Research review

Forest-tree population genomics and adaptive evolution

Santiago C. González-Martínez1, Konstantin V. Krutovsky2 and David B. Neale3,4
1Department of Forest Systems and Resources, Center of Forest Research (CIFOR-INIA), 28040 Madrid, Spain; 2Department of Forest Science, Texas A&M University, College Station, TX 77843-2135, USA; 3Institute of Forest Genetics, Pacific South-west Research Station, US Department of Agriculture Forest Service, Davis, CA 95616, USA; 4Department of Plant Sciences, University of California, Davis, CA 95616, USA

Summary

Forest trees have gained much attention in recent years as nonclassical model eukaryotes for population, evolutionary and ecological genomic studies. Because of low domestication, large open-pollinated native populations, and high levels of both genetic and phenotypic variation, they are ideal organisms to unveil the molecular basis of population adaptive divergence in nature. Population genomics, in its broad-sense definition, is an emerging discipline that combines genome-wide sampling with traditional population genetic approaches to understanding evolution. Here we briefly review traditional methods of studying adaptive genetic variation in forest trees, and describe a new, integrated population genomics approach. First, alleles (haplotypes) at candidate genes for adaptive traits and their effects on phenotypes need to be characterized via sequencing and association mapping. At this stage, functional genomics can assist in understanding gene action and regulation by providing detailed transcriptional profiles. Second, frequencies of alleles in native populations for causative single-nucleotide polymorphisms are estimated to identify patterns of adaptive variation across heterogeneous environments. Population genomics, through deciphering allelic effects on phenotypes and identifying patterns of adaptive variation at the landscape level, will in the future constitute a useful tool, if cost-effective, to design conservation strategies for forest trees.

doi: 10.1111/j.1469-8137.2006.01686.x

Introduction

Understanding the genetic basis of population divergence and adaptation is an important goal in population genetics and evolutionary biology. Forest geneticists have long been concerned with understanding the interplay of evolutionary factors, demography and population structure that, together, shape genetic variation and adaptation in tree species (Eriksson, 1998; Namkoong, 2001). Traditional methods such as provenance tests and screening of molecular genetic markers have been used to study and measure adaptive genetic diversity in forest-tree populations. However, adaptive traits...
are usually under multigenic control. Here we review how developments in forest genomics now provide us with tools to identify the genes controlling adaptive traits and methods to carry out new-generation population genetic studies. Population genomics combines genome-wide sampling with the population-genetics objective of understanding evolution (Luikart et al., 2003). This emerging discipline takes advantage of the availability of functional genetic markers and new tools of population analysis, such as association mapping and genome scans, to reveal adaptive patterns in nature. Population genomics is complementary to other new approaches, such as community and ecosystem genetics (Whitham et al., 2003); evolutionary and ecological functional genomics (Feder & Mitchell-Olfs, 2003); and landscape genetics (Manel et al., 2003). We briefly describe population genomics approaches applied to forest trees, and how they might be useful for understanding patterns of adaptive variation in forest-tree populations.

**Traditional methods to study adaptive genetic variation in forest trees**

**Field experiments**

Common-garden experiments (provenance, progeny and clonal tests) are commonly used to study adaptive evolution of quantitative traits in forest-tree populations. Such studies have focused on traits of economic interest such as survival, growth, wood properties, cold-hardiness, drought tolerance, and pest or disease resistance. They have often shown geographical patterns of adaptive genetic variation, such as steep latitudinal or altitudinal clines, resulting from natural selection and local adaptation (Campbell, 1979; Rehfeldt et al., 1999; García-Gil et al., 2003). Tree populations are usually well adapted to local environments, although it is not uncommon to find populations living in suboptimal conditions. This can occur in zones with temporal fluctuation of local conditions, such as climate changes, or in marginal populations that receive recurrent maladapted immigrants from neighbouring populations (Rehfeldt et al., 2001). Common-garden studies of various kinds are used in tree breeding, and can identify families and clones that are specifically adapted to particular environments or to a broad variety of environments. Much has been learned about patterns of adaptive variation in complex traits at both macro- and micro-environmental levels. However, field experiments are very time consuming and relatively expensive and, more importantly, are based solely on phenotypes. They can estimate genetic parameters, but only on measurable traits, not on individual genes. The common-garden approach can provide information on neither what particular genes are involved in adaptation, nor how much phenotypic variation can be explained by genetic variation in these genes.

**Molecular genetic markers**

Genetic marker studies have contributed greatly to the understanding of gene flow, hybridization, population structure, genetic drift and mating systems (Newton et al., 1999; Ouborg et al., 1999; Hamrick & Nason, 2000; Linhart, 2000). In forest trees specifically, common applications of molecular markers have been to measure genetic diversity (Petit et al., 2005); to test glacialization hypotheses related to patterns of migration (Petit et al., 2003); to characterize human-mediated spread of particular genotypes (Gil et al., 2004); and to describe the breeding structure and gene flow in plants with keystone ecological roles (Nason et al., 1998; Adams & Burczyk, 2000; Smouse & Sork, 2004).

However, molecular-marker studies have contributed little to our understanding of natural selection and adaptation in forest-tree populations. A classification of genetic markers that takes into account their most important features can be found in Table 1 of Krutovsky & Neale (2005a). Biochemical markers, such as allozymes, are a class of genetic marker widely used in the past, and although variation revealed by these markers is caused by amino acid variation, it is often unclear whether this variation is selectively neutral or has any adaptive significance. DNA variation that resides in noncoding genomic regions (although a fraction of it might have vital regulatory functions; Sandelin et al., 2004), or does not lead to a change in the amino acid sequence, is likely to be selectively neutral. Many modern genetic markers, such as microsatellites or simple sequence repeats (SSRs), random amplified polymorphic DNA (RAPDs) and amplified fragment-length polymorphisms (AFLPs) generally reveal noncoding DNA sequences and should be assumed then to be selectively neutral. Restriction fragment-length polymorphisms (RFLPs) are of two general types, based on (1) complementary DNA (cDNA); or (2) genomic DNA. Both types have been used in forest trees, although only cDNA-based markers might potentially reveal adaptive variation.

There are many studies showing adaptive differences in morphological, phenological or growth characteristics among populations of forest-tree species, but only rarely have accompanying differences for molecular markers been found (see references in Boshier & Young, 2000). Despite a few studies showing concordance of morphological and allozymic variation (Lagerkrantz & Ryman, 1990; Mitton et al., 1998; Mitton & Duran, 2004), most studies showed different patterns of molecular marker and quantitative variation (reviewed by Karhu et al., 1996; McKay & Latta, 2002). In conifers, molecular markers typically show far less variation than adaptive traits when sampled in the same populations or across the same range (Adams & Campbell, 1981; Merkle et al., 1988; Karhu et al., 1996; González-Martínez et al., 2004). Furthermore, it might have been assumed in the past that some (if not many) of these markers could be genetically linked to genes under natural selection and would thus reveal adaptive
patterns, but recent studies showing relatively weak linkage disequilibrium (LD) in tree populations indicate that the assumption was unrealistic.

Quantitative trait locus mapping

Quantitative trait locus (QTL) mapping is primarily a method of finding genetic regions that are responsible for variation in complex traits, although it can also be used to study adaptive traits in forest trees. Quantitative trait locus mapping is relatively straightforward, but requires (1) dense genetic maps with evenly distributed markers covering the entire genome; (2) appropriate statistical tools; and (3) sufficient progeny size segregating for both genetic markers and phenotypic traits (Paterson, 1998). First, genetic markers are genotyped and quantitative traits are phenotyped in all individuals of a segregating population. Phenotypic values are then statistically associated with genotypes, usually using multiple-regression or maximum-likelihood methods to identify markers that cosegregate with the quantitative trait. An association between a genetic marker and a phenotypic trait is usually the result of tight linkage between a marker and a gene or genes that control the phenotypic trait. Quantitative trait locus mapping depends heavily on dense genetic maps that are usually time-consuming and expensive to construct, and requires large sample sizes (over 500 individuals). Quantitative trait locus detection is often problematic, and has limited application because of: (1) instability of QTL associations across different environments and genetic backgrounds; (2) preferential detection of QTL with large phenotypic effect, and therefore underestimation of the number of genes with minor effects that also control a trait; (3) the multiplicity of epistatic QTL effects; and (4) caveats associated with statistical methods, such as assumption of normal distribution of phenotypic traits and multiple testing that can lead to detection of false-positive QTL (Doerge, 2002; Mauricio, 2001). For example, some QTL for spring cold-hardiness and other traits in Douglas fir were detected only in one environment, not in another (Jermstad et al., 2001b; Wheeler et al., 2005). This makes verification of QTL a very important requirement. Furthermore, QTL very rarely explain a significant part of the total phenotypic variation associated with a trait (Doerge, 2002). For instance, in conifers they usually explain only about 5–15% of phenotypic variation (Table 1 at http://www.pinegenome.org/pdf/workshop_summary.pdf). Nevertheless, QTL for several adaptive traits, such as growth rhythm, phenology, stem form, wood quality, disease resistance, cold hardiness, drought tolerance and others, have been detected and mapped in forest trees (for reviews see Sewell & Neale, 2000; Guevara et al., 2005). A high level of heterozygosity in forest trees, caused by large population sizes and the outcrossing mating system, allows the use of progeny from F1 crosses for QTL mapping, unlike many crop species that typically require F2 crosses. For example, Lerceteau et al. (2000) identified three QTL that explained 25.8% of the total phenotypic variance for the tree-height trait using an F1 full-sib progeny from two plus-trees originating in northern Sweden. Genetic regions harbouring QTL for adaptive traits have been identified in Douglas fir (Jermstad et al., 2001a, 2001b, 2003; Wheeler et al., 2005); pine hybrids (Weng et al., 2002); poplar (Freven et al., 2000; Ferris et al., 2002; Wu et al., 2003); willows (Tsarouhas et al., 2002, 2003, 2004; Ronnberg-Wastljung et al., 2005); loblolly pine (Neale et al., 2002; Sewell et al., 2002; Brown et al., 2003); maritime pine (Brendel et al., 2002; Markussen et al., 2003; Pot, 2004); Scots pine (Lerceteau et al., 2001; Yazdani et al., 2003); radiata pine (Devey et al., 2004a, 2004b); European beech (Scalfi et al., 2004); chestnut (Casasoli et al., 2004); oak (Scotti-Saintagne et al., 2004a); eucalyptus (Kirst et al., 2004; Thamarus et al., 2004).

However, QTL studies cannot reveal the specific genes underlying the adaptive traits. The QTL mapping data can be used to identify individual genes via positional cloning, but it is very challenging, if not impossible, in forest trees. Positional cloning requires a well defined, narrow QTL interval that can be achieved only by means of a large segregating population (over 1000 individuals) and a marker-saturated fine-linkage map. Then a large-insert genomic library (bacterial artificial chromosome, BAC or yeast artificial chromosome, YAC) should be screened to find a genomic fragment that corresponds to this QTL interval. The fragment can be progressively sequenced, but it may potentially contain many different genes. In forest trees, precision of QTL mapping is usually low (intervals under which QTL are mapped can include several hundred genes), and comprehensive BAC libraries have been developed only in a few tree species (e.g. Eucalyptus, Grattapaglia, 2004) with genomes much smaller than conifers. Nevertheless, QTL studies have shown the existence of loci with major effects on phenotypes, typically explaining 5–15% of the phenotypic variance (see review by Guevara et al., 2005). Furthermore, use of candidate genes for QTL mapping of adaptive traits can increase the chances of finding the target genes underlying these traits, because collocation between candidate gene and QTL might suggest that the candidate gene is directly involved in the control of the adaptive trait (Freven et al., 2000; Brown et al., 2003; Chagné et al., 2003; Wheeler et al., 2005).

Recent population and functional genomics approaches

New types of functional genomic markers

The ideal molecular marker for the study of adaptive variation should meet the following criteria: (1) be directly involved in the genetic control of adaptive traits; (2) have an identified DNA sequence and known function; and (3) have easily identifiable allelic variation. These criteria are not fully satisfied by any traditional marker, but new sequence-based...
markers that do so are rapidly being developed in several forest-tree species.

The past decade has seen an enormous increase in genomic resources publicly available for forest trees. Expressed sequence tag (EST) sequencing projects have provided numerous nucleotide sequences for pine, poplar, spruce (327 484; 260 997; 79 003, respectively, available at the July 2005 release of The Institute for Genomic Research, TIGR), and other tree species (see listing of EST projects and databases in forest trees at http://dendrome.ucdavis.edu/Gen_Page_body.htm and http://www.tigr.org/tdb/tgi/plant.shtml). Expressed sequence tags represent expressed genes with known or predicted function, and therefore can be considered as a new type of functional genomic marker (Andersen & Lubberstedt, 2003). Direct analysis of EST sequences has shown that approx. 2–8% of them contain SSR regions that can easily be used for developing hundreds, if not thousands, of SSR markers (Scotti et al., 2000; Moriguchi et al., 2003; see references in Gupta & Rustgi, 2004; Li et al., 2004; La Rota et al., 2005; Vasemägi et al., 2005), which are also readily transferable among related species (Chagné et al., 2004). For example, the poplar (*Populus trichocarpa*) complete genome contains more than 300 000 perfect-repeat SSRs. Tuskan et al. (2004) estimated that approx. 70–99% of these would be transferable within and across sections at the subgenus level within *Populus*. In outbreeding species with large population sizes and high recombination, such as several forest trees, SSRs found within noncoding (usually 3′) untranslated regions in ESTs represent the same genomic region, but might not necessarily be in LD with coding regions because of the short extent of LD in these species. Therefore the identification of selection signatures on EST-based SSRs provides a means to study whether nucleotide polymorphism patterns in functional regions are the result of selection or other factors, such as demographic processes.

Single-nucleotide polymorphisms (SNPs) are potentially the best type of genetic marker because of their abundance in the genome and their potential association with disease and adaptive traits. Typical SNP discovery projects are based on direct sequencing of amplicons from a set of individuals (the discovery panel) covering the range of variation of a given species (Kado et al., 2003; Brown et al., 2004; González-Martínez et al., 2006; Krutovsky & Neale, 2005b; Pot et al., 2005). The dinucleotide nature of most SNPs facilitates the development of automated high-throughput SNP-genotyping methods (see reviews in Kwok, 2001; Hirschhorn & Daly, 2005). In *silico* SNP discovery has also been implemented in several forest-tree EST databases to discover SNP variation in ESTs (loblolly pine, http://fungen.org/Projects/Pine/Pine.htm; maritime pine, http://www.pierroton.inra.fr/genetics/Pinesnps). Although highly efficient (e.g. for maritime pine, Le-Dantec et al., 2004), in *silico* SNP discovery can be biased because of the typically small number of individuals from a limited number of populations used to generate EST libraries (Gupta et al., 2005 and references therein).

The best candidates for population genomics and related approaches are SNPs that cause nonsynonymous substitutions, mark haploblocks, and are under positive selection (as shown by neutrality tests). Given the level of nucleotide diversity and within-gene LD found in trees (Table 1), genotyping

---

**Table 1** Nucleotide diversity, recombination and putative candidate gene loci under selection identified by analyses of DNA sequence variation patterns in forest trees

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of loci</th>
<th>Sample size</th>
<th>Nucleotide diversity (π)</th>
<th>Recombination rate (ρ)</th>
<th>Putative candidate genes under selection</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pinus pinaster</td>
<td>8</td>
<td>22–91</td>
<td>0.0024</td>
<td>NA</td>
<td>pp1 (glycine-rich protein); cesA3; korrigan</td>
<td>Pot et al. (2005); D. Pot (personal communication)</td>
</tr>
<tr>
<td>Pinus radiata</td>
<td>8</td>
<td>12–23</td>
<td>0.0019</td>
<td>NA</td>
<td>pp1 (glycine-rich protein); cesA3; korrigan</td>
<td>Pot et al. (2005); D. Pot (personal communication)</td>
</tr>
<tr>
<td>Pinus sylvestris</td>
<td>1–2</td>
<td>12–20</td>
<td>0.0007; 0.0014</td>
<td>NA</td>
<td>–</td>
<td>Dvornýk et al. (2002); García-Gil et al. (2003)</td>
</tr>
<tr>
<td>Pinus taeda</td>
<td>18–19</td>
<td>32</td>
<td>0.0040; 0.0051; 0.00175; 0.00326</td>
<td>ccoaomt-1; erd3</td>
<td>–</td>
<td>Brown et al. (2004); González-Martínez et al. (2006); our unpublished data</td>
</tr>
<tr>
<td>Pseudotsuga menziesii</td>
<td>18</td>
<td>27–39</td>
<td>0.0066</td>
<td>NA</td>
<td>f3h1; 4c1; mt-like</td>
<td>Krutovsky &amp; Neale (2005b)</td>
</tr>
<tr>
<td>Cryptomeria japonica</td>
<td>7</td>
<td>48</td>
<td>0.0025</td>
<td>NA</td>
<td>acl5</td>
<td>Kado et al. (2003)</td>
</tr>
<tr>
<td>Betula pendula</td>
<td>1</td>
<td>40</td>
<td>0.0023; 0.0054</td>
<td>NA</td>
<td>–</td>
<td>Järvinen et al. (2003)</td>
</tr>
<tr>
<td>Populus tremula</td>
<td>5</td>
<td>34–48</td>
<td>0.0111</td>
<td>NA</td>
<td>–</td>
<td>Ingvarsson (2005)</td>
</tr>
</tbody>
</table>

NA, not available.
of a few haplotype-tagging SNPs (htSNPs) might be sufficient to genotype all or most common alleles. Indeed, genotyping a subset (30–60%) of all SNP markers discovered in 18 abiotic-stress candidate genes would be sufficient to represent most common allelic variation within these genes in different conifer species (González-Martínez et al., 2006; Krutovsky & Neale, 2005b). New and highly efficient SNP-discovery and SNP-genotyping techniques (Table 1 of Pask et al., 2004; Prokunina & Alarcón-Riquelme, 2004; Table 2 of Hirschhorn & Daly, 2005) have provided an almost unlimited source of markers and genotyping capacity.

Selection of adaptive trait-related candidate genes in forest trees

Ideally, in a true population-genomics approach as many genes and traits as available should be studied, because all expressed genes are candidates for one or several quantitative traits. However, time and budget restrictions make it necessary to preselect putative candidate gene loci for the particular adaptive trait(s) under study. For a few tree species, where fine QTL mapping studies exist, collocation of candidate genes might be used. For instance, collocation of cold-tolerance candidate genes and QTL for cold hardiness were used in candidate-gene selection for association studies in Douglas fir (Krutovsky & Neale, 2005b; Wheeler et al., 2005). For most trees, however, selection of candidate genes will rely on transference of information from model species (functional candidates: genes of known function in model systems) or in gene-expression studies for forest trees (expression candidates: Watsonson et al., 2003 for loblolly pine; Dubos & Plomion, 2003 and Dubos et al., 2003 for maritime pine).

Standard neutrality tests applied to population nucleotide sequence data of a single or a few gene(s) can also be used in selecting candidate genes or SNPs that are potentially under selection for association-mapping or population genomics studies. Deviations of allele (haplotype) distributions from standard neutral expectations can be associated with balancing selection, purifying selection or selective sweeps caused by positive selection (reviewed by Kreitman, 2000; Ford, 2002; Rosenberg & Nordborg, 2002), as long as deviations are not caused by demographic changes or population structure. Several genes that have been identified following this approach were related to environmental-stress tolerance, disease resistance or general metabolism (Table 1 of Ford, 2002). In pines, the majority of genes that showed a departure from neutrality in DNA-sequence studies were related to biotic- and abiotic-stress tolerance or key metabolic pathways such as those responsible for the formation of lignin in plants (González-Martínez et al., 2006; Pot et al., 2005). The lignification pathway is associated with physical and chemical properties of wood, tree growth and tolerance to biotic and abiotic stresses (Pot et al., 2002; Peter & Neale, 2004), and thus might have adaptive importance.

Association mapping

Association mapping uses LD in populations to find statistical associations between molecular markers and phenotype. After many generations of recombination and random mating, only tightly linked loci will show statistical association, allowing a finer mapping than standard QTL approaches. If candidate genes are used as markers, then this approach can find individual alleles that are directly involved in the genetic control of phenotypes shaped by several generations of natural selection. Association mapping in natural populations has been proposed as a powerful method for the identification of genes that underlie complex traits and for characterizing their effect on complex phenotypes (Cardon & Bell, 2001; Jannink & Walsh, 2002; Neale & Savolainen, 2004, for conifers; Gupta et al., 2005; Hirschhorn & Daly, 2005).

Because statistical power in association studies increases with allele frequency, common variants are usually preferred (Wang et al., 2005), although common alleles might have lower phenotypic effects (Frank, 2004). Population stratification (for instance, resulting from historical migration patterns) is the most common systematic bias producing false-positives in association studies (Marchini et al., 2004; Hirschhorn & Daly, 2005). Nevertheless, methods have been developed that correct for population structure or take advantage of family structure in populations with known pedigrees, such as the transmission/disequilibrium test of Spielman et al. (1993). Therefore it is very important to test for population structure in association-mapping populations. For this purpose, neutral markers are readily available in forest trees. For instance, highly polymorphic nuclear microsatellite markers for more than 50 forest-tree species are available at the Molecular Ecology Notes Primer Database (July 2005. http://tomato.bio.trinity.edu/MENotes/home.html). Apart from lacking population substructure, the sample size and origin of the trees are important in an association population. Sample sizes of about 500 individuals are required, in most cases, to have sufficient power to detect causative polymorphisms (Long & Langley, 1999). Measurement of phenotypes with enough precision in such large populations is challenging, as has been noted in *Eucalyptus* genomic programmes (Grattapaglia, 2004). Adaptive variation in forest trees is often arrayed clinally, in response to latitudinal or altitudinal climate or soil gradients (García-Gil et al., 2003), and therefore sampling the edges of a steep cline can increase the chance of elucidating the molecular basis of the divergent adaptive trait, as extreme genotypes might be sampled.

Recent studies in aspen and different conifers (pines, Norway spruce and Douglas fir) have shown a rapid decay of LD within candidate genes for different adaptive traits (Brown et al., 2004; Rafalski & Morgante, 2004; González-Martínez et al., 2006; Ingvarsson, 2005; Krutovsky & Neale, 2005b). The short extent of LD within genes (from approx. 200–400 bp in Norway spruce and Douglas fir to approx. 800–1500 bp in loblolly pine and < 500 bp in aspen), along with
large genome sizes in tree species (poplar and pine genomes are approximately fourfold and 160-fold larger, respectively, than the Arabidopsis genome), prevents genome-wide association studies because of the large number of SNPs (approx. 20 million in pine) that would be needed to cover the full genome evenly with spacing short enough to effectively identify adaptive mutations through LD. Instead, a more feasible association-mapping strategy based on candidate genes and flanking promoter regions is suggested for forest trees (Neale & Savolainen, 2004).

Functional genomics in adaptation research

A great deal of progress has been made in recent years in functional genomics and technology for studying gene expression and function. Functional genomics helps us understand how genotypes influence complex phenotypes by providing detailed transcriptional profiles and insights on gene expression and regulatory control. Microarray technology is rapidly becoming available for studying gene expression in organisms other than model species. Several other techniques can also be used for transcriptome analysis, such as cDNA-AFLP; reverse transcription–polymerase chain reaction (RT–PCR); and differential-display RT–PCR (DDRT–PCR) (Kuhn, 2001; Dubos & Plomion, 2003 and references therein). Once significant association is found between a phenotype and a particular allele or haplotype, functional genomics approaches can be used to study the effects of SNP, allele or haplotype variation on expression. Microarray-based gene-expression studies can provide relevant information about genetic interactions among gene complexes in response to different environmental stresses (Seki et al., 2001; Watkinson et al., 2003, for Pinus taeda). In addition, expression data for a gene can be measured in individual trees as a quantitative trait (an expression level polymorphism, ELP) and thus can be used in an association or QTL study as any other phenotype (see insights for QTL mapping of ELPs in Doerge, 2002).

In forest trees, large-scale changes in transcript profiles have been studied in pines, aspen and other species. Changes in transcript profiles of P. taeda that reflected photosynthetic acclimation depending on drought intensity have been revealed using a microarray based on 2173 cDNA clones (Watkinson et al., 2003). In their study, cDNAs were classified in functional categories to analyse the co-response of different groups of genes to mild and severe drought stress. Several genes responded differently to the two levels of drought stress, including some that belonged to the same individual gene family. For example, late embryogenesis-abundant (LEA) group 2 genes (dehydrins) were specifically upregulated during mild drought stress, thus being associated to photosynthetic acclimation, whereas expression of LEA group 3 genes were more associated with severe stress conditions. In aspen, a major shift in gene expression, similar to the effects of senescence in annual plants, has been observed for autumn leaf senescence, coinciding with massive chlorophyll degradation, using a 13 490 clone cDNA microarray (Andersson et al., 2004). Transcriptional profiles can also vary among year seasons and seed sources. Yang et al. (2004) found different expression patterns in 569 (out of 1873) cDNAs in Robinia pseudoacacia when transcriptional profiles in autumn and summer were compared. Yang & Loopstra (2005) showed variation in gene expression between Arkansas and Louisiana loblolly pine origins that might be related to adaptation to different environments.

Unveiling adaptive genetic divergence in natural populations using outlier-detection approaches

The detection of loci with unusually high or low levels of variation and differentiation (outlier loci) is a powerful method to find loci under selection and to separate genome-wide effects that are caused by demographic processes from adaptive locus-specific effects (Luikart et al., 2003). For instance, lower than expected (from the neutral model) observed heterozygosity is a typical genome-wide signature of population expansion, but also a locus-specific signature of selective sweeps and directional selection (Payseur et al., 2002). On the other hand, certain cases of balancing selection, such as those caused by overdominance (heterozygous individuals are favoured) or frequency-dependent selection (in which single alleles confer higher fitness when rare and become less favoured at higher frequency) can result in a locus-specific excess of heterozygosity for the selected gene (Black et al., 2001).

The most widely used tests are based on the detection of outlier loci for multiple-population genetic differentiation estimates. One simple method is based on the comparison of differentiation estimates, such as $F_{st}$, for putatively neutral molecular markers (usually nuclear SSRs) and candidate gene markers (e.g., SNPs or EST-based markers). Markers that show higher (or lower) differentiation than putatively neutral ones can be considered as being under diversifying (or stabilizing) selection. A more sophisticated approach, which does not require screening of any neutral molecular marker, consists of the use of the coalescent theory to build, by means of simulation, a neutral expectation of genetic divergence among populations. Two competing methods for this approach are rapidly becoming widespread. First, Beaumont & Nichols (1996) developed a method based on the analytical framework of Lewontin & Krakauer (1973) that was further extended using Bayesian theory (Beaumont & Balding, 2004). This method constructs a theoretical neutral expectation of $F_{st}$ for each value of expected heterozygosity ($H_e$) based on the global genetic differentiation found in a sample. Simulation studies have shown an acceptable rate of identification of loci under positive selection, but also showed that this method can fail to detect loci under balancing selection (Beaumont & Balding, 2004). The second method, but less used to date, was developed by Vitalis et al. (2001) and is based on estimates ($F$) of shared ancestry among populations. This method computes estimates for pairs of populations, which might be advantageous for detecting selection at a local
scale (Vitalis et al., 2001). Some more advantages of the shared-ancestry method of Vitalis et al. (2001) include: (1) allowing for historical changes in effective population size, such as range expansions or reductions; (2) being robust when moderate gene flow among populations is considered; and (3) having a higher resolution to identify selection in a single or reduced number of populations through pairwise-population analysis (Akey et al., 2002). Outlier-detection approaches have been applied to several organisms, including oaks (Scotti-Saintagne et al., 2004b) and pine (our unpublished results). These studies have revealed that intraspecific positive selection might be widespread in nature. For instance, four out of 55 SNPs (approx. 7%) in loblolly pine (P. taeda) have shown a level of genetic differentiation among populations sevenfold the species average, and were probably affected by natural selection (Fig. 1).

Despite the potential of outlier-based methods to detect selection in natural populations, it must be noted that it can be extremely difficult to verify whether all variation in the genes that behave as outliers are genuinely under adaptive selection (Luikart et al., 2003). To overcome this drawback, these methods should be used in combination with coalescence-based methods, association mapping and gene-expression studies, and repeated in different environments and/or species. In addition, the existing statistical tools need to be improved in order to exploit the full power of outlier-detection methods.

Integrating population genomics and related approaches

Understanding the molecular basis of adaptation and the evolutionary processes responsible for shaping gene diversity in forest trees requires integrating population genomics approaches and related disciplines. The coming of age of forest-tree genomics and biotechnology (reviewed by Campbell et al., 2003; Krutovsky & Neale, 2005a) and new exciting developments of evolutionary theory, such as extended coalescence models (Rosenberg & Nordborg, 2002) and Bayesian inference (Beaumont & Rannala, 2004), make multistage integrated approaches possible. First, candidate loci for adaptive traits and control regions must be selected. Evolutionary and ecological functional genomics (Feder & Mitchell-Olds, 2003; Purugganan & Gibson, 2003) through, for instance, transcription profiling (Gibson, 2002) can provide valuable lists of target candidate genes. Second, alleles (haplotypes) at candidate gene loci and their effects on phenotypes need to be characterized via association mapping. At this stage, large-scale gene-expression studies, like those based on microarray technology, can provide detailed transcriptional profiles and insights on gene interactions and regulatory control. Third, frequencies of alleles in native populations must be estimated to identify patterns of adaptive variation across heterogeneous environments. Detailed knowledge of how landscape features structure populations is the subject of landscape genetics, a newborn discipline that addresses the interaction between the spatial ecological processes and microevolutionary processes, such as gene flow, genetic drift and selection (Manel et al., 2003). Allelic frequency distributions of candidate genes underlying adaptive traits might be correlated with edaphic or altitudinal clines, similarly to the clinal organization of phenotypic variation described in several forest trees (Hamrick, 2004 and references therein). Furthermore, genetic differentiation among populations might reveal local selective pressures resulting in adaptive divergence, or identify the geographical range where a previously characterized mutation has been favoured by natural selection (see Storz, 2005 for review).

Finally, community and ecosystem genetics approaches (Agrawal, 2003; Whitham et al., 2003; Vellend & Geber, 2005), which focus on how intraspecific genetic variation of keystone organisms can affect dependent species, community organization and ecosystem dynamics, are necessary to understand complex natural systems, extending single-tree studies to an ecosystem-wide level.

Forest trees are good models for population genomics

Forest trees are convenient study organisms for population genomic studies for several reasons: (1) they are relatively undomesticated and have abundant genetic and phenotypic variation, unlike many crop plants that have been through domestication bottlenecks; (2) they are open-pollinated and typically show low-to-moderate LD, making it easier to identify genes controlling complex traits; and (3) unlike other undomesticated plants, traditional tree breeding provides a large infrastructure for evaluating complex trait variation in replicated genetic tests across different environments. High levels of individual heterozygosity in forest trees facilitate the use of $F_t$ crosses in genetic and QTL mapping and more complex mating designs are usually not required to produce segregating mapping populations in forest trees.
Trees are long-lived, sessile organisms that occupy extensive landscapes. Many forest trees, including gymnosperms such as cycads and conifers, are among the most ancient seed plants, dating back to the Devonian period (400–360 million yr ago). As a result of recent speciation, there are also several modern tree species. For example, some species of the diverse *Inga* genus of neotropical rainforest trees might have evolved in the past approx. 2 million yr (Richardson *et al*., 2001). Intraspecific genetic diversity of dominant or keystone tree species may have ecosystem-wide consequences through their extended phenotypes (see review in Whitham *et al*., 2003), being relevant for global biodiversity conservation. Despite the ancient use of forests by humans, there is still abundant genetic variation present in natural populations of trees. Reviews based on molecular markers report higher genetic variation in trees than in other plant species (Hamrick *et al*., 1992; Nybom & Bartish, 2000). Recent studies based on DNA-sequence data for several loci also showed a considerable amount of genetic variation still present in trees (Table 1), even in intensively managed species such as loblolly or maritime pines.

Poplars (*Populus* spp.) and conifers (e.g. *Pinus* spp., *Pseudotsuga menziesii* and *Cryptomeria japonica*) are the best candidates for nonclassical model eukaryotes for population, evolutionary and ecological genomic studies (Feder & Mitchell-Olds, 2003; Neale & Savolainen, 2004). As a consequence, several multidisciplinary projects have been developed recently to unveil adaptive variation in these species (e.g. ADEPT, http://dendrome.ucdavis.edu/adept; TREESNIPS, http://cc.oulu.fi/~genetwww/treesnips; DIGENFOR, http://www.pierroton.inra.fr/biogecco/genetique/projets/europe/digenfor; POPYOMICS, http://www.soton.ac.uk/~popyomic). A complete poplar genome (*P. trichocarpa*), four times larger than the *Arabidopsis* genome, has been sequenced and made publicly available (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html). Conifers represent a widespread group with an important ecological role in terrestrial ecosystems, including some species that also have a high commercial value (e.g. *Pinus taeda*, *Pinus radiata*, *Pinus sylvestris*, *Pinus pinaster*, *Cryptomeria japonica*, *Picea abies*, *Pseudotsuga menziesii*). Conifers have a unique reproductive system with a haploid megagametophyte (the nutritious mother tissue of a seed) originating from a maternal gamete that can be used for direct sequencing and haplotype determination. The ancient evolutionary history, low domestication, large open-pollinated native populations and high levels of both genetic and phenotypic variation make conifers almost ideal species for the study of adaptive evolution using population genomics approaches.

### Population genomics and forest-tree conservation genetics

Preservation of adaptive polymorphisms and divergent populations are major goals of genetic conservation (Frankham *et al*., 2002; Moritz, 2002). Population genomics studies can play an important role in the selection of populations for *in situ* genetic reserves or for establishing *ex situ* conservation plantations. The widespread use of population genomics approaches in the future, along with new developments in functional genomics, would increase our understanding of the molecular basis of adaptation and also provide us tools for molecular breeding strategies in trees. Identification of allele-specific effects in hundreds or thousands of genes via association mapping and other population genomics approaches would help to understand local adaptive structure. The estimation of allele frequencies in natural populations would provide the spatial framework to unveil the action of natural selection in the wild and to correlate environmental and allelic variation.

The adoption of population genomics methods would also correct biases in current conservation genetic studies by, first, increasing the number of informative as well as neutral (such as nuSSRs) markers available for population analysis and guaranteeing a better representation of the genome. Current estimates of genetic diversity based on a limited number of markers (typically approx. 6–10 nuSSRs) might be severely biased (Mariette *et al*., 2002). Second, population genomics analysis can help to detect loci that are under strong selection and remove them from studies of demographic or historical processes. Otherwise, biases up to 60% in genetic differentiation estimates ($F_{st}$) could be obtained because of the inclusion of outlier loci that might be under selection (reviewed by Luikart *et al*., 2003). Removing outliers would be also useful to improve the adjustment of test statistics (such as Tajima’s $D$) to the distributions expected under different demographic models (Schmid *et al*., 2005).

Transference of information from model to nonmodel tree species, development of integrated approaches for understanding adaptive variation (such as those reviewed here), and deciphering allelic effects on single phenotypes are basic elements of new-generation conservation strategies based on population genomics.

### Acknowledgements

We thank M.T. Cervera, R. Alía and J. Climent for valuable comments on the manuscript. The work of S.C. González-Martín was supported by a Fulbright/MECD scholarship at University of California (Davis) and the ‘Ramón y Cajal’ fellowship RC02-2941. This research was supported by the ADEPT (Allele Discovery for Genes Controlling Economic Traits in Loblolly Pine) project funded in the framework of the Initiative for Future Agriculture and Food Systems (USDA, USA).

### References


About New Phytologist

- New Phytologist is owned by a non-profit-making charitable trust dedicated to the promotion of plant science, facilitating projects from symposia to open access for our Tansley reviews. Complete information is available at www.newphytologist.org.

- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication ‘as-ready’ via OnlineEarly – the 2004 average submission to decision time was just 30 days. Online-only colour is free, and essential print colour costs will be met if necessary. We also provide 25 offprints as well as a PDF for each article.

- For online summaries and ToC alerts, go to the website and click on ‘Journal online’. You can take out a personal subscription to the journal for a fraction of the institutional price. Rates start at £109 in Europe/$202 in the USA & Canada for the online edition (click on ‘Subscribe’ at the website).

- If you have any questions, do get in touch with Central Office (newphytol@lancaster.ac.uk; tel +44 1524 594691) or, for a local contact in North America, the US Office (newphytol@ornl.gov; tel +1 865 576 5261).