



Meetings

Genomic science provides new insights into the biology of forest trees

35th New Phytologist Symposium: The genomes of forest trees: new frontiers of forest biology, Arnold Arboretum of Harvard University, Boston, MA, USA, June 2015

Forest biology is undergoing a fundamental change fostered by the application of genomic science to longstanding questions surrounding the evolution, adaptive traits, development, and environmental interactions of tree species. Genomic science has made major technical leaps in recent years, most notably with the advent of 'next generation sequencing' but translating these technical advances into new biological insights can be challenging, especially in long-lived perennial forest trees. In a meeting of 80 delegates from around the world (Fig. 1), the 35th New Phytologist Symposium co-organized by Andrew Groover (US Forest Service and University of California Davis, USA) and William Friedman (Arnold Arboretum of Harvard University, Boston, MA, USA) showcased examples of both technical feats such as the sequencing of massive conifer genomes as well as important new insights into the biology of trees. Excitingly, genomics-based research is being extended beyond the handful of traditional model forest tree species, to encompass the amazing diversity of species we call trees. Additionally, delegates participated in discussion groups, which were challenged with questions concerning the scientific advancement of forest biology research as well as questions surrounding practical issues ranging from funding to career advancement for young researchers in this field. Here, highlights are presented both from the scientific presentations, as well as opinions shared by delegates regarding the needs and opportunities facing the forest biology community.

Technical advances in sequencing and assembly support new forest tree genomic resources

The recent advances in DNA sequencing technology are nothing short of astonishing. The increasing capacity and precision, and decreasing cost of sequencing (Wetterstrand, 2015) has made feasible whole genome sequencing projects in several forest tree species. Updates were provided on conifer genome sequencing by David Neale (University of California Davis) and Nathalie Isabel

(Natural Resources Canada, Quebec, Canada). Conifer genomes are huge by any standard – up to nearly 40 gigabases. This has presented new technical challenges for sequencing and assembly, and the current conifer genome assemblies are still relatively fragmented. However, these initial genome assemblies are providing new biological insights. Conifers present several stark contrasts to angiosperms, with conifers represented by a comparatively modest number of species, rare occurrence of ploidy changes, and high levels of heterozygosity within species and individuals. The available draft conifer genomes provide insights into these fundamental features of conifers: there is no evidence for recent whole genome duplications in conifers, there are large physical distances separating genes, introns can be very large, and there is extremely low linkage disequilibrium. Additionally, conifer genomes are characterized by the long-term accumulation of transposable elements that may indicate mechanistic differences in how gymnosperms and angiosperms silence and eliminate transposons and repeat sequences. Challenges to the research community include determining the mechanisms resulting in these basic features of conifer genomes, and ultimately how they account for the relatively ancient and conserved gene content of conifers and low rates of speciation in comparison to angiosperms.

Sometimes exceptions to the rules provide insight, and Alison Dawn Scott (University of Wisconsin, Madison, WI, USA) presented an interesting case of using genomic data to understand the evolutionary history of an unusual conifer, the polyploid coast redwood (*Sequoia sempervirens*). Using a comparative approach, she sequenced transcriptomes from *Sequoia* and related *Sequoiadendron* and *Metasequoia*. What emerges is a complicated story; with support for an autopolyploid origin for coast redwood some 24 million years ago (Ma).

Conifers are not the only species with complicated genomes, and Nathaniel Street (Umeå Plant Science Centre, Sweden) described an unexpectedly complex group of aspen genomes. In the same genus as *Populus trichocarpa* (with arguably the most complete and finished genome sequence to date for a tree), *P. tremula* and *P. tremuloides* show remarkable divergence and heterozygosity in nongenic regions. This has presented major technical hurdles in assembling these genomes, but also raises fascinating questions about the origins and the causal mechanisms and evolutionary consequences of this rapid divergence and heterozygosity.

New genomic approaches and new species

One change in direction for the field of forest genetics and genomics is a move away from focus on a limited number of 'model' tree species to a broader focus driven by both technological advances in sequencing, recognition of the power of comparative approaches, and pursuit of unique biological questions. Catherine Kidner (Royal Botanic Garden, Edinburgh, UK) introduced a fascinating



Fig. 1 Delegates from the 35th New Phytologist Symposium enjoying the sunshine at Arnold Arboretum in Boston (MA, USA).

project addressing the astounding diversity of tree species in tropical rain forests. *Inga* is a genus of *c.* 300 neotropical legume species which radiated over the past 6 million years. A major challenge is to produce robust phylogenies that resolve relationships among species with relatively low sequence divergence for commonly used molecular markers. Kidner and colleagues developed a hybrid capture protocol that allowed efficient sequencing of genes in a wide variety of *Inga* species and samples, including samples from herbarium specimens. This protocol enables a variety of approaches including comparisons of genetic diversity over time through comparison of current samples to historical herbarium samples from similar locations and species. The results from the project to date suggest that *Inga* diversity is driven by herbivore pressure, leading to divergence in the production of defense-related secondary metabolites.

Richard Buggs (Queen Mary University, London, UK) and colleagues are undertaking an ambitious project aimed at understanding variation within and among species for resistance to emerald ash borer and ash die back. *Fraxinus* (ash) genotypes and species with varying degrees of resistance to the pathogen and insect pest are being subjected to a comparative genomics approach, underpinned by a *de novo* assembled reference genome for *F. excelsior* (European ash). This project will serve as a valuable test case for the application of comparative genomics approaches to the study of invasive diseases and insect pests, which are a common and major threat to forest tree species.

Birch species (genus *Betula*) are being developed as a powerful new model system, as described by Kaisa Nieminen (Natural Resources Institute, Helsinki, Finland). In contrast to the long generation times of most trees, flowering can be experimentally induced in at least some birch. And because birch is monoecious, approaches requiring inbreeding (e.g. making homozygous mutations) are feasible. The significant variation among and within birch species was highlighted, and when paired with ongoing efforts to complete a reference genome for silver birch (*B. pendula*) will provide a comprehensive new model system for comparative and functional genomics.

Recent buzz surrounding the importance of microbiomes on human health are nothing new to tree biologists, who have long recognized the value of microbial symbionts in tree growth and forest health. Jonathan Plett (University of Western Sydney, Australia) described how whole genome sequencing of the *Populus*

ectomycorrhizal symbiont, *Laccaria bicolor*, is enabling a new era of research regarding how the plant and fungal organisms recognize and interact with each other. Interestingly, similar to disease-causing pathogens, it appears that the beneficial fungus secretes small effector-like proteins that help avoid triggering an immune detection by the plant. Ongoing research is providing additional molecular-level details as to the intricacies of signaling between the fungus and plant.

Evolutionary perspectives

Dobzhansky famously declared that ‘Nothing in biology makes sense except in the light of evolution’ (Dobzhansky, 1973). Arguably, evolutionary perspectives have been restrained in forest genetics by the availability of genomic data for only a limited number of taxa. Excitingly, evolutionary and comparative genomic approaches are now feasible for tree species by the application of next generation sequencing technologies, but questions and approaches must be developed within the best possible framework of phylogenetic relationships and evolutionary histories of traits. William Friedman (Arnold Arboretum of Harvard University) provided a fascinating tour of the evolution of secondary growth from vascular cambium. The multiple, independent origins of cambium in land plant evolution was highlighted, as were instances of cambial loss (e.g. in monocots). Friedman challenged the tree genomics research community with a number of evolutionarily-leveraged questions, including why there are no herbaceous gymnosperms and what genes and mechanisms are responsible for loss of cambium in lineages such as water lilies (Nymphaeales).

At the genome level, Steve DiFazio (West Virginia University, Morgantown, WV, USA) is examining the evolutionary dynamics of whole genome duplication in the Salicaceae by taking advantage of available genome sequences for the two member genera, *Populus* (aspens, cottonwoods, and poplars) and *Salix* (willow). Comparisons suggest a common genome duplication predating the divergence of *Salix* and *Populus*, allowing for examination of the selective retention of duplicated genes. Classically, duplicated genes have been suggested to provide the opportunity for neofunctionalization and thus play important roles in evolution and speciation. DiFazio presented analyses that seek to identify adaptive, neutral or divergent fates in duplicated genes based on differential retention of paralogs in species with contrasting life histories. The ability to

transform *Populus* and available RNAseq datasets for multiple *Salix* and *Populus* species provides experimental avenues to further determine the function and divergence of paralogs.

On the molecular level, Taku Demura (Nara Institute of Science and Technology, Ikoma, Japan) described the evolution of tracheary element and secondary cell wall development from the perspective of transcriptional regulators. Specific VASCULAR-RELATED NAC-DOMAIN (VND) transcription factors are master transcriptional regulators specifying metaxylem (AtVND6) and protoxylem (AtVND7) in *Arabidopsis*. Comparative analyses with poplar orthologs showed partial conservation of function, but with potential modification in the expanded gene families in the duplicated poplar genome. More surprisingly, Demura showed that hydroid (water-conducting) and stereid (supporting) cells in the moss, *Physcomitrella patens*, also rely on genes homologous to VND in their development despite lacking lignified cell walls. This observation provides an intriguing opportunity to determine at the molecular genetic levels what evolutionary steps led to water conducting cells with lignified, secondary cell walls within the Tracheophytes.

Population genomics, adaptive traits and climate change

Population and quantitative genetics have a long history in forest genetics research, and several major projects are underway to leverage genomics-based approaches for understanding the natural, standing genetic variation in tree populations and how it relates to variation in adaptive traits vital to the interactions of trees with the environment. Carl Douglas (University of British Columbia, Vancouver, Canada) described landscape genomics studies in *Populus trichocarpa* and the related *P. balsamifera* that used extensive SNP genotyping, RNAseq and whole genome resequencing paired with common garden studies to understand the genetic regulation of a number of adaptive traits. Evidence for genes with associations to latitude and other environmental variables were found, as well as evidence for introgression of alleles responsible for local adaptation that resulted from hybridization of *P. trichocarpa* and *P. balsamifera*. Karl Fetter (University of Vermont, Burlington, VT, USA) used the extensive range of *P. balsamifera* to ask whether populations in the extreme of the range of the species showed genomic regions associated with local adaptation. Because leading or trailing-edges of the slowly migrating species may show phenotypic extremes, this analysis could be relevant to models of speciation, adaptation to environmental extremes, and how the species responds to ongoing climate change. Evidence was presented for genomic regions associated with locally adapted loci in range-edge populations associated with warmer, drier environments.

Molecular mechanisms underlying forest tree development

Developmental biology with trees has been historically challenging, but genomics is providing new avenues for insights into novel biology in trees. Isabelle Henry (University of California Davis)

and colleagues used a sequencing-based approach to understand gender determination in the dioecious persimmon. Through comparison of genomic and transcriptome sequences from male vs female full siblings, an elegant mechanism was identified in which a small RNA encoded by the *OGI* locus represses a homologous autosomal gene, *MeGI*, which acts as a feminizing agent through repression of androecium development. Henry also described a new functional genomics resource in poplar based on induced dosage variation, which can be used to dissect gender determination and other phenotypic traits in these dioecious trees.

Also in the realm of small RNAs, Meng-Zhu Lu (Chinese Academy of Forestry, Beijing, China) described an experimental system for the regeneration of cambium after girdling in poplar. Using this system, Lu and his colleagues detailed the expression of small RNAs during cambium regeneration, identifying both known and novel miRNAs expressed during cambial development. The miRNAs include good candidates for regulating specific aspects of cambium and wood development that are being investigated in ongoing studies.

A precocious graduate student, Elizabeth Trippe (University of Georgia, Athens, GA, USA), gave a careful and well-reasoned account of studies aimed at understanding differing sucrose utilization in angiosperm trees. Using a comparative transcriptome approach, Trippe and colleagues found differences in expression of distinct plasma membrane vs tonoplast-localized sucrose transporters in trees such as eucalyptus with serotinous flowering vs trees such as poplar with early flowering. Additional studies are underway to further determine the mechanistic and evolutionary basis of these differences.

Computational approaches and how to handle 'big data'

A major challenge for forest genomics is how to extract the maximum information from genomic studies to uncover biological meaning. A related challenge is how to share and integrate genomic data across laboratories and databases. Jill Wegrzyn (University of Connecticut, Storrs, CT, USA) and her associates are using new methods for integrating databases and enabling advanced genomic analyses that minimize the need for end-user hardware investments or complicated reformatting and transfer of data among multiple platforms. Undoubtedly these efforts will be increasingly appreciated by the forest genomics community as comparative studies progress.

Computational approaches were highlighted by Siobhan Brady (University of California Davis) and Matthew Zinkgraf (US Forest Service). Brady described elegant *Arabidopsis* vascular transcriptional network models based on yeast one hybrid binding of transcription factors to target promoters. These models provide insights into features of the networks such as the predominance of feed-forward loops, as well as specific associations between regulators and target genes that allowed development of specific, testable hypotheses regarding the regulation of vascular development. Working in poplar, Zinkgraf and colleagues used perturbations of gene expression during tension wood formation to develop gene co-expression networks. Zinkgraf described how genes can be

assigned to co-expression modules, how gene modules can be correlated with phenotypes, and how modules can be further dissected to identify key regulators.

Opinions and perspectives

Delegates were asked to address a number of questions ranging from scientific to societal. Delegates identified a number of exciting scientific challenges that are now tractable, including comprehensive determination of the relationship between genetic and phenotypic variation, whether there are common genomic patterns or genes that define trees, and the nature of regulatory networks controlling tree development and responses to the environment. A major, urgent challenge in the face of ongoing climate change is to understand the regulation of adaptive traits of trees and the degree of phenotypic plasticity among traits and species. Delegates speculated that further advances in sequencing, genomic and computational approaches will fuel future research potential but will require additional tools for data storage and analysis, and institutional interactivity. Importantly, biological repositories including arboreta, herbaria, and long-term experimental plantings will become increasingly important as generating sequencing and other genomic data is no longer the bottleneck for research (Groover & Dosmann, 2012).

As a community, delegates expressed frustration with misperceptions and misrepresentation of forest genomics. While genomics can play a key role in addressing the conservation, ecological and societal pressures on forest, in general the public and policy-makers need to better understand the technology and what it represents to ultimately provide better support for research. Delegates cited challenges in securing stable funding for research programs, as well as concerns about the support for new researchers entering this competitive field of study. A call was made to study how recent graduates and postdocs are faring to provide reliable data and identify ways of better supporting the development of new researchers.

A true highlight of the symposium and ultimate perspective was provided by keynote speaker, Peter Crane (Yale University, New Haven, CT, USA). Referencing his recently published book on the ginkgo (Crane, 2015), Crane treated the delegates to the 250

million year story of the ginkgo, including a surprising twist of how humans may have helped save the tree from extinction. The interweaving of science, philosophy, history and personal experience made for a fascinating tale, and provided an exquisite example of how to bring the excitement and fascination of science to a broader audience.

Acknowledgements

Thanks to Helen Pinfield-Wells and *New Phytologist* Central Office, and the staff of the Arnold Arboretum for their instrumental roles in organizing and planning this successful symposium. Research in the Groover laboratory is supported by funding from the US Forest Service, USDA AFRI 2015-67013-22891, and DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0007183.

Andrew Groover^{1,2}

¹US Forest Service, Pacific Southwest Research Station, Davis, CA 95618, USA;

²Department of Plant Biology, University of California, Davis, CA, USA
(tel +530 759 1738; email agroover@fs.fed.us)

References

- Crane P. 2015. *Ginkgo: the tree that time forgot*. New Haven, CT, USA: Yale University Press.
- Dobzhansky H. 1973. Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher* 35: 125–129.
- Groover A, Dosmann M. 2012. The importance of living botanical collections for plant biology and the “next generation” of evo-devo research. *Frontiers in Plant Science* 3: 137, doi: 10.3389/fpls.2012.00137.
- Wetterstrand K. 2015. *DNA sequencing costs: data from the NHGRI genome sequencing program GSP*. [WWW document] URL www.genome.gov/sequencingcosts [accessed 15 June 2015].

Key words: adaptation, climate change, forest trees, population genomics, sequencing, tree genomics.