

OPINION

Open access to tree genomes: the path to a better forest

David B Neale^{1,*}, Charles H Langley², Steven L Salzberg³ and Jill L Wegrzyn¹

Abstract

An open-access culture and a well-developed comparative-genomics infrastructure must be developed in forest trees to derive the full potential of genome sequencing in this diverse group of plants that are the dominant species in much of the earth's terrestrial ecosystems.

Keywords: Forest tree genome, Open access, Sequencing, Genomics, Database

Opportunities and challenges in forest tree genomics are seemingly as diverse and as large as the trees themselves; however, here, we have chosen to focus on the potential significant impact on all of tree biology research if only an open-access culture and comparative-genomics infrastructure were developed. In earlier articles [1,2], we argued that the great diversity of forest trees found in both the undomesticated and domesticated state provides an excellent opportunity to understand the molecular basis of adaptation in plants and furthermore that comparative-genomic approaches will greatly facilitate discovery and understanding. We identified several priority research areas towards realizing these goals (Box 1), such as establishing reference genome sequences for important tree species, determining how to apply sequencing technologies to understand adaptation, and developing resources for storing and accessing forestry data. Significant progress has been made in many of these priorities, with the exception of investments in database resources and understanding ecological functions. Here, we briefly summarize the rapid progress in developing genomic resources in a small number of species and then offer our view on what we believe it will take to realize the final two priorities.

*Correspondence: dbneale@ucdavis.edu

¹Department of Plant Sciences, University of California, Davis, CA 95616, USA
Full list of author information is available at the end of the article

The great diversity found in forest trees

There are an estimated 60,000 tree species on earth, and approximately 30 of the 49 plant orders contain tree species. Clearly, the tree phenotype has evolved many times in plants. The diversity of plant structures, development, life history, environments occupied and so on in trees is nearly as broad as higher plants in general, but trees share the common characteristic that all are perennial and many are very long lived. Because of the sessile nature of plants, each tree must survive and reproduce in a specific environment over the seasonal cycles of its lifetime. This tight association between individual genotypes and their environment provides a powerful research setting, just as it has driven the evolution of a plethora of uniquely arboreal adaptations. Understanding these evolutionary strategies is a long-standing area of study of tree biologists, with many broader biological implications.

Completed and current genome-sequencing projects in forest trees are limited to about 25 species from just 4 of more than 100 families: Pinaceae (pines, spruces and firs), Salicaceae (poplars and willows), Myrtaceae (eucalyptus) and Fagaceae (oaks, chestnuts and beeches). Large-scale sequencing projects such as the 1000 Human Genomes [3], 1000 Plant Genomes (1KP) [4] or the 5000 Insect Genome (i5k) [5] projects have not yet been proposed for forest trees.

Rapidly developing genomic resources in forest trees

Genome resources are developing rapidly in forest trees in spite of the challenges associated with working with large, long-lived organisms and sometimes very large genomes [2]. Complete genome sequencing, however, has been slow to advance in forest trees owing to funding limitations and the large size of conifer genomes. Black cottonwood (*Populus trichocarpa* Torr. & Gray) was the first forest tree genome to be sequenced by the US Department of Energy Joint Genome Institute (DOE/JGI) [6] (Table 1). Black cottonwood has a relatively small genome (450 Mb) and is a target feedstock species for cellulosic ethanol production, and thus fits into the DOE/JGI priority of sequencing bioenergy feedstock species.

Box 1. Research priorities in forest tree genomics identified in earlier *Opinion* papers

From Neale and Ingvarsson [1]:

- Deep expressed-sequence tag (EST) sequencing in many species
- Comparative resequencing in many species
- Reference genome sequence for pine

From Neale and Kremer [2]:

- Reference genome sequences for several important species
- Greater investment in diverse species towards understanding ecological function
- Application of next-generation sequencing technologies to understand adaptation using landscape genomic approaches
- Greater investment in database resources and cyber-infrastructure development
- Development of new and high-throughput phenotyping technologies

The genus *Populus* has 30+ species (aspens and cottonwoods) with genome sizes of approximately 500 Mb. Several species are being sequenced by DOE/JGI, and other groups around the world, and it seems likely that all members of the genus will soon have a genome sequence (Table 1). The next forest tree to be sequenced was the flooded gum (*Eucalyptus grandis* BRASUZ1, which is a member of the Myrtaceae family), again by DOE/JGI. *Eucalyptus* species and their hybrids are important commercial species grown in their native Australia and many regions throughout the southern hemisphere. Several more eucalyptus species are being sequenced (Table 1), each with relatively small genomes (500 Mb), but it will probably take many years before all 700+ members of this genus are completed. Several members of the Fagaceae family are now being sequenced (Table 1). Members of this group include the oaks, beeches and chestnuts, with genome sizes less than 1 Gb.

The gymnosperm forest trees (such as the conifers) were the last to enter the world of genome sequencing. This was entirely due to their very large genomes (10 Gb and greater) as they are extremely important economically and ecologically, and phylogenetically they represent the ancient sister lineage to that of angiosperm species. Genome resources needed to support a sequencing project were reasonably well developed, but it was not until the introduction of next-generation sequencing (NGS) technologies that sequencing conifer genomes became tractable. Currently, there are at least ten conifer (Pinaceae) genome-sequencing projects under way (Table 1).

Aside from reference genome sequencing in forest trees, there is significant activity in transcriptome

sequencing and resequencing for polymorphism discovery (Tables 2 and 3). We have only listed the transcriptome and resequencing projects in Table 1 that are associated with a species that has an active genome-sequencing project.

The opportunity for comparative-genomic approaches in forest trees

The power of comparative-genomic approaches for understanding function in an evolutionary framework is well established [7-13]. Comparative genomics can be applied to sequence data (nucleotide and protein) at the level of individual genes or genome-wide. Genome-wide approaches provide insight into both chromosome evolution and the diversification of biological functions and interactions.

Understanding of gene function in forest tree species is challenged by the lack of standard reverse-genetic tools routinely used in other systems - for example, standard marker stocks, facile transformation and regeneration - and by the long generation times. Thus, comparative genomics becomes the more powerful approach to understanding gene function in trees.

Comparative genomics requires not only data availability but also cyber-infrastructure to support exchange and analysis. The TreeGenes database is the most comprehensive resource for comparative-genomic analyses in forest trees [14]. Several smaller databases have been created to facilitate collaborations, including: Fagaceae genomics web, hardwoodgenomics.org, Quercus portal, PineDB, ConiferGDB, EuroPineDB, PopulusDB, PoplarDB, EucalyptusDB and Eucanext (Tables 1, 2, and 3). These resources vary greatly in their scope, relevance and integration. Some are static and archival, whereas others focus on current sequence content for a specific species or a small number of related species. This results in overlapping and conflicting data among repositories. In addition, each database uses its own custom interfaces and back-end database technology to serve sequence to the user. The US National Science Foundation funding for large-scale infrastructure projects, such as iPlant, is leading efforts aimed towards centralizing resources for research communities [15]. Without centralized resources, researchers are forced to employ inefficient data-mining methods through queries of independently maintained databases or inconsistently formatted supplemental files on journal websites. Specific areas of interest for the forest tree genomic community include the ability to connect sequence, genotype and phenotype to individual, geo-referenced trees. This type of integration can only be achieved through web services that allow disparate resources to communicate in ways that are transparent to the user [16]. With the recent increase of genome sequences available for many of these species, there is a

Table 1. Genome resources in forest trees

Family	Genus	Species	Sequence access	Genome	Related publications
Pinaceae	<i>Pinus</i>	<i>taeda</i> (loblolly pine)	[14,19]	Resequenced amplicons	[21,22]
			[20]	BACs	
			[14,19]	Fosmids	
			[23]	Draft genome complete	
	<i>Pinus</i>	<i>lambertiana</i> (sugar pine)	[14,19]	Resequenced amplicons	[24]
			[23]	<i>Draft genome (in progress)</i>	
	<i>Pseudotsuga</i>	<i>menziesii</i> (Douglas-fir)	[14,19]	Resequenced amplicons	[25]
			[23]	<i>Draft genome (in progress)</i>	
	<i>Pinus</i>	<i>sylvestris</i> (Scots pine)	[26]	<i>Draft genome (in progress)</i>	[27]
			<i>Pinus</i>	<i>pinsater</i> (maritime pine)	
	<i>Pinus</i>	<i>sibirica</i> (Siberian pine)			[14,19]
			[28]	<i>Draft genome (in progress)</i>	
	<i>Pinus</i>	<i>radiata</i> (Monterey pine)	[14,19]	Resequenced amplicons	[29]
			[14,19]	Resequenced amplicons	[30]
<i>Picea</i>	<i>glauca</i> (white spruce)	[31]	Draft genome complete (<i>restricted</i>)	[33]	
		[32]	Draft genome complete		
<i>Larix</i>	<i>sibirica</i> (Siberian larch)	[14,19]	BACs	[33]	
		[28]	<i>Draft genome (in progress)</i>		
Salicaceae	<i>Populus</i>	<i>trichocarpa</i> (black cottonwood)	[34]	Genome complete	[6]
			[19]	BACs	[35]
				Genome resequencing (<i>restricted</i>)	[36]
	<i>Populus</i>	<i>tremula</i> (European aspen)	[37]	Draft genome complete	[38]
			<i>Populus</i>	<i>tremula x tremuloides</i>	
	<i>Populus</i>	<i>tremuloides</i> (quaking aspen)			[37]
			[19]	BACs	
	<i>Populus</i>	<i>grandidentata</i> (bigtooth aspen)	[37]	Draft genome complete	[38]
			<i>Populus</i>	<i>nigra</i> (black poplar)	
<i>Salix</i>	<i>purpurea</i> (purpleosier willow)	[40]			Draft genome complete (<i>restricted</i>)
Myrtaceae	<i>Eucalyptus</i>	<i>grandis</i> (rose gum)	[19]	BACs	[41]
			[34]	Draft genome complete	
	<i>Eucalyptus</i>	<i>globulus</i> (blue gum)	[42]	Draft genome (<i>in progress</i>)	[44]
			<i>Eucalyptus</i>	<i>camaldulensis</i> (river red gum)	
	<i>Corymbia</i>	<i>citriodora</i> (lemon-scented gum)			[45]
Fagaceae	<i>Quercus</i>	<i>robur</i> (English oak)	[19]	BACs	[46,47]
			[48]	<i>Draft genome (in progress)</i>	
	<i>Castanea</i>	<i>mollissima</i> (Chinese chestnut)	[49]	<i>Draft genome (in progress)</i>	[51]
			[50]	BACs	
Betulaceae	<i>Betula</i>	<i>nana</i> (dwarf birch)	[52]	Draft genome complete	[53]
Oleaceae	<i>Fraxinus</i>	<i>excelsior</i> (European ash)	[54]	Draft genome complete	

Details current genome sequencing projects in forest trees with sequence access information and relevant publications.

Table 2. Transcriptome resources in forest trees

Family	Genus	Species	Sequence access	Transcriptome	Related publications
Pinaceae	<i>Pinus</i>	<i>taeda</i> (loblolly pine)	[14,19,55]	EST sequencing (Sanger)	[56-59]
			[14,19,60]	EST sequencing (454)	[60]
			[19]	Exome resequencing	[61]
	<i>Pinus</i>	<i>lambertiana</i> (sugar pine)	[14,19,60]	EST sequencing (454)	[60]
	<i>Pseudotsuga</i>	<i>menziesii</i> (Douglas-fir)	[14,19]	EST sequencing (Sanger)	
			[14,19,62]	EST sequencing (454)	[60,63,64]
	<i>Pinus</i>	<i>sylvestris</i> (Scots pine)	[14,19,55]	EST sequencing (Sanger)	
	<i>Pinus</i>	<i>pinsater</i> (maritime pine)	[14,19,65]	EST sequencing (Sanger/454)	[65,66]
	<i>Pinus</i>	<i>radiata</i> (Monterey pine)	[14,19,55]	EST sequencing (Sanger)	[67-72]
	<i>Picea</i>	<i>abies</i> (Norway spruce)	[14,19,60]	EST sequencing (454)	[60]
[19]			EST sequencing (Next-Gen)	[73]	
[19]			UniGenes (Sanger/454)	[74]	
Salicaceae	<i>Populus</i>	<i>trichocarpa</i> (black cottonwood)	[14,19,75]	EST sequencing (Sanger)	[75-78]
			[19]	Exon capture	[79]
			[14, 19]	UniGenes (Sanger)	[80]
	<i>Populus</i>	<i>tremula</i> (European aspen)	[14,19,75]	EST sequencing (Sanger)	[75,81]
	<i>Populus</i>	<i>tremula x tremuloides</i>	[14,19,75]	EST sequencing (Sanger)	[75,76]
	<i>Populus</i>	<i>tremuloides</i> (quaking aspen)	[14,19]	EST sequencing (Next-Gen)	[82]
Myrtaceae	<i>Eucalyptus</i>	<i>grandis</i> (rose gum)	[14,19]	EST sequencing (Sanger)	[84-86]
			[14,19]	EST sequencing (NextGen)	[87]
	<i>Eucalyptus</i>	<i>globulus</i> (blue gum)	[14,19]	EST sequencing (Next-Gen)	[41,88]
	<i>Eucalyptus</i>	<i>camaldulensis</i> (river red gum)	[14,19]	EST sequencing (RNA-Seq)	[89]
Fagaceae	<i>Quercus</i>	<i>robur</i> (English oak)	[19,90]	EST sequencing (454)	[91]
	<i>Castanea</i>	<i>mollissima</i> (Chinese chestnut)	[50]	EST sequencing (454)	[92,93]
Oleaceae	<i>Fraxinus</i>	<i>excelsior</i> (European ash)	[19]	EST sequencing (454)	[94]

Details current transcriptome sequencing projects in forest trees with sequence access information and relevant publications.

need to facilitate community-level annotation and research support.

The need for a better-developed open-access culture in forest tree genomics research

The Human Genome Project established a culture of open access and data sharing in genomics research for both humans and animal models that has been extended to many other species, including *Arabidopsis*, rat, cow, dog, rice, maize and more than 500 other eukaryotes. Beginning in the late 1990s, these large-scale projects

released data very rapidly to the scientific community, often years before publication. This rapid release of data with few restrictions has allowed thousands of scientists to begin work on specific genes and gene families, and on functional studies, long before the genome papers have appeared. One of the driving motivations for this culture, and the reason that many scientists support it, is that large-scale sequencing can be done most efficiently when centers that have expertise in sequencing technology take the lead. With all the sequencing concentrated, the body of data needs to be shared freely in order to get it in the

Table 3. Polymorphism resources in forest trees

Family	Genus	Species	SNP access	Polymorphism	Related publications
Pinaceae	<i>Pinus</i>	<i>taeda</i> (loblolly pine)	[14,19]	GoldenGate & Infinium iSelect	[95,96]
	<i>Pinus</i>	<i>lambertiana</i> (sugar pine)	[14]	GoldenGate array	[24]
	<i>Pseudotsuga</i>	<i>menziesii</i> (Douglas-fir)	[14]	GoldenGate array	[25]
			[19]	Infinium iSelect	[64]
	<i>Pinus</i>	<i>sylvestris</i> (Scots pine)	[14]	GoldenGate array	
	<i>Pinus</i>	<i>pinsater</i> (maritime pine)	[14]	GoldenGate array	[27,90]
	<i>Pinus</i>	<i>radiata</i> (Monterey pine)	[14]	GoldenGate array	[29]
	<i>Picea</i>	<i>abies</i> (Norway spruce)	[14]	GoldenGate array	[30]
<i>Picea</i>	<i>glauca</i> (white spruce)	[19]	GoldenGate & Infinium iSelect	[97,98]	
Salicaceae	<i>Populus</i>	<i>trichocarpa</i> (black cottonwood)	[19,99]	Infinium iSelect	[100]
				Infinium iSelect (<i>Restricted</i>)	[36]
			[14]	GoldenGate array	[101]
		[19,99]	SNP assay	[102]	
	<i>Populus</i>	<i>nigra</i> (black poplar)	[14]	GoldenGate array	[103]
Myrtaceae	<i>Eucalyptus</i>	<i>grandis</i> (rose gum)	[104]	DArT high-density array	[105]
			[106]	GoldenGate array	[106]
	<i>Eucalyptus</i>	<i>camaldulensis</i> (river red gum)		RNA-Seq SNP discovery (<i>restricted</i>)	[107]

Details current genotyping projects in forest trees with data access information and relevant publications.

hands of the widely distributed experts. This open-access culture has dramatically accelerated scientific progress in biological research.

The path to success avoids delays

Careful inspection of Table 1 reveals that forest tree genome projects are very slow to release sequence data into the public domain. Once a project is finished and submitted for publication, a draft genome becomes available - for example, the poplar genome was released and published in 2006. However, pre-publication releases are infrequent, exceptions being the PineRefSeq project that has made three releases and the SMarTForest project that has made one (Table 1). This is unfortunate because good-quality sequence contigs and scaffolds could be made available years before publication, delivering an extremely important resource to the community. This delay can be understood from privately financed projects seeking commercial advantages, but nearly all the projects listed in Table 1 are financed by public funds whose stated mission is advancing science and development of community resources. Publication rights are easily protected by data-use policy statements such as the Ft Lauderdale [17] and Toronto agreements [18], but

unfortunately these conventions are not often used and data access is restricted by password-protected websites (Tables 1, 2, and 3). We hope the opinion offered here will lead to a discussion in the forest tree community, to a more open-access culture and thus to a more vibrant and rapidly advancing research area.

Competing interests

The authors declare they have no competing interests.

Acknowledgements

Writing of this paper was funded by US Department of Agriculture, National Institute of Food and Agriculture grant #2011-67009-30030.

Abbreviations

EST, expressed-sequence tag; Mb, mega-base; NGS, next-generation sequencing.

Author details

¹Department of Plant Sciences, University of California, Davis, CA 95616, USA. ²Department of Evolution and Ecology, University of California, Davis, CA 95616, USA. ³Center for Computational Biology, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD 21205, USA.

Published: 24 June 2013

References

1. Neale DB, Ingvarsson PK: **Population, quantitative and comparative genomics of adaptation in forest trees.** *Curr Opin Plant Biol* 2008, **11**:149-155.

2. Neale DB, Kremer A: **Forest tree genomics: growing resources and applications.** *Nat Rev Genet* 2011, **12**:111-122.
3. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA: **An integrated map of genetic variation from 1,092 human genomes.** *Nature* 2012, **491**:56-65.
4. **The 1KP Project** [<http://www.onekp.com/>]
5. **i5k Insect and other Arthropod Genome Sequencing Initiative** [<http://arthropodgenomes.org/wiki/i5K>]
6. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, et al.: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray).** *Science* 2006, **313**:1596-1604.
7. Arabidopsis Genome I: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature* 2000, **408**:796-815.
8. Batzoglu S, Pachter L, Mesirov JP, Berger B, Lander ES: **Human and mouse gene structure: comparative analysis and application to exon prediction.** *Genome Res* 2000, **10**:950-958.
9. Zdobnov EM, von Mering C, Letunic I, Torrents D, Suyama M, Copley RR, Christophides GK, Thomasova D, Holt RA, Subramanian GM, Mueller HM, Dimopoulos G, Law JH, Wells MA, Birney E, Charlab R, Halpern AL, Kokoza E, Kraft CL, Lai Z, Lewis S, Louis C, Barillas-Mury C, Nusskern D, Rubin GM, Salzberg SL, Sutton GG, Topalis P, Wides R, Wincker P, et al.: **Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*.** *Science* 2002, **298**:149-159.
10. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renaud H, Wortley EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Haas BJ, Tran AN, Wortman JR, Alsmark UC, Angiuoli S, Anupama A, Badger J, Brington F, Cadag E, Carlton JM, Cerqueira GC, Creasy T, Delcher AL, Djikeng A, Embley TM, Hauser C, Ivens AC, et al.: **Comparative genomics of trypanosomatid parasitic protozoa.** *Science* 2005, **309**:404-409.
11. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, Pollard DA, Sackton TB, Larracuentte AM, Singh ND, Abad JP, Abt DN, Adryan B, Aguade M, Akashi H, Anderson WW, Aquadro CF, Ardell DH, Arguello R, Artieri CG, Barbash DA, Barker D, Barsanti P, Batterham P, Batzoglu S, Begun D, et al.: **Evolution of genes and genomes on the *Drosophila* phylogeny.** *Nature* 2007, **450**:203-218.
12. Koonin EV, Wolf YI: **Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world.** *Nucleic Acids Res* 2008, **36**:6688-6719.
13. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P, Cheng Q, Coulson RM, Crabb BS, Del Portillo HA, Essien K, Feldblyum TV, Fernandez-Becerra C, Gilson PR, Gueye AH, Guo X, Kang'a S, Kooij TW, Korsinczky M, Meyer EV, Nene V, Paulsen I, White O, Ralph SA, Ren Q, Sargeant TJ, et al.: **Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*.** *Nature* 2008, **455**:757-763.
14. Wegrzyn JL, Lee JM, Tearse BR, Neale DB: **TreeGenes: A forest tree genome database.** *Int J Plant Genomics* 2008, **2008**:412875.
15. Goff SA, Vaughn M, McKay S, Lyons E, Stapleton AE, Gessler D, Matasci N, Wang L, Hanlon M, Lenards A, Muir A, Merchant N, Lowry S, Mock S, Helmke M, Kubach A, Narro M, Hopkins N, Micklos D, Hilgert U, Gonzales M, Jordan C, Skidmore E, Dooley R, Cazes J, McLay R, Lu Z, Pasternak S, Koesterke L, Piel WH, et al.: **The iPlant Collaborative: Cyberinfrastructure for Plant Biology.** *Front Plant Sci* 2011, **2**:34.
16. Vasquez-Gross HA, Yu JJ, Figueroa B, Gessler DD, Neale DB, Wegrzyn JL: **CartograTree: connecting tree genomes, phenotypes and environment.** *Mol Ecol Resour* 2013, **13**:528-537.
17. The Wellcome Trust: **Sharing data from large-scale biological research projects: a system of tripartite responsibility** [<http://www.genome.gov/pages/research/wellcomereport0303.pdf>]
18. Toronto International Data Release Workshop A, Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, Rogers J, Harris JR, Ehrlich SD, Apweiler R, Austin CP, Berglund L, Bobrow M, Bountra C, Brookes AJ, Cambon-Thomsen A, Carter NP, Chisholm RL, Contreras JL, Cooke RM, Crosby WL, Dewar K, Durbin R, Dyke SO, Ecker JR, El Emam K, Feuk L, Gabriel SB, Gallacher J, Gelbart WM, et al.: **Prepublication data sharing.** *Nature* 2009, **461**:168-170.
19. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW: **GenBank.** *Nucleic Acids Res* 2013, **41**:D36-D42.
20. **Accelerating Pine Genomics** [<http://www.pine.msstate.edu/>]
21. Kovach A, Wegrzyn JL, Parra G, Holt C, Bruening GE, Loopstra CA, Hartigan J, Yandell M, Langley CH, Korf I, Neale DB: **The *Pinus taeda* genome is characterized by diverse and highly diverged repetitive sequences.** *BMC Genomics* 2010, **11**:420.
22. Magbanua ZV, Ozkan S, Bartlett BD, Chouvarine P, Sasaki CA, Liston A, Cronn RC, Nelson CD, Peterson DG: **Adventures in the enormous: a 1.8 million clone BAC library for the 21.7 Gb genome of loblolly pine.** *PLoS One* 2011, **6**:e16214.
23. **PineRefSeq** [<http://www.pinegenome.org/pinerefseq/>]
24. Jermstad KD, Eckert A, Wegrzyn J, Delfino-Mix A, Davis DA, Burton DC, Neale DB: **Comparative mapping in *Pinus*: sugar pine (*Pinus lambertiana* Dougl.) and loblolly pine (*Pinus taeda* L.).** *Tree Genet Genomes* 2011, **7**:457-468.
25. Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV, St Clair JB, Neale DB: **Association genetics of coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits.** *Genetics* 2009, **182**:1289-1302.
26. **Promoting Conifer Genomic Resources** [<http://bfw.ac.at/rz/bfwcms.web?dok=9020>]
27. Lepoittevin C, Frigerio JM, Garnier-Gere P, Salin F, Cervera MT, Vornam B, Harvengt L, Plomion C: **In vitro vs in silico detected SNPs for the development of a genotyping array: what can we learn from a non-model species?** *PLoS One* 2010, **5**:e11034.
28. **Genome Research and Education center. Siberian Federal University** [<http://genome.sfu-kras.ru/en/main>]
29. Dillon SK, Nolan M, Li W, Bell C, Wu HX, Southerton SG: **Allelic variation in cell wall candidate genes affecting solid wood properties in natural populations and land races of *Pinus radiata*.** *Genetics* 2010, **185**:1477-1487.
30. Chen J, Kallman T, Ma X, Gyllenstrand N, Zaina G, Morgante M, Bousquet J, Eckert A, Wegrzyn J, Neale D, Lagercrantz U, Lascoux M: **Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*).** *Genetics* 2012, **191**:865-881.
31. **ConGenIE** [<http://congenie.org/>]
32. **Index of public *Picea glauca* Release** [ftp://ftp.bcgsc.ca/public/Picea_Glauca/Release_1/]
33. Hamberger B, Hall D, Yuen M, Oddy C, Keeling CI, Ritland C, Ritland K, Bohlmann J: **Targeted isolation, sequence assembly and characterization of two white spruce (*Picea glauca*) BAC clones for terpenoid synthase and cytochrome P450 genes involved in conifer defense reveal insights into a conifer genome.** *BMC Plant Biol* 2009, **9**:106.
34. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS: **Phytozome: a comparative platform for green plant genomics.** *Nucleic Acids Res* 2012, **40**:D1178-D1186.
35. Kelleher CT, Chiu R, Shin H, Bosdet IA, Krzywinski MI, Fjell2 CD, Wilkin J, Yin T, Difazio SP, Ali J, Asano JK, Chan S, Cloutier A, Girn N, Leach S, Lee D, Mathewson CA, Olson T, O'Connor K, Prabhu A-L, Smailus DE, Stott JM, Tsai M: **A physical map of the highly heterozygous *Populus* genome: integration with the genome sequence and genetic map and analysis of haplotype variation.** *Plant J* 2007, **50**:1066-1078.
36. Slavov GT, Difazio SP, Martin J, Schackwitz W, Muchero W, Rodgers-Melnick E, Lipphardt MF, Pennacchio CP, Hellsten U, Pennacchio LA, Gunter LE, Ranjan P, Vining K, Pomraning KR, Wilhelm LJ, Pellegrini M, Mockler TC, Freitag M, Gerales A, El-Kassaby YA, Mansfield SD, Cronk QC, Douglas CJ, Strauss SH, Rokhsar D, Tuskan GA: **Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*.** *New Phytol* 2012, **196**:713-725.
37. **Popgenie draft assemblies** [ftp://v22.popgenie.org/popgenie/UPSC_genomes/UPSC_Draft_Assemblies/]
38. Fladung M, Kaufmann H, Markussen T, Hoenicka H: **Construction of a *Populus tremuloides* Michx. BAC library.** *Silvae Genetica* 2008, **57**:65-69.
39. **IGA External Resources** [<https://services.appliedgenomics.org/>]
40. **Willowpedia** [<http://willow.cals.cornell.edu>]
41. Paiva JA, Prat E, Vautrin S, Santos MD, San-Clemente H, Brommonschenkel S, Fonseca PG, Grattapaglia D, Song X, Ammiraju JS, Kudrna D, Wing RA, Freitas AT, Berges H, Grima-Pettenati J: **Advancing Eucalyptus genomics: identification and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries.** *BMC Genomics* 2011, **12**:137.
42. **EUCAGEN** [<http://web.up.ac.za/eucagen/viewnews.aspx?id=33>]
43. ***Eucalyptus camaldulensis* Genome Database** [<http://www.kazusa.or.jp/eucaly/>]
44. Hirakawa H, Nakamura Y, Kaneko T, Isobe S, Sakai H, Kato T, Hibino T,

- Sasamoto S, Watanabe A, M, Y, S, N, Fujishiro T, Kishida Y, Kohara M, Tabata S, Sato S: Survey of the genetic information carried in the genome of *Eucalyptus camaldulensis*. *Plant Biotechnol J* 2011, **28**:471-480.
45. **Corymbia Genome Project** [http://www.scu.edu.au/scps/index.php/137]
46. Faivre Rampant P, Lesur I, Boussardou C, Bitton F, Martin-Magniette ML, Bodenes C, Le Provost G, Berges H, Fluch S, Kremer A, Plomion C: Analysis of BAC end sequences in oak, a keystone forest tree species, providing insight into the composition of its genome. *BMC Genomics* 2011, **12**:292.
47. Lesur I, Durand J, Sebastiani F, Gyllenstrand N, Bodénès C, Lascoux M, Kremer A, Vendramin GG, Plomion C: A sample view of the pedunculate oak (*Quercus robur*) genome from the sequencing of hypomethylated and random genomic libraries. *Tree Genet Genomes* 2011, **7**:1277-1285.
48. **Quercus portal, a European genetic and genomic web resource for Quercus** [https://w3.pierroton.inra.fr/QuercusPortal/]
49. **The Hardwood Genomics Project** [http://www.hardwoodgenomics.org/]
50. **Fagaceae Genomics Web** [http://www.fagaceae.org/]
51. Fang G, Blackmon B, Staton M, Nelson D, Kubisiak TL, Olukolu BA, Henry D, Zhebentyayeva T, Sasaki CA, Cheng CH, Monsanto M, Ficklin S, Atkins M, Georgi LL, Barakat A, Wheeler N, Carlson J, Sederoff R, Abbott A: A physical map of the Chinese chestnut (*Castanea mollissima*) genome and its integration with the genetic map. *Tree Genet Genomes* 2013, **9**:525-537.
52. **The Dwarf Birch Genome Project** [http://www.birchgenome.org/]
53. Wang N, Thomson M, Bodles WJ, Crawford RM, Hunt HV, Featherstone AW, Pellicer J, Buggs RJ: Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Mol Ecol* 2013, **22**:3098-3111.
54. **The British Ash Tree Genome Project** [http://www.ashgenome.org/]
55. **PineDB** [http://bioinfolab.muohio.edu/txid3352v1/]
56. Allona I, Quinn M, Shoop E, Swope K, St Cyr S, Carlis J, Riedl J, Retzel E, Campbell MM, Sederoff R, Whetten RW: Analysis of xylem formation in pine by cDNA sequencing. *Proc Natl Acad Sci U S A* 1998, **95**:9693-9698.
57. Kirst M, Johnson AF, Baucom C, Ulrich E, Hubbard K, Staggs R, Paule C, Retzel E, Whetten R, Sederoff R: Apparent homology of expressed genes from wood-forming tissues of loblolly pine (*Pinus taeda* L.) with *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 2003, **100**:7383-7388.
58. Cairney J, Zheng L, Cowels A, Hsiao J, Zismann V, Liu J, Ouyang S, Thibaud-Nissen F, Hamilton J, Childs K, Pullman GS, Zhang Y, Oh T, Buell CR: Expressed sequence tags from loblolly pine embryos reveal similarities with angiosperm embryogenesis. *Plant Mol Biol* 2006, **62**:485-501.
59. Lorenz WW, Sun F, Liang C, Kolychev D, Wang H, Zhao X, Cordonnier-Pratt MM, Pratt LH, Dean JF: Water stress-responsive genes in loblolly pine (*Pinus taeda*) roots identified by analyses of expressed sequence tag libraries. *Tree Physiol* 2006, **26**:1-16.
60. Lorenz WW, Ayyampalayam S, Bordeaux JM, Howe GT, Jermstad KD, Neale DB, Rogers D, Dean JF: Conifer DBMagic: a database housing multiple de novo transcriptome assemblies for 12 diverse conifer species. *Tree Genet Genomes* 2012, **8**:1477-1485.
61. Neves LG, Davis JM, Brad Barbazuk W, Kirst M: Whole-exome targeted sequencing of the uncharacterized pine genome. *Plant J* 2013 [Epub ahead of print].
62. **Treiversity** [http://www.treiversity.org/]
63. Muller T, Ensminger I, Schmid KJ: A catalogue of putative unique transcripts from Douglas fir (*Pseudotsuga menziesii*) based on 454 transcriptome sequencing of genetically diverse, drought stressed seedlings. *BMC Genomics* 2012, **13**:673.
64. Howe GT, Yu J, Knaus B, Cronn R, Kolpak S, Dolan P, Lorenz WW, Dean JF: A SNP resource for Douglas-fir: de novo transcriptome assembly and SNP detection and validation. *BMC Genomics* 2013, **14**:137.
65. Fernandez-Pozo N, Canales J, Guerrero-Fernandez D, Villalobos DP, Diaz-Moreno SM, Bautista R, Flores-Monterroso A, Guevara MA, Perdiguero P, Collada C, Cervera MT, Soto A, Ordas R, Canton FR, Avila C, Canovas FM, Claros MG: EuroPineDB: a high-coverage web database for maritime pine transcriptome. *BMC Genomics* 2011, **12**:366.
66. Sancho dos Santos CS, Wilton de Vasconcelos M: Identification of genes differentially expressed in *Pinus pinaster* and *Pinus pinea* after infection with the pine wood nematode. *Eur J Plant Pathol* 2012, **132**:407-418.
67. Walden AR, Walter C, Gardner RC: Genes expressed in *Pinus radiata* male cones include homologs to anther-specific and pathogenesis response genes. *Plant Physiol* 1999, **121**:1103-1116.
68. Li X, Wu HX, Dillon SK, Southerton SG: Generation and analysis of expressed sequence tags from six developing xylem libraries in *Pinus radiata* D. Don. *BMC Genomics* 2009, **10**:41.
69. Li X, Wu HX, Southerton SG: Seasonal reorganization of the xylem transcriptome at different tree ages reveals novel insights into wood formation in *Pinus radiata*. *New Phytol* 2010, **187**:764-776.
70. Li X, Wu HX, Southerton SG: Transcriptome profiling of *Pinus radiata* juvenile wood with contrasting stiffness identifies putative candidate genes involved in microfibril orientation and cell wall mechanics. *BMC Genomics* 2011, **12**:480.
71. Li X, Wu HX, Southerton SG: Transcriptome profiling of wood maturation in *Pinus radiata* identifies differentially expressed genes with implications in juvenile and mature wood variation. *Gene* 2011, **487**:62-71.
72. Li X, Wu HX, Southerton SG: Identification of putative candidate genes for juvenile wood density in *Pinus radiata*. *Tree Physiol* 2012, **32**:1046-1057.
73. Chen J, Uebbing S, Gyllenstrand N, Lagercrantz U, Lascoux M, Kallman T: Sequencing of the needle transcriptome from Norway spruce (*Picea abies* Karst L.) reveals lower substitution rates, but similar selective constraints in gymnosperms and angiosperms. *BMC Genomics* 2012, **13**:589.
74. Rigault P, Boyle B, Lepage P, Cooke JE, Bousquet J, MacKay JJ: A white spruce gene catalog for conifer genome analyses. *Plant Physiol* 2011, **157**:14-28.
75. Sterky F, Bhalerao RR, Unneberg P, Segerman B, Nilsson P, Brunner AM, Charbonnel-Campaa L, Lindvall JJ, Tandre K, Strauss SH, Sundberg B, Gustafsson P, Uhlen M, Bhalerao RP, Nilsson O, Sandberg G, Karlsson J, Lundeberg J, Jansson S: A *Populus* EST resource for plant functional genomics. *Proc Natl Acad Sci U S A* 2004, **101**:13951-13956.
76. Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, Holmberg A, Amini B, Bhalerao R, Larsson M, Villaruel R, Van Montagu M, Sandberg G, Olsson O, Teeri TT, Boerjan W, Gustafsson P, Uhlen M, Sundberg B, Lundeberg J: Gene discovery in the wood-forming tissues of poplar: analysis of 5,692 expressed sequence tags. *Proc Natl Acad Sci U S A* 1998, **95**:13330-13335.
77. Kohler A, Delaruelle C, Martin D, Encelot N, Martin F: The poplar root transcriptome: analysis of 7000 expressed sequence tags. *FEBS Lett* 2003, **542**:37-41.
78. Dejardin A, Leple JC, Lesage-Descauses MC, Costa G, Pilate G: Expressed sequence tags from poplar wood tissues—a comparative analysis from multiple libraries. *Plant Biol (Stuttg)* 2004, **6**:55-64.
79. Zhou L, Holliday JA: Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture. *BMC Genomics* 2012, **13**:703.
80. Ralph SG, Chun HJ, Cooper D, Kirkpatrick R, Kolosova N, Gunter L, Tuskan GA, Douglas CJ, Holt RA, Jones SJ, Marra MA, Bohlmann J: Analysis of 4,664 high-quality sequence-finished poplar full-length cDNA clones and their utility for the discovery of genes responding to insect feeding. *BMC Genomics* 2008, **9**:57.
81. Bhalerao R, Keskitalo J, Sterky F, Erlandsson R, Bjorkbacka H, Birve SJ, Karlsson J, Gardestrom P, Gustafsson P, Lundeberg J, Jansson S: Gene expression in autumn leaves. *Plant Physiol* 2003, **131**:430-442.
82. Rai H, Mock K, Richardson B, Cronn R, Hayden K, Wright J, Knaus B, Wolf P: Transcriptome characterization and detection of gene expression differences in aspen (*Populus tremuloides*). *Tree Genet Genomes* 2013, in press.
83. Nanjo T, Sakurai T, Totoki Y, Toyoda A, Nishiguchi M, Kado T, Igasaki T, Futamura N, Seki M, Sakaki Y, Shinozaki K, Shinohara K: Functional annotation of 19,841 *Populus nigra* full-length enriched cDNA clones. *BMC Genomics* 2007, **8**:448.
84. Paux E, Tamasloukht M, Ladouce N, Sivadon P, Grima-Pettenati J: Identification of genes preferentially expressed during wood formation in *Eucalyptus*. *Plant Mol Biol* 2004, **55**:263-280.
85. Ranik M, Creux NM, Myburg AA: Within-tree transcriptome profiling in wood-forming tissues of a fast-growing *Eucalyptus* tree. *Tree Physiol* 2006, **26**:365-375.
86. Foucart C, Paux E, Ladouce N, San-Clemente H, Grima-Pettenati J, Sivadon P: Transcript profiling of a xylem vs phloem cDNA subtractive library identifies new genes expressed during xylogenesis in *Eucalyptus*. *New Phytol* 2006, **170**:739-752.
87. Novaes E, Drost DR, Farmerie WG, Pappas GJ, Jr, Grattapaglia D, Sederoff RR, Kirst M: High throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics* 2008, **9**:312.
88. Salazar MM, Nascimento LC, Camargo EL, Goncalves DC, Neto JL, Marques WL, Teixeira PJ, Mieczkowski P, Mondego JM, Carazzolle MF, Deckmann AC, Pereira GA: Xylem transcription profiles indicate potential metabolic responses for economically relevant characteristics of *Eucalyptus* species. *BMC Genomics* 2013, **14**:201.

89. Thumma BR, Sharma N, Southerton SG: **Transcriptome sequencing of *Eucalyptus camaldulensis* seedlings subjected to water stress reveals functional single nucleotide polymorphisms and genes under selection.** *BMC Genomics* 2012, **13**:364.
90. Chancerel E, Lepoittevin C, Le Provost G, Lin YC, Jaramillo-Correa JP, Eckert AJ, Wegrzyn JL, Zelenika D, Boland A, Frigerio JM, Chaumeil P, Garnier-Gere P, Boury C, Grivet D, Gonzalez-Martinez SC, Rouze P, Van de Peer Y, Neale DB, Cervera MT, Kremer A, Plomion C: **Development and implementation of a highly-multiplexed SNP array for genetic mapping in maritime pine and comparative mapping with loblolly pine.** *BMC Genomics* 2011, **12**:368.
91. Ueno S, Le Provost G, Leger V, Klopp C, Noirot C, Frigerio JM, Salin F, Salse J, Abrouk M, Murat F, Brendel O, Derory J, Abadie P, Leger P, Cabane C, Barre A, de Daruvar A, Couloux A, Wincker P, Reviron MP, Kremer A, Plomion C: **Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak.** *BMC Genomics* 2010, **11**:650.
92. Barakat A, DiLoreto DS, Zhang Y, Smith C, Baier K, Powell WA, Wheeler N, Sederoff R, Carlson JE: **Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection.** *BMC Plant Biol* 2009, **9**:51.
93. Barakat A, Staton M, Cheng CH, Park J, Yassin NB, Ficklin S, Yeh CC, Hebard F, Baier K, Powell W, Schuster SC, Wheeler N, Abbott A, Carlson JE, Sederoff R: **Chestnut resistance to the blight disease: insights from transcriptome analysis.** *BMC Plant Biol* 2012, **12**:38.
94. Bai X, Rivera-Vega L, Mamidala P, Bonello P, Herms DA, Mittapalli O: **Transcriptomic signatures of ash (*Fraxinus* spp.) phloem.** *PLoS One* 2011, **6**:e16368.
95. Eckert AJ, Bower AD, Gonzalez-Martinez SC, Wegrzyn JL, Coop G, Neale DB: **Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae).** *Mol Ecol* 2010, **19**:3789-3805.
96. Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, Gonzalez-Martinez SC, Neale DB: **Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae).** *Genetics* 2010, **185**:969-982.
97. Pavy N, Gagnon F, Rigault P, Blais S, Deschenes A, Boyle B, Pelgas B, Deslauriers M, Clement S, Lavigne P, Lamothe M, Cooke JE, Jaramillo-Correa JP, Beaulieu J, Isabel N, Mackay J, Bousquet J: **Development of high-density SNP genotyping arrays for white spruce (*Picea glauca*) and transferability to subtropical and nordic congeners.** *Mol Ecol Resour* 2013, **13**:324-336.
98. Pavy N, Pelgas B, Beauseigle S, Blais S, Gagnon F, Gosselin I, Lamothe M, Isabel N, Bousquet J: **Enhancing genetic mapping of complex genomes through the design of highly multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce.** *BMC Genomics* 2008, **9**:21.
99. **Dryad** [<http://datadryad.org/>]
100. Geraldine A, Pang J, Thiessen N, Cezard T, Moore R, Zhao Y, Tam A, Wang S, Friedmann M, Birol I, Jones SJ, Cronk QC, Douglas CJ: **SNP discovery in black cottonwood (*Populus trichocarpa*) by population transcriptome resequencing.** *Mol Ecol Resour* 2011, **11** Suppl 1:81-92.
101. Wegrzyn JL, Eckert AJ, Choi M, Lee JM, Stanton BJ, Sykes R, Davis MF, Tsai CJ, Neale DB: **Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem.** *New Phytol* 2010, **188**:515-532.
102. Isabel N, Lamothe M, Thompson SL: **A second-generation diagnostic single nucleotide polymorphism (SNP)-based assay, optimized to distinguish among eight poplar (*Populus* L.) species and their early hybrids.** *Tree Genet Genomes* 2013, **9**:621-626.
103. Guerra FP, Wegrzyn JL, Sykes R, Davis MF, Stanton BJ, Neale DB: **Association genetics of chemical wood properties in black poplar (*Populus nigra*).** *New Phytol* 2013, **197**:162-176.
104. **Diversity Arrays Technology Pty Ltd (DArT P/L)** [<http://www.diversityarrays.com/>]
105. Sansaloni CP, Petroli CD, Carling J, Hudson CJ, Steane DA, Myburg AA, Grattapaglia D, Vaillancourt RE, Kilian A: **A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*.** *Plant Methods* 2010, **6**:16.
106. Grattapaglia D, Silva-Junior OB, Kirst M, de Lima BM, Faria DA, Pappas GJ Jr: **High throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species.** *BMC Plant Biol* 2011, **11**:65.
107. Hendre PS, Kamalakannan R, Varghese M: **High-throughput and parallel SNP discovery in selected candidate genes in *Eucalyptus camaldulensis* using Illumina NGS platform.** *Plant Biotechnol J* 2012, **10**:646-656.

doi:10.1186/gb-2013-14-6-120

Cite this article as: Neale DB, et al.: Open access to tree genomes: the path to a better forest. *Genome Biology* 2013, **14**:120.