Tree Physiology Advance Access published October 25, 2012

Tree Physiology 00, 1–8
doi:10.1093/treephys/tps081

Commentary

Revisiting the sequencing of the first tree genome: *Populus trichocarpa*

Stan D. Wullschleger[1,4], D. J. Weston[2], S. P. DiFazio[3] and G. A. Tuskan[2]

[1]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6301, USA; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6422, USA; [3]Department of Biology, West Virginia University, Morgantown, WV 26506-6057, USA; [4]Corresponding author (wullschlegsd@ornl.gov)

Received March 5, 2012; accepted August 1, 2012; handling Editor Danielle Way

Ten years ago, it was announced that the Joint Genome Institute with funds provided by the Department of Energy, Office of Science, Biological and Environmental Research would sequence the black cottonwood (*Populus trichocarpa* Torr. & Gray) genome. This landmark decision was the culmination of work by the forest science community to develop *Populus* as a model system. Since its public release in late 2006, the availability of the *Populus* genome has spawned research in plant biology, morphology, genetics and ecology. Here we address how the tree physiologist has used this resource. More specifically, we revisit our earlier contention that the rewards of sequencing the *Populus* genome would depend on how quickly scientists working with woody perennials could adopt molecular approaches to investigate the mechanistic underpinnings of basic physiological processes. Several examples illustrate the integration of functional and comparative genomics into the forest sciences, especially in areas that target improved understanding of the developmental differences between woody perennials and herbaceous annuals (e.g., phase transitions). Sequencing the *Populus* genome and the availability of genetic and genomic resources has also been instrumental in identifying candidate genes that underlie physiological and morphological traits of interest. Genome-enabled research has advanced our understanding of how phenotype and genotype are related and provided insights into the genetic mechanisms whereby woody perennials adapt to environmental stress. In the future, we anticipate that low-cost, high-throughput sequencing will continue to facilitate research in tree physiology and enhance our understanding at scales of individual organisms and populations. A challenge remains, however, as to how genomic resources, including the *Populus* genome, can be used to understand ecosystem function. Although examples are limited, progress in this area is encouraging and will undoubtedly improve as future research targets the many unique aspects of *Populus* as a keystone species in terrestrial ecosystems.

Keywords: comparative genomics, ecology, functional genomics, molecular biology, whole-tree physiology.

Introduction

Sequencing of the *Populus trichocarpa* Torr. & Gray genome (Tuskan et al. 2006) and the subsequent development of genetic and genomic resources have solidified the role of *Populus* as a model organism for molecular studies in woody perennial research and forestry (Taylor 2002, Wullschleger et al. 2002*a*, Tuskan et al. 2004, Jansson and Douglas 2007). The availability of *Populus* as a model system for woody perennial plant biology and the availability of the genome sequence promises to expand our understanding of wood development, flowering and dormancy (Hsu et al. 2011), interactions of woody perennials with other organisms (Gottel et al. 2011, Labbe et al. 2011) and the genetic underpinnings of natural variation and plant adaptation to environmental change (He et al. 2010). Although some molecular aspects of these topics can be explored using resources developed for *Arabidopsis* (Chaffey et al. 2002), there is general agreement

that woody perennials embody a distinct spectrum of structure and function that encompasses long life spans, extended generation times and perennial growth habits (Jansson and Douglas 2007). Therefore, many tree-specific questions can best be addressed using experimental approaches that combine physiological knowledge with information contained in the *Populus* genome. Although implementing such experiments can be difficult, there is encouraging evidence that such a promise is being realized in forest biology (Wullschleger et al. 2009, Yang et al. 2009, Neale and Kremer 2011).

A decade ago we emphasized that with the sequencing of the *Populus* genome a series of opportunities would arise for the tree physiologist to use this emerging resource (Wullschleger et al. 2002*b*). We speculated that rapid integration of functional and comparative genomics into the forest sciences would occur and thus facilitate research related to (i) developmental differences between woody perennials and herbaceous annuals, (ii) mechanisms that underlie phenotypic plasticity and adaptation of long-lived species to their edaphic and climatic environment, (iii) molecular insights into community ecology and ecosystem function and (iv) the use of new genomic tools and approaches to investigate causal relationships that are currently intractable using conventional methodologies. Here we revisit the sequencing of the *Populus* genome and, within the context of these topics, address how the tree physiologist has used this resource. A series of examples broadly illustrate where progress has been made and where opportunities remain. We find that the tree physiologist has used the *Populus* genome and associated molecular approaches to better understand phase transitions (e.g., flowering), interannual nutrient cycling and dormancy traits that are unique to woody perennials (Rennenberg et al. 2010), and to characterize the breadth and significance of genetic variation within natural populations (Slavov et al. 2012). Recent research has begun to emphasize plant–microbe interactions, focusing on *Populus* and its associated microbiome as a model system (Gottel et al. 2011). Additionally, we highlight how physiological studies have been facilitated by the rapid emergence of next-generation sequencing technology and the development of bioinformatic tools and capabilities. Finally, we show how sequencing of the *Populus* genome has enabled research relevant to ecosystem-scale processes, thereby opening opportunities for the tree physiologist to transcend organismal boundaries with powerful new sequencing techniques.

Trees differ in many ways from herbaceous annuals

The availability of the *Populus* genome makes it possible to ask, as did Groover (2005) and Petit and Hampe (2006), what genes make a tree a tree? We know that great size, long life span and phase transitions are distinguishing features of woody perennials compared with herbaceous annuals. We also know that size and longevity complicate almost every aspect of tree physiological studies. However, it is partly because of their size and life span that woody perennials differ in so many respects from herbaceous plants: apical dominance, growth habit, complex crown form, dormancy and seasonal nutrient reallocation. Several recent reviews have summarized progress in these areas and there is ample evidence for how the tree physiologist is integrating molecular approaches into studies of carbon uptake and allocation, nutrient dynamics, water transport, stress tolerance mechanisms and more (Wullschleger et al. 2009, Polle and Douglas 2012). In addition, developmental phase transitions (e.g., juvenile to mature and vegetative to reproductive) are also a distinguishing feature of woody perennials, with advances in our understanding of flowering as a good example for how the *Populus* genome has allowed insights to be gained by the tree physiologist.

In annuals and perennials, flowering is regulated by a complex network of pathways that integrate information from various endogenous and environmental cues (Huijser and Schmid 2011). Annual plants grow, reproduce and senesce within a single growing season, whereas woody perennials typically display multiple years of vegetative growth before they reach sexual maturity (Brunner and Nilsson 2004, Mohamed et al. 2010). In *Populus*, flowering occurs in late spring and early summer, typically before vegetative bud burst. The overwintering buds are either vegetative or reproductive so the decision to flower in the coming year is set by signals received before bud set (Jansson and Douglas 2007).

We now know that the transition between juvenile and reproductive maturity is genetically controlled. Using information contained in the *Populus* genome, Bohlenius et al. (2006) and Hsu et al. (2006) examined the role of specific genes in regulating the juvenile-to-adult phase transition in *Populus* (i.e., first time to flower, reproductive onset) and in the regulation of seasonal flowering in mature shoots. These authors showed that *FLOWERING LOCUS T1* (*FT1*) and *FLOWERING LOCUS T2* (*FT2*) could be manipulated to induce early flowering in *Populus*. In a series of subsequent studies, Hsu et al. (2011) used manipulative physiological and genetic experiments coupled with field studies, expression profiling and network analysis to reveal that reproductive onset is determined by *FT1* in response to winter temperatures, whereas vegetative growth and inhibition of bud set are promoted by *FT2* in response to warm temperatures and long days in the growing season. We now also have a working knowledge of microRNAs (miRNAs) and their occurrence across the genome. MicroRNAs are noncoding RNAs that appear to regulate target genes by binding to complementary sequences located in the transcripts produced by these genes (Aukerman and Sakai 2003). Two highly conserved miRNAs, miRNA156 and miRNA172, and their targets have been identified as key components of the genetic control

Table 1. Fundamental changes in approaches to physiological research have been enabled since the transition from the relatively low-throughput technologies that prevailed at the time of our first perspective in 2002, and the high-throughput technologies enabled by current technology.

	2002	2012	Implications
Genotyping	Amplicon sequencing; single-base extension; microsatellites; AFLP	Bead arrays; genotyping by sequencing; whole-genome resequencing	The availability of inexpensive, high-throughput technologies has made high-density genotyping of entire populations feasible. This can provide insights into the present and historical distribution of genetic variation in a species, a prerequisite for understanding physiological variation.
Genotype–phenotype associations	QTL analysis, candidate gene resequencing	Whole-genome association studies	This will enable direct identification of alleles that alter physiological traits, including regulatory elements located in non-coding sequences. Furthermore, whole-genome association analysis will lead to novel discoveries of genes involved in physiological functions.
Transcriptomics	ESTs; differential display; microarrays	RNASeq	Direct sequencing of transcripts has opened whole-genome expression analysis to any organism of physiological interest. Furthermore, the ability to monitor allele-specific expression and alternative splicing provides novel insights into physiological processes.
Community analysis	tRFLP; ITS sequencing; rRNA sequencing	Metagenome sequencing	The ability to quantify functionally significant community variation will enable exploration of tree physiology within the complete biotic context of trees in natural populations.

AFLP, amplified fragment length polymorphism; QTL, quantitative trait locus; tRFLP, terminal restriction-fragment-length polymorphism; ITS, internal transcribed spacer.

mechanisms that underlie plant phase changes (Huijser and Schmid 2011). Wang et al. (2011) demonstrated that changes in the relative abundance of miRNA156 and miRNA172 in *Populus* are associated with the transition from juvenile to mature phase.

Armed with this information, the tree physiologist is positioned to use the *Populus* genome in support of forest tree breeding and accelerated domestication of tree crops. Gaining a thorough understanding of the mechanisms that regulate competence to flowering is important, especially in perennial plants, since a shortened juvenile phase could greatly facilitate breeding programs (Meilan 1997, Bergonze and Albani 2011). One way that this can be accomplished is through altering the temporal constraints to the breeding cycle by shortening the time to reproductive maturity. Neale and Kremer (2011) argue that progress can be made in this area by using direct genetic engineering and indirect marker-assisted selection approaches. Evidence is not yet available to suggest that either approach can be used to appreciably shorten flowering time, but access to the *Populus* genome is providing a new tool with which to explore this possibility.

Mechanisms underlying genetic diversity in natural populations

In our original perspective (Wullschleger et al. 2002*b*), we commented that genomics tools would facilitate scaling physiological understanding from cells to ecosystems. While we are clearly far from meeting this monumental challenge, recent developments have supported the validity of this assertion. Notably, the recent revolution in high-throughput genotyping and sequencing has caused a fundamental change in the way we approach genomic research (Table 1). For example, the advent of population resequencing has revolutionized the study of natural variation in forest tree populations, thereby enabling elucidation of the molecular determinants of adaptive variation at a landscape scale. Keller et al. (2011) and Hall et al. (2011) surveyed nucleotide variation in relation to geographic location in the boreal species *Populus balsamifera* L. and *Populus tremula* L., respectively, and successfully identified genes and polymorphisms showing signatures of purifying and positive selection. Breen et al. (2009) reported nucleotide diversity among natural boreal and Arctic populations of *P. balsamifera*, whereas Slavov et al. (2012) show that polymorphisms (single-nucleotide polymorphisms (SNPs)) in and near certain genes show stronger latitudinal variation than non-coding SNPs, suggesting that natural selection is operating to form locally adapted populations in *P. trichocarpa*. Population-scale resequencing approaches will provide new insights into molecular adaptation as true whole-genome resequencing of large population samples becomes more widespread. Such studies would not be feasible without a high-quality reference genome, which enables rapid and unambiguous genotyping at population scales.

The forest tree genetics community has fully embraced the association of nucleotide variation, especially SNPs, to adaptive traits of interest (e.g., Wegrzyn et al. 2010, Eckert et al. 2012). This should continue to enable the tree physiologist to pursue a more complete understanding of how genetic variation translates to phenotypic plasticity in natural environments. Although SNPs are a major source of genetic and phenotypic variation, we envision that resequencing efforts will discover additional sources of genetic variation. This has been the case in human genetics where genome-scanning algorithms have revealed substantial structural genetic variation due to deletions,

duplications, insertions, inversions and translocations (Feuk et al. 2006). Current resequencing efforts are now being conducted at the population level and will enable the discovery and evaluation of structural variation contributing to genetic and phenotypic variation. A major limitation to the tree physiologist in exploring this source of variation is the elaborate and often customized bioinformatic workflows necessary to explore such data. However, a number of bioinformatics applications and computational resources are now available to analyze sequencing data using cloud and parallel computing (Schatz et al. 2010), and there are recent efforts to make these resources available to the plant biologist (e.g., http://kbase.us/). The combination of population resequencing efforts with sophisticated bioinformatic resources will surely lead the tree physiologist to novel insights on the sources of genetic variation underlying adaptation, including those driving physiological traits.

Ecosystem function and community genomics

Populus, as a perennial woody plant, occupies a keystone position within many ecosystems and as such creates an environment that promotes, forms and conserves assorted communities of associated species, including bacteria, fungi, insects and herbivores (Whitham et al. 2008). Shortly after the release of the *Populus* genome, there was a call to sequence many of the bacteria and fungi that comprise the *Populus* megagenome (Martin et al. 2004, Cheng and Tuskan 2009). As a result, the *Laccaria* genome (Martin et al. 2008), the *Melampsora* genome (Duplessis et al. 2011) and the *Glomus* transcriptome (Tisserant et al. 2012) have been sequenced, assembled, annotated and released to the public. In addition, numerous bacterial genomes, both endophytic and rhizospheric, have also been sequenced and released (van der Lelie et al. 2009, Brown et al. 2012).

These genomic resources have the potential to contribute to an unprecedented understanding of how *Populus* interacts with its associated microbial members (also known as the microbiome), which together act in a coordinated manner to shape host physiology and performance. Emerging results suggest that the number of potential *Populus* microbial associates is staggering. The *Populus* sequencing project identified numerous sequence reads that mapped to microorganisms spanning 3 kingdoms, 37 genera and 99 species (Tuskan et al. 2006, Appendix Table S3). Similarly, a metagenomics study by Gottel et al. (2011) found that washed roots of *Populus deltoides* W. Bartram ex Marshall contained a broad array of rhizospheric and endophytic bacteria and fungal communities, suggesting that the tissues within naturally occurring *Populus* roots represent a unique niche for microbial communities. Given these observations, and their significance to plant–microbe signaling pathways (Weston et al. 2012*b*), we contend that unlocking the genome sequence of not only the *Populus* host, but the associated fungi, bacteria and pathogens offers an exceptional opportunity to dissect how the interaction of genomes influences host plant function. We further contend that the sequencing of the *Populus* microbiome is not only a natural extension to the *Populus* genome project, but is a necessary step if we are to understand *Populus* physiology within the necessary context of interacting genomes. For example, it was recently shown that *Populus* leaf endophytes greatly reduced rust (*Melampsora* spp.) severity regardless of host genotype tested. This suggests that disease resistance is not entirely driven by the plant, but rather by a coordinated system comprised of *Populus* and its corresponding microbiome.

The rapid emergence of sequenced genomes and the genome-based technologies developed from those resources are already leading to new insights into the intimate relationship between plants and microbes. Likewise, these same resources could also lead to improved understanding of how genes shape the structure and function of terrestrial ecosystems and those insights could, in turn, help us better predict the response of plants, plant communities and ecosystems to biotic and abiotic stresses. In addition to the plant–microbiome connections, Whitham et al. (2008) proposed extending genomics to natural communities and ecosystems, arguing that feedback loops among intimately associated organisms cause the development of coevolving communities that are heritable at the level of the host tree, with large implications for ecosystem functions such as primary productivity and nutrient cycling. Indeed, Wymore et al. (2011) has taken the basic genes-to-ecosystem framework and put forth a series of postulates that describe how causal relationships between genes and their community or ecosystem consequences might be propagated. As a keystone species in many riparian habitats, *Populus* and other tree species occupy a central role in community or ecosystem dynamics and thus provide an excellent opportunity to evaluate this framework.

Genome-enabled technology overcomes constraints of existing methods

Microarrays were envisioned in 2002 as a tool that the tree physiologist would use to quantify gene expression for plants across a range of environmental conditions and thus learn more about the response of woody perennials to biotic and abiotic stress. Indeed, microarrays have proven useful for investigating high-throughout transcript profiling in response to various environmental and genetic conditions (Wullschleger et al. 2009, Yang et al. 2009). Such studies were enabled by the development of gene annotations contained in the *Populus* genome database. Various representations of cDNA expression arrays were created for *Populus* including the 13K POP1 (Andersson et al. 2004), the 25K POP2 (Moreau et al. 2005), a 6K

P. euphratica array (Brosche et al. 2005) and a 15.5K cDNA microarray developed by Ralph et al. (2006). Although tremendous insights were gained using spotted cDNA microarray technology including the development of a *Populus* gene expression database with cross-comparison analytical tools (Sjodin et al. 2008), whole-genome-based oligonucleotide microarrays were quickly developed once the genome sequence became available. The first oligoarray was developed by scientists at Oak Ridge National Laboratory in collaboration with NimbleGen (Madison, WI, USA) targeting 59,216 predicted genes. This was followed by an Affymetrix (Santa Clara, CA, USA) microarray targeting 40,444 genes and a four-plex format by Agilent (Santa Clara, CA, USA) targeting 43,363 genes (Tsai et al. 2011). These array platforms have facilitated numerous investigations including studies of organ differentiation (e.g., Drost et al. 2010, Rodgers-Melnick et al. 2012), nutritional status (e.g., Qin et al. 2008), carbon partitioning (Payyavula et al. 2011) and abiotic stress (Weston et al. 2011).

More quickly than one would have originally imagined, microarrays gave way to next-generation sequencing, specifically RNASeq, which relates the number of sequence tags of cDNA molecules (counts) to corresponding gene models, thereby providing a measure of transcript abundance (Wang et al. 2009). RNASeq opens the possibility of novel discoveries by allowing quantification of genetic variation from SNPs, allelic differences in expression levels and patterns of alternative splicing (Beaulieu et al. 2011, Geraldes et al. 2011). Furthermore, because it is not tied to a specific annotation, RNASeq permits genome-scale study of virtually any organism, thereby opening genomics approaches to the entire tree physiology community (e.g., Ueno et al. 2010, Hamanishi and Campbell 2011, Rigault et al. 2011). Not being tied to a specific annotation is important since RNASeq can therefore more easily keep pace with the fluid progression of genome annotation for recently sequenced organisms. For example, *Populus* gene call estimates changed from 56K to 41K in annotation v1.1 and v2.2 (http://www.phytozome.net/poplar.phpl), respectively, and a completely new annotation of the *Populus* genome will be released in 2012. This situation will be exacerbated for non-model organisms that lack high-quality reference genomes, which facilitate the assembly of RNASeq reads into transcriptional units (Yang et al. 2011*c*). In contrast to microarray analyses which are based on a fixed platform, RNASeq data can be readily reanalyzed as improved reference annotations become available, thereby providing a more flexible platform that is appropriate for rapidly changing genomics information. RNASeq also permits the identification and characterization of genome 'dark matter', regions of the genome lacking coding sequence (Nagano and Fraser 2011). Many of these regions are associated with regulatory networks and gene function.

Numerous RNASeq analyses have been conducted for forest trees in the past several years. In one of the early efforts, Geraldes et al. (2011) conducted an RNASeq investigation on the developing xylem transcriptome for 20 accessions collected throughout the natural range of *P. trichocarpa*. They identified nearly 0.5 million SNPs in 26,595 expressed genes within developing secondary xylem, and detected numerous cases of allele-specific expression and alternative splicing. RNASeq can also reveal patterns that are not apparent from small-scale expressed sequence tag (EST) studies or from annotation of primary genomic sequence. For example, using an RNASeq approach coupled with proteomic analysis, Yang et al. (2011*c*) reported finding 56 previously unknown small proteins based on RNASeq data in *P. deltoides*. Published RNASeq studies continue to be reported for *Populus* and other trees, and are likely to surpass published microarray studies in the near future.

A major remaining challenge for the tree physiologist is to go beyond the generation of lists of candidate genes and relate transcript abundance and genetic variation to enzyme kinetics and plant function. While this is certain to be a long-term challenge in plant biology, there have been recent computational developments to initiate this process. For example, computational tools to predict the consequences of SNPs on possible amino acid substitutions and protein structure and function have been developed (e.g., PolyPhenI; Adzhubei et al. 2010) and applied to plant genomes (e.g., Gramene; Jaiswal et al. 2006). Using computational approaches, Abraham et al. (2012) were able to identify and align 25% more peptides from a trypsin-digested xylem protein library using SNP information from *P. trichocarpa*. Additionally, network analysis tools have greatly aided our understanding of the plant as a system by linking regulatory interactions and groups of genes into pathways participating in similar biological systems. This approach was recently used to investigate organ differentiation in *Populus* (Drost et al. 2010) and link gene networks to thermal inhibition of photosynthesis (Weston et al. 2011). A continuing challenge will be how, or if, we can link network results and systems models to whole-plant processes of interest to the physiologist (Weston et al. 2012*a*).

Conclusion

Our retrospective analysis confirms that physiological studies with woody perennials have been enabled by the sequencing of the *Populus* genome. In its broadest application the impact of the genome sequence is undeniable, as evidenced by the widespread influence of the main publication describing the genome sequence (Tuskan et al. 2006), which has now been cited over 1000 times in a wide range of journals and embraced by many disciplines. Sequencing the first tree genome has sparked research relevant not only to *Populus* as a resource for the pulp and paper industry (e.g., Nelson and Johnsen 2008), as a keystone species in ecosystems (e.g., Whitham et al.

2008) and as an emerging biofuels feedstock (e.g., Yang et al. 2011*a*), but it has also stimulated molecular-scale research in other woody perennials (e.g., Eckert et al. 2010, Jin et al. 2011, Rampant et al. 2011) and comparative research with herbaceous and woody plants alike (e.g., Weston et al. 2011, Yang et al. 2011*b*). Polle and Douglas (2012) emphasize that significant strides have been taken to understand targeted characteristics of complex tree life histories and that analyses of mineral nutrition and stress responses in *Populus* will continue to benefit from a combination of molecular and ecophysiological approaches. Thus, the forest science community should be pleased with how the *Populus* genome has enabled breakthroughs in areas related not only to whole-tree physiology, but fields as diverse as cell biology and chromosome evolution (Yin et al. 2008) in woody perennials. This added-value contribution comes as a result of advances in comparative genomics where insights gained for one species can inform knowledge of physiological processes in other species.

Despite significant advances in recent years, gaps remain in our understanding of basic biological processes that underlie how plants, populations, communities and ecosystems respond to the environment. For example, the third postulate of Wymore et al. (2011) for demonstrating causal relationships between genes and ecosystem processes requires demonstrating differential effects of the alleles of the responsible gene. However, individual tree genes have not yet been unambiguously implicated as the causative determinants of community or ecosystem traits, and it is probably unrealistic to expect that single loci will be responsible for a substantial fraction of the variation in such complex phenotypes. Techniques to model metabolic networks and an ever-widening array of bioinformatics tools are available for integration of genomics into tree physiological research in an effort to bridge these knowledge gaps. High-throughput tools now exist for collecting data on thousands of genes for organisms living in natural and managed environments, and physiologists, ecologists and evolutionary biologists are beginning to use these tools in creative ways.

Our assessment of the current state of how the *Populus* genome has been used in forest science research also shows where we have fallen short of complete realization of our 2002 predictions (i.e., functional characterization of ecosystem-scale processes). As the *Populus* genome undergoes its third round of improved assembly and annotation, we should remember that a genome is simply a computational representation of nucleotides and that our ability to do fundamental research in the plant sciences requires that we can align sequences, detect SNPs, classify peptides, identify genes and characterize transcriptomes accurately across a wide range of soil and climatic conditions. Each round of improved assembly and annotation gets us closer to having a system-wide understanding of some of the largest and longest-living organisms on Earth. Furthermore, as a reference genome, the *P. trichocarpa* genome sequence represents a small fraction of the genetic variation that exists in this species. Therefore, we anticipate that progressively more physiology-based studies will rely on resequencing to inform biology-based questions. In Wullschleger et al. (2002*b*), we emphasized that embracing new technology and new research paradigms is never easy, and it is clear that many long-term challenges remain for the application of genomics to enhance understanding of tree physiology. However, our community has already made tremendous progress and we are confident that genome-enabled tree physiology will play an increasingly important role in enhancing our understanding of what makes a tree a tree.

Funding

This research was supported by the BioEnergy Science Center, a US Department of Energy Bioenergy Research Facility 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 US Department of Energy under the contract DE-AC05-00OR22725.

References

Abraham P, Adams R, Giannone R, Kalluri U, Ranjan P, Erickson B, Shah M, Tuskan G, Hettich R (2012) Defining the boundaries and characterizing the landscape of functional genome expression in vascular tissues of *Populus* using shotgun proteomics. J Proteome Res 11:449–460.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248–249.

Andersson A, Keskitalo J, Sjodin A, et al. (2004) A transcriptional timetable of autumn senescence. Genome Biol 5:R24.

Aukerman MJ, Sakai H (2003) Regulation of flowering time and floral organ identify by a microRNA and its *APETALA2*-like target genes. Plant Cell 15:2730–2741.

Beaulieu J, Doerksen T, Boyle B, et al. (2011) Association genetics of wood physical traits in the conifer white spruce and relationships with gene expression. Genetics 188:U197–U329.

Bergonze S, Albani MC (2011) Reproductive competence from an annual and a perennial perspective. J Exp Bot 62:4415–4422.

Bohlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O (2006) CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. Science 312:1040–1043.

Breen AL, Glenn E, Yeager A, Olson MS (2009) Nucleotide diversity among natural populations of a North American poplar (*Populus balsamifera*, Salicaceae). New Phytol 182:763–773.

Brosche M, Vinocur B, Alatalo ER, et al. (2005) Gene expression and metabolite profiling of *Populus euphratica* growing in the Negev Desert. Genome Biol 6:R101.

Brown SD, Klingeman DM, Lu TS, Johnson CM, Utturkar SM, Land ML, Schadt CW, Doktycz MJ, Pelletier DA (2012) Draft genome sequence of *Rhizobium* sp. strain PDO1–076, a bacterium isolated from *Populus deltoides*. J Bacteriol 194:2382–2384.

Brunner AM, Nilsson O (2004) Revisiting tree maturation and floral initiation in the poplar functional genomics era. New Phytol 164:43–51.

Chaffey N, Cholewa E, Regan S, Sundberg B (2002) Secondary xylem development in *Arabidopsis*: a model for wood formation. Physiol Plant 114:594–607.

Cheng Z-M, Tuskan GA (2009) *Populus* community mega-genomics: coming of age. Crit Rev Plant Sci 28:282–284.

Drost DR, Benedict CI, Berg A, et al. (2010) Diversification in the genetic architecture of gene expression and transcriptional networks in organ differentiation of *Populus*. Proc Natl Acad Sci USA 107:8492–8497.

Duplessis S, Cuomo CA, Lin Y-C, et al. (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc Natl Acad Sci USA 108:9166–9171.

Eckert AJ, Bower AD, Gonzalez-Martinez SC, Wegrzyn JL, Coop G, Neale DB (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). Mol Ecol 19:3789–3805.

Eckert AJ, Wegrzyn JL, Cumbie WP, Goldfarb B, Huber DA, Tolstikov V, Fiehn O, Neale DB (2012) Association genetics of the loblolly pine (*Pinus taeda*, Pinaceae) metabolome. New Phytol 193:890–902.

Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. Nat Rev Genet 7:85–97.

Geraldes A, Pang J, Thiesen N, Cezardt T, Moore R, Zhao Y, Tam A, Wang S, Friedmann M, Birol I (2011) SNP discovery in black cottonwood (*Populus trichocarpa*) by population transcriptome resequencing. Mol Ecol Res 11:81–92.

Gottel NR, Castro HF, Kerley M, et al. (2011) Distinct microbial communities within the endosphere and rhizosphere of *Populus deltoides* roots across contrasting soil types. Appl Environ Microbiol 77:5934–5944.

Groover AT (2005) What genes make a tree a tree? Trends Plant Sci 10:210–214.

Hall D, Ma X-F, Ingvarsson PK (2011) Adaptive evolution of the *Populus tremula* photoperiod pathway. Mol Ecol 20:1463–1474.

Hamanishi ET, Campbell MM (2011) Genome-wide responses to drought in forest trees. Forestry 84:273–283.

He CY, Zheng SX, Zhang JG, Duan AG, Zeng YF, Cui K (2010) Clonal reproduction and natural variation of *Populus canescens* patches. Tree Physiol 30:1383–1390.

Hsu C-Y, Liu Y, Luthe DS, Yuceer C (2006) Poplar *FT2* shortens the juvenile phase and promotes seasonal flowering. Plant Cell 18:1846–1861.

Hsu C-Y, Adams JP, Kim H, et al. (2011) FLOWERING LOCUS T duplication coordinates reproductive and vegetative growth in perennial poplar. Proc Natl Acad Sci USA 108:10756–10761.

Huijser P, Schmid M (2011) The control of developmental phase transitions in plants. Development 138:4117–4129.

Jaiswal P, Ni J, Yap I, et al. (2006) Gramene: a bird's eye view of cereal genomes. Nucleic Acids Res 34:D717–723.

Jansson S, Douglas CJ (2007) *Populus*: a model system for plant biology. Annu Rev Plant Biol 58:435–458.

Jin H, Do J, Moon D, Noh EW, Kim W, Kwon M (2011) EST analysis of functional genes associated with cell wall biosynthesis and modification in the secondary xylem of the yellow poplar (*Liriodendron tulipifera*) stem during early stage of tension wood formation. Planta 234:959–977.

Keller SR, Levsen N, Ingvarsson PK, Olson MS, Tiffin P (2011) Local selection across a latitudinal gradient shapes nucleotide diversity in balsam poplar, *Populus balsamifera* L. Genetics 188:941–U318.

Labbe J, Jorge V, Kohler A, Vion P, Marcais B, Bastien C, Tuskan GA, Martin F, Le Tacon F (2011) Identification of quantitative trait loci affecting ectomycorrhizal symbiosis in an interspecific F_1 poplar cross and differential expression of genes in ectomycorrhizas of the two parents: *Populus deltoides* and *Populus trichocarpa*. Tree Genet Genom 7:617–627.

Martin E, Tuskan GA, DiFazio SP, Lammers P, Newcombe G, Podila GK (2004) Symbiotic sequencing for the *Populus* mesocosm. New Phytol 161:330–335.

Martin F, Aerts A, Ahren D, et al. (2008) Symbiosis insights from the genome of the mycorrhizal basidiomycete *Laccaria bicolor*. Nature 452:88–93.

Meilan R (1997) Floral induction in woody angiosperms. New For 14:179–202.

Mohamed R, Wang CT, Ma C, et al. (2010) *Populus CEN*/*TFL1* regulates first onset of flowering, axillary meristem identity and dormancy release in *Populus*. Plant J 62:674–688.

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

Nagano T, Fraser P (2011) No-nonsense functions for long noncoding RNAs. Cell 145:178–181.

Neale DB, Kremer A (2011) Forest genomics: growing resources and applications. Nat Rev Genet 12:111–122.

Nelson CD, Johnsen KH (2008) Genomic and physiological approaches to advancing forest tree improvement. Tree Physiol 28:1135–1143.

Payyavula RS, Tay KHC, Tsai CJ, Harding SA (2011) The sucrose transporter family in *Populus*: the importance of a tonoplast PtaSUT4 to biomass and carbon partitioning. Plant J 65:757–770.

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. Annu Rev Ecol Evol Syst 37:187–214.

Polle A, Douglas C (2012) The molecular physiology of poplars: paving the way for knowledge-based biomass production. Plant Biol 12:239–241.

Qin H, Feng T, Harding SA, Tsai CJ, Zhang S (2008) An efficient method to identify differentially expressed genes in microarray experiments. Bioinformatics 24:1583–1589.

Ralph S, Oddy C, Cooper 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 defenses in poplar. Mol Ecol 15:1275–1297.

Rampant PF, Lesur I, Boussardon C, et al. (2011) Analysis of BAC end sequences in oak, a keystone forest tree species, providing insight into the composition of its genome. BMC Genomics 12:292.

Rennenberg H, Wildhagen H, Ehlting BB (2010) Nitrogen nutrition of poplar trees. Plant Biol 12:275–291.

Rigault P, Boyle B, Lepage P, Cooke JEK, Bousquet J, MacKay JJ (2011) A white spruce gene catalog for conifer genome analyses. Plant Physiol 157:14–28.

Rodgers-Melnick E, Mane SP, Dharmawardhana P, Slavov GT, Crasta OR, Strauss SH, Brunner AM, DiFazio SP (2012) Contrasting patterns of evolution following whole genome versus tandem duplication events in *Populus*. Genome Res 22:95–105.

Schatz MC, Langmead B, Salzberg SL (2010) Cloud computing and the DNA data race. Nat Biotechnol 28:691–693.

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

Slavov GT, DiFazio SP, Martin J, et al. (2012) Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. New Phytol (in press). doi:10.1111/j.1469-8137.2012.04258.x.

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

Tisserant E, Kohler A, Dozolme-Seddas P et al. (2012) The transcriptome of the arbuscular mycorrhizal fungus *Glomus intraradices*

(DAOM 197198) reveals functional trade offs in an obligate symbiont. New Phytol 193:755–769.

Tsai CJ, Ranjan P, DiFazio SP, Tuskan GA, Johnson V (2011) Poplar genome microarrays. In: Joshi CS, DiFazio SP, Kole C (eds) Genetics, genomics and breeding of poplar. Science Publishers, Enfield, NH, pp 112–127.

Tuskan GA, DiFazio SP, Teichmann T (2004) Poplar genomics is getting popular: the impact of the poplar genome project on tree research. Plant Biol 6:2–4.

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

Ueno S, Le Provost G, Leger V, et al. (2010) Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. BMC Genomics 11:24.

van der Lelie D, Taghavi S, Monchy S, et al. (2009) Poplar and its bacterial endophytes: coexistence and harmony. Crit Rev Plant Sci 28:346–358.

Wang JW, Park MY, Wang LJ, Koo Y, Chen XY, Weigel D, Poethig RS (2011) miRNA control of vegetative phase change in trees. PLoS Genet 7:e1002012.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63.

Wegrzyn JL, Eckert AJ, Choi M, Lee JM, Stanton BJ, Sykes R, Davis MF, Tsai C-J, Neale DB (2010) Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. New Phytol 188:515–532.

Weston DJ, Karve AA, Gunter LE, Jawdy SA, Yang X, Allen SM, Wullschleger SD (2011) Comparative physiology and transcriptional networks underlying the heat shock response in *Populus trichocarpa*, *Arabidopsis thaliana* and *Glycine max*. Plant Cell Environ 34:1488–1506.

Weston DJ, Hanson PJ, Norby RJ, Tuskan GA, Wullschleger SD (2012*a*) From systems biology to photosynthesis and whole-plant physiology: a conceptual model for integrating multi-scale networks. Plant Signal Behav 7:260–262.

Weston DJ, Pelletier DA, Morrell-Falvey JL, et al. (2012*b*) *Pseudomonas fluorescens* induces strain-dependent and strain-independent host plant responses in defense networks, primary metabolism, photosynthesis and fitness. Mol Plant Microbe Interact 25:765–778.

Whitham TG, DiFazio SP, Schweitzer JA, Shuster SM, Allan GJ, Bailey JK, Woolbright SA (2008) Extending genomics to natural communities and ecosystems. Science 320:492–495.

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

Wullschleger SD, Tuskan GA, DiFazio SP (2002*b*) Genomics and the tree physiologist. Tree Physiol 22:1273–1276.

Wullschleger SD, Weston DJ, Davis JM (2009) *Populus* responses to edaphic and climatic cues: emerging evidence from systems biology research. Crit Rev Plant Sci 28:368–374.

Wymore AS, Keeley ATH, Yturralde KM, Schroer ML, Popper CR, Whitham TG (2011) Genes to ecosystems: exploring the frontiers of ecology with one of the smallest biological units. New Phytol 191:19–36.

Yang X, Kalluri UC, DiFazio SP, Wullschleger SD, Tschaplinski TJ, Cheng ZM, Tuskan GA (2009) Poplar genomics: state of the science. Crit Rev Plant Sci 28:285–308.

Yang X, Ye C-Y, Bisaria A, Tuskan GA, Kalluri UC (2011*a*) Identification of candidate genes in *Populus* cell wall biosynthesis using text-mining, co-expression network and comparative genomics. Plant Sci 181:675–687.

Yang X, Kalluri UC, Jawdy S, Gunter LE, Yin TM, Tschaplinski TJ, Weston DJ, Ranjan P, Tuskan GA (2011*b*) The F-box gene family is expanded in herbaceous annual plants relative to woody perennial plants. Plant Physiol 148:1189–1200.

Yang X, Tschaplinski TJ, Hurst GB, et al. (2011*c*) Discovery and annotation of small proteins using genomics, proteomics, and computational approaches. Genome Res 21:634–641.

Yin T, DiFazio SP, Gunter LE, et al. (2008) Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. Genome Res 18:422–430.