# Current status and trends in forest genomics

Dulal Borthakur[1], Victor Busov[2], Xuan Hieu Cao[3], Qingzhang Du[4], Oliver Gailing[3], Fikret Isik[5], Jae-Heung Ko[6], Chenghao Li[7], Quanzi Li[8], Shihui Niu[4], Guanzheng Qu[7], Thi Ha Giang Vu[3], Xiao-Ru Wang[9], Zhigang Wei[10], Lin Zhang[11], and Hairong Wei[2*]

[1] Dulal Borthakur, Department of Molecular Biosciences and Bioengineering, University of Hawaii at Manoa, 1955 East-West Road, Honolulu, HI 96822, USA
[2] College of Forest Resources and Environmental Science, Michigan Technological University, Houghton, MI 49931, USA
[3] Forest Genetics and Forest Tree Breeding, Faculty for Forest Sciences and Forest Ecology, University of Göttingen, Büsgenweg 2, 37077 Göttingen, Germany
[4] National Engineering Research Center of Tree Breeding and Ecological Restoration, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, P.R. China
[5] Cooperative Tree Improvement Program, North Carolina State University, Raleigh, NC 27695, USA
[6] Department of Plant & Environmental New Resources, Kyung Hee University, 1732 Deogyeong-daero, Yongin 17104, Republic of Korea
[7] State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, Harbin 150040, P.R. China
[8] State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing 100093, P.R. China
[9] Department of Ecology and Environmental Science, Umeå Plant Science Centre, Umeå University, Umeå 90187, Sweden
[10] College of Life Sciences, Heilongjiang University, Harbin 150080, P. R. China
[11] Key Laboratory of Cultivation and Protection for Non-Wood Forest Trees, Ministry of Education, Central South University of Forestry and Technology, Changsha 410004, Hunan Province, P.R. China
All authors contributed equally and are listed in the alphabetical order of the last names with the correspondence as an exception.
* Corresponding author, E-mail: hairong@mtu.edu

## Abstract

Forests are not only the most predominant of the Earth's terrestrial ecosystems, but are also the core supply for essential products for human use. However, global climate change and ongoing population explosion severely threatens the health of the forest ecosystem and aggravtes the deforestation and forest degradation. Forest genomics has great potential of increasing forest productivity and adaptation to the changing climate. In the last two decades, the field of forest genomics has advanced quickly owing to the advent of multiple high-throughput sequencing technologies, single cell RNA-seq, clustered regularly interspaced short palindromic repeats (CRISPR)-mediated genome editing, and spatial transcriptomes, as well as bioinformatics analysis technologies, which have led to the generation of multidimensional, multilayered, and spatiotemporal gene expression data. These technologies, together with basic technologies routinely used in plant biotechnology, enable us to tackle many important or unique issues in forest biology, and provide a panoramic view and an integrative elucidation of molecular regulatory mechanisms underlying phenotypic changes and variations. In this review, we recapitulated the advancement and current status of 12 research branches of forest genomics, and then provided future research directions and focuses for each area. Evidently, a shift from simple biotechnology-based research to advanced and integrative genomics research, and a setup for investigation and interpretation of many spatiotemporal development and differentiation issues in forest genomics have just begun to emerge.

## Introduction

Forests are the predominant terrestrial component on Earth, and account for nearly three quarters of the total primary production of the Earth's biosphere. Global forests generate an estimated 21.9, 8.1 and 2.6 gigatonnes of net biomass per year by tropical, temperate and boreal forests, respectively[1]. Forests can absorb greenhouse gases, mainly carbon dioxide ($CO_2$), and sequester carbon into wood for long-term storage. Therefore, they are essential for the Earth's ecosystem and our living planet by carrying out air purification, ground water recharge, and climate abnormality control. In addition, forests play a leading role in the global cycling of energy, carbon, water and nutrients[2]. However, global climate change, or the long-term shift in temperatures and weather patterns, caused primarily by the planet's overall rising temperature since the pre-industrial period due to the increasing human activities like fossil fuel burning, threatens the health of forests.

Forests are also the core raw sources for many bioproducts, which include timber, pulp, fuel wood, feedstock, and a variety of traditional and novel bioproducts. As the world population is projected to reach 9–10 billion by 2050, demand for forest-based products will continue to increase rapidly, therefore, the pressure to protect forests from deforestation and destruction are mounting up, intensifying the challenge of maintaining the health of the Earth's biosphere.

One of the strategies to counter the aforementioned challenges is to increase forest productivity, adaptation, resilience and sustainability, which requires the development of multi-faced approaches where forest genomics play a key role[3]. Firstly, genomics-based approaches can significantly increase the productivity and adaptation directly through modification of one or multiple genes in tree genomes. Secondly, genomic-based approaches are particularly instrumental for identifying the genetic markers and genes for gene pyramiding in mole-

cular breeding. Therefore, significantly enhancing the efficiency of conventional breeding in targeting the complex and low heritable quantitative traits of interest in forest trees. Thirdly, genomic approaches have great potential to accelerate the generation of system biology knowledge that are essential for determination of the optimal strategies and plans for improving forests. Fourthly, computational genomics provides an alternate approach to identify genes regulating the complex traits, enabling gene stacking and genome editing for developing custom-designed trees for designated uses. Finally, genomics-based approaches can answer many unique biological questions of forest trees, including their unusual reproductive features, woody and perennial growth habit, and their mechanisms of adaptation to abiotic and biotic stress[4].

In this review, we summarized the status, progress and future directions of multiple facets of forest genomics research, and hope this review can aid and foster the developments in genomics to enhance the productivity and sustainability, and accelerate the domestication of forest trees for combating climate change and meeting the increasing demands for forest products caused by the continue explosion of the World's population.

## Wood formation

Environmentally friendly, renewable and sustainable woody biomass is needed to save humanity from the global climate crisis. Woody biomass, or xylem cells with thickened secondary cell walls (SCW), composed of cellulose, hemicellulose (xylan) and lignin (for a recent review, see Schultz et al.[5]), accounts for about 90% of the total biomass produced on Earth[6,7]. Cellulose and xylan are polymers of glucose and xylose, respectively, and lignin is a polymer of phenolic compounds. These three types of polymers can be utilized to produce not only biofuels but also various traditional and new materials, such as, bioplastics, carbon nanomaterials, pharmaceuticals, and various synthetic materials by using appropriate conversion technology (for a recent review, see de Vries et al.[8]). Therefore, the utilization of woody biomass can not only partially replace energy and materials provided by existing fossil fuels, but also reduce atmospheric carbon dioxide to mitigate the global climate crisis.

Wood is formed by complex but highly organized regulatory processes that include vascular cambium cell division and development, and xylem cell differentiation. Extensive studies have been performed to resolve molecular networks controlling vascular cambium development and xylem cell differentiation, mainly using plant species such as *Arabidopsis*, *Populus*, *Eucalyptus*, *Zinnia elegance*, and other species[9–11]. For xylem cell differentiation, vascular-related NAC domain (VND) subfamily proteins, VND6 and VND7, which are identified as master regulators for metaxylem (i.e., SCW deposited with pitted or reticular form) and protoxylem (i.e., SCW with helical form) differentiation, respectively[10], regulate transcription of SCW forming genes (e.g., MYB46 and MYB83, a paralog of MYB46) and programmed cell death (PCD)[12,13]. MYB46 and MYB83, the master regulators of SCW biosynthesis found in *Arabidopsis*, regulate directly/indirectly downstream transcription factors and structural SCW biosynthesis genes in a feed-forward manner through a highly complex and sophisticated regulatory network[12,13]. Recently, it was reported that PtrHB7, a member of HD-ZIP III family in poplar, is involved in the auxin-induced xylem differentiation regulatory network in woody stems[14].

Cylindrical vascular cambium is a secondary meristem containing bifacial stem cells[15,16]. Recently, considerable progress has been made in the molecular understanding of the development of vascular cambium, and it has been revealed that the coordinative regulatory mechanisms, including transcription factors, peptides and hormones, are all required for this process[13,17]. Both auxin and cytokinin (CK) play pivotal roles in regulating the initiation and maintenance of procambial cells. Auxin activates MONOPTEROS (MP), which positively regulates vascular initial cell divisions through auxin efflux carrier *PIN-FORMED1* (*PIN1*) and TARGET OF MONOPTEROS 5 (TMO5)/LONESOME HIGHWAY (LHW) dimer[18–20]. LONELY GUY 4 (LOG4), a rate-limiting enzyme in CK biosynthesis, is directly activated by the TMO5/LHW dimer, and the resulting CK promotes cell divisions through the DOF transcription factor, DOF2.1[19,21]. It has been known that the auxin-signaling IAA12-ARF5/MP module controls *Arabidopsis* provascular specification and patterning[22]. A similar module (PtoIAA9-PtoARF5) was recently found to operate in auxin-triggered xylem cell differentiation in *Populus*[14]. In addition to plant hormones, TRACHEARY ELEMENT DIFFERENTIATION INHIBITORY FACTOR (TDIF)-PHLOEM INTERCALATED WITH XYLEM (PXY)-WUSCHEL-RELATED HOMEOBOX (WOX) signaling module regulates cambial cell division and radial growth in trees (for a recent review see Wang et al.[11]).

Currently, the understanding of the underlying molecular mechanisms by which the procambium develops into the vascular cambium is still limited. In this process, the parenchymatous cells in the interfascicular region must be differentiated into cambium to connect the fascicular cambium. Recently, the HD-ZIP III gene *PtrHB4* was found to be involved in the induction of interfascicular cambial cell division in *Populus* stem[13]. Therefore, further analysis of target genes directly downstream of PtHB4 is expected to provide clues to this process.

Future research for wood formation may be focused on the cambium cell identity maintenance, and development and differentiation process of cambium cells into secondary xylem and phloem by combining sc-RNA-seq and spatiotemporal transcriptome-based technologies, which allows the determination of new cell- and tissue-specific regulators during these processes. Single cell RNA sequencing (scRNA-seq) has recently been used to investigate vascular cell specification and differentiation in trees at single-cell resolution using poplar stems[23,24]. The advent of spatiotemporal transcriptome technologies, for example, stereo-seq using nanoball array[25], will enable the mapping of spatiotemporal transcriptomic dynamics and changes during the formation of secondary xylem and phloem, which will significantly advance our understanding of this important process.

## CRISPR-mediated genome editing provides a powerful tool for forest tree improvement

With its versatility, high efficiency, and robustness, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) biotechnology has emerged as the most widely used genome editing tool. To date, SRISPR-based technologies have been harnessed to introduce precise alterations at the target sites in many given plant species. For recent reviews in forest trees, see[26,27]).

Genome editing and genetic manipulation of forest trees have become leading edge research fields. Several types of the engineered Cas9s have been utilized to establish the CRISPR/Cas9 application in several forest trees[28−34]. Among these, poplar and eucalypt genomes are the most common and successful genomes that have been subjected to genome editing due to well-established *Agrobacterium*-mediated stable transformation and abundant genetic and genomic resources and tools. Nevertheless, there is substantial room for the improvement of specificity and efficiency in CRISPR-edited tree experiments, given that the CRISPR toolbox for plant genome editing has been continuously expanded with unexplored or newly engineered CRISPR effectors (for reviews, see[35−37]). In addition, versatile methods for targeted insertion or replacement of genes have been demonstrated to improve the efficiency of HDR-mediated precise modification, such as tandem repeat-HDR[38] and transcript-templated HDR[39]. Morever, genome editing approaches without introducing DSBs via a variety of CRISPR-mediated base editing[40] and prime editing methods have recently been developed[41]. The first cytosine base editor for conversions of target cytosines to thymines was designed by the fusion of a nickase Cas9 and a cytidine deaminase enzyme[40,42].

In general, CRISPR-mediated genome editing holds great potential for genetic engineering of forest trees, studying and understanding their growth, resilience and adaptation during climate change. Several studies with genetically-modified or CRISPR-mediated improvement of wood quality[43], resistance to pests and diseases[44], tolerance to drought[45], salt and cold stresses[46] have been conducted in poplar, black locust[47], mulberries[48], and upland cotton trees[46] (for recent reviews, see[26,49,50]). We anticipate that many more CRISPR-based applications and studies will be performed in the following areas. Firstly, biological functions of candidate genes from quantitative trait loci (QTL) and association studies or homologous genes of interest from other plant species can be studied in loss-of-function mutants. Secondly, functions of allelic variants even with small effects in association studies can be investigated by precise knock-in CRISPR systems (i.e., HDR-mediated gene replacement, base editing, and prime editing). Thirdly, improvement of quantitative traits can be achieved by simultaneously inserting or exchanging different regulatory elements (i.e., enhancers and promoters) from functionally known genes, modulating the expression levels of genes in the known gene regulatory network as well as fine-tuning biological pathway switches and rate-limiting enzymes. Lastly, gain-of-function CRISPRed trees with better resistance to disease[44] and adaptation enhanced traits can be pursued by CRISPR-based transcriptional activation of endogenous genes. Altogether, the robustness of the CRISPR technology allows researchers in the field of forest genetics and forest tree breeding to take advantage of innovative ideas from all plant genome editing projects such as climate-resilient crops and fruit trees. With the significant and rapid progress in CRISPR technologies, a new green revolution in forest tree breeding and conservation with more climate resilient forests might become reality in the near future.

More recently, new precise DNA base substitutions have been established (for a review, see[51]), including A−G base transition[52], C−A[53], and C−G transversion[53−55]. Importantly, using a nickase Cas9 fused to a reverse-transcriptase (RT), the prime-editing tools[41] can mediate desired combinations of deletions, insertions, and base-to-base conversions to replace the target sequence. The RT is guided by a prime-editing guide RNA (pegRNA) that is made up of an extended gRNA specifying the target site and a RT template sequence containing desired edits. First successful demonstrations of base editing and prime editing in plant models offer great potential to solve practical problems for tree breeding by introducing novel quantitative traits with a gain-of-function mutation[56,57]. However, great challenges remain because many attractive target traits for tree improvement including tree growth, wood properties, stress and disease resistance are highly polygenic and under complex regulation networks. CRISPR-based gene manipulations can possibly generate unintended consequences due to, for example, imbalances in the gene interaction networks. These limitations might be mitigated by CRISPR-based pyramiding of multiple monogenic traits or polygenic traits, taking advantage of the robustness, precision and multiplexity of the CRISPR technology in combination with genetic crosses and marker-assisted selection. Nevertheless, a wider application of CRISPR-mediated gene editing in forestry requests a significant improvement of the low efficiency of the CRISPR-based editing technologies and to overcome the recalcitrance in transformation and regeneration of many tree species.

## Single-cell RNA-sequencing technology and implication in forest genomics

The development of multicellular plants involves the formation of organs that are comprised of different types of cells with high heterogeneity. During organ formation, cell fate determination and differentiation are precisely controlled by successive transcriptional regulations. However, the cell fate determination from the cambium to the secondary xylem or secondary phloem also remains elusive. Since the existing transcriptome analysis (RNA-seq) can only obtain information on a whole tissue, it is difficult to distinguish and analyze different cell types. With the successful isolation of single cells, scRNA-seq has become a powerful tool to study the gene expressions for individual cells among a heterogenous tissue. Because it can classify cells into different groups based on their types and quantify cell type-specific expression, and also enables cell trajectory analysis for cell differentiation and development. Thus, by reconstructing the spatiotemporal relationship between the cambium and surrounding cells, it is possible to track the progressive cell fate changes of secondary xylem and phloem, respectively, and to uncover new key cell type-specific regulators.

Single cells can be separated by using limiting dilution, micromanipulation, laser capture microdissection, flow cytometry and microfluidics[58]. In existing scRNA-seq experimental procedures, thousands of individual cells, upon being isolated with droplet-based microfluidics[59], can be barcoded to distinguish transcripts from different cells, allowing a high throughput gene expression profiling at single-cell resolution[60−62]. Based on the differential gene expression patterns, these cells can be grouped into different clusters, which can be classified to specific cell types. The analysis of cell clusters enables to understand the cell heterogeneity and provides additional markers for each cell type[60]. Since the cells that are undergoing the transition from one to another state could be

captured, novel cell types may be uncovered, and using pseudo-time analysis to make the order between single cells based on the similarity of gene expression patterns could deduce the developmental trajectories of the clusters (cell types). New regulators controlling developmental transition can be identified, improving the investigation of dynamic developmental processes.

Microfluidic-based scRNA-seq was first applied in *Arabidopsis* roots and showed the feasibility of high throughput scRNA-seq in plants[63]. Additional scRNA-seq studies in *Arabidopsis* roots used either drop-seq[64–67] or 10X genomics methods[68] to provide detailed spatiotemporal information for different cell types, including the quiescent center that has stem cells, and reconstruct the continuous differentiation trajectory of root epidermal, endodermical root hair cells, and other cell types. scRNA-seq was also applied to rice roots[69,70] and other tissues of plants, including shoot tips, lateral roots, gametophytes, anthers, etc.[71–81]. Compared to these model plant species, the application of scRNA-seq in trees is limited. The first scRNA-seq was performed in the differentiating xylem of *Populus alba* × *Populus glandulosa*[24]. This study profiled 9798 cells and identified 12 cell clusters, encompassing vessel cells, fiber cells, ray parenchyma cells, and xylem precursor cells. Further pseudo-time analysis revealed the differentiating trajectory of fiber cells, ray parenchyma cells and vessels. Chen et al.[23] performed scRNA-seq on protoplasts harvested separately from *P. alba* var. *pyramidalis* stem bark and wood. This study also identified the marker genes for the phloem, and reconstructed the cell differentiation trajectories for phloem and xylem development from cambium. The scRNA-seq technique was very recently used to investigate vascular cell specification and differentiation in trees at single-cell resolution using poplar stems[23,24]. Compared to spatial transcriptome technology like Stereo-seq, the information we can obtain from scRNA-seq is still very limited.

Most of the current studies in plants used protoplasts for scRNA-seq, which relies on successful protoplast preparation. It is challenging to isolate high quality protoplasts from the cells that are located in the inner tissue of an organ and have thick cell walls, especially in trees. High protoplast preparation efficiency needs to be improved for some specific tissues and in most species. Longer incubation times of enzymatic digestions to remove cell walls may lead to certain changes in transcriptional activity[82]. Single-nucleus RNA sequencing (snRNA-seq) can avoid the protoplasting effect, and snRNA-seq protocols have been established in *Arabidopsis* and rice[83–88], and recently in poplar[89]. Although it was observed that expression of fewer genes was captured per cell by snRNA-seq than scRNA-seq, it is still worth to expand snRNA-seq for other tissues. The scRNA-seq data can be integrated with other transcriptome data to explore key regulators and novel mechanisms. For example, the integration analysis of snRNA-seq and snATAC-seq elucidated cell-type-specific patterns of chromatin accessibility for the cell-type-specific markers, showing that chromatin accessibility can be used as molecular markers to indicate root hair and endodermal cell developmental states[83]. Through analysis of single cell transcriptome overexpression of *VND7*, *MYB46* and *MYB83*, Turco et al. identified that other four targets of VND7, but not MYB46 and MYB83, were involved in VND7-mediated switch of root cells to xylem cell identity[90]. Further wide applications of scRNA-seq in

mutants or transgenics will enhance the understanding on the transcriptional regulation of a gene of interest. However, scRNA-seq lacks spatial information, which can be overcome by combination with the spatial transcriptome[91,92], and this combination has not been reported in plants. Overall, increasing number of scRNA-seq studies have been performed in plants and provide insights into the transcriptional regulation of cellular state during organ formation. It is anticipated that other analyses, including chromatin immunoprecipitation (ChIP)-seq, DNA methylation, protein-protein interactions at single-cell level, will also be achieved successfully in plants in the near future. The limiting factors for implementation of scRNA-seq to forest trees include the challenges in target tissue isolation and removal of cell walls. Compared to other plant species, both target tissue isolation and cell wall removal in woody plants are more challenging due to rigidness, and lignified and thickened cell walls.

## Haploid induction and breeding

High-levels of heterozygosity are a significant obstacle for not only the genome sequencing and assembly but also tree breeding through inbred lines, which can be used to generate heterosis through crossing. Although the endosperms originated from female gametophyte tissue (n) in gymnosperms seeds and some angiosperm seeds are haploid, the haploid tissue in a single seed is generally insufficient for extracting enough DNA for haploid genome sequencing. Generating inbred lines for creating heterosis through cross-hybridization takes at least six generations to achieve approximate complete homozygosity in most agricultural crops. This can take even more generations and longer time for tree species due to their higher levels of heterozygosity, long juvenile phases, and long-generation cycles as well as self-incompatibility. Therefore, it is practically impossible to implement inbreeding improvement strategy in trees in the conventional manner. A strategy to overcome these obstacles is to obtain haploid plants, which can not only facilitate genome sequencing and assembly but also generate homozygosity through chromosomal doubling. Doubled haploid (DH) lines are even better than inbred lines, and thus can be of great value for tree breeding, genome sequencing and assembly, and also other genome-based research.

For agricultural crops, haploids can be induced in natural conditions or artificially by a physical or chemical treatment. Spontaneous haploids occurred in many species including maize, cotton, rice tomato, barley, and brassica[93]. However, the induction rates are very low in many species, and as a result, selection of spontaneous haploids from natural populations is not cost-effective. Subsequently, numerous endeavors have been made to increase the frequency of artificial haploid induction. It has been found that a specific maternal or paternal parent can give rise to a higher induction frequency. For example, a maize inbred designated as 'stock 6' could generate 3.23% haploids in 10,616 progenies[94]. Hence, selection of parents that have higher haploid induction frequency was effective and accelerated maize inbreed line generation. To date, various techniques have been employed in different plant species, some of which have been reviewed in several reviews[95–97]. One of these technique is distant pollination, namely wide crossing, which is one of the most effective

methods for female-derived haploid production. For example, haploid plants were obtained from *Lactuca sativa* pollinated with the pollen from distant species like *Helianthus annuus* and *H. tuberosus*[98]. The pollen tubes of the distant species may not release the sperm cells into the ovaries but can stimulate the development of ovaries into haploid seeds. In addition, the female-derived haploid induction by pollination with the pollen pretreated with various physical or chemical agents has also been used in many species. For example, haploid plants were obtained by pollination with irradiated pollen in melon[99,100]. In addition, gynogenesis induction by direct *in vitro* culture of unfertilized ovaries or ovules is another viable approach for haploid production for some species, such as cucumber and red beet[101,102]. Moreover, androgenesis induction by anther culture is one of the more frequently used methods in various plant species, such as wheat[103], cucurbita[104]; *Cannabis sativa*[105]. Those methods that were successfully used in agricultural crops need to be tested for trees which are mostly self-incompatible.

To date, some aforementioned approaches have been adapted or modified to produce haploids in forest trees. For example, cross-pollination with stress-treated pollen that stimulate parthenogenesis for inducing haploid trees; the pollen grains are often pretreated with physical stresses like high temperature[106], heat[107], irradiation[108], or chemical like toluidine blue[109]. In addition, the unpollinated ovaries were reported to develop into haploid poplar trees[110], which indicates that the female gametophyte and megaspore of angiosperms can be induced *in vitro* for sporophytic development, thereby creating a new avenue to haploid breeding. Moreover, the endosperms of most gymnosperms comprise haploid cells developed directly from the female gametophyte, and thus can be used to induce haploids directly. For example, haploid calli have been generated by culturing *Taxus chinensis* endosperm[111]. Haploid plants of *Eucommia ulmoides* were induced *via* parthenogenesis with heat treatment on female flower buds during the developmental stage of embryo sac formation[112]. Anther or pollen culture coupled with plant hormone treatments under tissue culture conditions was the most commonly used method in tree species[113–116]. In contrast to a wide variety of crop species that have been subjected to haploid induction, much fewer haploid tree species have been generated *in vitro* because of their recalcitrance to tissue culture and regeneration. Furthermore, the DH trees generated from haploids usually show poor vitality and can hardly survive for more than a few years until they reach flowering age. To date, no DH population of tree species has been reported to be established due to poor adaptation caused by high homozygosity. In the future, more efforts should be made to increase induction efficiency, establish effective procedures that can facilitate acquisition of a large number of DH lines, and increase their survival under greenhouse and field conditions. In recent years, a CENH3-mediated haploid induction system has been established, which has the potential to be extended to tree species since CENH3 is universally present in eukaryote species[117–119].

## Recalcitrance to genetic transformation and regeneration in trees

Generation of transgenic trees with new or improved desirable traits relies on genetic transformation and subsequent regeneration. Transformation of foreign genes for molecular breeding or gene/genome editing is the key to successful tree trait improvement. Various transformation protocols including *Agrobacterium*-mediated transformation, Polyethylene glycol (PEG)-mediated direct transfer and biolistic bombardment have been applied for plants[120] . Among the most widely used is *Agrobacterium*-mediated T-DNA transformation which, in general, has the highest transformation efficiency[121]. Subsequent regeneration is an essential process to recover transgenic plants. Callus induction, *de novo* organogenesis, and somatic embryogenesis of conventional tissue culture-based regeneration protocols have been widely developed in trees[122]. The earliest reports on transgenic poplar and European larch trees are dated 1987[123] and 1991[124], respectively. Although genetic transformation and regeneration have been widely applied to many other tree species, the practice for genetic transformation and regeneration in trees is still largely restricted to model forest (poplars and *Eucalyptus*)[125] and fruit trees[121,126], and has been rarely applied to industrially important trees such as rubber trees[127] and conifers[128] because of the recalcitrance of transformation and regeneration. Thus far, the following efforts have been made to overcome recalcitrance in genetic transformation and regeneration practices.

The most widely employed approach to overcome recalcitrance is the selection of most amenable genotypes and tissues such as juvenile leaves, petioles, cotyledons[129–133] and the use of phytohormones[134,135] such as auxin and cytokinin. Although these approaches work for some species, they have different effects in different species and have little general applicability. For example, for Amur cork trees[136–138] and willow trees[139,140], even the juvenile leaves and stem explants are recalcitrant to transformation and shoot regeneration, the shoot apical meristems germinated from the mature seeds exhibit amenability to transformation and subsequent regeneration. In conifers, somatic embryogenesis from the immature zygotic embryos is the only way to regenerate plantlets[141,142]. The amenability of explants to genetic transformation and regeneration differs among explant types (juvenile or mature) and genotypes; regeneration capacity of explants generally declines as age increases. Different genotypes have widely exhibited distinct regenerative responses even in the same species. Loss of phytohormone responsiveness is a putative cause for declined regenerative capacity. As plants get older, the developmental stages and differentiation as well as epigenetic modifications deactivate phytohormone responsiveness, thereby resulting in different degrees of recalcitrancy[143]. Some genes that control transformation amenability/regeneration have been identified, e.g. *Baby boom* (*BBM*), *WUSCHEL*, *SHOOT MERISTEMLESS* (*STM*), *WOX*[144,145], and the synthetic gene for the GROWTH-REGULATING FACTOR (GRF)- GRF-INTERACTING FACTOR (GIF) chimeric protein[146], and these genes have been used to improve regeneration potential. Co/pre-transformation has also been reported to boost genetic transformation and regeneration in trees[147,148].

In general, phytohormones including auxin, cytokinin, and abscisic acid trigger signaling events for plant regeneration which then relieve epigenetic constraints and activates the above developmental regulator genes to initiate cell-fate transitions (dedifferentiation) and expression of downstream genes for plant development - usually the hormone biosynthesis and signaling genes (redifferentiation)[143,149,150]. GRF-GIF

chimeric protein has recently been highlighted for its capacity to remove epigenetic regulatory barriers against plant regeneration. The expression of developmental genes driving the cell transition processes for dedifferentiation and redifferentiation during plant regeneration are strictly regulated by a closed chromatin state that can be opened by epigenetic factors such as SWITCH/SUCROSE NONFERMENTING (SWI/SNF) complexes. GRF and its cofactor GIF form chimera which then can recruit SWI/SNF complexes to regeneration-related genes to remodeling chromatin structures and states, relieving the constraints for the transcription of those genes to promote regeneration[151].

Owing to the above efforts, genetic transformation and regeneration have significantly improved; however, the following questions on recalcitrance remain to be answered:

(1) What are the upstream regulatory genes for the developmental genes that control transformation amenability/regeneration including *BBM*, *WUSCHEL*, *STM*, and *WOX,* and how are the developmental genes upon a phytohormone treatment triggered and regulated?

(2) What factors cause loss of phytohormone responsiveness during aging and how to revert the process?

(3) What are the key factors that determine embryogenic competence during somatic embryogenesis?

(4) How are the external signals such as stresses and phytohormones transduced to initiate upregulation of those factors during somatic embryogenesis?

(5) How does the aging or subculture affect embryogenic competence of explants?

Providing answers to these questions will significantly advance our understanding of transformation and regeneration. Moreover, conventional regeneration protocols based on tissue culture are laborious and time-consuming, especially for trees which generally have a longer generation time than crops. Some scientists have tried to bypass the regeneration process to circumvent regeneration-recalcitrant issues. For example, tissue culture-free transformation systems have been established in *Arabidopsis* and tobacco[152,153]. However, these novel protocols cannot yet completely bypass the tissue culture in the whole transformation process, and are not yet been applied to tree species. Overall, the efforts to overcome the recalcitrancy issue have greatly advanced the genetic transformation and regeneration in trees, but there is still a long way to go. There is little evidence that substantiates the molecular mechanism underlying recalcitrance and further work should be continued to obtain in-depth insights into the process. It will be of great significance in gene transformation and regeneration practices in trees.

## Nitrogen fixing: A special feature of *Leucaena*-Rhizobium symbiosis

Legume-*Rhizobium* symbiosis is a beneficial mutualistic interaction between legume hosts and compatible rhizobia that exhibit species specificity. Most current knowledge of legume-*Rhizobium* symbiosis has been derived from studies of nodulation and nitrogen fixation in field legumes, such as alfalfa, peas, beans, and soybean, and model legume species, such as *Lotus japonicus* and *Medicago truncatula*, through their interactions with compatible rhizobia[154]. The knowledge of nitrogen-fixing symbiosis between tree legumes and rhizobia is

relatively scarce. Moreover, some tree legumes have uncommon features, for example, *Acacia koa*, a tree legume endemic to the Hawaiian Islands, produces both root and canopy nodulation in symbiosis with *Bradyrhizobium*[155]. Symbiotic nitrogen fixation in the tree-legume *Leucaena* in association with specific fast-growing rhizobia has an additional feature. Leucaena produces a toxic compound, mimosine, which is present in all parts of the plant, including roots and root nodules. Mimosine is also secreted to the rhizosphere in the root exudate[156]. Rhizobia in the rhizosphere or in the nodule have to overcome mimosine toxicity. Rhizobia that effectively nodulate *Leucaena* degrade mimosine in both rhizosphere and inside root nodules.

The genes for mimosine degradation have been well-characterized in *Rhizobium* sp. strain TAL1145 that effectively nodulates *Leucaena*. Mimosine degradation by *Rhizobium* takes place in two steps. First, mimosine is degraded to ammonia, pyruvate and 3-hydroxy-4-pyridone (3H4P) by a C–N lyase, called 'rhizomimosinase' encoded by the *Rhizobium midD* gene[157,158]. 3H4P is further degraded to ammonia, pyruvate, and formate by a dioxygenase and a hydrolase encoded by the *Rhizobium pydA* and *pydB* genes, respectively[159]. The *midD* gene for *rhizomimosinase* is a part of the *midABCD* operon, where *midABC* encodes three constituent proteins of ABC transporters: a periplasmic mimosine-binding protein, a permease, and the ATP-binding protein, for transporting mimosine into *Rhizobium* cytoplasm. By constructing several *midA::gus*, *midC::gus* and *midD::gus* fusions, it was demonstrated that the *midABCD* operon is expressed in *Rhizobium* in the nodule and also inducible by mimosine[157]. Similarly, by constructing *midA::phoA* insertion mutant derivatives of TAL1145, it was shown that the *midABCD* operon expressed in *Rhizobium* bacteroids inside the *Leucaena* root nodules (Fig. 1). The expression of the *midABCD* operon is regulated by a positive regulatory protein encoded by *midR*, which has high sequence similarity with the LysR family of positive regulators. In the rhizosphere, some amounts of mimosine may be degraded to 3H4P, which can be also taken up by *Rhizobium* using a specific ABC transporter encoded by *pydC*, *pydD*, and *pydE* genes. The *pydA*, *pydB*, *pydC*, *pydD*, and *pydE* genes are inducible by mimosine, 3H4P, and several analogs of 3H4P[6].

The ability to degrade mimosine is not essential for forming nitrogen-fixing nodules on *Leucaena*. However, mimosine-degrading ability provides a competitive advantage for nodule occupancy and efficient nitrogen fixation in *Leucaena* nodules[160]. Mimosine is a bacteriostatic compound and the mimosine-non-degrading *Rhizobium* strains, such as CTAT899, do not have a mimosine-specific ABC transporter. Therefore, these rhizobia can survive and grow in the *Leucaena* rhizosphere or inside root-nodules even in the presence of mimosine. Mimosine chelates $Fe^{3+}$ ions in the rhizosphere by forming $Fe^{3+}$-mimosine complexes, thereby depriving these rhizobia of $Fe^{3+}$ and reducing their growth. On the other hand, mimosine-degrading strains, such as TAL1145, can uptake $Fe^{3+}$-mimosine complexes and utilize them as a source of carbon and nitrogen. Therefore, they can grow in the presence of mimosine and occupy nodules more effectively. In the absence of mimosine-degrading rhizobia, non-degrading rhizobia can occupy *Leucaena* nodules. About 40% of *Rhizobium* isolates obtained from the nodules of *Leucaena* growing in several locations on the island of Oahu contained mimosine-degrading
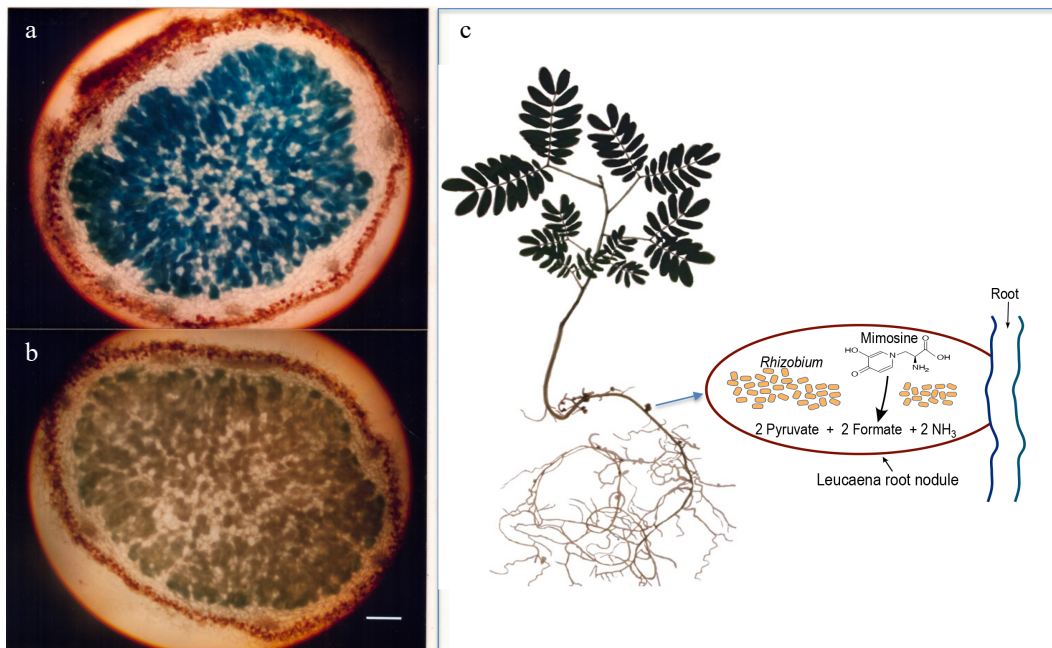
**Fig. 1** Light microscopy of 20-μM thick sections of 5-week-old *Leucaena leucocephala* nodules. (a) Induction of alkaline phosphatase activity in the *Leucaena* nodule formed by the *midA::phoA* mutant RUH128 of *Rhizobium* TAL1145. (b) Another mutant RUH129, in which the *phoA* insertion on *midA* was in the opposite orientation, was used as a negative control. Nodule sections were stained for phosphatase activity. RUH128 cells inside the nodule expressed phosphatase activity whereas RUH129 cells inside the nodule did not show any detectable phosphatase activity. Bar represents 400 μM. (c) Schematic of a *Leucaena* nodule occupied by a *Rhizobium* strain that degrades mimosine into pyruvate, formate and ammonia.

rhizobia[161]. Our current understanding of the role of mimosine-degrading free-living rhizobia in the *Leucaena* rhizosphere is depicted in Fig. 2. Leucaena can grow successfully even in alkaline soils where iron forms insoluble hydroxides, which cannot be taken up by plants. Mimosine, secreted by *Leucaena* roots, has high binding affinity for $Fe^{3+}$ at alkaline pH, where it forms $Fe^{3+}$-mimosine complexes that are water soluble at alkaline pH. Depending on the availability of mimosine in the soil, three types of $Fe^{3+}$-mimosine complexes are possible, where the ratios of $Fe^{3+}$ and mimosine in the complexes may be 1:1, 1:2, or 1:3[3] (see the box in Fig. 2). The water-soluble $Fe^{3+}$-mimosine complexes are taken up by *Leucaena* roots through transporter proteins belonging to the oligopeptide transporter family, which includes YSL transporters. Such an iron uptake system using phytosiderophore and oligopeptide transporters is common in grasses and is known as strategy II of iron uptake. Leucaena also employs strategy I of iron uptake that involves a membrane-bound ferric chelate reductase for converting $Fe^{3+}$ to $Fe^{2+}$, which is then taken up by the plant through an iron-regulated transporter (IRT). Free-living rhizobia inhabiting in the *Leucaena* rhizosphere may also uptake $Fe^{3+}$-mimosine complexes and degrade mimosine, releasing $Fe^{2+}$ to the rhizosphere, where it is taken up by the plant using an IRT transporter (Fig. 2). Thus, besides providing a selective advantage to occupy *Leucaena* root nodules, such mimosine-degrading rhizobia inhabiting in close proximity of the Leucaena's root system, enhance iron availability in the *Leucaena* rhizosphere. Future research will apply genome editing technology to develop new varieties of *Leucaena* for regulated expression or inhibition of the mimosine biosynthesis genes.

## Genome assembly and evolution

The publication of the genome sequence of the first tree species *Populus trichocarpa* in 2006[162] marked the beginning of the genomic era of tree species. In the first decade after the poplar genome was released, a total of 47 tree species were sequenced. With the technical advantages brought by more powerful DNA sequencing technology, draft or reference genomes of 357 tree species, including 266 arboreal, 71 shrub and 20 vine species, were assembled by the end of February 2022, which accounts for about 36% of all sequenced plants.

The advantage of third-generation long-read sequencing technology is the key driving force behind the substantial improvement of genome assembly continuity in recent years. The contig N50 of published plant genome was increased from $99.5 \pm 48.1$ kb in 2010 to $3,395.2 \pm 735.4$ kb in 2020[163]. The Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) are the two mainly long-read sequencing platforms used in the tree genome project, and to date, the adoption ratio of PacBio to ONT is five to one. The lower sequencing error rate of PacBio is the main reason for its preference, especially its recently updated circular consensus sequencing (CCS) mode which can generate highly accurate long high-fidelity (HiFi) reads[164]. In comparison with continuous long-read (CLR) mode, the CCS-based assembly is more efficient and time-saving. It is reported that a 27 Gb-sized hexaploidy coast redwood (*Sequoia sempervirens*) has been sequenced and assembled in only two weeks with a high contig N50 value of 1.92 Mb[165]. Due to its advantages, CCS will be more widely applied for tree genome sequencing, especially for giant genomes and genome phasing.
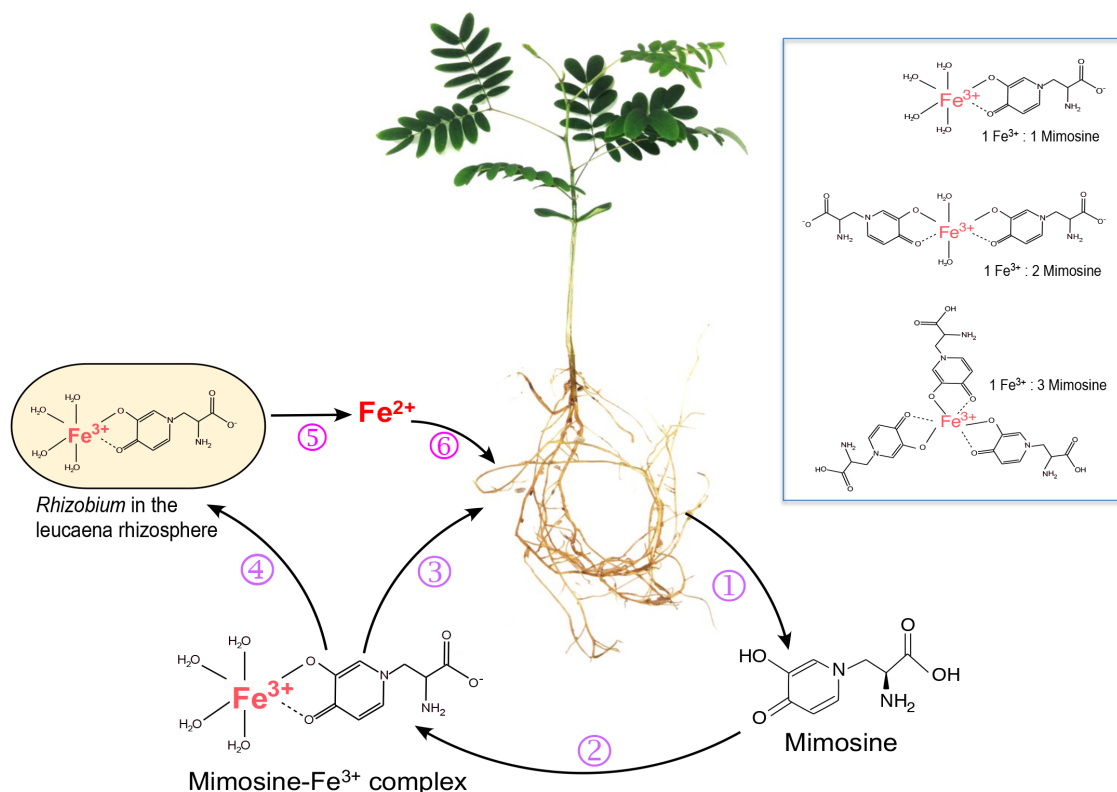
**Fig. 2** Roles of mimosine and free-living rhizobia in iron uptake by *Leucaena*. (1) Leucaena root exudates contain mimosine, which chelates $Fe^{3+}$ to form $Fe^{3+}$-mimosine complexes (2). Three types of $Fe^{3+}$-mimosine complexes may be formed depending on the amount of mimosine available in the rhizosphere (box). The $Fe^{3+}$-mimosine complexes are taken up by *Leucaena* roots using an oligopeptide-type of transporter (3). $Fe^{3+}$-mimosine complexes can also be taken up by mimosine-degrading rhizobia in the rhizosphere using an ABC transporter (4). Rhizobia degrade $Fe^{3+}$-mimosine complexes and release $Fe^{2+}$ into the rhizosphere (5). $Fe^{2+}$ is taken up by *Leucaena* through an IRT transporter (6).

Furthermore, cost and time-effective methods as high throughput chromatin conformation capture (Hi-C)[166] and BioNano optical mapping[167] have greatly accelerated the scaffolding of contigs to generate chromosome-level genome assemblies, especially for forest trees in which building a high-density genetic map is generally very time-consuming and labor-intensive.

With these latest technological advances, we now have the ability to assemble near-complete genomes for any tree species, including conifers which account for 39% of the world's forests but with giant genomes. Recently, based on PacBio and Hi-C technology, the 25.4 Gb chromosome-level assembly of Chinese pine (*Pinus tabuliformis*) was published[168] which is 60 times larger than that of the first sequenced *P. trichocarpa* genome and represents a new milestone in the development of forest tree genomics.

Despite the recent progress, forest trees comprise an estimated ~73,000 species globally[169]. Thus the decoded genomes of 357 tree species merely represent a very small portion of diverse forest tree species. Moreover, there are still great challenges for the haplotype-solved assembly of complex genomes, especially for autopolyploid genomes owing to the high similarity of their subgenomes[170]. Although several freely available pipelines including TrioCanu[171], ALLHiC[172], and FALCON-Phase[173] have been developed for plant genomes assembled as chromosome-scale haplotypes, only four autoploid genomes, i.e., the sugarcane *Saccharum spontaneum* genomes (2n = 8x = 64, 2n = 4x = 40)[174], the cultivated alfalfa

genome (2n = 4x = 32)[170], and tetraploid potato[175] were *de novo* assembled at the chromosome level.

In addition to genome sequencing and assembly, the annotation of gene space has also received more attention. Recent RNA-seq studies hint that the transcriptomes are often substantially underestimated, even in the extensively studied important model organisms like Arabidopsis[176] and rat (*Rattus norvegicus*)[177]. Large-scale RNA-seq data of at least a few hundred samples of different tissues conditions can provide direct transcript resources and enable an unbiased gene space annotation with high resolution[168]. However, obtaining diverse samples including different tissues and various induction conditions for RNA-seq analysis remains a labor-intensive challenge.

As more complete and accurate reference genomes become available, comparative genomics is rising. Due to newly developed sequencing technologies and updated assembly tools for producing longer reads and reducing assembly errors, many previous evolutionary findings are also worth re-examining. As all earlier draft released gymnosperm genomes reported a much older LTR outbreaks time which were later found to be overestimated by fragmented and less accurate assembly[178]. In addition, as more genomes of related species become available, pangenome-oriented studies are imminent.

## Perennial growth and seasonality regulation

Boreal and temperate climates are characterized with annual alterations of seasons that are favorable (spring and summer)

and unfavorable or prohibitive (fall and winter) to growth. To survive the freezing and dehydration stress during fall/winter months, woody perennials from temperate latitudes temporarily suspend growth and protect the shoot apical meristem and the subtending leaf primordia in a specialized organ, known as the bud. The morphological, physiological and developmental processes taking place during this period are known collectively as dormancy.

Significant progress has been made in understanding the different stages of dormancy including the transitions from, and to, active growth. Here we focus only on the genomics aspects and prospects in forest trees. The molecular mechanisms regulating dormancy in fruit trees and vines, have been well reviewed[179–187].

Although genomics attempts to study dormancy-related traits in forest trees were initiated *via* sequencing expressed sequence libraries (EST)[188], it burgeoned only after the sequencing of the first tree genome in 2006[162] and subsequently other forest tree genomes[189–193] when these studies were scaled to the genome-wide level. Most prominently, availability of genome sequences, enabled for the first-time, a glimpse at the gene level landscapes of the QTL associated with dormancy[194]. Candidate genes found within the intervals were consistent with the molecular studies and pointing to genes encoding light receptors, circadian clock oscillators and regulators of the flowering time genes. In addition to QTL methods, the progress in genome sequencing spurred development of entirely new approaches known as association genetics and bringing the mapping resolution to the nucleotide level (e.g., single nucleotide polymorphism (SNP))[195]. Similar to the QTL studies, these approaches brought validation to the molecular studies and also identified new regulators of unknown functions[195].

Perhaps the most wide-spread application of genomics approaches to dissection of dormancy traits is the characterization of transcriptomes associated with different stages. These were first done using microarrays and more recently by employing the more advanced RNA-seq technology[196–204]. Because of the advanced genomics tools and early adoption as a model for forest tree biology, many of the early and major transcriptomic characterization of dormancy were undertaken in poplar[196,201,205,206]. These studies have been expanded into many forest trees including angiosperms and gymnosperms and in different continents including North America, Europe and Asia[196–204]. These studies, when focusing on transcription factors, have employed chromatin immunoprecipitation (ChIP) to identify the direct targets among the many differentially expressed genes[207,208]. This has allowed insights into the hierarchical structure of the underlying networks and has led to discovery of new key regulatory factors. For example, transcription profiling of transgene-modified for the Early Bud Break 1 (EBB1) transcription factor of the AP2/ERF family has led to the discovery of Short Vegetative Phase-like (SVL) transcription factor[206]. This was later found to be another key regulator of bud-break and other dormancy stages[207,208]. However, although the above referenced studies only employed ChIP, the use of ChIP-seq should further expand the horizon of studies in this area and realize greater potential.

Role of epigenetics in regulation of dormancy traits is often speculated but rarely studied. Advances of epigenomics technologies could potentially address this void, including characterization of DNA methylation *via* bisulfide sequencing,

posttranslational modifications of histones using ChIP-seq and sequencing of small/non-coding (nc) RNA libraries. However, these approaches have been rarely applied to dormancy traits. A notable exception is a study looking at the role of the DEMETER DNA demethylase on the genome-wide methylation pattern during the acquisition of growth competence after chilling and its correlation with transcriptomic changes[209]. This study pointed to key genes and processes affected during reactivation of growth including key regulators of meristem activity.

Methylation and acetylation marks at specific amino acids of the histone proteins are important cues regulating gene expression, and their effects can be studied via ChIP-seq approaches employing antibodies binding specifically to these modifications. Although the histone marks at individual gene loci in forest trees and mostly fruit trees have been studied[210,211], a genome-wide map through different stages is still unavailable. Such maps combined with studies on methylation patterns as referred above can provide significant insights into the role of chromatin modification in regulating dormancy.

Small RNAs and ncRNAs have been recently linked to regulation of many processes, including vernalization[212], a process suggested to be regulated in a similar fashion as dormancy[182]. Roles of these regulatory RNAs and their links to chromatin regulation are severely understudied in forest trees. miRNAs, a class of small regulatory RNAs have been linked to the 'memory' spruce embryos retain about the temperatures they experience during their development[213]. The temperatures at the time of embryo development can change phenological traits in one generation and these changes are stable over many years. For example, high temperatures during embryo development lead to later growth cessation and cold acclimation. It was speculated that epigenetic mechanisms are responsible and allow a high level of plasticity and adaptation that does not require lengthy cycles of natural selection. Sequencing of microRNA libraries have found correlation of the abundance of these molecules with different stages of this 'memory' acquisition[214–216].

As genomics technologies become more advanced and affordable, there will be an even further increase in amount of genomics data. Although very useful, this data at the moment is largely descriptive. Development of advanced machine learning and artificial intelligence algorithms that can take advantage of this large data and provide insights into the causality, hierarchy and the regulatory landscape of the underlying mechanisms are needed.

Application and integration of multi-genomics and multi-omics approaches are required to provide system level knowledge of the underlying processes, however, these studies are still limited[201]. Conde et al.[209] compared the genes with differential methylation pattern and changed expression to the gene set that was found with polymorphisms linked to bud-break[195]. Incorporation of such systems-level approaches spanning different regulatory and omics levels can bring novel insights and applications.

## Quantitative trait loci and association studies in forest trees

In order to design optimal early selection methods and breeding strategies, it is imperative to dissect the genetic basis

and identify favorable alleles that underlie economically important and ecologically relevant traits. Most forest trees, unlike annual plants such as *Arabidopsis* and rice that would allow the use of reverse-genetic approaches to detect underlying genes, lack visible large collections of mutations of quantitatively inherited traits. Therefore, forward-genetic approaches such as QTLs and linkage disequilibrium (LD)-based association studies (AS) have been developed for identifying and dissecting quantitative traits in forest trees[217]. QTLs and AS have been broadly performed on traits of interest in numerous tree species, including growth and yield, wood properties, resistance to biotic and tolerance to abiotic stresses, and adaptive traits[217,218], to promote the progress of molecular marker-assisted selection (MAS) breeding in trees[219].

Due to the long juvenile periods and high heterozygosity of most forest trees, QTL mapping usually employs $F_1$ full- or half-sib progenies. The availability of two-way pseudo-testcross strategy has promoted the construction of linkage maps in major timber species, but the limited resolution of these maps may not warrant a successful positional cloning of QTLs. Therefore, the use of high-throughput sequencing technologies can improve the accuracy and resolution of genetic maps, and allows the validation of the quality of scaffold anchoring and whole-genome assembly in trees[220]. Researchers developed several practical mapping algorithms to improve the speed and reliability of genetic distance estimation, which might alleviate the computing complexity and burden caused by map marker redundancy[221,222]. Currently, the bin mapping strategy[223] has become the method of choice to construct dense genetic linkage maps.

QTL mapping has been conducted on a large scale in all major tree species groups for more than two decades. The classic strategies for QTL analyses include bulk segregation analysis (BSA), interval mapping and multiple interval mapping[224,225]. Currently, ultrahigh-density genetic maps based QTLs mapping in combination with systems genetic methods has provided the possibility of fine mapping and identifying genes in trees[226,227]. For instance, an integrated linkage-LD mapping was developed to improve the resolution and effect sizes of dynamic QTLs during stem growth in *P. tomentosa*[228,229]. The mapping noise caused by heterozygous genetic background can be greatly reduced by constructing haplotype modules based on linkage information and employing haplotype-based AS in QTL intervals[219]. In addition, genetic dissection of the segregation distortion of allelic variations is regarded as a key question of fine mapping and will likely be used in breeding in forest trees.

Forest trees are mostly undomesticated populations with wide geographical distribution, high maintenance of genetic variation, and low population differentiation, and thus are ideal systems for conducting AS and MAS breeding. Since Porth et al. conducted the first genome-wide association study (GWAS) for wood chemistry traits in black cottonwood (*P. trichocarpa*)[230], hundreds of genetic marker loci have been found significantly associated with wood chemical composition and ultrastructural traits in *Populus*[231,232], *Eucalyptus*[217,233], *Pinus*[234], and *Picea*[235]. Next-generation sequencing (NGS) has enabled more types of genetic variation such as insertion/deletions (InDels), structural variation (SV), and copy number variants (CNV) to be applied to GWAS to solve the problem of 'missing heritability'. For example, InDels can explain 14% of average phenotypic

variance in growth and wood property-related traits in *P. tomentosa*[236], and haplotype-based association analysis of multiple variation loci was also performed in *Populus*[219] and *Eucalyptus*[237]. In addition, by integrating heterogeneous phenotypic data from different ages, loci, and pedigrees in 120 field experiments of 483,424 progenies of Norway spruce, the accuracy of GWAS phenotypic values were largely improved[235]. In recent years, the analytical strategy combining GWAS with multiple omics data has been applied widely in the investigation of the genetic architecture of wood variation[228,238,239]. Various new methods and strategies have been gradually applied to GWAS to improve the analytical accuracy, such as the development of high-throughput phenotyping techniques[240], and to the detection of rare allelic and major structural variation[217,241].

In boreal and temperate ecosystems, the adaptation of perennial plants to their surroundings is the major cue to regulate the seasonally synchronous annual growth cycle[207], particularly, the key factors for the adaptation to changing conditions and environmental stressors over generations and the maintenance of standing genetic variation[242,243]. The application of environmental association analysis (EAA) allows the detection of the candidate genes involved in the environmental adaptation that results from populations across environmental gradients[244]. Many adaptive genetic variations have been determined in diverse and widespread woody plant genera. For example, a locus centered on *PtFT2* was identified using fitting latent factor mixed models (LFMM) in *Populus*, whose allele frequencies displayed a strong clinical pattern in latitude and had major effects on bud set[245]. In the *Quercus* range-wide model, six candidate SNPs were explored among the strongest environment-associated SNPs, offering robust evidence for local adaptation at multiple spatial scales[246]. These studies have determined the functional genes involved in the local adaptation that can be regarded as a gain or maintenance of divergent selection associated with complex environmental variables, and could potentially be utilized in ecological molecular breeding.

In reent years, our knowledge on the roles of epigenetic variation in diverse populations that underlie phenotypic variation has been growing. The rise of epigenetic quantitative trait loci (QTL$^{epi}$) mapping and epigenome-wide association studies (EWAS) allow the detection of causal QTLs$^{epi}$ and candidate genes, to overcome the insufficient effect of genetic dissection in explaining the 'missing heritability'. The first high-density linkage epigenetic map in tree species was constructed in *P. tomentosa*, and the QTLs$^{epi}$ that control growth and wood property traits were detected[247]. This study may serve as a paradigm for making headway in the systematic mapping of complex traits in trees. Likewise, EWAS has deepened epigenetics variation research in trait regulation[248,249]. The discovery of the *Bad karma* locus in somaclonal variants of oil palms is one of the more prominent EWAS examples[250]. DNA methylation variation associated with climate gradients is also important for phenotypic plasticity and their ecological adaptation of forest trees[251]. The role of CG methylation in adaptation to climate and spatial variation was approved in natural oak populations[252]. Recently, DNA methylation level was reported to be affected by genetic variation cues. Trait-associated differentially methylated regions (DMRs) were reported to show evidence of chromatin interactions, enhancer

activities[253] or in linkage[248] disequilibrium (LD) with nearby SNPs[254,255]. The interaction between DNA and methylation variations remains to be further investigated. The integrated linkage-LD mapping method using F1, F2, and backcross1 (BC1) populations will help to disentangle the causal methylation loci or regions that are responsible for heritable morphological variation from parents to progenies[256]. This will specifically reveal the contribution of allelic variants and overcome the inherent limitations of QTL$^{epi}$ mapping.

'Missing heritability' has occurred in many QTL studies of tree species due to a variety of reasons that include, but are not limited to, small sample sizes, rare alleles, allelic heterogeneity, and epistasis. The genetic contribution of rare functional variants with high proportions of low-frequency genetic polymorphisms in forest trees remains unexplored[232,234]. Future studies on rare alleles in genetically diverse forest populations may perform experimental segregating validation in family-populations, and also improve detection efficiency using new statistical models and methods, such as the sequence-based kernel association test (SKAT)[221], regional genetic mapping (RHM)[218] and haplotype association analysis[235]. We are increasingly capable to address questions through a combination of alleles and haplotypes for traits of interest. However, functional interpretation and annotation of loci is still a critical challenge, especially as the majority of susceptible loci are located in non-coding or intergenic regions. The pleiotropic effects of genetic epistasis on gene expression, metabolites, and growth can explain the biological regulatory mechanisms beneath the statistical associations. Systems genetic approaches can help provide a comprehensive understanding of quantitative traits and enable the molecular design of new cultivars.

For the quantitative trait loci and association studies in forest trees, we expect that advances in systems genetics, including dense genetic linkage mapping, GWAS with multiple types of genetic variations, EAA, QTL$^{epi}$ mapping, EWAS, and dissection of rare alleles, will promote data integration, upgradation and renovation, and provide a more powerful breeding guiding system for forest trees. The advent of precision genome editing such as CRISPR system in either genetic or epigenetic backgrounds will validate the causality of mutual exclusion (or heritable covariation) of traits and linkage drag, and will be an ideal method for the development of superior cultivars of forest trees.

## Genetic diversity and climate adaptation

Genetic diversity in forest trees is the foundation supporting their evolutionary potential in future generations. Populations with low levels of standing genetic diversity are predicted to have reduced responses to selection[257], which could lead to increased risks of population decline or extinction under rapid environmental change. The genetic reservoir in a species is shaped by both stochastic neutral demographic events and natural selection over the course of species' evolutionary history. Species with wide distribution ranges often show a linear relationship between population genetic distances and geographical distances, a pattern defined as isolation by distance (IBD)[258], which reflects decreasing rates of gene flow between more distantly located populations. IBD causes allele frequency clines among populations. Additionally, repeated

founder events along species' migration routes can produce discrete genetic clusters and further amplify IBD[259,260]. In species that are widely distributed over heterogeneous landscapes, locally adapted populations are expected to harbor genetic compositions selected by the local environments. Gene flow among locally adapted populations is limited by selection because of lower fitness of immigrants, a process defined as isolation by environment (IBE)[261,262]. The consequences of these dispersal-demographic factors and selective forces are not mutually exclusive, and they often act together in generating population differentiation in natural systems[259,261,263,264]. Dissecting the separate contributions of these neutral and selective processes to population diversity is thus the first step towards a mechanistic understanding about the distribution of genetic diversity across a landscape.

Quantifying genetic variation in forest tree populations has been transformed by high throughput sequencing technologies in recent years, and transitioned from limited organellar and nuclear SSR markers to genome-wide scans. A variety of genome-wide genotyping methods, such as restriction-site associated DNA sequencing, genotyping-by-sequencing, whole exome capture sequencing, SNP array and resequencing, have provided improved spatial and temporal resolution of evolutionary dynamics of major forest tree species[265−271]. However, defining the genetic basis of local adaptation is not a straightforward task due to the complex biological processes involved. The detection of adaptive variation from genome-wide data is often approached by using outlier detection, association of allele frequencies with environmental variables (GEA), and evaluation of the contribution of IBD and IBE to population differentiation using redundancy analysis (RDA). By constraining either geography or environment in partial RDA models, the independent contributions of the two factors can be quantified. One pattern emerging from conifer species is that environment alone explains less than 10% of the allele frequency shifts among populations at all genomic SNPs, however this portion can go up to ~20% at GEA and $F_{ST}$ outlier SNPs, although still leaving a large portion of the genetic diversity due to joint effects of IBE and IBD[259,263,269,271,272]. The strong confounding effect of IBE and IBD in natural populations makes disentangling adaptive variations among loci from the geographic relationships difficult[273].

When both phenotypic and genomic data are available in parallel, phenotype-genotype associations can be evaluated using genome-wide association studies (GWAS), which offers a possibility of identifying the genetic causes of phenotypic variation. GWAS have been successful in pedigree materials, but association mapping in natural populations often end up with weak predictive powers of causal loci due to the confounding factors mentioned above and complex genetic interactions for polygenic traits[274,275]. Hall et al.[269] illustrates this problem in their study on frost hardiness variation among Scots pine populations. They analyzed genotype–phenotype associations across 10 000 SNPs, and found a high marker-estimated heritability of the hardiness variation (0.56), illustrating the ability of genomic SNPs to capture the genetic variation in the trait. However, few loci appeared to have identifiable effects on the trait. The promising message from these studies is that as long as neutral-processes causing population structure are properly controlled and false positive rates are set at a reasonable level, GEA and GWAS can provide

valuable information about adaptation dynamics across environmental gradients.

Knowledge about genetic adaptations across landscapes is informative for predicting the degree of genetic offset of current populations to future conditions[276]. The offset is often presented as a genetic distance between the extant and required genomic compositions for matching a future condition, assuming the current genotype-environment relationships in local populations are at equilibrium[276,277]. Gradient forest (GF), a machine learning algorithm, has been increasingly used to characterize GEA and to predict regions in a distribution range that are most vulnerable to climate change[259,276,278]. Offset projections can be simulated under different climate change, migration and gene flow scenarios to gain a preliminary assessment of potential actions needed to track changing conditions. Although a very attractive approach, many assumptions in offset simulations are difficult to meet with certainty, e.g., defining adaptive and neutral genetic variation and the position of local populations on the adaptive landscape[259,279]. As illustrated in Fig. 3, genomic offset projections using different SNP sets convey different biological implications, thus may not align well with each other. A future improvement is to incorporate fitness effect estimates of allele frequency changes into GF models of species distribution shifts[279]. In this regard, long-term large-scale provenance trials remain irreplaceable even in the genomic era for model validation of environmental responses of forest trees.

To conclude, rapid climate changes represent a challenge for many forest tree species. Due to their long generation time and the often-limited seed dispersal, genetic adaptation in forest populations lags behind environmental change. Knowledge about genetic diversity and genetic adaptation in natural populations is therefore essential for sustainable management of forest ecosystems. Applications of this research include the development of guidelines for assisted gene flow[280], breeding zones, seed transfer for forest restoration, and conservation of endangered species and populations.

## Genotyping efforts and genomic selection in forest trees

Intensively managed forests are essential sources of fiber, biomass, pulp/paper, and timber in many countries worldwide[281]. Managed forests provide raw materials to industry sustainably while reducing the pressure on natural forests. For example, in the southern United States, about 20% of forestland is managed for softwood production, yet the region produces more than 55% of the timber[282]. Managed forests are significant sources of carbon storage, and they help mitigate greenhouse gas emissions. Forest tree breeding is an essential part of growing woody biomass sustainably. Moreover, genetics and breeding are usually the only way to improve the forests for biotic (pests and pathogens) and abiotic (air pollution, climate-change-related factors) stresses.

Despite the critical role of forest tree breeding for planted forests, forest tree breeding has not received the needed resources and support as crop and animal breeding programs. Forest tree breeding is still in its early stages. Tree breeding activities in some developing countries (e.g., Sweden, Finland, and the USA) started in the 1950s; however, modern tree breeding did not start in many countries until the 1980s. Even though the breeding of the major conifers started more than 60 years ago, the progress has been limited. The slow progress of tree breeding is due to the lack of resources, as tree breeding has not been considered as important as crops. Other reasons are biological. Forest trees, especially conifers, take many years
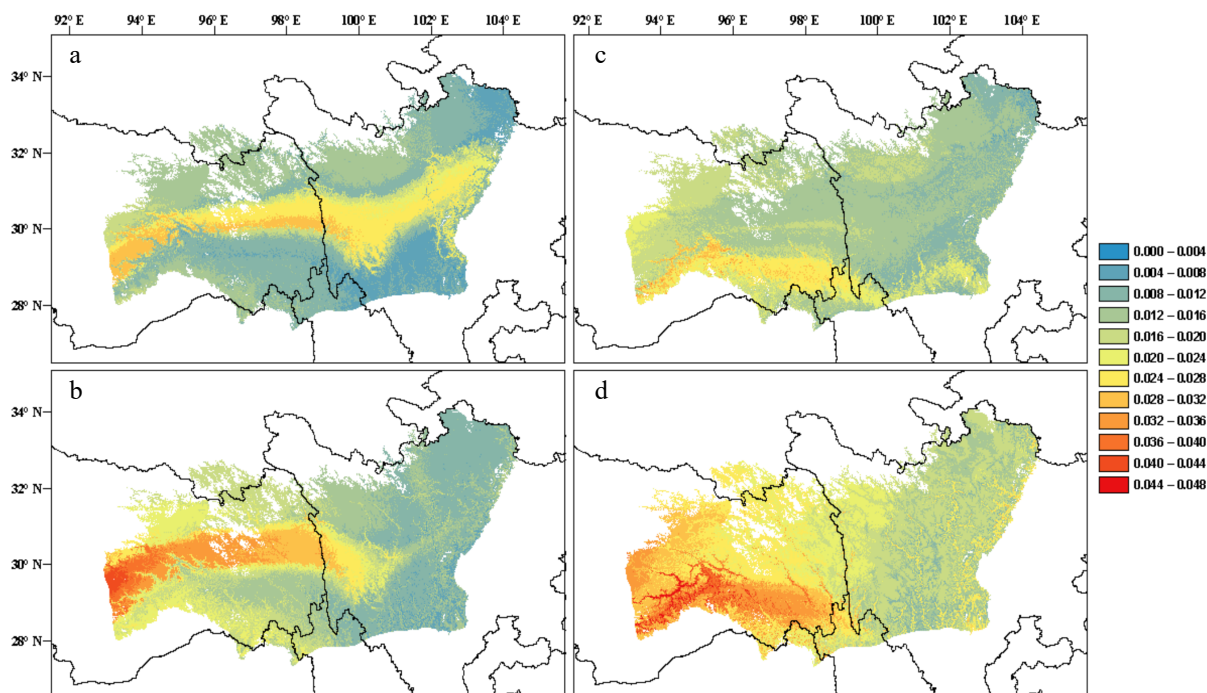


**Fig. 3** (a), (b) Predictions of genomic offset to future climate change in *Pinus densata* using a full set of 47,612 SNPs in exome sequences. (c), (d) Subset of 2,025 significant GEA SNPs. (a) and (c) reflect scenario representative concentration pathway (RCP) 2.6 2070; (b) and (d) reflect scenario RCP 8.5 2070. Red and blue indicate high and low genomic offset, respectively (Adapted from Zhao et al.[259]).

to mature and flower. The flowering is not frequent for many northern latitude conifer species. Field testing takes years, sometimes a decade, to collect data and make selection decisions. Shifting program priority and scientist turnover, may also disrupt the long breeding effort.

Advances in DNA sequencing technologies have impacted plant and animal breeding since 2008. Breeders are now equipped with thousands of single nucleotides polymorphic (SNP) markers to fundamentally change plant and animal breeding. SNP markers are promising to make selection of breeding germplasm at a juvenile stage, a process called genomic selection. Genomic selection has doubled genetic gain per unit of time in cattle breeding[283]. The impact of genomic selection and other applications of DNA markers in forest tree breeding is expected to be even higher because of longer breeding cycles of forest trees[284]. SNP arrays and genotyping-by-sequencing are the most common genotyping platforms for forest trees[285–288]. Forest tree breeders have shown a great interest in genomic selection (GS) to reduce the long breeding cycles[289,290]. Several studies on cost-benefit analysis suggested that GS selection is feasible[291,292]. However, the progress in the implementation of GS in forest tree breeding is still behind animal and major crop breeding. High genotyping cost per sample is one of the limiting factors. Developing more cost-efficient genotyping platforms is still active research[293]. For example, at North Carolina State University Tree Improvement Program, AgriSeq Targeted GBS panel designed by Thermo Fisher Scientific is promising for some molecular applications in *Pinus taeda* breeding (unpublished). The panel is based on a small subset of SNP markers selected from *Pinus taeda* SNP array Pita50K[285]. The quality control statistics are comparable to the SNP array. For example, the sample call rate is 86%, and sample uniformity is 89%. The following plot shows how the AgriSeq Targeted GBS panel clustered eight full-sib families compared to the Pita50K SNP array and a subset of array markers amplified for AgriSeq panel (Pita995) (Fig. 4).

A routine application of GS requires a reliable and cost-efficient genotyping platform. Although the cost of DNA sequencing has dropped sharply since 2008, the cost per sample using SNP arrays is still somewhat high, especially for many developing countries. Targeted sequencing-based methods are promising genotyping platforms to reduce the cost. Research in developing cost-efficient genotyping platforms will continue to be an important subject. Developing and updating high-quality reference genomes of forest trees needs to be prioritized. Improved reference genomes can greatly enhance marker discovery, annotation, haplotype construction, imput of missing genotypes, and QTL discovery. The current reference genomes of forest trees, especially conifers, are highly fragmented[294]. To be useful for many applications, they need substantial improvement using the latest long-read sequencing technologies, such as PacBio sequencing.

## Computational biology approaches in identifying genes controlling biological processes and complex traits in forest genomics-based research

The advent of microarray and RNA sequencing (RNA-seq) technologies has generated an enormous high-throughput gene expression data, which are usually analyzed with various statistical methods, leading to differentially expressed genes (DEGs). Following that, DEGs-based enrichment analyses including gene ontology (GO), protein domain and pathway enrichment analyses can be performed using Fisher Exact Test or hypergeometric distribution to reveal enriched moieties or entities[295]. These analyses are useful but provide little information about the organization of the genes in regulatory, collaborative, or interactive networks, which are essential for discovering underlying and novel molecular mechanisms essential for advancing our understanding. To build gene regulatory networks, many algorithms based on various mechanistic and statistical modelings have been developed. These methods can be, by and large, classified into two categories: dynamic and static methods. Dynamic methods include differential equation[296], finite state[297], dynamic Bayesian[298], control logic[299], Boolean[300], and stochastic networks[301], which requires true time-course data with small intervals. This kind of data can be readily generated for prokaryotes and unicellular organisms like yeast. Static methods, which do not necessarily require time-course data, are represented by graphical Gaussian models (GGM)[302], mutual information based relevance networks[303], Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE)[304], Context Likelihood of Relatedness (CLR)[305], C3NET[306], Mutual Information 3 (MI3)[307], and Bayesian probabilistic network[299]. Although
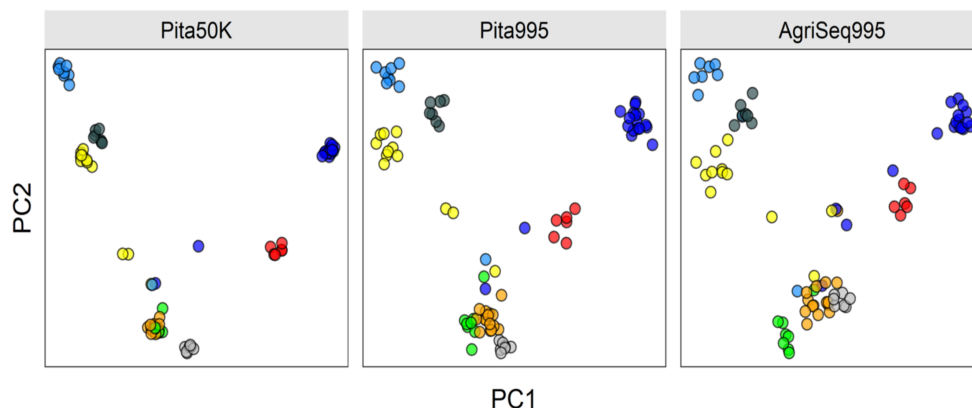


**Fig. 4** Clustering eight *Pinus taeda* full-sib families (colored circles) based on three different SNP marker sets. AgriSeq Targeted GBS panel (on the right) clustered the full-sib families similar to the same SNP markers selected from the *Pinus taeda* SNP array (in the middle). Trees not clustered are likely pedigree errors or the half-sibs (sharing one parent of another full-sib family).

static methods can also be used to analyze time-course data, they do not take temporal dependance and causality into consideration, which may cause some information loss. Although these dynamic and static methods can broaden our understanding of underlying gene regulatory networks (GRNs), they are not designed and tailored to generate inferences about the hierarchical architecture of the GRNs.

Recently, algorithms for construction of hierarchical GRNs from high-throughput gene expression have been established. These approaches can be used to infer the multilayered hierarchical network mediated by a transcription factor. For example, Top-down GGM Algorithm[308] for constructing a multilayered gene regulatory network mediated (ML-hGRN) by a regulatory gene like a transcription factor (TF). Top-down GGM algorithm has been employed to construct a ML-hGRN mediated by PtrSND1[308], PuHox52[309], PpnGRF5[310], Ptr-miRNA319a[311] and BplERF1[312]. The method is especially valuable when being used in conjunction with the perturbation of a regulator gene/TF followed by RNA-seq assay; the method has been shown to capacitate the separation of direct from indirect target genes with high accuracy[308]. On the contrary, Bottom-up GGM Algorithm[311,313] and Backward Elimination Random Forest Algorithm[314] have been developed for inferring ML-hGRN operating above a biological process or a pathway. From the ML-hGRN built with Top-down GGM Algorithm, the hierarchical regulators at different levels can be identified. Thus far, Bottom-up GGM Algorithm has been used to identify Ptr-miRNA319a[311] functioning above lignin polymerization pathway, and PuMYB40 and PuWRKY75 functioning as high hierarchical regulators above the biological process of low phosphorus (LP)-mediated adventitious root (AR) formation. Bottom-up GGM Algorithm and Bottom-up GGM can be used synergistically to infer the ML-hGRN encompassing a TF[315], as shown in Fig. 5.

One of the central tasks of genomics is to identify the genes regulating a biological process or a complex trait directly from high-throughput transcriptomic data. To do this, the high-throughput data needs to be generated from a tissue or an organ in which a trait of interest is under a transition. This kind of transcriptome is easy to produce. However, until now the genes controlling a trait have been primarily identified through linkage mapping and genome-wide association studies, which is labor-intensive and time-consuming, and may not guarantee a success in most cases where pleiotropy is dominant. Nevertheless, a new theory has been proposed for identifying the genes regulating a biological process or a complex trait from transcriptomes, which assumes that the genes that regulate a biological process or a complex trait often collaborate throughout the time period during which the biological process or the complex train is under a transition[316,317]. Based on this theory, a method was developed for building the collaborative gene regulatory network which was then decomposed into many small subnetworks, each containing regulatory genes collaboratively controlling a biological process or a trait. The method has been applied to multiple transcriptomic data sets in multiple species and the results showed that regulatory genes governing a biological process or a complex trait were captured in almost every case.

Recently, a method for joint construction of multiple gene GRNs using gene expression data from multiple tissues or conditions was developed[318]. The method allows for identifying the common regulators across multiple tissues/conditions and unique regulators peculiar to just one tissue. In addition to this, integrative analysis of spatial transcriptome, single-cell RNA-seq and RNA-seq data has emerged[25,91], which indicates a new era for interpretation of spatiotemporal data yielded from development and differentiation processes has arrived. Obviously, more advanced software pipelines and tools for analyzing multi-sourced data or real spatiotemporal data are needed.

## Concluding remarks

It becomes increasingly clear that high-throughput sequencing technologies, together with the newly available scRNA-seq, CRISPR-mediated genome editing, CRISPR-mediated upregulation/downregulation, spatial transcriptome and advanced
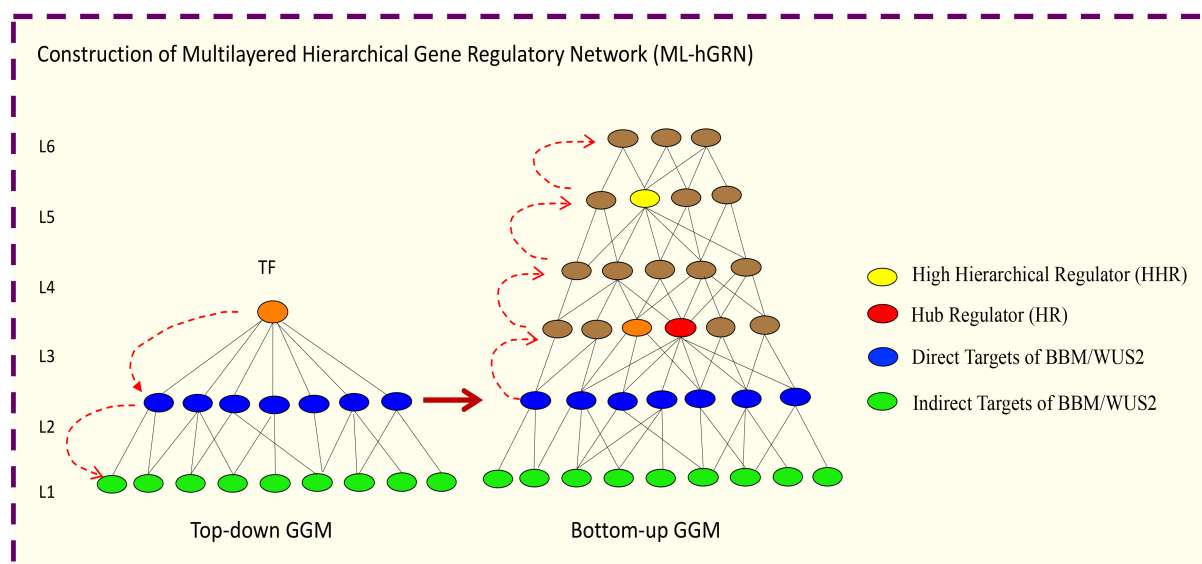


**Fig. 5**  Illustration of how to construct a multilayered hierarchical gene regulatory network (ML-hGRN) to encompass a given transcription factor (TF) using Top-down GGM and Bottom-up GGM Algorithms synergistically.

Page 14 of 22

Borthakur et al. Forestry Research 2022, 2:11

bioinformatics analysis technologies have provided unprecedented opportunities to leverage the important and unique development and differentiation-related issues of forest trees. The large amounts of data will help to study molecular mechanisms of tree growth and development as well as genome-assisted forest tree breeding. Future research in forest genomics is being refocused on some important developmental, evolutionary, and adaptive problems, for example, stem cell entity maintenance, tissue, organ and architecture formation, secondary growth, transformation recalcitrance, nitrogen fixing, perennial growth and seasonality regulation, genome evolution, genetic diversity and climate changes, QTL mapping and genomic selection, and regulatory mechanisms underlying various complex traits. Forest genomics research is moving towards more interdisciplinary endeavors. With effective collaboration among global researchers from different disciplines, we can optimistically foresee the even greater, and faster advancement in forest genomics and systems biology, which will maximize the potential of developing 'desirable, customer-designed forest trees', for improving our environment, and fighting climate change.

## Conflict of interest

The authors declare that they have no conflict of interest.

## Dates

## REFERENCES

1. Pan Y, Birdsey RA, Phillips OL, Jackson RB. 2013. The structure, distribution, and biomass of the World's forests. *Anual Review of Ecology, Evolution, and Systematics* 44:593−622
2. Wei H. 2021. Inaugural Editorial. *Forestry Research* 1:1
3. Plomion C, Bastien C, Bogeat-Triboulot MB, Bouffier L, Déjardin A, et al. 2016. Forest tree genomics: 10 achievements from the past 10 years and future prospects. *Annals of Forest Science* 73:77−103
4. Allona I, Kirst M, Boerjan W, Strauss S, Sederoff R. 2019. Editorial: Forest Genomics and Biotechnology. *Frontiers in Plant Science* 10:1187
5. Schultz JA, Coleman HD. 2021. Pectin and xylan biosynthesis in poplar: Implications and opportunities for biofuels production. *Frontiers in Plant Science* 12:712083
6. Cairns MA, Brown S, Helmer EH, Baumgardner GA. 1997. Root biomass allocation in the world's upland forests. *Oecologia* 111:1−11
7. Houghton RA, Hall F, Goetz SJ. 2009. Importance of biomass in the global carbon cycle. *Biogeosciences* 114:G00E03
8. de Vries L, Guevara-Rozo S, Cho M, Liu LY, Renneckar S, et al. 2021. Tailoring renewable materials via plant biotechnology. *Biotechnology For Biofuels* 14:167
9. Fukuda H, Komamine A. 1980. Establishment of an experimental system for the study of tracheary element differentiation from single cells isolated from the mesophyll of *Zinnia elegans*. *Plant Physiology* 65:57−60
10. Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, et al. 2005. Transcription switches for protoxylem and metaxylem vessel formation. *Genes & Development* 19:1855−60
11. Wang D, Chen Y, Li W, Li Q, Lu M, et al. 2021. Vascular Cambium: The Source of Wood Formation. *Frontiers in Plant Science* 12:700928
12. Ko JH, Kim WC, Kim JY, Ahn SJ, Han KH. 2012. MYB46-mediated transcriptional regulation of secondary wall biosynthesis. *Molecular Plant* 5:961−63
13. Zhu Y, Li L. 2021. Multi-layered regulation of plant cell wall thickening. *Plant and Cell Physiology* 62:1867−73
14. Xu C, Shen Y, He F, Fu X, Yu H, et al. 2019. Auxin-mediated Aux/IAA-ARF-HB signaling cascade regulates secondary xylem development in *Populus*. *New Phtologist* 222:752−67
15. Shi D, Lebovka I, López-Salmerón V, Sanchez P, Greb T. 2019. Bifacial cambium stem cells generate xylem and phloem during radial plant growth. *Development* 146:dev171355
16. Smetana O, Mäkilä R, Lyu M, Amiryousefi A, Sanchez Rodriguez F, et al. 2019. High levels of auxin signalling define the stem-cell organizer of the vascular cambium. *Nature* 565:485−89
17. Chiang MH, Greb T. 2019. How to organize bidirectional tissue production. *Current Opinion in Plant Biology* 51:15−21
18. Wenzel CL, Schuetz M, Yu Q, Mattsson J. 2007. Dynamics of *MONOPTEROS* and PIN-FORMED1 expression during leaf vein pattern formation in *Arabidopsis thaliana*. *The Plant Journal* 49:387−98
19. De Rybel B, Adibi M, Breda AS, Wendrich JR, Smit ME, et al. 2014. Plant development. Integration of growth and patterning during vascular tissue formation in *Arabidopsis*. *Science* 345:1255215
20. Ohashi-Ito K, Saegusa M, Iwamoto K, Oda Y, Katayama H, et al. 2014. A bHLH complex activates vascular cell division via cytokinin action in root apical meristem. *Current Biology* 24:2053−58
21. Smet W, Sevilem I, de Luis Balaguer MA, Wybouw B, Mor E, et al. 2019. DOF2.1 controls cytokinin-dependent vascular cell proliferation downstream of TMO5/LHW. *Current Biology* 29:520−9.e6
22. Hardtke CS, Berleth T. 1998. The *Arabidopsis* gene *MONOPTEROS* encodes a transcription factor mediating embryo axis formation and vascular development. *The EMBO Journal* 17:1405−11
23. Chen Y, Tong S, Jiang Y, Ai F, Feng Y, et al. 2021. Transcriptional landscape of highly lignified poplar stems at single-cell resolution. *Genome Biology* 22:319
24. Li H, Dai X, Huang X, Xu M, Wang Q, et al. 2021. Single-cell RNA sequencing reveals a high-resolution cell atlas of xylem in *Populus*. *Journal of Integrative Plant Biology* 63:1906−21
25. Chen A, Liao S, Cheng M, Ma K, Wu L, et al. 2022. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* 185:1777−92 e21
26. Cao HX, Vu GTH, Gailing O. 2022. From Genome sequencing to CRISPR-based genome editing for climate-resilient forest trees. *International Journal of Molecular Sciences* 23:966
27. Pak S, Li C. 2022. Progress and challenges in applying CRISPR/Cas techniques to the genome editing of trees. *Forestry Research* 2:6
28. Fan D, Liu T, Li C, Jiao B, Li S, et al. 2015. Efficient CRISPR/Cas9-mediated targeted mutagenesis in Populus in the first generation. *Scientific Reports* 5:12217
29. van Zeijl A, Wardhani TAK, Seifi Kalhor M, Rutten L, Bu F, et al. 2018. CRISPR/Cas9-mediated mutagenesis of four putative symbiosis genes of the tropical tree *Parasponia andersonii* reveals novel phenotypes. *Frontiers in Plant Science* 9:284
30. Dai Y, Hu G, Dupas A, Medina L, Blandels N, et al. 2020. Implementing the CRISPR/Cas9 technology in *Eucalyptus* hairy roots using wood-related genes. *International Journal of Molecular Sciences* 21:3408
31. Dai X, Yang X, Wang C, Fan Y, Xin S, et al. 2021. CRISPR/Cas9-mediated genome editing in *Hevea brasiliensis*. *Industrial Crops And Products* 164:113418
32. Fan Y, Xin S, Dai X, Yang X, Huang H, et al. 2020. Efficient genome editing of rubber tree (*Hevea brasiliensis*) protoplasts using CRISPR/Cas9 ribonucleoproteins. *Industrial Crops And Products* 146:112146
33. Poovaiah C, Phillips L, Geddes B, Reeves C, Sorieul M, Thorlby G. 2021. Genome editing with CRISPR/Cas9 in *Pinus radiata* (D. Don). *BMC Plant Biology* 21:363
34. Pavese V, Moglia A, Corredoira E, Martínez MT, Torello Marinoni D, et al. 2021. First report of CRISPR/Cas9 gene editing in *Castanea sativa* Mill. *Frontiers in Plant Science* 12:728516

35. Pramanik D, Shelake RM, Kim MJ, Kim JY. 2021. CRISPR-mediated engineering across the central dogma in plant biology for basic research and crop improvement. *Molecular Plant* 14:127–50

36. Hassan MM, Zhang Y, Yuan G, De K, Chen JG, et al. 2021. Construct design for CRISPR/Cas-based genome editing in plants. *Trends in Plant Science* 26:1133–52

37. Nidhi S, Anand U, Oleksak P, Tripathi P, Lal JA, et al. 2021. Novel CRISPR–cas systems: An updated review of the current achievements, applications, and future research perspectives. *International Journal of Molecular Sciences* 22:3327

38. Lu Y, Tian Y, Shen R, Yao Q, Wang M, et al. 2020. Targeted, efficient sequence insertion and replacement in rice. *Nature Biotechnology* 38:1402–7

39. Li S, Li J, He Y, Xu M, Zhang J, et al. 2019. Precise gene replacement in rice by RNA transcript-templated homologous recombination. *Nature Biotechnology* 37:445–50

40. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, et al. 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353:aaf8729

41. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, et al. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576:149–57

42. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533:420–24

43. Chanoca A, de Vries L, Boerjan W. 2019. Lignin engineering in forest trees. *Frontiers in Plant Science* 10:912

44. Dort EN, Tanguay P, Hamelin RC. 2020. CRISPR/Cas9 gene editing: An unexplored frontier for forest pathology. *Frontiers in Plant Science* 11:1126

45. Fang Q, Wang X, Wang H, Tang X, Liu C, et al. 2020. The poplar R2R3 MYB transcription factor *PtrMYB94* coordinates with abscisic acid signaling to improve drought tolerance in plants. *Tree Physiology* 40:46–59

46. Su Y, Guo A, Huang Y, Wang Y, Hua J. 2020. *GhCIPK6a* increases salt tolerance in transgenic upland cotton by involving in ROS scavenging and MAPK signaling pathways. *BMC Plant Biology* 20:421

47. Xiu Y, Iqbal A, Zhu C, Wu G, Chang Y, et al. 2016. Improvement and transcriptome analysis of root architecture by overexpression of *Fraxinus pennsylvanica* DREB2A transcription factor in *Robinia pseudoacacia* L. 'Idaho'. *Plant Biotechnology Journal* 14:1456–69

48. Li R, Liu L, Dominic K, Wang T, Fan T, et al. 2018. Mulberry (*Morus alba*) MmSK gene enhances tolerance to drought stress in transgenic mulberry. *Plant Physiology and Biochemistry* 132:603–11

49. Bewg WP, Ci D, Tsai CJ. 2018. Genome editing in trees: from multiple repair pathways to long-term stability. *Frontiers in Plant Science* 9:1732

50. Polle A, Chen SL, Eckert C, Harfouche A. 2019. Engineering Drought Resistance in Forest Trees. *Frontiers in Plant Science* 9:1875

51. Molla KA, Yang Y. 2019. CRISPR/Cas-mediated base editing: technical considerations and practical applications. *Trends in Biotechnology* 37:1121–42

52. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, et al. 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551:464–71

53. Zhao D, Li J, Li S, Xin X, Hu M, et al. 2021. Glycosylase base editors enable C-to-A and C-to-G base changes. *Nature Biotechnology* 39:35–40

54. Kurt IC, Zhou R, Iyer S, Garcia SP, Miller BR, et al. 2021. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nature Biotechnology* 39:41–6

55. Chen L, Park JE, Paa P, Rajakumar PD, Prekop HT, et al. 2021. Programmable C:G to G:C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nature Communication* 12:1348

56. Choi M, Yun JY, Kim JH, Kim JS, Kim ST. 2021. The efficacy of CRISPR-mediated cytosine base editing with the RPS5a promoter in *Arabidopsis thaliana*. *Scientific Reports* 11:8087

57. Molla KA, Sretenovic S, Bansal KC, Qi Y. 2021. Precise plant genome editing using base editors and prime editors. *Nature Plants* 7:1166–87

58. Gross A, Schoendube J, Zimmermann S, Steeb M, Zengerle R, et al. 2015. Technologies for single-cell isolation. *International Journal of Molecular Sciences* 16:16897–919

59. Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, et al. 2017. Single-cell barcoding and sequencing using droplet microfluidics. *Nature Protocols* 12:44–73

60. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, et al. 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161:1187–201

61. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nature Communications* 8:14049

62. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, et al. 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–14

63. Ryu KH, Huang L, Kang HM, Schiefelbein J. 2019. Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiology* 179:1444–56

64. Zhang T, Xu Z, Shang G, Wang J. 2019. A single-cell RNA sequencing profiles the developmental landscape of *Arabidopsis* root. *Molecular Plant* 12:648–60

65. Shulse CN, Cole BJ, Ciobanu D, Lin J, Yoshinaga Y, et al. 2019. High-throughput single-cell transcriptome profiling of plant cell types. *Cell Reports* 27:2241–2247.E4

66. Denyer T, Ma X, Klesen S, Scacchi E, Nieselt K, et al. 2019. Spatiotemporal developmental trajectories in the Arabidopsis root revealed using high-throughput single-cell RNA sequencing. *Developmental Cell* 48:840–852.E5

67. Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, et al. 2019. Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. *The Plant Cell* 31:993–1011

68. Shahan R, Hsu CW, Nolan TM, Cole BJ, Taylor IW, et al. 2022. A single-cell *Arabidopsis* root atlas reveals developmental trajectories in wild-type and cell identity mutants. *Developmental Cell* 57:543–560.e9

69. Zhang T, Chen Y, Liu Y, Lin W, Wang J. 2021. Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. *Nature Communications* 12:2053

70. Liu Q, Liang Z, Feng D, Jiang S, Wang Y, et al. 2021. Transcriptional landscape of rice roots at the single-cell resolution. *Molecular Plant* 14:384–94

71. Zhang T, Chen Y, Wang J. 2021. A single-cell analysis of the Arabidopsis vegetative shoot apex. *Developmental Cell* 56:1056–74

72. Song Q, Ando A, Jiang N, Ikeda Y, Chen ZJ. 2020. Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in Arabidopsis female gametophytes. *Genome Biology* 21:178

73. Nelms B, Walbot V. 2019. Defining the developmental program leading to meiosis in maize. *Science* 364:52–56

74. Satterlee JW, Strable J, Scanlon MJ. 2020. Plant stem-cell organization and differentiation at single-cell resolution. *PNAS* 117:33689–99

75. Xu X, Crow M, Rice BR, Li F, Harris B, et al. 2021. Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Developmental Cell* 56:557–568.E6

76. Kim JY, Symeonidi E, Pang TY, Denyer T, Weidauer D, et al. 2021. Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *The Plant Cell* 33:511–30

77. Serrano-Ron L, Perez-Garcia P, Sanchez-Corrionero A, Gude I,

Cabrera J, et al. 2021. Reconstruction of lateral root formation through single-cell RNA sequencing reveals order of tissue initiation. *Molecular Plant* 14:1362–78

78. Liu Z, Zhou Y, Guo J, Li J, Tian Z, et al. 2020. Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. *Molecular Plant* 13:1178–93

79. Liu H, Hu D, Du P, Wang L, Liang X, et al. 2021. Single-cell RNA-seq describes the transcriptome landscape and identifies critical transcription factors in the leaf blade of the allotetraploid peanut (*Arachis hypogaea* L.). *Plant Biotechnology Journal* 19:2261–76

80. Gala HP, Lanctot A, Jean-Baptiste K, Guizou S, Chu JC, et al. 2021. A single-cell view of the transcriptome during lateral root initiation in Arabidopsis thaliana. *The Plant Cell* 33:2197–220

81. Lopez-Anido CB, Vatén A, Smoot NK, Sharma N, Guo V, et al. 2021. Single-cell resolution of lineage trajectories in the Arabidopsis stomatal lineage and developing leaf. *Developmental Cell* 56:1043–1055.E4

82. Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, et al. 2003. A gene expression map of the *Arabidopsis* root. *Science* 302:1956–60

83. Farmer A, Thibivilliers S, Ryu KH, Schiefelbein J, Libault M. 2021. Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in Arabidopsis roots at the single-cell level. *Molecular Plant* 14:372–83

84. Deal RB, Henikoff S. 2010. A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Developmental Cell* 18:1030–40

85. Reynoso MA, Pauluzzi GC, Kajala K, Cabanlit S, Velasco J, et al. 2018. Nuclear transcriptomes at high resolution using retooled INTACT. *Plant Physiology* 176:270–81

86. Palovaara J, Weijers D. 2019. Adapting INTACT to analyse cell-type-specific transcriptomes and nucleocytoplasmic mRNA dynamics in the *Arabidopsis* embryo. *Plant Reproduction* 32:113–21

87. Del Toro-De León G, Köhler C. 2019. Endosperm-specific transcriptome analysis by applying the INTACT system. *Plant Reproduction* 32:55–61

88. Pirrello J, Deluche C, Frangne N, Gévaudant F, Maza E, et al. 2018. Transcriptome profiling of sorted endoreduplicated nuclei from tomato fruits: how the global shift in expression ascribed to DNA ploidy influences RNA-Seq data normalization and interpretation. *The Plant Journal* 93:387–98

89. Conde D, Triozzi PM, Balmant KM, Doty AL, Miranda M, et al. 2021. A robust method of nuclei isolation for single-cell RNA sequencing of solid tissues from the plant genus *Populus*. *PLoS One* 16:e0251149

90. Turco GM, Rodriguez-Medina J, Siebert S, Han D, Valderrama-Gómez MÁ, et al. 2019. Molecular mechanisms driving switch behavior in xylem cell differentiation. *Cell Reports* 28:342–351.E4

91. Giacomello S, Salmén F, Terebieniec BK, Vickovic S, Navarro JF, et al. 2017. Spatially resolved transcriptome profiling in model plant species. *Nature Plants* 3:17061

92. Giacomello S, Lundeberg J. 2018. Preparation of plant tissue to enable Spatial Transcriptomics profiling using barcoded microarrays. *Nature Protocols* 13:2425–46

93. Dunwell JM. 2010. Haploids in flowering plants: origins and exploitation. *Plant Biotechnology Journal* 8:377–424

94. Coe EH Jr . 1959. A line of maize with high haploid frequency. *The American Naturalist* 93:381–2

95. Jain SM, Sopory SK, Veilleux R. 1996. *In Vitro Haploid Production in Higher Plants: Volume 2: Applications*. Springer Science & Business Media

96. Maluszynski M, Kasha K, Forster B, Szarejko I (eds.). 2003. *Doubled haploid production in crop plants: anual*. New York: Springer Science & Business Media

97. Xu L, Najeeb U, Tang G, Gu H, Zhang G, et al. 2007. Haploid and doubled haploid technology. *Advances in Botanical Research* 45:181–216

98. Piosik Ł, Zenkteler E, Zenkteler M. 2016. Development of haploid embryos and plants of *Lactuca sativa* induced by distant pollination with *Helianthus annuus* and *H. tuberosus*. *Euphytica* 208:439–51

99. Godbole M, Murthy HN. 2012. In vitro production of haploids via parthenogenesis in culinary melon (*Cucumis melo* var. acidulus). *Indian Journal of Biotechnology* 11:495–97

100. Soleimani A. 2012. Production of haploid lines from parthenogenetic Iranian melon plants obtained of irradiated pollen (*Cucumis melo* L.). *International Research Journal of Applied and Basic Sciences* 3:1585–89

101. Deng Y, Tang B, Zhou X, Fu W, Tao L, et al. 2020. Direct regeneration of haploid or doubled haploid plantlets in cucumber (*Cucumis sativus* L.) through ovary culture. *Plant Cell, Tissue and Organ Culture (PCTOC)* 142:253–68

102. Zayachkovskaya T, Domblides E, Zayachkovsky V, Kan L, Domblides A, et al. 2021. Production of Gynogenic Plants of Red Beet (*Beta vulgaris* L.) in Unpollinated Ovule Culture In Vitro. *Plants* 10:2703

103. Castillo AM, Valero-Rubira I, Allué S, Costar MA, Vallés MP. 2021. Bread Wheat Doubled Haploid Production by Anther Culture. *Methods in Molecular Biology* 2287:227–44

104. Kurtar ES, Seymen M. 2021. Gynogenesis in *Cucurbita* Species. *Methods in Molecular Biology* 2289:123–33

105. Galán-Ávila A, García-Fortea E, Prohens J, Herraiz FJ. 2021. Microgametophyte Development in Cannabis sativa L. and First Androgenesis Induction Through Microspore Embryogenesis. *Frontiers in Plant Science* 12:669424

106. Kang X, Wei H. 2022. Breeding polyploid *Populus*: progress and perspective. *Forestry Research* 2:4

107. Winton LL, Einspahr DW. 1968. The use of heat-treated pollen for aspen haploid production. *Forest Science* 406–7

108. Stettler R, Bawa K. 1971. Experimental induction of haploid parthenogenesis in black cottonwood (*Populus trichocarpa* T. & G. ex Hook.). *Silvae Genetica* 20:15–25

109. Illies Z. 1974. Induction of haploid parthenogenesis in aspen by postpollination treatment with Toluidine-blue. *Silvae Genetica* 23:221–26

110. Wu K, Xu M. 1984. Induction of matrilinear haploid plants from unpollinated ovaries of poplar invitro. *Kexue Tongbao* 29:141–42

111. Li Y, Huang S, Zhang J, Bu F, Lin T, et al. 2016. A protocol of homozygous haploid callus induction from endosperm of Taxus chinensis Rehd. var. mairei. *SpringerPlus* 5:1–9

112. Li Y, Wei H, Yang J, Du K, Li J, et al. 2020. High-quality de novo assembly of the *Eucommia ulmoides* haploid genome provides new insights into evolution and rubber biosynthesis. *Horticulture Research* 7:183

113. Wang C, Chu Z, Sun C. 1975. The induction of pollen plants of Populus. *Acta Botanica Sinica* 17:56–62

114. Deutsch F, Kumlehn J, Ziegenhagen B, Fladung M. 2004. Stable haploid poplar callus lines from immature pollen culture. *Physiologia Plantarum* 120:613–22

115. Li Y, Li H, Chen Z, Ji L, Ye M, et al. 2013. Haploid plants from anther cultures of poplar (*Populus × beijingensis*). *Plant Cell, Tissue and Organ Culture (PCTOC)* 114:39–48

116. Yang J, Li K, Li C, Li J, Zhao B, et al. 2018. In vitro anther culture and *Agrobacterium*-mediated transformation of the *AP1* gene from Salix integra Linn. in haploid poplar (*Populus simonii × P. nigra*). *Journal of Forestry Research* 29:321–30

117. Ravi M, Chan SWL. 2010. Haploid plants produced by centromere-mediated genome elimination. *Nature* 464:615–18

118. Kelliher T, Starr D, Su X, Tang G, Chen Z, et al. 2019. One-step genome editing of elite crop germplasm during haploid induction. *Nature Biotechnology* 37:287–92

119. Wang N, Gent JI, Dawe RK. 2021. Haploid induction by a maize *cenh3* null mutant. *Science Advances* 7:eabe2299

120. Castellanos-Hernández OA, Rodríguez-Sahagún A, Acevedo-Hernández GJ, Herrera-Estrella LR. 2011. Genetic Transformation of Forest Trees. In *Genetic Transformation*, ed. Alvarez M. Rijeka: IntechOpen. pp. 191–214. http://doi.org/10.5772/24354

121. Song G, Prieto H, Orbovic V. 2019. *Agrobacterium*-Mediated

Transformation of Tree Fruit Crops: Methods, Progress, and Challenges. *Frontiers in Plant Science* 10:266

122. Giri CC, Shyamkumar B, Anjaneyulu C. 2004. Progress in tissue culture, genetic transformation and applications of biotechnology to trees: an overview. *Trees* 18:115–35

123. Fillatti JJ, Sellmer J, Mccown B, Haissig B, Comai L. 1987. Agrobacterium Mediated Transformation and Regeneration of Populus. *Molecular and General Genetics MGG* 206:192–9

124. Huang Y, Diner AM, Karnosky DF. 1991. *Agrobacterium rhizogenes*-mediated genetic transformation and regeneration of a conifer: *Larix decidua*. *In Vitro Cellular & Developmental Biology - Plant* 27:201–7

125. Bruegmann T, Polak O, Deecke K, Nietsch J, Fladung M. 2019. Poplar Transformation. In *Transgenic Plants, Methods in Molecular Biology*, eds. Kumar S, Barone P, Smith M. vol 1864. New York: Humana Press. pp. 165–77 https://doi.org/10.1007/978-1-4939-8778-8_12

126. Litz RE, Padilla G. 2012. Genetic Transformation of Fruit Trees. In *Genomics of Tree Crops*, eds. Schnell R, Priyadarshan P. New York: Springer. pp. 117–53. https://doi.org/10.1007/978-1-4614-0920-5_5

127. Jayashree R, Rekha K, Venkatachalam P, Uratsu SL, Dandekar AM, et al. 2003. Genetic transformation and regeneration of rubber tree (*Hevea brasiliensis* Muell. Arg) transgenic plants with a constitutive version of an anti-oxidative stress superoxide dismutase gene. *Plant Cell Reports* 22:201–9

128. Tang W, Newton RJ. 2003. Genetic transformation of conifers and its application in forest biotechnology. *Plant Cell Reports* 22:1–15

129. Cervera M, Navarro A, Navarro L, Peña L. 2008. Production of transgenic adult plants from clementine mandarin by enhancing cell competence for transformation and regeneration. *Tree Physiology* 28:55–66

130. Dong JZ, Jia SR. 1991. High efficiency plant regeneration from cotyledons of watermelon (*Citrullus vulgaris* Schrad). *Plant Cell Reports* 9:559–62

131. Baker BS, Bhatia SK. 1993. Factors effecting adventitious shoot regeneration from leaf explants of quince (*Cydonia oblonga*). *Plant Cell, Tissue and Organ Culture* 35:273–77

132. Becerra D, Forero A, Góngora GA. 2004. Age and physiological condition of donor plants affect in vitro morphogenesis in leaf explants of *Passiflora edulis* f. *flavicarpa*. *Plant Cell, Tissue and Organ Culture* 79:87–90

133. Zhang T, Lian H, Tang H, Dolezal K, Zhou C, et al. 2015. An intrinsic microRNA timer regulates progressive decline in shoot regenerative capacity in plants. *The Plant Cell* 27:349–60

134. Raspor M, Motyka V, Kaleri AR, Ninković S, Tubić L, et al. 2021. Integrating the roles for cytokinin and auxin in de novo shoot organogenesis: from hormone uptake to signaling outputs. *International Journal of Molecular Sciences* 22:8554

135. Hill K, Schaller GE. 2013. Enhancing plant regeneration in tissue culture: a molecular approach through manipulation of cytokinin sensitivity. *Plant Signaling and Behavior* 8:e25709

136. Yang J, Zhao B, Kim YB, Zhou C, Li C, et al. 2013. *Agrobacterium tumefaciens*-mediated transformation of *Phellodendron amurense* Rupr. using mature-seed explants. *Molecular Biology Reports* 40:281–88

137. Yang J, Zhao S, Zhao B, Li C. 2018. Overexpression of *TaLEA3* induces rapid stomatal closure under drought stress in *Phellodendron amurense* Rupr. *Plant Science* 277:100–9

138. Yang J, Zhou C, Zhao B, Jin C, Zhang T, et al. 2011. Rapid direct adventitious shoot organogenesis and plant regeneration from mature seed explants of *Phellodendron amurense* Rupr. *Journal of Medicinal Plant Research* 5:4560–65

139. Yang J, Yi J, Yang C, Li CJTP. 2013. *Agrobacterium tumefaciens*-mediated genetic transformation of *Salix matsudana* Koidz. using mature seeds. *Tree Physiology* 33:628–39

140. Yang J, Chen Z, Wu S, Cui Y, Zhang L, et al. 2015. Overexpression of the *Tamarix hispida ThMT3* gene increases copper tolerance and adventitious root induction in *Salix matsudana* Koidz. *Plant Cell, Tissue and Organ Culture* 121:469–79

141. Teixeira da Silva JA, Malabadi RB. 2012. Factors affecting somatic embryogenesis in conifers. *Journal of Forestry Research* 23:503–15

142. Hazubska-Przyby T, Bojarczuk K. 2016. Tree somatic embryogenesis in science and forestry. *Dendrobiology* 76:105–16

143. Ikeuchi M, Ogawa Y, Iwase A, Sugimoto K. 2016. Plant regeneration: cellular origins and molecular mechanisms. *Development* 143:1442–51

144. Duan H, Maren NA, Ranney TG, Liu WJOPR. 2022. New opportunities for using *WUS/BBM* and *GRF-GIF* genes to enhance genetic transformation of ornamental plants. *Ornamental Plant Research* 2:4

145. Lowe K, Wu E, Wang N, Hoerster G, Hastings C, et al. 2016. Morphogenic regulators *Baby boom* and *Wuschel* improve monocot transformation. *The Plant Cell* 28:1998–2015

146. Debernardi JM, Tricoli DM, Ercoli MF, Hayta S, Ronald P, et al. 2020. A GRF-GIF chimeric protein improves the regeneration efficiency of transgenic plants. *Nature Biotechnology* 38:1274–79

147. Liu B, Zhang J, Yang Z, Matsui A, Seki M, et al. 2018. *PtWOX11* acts as master regulator conducting the expression of key transcription factors to induce de novo shoot organogenesis in poplar. *Plant Molecular Biology* 98:389–406

148. Chen J, Tomes S, Gleave AP, Hall W, Luo Z, et al. 2022. Significant improvement of apple (*Malus domestica* Borkh.) transgenic plant production by pre-transformation with a Baby boom transcription factor. *Horticulture Research* 9:uhab014

149. Xu L, Huang H. 2014. Genetic and epigenetic controls of plant regeneration. *Current Topics In Developmental Biology* 108:1–33

150. Nagle M, Déjardin A, Pilate G, Strauss SH. 2018. Opportunities for Innovation in Genetic Transformation of Forest Trees. *Frontiers in Plant Science* 9:1443

151. Luo G, Palmgren M. 2021. GRF-GIF Chimeras Boost Plant Regeneration. *Trends in Plant Science* 26:201–4

152. Ellison EE, Nagalakshmi U, Gamo ME, Huang PJ, Dinesh-Kumar S, et al. 2020. Multiplexed heritable gene editing using RNA viruses and mobile single guide RNAs. *Nature Plants* 6:620–24

153. Ji X, Yang B, Wang D. 2020. Achieving plant genome editing while bypassing tissue culture. *Trends in Plant Science* 25:427–29

154. Yang J, Lan L, Jin Y, Yu N, Wang D, et al. 2022. Mechanisms underlying legume-rhizobium symbioses. *Journal of Integrative Plant Biology* 64:244–67

155. Leary JK, Singleton PW, Borthakur D. 2004. Canopy nodulation of the endemic tree legume Acacia koa in the mesic forests of Hawaii. *Ecology* 85:3151–75

156. Honda MDH, Borthakur D. 2020. Mimosine facilitates metallic cation uptake by plants through formation of mimosine-cation complexes. *Plant Molecular Biology* 102:431–45

157. Borthakur D, Soedarjo M, Fox PM, Webb DT. 2003. The *mid* genes of *Rhizobium* sp strain TAL1145 are required for degradation of mimosine into 3-hydroxy-4-pyridone and are inducible by mimosine. *Microbiology* 149:537–46

158. Negi VS, Bingham JP, Li QX, Borthakur D. 2013. *midD*-encoded 'rhizomimosinase' from *Rhizobium* sp. strain TAL1145 is a C-N lyase that catabolizes L-mimosine into 3-hydroxy-4-pyridone, pyruvate and ammonia. *Amino Acids* 44:1537–47

159. Awaya JD, Fox PM, Borthakur D. 2005. *pyd* genes of Rhizobium sp. strain TAL1145 are required for degradation of 3-hydroxy-4-pyridone, an aromatic intermediate in mimosine metabolism. *Journal of Bacteriology* 187:4480–7

160. Soedarjo M, Borthakur D. 1998. Mimosine, a toxin produced by the tree-legume *Leucaena* provides a nodulation competition advantage to mimosine-degrading *Rhizobium* strains. *Soil Biology and Biochemistry* 30:1605–13

161. Soedarjo M, Hemscheidt TK, Borthakur D. 1994. Mimosine, a toxin present in leguminous trees (*Leucaena* spp.), induces a mimosine-degrading enzyme activity in some *Rhizobium* strains. *Applied and Environmental Microbiology* 60:4268–72

162. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, et al.

2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596–604

163. Marks RA, Hotaling S, Frandsen PB, VanBuren R. 2021. Representation and participation across 20 years of plant genome sequencing. *Nature Plants* 7:1571–78

164. Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, et al. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology* 37:1155–62

165. Neale DB, Zimin AV, Zaman S, Scott AD, Shrestha B, et al. 2022. Assembled and annotated 26.5 Gbp coast redwood genome: a resource for estimating evolutionary adaptive potential and investigating hexaploid origin. *G3 Genes|Genomes|Genetics* 12:jkab380

166. Wang C, Liu C, Roqueiro D, Grimm D, Schwab R, et al. 2015. Genome-wide analysis of local chromatin packing in Arabidopsis thaliana. *Genome Research* 25:246–56

167. Michael TP, Bryant D, Gutierrez R, Borisjuk N, Chu P, et al. 2017. Comprehensive definition of genome features in *Spirodela polyrhiza* by high-depth physical mapping and short-read DNA sequencing strategies. *The Plant Journal* 89:617–35

168. Niu S, Li J, Bo W, Yang W, Zuccolo A, et al. 2022. The Chinese pine genome and methylome unveil key features of conifer evolution. *Cell* 185:204–217.E14

169. Cazzolla Gatti R, Reich PB, Gamarra JGP, Crowther T, Hui C, et al. 2022. The number of tree species on Earth. *PNAS* 119:.e2115329119

170. Chen H, Zeng Y, Yang Y, Huang L, Tang B, et al. 2020. Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nature Communications* 11:2494

171. Koren S, Rhie A, Walenz BP, Dilthey AT, Bickhart DM, et al. 2018. *De novo* assembly of haplotype-resolved genomes with trio binning. *Nature Biotechnology* 36:1174–82

172. Zhang X, Zhang S, Zhao Q, Ming R, Tang H. 2019. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nature Plants* 5:833–45

173. Kronenberg ZN, Rhie A, Koren S, Concepcion GT, Peluso P, et al. 2021. Extended haplotype-phasing of long-read de novo genome assemblies using Hi-C. *Nature Communications* 12:1935

174. Zhang Q, Qi Y, Pan H, Tang H, Wang G, et al. 2022. Genomic insights into the recent chromosome reduction of autopolyploid sugarcane *Saccharum spontaneum*. *Nature Genetics* 54:885–96

175. Sun H, Jiao W, Krause K, Campoy JA, Goel M, et al. 2022. Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nature Genetics* 54:342–48

176. Zhang S, Li R, Zhang L, Chen S, Xie M, et al. 2020. New insights into *Arabidopsis* transcriptome complexity revealed by direct sequencing of native RNAs. *Nucleic Acids Research* 48:7700–11

177. Ji X, Li P, Fuscoe JC, Chen G, Xiao W, et al. 2020. A comprehensive rat transcriptome built from large scale RNA-seq-based annotation. *Nucleic Acids Research* 48:8320–31

178. Liu H, Wang X, Wang G, Cui P, Wu S, et al. 2021. The nearly complete genome of Ginkgo biloba illuminates gymnosperm evolution. *Nature Plant* 7:748–56

179. Cooke JEK, Eriksson ME, Junttila O. 2012. The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. *Plant, Cell & Environment* 35:1707–28

180. Triozzi PM, Ramos-Sánchez JM, Hernández-Verdeja T, Moreno-Cortés A, Allona I, et al. 2018. Photoperiodic regulation of shoot apical growth in poplar. *Frontiers in Plant Science* 9:1030

181. Lloret A, Badenes ML, Ríos G. 2018. Modulation of dormancy and growth responses in reproductive buds of temperate trees. *Frontiers in Plant Science* 9:1368

182. Brunner AM, Evans LM, Hsu CY, Sheng X. 2014. Vernalization and the chilling requirement to exit bud dormancy: shared or separate regulation. *Frontiers in Plant Science* 5:732

183. Liu J, Sherif SM. 2019. Hormonal orchestration of bud dormancy cycle in deciduous woody perennials. *Frontiers in Plant Science* 10:1136

184. Yang Q, Gao Y, Wu X, Moriguchi T, Bai S, et al. 2021. Bud endodormancy in deciduous fruit trees: advances and prospects. *Horticulture Research* 8:139

185. Maurya JP, Bhalerao RP. 2017. Photoperiod- and temperature-mediated control of growth cessation and dormancy in trees: a molecular perspective. *Annals of Botany* 120:351–60

186. Maurya JP, Triozzi PM, Bhalerao RP, Perales M. 2018. Environmentally sensitive molecular switches drive poplar phenology. *Frontiers in Plant Science* 9:1873

187. Rohde A, Bhalerao RP. 2007. Plant dormancy in the perennial context. *Trends in Plant Science* 12:217–23

188. Andersson A, Keskitalo J, Sjodin A, Bhalerao R, Sterky F, et al. 2004. A transcriptional timetable of autumn senescence. *Genome Biology* 5:R24

189. Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, et al. 2014. The genome of *Eucalyptus grandis*. *Nature* 510:356–62

190. Chen S, Wang Y, Yu L, Zheng T, Wang S, et al. 2021. Genome sequence and evolution of *Betula platyphylla*. *Horticulture Research* 8:37

191. Plomion C, Aury JM, Amselem J, Leroy T, Murat F, et al. 2018. Oak genome reveals facets of long lifespan. *Nature Plants* 4:440–52

192. Sollars ESA, Harper AL, Kelly LJ, Sambles CM, Ramirez-Gonzalez RH, et al. 2017. Genome sequence and genetic diversity of European ash trees. *Nature* 541:212–6

193. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, et al. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* 497:579–84

194. Rohde A, Storme V, Jorge V, Gaudet M, Vitacolonna N, et al. 2011. Bud set in poplar - genetic dissection of a complex trait in natural and hybrid populations. *New Phytologist* 189:106–21

195. Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, et al. 2014. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nature Genetics* 46:1089–96

196. Howe GT, Horvath DP, Dharmawardhana P, Priest HD, Mockler TC, et al. 2015. Extensive transcriptome changes during natural onset and release of vegetative bud dormancy in *Populus*. *Frontiers in Plant Science* 6:989

197. Karlberg A, Englund M, Petterle A, Molnár G, Sjödin A, et al. 2010. Analysis of global changes in gene expression during activity-dormancy cycle in hybrid aspen apex. *Plant Biotechnology* 27:1–16

198. Ueno S, Klopp C, Leplé JC, Derory J, Noirot C, et al. 2013. Transcriptional profiling of bud dormancy induction and release in oak by next-generation sequencing. *BMC Genomics* 14:236

199. Nose M, Kurita M, Tamura M, Matsushita M, Hiraoka Y, et al. 2020. Effects of day length- and temperature-regulated genes on annual transcriptome dynamics in Japanese cedar (*Cryptomeria japonica* D. Don), a gymnosperm indeterminate species. *PLoS One* 15:e0229843

200. Wu K, Duan X, Zhu Z, Sang Z, Zhang Y, et al. 2021. Transcriptomic analysis reveals the positive role of abscisic acid in endodormancy maintenance of leaf buds of *Magnolia wufengensis*. *Frontiers in Plant Science* 12:742504

201. Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, et al. 2007. A molecular timetable for apical bud formation and dormancy induction in poplar. *The Plant Cell* 19:2370–90

202. Lesur I, Le Provost G, Bento P, Da Silva C, Leplé JC, 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

203. Santamaría ME, Rodríguez R, Cañal MJ, Toorop PE. 2011. Transcriptome analysis of chestnut (*Castanea sativa*) tree buds suggests a putative role for epigenetic control of bud dormancy. *Annals of Botany* 108:485–98

204. Li W, Kang Y, Zhang Y, Zang Q, Qi L. 2021. Concerted control of

the *LaRAV1-LaCDKB1;3* module by temperature during dormancy release and reactivation of larch. *Tree Physiology* 41:1918–37

205. Hsu CY, Adams JP, No K, Liang H, Meilan R, et al. 2012. Overexpression of *CONSTANS* homologs *CO1* and *CO2* fails to alter normal reproductive onset and fall bud set in woody perennial poplar. *PloS One* 7:e45448-e

206. Yordanov YS, Ma C, Strauss SH, Busov VB. 2014. EARLY BUD-BREAK 1 (*EBB1*) is a regulator of release from seasonal dormancy in poplar trees. *PNAS* 111:10001–6

207. Singh RK, Maurya JP, Azeez A, Miskolczi P, Tylewicz S, et al. 2018. A genetic network mediating the control of bud break in hybrid aspen. *Nature Communications* 9:4173

208. Singh RK, Miskolczi P, Maurya JP, Bhalerao RP. 2019. A Tree ortholog of *SHORT VEGETATIVE PHASE* floral repressor mediates photoperiodic control of bud dormancy. *Current Biology* 29:128–33

209. Conde D, Le Gac AL, Perales M, Dervinis C, Kirst M, et al. 2017. Chilling-responsive DEMETER-LIKE DNA demethylase mediates in poplar bud break. *Plant, Cell & Environment* 40:2236–49

210. Azeez A, Zhao YC, Singh RK, Yordanov YS, Dash M, et al. 2021. EARLY BUD-BREAK 1 and EARLY BUD-BREAK 3 control resumption of poplar growth after winter dormancy. *Nature Communications* 12:1123

211. Anh Tuan P, Bai S, Saito T, Imai T, Ito A, et al. 2016. Involvement of *EARLY BUD-BREAK*, an AP2/ERF transcription factor gene, in bud break in Japanese pear (*Pyrus pyrifolia* Nakai) lateral flower buds: expression, histone modifications and possible target genes. *Plant Cell Physiology* 57:1038–47

212. Csorba T, Questa JI, Sun Q, Dean C. 2014. Antisense *COOLAIR* mediates the coordinated switching of chromatin states at *FLC* during vernalization. *PNAS* 111:16160–5

213. Johnsen Ø, Fossdal CG, Nagy N, Mølmann JA, Dæhlen OG, et al. 2005. Climatic adaptation in Picea abies progenies is affected by the temperature during zygotic embryogenesis and seed maturation. *Plant, Cell & Environment* 28:1090–102

214. Yakovlev I, Fossdal CG, Skrøppa T, Olsen JE, Jahren AH, Johnsen Ø. 2012. An adaptive epigenetic memory in conifers with important implications for seed production. *Seed Science Research* 22:63–76

215. Yakovlev IA, Fossdal CG, Johnsen Ø. 2010. MicroRNAs, the epigenetic memory and climatic adaptation in Norway spruce. *New Phytologist* 187:1154–69

216. Yakovlev IA, Fossdal CG, Johnsen Ø, Junttila O, Skrøppa T. 2006. Analysis of gene expression during bud burst initiation in Norway spruce via ESTs from subtracted cDNA libraries. *Tree Genetics & Genomes* 2:39–52

217. Du Q, Lu W, Quan M, Xiao L, Song F, et al. 2018. Genome-wide association studies to improve wood properties: challenges and prospects. *Frontiers in Plant Science* 9:1912

218. Resende RT, Resende MDV, Silva FF, Azevedo CF, Takahashi EK, et al. 2017. Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in *Eucalyptus*. *New Phtologist* 213:1287–300

219. Lv C, Lu W, Quan M, Xiao L, Li L, et al. 2021. Pyramiding superior haplotypes and epistatic alleles to accelerate wood quality and yield improvement in poplar breeding. *Industrial Crops and Products* 171:113891

220. Bartholomé J, Mandrou E, Mabiala A, Jenkins J, Nabihoudine I, et al. 2015. High-resolution genetic maps of Eucalyptus improve Eucalyptus grandis genome assembly. *New Phtologist* 206:1283–96

221. Wu Y, Close TJ, Lonardi S. 2011. Accurate construction of consensus genetic maps via integer linear programming. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 8:381–94

222. Rastas P. 2017. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* 33:3726–32

223. Qian Z, Zhang B, Chen H, Lu L, Duan M, et al. 2021. Identification of quantitative trait loci controlling the development of prickles in eggplant by genome re-sequencing analysis. *Frontiers in Plant Science* 12:731079

224. Li Z, Xu Y. 2022. Bulk segregation analysis in the NGS era: a review of its teenage years. *The Plant Journal* 109:1355–74

225. Zhuo X, Zheng T, Li S, Zhang Z, Zhang M, et al. 2021. Identification of the *PmWEEP* locus controlling weeping traits in *Prunus mume* through an integrated genome-wide association study and quantitative trait locus mapping. *Horticulture Research* 8:131

226. Gattolin S, Cirilli M, Pacheco I, Ciacciulli A, Da Silva Linge C, et al. 2018. Deletion of the miR172 target site in a TOE-type gene is a strong candidate variant for dominant double-flower trait in Rosaceae. *The Plant Journal* 96:358–71

227. An Z, Zhao Y, Zhang X, Huang X, Hu Y, et al. 2019. A high-density genetic map and QTL mapping on growth and latex yield-related traits in *Hevea brasiliensis* Müll. Arg. *Industrial Crops and Products* 132:440–48

228. Du Q, Gong C, Wang Q, Zhou D, Yang H, et al. 2016. Genetic architecture of growth traits in *Populus* revealed by integrated quantitative trait locus (QTL) analysis and association studies. *New Phtologist* 209:1067–82

229. Du Q, Yang X, Xie J, Quan M, Xiao L, et al. 2019. Time-specific and pleiotropic quantitative trait loci coordinately modulate stem growth in *Populus*. *Plant Biotechnology Journal* 17:608–24

230. Porth I, Klapste J, Skyba O, Hannemann J, McKown AD, et al. 2013. Genome-wide association mapping for wood characteristics in *Populus* identifies an array of candidate single nucleotide polymorphisms. *New Phytologist* 200:710–26

231. Fahrenkrog AM, Neves LG, Resende MFR Jr, Vazquez AI, de Los Campos G, et al. 2017. Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*. *New Phtologist* 213:799–811

232. Quan M, Liu X, Du Q, Xiao L, Lu W, et al. 2021. Genome-wide association studies reveal the coordinated regulatory networks underlying photosynthesis and wood formation in Populus. *Journal of Experimental Botany* 72:5372–89

233. Müller BSF, de Almeida Filho JE, Lima BM, Garcia CC, Missiaggia A, et al. 2019. Independent and Joint-GWAS for growth traits in Eucalyptus by assembling genome-wide data for 3373 individuals across four breeding populations. *New Phtologist* 221:818–33

234. De La Torre AR, Puiu D, Crepeau MW, Stevens K, Salzberg SL, et al. 2019. Genomic architecture of complex traits in loblolly pine. *New Phtologist* 221:1789–801

235. Chen M, Fan W, Ji F, Hua H, Liu J, et al. 2021. Genome-wide identification of agronomically important genes in outcrossing crops using OutcrossSeq. *Molecular Plant* 14:556–70

236. Gong C, Du Q, Xie J, Quan M, Chen B, et al. 2018. Dissection of insertion-deletion variants within differentially expressed genes involved in wood formation in *Populus*. *Frontiers in Plant Science* 8:2199

237. Valenzuela CE, Ballesta P, Ahmar S, Fiaz S, Heidari P, et al. 2021. Haplotype- and SNP-based GWAS for growth and wood quality traits in *Eucalyptus cladocalyx* trees under arid conditions. *Plants* 10:148

238. Ding X, Diao S, Luan Q, Wu HX, Zhang Y, et al. 2022. A transcriptome-based association study of growth, wood quality, and oleoresin traits in a *slash pine* breeding population. *PloS Genetics* 18:e1010017

239. Francisco FR, Aono AH, da Silva CC, Gonçalves PS, Scaloppi Junior EJ, et al. 2021. Unravelling Rubber Tree Growth by Integrating GWAS and Biological Network-Based Approaches. *Frontiers in Plant Science* 12:768589

240. Sarić R, Nguyen VD, Burge T, Berkowitz O, Trtílek M, et al. 2022. Applications of hyperspectral imaging in plant phenotyping. *Trends in Plant Science* 27:301–15

241. Sawitri, Tani N, Na'iem M, Widiyatno, Indrioko S, et al. 2020.

Potential of genome-wide association studies and genomic selection to improve productivity and quality of commercial timber species in tropical rainforest, a case study of *Shorea platyclados*. *Forests* 11:239

242. Chhatre VE, Fetter KC, Gougherty AV, Fitzpatrick MC, Soolanayakanahally RY, et al. 2019. Climatic niche predicts the landscape structure of locally adaptive standing genetic variation. *bioRxiv* Preprint

243. Guzella TS, Dey S, Chelo IM, Pino-Querido A, Pereira VF, et al. 2018. Slower environmental change hinders adaptation from standing genetic variation. *PLoS Genetics* 14:e1007731

244. Filipe JC, Rymer PD, Byrne M, Hardy G, Mazanec R, et al. 2022. Signatures of natural selection in a foundation tree along Mediterranean climatic gradients. *Molecular Ecology* 31:1735−52

245. Wang J, Ding J, Tan B, Robinson KM, Michelson IH, et al. 2018. A major locus controls local adaptation and adaptive life history variation in a perennial plant. *Genome Biology* 19:72

246. Gugger PF, Fitz-Gibbon ST, Albarrán-Lara A, Wright JW, Sork VL. 2021. Landscape genomics of *Quercus lobata* reveals genes involved in local climate adaptation at multiple spatial scales. *Molecular Ecology* 30:406−23

247. Lu W, Xiao L, Quan M, Wang Q, El-Kassaby YA, et al. 2020. Linkage-linkage disequilibrium dissection of the epigenetic quantitative trait loci (epiQTLs) underlying growth and wood properties in *Populus*. *New Phtologist* 225:1218−33

248. Ci D, Song Y, Du Q, Tian M, Han S, Zhang D. 2016. Variation in genomic methylation in natural populations of *Populus simonii* is associated with leaf shape and photosynthetic traits. *Journal of Experimental Botany* 67:723−37

249. Ma K, Sun L, Cheng T, Pan H, Wang J, Zhang Q. 2018. Epigenetic Variance, Performing Cooperative Structure with Genetics, Is Associated with Leaf Shape Traits in Widely Distributed Populations of Ornamental Tree *Prunus mume*. *Frontiers in Plant Science* 9:41

250. Ong-Abdullah M, Ordway JM, Jiang N, Ooi SE, Kok SY, et al. 2015. Loss of *Karma* transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* 525:533−7

251. Sáez-Laguna E, Guevara MÁ, Díaz LM, Sánchez-Gómez D, Collada C, et al. 2014. Epigenetic variability in the genetically uniform forest tree species *Pinus pinea* L. *PLoS One* 9:e103145

252. Gugger PF, Fitz-Gibbon S, PellEgrini M, Sork VL. 2016. Species-wide patterns of DNA methylation variation in *Quercus lobata* and their association with climate gradients. *Molecular Ecology* 25:1665−80

253. Xu G, Lyu J, Li Q, Liu H, Wang D, et al. 2020. Evolutionary and functional genomics of DNA methylation in maize domestication and improvement. *Nature Communications* 11:5539

254. Xu J, Chen G, Hermanson PJ, Xu Q, Sun C, et al. 2019. Population-level analysis reveals the widespread occurrence and phenotypic consequence of DNA methylation variation not tagged by genetic variation in maize. *Genome Biology* 20:243

255. Eichten SR, Briskine R, Song J, Li Q, Swanson-Wagner R, et al. 2013. Epigenetic and genetic influences on DNA methylation variation in maize populations. *The Plant Cell* 25:2783−97

256. Motte H, Vercauteren A, Depuydt S, Landschoot S, Geelen D, et al. 2014. Combining linkage and association mapping identifies *RECEPTOR-LIKE PROTEIN KINASE1* as an essential Arabidopsis shoot regeneration gene. *PNAS* 111:8305−10

257. Willi Y, Kristensen TN, Sgrò CM, Weeks AR, Ørsted M, et al. 2022. Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. *PNAS* 119:e2105076119

258. Wright S. 1943. Isolation by distance. *Genetics* 28:114−38

259. Zhao W, Sun YQ, Pan J, Sullivan AR, Arnold ML, et al. 2020. Effects of landscapes and range expansion on population structure and local adaptation. *New Phytologist* 228:330−43

260. Excoffier L, Foll M, Petit RJ. 2009. Genetic consequences of range expansions. *Annual Review of Ecology Evolution and Systematics* 40:481−501

261. Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. 2013. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology* 22:5983−99

262. Wang IJ, Bradburd GS. 2014. Isolation by environment. *Molecular Ecology* 23:5649−62

263. Xia H, Wang B, Zhao W, Pan J, Mao J, et al. 2018. Combining mitochondrial and nuclear genome analyses to dissect the effects of colonization, environment, and geography on population structure in *Pinus tabuliformis*. *Evolutionary Applications* 11:1931−45

264. Nadeau S, Meirmans PG, Aitken SN, Ritland K, Isabel N. 2016. The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. *Ecology and Evolution* 6:8649−64

265. Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, et al. 2016. Convergent local adaptation to climate in distantly related conifers. *Science* 353:1431−33

266. Chen J, Li L, Milesi P, Jansson G, Berlin M, et al. 2019. Genomic data provide new insights on the demographic history and the extent of recent material transfers in Norway spruce. *Evolutionary Applications* 12:1539−51

267. Eckert AJ, Wegrzyn JL, Liechty JD, Lee JM, Cumbie WP, et al. 2013. The evolutionary genetics of the genes underlying phenotypic associations for loblolly pine (*Pinus taeda*, Pinaceae). *Genetics* 195:1353−72

268. Grivet D, Avia K, Vaattovaara A, Eckert AJ, Neale DB, et al. 2017. High rate of adaptive evolution in two widespread European pines. *Molecular Ecology* 26:6857−70

269. Hall D, Olsson J, Zhao W, Kroon J, Wennström U, et al. 2021. Divergent pattern between phenotypic and genetic variation in Scots pine. *Plant Communications* 2:100139

270. Rellstab C, Zoller S, Walthert L, Lesur I, Pluess AR, et al. 2016. Signatures of local adaptation in candidate genes of oaks (*Quercus* spp. ) with respect to present and future climatic conditions. *Molecular Ecology* 25:5907−24

271. Jia KH, Zhao W, Maier PA, Hu XG, Jin YQ, et al. 2020. Landscape genomics predicts climate change-related genetic offset for the widespread *Platycladus orientalis* (Cupressaceae). *Evolutionary Applications* 13:665−76

272. Guo J, Wang B, Liu Z, Mao J, Wang X, et al. 2022. Low genetic diversity and population connectivity fuelvulnerability to climate change for the Tertiary relict pinePinus bungeana. *Journal of Systematics and Evolution* 00:1−14

273. Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, et al. 2016. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist* 188:379−97

274. Barton N, Hermisson J, Nordborg M. 2019. Population Genetics: Why structure matters. *eLife* 8:e45380

275. Gienapp P. 2020. Is gene mapping in wild populations useful for understanding and predicting adaptation to global change. *Global Change Biology* 26:2737−49

276. Fitzpatrick MC, Keller SR. 2015. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters* 18:1−16

277. Bay RA, Harrigan RJ, Underwood VL, Gibbs HL, Smith TB, Ruegg K. 2018. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* 359:83−6

278. Gougherty AV, Keller SR, Fitzpatrick MC. 2021. Maladaptation, migration and extirpation fuel climate change risk in a forest tree species. *Nature Climate Change* 11:166−71

279. Láruson ÁJ, Fitzpatrick MC, Keller SR, Haller BC, Lotterhos KE. 2022. Seeing the Forest for the trees: Assessing genetic offset predictions from Gradient Forest. *Evolutionary Applications* 15:403−16

280. Aitken SN, Whitlock MC. 2013. Assisted gene flow to facilitate

local adaptation to climate change. *Annual Review of Ecology Evolution and Systematics* 44:367−88

281. Ruotsalainen S. 2014. Increased forest production through forest tree breeding. *Scandinavian Journal of Forest Research* 29:49

282. Butler BJ, Wear DN. 2013. Forest Ownership Dynamics of Southern Forests. In *The Southern Forest Futures Project: Technical Report*, eds. Wear DN, Greispp JG. USA: USDA Forest Service, Southern Research Station. pp.103−22

283. Meuwissen T, Hayes B, Goddard M. 2016. Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers* 6:6−14

284. Isik F. 2014. Genomic selection in forest tree breeding: the concept and an outlook to the future. *New Forests* 45:379−401

285. Caballero M, Lauer E, Bennett J, Zaman S, McEvoy S, et al. 2021. Toward genomic selection in *Pinus taeda*: Integrating resources to support array design in a complex conifer genome. *Applications in Plant Sciences* 2021:e11439

286. Jackson C, Christie N, Reynolds SM, Marais GC, Tii-Kuzu Y, et al. 2022. A genome-wide SNP genotyping resource for tropical pine tree species. *Molecular Ecology Resources* 22:695−710

287. Graham N, Telfer E, Frickey T, Slavov G, Ismael A, et al. 2022. Development and validation of a 36K SNP array for radiata pine (*Pinus Radiata* D. Don). *Forests* 13:176

288. Silva-Junior OB, Faria DA, 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 Phtologist* 206:1527−40

289. Grattapaglia D, Silva-Junior OB, Resende RT, Cappa EP, Müller BSF, et al. 2018. Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in Plant Science* 9:1693

290. Lebedev VG, Lebedeva TN, Chernodubov AI, Shestibratov K. 2020. Genomic selection for forest tree improvement: methods, achievements and perspectives. *Forests* 11:1190

291. Mphahlele MM, Isik F, Mostert-O'Neill MM, Reynolds SM, Hodge GR, et al. 2020. Expected benefits of genomic selection for growth and wood quality traits in *Eucalyptus grandis*. *Tree Genetics & Genomes* 16:49

292. Chamberland V, Robichaud F, Perron M, Gélinas N, Bousquet J, et al. 2020. Conventional versus genomic selection for white spruce improvement: a comparison of costs and benefits of plantations on Quebec public lands. *Tree Genetics & Genomes* 16:17

293. Telfer E, Graham N, Macdonald L, Li Y, Klápště J, et al. 2019. A high-density exome capture genotype-by-sequencing panel for forestry breeding in *Pinus radiata*. *PLoS One* 14:e0222640

294. Wegrzyn JL, Liechty JD, Stevens KA, Wu LS, Loopstra CA, et al. 2014. Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* 196:891−909

295. Li X, Gunasekara C, Guo Y, Zhang H, Lei L, et al. 2014. Pop's Pipes: poplar gene expression data analysis pipelines. *Tree Genetics & Genomes* 10:1093−101

296. Chen T, He HL, Church GM. 1999. Modeling gene expression with differential equations. *Pacific Symposium on Biocomputing* 1999:29−40

297. Ruklisa D, Brazma A, Viksna J. 2005. Reconstruction of gene regulatory networks under the finite state linear model. *Genome Information* 16:225−36

298. Dojer N, Gambin A, Mizera A, Wilczyński B, Tiuryn J. 2006. Applying dynamic Bayesian networks to perturbed gene expression data. *BMC Bioinformatics* 7:249

299. Louis M, Becskei A. 2002. Binary and graded responses in gene networks. *Science's STKE* 43:PE33

300. Kauffman S. 1969. Homeostasis and differentiation in random genetic control networks. *Nature* 224:177−78

301. Chen BS, Chang CH, Wang YC, Wu CH, Lee HC. 2011. Robust model matching design methodology for a stochastic synthetic gene network. *Mathematical Biosciences* 230:23−36

302. Schäfer J, Strimmer K. 2005. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* 21:754−64

303. Butte AJ, Kohane IS. 2000. Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. *Proceedings of Pacific Symposium on Biocomputing* 5:418−29

304. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, et al. 2006. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7:S7

305. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, et al. 2007. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biology* 5:e0050008

306. Altay G, Emmert-Streib F. 2010. Inferring the conservative causal core of gene regulatory networks. *BMC Systems Biology* 4:132

307. Luo W, Hankenson KD, Woolf PJ. 2008. Learning transcriptional regulatory networks from high throughput gene expression data using continuous three-way mutual information. *BMC Bioinformatics* 9:467

308. Lin YC, Li W, Sun YH, Kumari S, Wei H, et al. 2013. SND1 transcription factor-directed quantitative functional hierarchical genetic regulatory network in wood formation in *Populus trichocarpa*. *The Plant Cell* 25:4324−41

309. Wei M, Liu Q, Wang Z, Yang J, Li W, et al. 2020. PuHox52-mediated hierarchical multilayered gene regulatory network promotes adventitious root formation in *Populus ussuriensis*. *New Phytologist* 228:1369−85

310. Wu W, Li J, Wang Q, Lv K, Du K, et al. 2021. Growth-regulating factor 5 (GRF5)-mediated gene regulatory network promotes leaf growth and expansion in poplar. *New Phytologist* 230:612−28

311. Lu S, Li Q, Wei H, Chang MJ, Tunlaya-Anukit S, et al. 2013. Ptr-miR397a is a negative regulator of laccase genes affecting lignin content in Populus trichocarpa. *Proceedings of the National Academy of Sciences The United States of America* 110:10848−53

312. Lv K, Wu W, Wei H, Liu G. 2021. A systems biology approach identifies a regulator, *BplERF1*, of cold tolerance in *Betula platyphylla*. *Forestry Research* 1:11

313. Kumari S, Deng W, Gunasekara C, Chiang V, Chen HS, et al. 2016. Bottom-up GGM algorithm for constructing multilayered hierarchical gene regulatory networks that govern biological pathways or processes. *BMC Bioinformatics* 17:132

314. Deng W, Zhang K, Busov V, Wei H. 2017. Recursive random forest algorithm for constructing multilayered hierarchical gene regulatory networks that govern biological pathways. *PloS ONE* 12:e0171532

315. Wei H. 2019. Construction of a hierarchical gene regulatory network centered around a transcription factor. *Briefings in Bioinformatics* 20:1021−31

316. Nie J, Stewart R, Zhang H, Thomson JA, Ruan F, et al. 2011. TF-Cluster: a pipeline for identifying functionally coordinated transcription factors via network decomposition of the shared coexpression connectivity matrix (SCCM). *BMC Systems Biology* 5:53

317. Ji X, Chen S, Li J, Deng W, Wei Z, et al. 2017. SSGA and MSGA: two seed-growing algorithms for constructing collaborative subnetworks. *Scientific Reports* 7:1446

318. Deng W. 2018. *Algorithms for reconstruction of gene regulatory networks from high -throughput gene expression data*. Thesis. Michigan Technological University, USA. 101 pp.