Population, quantitative and comparative genomics of adaptation in forest trees
David B Neale¹ and Pär K Ingvarsson²

High-throughput DNA sequencing and genotyping technologies have enabled a new generation of research in plant genetics where combined quantitative and population genetic approaches can be used to better understand the relationship between naturally occurring genotypic and phenotypic diversity. Forest trees are highly amenable to such studies because of their combined undomesticated and partially domesticated state. Forest geneticists are using association genetics to dissect complex adaptive traits and discover the underlying genes. In parallel, they are using resequencing of candidate genes and modern population genetics methods to discover genes under natural selection. This combined approach is identifying the most important genes that determine patterns of complex trait adaptation observed in many tree populations.

Addresses
1 Department of Plant Sciences, University of California, Davis, Mail Stop 6 and Institute of Forest Genetics, Pacific Southwest Research Station, USDA Forest Service, Davis, CA 95616, USA
2 Department of Ecology and Environmental Science, Umeå Plant Sciences Centre, University of Umeå, Umeå SE-901 87, Sweden

Corresponding author: Neale, David B (dnoneale@ucdavis.edu) and Ingvarsson, Pär K (pelle@wallace.emg.umu.se)

Introduction
Forest trees are rarely viewed as model systems for most lines of investigation in plant biology, however, in this short review we will argue that they provide an excellent system for understanding the relationship between naturally occurring genotypic and phenotypic variation in plants. We must first distinguish between angiosperm forest trees such as poplars (Populus) and gymnosperm (conifer) species such as pines (Pinus) because there are large differences between these groups in their basic genetic and life history characteristics that make them more or less amenable to genetic studies (Table 1). The genus Populus contains approximately 29 species and is from the family Salicaceae (2 genera, ~450 species) [2]. Pines are monoecious with a mixed mating system that is primarily outcrossing. Poplars are dioecious and are thus obligate outcrossers. The time to natural reproductive maturity in pines can be quite long (more than 10 years) but can be shortened under artificial conditions to just a few years. The time to reproductive maturity in poplar is much less, only four or five years in some cases. Both pines (2N = 38) and poplars (2N = 38) are diploids, though poplar has undergone genome duplications during its evolutionary history [3⁷]. The poplar genome size is relatively small (450 Mb), whereas pine genomes are enormous (10–40 Gb) [4]. This explains why the genome of a poplar has been completely sequenced [3] and has not for any pine. Poplar can easily be transformed using Agrobacterium which facilitates functional studies, however the number of species and background genotypes within species that can be transformed is limited [5]. All forest tree species take many years to complete development and reveal their full phenotypic expression. In addition, trees are large and require extensive field plantations for phenotypic evaluations. The factors listed above make forest trees difficult organisms for most lines of genetic study, especially questions related to development, but for studies seeking to understand the relationship between naturally occurring genotypic and phenotypic variation forest trees are somewhat ideal, as we will describe in the next section.

Experimental advantages of combined quantitative and population genetic approaches in forest trees
Forest trees have the somewhat unique distinction in that many species can be found in both a domesticated and undomesticated state. This situation provides extensive experimental opportunity for the study of the relationship between naturally occurring genotypic and phenotypic variation. Having access to populations with little or no human disturbance means that extant populations are the result of natural evolutionary forces and questions pertaining to speciation, adaptation, and demography will not be anthropomorphically confounded. This contrasts with the large number of agricultural species that cannot be found in natural populations and have often been through large domestication bottlenecks. At the same time however, many forest trees are in the early stages of domestication which means large numbers of genotypes can be found in replicated genetic tests for trait evaluation and selection (Table 2).
Nevertheless, large natural populations remain for the few species that are in the early stages of domestication. These resources are extremely valuable for complex trait dissection using either quantitative trait locus (QTL) or association mapping. This contrasts with other noneconomic model systems such as Arabidopsis or Mimulus where such populations for phenotypic evaluation and trait dissection do not exist or must be assembled de novo by researchers. Furthermore, tree populations can be evaluated over their lifetimes, thus enabling ‘cradle to grave’ complex trait dissection.

The study of natural genetic variation in forest trees has traditionally been investigated using either of two approaches: first, quantitative genetics using common gardens or second, population genetics using markers. The common garden approach estimates genetic variation based on the measurement of phenotypic variation. Common gardens can be replicated over many different environments using either clonal or family-based testing. This allows estimation of genotype by environment interactions. Quantitative genetic parameters such as heritability and additive and dominance variance components are used to characterize genetic variation. This approach is extremely useful for characterizing broad patterns of adaptive genetic variation and has been used in a practical manner to define seed or breeding zones in reforestation programs. However, the individual genes underlying complex adaptive traits are not known, so single-locus population genetic theory cannot be applied. The alternative has been to use genetic markers for study of natural genetic variation. Markers such as isozymes, RFLPs, RAPDs, AFLPs, SSRs, and ESTPs have all been used but for the most part all reveal neutral genetic variation. Such markers are useful for characterizing demographic patterns of variation (migration and drift) but are not instructive of adaptive patterns of genetic variation. So before population genetic theory can be applied to genes controlling complex adaptive traits, they must first be discovered by complex trait dissection experiments. The QTL approach was first used, but because of low map resolution of QTLs, the underlying genes could not be determined [6]. The association genetics approach does provide much higher-level map resolution and it potentially can reveal individual genes underlying complex traits. Once the genes are identified that underly adaptive traits, then it is just one more step to discover the naturally occurring allelic variation in populations and test for the presence of selection using modern population genetic methods [7].

Neale and Savolainen [6] outlined the advantages of using the association genetic approach for dissecting complex traits in conifers and will be briefly reviewed again here for completeness. The primary advantages are first, large and random-mating populations with minimal population structure; second, adequate levels for nucleotide diversity for single nucleotide polymorphism (SNP) markers; third, rapid decay of linkage disequilibrium (LD); fourth, direct determination of haploids from DNA sequencing of haploid seed megagametophyte tissue; and fifth, access to large clonal or family-based genetic tests for precise evaluation of phenotypes. All but item four, pertain to poplars and many other angiosperm forest tree species. Two association genetic studies in trees have been published [8,9] and several more have been completed (see Table 1 in [10]). Some of the different types of genetic resources available from four conifer tree improvement programs in the US are shown in Table 1.

Neale and Savolainen [6] also outlined the advantages of using the association genetic approach for dissecting complex traits in conifers and will be briefly reviewed again here for completeness. The primary advantages are first, large and random-mating populations with minimal population structure; second, adequate levels for nucleotide diversity for single nucleotide polymorphism (SNP) markers; third, rapid decay of linkage disequilibrium (LD); fourth, direct determination of haploids from DNA sequencing of haploid seed megagametophyte tissue; and fifth, access to large clonal or family-based genetic tests for precise evaluation of phenotypes. All but item four, pertain to poplars and many other angiosperm forest tree species. Two association genetic studies in trees have been published [8,9] and several more have been completed (see Table 1 in [10]). Some of the different types of genetic resources available from four conifer tree improvement programs in the US are shown in Table 1.

The most outstanding difference is in the size of genomes. Lack of a reference genome sequence has hindered progress in Pinus genetics, but once obtained, Pinus and Populus will be comparable and excellent plant systems for studying the relationship between natural genetic and phenotypic diversity.
reported associations are between candidate gene SNPs and biotic (disease resistance) and abiotic (water use efficiency, cold tolerance) stresses. These quantitative genetic studies are among the first to identify individual genes controlling complex adaptive traits. We will review the progress using this approach in the next section.

**Population genetics of genes controlling adaptive traits in trees**

DNA sequencing of candidate genes in small panels of individuals to discover SNPs in trees began just a few years ago [11,12]. A number of studies in conifers and poplars have been published and have been recently reviewed [10,13]. Two general types of inference can be made from resequencing data. First, are the estimates of the neutral mutation rate based on the amount of synonymous nucleotide diversity. The second type of inference made from resequencing data is departures from neutrality. A suite of neutrality tests have been derived in recent years and these have been applied to tree resequencing data to identify genes potentially under natural selection [11,12,14–16,17,18–21]. Depending on the test statistic used, evidence for either balancing or directional selection can be inferred. Of the approximately 290 genes in trees that have been subjected to such tests, 55 (20%) showed departures from neutrality. These are very preliminary data but in these early stages we might conclude that one in five genes in tree genomes may be under some form of natural selection. Given the highly heterogeneous environments that many species are adapted to and the wealth of data demonstrating adaptive patterns of variation using quantitative genetic approaches, 20% of the genes being under selection would seem to be a lower limit. One caveat to bear in mind is that many demographic processes, such as population expansion, bottlenecks or population subdivision, can result in patterns of polymorphism that mimic signatures of natural selection. Demographic processes are expected to affect the entire genome of a species whereas selection is likely to affect only one or a few genes. If past demographic changes have led to systematic departures from the neutral model, this increases the probability that a gene shows departures from neutrality and will therefore inflate the number of false positives. As multilocus sequence data sets are becoming available for more tree species, it should be possible to move away from using a strictly neutral model as a reference point and instead use empirically derived distributions of genome-wide polymorphism as a baseline against which unusual patterns of polymorphism in genes of interest can be tested, as have recently been advocated for *Arabidopsis thaliana* [22].

The most interesting genes from the perspective of understanding the relationship between natural genotypic and phenotypic variation are those that are implicated using both quantitative and population genetic approaches. Till date there are only two such examples, one in poplar and one in Douglas-fir. The example in poplar is a gene coding for phytochrome and the timing of bud set (Figure 1). Several QTL mapping studies in *Populus* have found that QTLs for both bud flush and bud set co-locate with a phytochrome gene, *phytochromeB2 (phyB2)* [23,24]. A resequencing study of 48 *phyB2* haplotypes in European aspen (*Populus tremula*) found several SNPs, both non-synonymous and non-coding, that showed significant clinal variation with latitude [25]. These SNPs were tested for associations with both bud set and bud flush and two non-synonymous SNPs were found to be independently associated with bud set. These SNPs explain between 1% and 5% of the phenotypic variation in bud set in *P. tremula*. This example shows how association mapping can be used to rapidly identify causal mutations in candidate genes identified in traditional QTL studies.

The example from Douglas-fir is a gene coding for a flavonone 3-hydroxylase (F3H) and stem cold hardiness (Figure 2). The estimated Tajima’s D = −1.194 in F3H [26], suggestive of directional selection and a SNP in this same gene associated with stem cold hardiness. That SNP accounted for only a small portion of the total phenotypic variance, nevertheless it is an early example of one gene controlling an important adaptive trait that is under natural selection.

**Comparative genomic analyses in trees**

Comparative genomic analyses are used in many plant and animal systems to fully understand the evolutionary basis of gene function and developmental processes [27]. Comparative genomic approaches can also be used to understand the evolution of adaptation. Forest trees, notably the pines and poplars, are well suited to comparative genomics. In contrast to some of the important agricultural crops where researchers around the world might study the same species (e.g. corn, wheat, rice, and many others), there is no one species of forest tree that is the object of study throughout the world. Forest geneticists have generally worked on the species of regional importance. This situation has been viewed as a limitation because of dispersed and lack of focused investigation. However, as the power of comparative genomics is realized, this situation may soon be viewed as an asset, rather than a liability. To that end, forest genomics researchers have taken early steps toward developing collaborative networks and a comparative genomic framework.

In conifers, the Conifer Comparative Genomics Project (http://dendrome.ucdavis.edu/ccgp) was started several years ago and led to a series of low-density comparative maps [28–33]. These maps suggest that macrosynteny is conserved in conifers, making it possible to easily navigate across genomes. More recently, a comparative resequencing project (http://dendrome.ucdavis.edu/crsp) was...
begun that will provide novel insight into within and between species genetic diversity and will lead to a deeper understanding of speciation and the genetic basis of phenotypic differences among species.

Tools for comparative genomics are also increasingly becoming available in *Populus*. Comparative linkage maps have been created for several species of *Populus* and they show a large degree of synteny across species. Recently, the first comparative map between *Populus* and its sister genus *Salix* was constructed by anchoring a *Salix* map to the *Populus* genome sequence. The alignment of the *Salix* and *Populus* linkage maps suggests that large-scale synteny is conserved between the two genera [34]. In addition, a physical map with approximately $9 \times$ coverage of the genome is also available for *P. trichocarpa* [35].

(a) Length of the growing season across Sweden, scored as the number of days with temperatures $>\, 5\, ^\circ\text{C}$. Gray circles show the geographic origin of the trees. (b) The distribution of genotypes for two nonsynonymous SNPs in *phyB2* and number of days until budset for trees collected across a latitudinal gradient in Sweden. (c) Genotypic effects of the two nonsynonymous *phyB2* SNPs. The map is reprinted with permission from the Swedish Meteorological and Hydrological Institute.
There are also abundant EST resources available for *Populus* [36], with over 360,000 ESTs from 10 different species available in GenBank as of November 2007. There are currently no comparative resequencing studies available for *Populus*, where the same homologous regions have been sequenced from multiple species. Such studies are much needed, however, as they will open up possibilities for a truly comparative genomics approach to adaptation. This will provide important insights into the relative importance of convergence, parallelism or completely independent evolution of the genetic architecture of adaptive phenotypic responses in different species.

The genus *Populus* is also, from an evolutionary perspective, relatively closely related to the model plant species *A. thaliana* and hopes have been raised that research in *Populus* will provide insights into traits that are poorly developed in *Arabidopsis*, such as wood formation, the transition from juvenility to maturity and adaptation to seasonal and other environmental changes [37]. Given the extreme differences in life history between *Populus* and *Arabidopsis*, the former being a long-lived perennial and the latter an extremely short-lived annual, comparative genomic studies should provide important insights both into mechanisms that are conserved across dicots but also how different evolutionary lineages have used common developmental pathways to solve different evolutionary problems. The latter is nicely illustrated by a recent comparative study of the photoperiodic pathway in *Populus* and *Arabidopsis*. This pathway regulates photoperiodic control of flowering in *Arabidopsis*, but in...
Populus the same pathway has also been co-opted to control seasonal growth cessation, a process that is absent in Arabidopsis [38]. Interestingly, the photoperiodic pathway has recently been implicated in control of growth cessation also in Norway spruce (Picea abies) [39], suggesting that some aspects of the photoperiodic pathway are conserved also between angiosperms and gymnosperms.

Conclusion

We have argued that forest trees make excellent experimental systems for the study of the relationship between naturally occurring genotypic and phenotypic diversity in plant populations. Some of the advantages of model systems such as A. thaliana and rice (small genomes, reference sequences, mutants, and transformation) might be superseded by the life history characteristics of forest trees (large natural populations, outbreeding, large infrastructure for phenotyping). However, the genomics resources for forest trees are still badly underdeveloped and will need considerable investment. In the short term, much can be accomplished by focusing on the gene space of tree genomes and leaving the large, noncoding regions for later study as technology develops and costs decrease. To that end, immediate priorities should be: first, the development of deep EST databases in multiple species; second, comparative resequencing in multiple species; and third, a reference genome sequence from a member of Pinus. Such resources will better enable genomic research in forest trees and better exploit the wealth of basic biological knowledge that can be learned from trees.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


This paper was the first to suggest that the association genetics approach be used for dissecting complex traits in trees and outlined some of the advantages of this approach over standard QTL mapping and the power of the approach in trees versus many of the agricultural crops.

19. The first study on nucleotide diversity, linkage disequilibrium and tests of neutrality in spruce (Picea), a genus of conifer that shows complex patterns of adaptation.


