* (six sequences). Their phylogenetic analysis showed that plant Neks were closely related to each other and contained paralogous genes. They examined the chromosomal distribution and exon-intron structure of these genes, concluding that the plant Nek family was derived from a single representative followed by large segmental duplication events.

6. Biosynthesis of Structural Components

Tubulins, as the major structural component of microtubules, consist of alpha/beta heterodimers (Jost *et al.*, 2004). Oakley

et al. (2007) identified eight alpha-TUBULIN (TUA) and 20 beta-TUBULIN (TUB) genes in the *Populus* genome. Their phylogenetic analysis of representative algal and plant full-length TUA proteins defined two distinct classes, with the eight *Populus*, six *Arabidopsis* and four *Oryza* isoforms evenly distributed between the two classes. According to a phylogenetic tree created from TUB proteins, Oakley *et al.* (2007) defined at least four distinct classes of plant TUBs: a Class I and Class I-like group containing half of the *Populus* (10 out of 20) and known maize (4 out of eight8) TUB families, along with a single *Arabidopsis* member (ArathTUB6); Class II containing primarily dicot TUBs, including six *Populus* and four *Arabidopsis* isoforms and one each from *Oryza*, *Zea* and green foxtail; Class III and Class IV each containing two *Populus* and two *Arabidopsis* isoforms arising from genome-wide duplications. They also reported that a number of features, including gene number, alpha:beta gene representation, amino acid changes at the C terminus, and transcript abundance in wood-forming tissue, distinguish the *Populus* tubulin suite from that of *Arabidopsis*.

7. *Biosynthesis of Carbohydrates*

The cellulose synthase gene superfamily of *Arabidopsis* and other seed plants is comprised of the cellulose synthase (CesA) family, which encodes the catalytic subunits of cellulose synthase, and eight families of CesA-like (Csl) genes, which have been proposed to encode processive beta-glycosyl transferases that synthesize noncellulosic cell wall polysaccharides (Roberts and Bushoven, 2007). Suzuki *et al.* (2006) identified 48 members of the cellulose synthase superfamily, which includes CesA and Csl genes in the *Populus* genome. Based on phylogenetic analysis, they divided the 48 *Populus* CesA/Csl protein sequences into nine groups, of which eight groups contained a pair of CesA genes with a nearly identical sequence and the remaining group contained 30 Csl gene family members that were further classified into PtCslA, B, C, D, E, and G subfamilies according to their protein sequence homology with the 29 known AtCsl members that define these subfamilies. They also examined the absolute transcript copy numbers of cellulose synthase superfamily genes in *Populus* and showed that 37 genes were expressed in various tissues, with seven CesA and four Csl genes being xylem specific. Geisler-Lee *et al.* (2006) have carried out a more comprehensive analysis of carbohydrate-active enzyme gene families. Their study based on glycosyltransferases, glycoside hydrolases, carbohydrate esterases, polysaccharide lyases, and expansins suggests that differential mechanisms of carbon flux regulation exist between a tree and an herb.

Endo-ß-mannanase is a hemicellulase that cleaves the ß-1,4 links between the mannose residues in the backbone of mannans, the widespread hemicellulosic polysaccharides in *Plant Cell* walls (Mo and Bewley, 2003; Yuan *et al.*, 2007). Yuan *et al.* (2007) identified 28 endo-beta-mannanase genes in the genomes of *Arabidopsis* (eight genes), *Oryza* (nine genes), and *Populus* (eleven genes). Their phylogenetic analysis of the endo-beta-mannanases from the three plant species implies that the existence of endo-β-mannanases predates the divergence of monocots and dicots, and in each orthologous group, the *Arabidopsis* gene(s) is/are more related to the *Populus* gene(s) than to the *Oryza* gene(s), consistent with the evolutionary relationships between the three plant species.

The polysaccharide xyloglucan (XG) plays an important structural role in the primary cell wall of dicotyledons. The xyloglucan endotransglucosylase/hydrolase (XTH) family proteins are known to have xyloglucan endotransglucosylase (XET) activity and xyloglucan endohydrolase activity (Rose *et al.*, 2002). XET enzymes play a key role in plant morphogenesis: non-hydrolytic cleavage and re-ligation of XG in the cell wall allowing transient, turgor-driven expansion and plant endoxyloglucanases are associated with the hydrolytic mobilization of seed storage XG during germination (Gilbert *et al.*, 2008). Baumann *et al.* (2007) performed phylogenetic analysis on ~130 full-length protein sequences that reflected the diversity of XETs and xyloglucanases in the archetypal glycoside hydrolase family 16 (GH16), including all genome-derived sequences from *A. thaliana* (33 sequences), *O. sativa* (29 sequences) and *P. trichocarpa* (36 sequences), full-length ESTs from *Solanum lycopersicum* (16 sequences) and *P. tremula* × *P. tremuloides* (13 sequences), as well as individual sequences from papaya, *Tropaeolum majus*, *Litchi chinensis* and *Vitis labrusca* × *V. vinifera*. Their phylogenetic analysis, together with kinetic data, suggest that xyloglucanase activity has evolved as a gain-of-function in an ancestral GH16 XET to meet specific biological requirements during seed germination, fruit ripening and rapid wall expansion.

The starch granule is composed of two structurally distinct homopolymers: amylase, which is essentially linear, and amylopectin, which is a moderately branched macromolecule (usually 6% of α-1,6 bonds within the polymer). Starch branching enzymes (SBEs) catalyze the formation of the α-1,6 linkages within amylopectin (Dumez *et al.*, 2006). Han *et al.* (2007) performed phylogenetic analysis of 47 SBE amino acid sequences of various plant species, and identified three orthologs encoding SBEs in *A. thaliana* (AtSBEIII), *O. sativa* (OsSBEIII) and *P. trichocarpa* (PtSBEIII), which represent a new SBE family (SBEIII) with structural features quite different from those of genes of both SBEI and SBEII families in plants.

Bocock *et al.* (2008) identified 24 invertase genes in the *Populus* genome including eight acid invertase genes and 16 neutral/alkaline invertase genes. The phylogenetic tree constructed using protein sequences of *Populus* and *Arabidopsis* acid invertases classified the eight acid invertase genes in *Populus* into two clades: cell wall invertases and vacuolar invertases. The phylogenetic tree constructed using protein sequences of *Populus* and *Arabidopsis* neutral/alkaline invertases classified the neutral/alkaline invertase genes in *Populus* into two clades.

8. *Disease Resistance*

The majority of disease resistance genes in plants encode nucleotide-binding site leucine-rich repeat (NBS-LRR)

proteins. This large family is encoded by hundreds of diverse genes per genome and can be divided into the functionally distinct TIR-domain-containing and CC-domain-containing subfamilies (McHale *et al.*, 2006). Kohler *et al.* (2008) identified 402 *Populus* NBS genes in *Populus*, which were distributed over 228 loci, with 170 sequences located on linkage groups and 232 genes on as yet unmapped scaffolds. According to the phylogenetic tree constructed using the NBS portion of the predicted protein sequences of 117 selected *Populus* R proteins, 11 *Arabidopsis* sequences and five *Oryza* sequences, they divided the NBS family into multiple subfamilies with distinct domain organizations, including Coiled-Coil-NBS-LRR genes, TIR-NBS-LRR genes and BED-finger-NBS-LRR, as well as truncated and unusual NBS- and NBS-LRR-containing genes. Surprisingly, the expansion of R genes in *Populus* relative to Arabidopsis cannot be accounted for by the large-scale duplication of the *Populus* genome. Instead, it appears that R genes have been expanding due to small-scale tandem duplications, and this rapidly-evolving gene family appears to be under strong diversifying selection, perhaps driven by the complex biotic interactions that develop over the long lifespan of a tree (Kohler *et al.*, 2008).

9. Fatty Acid Metabolism

Acyl-coenzyme A synthetases (ACS) catalyze the fundamental, initial reaction in fatty acid metabolism (Watkins *et al.*, 2007). 4-Coumarate:CoA ligase (4CL) is a branch point enzyme of plant phenylpropanoid metabolism, catalyzing the activation of 4-coumaric acid and various other hydroxylated and methoxylated cinnamic acid derivatives to the corresponding CoA esters in a two-step reaction (Pietrowska-Borek *et al.*, 2003). Souza Cde *et al.* (2008) performed phylogenetic analysis of 104 ACS related gene sequences from various organisms, including *bona fide* 4CL sequences from *Arabidopsis*, *Populus* and *Oryza*. Their phylogenetic tree revealed two general groups of adenylate-forming proteins: one large group containing representatives from all organisms analyzed, including bacteria, fungi, *Chlamydomonas*, *Physcomitrella* and angiosperm plants and a second group containing land plant-specific ACS proteins including both *bona fide* 4CL proteins and *Arabidopsis* 4CL-like ACS proteins. Further, they performed a separate phylogenetic analysis of the plant-specific ACSs protein sequences from *Arabidopsis* (nine sequences), *Populus* (13 sequences) and *Oryza* (12 sequences) and were able to divide the plant-specific ACSs into five clades. The phylogenetic tree revealed that the number of plant-specific ACS genes in each clade varied between species while the total number of plant-specific ACS genes within each genome was similar. They hypothesized that ACS genes have undergone differential expansion in each angiosperm lineage, perhaps reflecting differences in life histories that placed varying selective pressures on the elaboration of biochemical pathways requiring ACS activity (Souza Cde *et al.*, 2008).

10. Phenylpropanoid Pathway

Populus produces a rich array of natural products such as the secondary metabolites derived from phenylalanine via phenol and phenylpropanoid metabolism (Hamberger *et al.*, 2007). More than 100 *Populus* phenylpropanoid pathway genes were phylogenetically compared with homologs in *Arabidopsis* and *Oryza* by Tsai *et al.* (2006) and Hamberger *et al.* (2007). Gene families included in their analyses were: arogenate dehydratase, beta-alanine N-methyltransferase, flavonoid 3′,5′-hydroxylase, flavone synthase II, flavanone 3-hydroxylase, flavonol synthase, dihydroflavonol 4-reductase, anthocyanidin synthase, anthocyanidin reductase, leucoanthocyanidin reductase, flavonoid O-methyltransferase, trans-cinnamate 4-hydroxylase, coumarate 3-hydroxylase, caffeic acid O-methyltransferase, ferulate 5-hydroxylase, caffeoyl-CoA O-methyltransferase, hydroxycinnamoyltransferase, aldehyde dehydrogenase, cinnamoyl CoA reductase, NADPH-cytochrome P450 oxydoreductase, cinnamyl alcohol dehydrogenase-related, phenylalanine ammonia lyase, and 4-coumarate:CoA ligase.

11. Photorespiration

The glycine decarboxylase multi-enzyme complex (GDC) catalyzes in a multi-step reaction the rapid 'cracking' of glycine molecules flooding out of the peroxisomes during the course of photorespiration (Douce *et al.*, 2001). Rajinikanth *et al.* (2007) identified ten GDC proteins in *Populus*, of which eight are localized in mitochondria and two in plastids. Their phylogenetic analysis of GDC protein sequences from representative dicot, monocot and gymnosperm species clearly revealed two distinct GDC classes, with class I corresponding to photorespiratory isoforms and class II associated with one-carbon metabolism, indicating that the functional uniqueness is consistent with phylogenetic classification.

C. Databases

Just like other plant genomics models such as *Arabidopsis* and *Oryza*, *Populus* has become a new theme of genomics databases. To date, more than ten public genomics databases have been established for *Populus* or related to *Populus* (Table 2).

1. Populus-Specific Databases

The official release of the *Populus* genome information is provided by the DOE's Joint Genome Institute (JGI) *Populus* genome browser (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html). In addition, Sterky *et al.* (2004) established a public web-based EST database (POPULUSDB) which provides digital expression profiles for 18 tissues that comprise the majority of differentiated organs (http://www.populus.db.umu.se/). The RIKEN *Populus* database (http://rpop.psc.riken.jp) contains information covering 10 *Populus* species (*P. deltoides*, *P. euphratica*, *P. tremula*, *P. tremula* × *P. alba*, *P. tremula* × *P. tremuloides*, *P.*

TABLE 2
Populus genomics databases

Category	Database	URL	Reference
Populus-specific	PopulusDB	http://www.populus.db.umu.se/	Sterky *et al.* (2004)
Populus-specific	RPOPDB	http://rpop.psc.riken.jp/	
Populus-specific	RepPop	http://csbl.bmb.uga.edu/∼ffzhou/RepPop/	Zhou and Xu (2009)
Populus-specific	Transcription Factors	http://dptf.cbi.pku.edu.cn/	Zhu *et al.* (2007)
Populus-specific	POPARRAY	http://popgenome.ag.utk.edu/mdb/index.php	
Populus-specific	PopGenIE	http://www.popgenie.db.umu.se/popgenie/	Sjödin *et al.* (2008)
Populus-specific	JGI *Populus* genome	http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html	Tuskan *et al.* (2006)
Comprehensive	ChromDB	http://www.chromdb.org/org_specific.html?o = POPTR	Gendler *et al.* (2008)
Comprehensive	PlantGDB	http://www.plantgdb.org/	Duvick *et al.* (2008)
Comprehensive	Gramene	http://www.gramene.org/	Liang *et al.* (2008)
Comprehensive	PLEXdb	http://www.plexdb.org/index.php	Wise *et al.* (2007)
Comprehensive	Phytozome	http://www.phytozome.net/index.php	

tremuloides, *P. trichocarpa*, *P. trichocarpa* × *P. deltoides*, *P. trichocarpa* × *P. nigra*, and *P.* × *canadensis*). To facilitate functional studies on the regulation of gene expression in *Populus*, Zhu *et al.* (2007) created a *Populus* transcription factor (TF) database (DPTF; http://dptf.cbi.pku.edu.cn/) containing 2576 putative *Populus* TFs, distributed in 64 families. It provides comprehensive information for the *Populus* TFs such as sequence features, functional domains, GO assignment, expression evidence, phylogenetic tree of each family, and homologs in *Arabidopsis* and *Oryza*. Basing on the genome-wide analysis of 9,623 repetitive elements in the *P. trichocarpa* genome, Zhou and Xu (2009) created a web-browsable database, RepPop (http://csbl.bmb.uga.edu/∼ffzhou/RepPop/), which offers resources on DNA transposons, RNA retrotransposons, Miniature Inverted-repeat Transposable Elements (MITE), Simple Sequence Repeats (SSR), segmental duplications, etc. This database also provides various search capabilities and a Wiki system to facilitate functional annotation and curation of the repetitive elements. To facilitate the exploration of genes and gene function in *Populus*, Sjödin *et al.* (2008) developed an integrative *Populus* functional genomics database, PopGenIE (www.popgenie.org), which includes several browser tools for viewing genome sequence, QTL and synteny; an expression tool for displaying multiple-tissue expression patterns based on microarray experiments, Digital Northern analysis of EST data, and co-expressed genes; a category tool for information about gene families, pathways, GO annotations; and a sequence tool for similarity searches and data downloads.

2. *Comprehensive Databases Involving Populus*

To facilitate comparative genomic studies amongst green plants, JGI and the Center for Integrative Genomics developed Phytozome (http://www.phytozome.net/index.php), which provides access to eleven sequenced and annotated higher plant genomes, eight of which have been clustered into gene families at six evolutionarily significant nodes, along with PFAM, KOG, KEGG and PANTHER assignments, offering resources for clusters of orthologous and paralogous genes that represent the modern descendents of ancestral gene sets, as well as clade specific genes and gene expansions. Duvick *et al.* (2008) developed a comprehensive plant genomics database, PlantGDB (http://www.plantgdb.org/) which provides annotated transcript assemblies for >100 plant species including *Populus*, with transcripts mapped to their cognate genomic context integrated with a variety of sequence analysis tools and web services. PlantGDB also hosts a plant genomics research outreach portal that facilitates access to a large number of training resources. Another plant comparative genomics database is Gramene (Liang *et al.*, 2008), which contains genomic information for *Oryza sativa* var. *indica* and *japonica*, *O. glaberrima*, *O. rufipogon*, *Zea mays*, *Sorghum bicolor*, *Arabidopsis thaliana*, *Vitis vinifera* and *P. trichocarpa*, including genome assembly and annotations, cDNA/mRNA sequences, genetic and physical maps/markers, genes, quantitative trait loci, proteins, and comparative ontologies. To display sets of curated plant genes predicted to encode proteins associated with chromatin remodeling, Gendler *et al.* (2008) established the ChromDB database (http://www.chromdb.org) which displays chromatin-associated proteins, including RNAi-associated proteins, for a broad range of organisms such as *A. thaliana*, *O. sativa* ssp. *japonica*, *P. trichocarpa*, *Zea mays*, and model animal and fungal species. ChromDB contains three types of sequences: genomic-based (predominantly plant sequences); transcript-based (EST contigs or cDNAs for plants lacking a sequenced genome) and NCBI RefSeq sequences for a variety of model animal organisms. To facilitate large-scale gene expression analysis for plants, Wise *et al.* (2007) created PLEXdb (http://plexdb.org/), which offers a unified web interface to support the functional interpretation of highly parallel microarray experiments integrated with traditional structural genomics and phenotypic data. PLEXdb contains information for 13 plant species, such as *Arabidopsis*, barley, *Citrus*, cotton, *Vitis*, *Zea*,

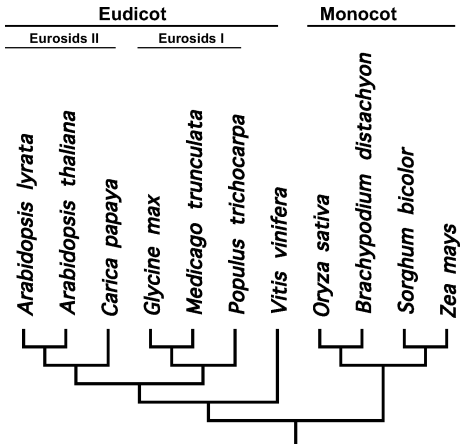

FIG. 2. Higher plant species whose genomes have been sequenced.

Medicago, *Populus*, *Oryza*, *Glycine*, sugarcane, tomato, and wheat.

D. Evolutionary Genomics

1. *Genome Evolution*

Knowledge of plant evolutionary history at the genome level forms an excellent context for understanding genomics at subgenome levels. Genomes of nine higher plant species have been sequenced (Fig. 2). Genome-wide sequence resources have facilitated large-scale computational analysis of genome evolution. Tuskan *et al.*, 2006 reported that *Populus* experienced two rounds of genome duplication, with the most recent event ('salicoid' duplication) contained within the Salicaceae and a second whole-genome event ('eurosid' duplication) apparently shared among the Eurosids. Using a robust computational framework that combines information from multiple orthologous and duplicated regions to construct local syntenic networks, Tang *et al.* (2008b) showed that a shared ancient hexaploidy event (γ triplication) can be inferred based on the genome sequences of *Arabidopsis*, *Carica*, *Populus*, *Vitis* and *Oryza*. They conclude that "paleo-hexaploidy" clearly preceded the rosid-asterid split, but it remains equivocal whether it also affected monocots. In short, three known genome duplication events have been inferred in *Populus*: salicoid, eurosid, and γ triplication (Fig. 3).

2. *Evolutionary Dynamics of Duplicated Genes*

Gene duplication and diversification contribute to the novelty of molecular functions. To study the evolutionary dynamics of gene families in plants, Yang *et al.* (2006) identified 27 pairs of paralogous Dof (DNA binding with one finger) genes in a phylogenetic tree constructed from protein sequences in *P. trichocarpa*, *A. thaliana* and *O. sativa*. The comparison of protein motif structure of the Dof paralogs and their ancestors revealed six different gene fates after gene duplication as well as epigenetic modification via protein methylation. Multiple modes of evolutionary dynamics including neo/nonfunctionalization and subfunctionalization were also reported at the promoter level by De Bodt *et al.* (2006), who investigated the evolution of MADS-box genes in *Arabidopsis* and *Populus* using phylogenetic footprinting. This study showed that many genes have diverged in their regulatory sequences after duplication and/or speciation. Based on microarray and genome sequence information, Kalluri *et al.* (2007) showed that segmental duplicate gene pairs in the *Populus* ARF

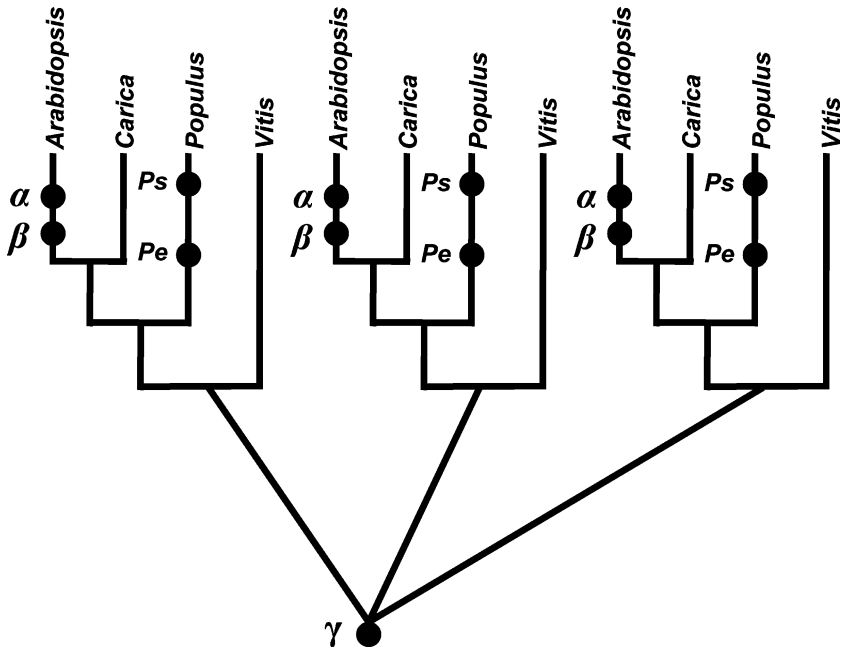

FIG. 3. The duplication events in the plant genomes. Adapted from Tang *et al.* (2008a, 2008b) and Tuskan *et al.* (2006).

gene family, *PoptrARF6.1–6.2* and *PoptrARF6.4–6.5*, have differential expression profiles. For example, *PoptrARF6.1* was expressed preferentially in mature leaves and phloem-cortex samples whereas *PoptrARF6.2* was expressed strongly in xylem, phloem, and vegetative and reproductive meristems, suggesting subfunctionalization after gene duplication. Subfunctionalization in expression of the duplicated genes has also been revealed in the some gene families involved in phenylpropanoid metabolism, such as phenylalanine ammonia lyase, coumaroyl 3-hydroxylase, cinnamyl alcohol dehydrogenase-related and caffeoyl-CoA 3-O-methyltransferase (Hamberger *et al.*, 2007), in the *Populus* MONOPTEROS gene family in which two duplicated MONOPTEROS genes, PoptrMP1 and PoptrMP2, have overlapping but distinct expression patterns (Johnson and Douglas, 2007), and in *Populus* LHY/CCA1 gene family in which two LHYs produced by the Salicoid polyploidy event showed asymmetric expressions (Takata *et al.*, 2009).

IV. CONCLUSION AND FUTURE PERSPECTIVES

A. Conclusion

Our primary objective in writing this review was to highlight how the recent sequencing of the *Populus* genome (Tuskan *et al.*, 2006) has spurred applications of state-of-the-art technologies in *Populus* research, including both experiment-based functional genomics and newly developed computational genomics. Technologies are in place for the high-throughput analysis of gene expression, coupled with protein and metabolite profiling and verification of gene function via plant transformation and incorporation of candidate genes into transgenic *Populus* plants. These capabilities, mostly developed in the last five years, provide much-needed insights into the growth, morphology and reproductive strategies for some of the largest and longest-living organisms on Earth.

Equally impressive have been the major inroads made in developing *in silico* resources for use in comparative genomics. Central data repositories and on-line bioinformatics tools now make possible identification of lineage-specific motifs and genes and improved understanding of gene family evolution and regulation. Studies based largely on *in silico* analyses are already demonstrating how gene duplication and diversification can contribute to the novelty of molecular functions, and how gene families develop and expand along unique trajectories in different phylogenetic lineages. This information highlights how plant evolutionary history at the genome level forms a context for understanding genome organization, evolution and function at the sub-genome level. Insights derived from functional and computational genomics will continue to yield improved understanding at the organismal, population, community and ecosystem scales. It will be in this arena that molecular biologists, physiologists and ecologists will derive maximum mutual benefits from investing in the development of genetic and genomics resources for *Populus* (DiFazio, 2005; Whitham *et al.*, 2008).

Finally, several articles published while the *Populus* genome was being sequenced (Taylor, 2002; Wullschleger *et al.*, 2002; Wullschleger *et al.*, 2002; Tuskan *et al.*, 2004) highlighted the notion that embracing new technology and new research paradigms will not be easy. To our surprise, and as evidenced in this review, the community has quickly risen to the challenge provided by the development and application of technologies in *Populus* genomics research. We see no reason why this enthusiasm should wane. Only time will tell, however, how our investments in functional and computational genomics will aid in identifying the suite of genes and gene families that underlie plant growth and development. Evidence exists that insights derived over a relatively short time are already being applied to enhance production of short-rotation bioenergy crops as a renewable source of biomass for transportation fuels. This advancement will potentially lessen our dependence on foreign oil and mitigating rising CO_2 concentrations in the atmosphere (Tuskan and Walsh, 2001).

B. Future Perspectives

The trend in biological investigations is shifting from reductionism to evolutionary system-inclusive biology. This trend is driven in part by the capabilities that genomic science is now providing for powerful cross-species comparative studies. Comparative genomics will be an essential component of investigating genetic features and molecular processes underlying core conserved, as well as unique divergent, plant properties. Comparative genomics studies of *Populus*, *Arabidopsis*, *Oryza*, papaya, *Vitis*, *Sorghum*, *Glycine*, *Brachypodium* and others have already begun shedding light on the evolutionary events that have resulted in divergent and conserved patterns at the genetic, molecular and organism levels. Moreover, comparative genomics research is expected to move beyond the narrow scope of studying isolated plants to the broader context of a community (i.e., the "metagenome") (Tringe and Rubin, 2005; Whitham *et al.*, 2008). The identity as well as roles of *Populus* endophytic and rhizosphere microbiota will also be important co-considerations in sustainable development and the establishment of plantations of suitably improved plants (Martin *et al.*, 2004).

It is not hard to foresee that the availability of genome sequences from dozens of plant species will propel science towards predictive outcomes facilitated by the 1) creation of open-source community resource and data repositories, 2) sharing of resources and discoveries, 3) emergence of new technological innovations, and 4) synergy in large interdisciplinary collaborative efforts. The new era of plant research will witness innovative hypotheses, better experimental design, and more sophisticated toolsets.

It is clear that plant biotechnology will play an increasingly large role in meeting soaring demands for food, fiber and energy. Therefore, it is important that the research community raise its awareness regarding future regulatory and consumer acceptance

issues. In order to realize the potential of the long-range agro- and forest biotechnology goals, molecular geneticists will need to understand the economic, environmental and social contexts, as well as biological implications, of their research (Strauss, 2003).

The attractiveness of *Populus* as a model woody crop has been acknowledged, and efforts are underway worldwide to elucidate genetic features that are key to generating suitably tailored plants. The 'book of life' awaits future examination and interpretations directed towards basic and applied research related to the development of tree-based strategies for fiber production, carbon sequestration, phytoremediation, and accelerated and sustainable domestication.

ACKNOWLEDGMENTS

We thank David J. Weston for the constructive comments on the manuscript and Tara A. Hall for assisting with proofreading. The writing of this review was supported by the BioEnergy Science Center (BESC), a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract Number DE–AC05–00OR22725.

REFERENCES

Alstrom-Rapaport, C. L. M., Wang, Y. C., Roberts, G., and Tuskan, G. A. 1998. Identification of a RAPD marker linked to sex determination in the basket willow (*Salix viminalis* L.). *J. Heredity* **89**: 44–49.

Andersson, A., Keskitalo, J., Sjodin, A., Bhalerao, R., Sterky, F., Wissel, K., Tandre, K., Aspeborg, H., Moyle, R., Ohmiya, Y., *et al.* 2004. A transcriptional timetable of autumn senescence. *Genome Biol.* **5**: R24.

Andersson-Gunneras, S., Mellerowicz, E. J., Love, J., Segerman, B., Ohmiya, Y., Coutinho, P. M., Nilsson, P., Henrissat, B., Moritz, T., and Sundberg, B. 2006. Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *Plant J.* **45**: 144–165.

Arnaud, D., Dejardin, A., Leple, J. C., Lesage-Descauses, M. C., and Pilate, G. 2007. Genome-wide analysis of LIM gene family in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*. *DNA Res.* **14**: 103–116.

Baek, J. M., Han, P., Iandolino, A., and Cook, D. R. 2008. Characterization and comparison of intron structure and alternative splicing between *Medicago truncatula*, *Populus trichocarpa*, *Arabidopsis* and rice. *Plant Mol. Biol.* **67**: 499–510.

Barakat, A., Wall, P. K., Diloreto, S., dePamphilis, C. W., and Carlson, J. E. 2007. Conservation and divergence of microRNAs in *Populus*. *BMC Genomics* **8**: 481.

Baumann, M. J., Eklof, J. M., Michel, G., Kallas, A. M., Teeri, T. T., Czjzek, M., and Brumer, H., 3rd. 2007. Structural evidence for the evolution of xyloglucanase activity from xyloglucan endo-transglycosylases: biological implications for cell wall metabolism. *Plant Cell* **19**: 1947–1963.

Beavis, W. D. 1998. QTL analyses: Power, precision, and accuracy. **In:** AH Paterson, ed, *Molecular Dissection of Complex Traits*. CRC Press, Boca Raton, FL, pp. 145–162.

Becker, A., and Theissen, G. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylo. Evol.* **29**: 464–489.

Benedict, C., Drost, D., Novaes, E., Boaventura-Novaes, C. D., Yu, Q., Dervinis, C., Maia, J. M., Yap, J., Miles, B., Peter, G. F., *et al.* 2009. Genome-wide eQTL mapping in xylem, leaf and root identifies tissue-specific hotspots in a *Populus trichocarpa* x *deltoides* pedigree. Plant & Animal Genomes XVII Conference W181.

Bocock, P. N., Morse, A. M., Dervinis, C., and Davis, J. M. 2008. Evolution and diversity of invertase genes in *Populus trichocarpa*. *Planta* **227**: 565–576.

Bogeat-Triboulot, M. B., Brosche, M., Renaut, J., Jouve, L., Le Thiec, D., Fayyaz, P., Vinocur, B., Witters, E., Laukens, K., Teichmann, T., *et al.* 2007. Gradual soil water depletion results in reversible changes of gene expression, protein profiles, ecophysiology, and growth performance in *Populus euphratica*, a poplar growing in arid regions. *Plant Physiol.* **143**: 876–892.

Bradshaw, H. D., and Stettler, R. F. 1995. Molecular genetics of growth and development in *Populus*. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. *Genetics* **139**: 963–973.

Bradshaw, H. D., Villar, M., Watson, B. D., Otto, K. G., Stewart, S., and Stettler, R. F. 1994. Molecular-genetics of growth and development in *Populus*. III. A genetic-linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. *Theoreti. Applied Genetics* **89**: 167–178.

Brosche, M., Vinocur, B., Alatalo, E. R., Lamminmaki, A., Teichmann, T., Ottow, E. A., Djilianov, D., Afif, D., Bogeat-Triboulot, M. B., Altman, A., *et al.* 2005. Gene expression and metabolite profiling of *Populus euphratica* growing in the Negev desert. *Genome Biol.* 6: –

Busov, V., Meilan, R., Pearce, D. W., Rood, S. B., Ma, C., Tschaplinski, T. J., and Strauss, S. H. 2006. Transgenic modification of gai or rgl1 causes dwarfing and alters gibberellins, root growth, and metabolite profiles in *Populus*. *Planta* **224**: 288–299.

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

Caruso, A., Chefdor, F., Carpin, S., Depierreux, C., Delmotte, F. M., Kahlem, G., and Morabito, D. 2008. Physiological characterization and identification of genes differentially expressed in response to drought induced by PEG 6000 in *Populus canadensis* leaves. *J. Plant Physiol.* **165**: 932–941.

Creux, N. M., Ranik, M., Berger, D. K., and Myburg, A. A. 2008. Comparative analysis of orthologous cellulose synthase promoters from *Arabidopsis*, *Populus* and *Eucalyptus*: evidence of conserved regulatory elements in angiosperms. *New Phytol.* **179**: 722–737.

Cseke, L. J., Cseke, S. B., and Podila, G. K. 2007. High efficiency poplar transformation. *Plant Cell Rep.* **26**: 1529–1538.

Danilevskaya, O. N., Meng, X., Hou, Z. L., Ananiev, E. V., and Simmons, C. R. 2008. A genomic and expression compendium of the expanded PEBP gene family from maize. *Plant Physiol.* **146**: 250–264.

De Bodt, S., Theissen, G., and Van de Peer, Y. 2006. Promoter analysis of MADS-box genes in eudicots through phylogenetic footprinting. *Mol. Biol. Evol.* **23**: 1293–1303.

Diaz-Riquelme, J., Lijavetzky, D., Martinez-Zapater, J. M., and Carmona, M. J. 2009. Genome-wide analysis of MIKCC-type MADS box genes in grapevine. *Plant Physiol.* **149**: 354–369.

DiFazio, S. P. 2005. A pioneer perspective on adaptation. *New Phytol.* 165**:** 661–664.

Douce, R., Bourguignon, J., Neuburger, M., and Rebeille, F. 2001. The glycine decarboxylase system: a fascinating complex. *Trends Plant Sci.* **6**: 167–176.

Druart, N., Johansson, A., Baba, K., Schrader, J., Sjodin, A., Bhalerao, R. R., Resman, L., Trygg, J., Moritz, T., and Bhalerao, R. P. 2007. Environmental and hormonal regulation of the activity-dormancy cycle in the cambial meristem involves stage-specific modulation of transcriptional and metabolic networks. *Plant J.* **50**: 557–573.

Du, J., Xie, H. L., Zhang, D. Q., He, X. Q., Wang, M. J., Li, Y. Z., Cui, K. M., and Lu, M. Z. 2006. Regeneration of the secondary vascular system in poplar as a novel system to investigate gene expression by a proteomic approach. *Proteomics* **6**: 881–895.

Dumez, S., Wattebled, F., Dauvillee, D., Delvalle, D., Planchot, V., Ball, S. G., and D'Hulst, C. 2006. Mutants of *Arabidopsis* lacking starch branching enzyme II substitute plastidial starch synthesis by cytoplasmic maltose accumulation. *Plant Cell* **18**: 2694–2709.

Duvick, J., Fu, A., Muppirala, U., Sabharwal, M., Wilkerson, M. D., Lawrence, C. J., Lushbough, C., and Brendel, V. 2008. PlantGDB: a resource for comparative plant genomics. *Nucleic Acids Res.* **36**: D959–965.

Emrich, S. J., Barbazuk, W. B., Li, L., and Schnable, P. S. 2007. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Res.* **17**: 69–73.

Espinosa-Ruiz, A., Saxena, S., Schmidt, J., Mellerowicz, E., Miskolczi, P., Bako, L., and Bhalerao, R. P. 2004. Differential stage-specific regulation of cyclin-dependent kinases during cambial dormancy in hybrid aspen. *Plant J.* **38**: 603–615.

Ferreira, S., Hjerno, K., Larsen, M., Wingsle, G., Larsen, P., Fey, S., Roepstorff, P., and Salome Pais, M. 2006. Proteome profiling of *Populus euphratica* Oliv. upon heat stress. *Ann. Bot.* (Lond) **98**: 361–377.

Ferris, R., Long, L., Bunn, S. M., Robinson, K. M., Bradshaw, H. D., Rae, A. M., and Taylor, G. 2002. Leaf stomatal and epidermal cell development: identification of putative quantitative trait loci in relation to elevated carbon dioxide concentration in poplar. *Tree Physiol.* **22**: 633–640.

Fiehn, O. 2002. Metabolomics – the link between genotypes and phenotypes. *Plant Mol. Biol.* 48**:** 155–171.

Gagne, J. M., Downes, B. P., Shiu, S. H., Durski, A. M., and Vierstra, R. D. 2002. The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **99**: 11519–11524.

Garcia-Lorenzo, M., Sjodin, A., Jansson, S., and Funk, C. 2006. Protease gene families in *Populus* and *Arabidopsis*. *BMC Plant Biol.* 6**:** 30.

Gaudet, M., Jorge, V., Paolucci, I., Beritognolo, I., Mugnozza, G. S., and Sabatti, M. 2008. Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait. *Tree Genetics & Genomes* **4**: 25–36.

Geisler-Lee, J., Geisler, M., Coutinho, P. M., Segerman, B., Nishikubo, N., Takahashi, J., Aspeborg, H., Djerbi, S., Master, E., Andersson-Gunneras, S., *et al.* 2006. Poplar carbohydrate-active enzymes. Gene identification and expression analyses. *Plant Physiol.* **140**: 946–962.

Gendler, K., Paulsen, T., and Napoli, C. 2008. ChromDB: the chromatin database. *Nucleic Acids Res.* **36**: D298–302.

Gilbert, H. J., Stalbrand, H., and Brurner, H. 2008. How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. *Curr. Opin. Plant Biol.* **11**: 338–348.

Gilchrist, E. J., Haughn, G. W., Ying, C. C., Otto, S. P., Zhuang, J., Cheung, D., Hamberger, B., Aboutorabi, F., Kalynyak, T., Johnson, L., *et al.* 2006. Use of Ecotilling as an efficient SNP discovery tool to survey genetic variation in wild populations of *Populus trichocarpa*. *Mol. Ecol.* **15**: 1367–1378.

Gollery, M., Harper, J., Cushman, J., Mittler, T., Girke, T., Zhu, J. K., Bailey-Serres, J., and Mittler, R. 2006. What makes species unique? The contribution of proteins with obscure features. *Genome Biol.* 7: R57.

Gollery, M., Harper, J., Cushman, J., Mittler, T., and Mittler, R. 2007. POFs: what we don't know can hurt us. *Trends Plant Sci.* **12**: 492–496.

Gou, J. Y., Park, S., Yu, X. H., Miller, L. M., and Liu, C. J. 2008. Compositional characterization and imaging of "wall-bound" acylesters of *Populus trichocarpa* reveal differential accumulation of acyl molecules in normal and reactive woods. *Planta* **229**: 15–24.

Gowda, M., Li, H., Alessi, J., Chen, F., Pratt, R., and Wang, G. L. 2006. Robust analysis of 5′-transcript ends (5′-RATE): a novel technique for transcriptome analysis and genome annotation. *Nucleic Acids Res.* **34**: e126.

Groover, A. T., Mansfield, S. D., DiFazio, S. P., Dupper, G., Fontana, J. R., Millar, R., and Wang, Y. 2006. The *Populus* homeobox gene ARBORKNOX1 reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. *Plant Mol. Biol.* **61**: 917–932.

Gu, Z. M., Ma, B. J., Jiang, Y., Chen, Z. W., Su, X., and Zhang, H. S. 2008. Expression analysis of the calcineurin B-like gene family in rice (*Oryza sativa* L.) under environmental stresses. *Gene* **415**: 1–12.

Hamberger, B., Ellis, M., Friedmann, M., Souza, C. D. A., Barbazuk, B., and Douglas, C. J. 2007. Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the *Populus* lignin toolbox and conservation and diversification of angiosperm gene families. *Can. J. Bot.* **85**: 1182–1201.

Han, K.-H., Gordon, M. P., and Strauss, S. H. 1996. Cellular and molecular biology of *Agrobacterium*-mediated transformation of plants and its application to genetic transformation of *Populus*. **In**: Stettler, R. F., Bradshaw, H. D., Heilman, P. E. Hinckley, and T. M., Eds., *Biology of* Populus *and Its Implications for Management and Conservation.* National Research Council of Canada, Ottawa, Ontario, Canada, pp. 201–222.

Han, Y., Sun, F. J., Rosales-Mendoza, S., and Korban, S. S. 2007. Three orthologs in rice, *Arabidopsis*, and *Populus* encoding starch branching enzymes (SBEs) are different from other SBE gene families in plants. *Gene* **401**: 123–130.

Harding, S. A., Jiang, H. Y., Jeong, M. L., Casado, F. L., Lin, H. W., and Tsai, C. J. 2005. Functional genomics analysis of foliar condensed tannin and phenolic glycoside regulation in natural cottonwood hybrids. *Tree Physiol.* **25**: 1475–1486.

Harrison, E. J., Bush, M., Plett, J. M., McPhee, D. P., Vitez, R., O'Malley, B., Sharma, V., Bosnich, W., Seguin, A., and MacKay, J., *et al.* 2007. Diverse developmental mutants revealed in an activation-tagged population of poplar. *Can. J. Bot.* **85**: 1071–1081.

Hayashi, H., Czaja, I., Lubenow, H., Schell, J., and Walden, R. 1992. Activation of a plant gene by T-DNA tagging: auxin-independent growth in vitro. *Science* **258**: 1350–1353.

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

Igasaki, T., Watanabe, Y., Nishiguchi, M., and Kotoda, N. 2008. The FLOWERING LOCUS T/TERMINAL FLOWER 1 family in Lombardy poplar. *Plant Cell Physiol.,* **49**: 291–300.

Ingvarsson, P. K. 2008. Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics* **180**: 329–340.

Ingvarsson, P. K., Garcia, M. V., Luquez, V., Hall, D., and Jansson, S. 2008. Nucleotide polymorphism and phenotypic associations within and around the phytochrome B2 Locus in European aspen (*Populus tremula*, Salicaceae). *Genetics* **178**: 2217–2226.

Jain, M., Nijhawan, A., Arora, R., Agarwal, P., Ray, S., Sharma, P., Kapoor, S., Tyagi, A. K., and Khurana, J. P. 2007. F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiol.* **143**: 1467–1483.

Johnson, L. A., and Douglas, C. J. 2007. *Populus trichocarpa* MONOPTEROS/AUXIN RESPONSE FACTOR5 (ARF5) genes: comparative structure, subfunctionalization, and *Populus Arabidopsis* microsynteny. *Can. J. Bot.* **85**: 1058–1070.

Jorge, V., Dowkiw, A., Faivre-Rampant, P., and Bastien, C. 2005. Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *New Phytol.* **167**: 113–127.

Jost, W., Baur, A., Nick, P., Reski, R., and Gorr, G. 2004. A large plant beta-tubulin family with minimal C-terminal variation but differences in expression. *Gene* **340**: 151–160.

Kalluri, U. C., DiFazio, S. P., Brunner, A. M., and Tuskan, G. A. 2007. Genome-wide analysis of Aux/IAA and ARF gene families in *Populus trichocarpa*. *BMC Plant Biol.* **7**: 59.

Kalluri, U. C., and Joshi, C. P. 2004. Differential expression patterns of two cellulose synthase genes are associated with primary and secondary cell wall development in aspen trees. *Planta* **220**: 47–55.

Kawaoka, A., and Ebinuma, H. 2001. Transcriptional control of lignin biosynthesis by tobacco LIM protein. *Phytochemistry* **57**: 1149–1157.

Khurana, T., Khurana, B., and Noegel, A. A. 2002. LIM proteins: association with the actin cytoskeleton. *Protoplasma* **219**: 1–12.

Kirst, M., Basten, C. J., Myburg, A. A., Zeng, Z. B., and Sederoff, R. R. 2005. Genetic architecture of transcript-level variation in differentiating xylem of a *Eucalyptus* hybrid. *Genetics* **169**: 2295–2303.

Ko, J. H., Prassinos, C., and Han, K. H. 2006. Developmental and seasonal expression of PtaHB1, a *Populus* gene encoding a class III HD-Zip protein, is closely associated with secondary growth and inversely correlated with the level of microRNA (miR166). *New Phytol.* **169**: 469–478.

Kohler, A., Delaruelle, C., Martin, D., Encelot, N., and Martin, F. 2003. The poplar root transcriptome: analysis of 7000 expressed sequence tags. *Febs Lett.* **542**: 37–41.

Kohler, A., Rinaldi, C., Duplessis, S., Baucher, M., Geelen, D., Duchaussoy, F., Meyers, B. C., Boerjan, W., and Martin, F. 2008. Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Mol. Biol.* **66**: 619–636.

Krom, N., and Ramakrishna, W. 2008. Comparative analysis of divergent and convergent gene pairs and their expression patterns in rice, *Arabidopsis*, and *Populus*. *Plant Physiol.* **147**: 1763–1773.

Leple, J. C., Dauwe, R., Morreel, K., Storme, V., Lapierre, C., Pollet, B., Naumann, A., Kang, K. Y., Kim, H., Ruel, K., *et al.* 2007. Downregulation of cinnamoyl-coenzyme A reductase in poplar: multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. *Plant Cell* **19**: 3669–3691.

Leseberg, C. H., Li, A., Kang, H., Duvall, M., and Mao, L. 2006. Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*. *Gene* 378: 84–94.

Li, J., Brader, G., and Palva, E. T. 2008. Kunitz trypsin inhibitor: An antagonist of cell death triggered by phytopathogens and fumonisin B1 in *Arabidopsis*. *Mol. Plant* **1**: 482–495.

Liang, C., Jaiswal, P., Hebbard, C., Avraham, S., Buckler, E. S., Casstevens, T., Hurwitz, B., McCouch, S., Ni, J., Pujar, A., *et al.* 2008. Gramene: a growing plant comparative genomics resource. *Nucleic Acids Res.* **36**: D947–953.

Liang, P. 2002. A decade of differential display. *Biotechniques* **33**: 338–344, 346.

Liscum, E., and Reed, J. W. 2002. Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Mol. Biol.* **49**: 387–400.

Lu, S., Sun, Y. H., Shi, R., Clark, C., Li, L., and Chiang, V. L. 2005. Novel and mechanical stress-responsive MicroRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. *Plant Cell* **17**: 2186–2203.

Ma, C., Strauss, S. H., and Meilan, R. 2004. *Agrobacterium*-mediated transformation of the genome-sequenced poplar clone, Nisqually-1 (*Populus trichocarpa*). *Plant Mol. Biol. Reporter* **22**: 311–312.

Major, I. T., and Constabel, C. P. 2008. Functional analysis of the Kunitz trypsin inhibitor family in poplar reveals biochemical diversity and multiplicity in defense against herbivores. *Plant Physiol.* **146**: 888–903.

Markussen, T., Pakull, B., and Fladung, M. 2007. Positioning of sex-correlated markers for *Populus* in a AFLP- and SSR-Marker based genetic map of *Populus tremula* x *tremuloides*. *Silvae Genetica* **56**: 180–184.

Martin, F., Aerts, A., Ahren, D., Brun, A., Danchin, E. G. J., Duchaussoy, F., Gibon, J., Kohler, A., Lindquist, E., Pereda, V., *et al.* 2008. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature*. **452**: 88–92.

Martin, F., Tuskan, G. A., DiFazio, S. P., Lammers, P., Newcombe, G., and Podila, G. K. 2004. Symbiotic sequencing for the *Populus mesocosm*. *New Phytol.* **161**: 330–335.

Mason, M. G., Mathews, D. E., Argyros, D. A., Maxwell, B. B., Kieber, J. J., Alonso, J. M., Ecker, J. R., and Schaller, G. E. 2005. Multiple type-B response regulators mediate cytokinin signal transduction in *Arabidopsis*. *Plant Cell* **17**: 3007–3018.

Matsubara, S., Hurry, V., Druart, N., Benedict, C., Janzik, I., Chavarria-Krauser, A., Walter, A., and Schurr, U. 2006. Nocturnal changes in leaf growth of *Populus deltoides* are controlled by cytoplasmic growth. *Planta* **223**: 1315–1328.

McHale, L., Tan, X. P., Koehl, P., and Michelmore, R. W. 2006. Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* **7**: 212.

Meijer, M., and Murray, J. A. H. 2000. The role and regulation of D-type cyclins in the *Plant Cell* cycle. *Plant Mol. Biol.* **43**: 621–633.

Menges, M., Pavesi, G., Morandini, P., Bogre, L., and Murray, J. A. 2007. Genomic organization and evolutionary conservation of plant D-type cyclins. *Plant Physiol.* **145**: 1558–1576.

Meyermans, H., Morreel, K., Lapierre, C., Pollet, B., De Bruyn, A., Busson, R., Herdewijn, P., Devreese, B., Van Beeumen, J., Marita, J. M., *et al.* 2000. Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A O-methyltransferase, an enzyme involved in lignin biosynthesis. *J. Biol. Chem.* **275**: 36899–36909.

Miranda, M., Ralph, S. G., Mellway, R., White, R., Heath, M. C., Bohlmann, J., and Constabel, C. P. 2007. The transcriptional response of hybrid poplar (*Populus trichocarpa* × *P. deltoides*) to infection by *Melampsora medusae* leaf rust involves induction of flavonoid pathway genes leading to the accumulation of proanthocyanidins. *Mol. Plant-Microbe Interact.* **20**: 816–831.

Mo, B., and Bewley, J. D. 2003. The relationship between beta-mannosidase and endo-beta-mannanase activities in tomato seeds during and following germination: a comparison of seed populations and individual seeds. *J. Exp. Bot.* **54**: 2503–2510.

Moreau, C., Aksenov, N., Lorenzo, M. G., Segerman, B., Funk, C., Nilsson, P., Jansson, S., and Tuominen, H. 2005. A genomic approach to investigate developmental cell death in woody tissues of *Populus* trees. *Genome Biol.* **6**: R34.

Morreel, K., Goeminne, G., Storme, V., Sterck, L., Ralph, J., Coppieters, W., Breyne, P., Steenackers, M., Georges, M., Messens, E., *et al.* 2006. Genetical metabolomics of flavonoid biosynthesis in *Populus*: a case study. *Plant J.* **47**: 224–237.

Morse, A. M., Tschaplinski, T. J., Dervinis, C., Pijut, P. M., Schmelz, E. A., Day, W., and Davis, J. M. 2007. A salicylate hydroxylase transgene in poplar induces compensatory mechanisms in the shikimate and phenylpropanoid pathways. *Phytochemistry* **68**: 2043–2052.

Nakano, T., Suzuki, K., Fujimura, T., and Shinshi, H. 2006. Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. *Plant Physiol.* **140**: 411–432.

Nanjo, T., Futamura, N., Nishiguchi, M., Igasaki, T., Shinozaki, K., and Shinohara, K. 2004. Characterization of full-length enriched expressed sequence tags of stress-treated poplar leaves. *Plant Cell Physiol.* **45**: 1738–1748.

Nanjo, T., Sakurai, T., Totoki, Y., Toyoda, A., Nishiguchi, M., Kado, T., Igasaki, T., Futamura, N., Seki, M., Sakaki, Y., *et al.* 2007. Functional annotation of 19,841 *Populus nigra* full-length enriched cDNA clones. *BMC Genomics* 8:448.

Neale, D. B., and Savolainen, O. 2004. Association genetics of complex traits in conifers. *Trends Plant Sci.* **9**: 325–330.

Nilsson, J., Karlberg, A., Antti, H., Lopez-Vernaza, M., Mellerowicz, E., Perrot-Rechenmann, C., Sandberg, G., and Bhalerao, R. P. 2008. Dissecting the molecular basis of the regulation of wood formation by auxin in hybrid aspen. *Plant Cell* **20**: 843–855.

Oakley, R. V., Wang, Y. S., Ramakrishna, W., Harding, S. A., and Tsai, C. J. 2007. Differential expansion and expression of alpha- and beta-tubulin gene families in *Populus*. *Plant Physiol.* **145**: 961–973.

Okamuro, J. K., Caster, B., Villarroel, R., Van Montagu, M., and Jofuku, K. D. 1997. The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **94**: 7076–7081.

Pandey, A., and Mann, M. 2000. Proteomics to study genes and genomes. *Nature*. **405**: 837–846.

Parenicova, L., de Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., Cook, H. E., Ingram, R. M., Kater, M. M., Davies, B., *et al.* 2003. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: New openings to the MADS world. *Plant Cell* **15**: 1538–1551.

Parker, J. D., Bradley, B. A., Mooers, A. O., and Quarmby, L. M. 2007. Phylogenetic analysis of the Neks reveals early diversification of ciliary-cell cycle kinases. PLoS ONE **2**: e1076.

Pietrowska-Borek, M., Stuible, H. P., Kombrink, E., and Guranowski, A. 2003. 4-Coumarate: coenzyme A ligase has the catalytic capacity to

synthesize and reuse various (di)adenosine polyphosphates. *Plant Physiol* **131**: 1401–1410.

Plomion, C., Lalanne, C., Claverol, S., Meddour, H., Kohler, A., Bogeat-Triboulot, M. B., Barre, A., Le Provost, G., Dumazet, H., Jacob, D., *et al.* 2006. Mapping the proteome of poplar and application to the discovery of drought-stress responsive proteins. *Proteomics* 6**:** 6509–6527.

Qin, H., Feng, T., Harding, S. A., Tsai, C. J., and Zhang, S. 2008. An efficient method to identify differentially expressed genes in microarray experiments. *Bioinformatics* **24**: 1583–1589.

Quesada, T., Li, Z., Dervinis, C., Li, Y., Bocock, P. N., Tuskan, G. A., Casella, G., Davis, J. M., and Kirst, M. 2008. Comparative analysis of the transcriptomes of *Populus trichocarpa* and *Arabidopsis thaliana* suggests extensive evolution of gene expression regulation in angiosperms. *New Phytol.* **180**: 408–420.

Rae, A. M., Pinel, M. P. C., Bastien, C., Sabatti, M., Street, N. R., Tucker, J., Dixon, C., Marron, N., Dillen, S. Y., and Taylor, G. 2008. QTL for yield in bioenergy *Populus*: identifying G×E interactions from growth at three contrasting sites. *Tree Genetics & Genomes* **4**: 97–112.

Rae, A. M., Tricker, P. J., Bunn, S. M., and Taylor, G. 2007. Adaptation of tree growth to elevated CO2: quantitative trait loci for biomass in *Populus*. *New Phytol.* **175**: 59–69.

Rajinikanth, M., Harding, S. A., and Tsai, C. J. 2007. The glycine decarboxylase complex multienzyme family in *Populus*. *J. Exp. Bot.* **58**: 1761–1770.

Ralph, S., Oddy, C., Cooper, D., Yueh, H., Jancsik, S., Kolosova, N., Philippe, R. N., Aeschliman, D., White, R., Huber, D., *et al.* 2006. Genomics of hybrid poplar (*Populus trichocarpa* × *deltoides*) interacting with forest tent caterpillars (*Malacosoma disstria*): normalized and full-length cDNA libraries, expressed sequence tags, and a cDNA microarray for the study of insect-induced defences in poplar. *Mol. Ecol.* **15**: 1275–1297.

Ralph, S. G., Chun, H. J. E., Cooper, D., Kirkpatrick, R., Kolosova, N., Gunter, L., Tuskan, G. A., Douglas, C. J., Holt, R. A., Jones, S. J. M., *et al.* 2008. Analysis of 4,664 high-quality sequence-finished poplar full-length cDNA clones and their utility for the discovery of genes responding to insect feeding. *BMC Genomics* **9**: 57.

Ramirez-Carvajal, G. A., Morse, A. M., and Davis, J. M. 2008. Transcript profiles of the cytokinin response regulator gene family in *Populus* imply diverse roles in plant development. *New Phytol.* **177**: 77–89.

Ranjan, P., Kao, Y. Y., Jiang, H., Joshi, C. P., Harding, S. A., and Tsai, C. J. 2004. Suppression subtractive hybridization-mediated transcriptome analysis from multiple tissues of aspen (*Populus tremuloides*) altered in phenylpropanoid metabolism. *Planta* **219**: 694–704.

Rinaldi, C., Kohler, A., Frey, P., Duchaussoy, F., Ningre, N., Couloux, A., Wincker, P., Le Thiec, D., Fluch, S., Martin, F., *et al.* 2007. Transcript profiling of poplar leaves upon infection with compatible and incompatible strains of the foliar rust *Melampsora larici-populina*. *Plant Physiol.* **144**: 347–366.

Roberts, A. W., and Bushoven, J. T. 2007. The cellulose synthase (CESA) gene superfamily of the moss *Physcomitrella patens*. *Plant Mol. Biol.* **63**: 207–219.

Rohde, A., Ruttink, T., Hostyn, V., Sterck, L., Van Driessche, K., and Boerjan, W. 2007. Gene expression during the induction, maintenance, and release of dormancy in apical buds of poplar. *J. Exp. Bot.* **58**: 4047–4060.

Rose, J. K. C., Braam, J., Fry, S. C., and Nishitani, K. 2002. The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: Current perspectives and a new unifying nomenclature. *Plant Cell Physiol.* **43**: 1421–1435.

Ruttink, T., Arend, M., Morreel, K., Storme, V., Rombauts, S., Fromm, J., Bhalerao, R. P., Boerjan, W., and Rohde, A. 2007. A molecular timetable for apical bud formation and dormancy induction in poplar. *Plant Cell* **19**: 2370–2390.

Sasaki, S., Shimizu, M., Wariishi, H., Tsutsumi, Y., and Kondo, R. 2007. Transcriptional and translational analyses of poplar anionic peroxidase isoenzymes. *J. Wood Sci*. **53**: 427–435.

Scarascia-Mugnozza, G. E., Hinckley, T. M., Stettler, R. F., Heilman, P. E., and Isebrands, J. G. 1999. Production physiology and morphology of *Populus* species and their hybrids grown under short rotation. III. Seasonal carbon allocation patterns from branches. *Can. J. Forest Res*. **29**: 1419–1432.

Schrader, J., Nilsson, J., Mellerowicz, E., Berglund, A., Nilsson, P., Hertzberg, M., and Sandberg, G. 2004. A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. *Plant Cell* **16**: 2278–2292.

Service, R. F. 2006. Gene sequencing. The race for the $1000 genome. *Science* **311**: 1544–1546.

Sjödin, A., Street, N., Sandberg, G., Gustafsson, P., and Jansson, S. 2008. PopGenIE: The *Populus* Genome Integrative Explorer. A new tool for exploring the *Populus* genome. www.popgenie.org.

Sjodin, A., Wissel, K., Bylesjo, M., Trygg, J., and Jansson, S. 2008. Global expression profiling in leaves of free-growing aspen. *BMC Plant Biol.* **8**: 61.

Song, J., Lu, S., Chen, Z. Z., Lourenco, R., and Chiang, V. L. 2006. Genetic transformation of *Populus trichocarpa* genotype Nisqually-1: a functional genomic tool for woody plants. *Plant Cell Physiol*. **47**: 1582–1589.

Souza Cde, A., Barbazuk, B., Ralph, S. G., Bohlmann, J., Hamberger, B., and Douglas, C. J. 2008. Genome-wide analysis of a land plant-specific acyl:coenzyme A synthetase (ACS) gene family in *Arabidopsis*, poplar, rice and *Physcomitrella*. *New Phytol.* **179**: 987–1003.

Sterky, F., Bhalerao, R. R., Unneberg, P., Segerman, B., Nilsson, P., Brunner, A. M., Charbonnel-Campaa, L., Lindvall, J. J., Tandre, K., Strauss, S. H., *et al.* 2004. A *Populus* EST resource for plant functional genomics. *Proc. Nat. Acad. Sciences USA* **101**: 13951–13956.

Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A., Holmberg, A., Amini, B., Bhalerao, R., Larsson, M., and Villarroel, R. 1998. Gene discovery in the wood-forming tissues of poplar: analysis of 5,692 expressed sequence tags. *Proc. Natl. Acad. Sci. USA* **95**: 13330–13335.

Stracke, R., Werber, M., and Weisshaar, B. 2001. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* **4**: 447–456.

Strauss, S. H. 2003. Genetic technologies – Genomics, genetic engineering, and domestication of crops. *Science* **300**: 61–62.

Strauss, S. H., Lande, R., and Namkoong, G. 1992. Limitations of molecular-marker-aided selection in forest tree breeding. *Can. J. Forest Res.* **22**: 1050–1061.

Street, N. R., Skogstrom, O., Sjodin, A., Tucker, J., Rodriguez-Acosta, M., Nilsson, P., Jansson, S., and Taylor, G. 2006. The genetics and genomics of the drought response in *Populus*. *Plant J.* **48**: 321–341.

Sun, M., Zhou, G., Lee, S., Chen, J., Shi, R. Z., and Wang, S. M. 2004. SAGE is far more sensitive than EST for detecting low-abundance transcripts. *BMC Genomics* **5**: 1.

Suzuki, S., Li, L., Sun, Y. H., and Chiang, V. L. 2006. The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. *Plant Physiol.* **142**: 1233–1245.

Tagu, D., Bastien, C., Faivre-Rampant, P., Garbaye, J., Vion, P., Villar, M., and Martin, F. 2005. Genetic analysis of phenotypic variation for ectomycorrhiza formation in an interspecific F1 poplar full-sib family. *Mycorrhiza* **15**: 87–91.

Takata, N., Saito, S., Tanaka Saito, C., Nanjo, T., Shinohara, K., and Uemura, M. 2009. Molecular phylogeny and expression of poplar circadian clock genes, LHY1 and LHY2. *New Phytol.* **181**: 808–819.

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. 2008a. Synteny and collinearity in plant genomes. *Science* **320**: 486–488.

Tang, H., Wang, X., Bowers, J. E., Ming, R., Alam, M., and Paterson, A. H. 2008b. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res*. **18**: 1944–1954.

Taylor, G. 2002. *Populus*: Arabidopsis for forestry. Do we need a model tree? *Ann. Bot.* **90**: 681–689.

To, J. P., Deruere, J., Maxwell, B. B., Morris, V. F., Hutchison, C. E., Ferreira, F. J., Schaller, G. E., and Kieber, J. J. 2007. Cytokinin regulates type-A *Arabidopsis* Response Regulator activity and protein stability via two-component phosphorelay. *Plant Cell* **19**: 3901–3914.

Tringe, S. G., and Rubin, E. M. 2005. Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics* **6**: 805–814.

Tsai, C. J., Harding, S. A., Tschaplinski, T. J., Lindroth, R. L., and Yuan, Y. N. 2006. Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. *New Phytol.* **172**: 47–62.

Tsai, C. J., Popko, J. L., Mielke, M. R., Hu, W. J., Podila, G. K., and Chiang, V. L. 1998. Suppression of O-methyltransferase gene by homologous sense transgene in quaking aspen causes red-brown wood phenotypes. *Plant Physiol.* **117**: 101–112.

Tsai, C. J., Ranjan, P., DiFazio, S. P., Tuskan, G. A., Johnson, V., and Joshi, C. P. 2009. Poplar genome microarrays. **In**: Joshi, C. P., and DiFazio, S. P., Eds., *Genetics*, *Genomics and Breeding of Crop Plants: Poplar.* Science Publishers, Enfield, New Hampshire.

Tschaplinski, T. J., Tuskan, G. A., Sewell, M. M., Gebre, G. M., Donald, E. T. I., and Pendleyi, C. 2006. Phenotypic variation and quantitative trait locus identification for osmotic potential in an interspecific hybrid inbred F-2 poplar pedigree grown in contrasting environments. *Tree Physiol.* **26**: 595–604.

Tschaplinski, T. J., Yin, T.-M., and Engle, N. 2005. Combining metabolomics and QTL analysis for identifying mQTL and gene discovery in poplar. Breakthrough Technologies Forum of IUFRO Tree Biotechnology Meeting, Pretoria, South Africa, November 6–11, 2005. Abstract S3.5.

Turck, F., Fornara, F., and Coupland, G. 2008. Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. *Ann. Rev. Plant Biol*. **59**: 573–594.

Tuskan, G., and Walsh, M. 2001. Short-rotation woody crop systems, atmospheric carbon dioxide and carbon management: a US case study. *Forestry Chronicle* **77**: 259–264.

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., *et al.* 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.

Tuskan, G. A., DiFazio, S. P., and Teichmann, T. 2004. Poplar genomics is getting popular: The impact of the poplar genome project on tree research. *Plant Biol.* **6**: 2–4.

Unneberg, P., Stromberg, M., Lundeberg, J., Jansson, S., and Sterky, F. 2005. Analysis of 70,000 EST sequences to study divergence between two closely related *Populus* species. *Tree Genetics & Genomes* **1**: 109–115.

van der Hoorn, R. A. L., and Jones, J. D. 2004. The plant proteolytic machinery and its role in defence. *Curr. Opin. Plant Biol.* **7**: 400–407.

Vigneault, F., Lachance, D., Cloutier, M., Pelletier, G., Levasseur, C., and Seguin, A. 2007. Members of the plant NIMA-related kinases are involved in organ development and vascularization in poplar, Arabidopsis and rice. *Plant J.* **51**: 575–588.

Wan, J., Zhang, X. C., Neece, D., Ramonell, K. M., Clough, S., Kim, S. Y., Stacey, M. G., and Stacey, G. 2008. A LysM receptor-like kinase plays a critical role in chitin signaling and fungal resistance in *Arabidopsis*. *Plant Cell* **20**: 471–481.

Watkins, P. A., Maiguel, D., Jia, Z., and Pevsner, J. 2007. Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome. *J. Lipid Res.* **48**: 2736–2750.

Weckwerth, W. 2003. Metabolomics in systems biology. *Ann. Rev. Plant Biol*. **54**: 669–689.

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., *et al.* 2000. Activation tagging in *Arabidopsis*. *Plant Physiol.* **122**: 1003–1013.

Whitham, T. G., DiFazio, S. P., Schweitzer, J. A., Shuster, S. M., Allan, G. J., Bailey, J. K., and Woolbright, S. A. 2008. Perspective – Extending genomics to natural communities and ecosystems. *Science* **320**: 492–495.

Wilkins, O., Nahal, H., Foong, J., Provart, N. J., and Campbell, M. M. 2009. Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiol.* **149**: 981–993.

Wise, R. P., Caldo, R. A., Hong, L., Shen, L., Cannon, E., and Dickerson, J. A. 2007. BarleyBase/PLEXdb. *Methods Mol. Biol.* **406**: 347–363.

Wu, R., Bradshaw, H. D. J., and Stettler, R. F. 1997. Molecular genetics of growth and development in *Populus* (Salicaceae). V. Mapping quantitative trait loci affecting leaf variation. *American Journal of Beauty*. **84**: 143–153.

Wu, R. L. 1998. Genetic mapping of QTLs affecting tree growth and architecture in *Populus*: implication for ideotype breeding 3079. *Theoret. Applied Genet.* **96**: 447–457.

Wu, R. L., and Lin, M. 2006. Opinion – Functional mapping – how to map and study the genetic architecture of dynamic complex traits. *Nature Rev. Genetics* **7**: 229–237.

Wullschleger, S., Yin, T. M., DiFazio, S. P., Tschaplinski, T. J., Gunter, L. E., Davis, M. F., and Tuskan, G. A. 2005. Phenotypic variation in growth and biomass distribution for two advanced-generation pedigrees of hybrid poplar. *Can. J. Forest Res.* **35**: 1779–1789.

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

Wullschleger, S. D., Tuskan, G. A., and DiFazio, S. P. 2002. Genomics and the tree physiologist. *Tree Physiol.* **22**: 1273–1276.

Yang, X. H., Jawdy, S., Tschaplinski, T. J., and Tuskan, G. A. 2009. Genome-wide identification of lineage-specific genes in *Arabidopsis*, *Oryza* and *Populus*. *Genomics* **93**: 473–480.

Yang, X. H., Kalluri, U. C., Jawdy, S., Gunter, L. E., Yin, T. M., Tschaplinski, T. J., Weston, D. J., Ranjan, P., and Tuskan, G. A. 2008. The F-box gene family is expanded in herbaceous annual plants relative to woody perennial plants. *Plant Physiol.* **148**: 1189–1200.

Yang, X. H., Tuskan, G. A., and Cheng, Z. M. 2006. Divergence of the Dof gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. *Plant Physiol.* **142**: 820–830.

Ye, X., Kang, B. G., Osburn, L. D., and Cheng, Z. M. 2009a. The COBRA gene family in *Populus* and gene expression in vegetative organs and in response to hormones and environmental stresses. *Plant Growth Regulation* **58**: 211–223.

Ye, X., Kang, B. G., Osburn, L. D., Li, Y., and Cheng, Z. M. 2009b. Identification of the flavindependent monooxygenase-encoding YUCCA gene family in *Populus trichocarpa* and their expression in vegetative tissues and in response to hormone and environmental stresses. *Plant Cell Tissue and Organ Culture* **97**: 271–283.

Yin, T., DiFazio, S. P., Gunter, L. E., Zhang, X., Sewell, M. M., Woolbright, S. A., Allan, G. J., Kelleher, C. T., Douglas, C. J., Wang, M., *et al.* 2008. Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. *Genome Res.* **18**: 422–430.

Yuan, J. S., Yang, X. H., Lai, J. R., Lin, H., Cheng, Z. M., Nonogaki, H., and Chen, F. 2007. The endo-beta-mannanase gene families in *Arabidopsis*, rice, and poplar. *Funct. Integ. Genomics* **7**: 1–16.

Zhang, H. C., Yin, W. L., and Xia, X. L. 2008. Calcineurin B-Like family in *Populus*: comparative genome analysis and expression pattern under cold, drought and salt stress treatment. *Plant Growth Reg.* **56**: 129–140.

Zhang, X. C., Wu, X., Findley, S., Wan, J., Libault, M., Nguyen, H. T., Cannon, S. B., and Stacey, G. 2007. Molecular evolution of lysin motif-type receptor-like kinases in plants. *Plant Physiol.* **144**: 623–636.

Zhou, F., and Xu, Y. 2009. RepPop: a database for repetitive elements in *Populus trichocarpa*. *BMC Genomics* **10**: 14.

Zhu, Q. H., Guo, A. Y., Gao, G., Zhong, Y. F., Xu, M., Huang, M., and Luo, J. 2007. DPTF: a database of poplar transcription factors. *Bioinformatics* **23**: 1307–1308.

Zhuang, J., Cai, B., Peng, R. H., Zhu, B., Jin, X. F., Xue, Y., Gao, F., Fu, X. Y., Tian, Y. S., Zhao, W., *et al.* 2008. Genome-wide analysis of the AP2/ERF gene family in *Populus trichocarpa*. *Biochem. Biophys. Res. Commun*. **371**: 468–474.

Zhuang, Y., and Adams, K. L. 2007. Extensive allelic variation in gene expression in *Populus* F_1 hybrids. *Genetics* **177**: 1987–1996.