Hardwood tree genomics: unlocking woody plant biology

Introduction

Forest trees have historically been classified as softwood or hardwoods for practical purposes. Botanically, softwood trees are gymnosperms and hardwood trees are angiosperms (i. e., flowering plants) and, as the names indicate, there are differences in reproductive biology and overall habit between these two groups. Around the globe, hardwood trees underpin complex ecosystems, sequester carbon (e. g., rainforests), provide raw materials for the forest products industry, are a source of primary energy in developing countries and are increasingly utilized as a source of renewable bioenergy, biomaterials, and other bioproducts (Ragauskas et al., 2014 ; FAO, 2016).

Woody perennial angiosperms, i. e., hardwood trees, are amazingly diverse in terms of morphology, life habits, physiology, among many other traits. Hardwood trees show extensive variation in wood anatomy, leaf morphology, whole-tree architecture, secondary metabolism and numerous adaptive traits (Groover and Crook, 2017). The classification of " tree" or arboreal form reflects several distinctive features: e. g., extensive wood formation produced by a vascular or secondary cambium, persistent interannual growth, single or multiple stems exceeding two meters in height, accompanied by perennial lateral branches. Individually, these defining features can however occur outside of the classification as a tree, e. g., some annuals that can produce woody tissue or achieve heights greater than 2 m. Exception aside, hardwood trees can establish extensive root systems that enable vigorous interannual growth, store energy and nutrients overwinter in

stem and root tissues, and in many scenarios, create dominant ecosystem overstories.

The ancestral habit for angiosperms is believed to be reflected in the most basal extant angiosperm, *Amborella trichopoda* (Soltis et al., 2008 ; Albert et al., 2013), which is a small tree native to New Caledonia. During angiosperm evolution, woody perennial habit has been accentuated in some lineages, lost in others (e. g., monocots), and lost and then regained in still others (Spicer and Groover, 2010). To a large extent, these evolutionary and developmental changes are contained in the genomes of the extant hardwood tree species.

Reflecting their evolutionary history, hardwood trees genomes display divergent chromosomal architectures, with multiple ancestral whole-genome duplications and rearrangements, gene family expansions, and many segmental deletions across lineages (Tuskan et al., 2006 ; Soltis et al., 2008 ; Dai et al., 2014 ; Myburg et al., 2014 ; Plomion et al., 2018). Understanding comparative connections among genomes and the associated characteristics defining hardwood trees remains a substantial research challenge. In this review we will cover the current state of the science in five major groups of hardwood trees outlining the current assembly and annotation metrics, as well as presenting on-going application of these resources. We will highlight on-line resources for these genomes and provide comparative resources for other angiosperm species. We will provide projections of near-term technologies and analytics that will increase the availability and research value of genomes of hardwood trees and conclude with a summary of

several unanswered questions related to macro- and micro-evolution of woody perennials that can be addressed using genomic approaches in future.

State of the Science

Populus

When the *Populus trichocarpa* genome, ' Nisquallly-1,' was released in 2006 it was the first plant genome to use shotgun sequencing assembly approaches, the first genome to create chromosome-level assemblies based on genetic mapping, the first woody perennial genome to be assembled and annotated, as well as the first metagenomic assembly of associated endophytic bacteria and fungi (Tuskan et al., 2006). Today, v3. 2 contains 423 Mb [of the ~485 Mb genome] in 1, 446 scaffolds with 2, 585 gaps 1 . The scaffold N50 metric is 8 Mb and contig N50 is 205 kb, with roughly 98% of the genome in 181 scaffolds. Trained on deep RNAseq data from the Gene Atlas project 2 , there are 41, 335 predicted loci and 73, 013 protein coding transcripts. Methylation maps are available for 10 tissue types, including bud tissue, callus tissue, female catkins, internode explants, male catkins, phloem, xylem, and roots (Vining et al., 2012). Over 28 million single nucleotide polymorphisms (SNPs) from resequenced data representing over 880 native *P* . *trichocarpa* genotypes covering the core distribution of the species range are publicly available (Geraldes et al., 2013 3). This SNP resource has been used to characterize geographic structure and linkage disequilibrium (Slavov et al., 2012), detect signatures of selection across the genome (Evans et al., 2014) and identify genetic loci associated with various phenotypes based on genome-wide association approaches (Porth et

al., 2013 ; Evans et al., 2014 ; McKown et al., 2014 ; Muchero et al., 2015 ; Fahrenkrog et al., 2017 ; Liu et al., 2018). The resequenced data is currently being assembled into a *Populus* pan genome, with preliminary data suggesting that there may be as many as 20, 000 additional gene models that are not included in the current *P. trichocarpa* reference genome annotation (Pinosio et al., 2016). The first draft assemblies of *Populus deltoides* , ' WV94' and *Populus tremula × alba* hybrid, ' 717-1B4,' have been sequenced, assembled and annotated [1] . The assembly for WV94 is approximately 446 Mb in 1, 375 scaffolds, with scaffold N50 of 21. 7 Mb and contig N50 of 590 kb and there are 44, 853 protein coding loci. CRISPR-related PAM (protospacer adjacent motif) sites for *P. tremula × alba* 717 have been published are on line at http://aspendb. uga. edu/s717 . Highly contiguous, *de novo* genome assemblies have also been produced for *P* . *euphratica* (Ma et al., 2014) and *P* . *tremula* 4 , which greatly widen the phylogenetic sampling of the genus. Ongoing resource development work in *Populus* includes efforts to expand the tissue type and experimental conditions in the Gene Atlas, enlarge the number resequenced genotypes from the southern and eastern portions of the range, and release of V4. 0 of the *P* . *trichocarpa* genome based on long-read sequencing and dense genetic maps derived from resequencing 1000 progeny from 49 half-sib families. Dozens of genome-wide association studies continue across a broad array of phenotypes and efforts to estimate the somatic mutation rate found in old-growth *P* . *trichocarpa* are currently underway. Methods to integrate SNP, gene expression, transcription factor binding, gene dosage, methylation, metabolite expression, phenome, and co-evolutionary

relationships have been developed to generate a systems biology view of the molecular and regulatory interactions that lead to organismal scale traits (Henry et al., 2015 ; Liu et al., 2015 ; Joubert et al., 2017 , 2018 ; Weighill et al., 2018). New algorithms are being developed to use signal processing techniques as well as explainable artificial intelligence to build better systems biology models, which include methods to discover and model genome-wide epistasis and pleiotropy. Efforts are also underway to gain a comprehensive view of chromatin structure and transcription factor binding sites (Rao et al., 2014).

Eucalyptus

Eucalyptus species and hybrids represent the most widely cultivated hardwood biomass crop globally and *Eucalyptus grandis* (rose gum) clonal genotype ' BRASUZ1,' was selected as the reference for the genus. Published in 2014, the *E. grandis* genome was the first for the rosid order Myrtales and for the family Myrtaceae. Key features of the genome included a very high proportion (34%) and number of genes (over 12, 500) in tandem duplicate arrays, an ancient (109 Mya) genome-wide duplication event and high diversity of gene families encoding plant specialized metabolites such as phenylpropanoids and terpenes important for plant defense and pharmaceutical uses such as eucalyptus oils (Mewalal et al., 2017). BRASUZ1 was one of the last plant genomes produced exclusively with Sanger technology combined with extensive BAC-end coverage, resulting in a high-quality assembly (currently V2. 0 [1]) comprising 691 Mb in 4, 943 scaffolds with a scaffold N50 of 57. 5 Mb, contig N50 of 67 kb and 94% of the estimated 640 Mb genome (Grattapaglia and Bradshaw, 1994) in 11

pseudomolecules containing 288 scaffolds longer than 50 kb. A total of 36, 349 protein coding loci are annotated and the assembly has served as reference for extensive transcriptomics studies targeting the regulation and biosynthesis of lignocellulosic biomass (Mizrachi et al., 2014 ; Carocha et al., 2015 ; Hussey et al., 2015 ; Soler et al., 2015) and in comparative analyses with *Arabidopsis* (Davin et al., 2016) and *Populus* (Hefer et al., 2015 ; Pinard et al., 2015). Other research targets have included terpene biosynthesis (Myburg et al., 2014 ; Kulheim et al., 2015), reproductive biology (Vining et al., 2015), plant defense (Christie et al., 2015 ; Mangwanda et al., 2015 ; Oates et al., 2015 ; Meyer et al., 2016) and abiotic interactions (Plett et al., 2015 ; Spokevicius et al., 2017). High-throughput transcriptome sequencing in 100s of segregating interspecific hybrids of *E. grandis × E. urophylla* have been used to perform the first comprehensive systems genetics analysis of wood development in *Eucalyptus* (Mizrachi et al., 2017). The *E. grandis* genome has also served as a reference for whole-genome analysis of the causes of inbreeding depression (Hedrick et al., 2016) and development of a commercial, multi-species 60K SNP genotyping chip tagging 96% of the genome with 1 SNP every 12–20 kb (Silva-Junior et al., 2015). The EuCHIP60K has been a key resource to investigate genome-wide recombination, linkage disequilibrium, and nucleotide diversity (Silva-Junior and Grattapaglia, 2015), carry out genome-wide association (Resende et al., 2016 ; Muller et al., 2017) and drive genomic selection in different *Eucalyptus* species (Duran et al., 2017 ; Resende et al., 2017 ; Tan et al., 2017 , 2018). *Eucalyptus* transcriptomics resources are available at https://eucgenie. org/ . Ongoing and future efforts is focusing on

understanding genome diversity and evolution in this species-rich genus and other members of the Myrtaceae (Grattapaglia et al., 2012), including sister genera such as *Corymbia* for which a genome assembly, soon to be released, revealed conservation but dynamic evolution of terpene genes relative to *Eucalyptus* (Butler et al., 2018).

Salix

Whole-genome analysis in willow (*Salix* spp.), including development of reference genome assemblies, have focused on four species: *Salix purpurea* L. (Section Helix), *S. suchowensis* Cheng Section Helix), *S. wilsonii* Seemen (Section Wilsonianae) and *S. viminalis* L. (Section Vimen). The genome of a female *S. purpurea* ' 94006' was assembled at the J. Craig Venter Institute from Illumina sequence generated at the DOE Joint Genome Institute, including mate-pair library reads, and scaffolds were ordered using a genetic map generated from an F $_2$ mapping population genotyped using genotyping-by-sequencing at Cornell University [1] , placing ~70% of the genome (276 Mb) in chromosome-scale pseudomolecules. The *S. suchowensis* genome was sequenced using a combination of Roche 454 technology and Illumina/HiSeq 2000 reads (Dai et al., 2014). Additional genome sequencing projects are currently underway in *S. purpurea* , *S. wilsonii* , and *S. viminalis* using deep PacBio sequencing, which should result in much more contiguous and complete genomes. Sequence information and gene assemblies for multiple *Salix* species are now available at TreeGenes 5 .

The genome sequencing projects have revealed three notable differences between the *Populus* and *Salix* genomes. First, the *Salix* genomes are clearly smaller than the *Populus* genomes. Using 17-mer analysis, the *S. suchowensis* genome size was estimated at ~425 Mb (Dai et al., 2014), while the *S. purpurea* genome was estimated to be ~379 Mb by 25-mer analysis, both of which are slightly smaller than the estimate of ~429 and 450 Mb from flow cytometry of each species, respectively. Notwithstanding, all of these estimates are smaller than the genome size of *P. trichocarpa* (~485 Mb). The *Salix* genomes also show evidence of more extensive fractionation following the whole-genome Salicoid duplication that is shared by *Salix* and *Populus* (Tuskan et al., 2006), resulting in a reduced number of predicted genes and overall smaller genome size (Dai et al., 2014). Second, it is clear that the large chromosome 1 in *Populus* corresponds to portions of chromosome 1 and chromosome 16 in *Salix* , suggesting a series of major chromosomal fission and fusion events occurred after the Salicoid duplication leading to the divergence of these lineages (Berlin et al., 2010 ; Hou et al., 2016). Finally, the sex determination locus is located on chromosome 15 in *Salix* , while it is on chromosome 19 in all *Populus* species studied to date, supporting the hypothesis that there has been greater genome fractionation in *Salix* following the Salicoid duplication event. Furthermore, most *Populus* species have an XY sex determination system (Tuskan et al., 2012 ; Geraldes et al., 2015), while all sequenced *Salix* species have a ZW system (Hou et al., 2015 ; Pucholt et al., 2015 ; Chen et al., 2016).

Given the apparently dynamic nature of sex determination in this dioecious family, rigorous efforts to understand this trait have been undertaken in recent years. Extensive differential gene expression has been documented in shoot tips of *S. purpurea* , which display an abundance of female-biased gene expression (Carlson et al., 2017). Expression analysis in *S. viminalis* suggests a predominance of male-biased gene expression, perhaps due to masculinization of the Z chromosome (Pucholt et al., 2017). Analysis of differentially expressed genes in unflushed buds of *S. suchowensis* identified 806 differentially expressed genes between males and females (Liu et al., 2015).

There are also extensive genetic mapping resources to support efforts to relate genotype to phenotype and accelerate willow breeding using molecular tools. In addition to the F $_2$ map used to anchor the *S. purpurea* genome assemblies (Kopp et al., 2002), more than a dozen linkage mapping populations have been developed for *S. viminalis* (Karp et al., 2011), as well as an association mapping population (Hallingbäck et al., 2016). New mapping populations have also been developed based on an approach of using *S. purpurea* male ' 94001' and female ' 94006' reference individuals as common parents crossed with a number of other species, including *S. viminalis* , *S. suchowensis* , *S. integra* , *S. koriyanagi* , and *S. udensis* at Cornell University. At Nanjing Forestry University, an effort has been employed to establish recombinant inbred lines for *S. suchowensis* , with an inbred F $_4$ pedigree currently established in the field.

Quercus

A highly contiguous haploid genome of a heterozygous pedunculate oak (*Quercus robur*) was recently generated based on Illumina synthetic long-reads and 454 (Roche) sequences (Plomion et al., 2018). The assembly contains 1, 409 scaffolds totaling 814 Mb (N50, 1. 34 Mb) 6 . Overall 871 scaffolds, representing 96% of the physical size of the genome, were anchored to the 12 chromosomes. Based on RNA-seq and protein homologies, 25, 808 genes were predicted. Two other draft genomes, while of lower quality, have also been published (Sork et al., 2016 ; Schmid-Siegert et al., 2017). As in *Eucalyptus* , one of the main features of the oak genome is its remarkably high level of proximal tandem duplication (35. 6% of the gene models). The tight relationships between duplicated genes and lineage-specific selection already reported in other species was found to be particularly exacerbated in oak, as three quarters of the genes in expanded orthogroups were also found in tandem duplications. Another interesting feature of long-lived tree species rests in the accumulation of somatic mutations. The presence and transmission of somatic mutations was recently demonstrated in oak (Schmid-Siegert et al., 2017 ; Plomion et al., 2018), which raises interesting questions related to the role of somatic mutation in the evolution of long-lived species with a potentially high mutational load.

One of the major challenges in plant biology is identifying and characterizing the genes responsible for variation in ecologically and economically important traits. Genetic architecture of traits related to growth, phenology, water metabolism, and defense against pathogens has been explored in different full-sib oak pedigrees (e. g., Brendel et al., 2008) and natural

populations (e. g., Alberto et al., 2013). Linkage and association mapping revealed many underlying loci with low to moderate contributions to the trait of interest. The characterization of the oak transcriptome is more recent. Gene repertoires were first constructed (Ueno et al., 2010 ; Pereira-Leal et al., 2014 ; Cokus et al., 2015 ; Lesur et al., 2015) and RNA-seq data were then used to (1) analyze molecular plasticity under abiotic stress (e. g., drought, Spieß et al., 2012) and biotic interactions (e. g., with insects, Kersten et al., 2013) and (2) identify genes and gene networks underlying developmental mechanisms (e. g., bud phenology, Ueno et al., 2013 and acorn development, Miguel et al., 2015) and evolutionary processes including reproductive isolation (Le Provost et al., 2012 , 2016) and local adaptation (Gugger et al., 2016a). Recent studies provide evidence for epigenetic variation in phenotypic plasticity and evolutionary processes in oaks (Gugger et al., 2016b), establishing new avenues for research into the role of epigenetics in trait plasticity for these long-lived species.

Shortlisted among the ' botanical horror stories' (Rieseberg et al., 2006), the *Quercus* genus constitutes an ideal taxon to investigate the dynamics of lineage diversification along a wide and fluid continuum of speciation. Several efforts to illuminate oak evolutionary histories using population genomic approaches have recently emerged (Ortego et al., 2016 , 2018 ; Leroy et al., 2017 , 2018). In European white oaks for example, the best-supported scenario of divergence is consistent with a long-term pervasive effect of interspecific gene flow, with the exception of some narrow genomic regions responsible for reproductive isolation. Further work will be required to identify genes targeted by intrinsic or ecological selection, but early

attempts appear promising and could enable predictions of the evolutionary responses of oaks to climate change (Gugger et al., 2016a ; Rellstab et al., 2016). An exciting perspective in population genomics is presented by allochronic approaches that compare ancient and modern DNA to derive evolutionary changes at targeted genomic regions. Wagner et al. (2018) have suggested ancient DNA could be retrieved from waterlogged archeological or fossil samples, thus enabling the incremental reconstruction of evolutionary trajectories.

Chestnut, Pecan, and Other Early Draft Angiosperm Genomes

Other reference genome projects for hardwoods are underway, including efforts to produce reference sequences for chestnut and pecans. The chestnut project aims to produce reference genomes of both *Castanea mollissima* (Chinese chestnut), the donor of resistance to *Cryphonectria parasitica* (chestnut blight) and *Castanea dentata* (American chestnut) the species destroyed by the blight in the early 1900s. The Forest Health Initiative (Nelson et al., 2014) supported a project to sequence the genome of the Chinese chestnut donor parent tree ' Vanuxem' used by The American Chestnut Foundation's (TACF). The first draft of the Chinese chestnut genome assembly was released in January 2014 at: https://www. hardwoodgenomics. org/chinese-chestnut-genome , consisting of 41, 260 scaffolds averaging 39. 6 kb in length and covering 98% (~724 Mb) of the genome. Targeted regions of mapped QTL resistance loci from various clones have also been produced (Staton et al., 2015). In addition, work is underway to prepare chromosome-length sequences for the Chinese

chestnut genome. Long, single-molecule PacBio data are being used to merge contigs, followed by anchoring of scaffolds to high-density linkage maps for the cv. Vanuxem. Fluorescent *in situ* hybridization is being used to validate the chromosome-level assemblies and subsequently to identify structural rearrangements among genotypes and species (Islam-Faridi et al., 2016). The Chinese chestnut reference genome is being utilized to develop genomic selection models for blight resistance in American chestnut backcross populations (Westbrook, 2018). The project to sequence the American chestnut has produced a contig assembly, based on PacBio sequencing, of 8. 1 Mb contig N50 and efforts are now underway to use new technologies (see below) and genetic mapping to produce chromosome-scale assemblies for other *C* . *mollissima* and *C* . *dentata* genotypes.

Initial efforts at generating a *Carya illinoinensis* (pecan) reference genome, released in 2013 and based on Illumina short-read sequencing (Jenkins et al., 2015) resulted in a contig N50 of 6. 5 kb and scaffold N50 of 11. 2 kb. The Mexican accession ' 87MX3-2. 11' from the USDA-ARS facility in Somerville, Texas was selected as the reference genome based on a reduced heterozygosity. Efforts are underway to generate PacBio-based genomes for pecans that include 87MX3-2 and three production cultivars [7]. The current version of the pre-chromosome assembly for 87MX3-2 covers 705. 6 Mb of the genome with a contig N50 of 2. 5 Mb with 36, 489 genes annotated [1] . Multiple early efforts are also underway to sequenced the European ash (*Fraxinus excelsior*) and American ash (*F. americana*) [8,9] , as well as several birch genomes (*Betula nana* , *B* . *pubescens* , and *B. pendula*) (Wang et al., 2013). Finally, several angiosperm genomes are in early draft

form and include species of ash [5] and walnut [9] , as well as several comparative angiosperms from horticultural species [10] .

A Look to the Near Future

The continued development of new technologies that can be applied to generate high-quality hardwood reference genomes has been fostered by single-molecule, real-time sequencing (SMRT) and related assembly algorithms from PacBio (Eid et al., 2009). These long reads, which average 10–15 kb in length, but are frequently 20–50 kb in length, can be used to resolve complex genomic repeats and improve contiguity of genomes. An early example of a small, inbred, grass genome (*Oropetium thomaeum*) achieved a contig N50 of 2. 4 Mb, while capturing almost the entire genome space (244 out of 245 Mb) (VanBuren et al., 2015). The PacBio long reads were also used to produce a reference for the large, complex sunflower (*Helianthus annuus*) genome (Badouin et al., 2017) at 3. 6 Gb with a contig N50 of 524 kb. An alternative reference to the *Zea mays* B73 supports using PacBio Iso-Seq sequencing for collecting full-length cDNAs to enhance the reference annotation and a genome (Jiao et al., 2017), which despite large-scale nested transposon activity, was completed at a contig N50 of 1. 18 Mb. Because of the low accuracy rate of individual PacBio reads, these reads must be combined with an alternative technology, such as Illumina short reads, to polish the consensus sequence. For hardwood tree genomes, the longer reads facilitate the generation of outbred reference sequences to sort haplotypes and produce a single reference despite high heterozygosity rates (see *Q. robur* , *P. deltoides* WV94 and *Castanea dentata* described above). In addition to long-read sequencing, which has brought us back to the

completeness of Sanger-based sequencing, new tools are emerging to achieve accurate chromosome-scale contiguity. These tools for genomic mapping and assembly include improved single-molecule mapping from BioNano (Staňková et al., 2016), *in vitro* reconstituted chromatin (Putnam et al., 2016) and binned sequencing approaches such as 10x genomics (Coombe et al., 2016). The most promising of these technologies is Hi-C for ordering large contigs from genome assemblies. Originally developed to perform " contact mapping" in human cell lines to show genes adjacent to promoter or regulatory elements (Suhas, 2014), it has been repurposed as a general solution to determining order in genome assemblies (Dudchenko et al., 2017). A recent example, the 4. 8 Gb genome of *Hordeum vulgare* , has shown the ability to use Hi-C to precisely place 95% of contigs on pseudomolecules (Mascher et al., 2017). All of these tools are actively being applied to improve hardwood genomic references as a foundation for accurate population and functional analysis.

These new technologies will enable greater within and across species comparisons. An example of unique tree biology, that can be explored, is the characterization of somatic mutations. Due to their perennial habit, trees can accumulate somatic mutations in alternate vegetative lineages (as noted in the *Populus* and *Quercus* sections above). However, the effect of somatic variation on the generation of new genetic variation at the population-level and/or during reproduction remains largely unknown. Since, in plants, germlines are not segregated from the vegetative lineages, somatic mutations can be transmitted to the next generation. The frequency with which a given somatic mutation propagates to the next generation depends

both on the overall fitness of the tree and the probability of the somatic sector giving rise to flowers and gametes. Because branches/tissue types do not make equal contributions to the resulting gene pool, somatic mutations may alter the adaptive balance between branches. This mosaic genetic architecture raises the possibility that selection acts simultaneously on both the branch and tree (Hadany, 2001 ; Clarke, 2011). Understanding the role of multi-level selection within a single tree will require answering several questions. That is, to what extent is the effect of a somatic mutation in a single branch shared with the entire tree (Folse and Roughgarden, 2011)? And, if a branch acquires a broadly-beneficial mutation, such as one conferring resistance to herbivory (e. g., Edwards et al., 1990), does the resistant branch gain a larger fitness benefit than the remainder of the tree? Conversely, can a mutation increase the fitness of a single branch despite imposing cost on the rest of the tree? For example, a mutation that increases flowering in a branch might increase seed production from that branch while depressing production from the tree as a whole by acting as a metabolic sink (Walbot, 1985). Finally, should these conflicts occur, do trees have mechanisms to mediate resource allocation among branches, and, if so, are these mechanisms common between independent tree lineages? New techniques (described above) for finding or constructing whole-genome genetic mosaic and chimeric trees with phenotypically-relevant somatic mutations are now allowing these questions to be addressed.

In summary, the assembly and annotation of multiple hardwood tree genomes has facilitated an increase in (1) functional characterization of genes and gene networks related to tree habit, (2) GWAS and genomic

selection investigations and applications, and (3) comparative genomic efforts among various tree species. The rapidly expanding new technologies will add even greater number of hardwood species to these efforts. The power of comparative genomics will increase our understanding of how these highly dynamic genomes have evolved and resulted in the amazing array of phenotypic diversity found among and within hardwood species. To broaden the raft of available hardwood genomes, resources should be directed toward additional candidate hardwood genera, including but is not limited to *Liriodendron* , *Liquidambar* , *Swietenia* , and *Acer* . In sum total, most hardwoods are undomesticated long-lived plants that provide many ecological and commercial benefits, whose management, conservation and domestication for economic and ecological purposes will benefit from a set of rich genomics resources. Ultimately such resources will favorably impact the pressing problems of climate change, soil and water conservation, bioenergy and biomaterials production, and maintenance of heathy ecosystem functions. As a result, we may finally answer the questions of: (1) why has the tree habit evolved repeatedly in the angiosperms and (2) what is the connection between the genomes and the defining characteristics of long-lived perennial plants?

Author Contributions

All authors contributed equally to the formatting and writing contained in this Mini-review.

Funding

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Footnotes

1. ^ https://phytozome. jgi. doe. gov

2. ^ https://jgi. doe. gov/doe-jgi-plant-flagship-gene-atlas/

3. ^ https://cbi. ornl. gov/

4. ^ http://popgenie. org/aspseq

5. ^ https://www. treegenesdb. org/

6. ^ http://www. oakgenome. fr/

7. ^ pecantoolbox. nmsu. edu

8. ^ http://www. ashgenome. org/

9. ^ http://hardwoodgenomics. org

10.   ^ https://www. rosaceae. org/

References

Albert, V. A., Barbazuk, W. B., Der, J. P., Leebens-Mack, J., Ma, H., Palmer, J. D., et al. (2013). The *Amborella* genome and the evolution of flowering plants. *Science* 342: 1241089. doi: 10. 1126/science. 1241089

Alberto, F. J., Derory, J., Boury, C., Frigério, J. M., Zimmermann, N. E., and Kremer, A. (2013). Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea* . *Genetics* 195, 495–512. doi: 10. 1534/genetics. 113. 153783

Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., et al. (2017). The sunflower genome provides insights into oil metabolism, flowering and asterid evolution. *Nature* 546, 148–152. doi: 10. 1038/nature22380

Berlin, S., Lagercrantz, U., von Arnold, S., Öst, T., and Rönnberg-Wästljung, A. C. (2010). High-density linkage mapping and evolution of paralogs and orthologs in *Salix* and *Populus* . *BMC Genomics* 11: 129. doi: 10. 1186/1471-2164-11-129

Brendel, O., Le Thie, D., Scotti-Saintagne, C., Bodénès, C., Kremer, A., and Guehl, J. M. (2008). Detection of quantitative trait loci controlling water use efficiency and related traits in *Quercus robur* L. *Tree Genet. Genomes* 4, 263–278. doi: 10. 1007/s11295-007-0107-z

Butler, J. B., Freeman, J. S., Potts, B. M., Vaillancourt, R. E., Grattapaglia, D., Silva-Junior, O. B., et al. (2018). Annotation of the *Corymbia* terpene synthase gene family shows broad conservation but dynamic evolution of

physical clusters relative to *Eucalyptus* . *Heredity* 121, 87–104. doi: 10. 1038/s41437-018-0058-1

Carlson, C. H., Choi, Y., Chan, A. P., Serapiglia, M. J., Town, C. D., and Smart, L. B. (2017). Dominance and sexual dimorphism pervade the *Salix purpurea* l. transcriptome. *Genome Biol. Evol.* 9, 2377–2394. doi: 10. 1093/gbe/evx174

Carocha, V., Soler, M., Hefer, C., Cassan-Wang, H., Fevereiro, P., Myburg, A. A., et al. (2015). Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis* . *New Phytol.* 206, 1297–1313. doi: 10. 1111/nph. 13313

Chen, Y., Wang, T., Fang, L., Li, X., and Yin, T. (2016). Confirmation of single-locus sex determination and female heterogamety in willow based on linkage analysis. *PLoS One* 11: e0147671. doi: 10. 1371/journal. pone. 0147671

Christie, N., Tobias, P. A., Naidoo, S., and Kulheim, C. (2015). The *Eucalyptus grandis* NBS-LRR gene family: physical clustering and expression hotspots. *Front. Plant. Sci.* 6: 1238. doi: 10. 3389/fpls. 2015. 01238

Clarke, E. (2011). " Plant individuality and multilevel selection theory," in *Major Transitions in Evolution Revisited* , eds B. Calcott and K. Sterelny (Cambridge, MA: The MIT Press).

Cokus, S. J., Gugger, P. F., and Sork, V. L. (2015). Evolutionary insights from de novo transcriptome assembly and SNP discovery in California white oaks. *BMC Genomics* 16: 552. doi: 10. 1186/s12864-015-1761-4

Coombe, L., Warren, R. L., Jackman, S. D., Yang, C., Vandervalk, B. P., Moore, R. A., et al. (2016). Assembly of the complete Sitka spruce chloroplast

genome using 10x genomics gemcode sequencing data. *PLoS One* 11: e0163059. doi: 10. 1371/journal. pone. 0163059

Dai, X., Hu, Q., Cai, Q., Feng, K., Ye, N., Tuskan, G. A., et al. (2014). The willow genome and divergent evolution from poplar after the common genome duplication. *Cell Res.* 24, 1274–1277. doi: 10. 1038/cr. 2014. 83

Davin, N., Edger, P. P., Hefer, C. A., Mizrachi, E., Schuetz, M., Smets, E., et al. (2016). Functional network analysis of genes differentially expressed during xylogenesis in soc1ful woody *Arabidopsis* plants. *Plant J.* 86, 376–390. doi: 10. 1111/tpj. 13157

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., et al. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356: 92. doi: 10. 1126/science. aal3327

Duran, R., Isik, F., Zapata-Valenzuela, J., Balocchi, C., and Valenzuela, S. (2017). Genomic predictions of breeding values in a cloned *Eucalyptus globulus* population in Chile. *Tree Genet. Genomes* 13: 74. doi: 10. 1007/s11295-017-1158-4

Edwards, P., Wanjura, W., Brown, W., and Dearn, J. (1990). Mosaic resistance in plants. *Nature* 347: 434. doi: 10. 1038/347434a0

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138. doi: 10. 1126/science. 1162986

Evans, L. M., Slavov, G. T., Rodgers-Melnick, E., Martin, J., Ranjan, P., Muchero, W., et al. (2014). Population genomics reveals signatures of selection coupled with adaptive trait associations in the model tree, *Populus trichocarpa. Nat. Genet.* 46, 1089–1096. doi: 10. 1038/ng. 3075

Fahrenkrog, A. M., Neves, L. G., Resende, M. F. R. Jr, Gustavo de los Campos AIV, Dervinis, C., et al. (2017). Genome-wide association study reveals putative regulators of bioenergy traits in *Populus* deltoides. *New Phytol.* 213, 799–811. doi: 10. 1111/nph. 14154

FAO (2016). *State of the World's Forests 2016.* Available at: http://www. fao. org/publications/sofo/2016/en/

Folse, H. J. III, and Roughgarden, J. (2011). Direct benefits of genetic mosaicism and intraorganismal selection: modeling coevolution between a long-lived tree and a short-lived herbivore. *Evolution* 66, 1091–1113. doi: 10. 1111/j. 1558-5646. 2011. 01500. x

Geraldes, A., Hannemann, J., Grassa, C., Farzaneh, N., Porth, I., McKown, A., et al. (2013). A 34K SNP genotyping array for *Populus trichocarpa* : design, application to the study of natural populations and transferability to other *Populus* species. *Mol. Ecol. Resour.* 13, 306–323. doi: 10. 1111/1755-0998. 12056

Geraldes, A., Hefer, C. A., Capron, A., Kolosova, N., Martinez-Nuñez, F., Soolanayakanahally, R. Y., et al. (2015). Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). *Mol. Ecol.* 24, 3243– 3256. doi: 10. 1111/mec. 13126

Grattapaglia, D., Vaillancourt, R. E., Shepherd, M., Thumma, B. R., Foley, W., Külheim, C., et al. (2012). Progress in *Myrtaceae* genetics and genomics: *Eucalyptus* as the pivotal genus. *Tree Genet. Genomes* 8, 463–508. doi: 10. 1007/s11295-012-0491-x

Grattapaglia, D., and Bradshaw, H. D. (1994). Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. *Can. J. For. Res.* 24, 1074–1078. doi: 10. 1139/x94-142

Groover, A., and Crook, Q., (eds) (2017). *Comparative and Evolutionary Genomics of Angiosperm Trees.* Berlin: Springer Publishing. 366. doi: 10. 1007/978-3-319-49329-9

Gugger, P. F., Cokus, S. J., and Sork, V. L. (2016a). Association of transcriptome-wide sequence variation with climate gradients in valley oak (*Quercus lobata*). *Tree Genet. Genomes* 12: 15. doi: 10. 1007/s11295-016-0975-1

Gugger, P. F., Fitz-Gibbon, S., Pellegrini, M., and Sork, V. L. (2016b). Species-wide patterns of DNA methylation variation in *Quercus lobata* and their association with climate gradients. *Mol. Ecol.* 25, 1665–1680. doi: 10. 1111/mec. 13563

Hadany, L. (2001). A conflict between two evolutionary levels in trees. *J. Theor. Biol.* 208, 507–521. doi: 10. 1006/jtbi. 2000. 2236

Hallingbäck, H. R., Fogelqvist, J., Powers, S. J., Turrion-Gomez, J., Rossiter, R., Amey, J., et al. (2016). Association mapping in *Salix viminalis* L. (Salicaceae)

– identification of candidate genes associated with growth and phenology. *GCB Bioenergy* 8, 670–685. doi: 10. 1111/gcbb. 12280

Hedrick, P., Hellesten, U., and Grattapaglia, D. (2016). Examining the cause of high inbreeding depression: analysis of whole-genome sequence data in 28 selfed progeny of *Eucalyptus grandis* . *New Phytol.* 209, 600–611. doi: 10. 1111/nph. 13639

Hefer, C. A., Mizrachi, E., Myburg, A. A., Douglas, C. J., and Mansfield, S. D. (2015). Comparative interrogation of the developing xylem transcriptomes of two wood-forming species: *Populus trichocarpa* and *Eucalyptus grandis* . *New Phytol.* 206, 1391–1405. doi: 10. 1111/nph. 13277

Henry, I. M., Zinkgraf, M. S., Groover, A. T., and Comai, L. (2015). A system for dosage-based functional genomics in poplar. *Plant Cell* 27, 2370–2383.

Hou, J., Ye, N., Dong, Z., Lu, M., Li, L., and Yin, T. (2016). Major chromosomal rearrangements distinguish willow and poplar after the ancestral " Salicoid" genome duplication. *Genome Biol. Evol.* 8, 1868–1875. doi: 10. 1093/gbe/evw127

Hou, J., Ye, N., Zhang, D., Chen, Y., Fang, L., Dai, X., et al. (2015). Different autosomes evolved into sex chromosomes in the sister genera of *Salix* and *Populus* . *Sci. Rep.* 5: 9076. doi: 10. 1038/srep09076

Hussey, S. G., Saidi, M. N., Hefer, C. A., Myburg, A. A., and Grima-Pettenati, J. (2015). Structural, evolutionary and functional analysis of the NAC domain

protein family in *Eucalyptus* . *New Phytol.* 206, 1337–1350. doi: 10. 1111/nph. 13139

Islam-Faridi, N., Majid, M. A., Zhebentyayeva, T., Georgi, L. L., Fan, S., Hebard, V., et al. (2016). FISH confirmation of a reciprocal translocation in chestnut. *Cytogenet. Genome Res.* 148, 144–144.

Jenkins, J., Wilson, B., Grimwood, J., Schmutz, J., and Grauke, L. (2015). Towards a reference pecan genome sequence. *Acta Hortic.* 1070, 101–108. doi: 10. 17660/ActaHortic. 2015. 1070. 11

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., et al. (2017). Improved maize reference genome with single-molecule technologies. *Nature* 546, 524–527. doi: 10. 1038/nature22971

Joubert, W., Nance, J., Climer, S., Weighill, D., and Jacobson, D. (2017). Parallel accelerated custom correlation coefficient calculations for genomics applications. *arXiv* [Preprint]. arXiv: 1705. 08213

Joubert, W., Nance, J., Weighill, D., and Jacobson, D. (2018). Parallel accelerated vector similarity calculations for genomics applications. *Parallel Comput.* 75, 130–145. doi: 10. 1016/j. parco. 2018. 03. 009

Karp, A., Hanley, S. J., Trybush, S. O., Macalpine, W., Pei, M., and Shield, I. (2011). Genetic improvement of willow for bioenergy and biofuels. *J. Integr. Plant Biol.* 53, 151–165. doi: 10. 1111/j. 1744-7909. 2010. 01015. x

Kersten, B., Ghirardo, A., Schnitzler, J. P., Kanawati, B., Schmitt-Kopplin, P., Fladung, M., et al. (2013). Integrated transcriptomics and metabolomics

decipher differences in the resistance of pedunculate oak to the herbivore *Tortrix viridana* L. *BMC Genomics* 14: 737. doi: 10. 1186/1471-2164-14-737

Kopp, F., Smart, B., Maynard, A., Tuskan, A., and Abrahamson, P. (2002). Predicting within-family variability in juvenile height growth of *Salix* based upon similarity among parental AFLP fingerprints. *Theor. Appl. Genet.* 105, 106–112.

Kulheim, C., Padovan, A., Hefer, C., Krause, S. T., Kollner, T. G., Myburg, A. A., et al. (2015). The *Eucalyptus terpene* synthase gene family. *BMC Genomics* 16: 450. doi: 10. 1186/s12864-015-1598-x

Le Provost, G., Sulmon, C., Frigerio, J. M., Bodenes, C., Kremer, A., and Plomion, C. (2012). Role of waterlogging genes in shaping interspecific differentiation between two sympatric oak species. *Tree Physiol.* 32, 119–134. doi: 10. 1093/treephys/tpr123

Le Provost, G., Lesur, I., Lalanne, C., Da Silva, C., Labadie, K., Aury, J. M., et al. (2016). Implication of the suberin pathway in adaptation to waterlogging and hypertrophied lenticels formation in pedunculate oak (*Quercus robur* L.) *Tree Physiol.* 36, 1330–1342. doi: 10. 1093/treephys/tpw056

Leroy, T., Roux, C., Villate, L., Bodénès, C., Romiguier, J., Paiva, J. A. P., et al. (2017). Extensive recent secondary contacts between four European white oak species. *New Phytol.* 214, 865–878. doi: 10. 1111/nph. 14413

Leroy, T., Rougemont, Q., Bodénes, C., Lalanne, C., Belser, C., Labadie, K., et al. (2018). Secondary contacts between European white oaks reveal genes underlying reproductive isolation. *bioRxiv* [Preprint]. doi: 10. 1101/246637

Lesur, I., Le Provost, G., Bento, P., Da Silva, C., Leplé, J. C., Murat, F., et al. (2015). The oak gene expression Atlas: insights into Fagaceae genome evolution and the discovery of genes regulated during bud dormancy release. *BMC Genomics* 16: 112. doi: 10. 1186/s12864-015-1331-9

Liu, J., Ye, M., Zhu, S., Jiang, L., Sang, M., Gan, J., et al. (2018). Two-stage identification of SNP effects on dynamic poplar. *Plant J.* 93, 286–296. doi: 10. 1111/tpj. 13777

Liu, L., Ramsay, T., Zinkgraf, M., Sundell, D., Street, N. R., Filkov, V., et al. (2015). A resource for characterizing genome-wide binding and putative target genes of transcription factors expressed during secondary growth and wood formation in *Populus* . *Plant J.* 82, 887–898.

Ma, T., Wang, J., Zhou, G., Yue, Z., Hu, Q., Chen, Y., et al. (2014). Genomic insights into salt adaptation in a desert poplar. *Nat. Commun.* 5: 2797. doi: 10. 1038/ncomms3797

Mangwanda, R., Myburg, A. A., and Naidoo, S. (2015). Transcriptome and hormone profiling reveals *Eucalyptus grandis* defence responses against *Chrysoporthe austroafricana* . *BMC Genomics* 16: 319. doi: 10. 1186/s12864-015-1529-x

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544: 427. doi: 10. 1038/nature22043

McKown, A. D., Klápštì, J., Guya, R. D., Geraldes, A., Portha, I., Hannemanne, J., et al. (2014). Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa* . *New Phytol.* 203, 535–553. doi: 10. 1111/nph. 12815

Mewalal, R., Rai, D. K., Kainer, D., Chen, F., Külheim, C., Peter, G. F., et al. (2017). Plant-derived terpenes: a feedstock for specialty biofuels. *Trends Biotechnol.* 35, 227–240. doi: 10. 1016/j. tibtech. 2016. 08. 003

Meyer, F. E., Shuey, L. S., Naidoo, S., Mamni, T., Berger, D. K., Mybur, A. A., et al. (2016). Dual RNA-sequencing of *Eucalyptus nitens* during *Phytophthora cinnamomi* challenge reveals pathogen and host factors influencing compatibility. *Front. Plant Sci.* 7: 191. doi: 10. 3389/fpls. 2016. 00191

Miguel, A., Vega Bartol, J., Marum, L., Chaves, I., Santo, T., Leitão, J., et al. (2015). Characterization of the cork oak dynamics during acorn development. *BMC Plant Biol.* 15: 258. doi: 10. 1186/s12870-015-0534-1

Mizrachi, E., Maloney, V. J., Silberbauer, J., Hefer, C. A., Berger, D. K., Mansfield, S. D., et al. (2014). Investigating the molecular underpinnings underlying morphology and changes in carbon partitioning during tension wood formation in *Eucalyptus* . *New Phytol.* 206, 1351–1363. doi: 10. 1111/nph. 13152

Mizrachi, E., Verbeke, L., Christie, N., Fierro, A. C., Mansfield, S. D., Davis, M. F., et al. (2017). Network-based integration of systems genetics data reveals pathways associated with lignocellulosic biomass accumulation and processing. *Proc. Natl. Acad. Sci. U. S. A.* 114, 1195–1200. doi: 10. 1073/pnas. 1620119114

Muchero, W., Guo, J., DiFazio, S. P., Chen, J. G., Ranjan, P., Slavov, G. T., et al. (2015). High-resolution genetic mapping of allelic variants associated with cell wall chemistry in *Populus* . *BMC Genomics* 16: 24. doi: 10. 1186/s12864-015-1215-z

Muller, B. S. F., Neves, L. G., Almeida-Filho, J. E., Resende, M. F. R. J., Munoz Del, Valle, P., et al. (2017). Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of *Eucalyptus* . *BMC Genomics* 18: 524. doi: 10. 1186/s12864-017-3920-2

Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., et al. (2014). The genome of *Eucalyptus grandis* . *Nature* 510, 356–362. doi: 10. 1038/nature13308

Nelson, C. D., Powell, W. A., Maynard, C. A., Baier, K. M., Newhouse, A., Merkle, S. A., et al. (2014). The forest health initiative, American chestnut (*Castanea dentata*) as a model for forest tree restoration: biological research program. *Acta Hortic.* 1019, 179–189.

Oates, C. N., Kulheim, C., Myburg, A. A., Slippers, B., and Naidoo, S. (2015). The transcriptome and terpene profile of *Eucalyptus grandis* reveals

mechanisms of defense against the insect pest, *Leptocybe invasa. Plant Cell Physiol.* 56, 1418–1428. doi: 10. 1093/pcp/pcv064

Ortego, J., Gugger, P. F., and Sork, V. L. (2018). Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section *Protobalanus*). *New Phytol.* 2, 804–818. doi: 10. 1111/nph. 14951

Ortego, J., Gutter, P. F., and Sork, V. L. (2016). Impacts of human-induced environmental disturbances on hybridization between two ecologically differentiated Californian oak species. *New Phytol.* 213, 942–955. doi: 10. 1111/nph. 14182

Pereira-Leal, J. B., Abreu, I. A., Alabaça, C. S., Almeida, M. H., Almeida, P., Almeida, T., et al. (2014). A comprehensive assessment of the transcriptome of cork oak (*Quercus* suber) through EST sequencing. *BMC Genomics* 15: 371. doi: 10. 1186/1471-2164-15-371

Pinard, D., Mizrachi, E., Hefer, C. A., Kersting, A. R., Joubert, F., Douglas, C. J., et al. (2015). Comparative analysis of plant carbohydrate active enzymes and their role in xylogenesis. *BMC Genomics* 16: 402. doi: 10. 1186/s12864-015-1571-8

Pinosio, S., Giacomello, S., Faivre-Rampant, P., Taylor, G., Jorge, V., Le Paslier, M. C., et al. (2016). Characterization of the poplar pan-genome by genome-wide identification of structural variation. *Mol. Biol. Evol.* 33, 2706–2719. doi: 10. 1093/molbev/msw161

Plett, J. M., Kohler, A., Khachane, A., Keniry, K., Plett, K. L., Martin, F., et al. (2015). The effect of elevated carbon dioxide on the interaction between *Eucalyptus grandis* and diverse isolates of *Pisolithus* sp. is associated with a complex shift in the root transcriptome. *New Phytol.* 206, 1423–1436. doi: 10. 1111/nph. 13103

Plomion, C., Aury, J. M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., et al. (2018). Oak genome reveals facets of long lifespan. *Nat. Plants* 4, 440–452. doi: 10. 1038/s41477-018-0172-3

Porth, I., Klapšte, J., Skyba, O., Geraldes, A., Hannemann, J., McKown, A. D., et al. (2013). Genome-wide association mapping for wood characteristics in *Populus* identifies an array of candidate SNPs. *New Phytol.* 200, 710–726. doi: 10. 1111/nph. 12422

Pucholt, P., Hallingbäck, H. R., and Berlin, S. (2017). Allelic incompatibility can explain female biased sex ratios in dioecious plants. *BMC Genomics* 18: 251. doi: 10. 1186/s12864-017-3634-5

Pucholt, P., Rönnberg-Wästljung, A.-C., and Berlin, S. (2015). Single locus sex determination and female heterogamety in the basket willow (*Salix viminalis* L.). *Heredity* 114, 575–583. doi: 10. 1038/hdy. 2014. 125

Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., et al. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* 26: 342. doi: 10. 1101/gr. 193474. 115

Ragauskas, A. J., Beckham, G., Biddy, M., Chandra, R., Davis, M., Davison, B., et al. (2014). Lignin valorization in the biorefinery. *Science* 344, 709–714.

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. doi: 10. 1016/j. cell. 2014. 11. 021

Rellstab, C., Zoller, S., Walthert, L., Lesur, I., Pluess, A. R., Graf, R., et al. (2016). Signatures of local adaptation in candidate genes of oaks (*Quercus* spp.) with respect to present and future climatic conditions. *Mol. Ecol.* 25, 5907–5924. doi: 10. 1111/mec. 13889

Resende, R. T., Resende, M. D. V., Silva, F. F., Azevedo, C. F., Takahashi, E. K., Silva-Junior, et al. (2017). Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. *Heredity* 119, 245–255. doi: 10. 1038/hdy. 2017. 37

Resende, R. T., Resende, M. D. V., Silva, F. F., Azevedo, C. F., Takahashi, E. K., Silva-Junior, O. B., et al. (2016). Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in *Eucalyptus* . *New Phytol.* 213, 1287–1300. doi: 10. 1111/nph. 14266

Rieseberg, L. H., Wood, T. E., and Baack, E. J. (2006). The nature of plant species. *Nature* 440, 524–527. doi: 10. 1038/nature04402

Schmid-Siegert, E., Sarkar, N., Iseli, C., Calderon, S., Gouhier-Darimont, C., Chrast, J., et al. (2017). Low number of fixed somatic mutations in a long-lived oak tree. *Nat. Plants* 3, 926–929. doi: 10. 1038/s41477-017-0066-9

Silva-Junior, O. B., Faria, D. A., and Grattapaglia, D. (2015). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytol.* 206, 1527–1540. doi: 10. 1111/nph. 13322

Silva-Junior, O. B., and Grattapaglia, D. (2015). Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis* . *New Phytol.* 208, 830–845. doi: 10. 1111/nph. 13505

Slavov, G. T., DiFazio, S. P., Martin, J., Schackwitz, W., Muchero, W., Rodgers-Melnick, E., et al. (2012). Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa* . *New Phytol.* 196, 713–725. doi: 10. 1111/j. 1469-8137. 2012. 04258. x

Soler, M., Camargo, E. L., Carocha, V., Cassan-Wang, H., San Clemente, H., Savelli, B., et al. (2015). The *Eucalyptus grandis* R2R3-MYB transcription factor family: evidence for woody growth-related evolution and function. *New Phytol.* 206, 1364–1377. doi: 10. 1111/nph. 13039

Soltis, D. E., Albert, V. A., Leebens-Mack, J., Palmer, J. D., Wing, R. A., Ma, H., et al. (2008). The *Amborella* genome: an evolutionary reference for plant biology. *Genome Biol.* 9: 402.

Sork, V. L., Fitz-Gibbon, S. T., Puiu, D., Crepeau, M., Gugger, P. F., Sherman, R., et al. (2016). First draft assembly and annotation of the genome of a California endemic oak *Quercus lobata* Née (Fagaceae). *G3* 6, 3485–3495. doi: 10. 1534/g3. 116. 030411

Spicer, R., and Groover, A. (2010). The evolution of development of the vascular cambium and secondary growth. *New Phytol.* 186, 577–592.

Spieß, N., Oufir, M., Matušíková, I., Stierschneider, M., Kopecky, D., Homolka, A., et al. (2012). Ecophysiological and transcriptomic responses of oak (*Quercus robur*) to long-term drought exposure and rewatering. *Environ. Exp. Bot.* 77, 117–126. doi: 10. 1016/j. envexpbot. 2011. 11. 010

Spokevicius, A. V., Tibbits, J., Rigault, P., Nolin, M. A., Muller, C., and Merchant, A. (2017). Medium term water deficit elicits distinct transcriptome responses in *Eucalyptus* species of contrasting environmental origin. *BMC Genomics* 18: 284. doi: 10. 1186/s12864-017-3664-z

Staňková, H., Hastie, A. R., Chan, S., Vrána, J., Tulpová, Z., Kubaláková, M., et al. (2016). BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. *Plant Biotechnol. J.* 14, 1523–1531. doi: 10. 1111/pbi. 12513

Staton, M., Zhebentyayeva, T., Olukolu, B., Fang, G. C., Nelson, D., Carlson, J. E., et al. (2015). Substantial genome synteny preservation among woody angiosperm species: comparative genomics of Chinese chestnut (*Castanea mollissima*) and plant reference genomes. *BMC Genomics* 16: 744. doi: 10. 1186/s12864-015-1942-1

Tan, B., Grattapaglia, D., Martins, G. S., Ferreira, K. Z., Sundberg, B., and Ingvarsson, P. K. (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. *BMC Genomics* 17: 110. doi: 10. 1186/s12870-017-1059-6

Tan, B., Grattapaglia, D., Wu, H. X., and Ingvarsson, P. K. (2018). Genomic relationships reveal significant dominance effects for growth in hybrid *Eucalyptus* . *Plant Sci.* 267, 84–93. doi: 10. 1016/j. plantsci. 2017. 11. 011

Tuskan, G. A., DiFazio, S. P., Faivre-Rampant, P., Gaudet, M., Harfouche, A., Jorge, V., et al. (2012). The obscure events contributing to the evolution of an incipient sex chromosome in *Populus* : a retrospective working hypothesis. *Tree Genet. Genomes* 8, 559–571. doi: 10. 1007/s11295-012-0495-6

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604. doi: 10. 1126/science. 1128691

Ueno, S., Klopp, C., Leplè, J. C., Derory, J., Noirot, C., Leger, V., et al. (2013). Transcriptional profiling of bud dormancy and release in oak by next-

generation sequencing. *BMC Genomics* 14: 236. doi: 10. 1186/1471-2164-14-236

Ueno, S., Le Provost, G., Léger, V., Klopp, C., Noirot, C., Frigerio, J. M., et al. (2010). Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. *BMC Genomics* 11: 650. doi: 10. 1186/1471-2164-11-650

VanBuren, R., Bryant, D., Edger, P. P., Tang, H. B., Burgess, D., Challabathula, D., et al. (2015). Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum* . *Nature* 527, 508–511. doi: 10. 1038/nature15714

Vining, K. J., Pomraning, K. R., Wilhelm, L. J., Priest, H. D., Pellegrini, M., Mockler, T. C., et al. (2012). Dynamic DNA cytosine methylation in the *Populus trichocarpa* genome: tissue-level variation and relationship to gene expression. *BMC Genomics* 13: 27. doi: 10. 1186/1471-2164-13-27

Vining, K. J., Romanel, E., Jones, R. C., Klocko, A., Alves-Ferreira, M., Hefer, C. A., et al. (2015). The floral transcriptome of *Eucalyptus grandis* . *New Phytol.* 206, 406–1422. doi: 10. 1111/nph. 13077

Wagner, S., Lagane, F., Seguin-Orlando, A., Schubert, M., Leroy, T., Guichoux, E., et al. (2018). High-throughput DNA sequencing of ancient wood. *Mol. Ecol.* 27, 1138–1154. doi: 10. 1111/mec. 14514

Walbot, V. (1985). On the life strategies of plants and animals. *Trends Genet.* 1, 165–169. doi: 10. 1016/0168-9525(85)90071-X

Wang, N., Thomson, M., Bodles, W. J. A., Crawford, R. M. M., Hunt, H. V., Featherstone, A. W., et al. (2013). Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Mol. Ecol.* 22, 3098–3111.

Weighill, D., Jones, P., Shah, M., Ranjan, P., Muchero, W., Schmutz, J., et al. (2018). Pleiotropic and epistatic network-based discovery: integrated networks for target gene discovery. *Front. Energy Res.* 6: 30. doi: 10. 3389/fenrg. 2018. 00030

Westbrook, J. (2018). *Merging Backcross Breeding and Transgenic Blight Resistance to Accelerate Restoration of the American Chestnut: The American Chestnut Foundation's Breeding and Selection Plan 2015–2025.* Asheville, NC: The American Chestnut Foundation. Available at: www. acf. org