Physiologia Plantarum
An International Journal for Plant Biology

MINIREVIEW

Modeling transcriptional networks regulating secondary growth and wood formation in forest trees

Lijun Liu[a], Vladimir Filkov[b] and Andrew Groover[a,c,*]

[a]US Forest Service, Pacific Southwest Research Station, Davis, CA, USA
[b]Department of Computer Science, University of California, Davis, CA, USA
[c]Department of Plant Biology, University of California, Davis, CA, USA

Correspondence
*Corresponding author,
e-mail: agroover@fs.fed.us

Received 25 July 2013;
revised 23 August 2013

doi:10.1111/ppl.12113

The complex interactions among the genes that underlie a biological process can be modeled and presented as a transcriptional network, in which genes (nodes) and their interactions (edges) are shown in a graphical form similar to a wiring diagram. A large number of genes have been identified that are expressed during the radial woody growth of tree stems (secondary growth), but a comprehensive understanding of how these genes interact to influence woody growth is currently lacking. Modeling transcriptional networks has recently been made tractable by next-generation sequencing-based technologies that can comprehensively catalog gene expression and transcription factor-binding genome-wide, but has not yet been extensively applied to undomesticated tree species or woody growth. Here we discuss basic features of transcriptional networks, approaches for modeling biological networks, and examples of biological network models developed for forest trees to date. We discuss how transcriptional network research is being developed in the model forest tree genus, *Populus*, and how this research area can be further developed and applied. Transcriptional network models for forest tree secondary growth and wood formation could ultimately provide new predictive models to accelerate hypothesis-driven research and develop new breeding applications.

Introduction

Reductionist developmental genetics approaches traditionally focus on one gene, and through experimentation attempt to determine function through the change of phenotype after modulating expression of the gene (e.g. though the study of gene knockouts). While this has been successful in a variety of model plant and animal species, the reductionist approach has proven difficult to extend to long-lived, undomesticated perennials such as trees. There are two main reasons for this. First, long generation times, barriers to inbreeding and other practical considerations make it difficult to make homozygous mutations in trees. Second, many of the traits of interest such as wood formation are quantitative traits that are controlled by many genes, necessitating technologies that can comprehensively describe the contributions and interactions of large numbers of genes in conditioning the phenotype.

New, 'next-generation' DNA sequencing technologies are driving major advances in developmental genetics and genomics research of woody growth in trees, including availability of full genome sequence for a growing number of tree species (Tuskan et al. 2006, Ming et al. 2008, Nystedt et al. 2013). Additionally,

Abbreviations – ChIP-seq, chromatin immunoprecipitation sequencing; mRNA-seq, mRNA-sequencing; SND, Secondary Wall-associated NAC Domain; Y1H, yeast one hybrid.

'next-generation' sequencing provides the ability to develop extensive genome-wide datasets describing changes of gene expression across tissues or in experimental treatments, the binding of transcriptional regulators to target gene promoters, protein–protein interactions among transcriptional regulators and patterns of chromatin structure and epigenetic marks. Together, these data types can be integrated to provide models of the transcriptional networks that control gene expression. Transcriptional network models could be used to optimize research and predict phenotypic outcomes resulting from altering gene expression, as is being demonstrated in medical and pharmaceutical research (Schadt and Björkegren 2012). Transcriptional network models also represent an important framework on which data and results from other fields of study (e.g. developmental genetics, population genomics and evolutionary genomics) can be integrated and visualized.

Secondary growth is the process by which woody stems increase in girth. This growth is ultimately supported by the vascular cambium (Larson 1994), whose cells divide to provide daughter cells both to the inside of the stem that differentiate into secondary xylem (wood) and to the outside of the stem that differentiate into secondary phloem (inner bark). Transcript profiling in the model tree genus, *Populus*, has characterized gene expression across these tissues (Schrader et al. 2004), and a modest number of studies have provided functional characterization of transcription factors regulating specific aspects of secondary growth (reviewed in Groover et al. 2010, Matte Risopatron et al. 2010). However, we currently lack comprehensive models describing how the large number of genes involved in secondary growth act together to affect growth and development. Better understanding of secondary growth is important not only as a fundamental part of plant biology, but also because secondary growth is central to developing solutions for problems ranging from forest conservation, meeting demands for forest products, producing feedstocks for cellulosic biofuels and mitigating impacts of climate change on forests. Importantly, transcriptional network models are predictive, and could improve the efficiency of both research and application development.

Transcriptional regulatory networks and their features

The precise transcriptional interactions among the large number of genes in an organism are conveniently summarized in a network representation. A network (formally, graph) has nodes connected by links, or edges. In a transcriptional regulatory network the nodes are the entities involved in transcription, i.e. DNA (genes), proteins and RNA molecules (e.g. microRNAs). The links describe the physical interaction among the nodes, e.g. protein–DNA interactions. Most immediately, having a comprehensive network of such interactions allows efficient lookup of genes and proteins of interest, and sometimes also the functions to which they are contributing in the cell. In that sense, a visual interface to such a network can be an invaluable reference to the life scientist, especially if it is richly annotated and hyperlinks to other resources which are embedded in it. Currently, PopGenie (Sjödin et al. 2009) provides co-expression network visualization tools, while Phytozome (Goodstein et al. 2012) is the primary repository for *Populus* genome and annotations.

On a more fundamental level, the network wiring architecture, or its topology, can reveal the systemic properties of the organism, and provide models of complex organizational features (Babu et al. 2004). For example, linear paths connecting multiple genes can model cascades of genes turning on or off other genes in a sequence (Zhang and Klessig 2001); feed-forward loops can model precise temporal regulation (Mangan and Alon 2003); and alternating inhibition and activation in a fan-out fashion can model precise spatial regulation (Oliveri et al. 2008). From a more global, or top-down perspective, the in- and out-degree of a node, i.e. the number of its regulators and number of others it regulates, when aggregated over all nodes provides a measure of the efficiency of signal exchange in the network (Proulx et al. 2005, Costanzo et al. 2010), and contributes to gene expression robustness (MacNeil and Walhout 2011).

Approaches for modeling transcriptional networks

Transcriptional network modeling is being facilitated by new genomic and sequencing technologies. Importantly, massively parallel 'next-generation' sequencing now allows the comprehensive and quantitative measure of mRNA transcripts in any species or tissues for which mRNA can be extracted in reasonable quantity and quality. This approach of mRNA-sequencing (or mRNA-seq) (Wang et al. 2009) can be used to establish datasets describing transcript levels in secondary vascular tissues of trees that have been subjected to experimental treatments, expressing transgenes or undergoing time-course response to environmental changes or experimental treatments. Next-generation sequencing can also be used to identify the binding sites of individual transcription factors genome-wide, using the approach of chromatin immunoprecipitation

sequencing (ChIP-seq). ChIP-seq can also be used to map genome-wide location of DNA methylation and histone variants that influence gene expression (Consortium 2012). Sequencing-based data can also be integrated with other data types, such as yeast one hybrid (Y1H) data identifying transcription factors that bind to a promoter of interest. In the Y1H approach, a library is made of genes encoding transcription factors (prey) of interest, which is screened for prey that can bind a promoter (bait) sequence of interest that in turn drives expression of selectable and/or visible marker genes in yeast.

Integrating experimental and computational approaches

To achieve good qualitative and ideally even quantitative models of regulatory networks, computational approaches should work hand in hand with the genomics technologies. One approach, which we adopted in an ongoing project focused on *Populus* secondary growth transcriptional networks, is to interleave the two: the biological experiments are informed by current state-of-the-art knowledge and analytics from the computational approaches; likewise, the computational approaches are deeply embedded at all steps of the project, and are both informed by and guiding the experimental work. We describe our existing and planned approach next, and summarize it in Fig. 1.

RNA-seq and ChIP-seq read mapping

mRNA and ChIP-seq technologies yield millions of short reads which are mapped onto a reference genome. The precise pipelines used vary among research labs, but most employ some sort of: (1) pre-processing to trim adapter sequences, and filter the sequencing reads for quality, followed by (2) alignment of sequencing reads to map them to a reference sequence (in our case, *Populus trichocarpa* v3.0 from http://www.phytozome.net/poplar.php) using ultrafast aligners [e.g. BowTie, based on a Burrows–Wheeler transform (Langmead et al. 2009), and TopHat (Trapnell et al. 2009)]. The last step before analyzing data for biological discovery is (3) normalization to make the distributions of the sequencing read coverage data similar and thus make data comparable within and across sequencing lanes and experiments. A common normalization method is quantile normalization (quantiles are points at regular intervals in the cumulative distribution, dividing the data into equal sized subsets; Bullard et al. 2010).

Discovery: gene sets, function and network links

The sequencing reads from mRNA-seq then typically go through a differential expression discovery. In this step, methods are employed that compensate for the variability among biological replicates and provide higher sensitivity at lower coverage levels. EDGER (Robinson et al. 2010) and DESEQ (Anders and Huber 2010) are popular software packages for differential expression analysis.

Having data from multiple experiments, time points or treatment levels enable gene sets analysis. Genes with indistinguishable expression profiles are combined into gene sets, or groups, using short-time series analysis methods and clustering algorithms (the statistical package R includes clustering algorithms, e.g. k-means or agnes). For *Populus*, functional enrichment in gene sets can be evaluated using the Gene Ontology (Ashburner et al. 2000) and Phytozome (http://www.phytozome.net/poplar.php) functional annotations. A variety of tools exist for this analysis: SerbGO searches among many enrichment tools to find the best one for a given analysis (Mosquera and Sanchez-Pla 2008). In addition, shared upstream motifs, if they exist, can be identified for the gene groups using multiple alignment and motif searches, e.g. MEME (Bailey et al. 2006).

As precursor to network modeling, transcription factor DNA-binding events are identified from ChIP-seq reads aligned to a reference genome. To do this, typically, aligned ChIP-seq reads are evaluated for compact peaks of sequence reads pileup, representing potential loci of transcription factor binding (Pepke et al. 2009, Blahnik et al. 2010). As this is an area of active research, there are a variety of methods for peak detection from ChIP-seq data, many are able to estimate and correct for background noise using a control (e.g. un-ChIP'd input DNA) vs experimental sequences, e.g. MACS (Pepke et al. 2009), which can estimate binding site locations with high sensitivity.

Transcriptional gene modules (subsets of co-expressed genes which have a common regulator) can be identified by statistically integrating the ChIP-seq and RNA-seq data, using, e.g. the GRAM algorithm (Bar-Joseph et al. 2003) and models based on regulation hierarchies (Filkov and Shah 2008). The expression of genes in such modules is more highly correlated with the binding transcription factors' activity than those which do not show differential expression (Bar-Joseph et al. 2003, Moreno-Risueno et al. 2010).

To model a more complete regulatory network, in addition to identifying downstream links, upstream links

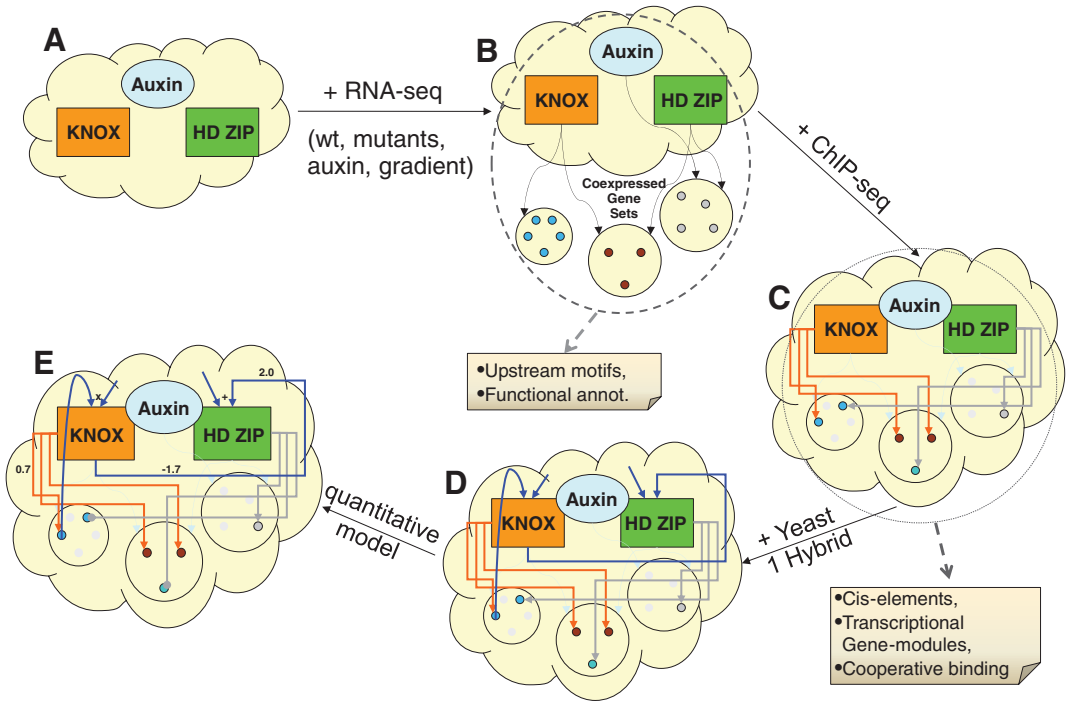

Fig. 1. Summary of our integrated experimental and computational approaches. Starting at the left top cloud (A), the initial experimental design acknowledges close relationships among auxin, class I KNOX and class III HD ZIP transcription factors in regulating secondary growth, suggesting they may act within short path lengths in a secondary growth transcriptional network. RNA sequencing is used to profile transcript levels in mutants of each transcription factor, as well as in developmental gradients within stems and time-course treatments of stem by auxin. These transcript data are used to identify groups of co-expressed genes (B), which can be mined for shared upstream *cis*-elements and evaluated for enrichment of genes of common functional categories (e.g. Gene Ontologies). Next, ChIP-seq is used to determine direct targets of each transcription factor, which can be overlayed on the co-expression network (C). Y1H data for the target transcription factors is then integrated, identifying factors that regulate these key regulators (D). Last, the integrated data types are used in computational analyses that describe motifs (e.g. feed-forward loops), positive vs negative regulation, signal aggregation (e.g. sum or product, shown as + or ×), and provide quantitative measures (shown as continuous numbers) of interactions among nodes in the final network model (E). By adding data from different technologies, a better-resolved outcome is realized at consecutive steps.

can be elucidated using Y1H technology. Using Y1H, transcription factors can be implicated that directly regulate the expression of genes encoding transcription factors of interest (Gaudinier et al. 2011).

Quantitative modeling

A precise quantitative transcriptional regulatory network model overlays the links and nodes structure with temporal and spatial parameters quantifying the context and size of the effect one node has on another's gene expression. With such a quantitative network, modeling the emergence of phenotypes may be possible.

One way to build a predictive model of gene expression is in terms of the binding events modeled as either overrepresented motifs or ChIP-seq binding events in gene's *cis*-regions (Bussemaker et al. 2001, Filkov and Shah 2008). These models can be extended to get more accurate results by accounting for transcriptional modules (e.g. modeling all genes in the same module as a single gene) and cooperative regulation (by using same regression coefficients for cooperative factors). These two approaches can increase the specificity of the predictions by decreasing the number of free parameters in the model.

Advantages and challenges of secondary growth

Secondary growth presents both advantages and challenges for transcriptional network approaches. A major advantage is the radial organization of the stem. The cells of the cambial zone have thin cell walls, and during active periods of growth the bark can be easily removed from the stem, separating away at the fragile cambial zone [in *Populus*, separation is in the early differentiating xylem (Du et al. 2006, Zhang et al. 2011)]. Horticulturalists recognize this when remarking that the

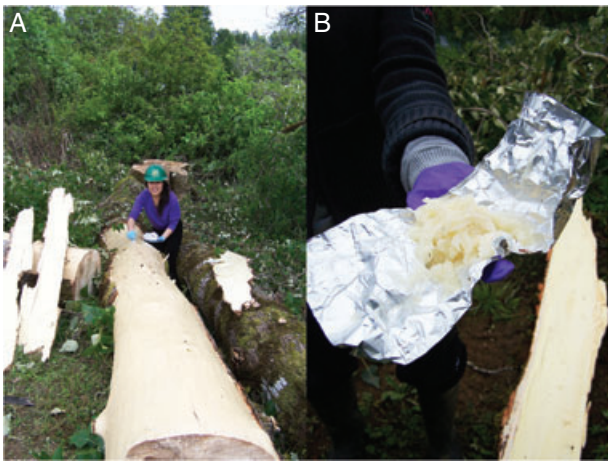

Fig. 2. Gram quantities of cambium and recent derivatives can be harvested for chromatin immunoprecipitation and biochemical assays. (A) The radial organization of the tree stem facilitates harvest of the cambium meristem. During active growth, bark can be peeled from the stem, with the bark separating at the fragile cambial zone or early differentiating xylem. (B) Simple scraping of the surface of the bark and wood allows harvest of copious cambium and recent cell derivatives.

bark of a tree is 'slipping' and the same phenomenon allows for easy harvest of large quantity of the cambium and recently derived cells (Fig. 2) sufficient for ChIP and other biochemical approaches. This stands in sharp contrast with the difficulty in harvesting significant amounts of apical or root meristem tissues. Radial organization also allows for harvest of cells at different stages of differentiation, and the historical record of differentiation is laid down in the form of lignified cell files in the secondary xylem (wood). Wood development has been well studied at the anatomical, biochemical and genetic levels (reviewed for *Populus* in Groover et al. 2010). Additionally, significant standing genetic variation exists for most tree species for wood development traits, providing a 'natural' resource for understanding the development and evolution of secondary growth.

There are also significant challenges to secondary growth genomics research. Making transgenic plants that alter the expression of candidate genes is a relatively cumbersome task, and typically only some species and/or genotypes are amenable for transformation. For example, in *Populus*, the current reference sequence is from a species (*P. trichocarpa*) and genotype (*Nisqually*) that is difficult to transform. The preferred laboratory genotypes for transformation are from hybrids of other species, thus requiring mapping of sequence reads from the laboratory species/genotype to the reference species/genotype. While typically not problematic for translated regions, cross species mapping can pose significant limitations for ChIP and other methods that probe non-genic regions. Additionally, there are typically not large community resources like knockout lines or reagent repositories as might be found for models such as Arabidopsis, which limits genomics research for trees.

Examples of network approaches in trees

Network biology is beginning to be extended to forest trees, however, and a few examples of the approaches taken are described below for the model genus, *Populus*. For the most part, the approaches described here were first developed for human and model plant/animal research. Forest tree researchers are thus leveraging the large investments in genomic technology for medical and model organism research.

Wood formation includes the differentiation of cell types such as fibers that synthesize a thick, lignified secondary cell wall. Cell types with lignified secondary cell walls also differentiate in the model plant, Arabidopsis, and are homologous (have shared evolutionary origin) with the same cell types in *Populus*. This raises the question of whether knowledge of transcriptional networks regulating secondary cell wall biosynthesis in Arabidopsis could be extended to *Populus*, and how evolution has acted to rewire the homologous networks. One hundred twenty-one Arabidopsis genes implicated in secondary cell wall biosynthesis were identified using text mining of published literature (Yang et al. 2011). These genes were then used to query a co-expression database to identify additional candidate genes with correlated expression. In this way a total of 694 *Arabidopsis* genes were implicated in cell wall biosynthesis or remodeling, for which there were 817 corresponding *Populus* orthologs. Microarray data from both species revealed that the tissue-specific pattern of expression was highly correlated across species, and evidence of conservation was found for co-expression of at least some gene clusters between the two species (Yang et al. 2011). Another example of conserved transcriptional networks is provided by Secondary Wall-associated NAC Domain (SND) transcription factors. SNDs have been well characterized as master regulators of secondary cell wall biosynthesis in Arabidopsis, where they activate expression of other transcription factors (notably MYBs) that in turn regulate suites of cell wall-related genes (Yamaguchi and Demura 2010). A group of *Populus* SNDs have been shown to act as functional orthologs of Arabidopsis SND1 and were able to activate the entire secondary cell wall biosynthesis program (Zhong et al. 2010, Zhong et al. 2011), again providing evidence for conserved function of both individual transcription factors and network modules across these species.

A systems approach was used to provide insight into the evolution of transcriptional regulation of the lignin biosynthesis pathway (Shi et al. 2010). Of 95 *Populus* genes encoding enzymes for monolignol biosynthesis, 18 were expressed specifically in xylem tissues. *Cis*-motif analysis found five core *cis*-motifs that were responsible for xylem expression of all 18 xylem-specific genes, and was consistent with the notion that these core motifs were both redundant and quantitative in their contribution to tissue specific expression. This study provides a model pathway for understanding the role of *cis*-elements in the evolution of gene expression patterns in wood formation.

Perhaps the most comprehensive transcriptional network models to date were developed using large numbers of microarray datasets from *Populus* leaves subjected to different treatments and at different stages of development (Street et al. 2011). Starting from 562 leaf-specific *Populus* genes with quantified transcription profiles across 456 samples in various experiments, transcriptional modules containing co-expressed genes with shared *cis*-motifs were first identified, and then the most likely transcription factors regulating each module were inferred based on gene expression predictability. A regulatory network was then inferred using regression models. The identified modules contained genes that were significantly co-expressed during developmental processes, and 71% of the modules had biological relevance based on GO overrepresentation of member genes. Impressively, the regulatory network could significantly predict gene expression in new experiments and identified new regulators of leaf development. Additionally, comparison with Arabidopsis identified conserved regulatory mechanisms, and provides another example of how comparative network approached could be useful for future evolution and development studies.

Conclusions and perspectives

While just beginning, there are good reasons to be optimistic for future network-based research and applications for woody growth in trees. The increasing power and decreasing cost of DNA sequencing supports the generation of datasets suitable for network reconstruction, even in woody species lacking full genome sequence. The development of computational approaches and bioinformatics tools in model species and medical research can be adapted for use in trees, allowing even smaller research groups to make use of these technologies. Importantly, the proof of concept that complex biological systems can be modeled using network approaches, and that the predictions of network models can drive research and application development is being provided in medical research. In forest trees, most traits of interest are quantitative in nature and are thus amenable to modeling using network approaches. The next steps for modeling transcriptional networks in trees include amassing the critical datasets required, developing or modifying tools to analyze the data and produce network models, and experimentally testing and refining the models in a reiterative fashion.

Network approaches also integrate with other genomic approaches and data types. For example, data from ongoing genome-wide association studies in tree species could potentially be better leveraged by incorporating the knowledge of gene networks to reduce false discovery and increase power for detecting alleles influencing a phenotype (Schadt and Björkegren 2012). Additionally, transcriptional network graphs provide an intuitive visualization of genomic data, as an alternative to gene and chromosome-based methods (e.g. viewing quantitative trait loci using genome browsers). Ultimately, once established, network approaches and models could provide numerous benefits by better guiding basic research, and providing predictive tools for breeding, conservation and forest management.

Acknowledgements – This work received funds from US Forest Service, USDA AFRI and DOE IAA Award No. DE-SC0005581.

References

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25–29

Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA (2004) Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol 14: 283–291

Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34: W369–W373

Bar-Joseph Z, Gerber GK, Lee TI, Rinaldi NJ, Yoo JY, Robert F, Gordon DB, Fraenkel E, Jaakkola TS, Young RA, Gifford DK (2003) Computational discovery of gene modules and regulatory networks. Nat Biotech 21: 1337–1342

Blahnik KR, Dou L, O'Geen H, McPhillips T, Xu X, Cao AR, Iyengar S, Nicolet CM, Ludascher B, Korf I, Farnham PJ (2010) Sole-Search: an integrated analysis program for

peak detection and functional annotation using ChIP-seq data. Nucleic Acids Res 38: e13
Bullard J, Purdom E, Hansen K, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11: 94
Bussemaker HJ, Li H, Siggia ED (2001) Regulatory element detection using correlation with expression. Nat Genet 27: 167–171
ConsortiumTEP (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57–74
Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JLY, Toufighi K, Mostafavi S, Prinz J, St. Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin Z-Y, Liang W, Marback M, Paw J, San Luis B-J, Shuteriqi E, Tong AHY, van Dyk N, Wallace IM, Whitney JA, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras A-C, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C (2010) The genetic landscape of a cell. Science 327: 425–431
Du J, Xie HL, Zhang DQ, He XQ, Wang MJ, Li YZ, Cui KM, Lu MZ (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
Filkov V, Shah N (2008) A simple model of the modular structure of transcriptional regulation in yeast. J Comput Biol 15: 393–405
Gaudinier A, Zhang L, Reece-Hoyes JS, Taylor-Teeples M, Pu L, Liu Z, Breton G, Pruneda-Paz JL, Kim D, Kay SA, Walhout AJM, Ware D, Brady SM (2011) Enhanced Y1H assays for Arabidopsis. Nat Meth 8: 1053–1055
Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40: D1178–D1186
Groover A, Nieminen K, Helariutta Y, Mansfield SD (2010) Wood formation in *Populus*. In: Jansson S, Bhalerao R, Groover A (eds) Genetics and Genomics of *Populus*. Springer, New York, London
Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25
Larson PR (1994) The Vascular Cambium, Springer-Verlag, Berlin, Heidelberg
MacNeil LT, Walhout AJM (2011) Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome Res 21: 645–657
Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci USA 100: 11980–11985
Matte Risopatron JP, Sun Y, Jones BJ (2010) The vascular cambium: molecular control of cellular structure. Protoplasma 247: 145–161
Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KLT, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall K, Rice DW, Albert H, Wang M-L, Zhu YJ, Schatz M, Nagarajan N, Acob RA, Guan P, Blas A, Wai CM, Ackerman CM, Ren Y, Liu C, Wang J, Wang J, Na J-K, Shakirov EV, Haas B, Thimmapuram J, Nelson D, Wang X, Bowers JE, Gschwend AR, Delcher AL, Singh R, Suzuki JY, Tripathi S, Neupane K, Wei H, Irikura B, Paidi M, Jiang N, Zhang W, Presting G, Windsor A, Navajas-Perez R, Torres MJ, Feltus FA, Porter B, Li Y, Burroughs AM, Luo M-C, Liu L, Christopher DA, Mount SM, Moore PH, Sugimura T, Jiang J, Schuler MA, Friedman V, Mitchell-Olds T, Shippen DE, dePamphilis CW, Palmer JD, Freeling M, Paterson AH, Gonsalves D, Wang L, Alam M (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature 452: 991–996
Moreno-Risueno MA, Busch W, Benfey PN (2010) Omics meet networks – using systems approaches to infer regulatory networks in plants. Curr Opin Plant Biol 13: 126–131
Mosquera JL, Sanchez-Pla A (2008) SerbGO: searching for the best GO tool. Nucleic Acids Res 36: W368–W371
Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hallman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Kaller M, Luthman J, Lysholm F, Niittyla T, Olson A, Rilakovic N, Ritland C, Rossello JA, Sena J, Svensson T, Talavera-Lopez C, Theiszen G, Tuominen H, Vanneste K, Wu Z-Q, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Lee Thompson S, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497: 579–584
Oliveri P, Tu Q, Davidson EH (2008) Global regulatory logic for specification of an embryonic cell lineage. Proc Natl Acad Sci USA 105: 5955–5962
Pepke S, Wold B, Mortazavi A (2009) Computation for ChIP-seq and RNA-seq studies. Nat Meth 6: S22–S32
Proulx SR, Promislow DEL, Phillips PC (2005) Network thinking in ecology and evolution. Trends Ecol Evol 20: 345–353

Robinson M, McCarthy DJ, Smyth G (2010) edgeR a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140

Schadt EE, Björkegren JLM (2012) NEW: Network-Enabled Wisdom in Biology, Medicine, and Health Care. Sci Transl Med 4: 115rv111

Schrader J, Nilsson J, Mellerowicz E, Berglund A, Nilsson P, Hertzberg M, 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

Shi R, Sun YH, Li Q, Heber S, Sederoff R, Chiang VL (2010) Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. Plant Cell Physiol 51: 144–163

Sjödin A, Sandberg G, Gustafsson P, Jansson S (2009) The Populus Genome Integrative Explorer (PopGenIE) a new resource for exploring the *Populus* genome. New Phytol 182: 1013–1025

Street N, Jansson S, Hvidsten T (2011) A systems biology model of the regulatory network in *Populus* leaves reveals interacting regulators and conserved regulation. BMC Plant Biol 11: 13

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105–1111

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G-L, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple J-C, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai C-J, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313: 1596–1604

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

Yamaguchi M, Demura T (2010) Transcriptional regulation of secondary wall formation controlled by NAC domain proteins. Plant Biotechnol 27: 237–242

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

Zhang S, Klessig DF (2001) MAPK cascades in plant defense signaling. Trends Plant Sci 6: 520–527

Zhang J, Gao G, Chen J-J, Taylor G, Cui K-M, He X-Q (2011) Molecular features of secondary vascular tissue regeneration after bark girdling in *Populus*. New Phytol 192: 869–884

Zhong R, Lee C, Ye ZH (2010) Functional characterization of poplar wood-associated NAC domain transcription factors. Plant Physiol 152: 1044–1055

Zhong R, McCarthy RL, Lee C, Ye ZH (2011) Dissection of the transcriptional program regulating secondary wall biosynthesis during wood formation in poplar. Plant Physiol 157: 1452–1468

Edited by Y. Helariutta