Review on single nucleotide polymorphisms (SNPs) and population genetic studies in conifer species

Federico Sebastiani, Santiago C. González-Martínez, Giovanni Giuseppe Vendramin

Bioversity International

Conifers Network

Report of the fourth meeting (18–20 October 2003, Pitlochry, United Kingdom)

J. Koskela, C.J.A. Samuel, Cs. Mátyás, and B. Fady, *compilers*

EUFORGEN

European Forest Genetic Resources Programme (EUFORGEN)

IPGRI and INIBAP operate under the name Bioversity International

Supported by the CGIAR

Conifers Network

Report of the fourth meeting (18–20 October 2003, Pitlochry, United Kingdom)

J. Koskela, C.J.A. Samuel, Cs. Mátyás, and B. Fady, *compilers*

European Forest Genetic Resources Programme (EUFORGEN)

Bioversity International is an independent international scientific organization that seeks to improve the well-being of present and future generations of people by enhancing conservation and the deployment of agricultural biodiversity on farms and in forests. It is one of 15 centres supported by the Consultative Group on International Agricultural Research (CGIAR), an association of public and private members who support efforts to mobilize cutting-edge science to reduce hunger and poverty, improve human nutrition and health, and protect the environment. Bioversity has its headquarters in Maccarese, near Rome, Italy, with offices in more than 20 other countries worldwide. The Institute operates through four programmes: Diversity for Livelihoods, Understanding and Managing Biodiversity, Global Partnerships, and Commodities for Livelihoods.

The international status of Bioversity is conferred under an Establishment Agreement which, by January 2007, had been signed by the Governments of Algeria, Australia, Belgium, Benin, Bolivia, Brazil, Burkina Faso, Cameroon, Chile, China, Congo, Costa Rica, Côte d'Ivoire, Cyprus, Czech Republic, Denmark, Ecuador, Egypt, Greece, Guinea, Hungary, India, Indonesia, Iran, Israel, Italy, Jordan, Kenya, Malaysia, Mali, Mauritania, Morocco, Norway, Pakistan, Panama, Peru, Poland, Portugal, Romania, Russia, Senegal, Slovakia, Sudan, Switzerland, Syria, Tunisia, Turkey, Uganda and Ukraine.

Financial support for Bioversity's research is provided by more than 150 donors, including governments, private foundations and international organizations. For details of donors and research activities please see Bioversity's Annual Reports, which are available in printed form on request from bioversity-cgiar.org or from Bioversity's website (www.bioversityinternational.org).

European Forest Genetic Resources Programme (EUFORGEN) is a collaborative programme among European countries aimed at ensuring effective conservation and sustainable utilization of forest genetic resources in Europe. It was established in 1994 to implement Resolution 2 of the Strasbourg Ministerial Conference for the Protection of Forests in Europe. EUFORGEN is financed by participating countries and is coordinated by Bioversity International, in collaboration with the Forestry Department of FAO. It facilitates the dissemination of information and various collaborative initiatives. The Programme operates through Networks in which scientists, managers and policy-makers work together to analyse needs, exchange experiences and develop conservation methods for selected tree species. The Networks also contribute to the development of appropriate conservation strategies for the ecosystems to which these species belong and promote integration of gene conservation into sustainable forest management. Network members and other experts from the participating countries carry out an agreed workplan with their own resources as inputs in kind to the Programme. EUFORGEN is overseen by a Steering Committee composed of National Coordinators nominated by the member countries. Further information on EUFORGEN can be found from its website (www.euforgen.org).

Citation: Koskela, J., Samuel, C.J.A., Mátyás, Cs. and Fady, B., compilers. 2007. Conifers Network, Report of the fourth meeting (18-20 October 2003, Pitlochry, United Kingdom). Bioversity International, Rome, Italy. iv+71 pp.

ISBN 978-92-9043-768-0

Bioversity International
Via dei Tre Denari, 472/a
00057 Maccarese
Rome, Italy

Contents

Review on single nucleotide polymorphisms (SNPs) and population genetic studies in conifer species

Federico Sebastiani,[1] Santiago C. González-Martínez[2] and Giovanni G. Vendramin[3]
[1]Dipartimento di Biotecnologie Agrarie, Genexpress, Università degli Studi di Firenze, via della Lastruccia 14, 50019 Sesto Fiorentino (FI), Italy
[2]Departamento de Sistemas y Recursos Forestales, Centro de Investigación Forestal, INIA, Carretera de La Coruña, km 7.5, 28040 Madrid, Spain
[3]Istituto di Genetica Vegetale, CNR, via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), Italy

Introduction

Forest biodiversity is one of the main components of terrestrial ecosystems. All together, tropical, temperate and boreal forests offer diverse habitats for plants, animals and micro-organisms, holding a vast amount of the World's terrestrial biodiversity. In view of growing concern about human impact and expected climatic changes, the maintenance of mechanisms generating diversity in forests has become a central issue as it determines the stability of terrestrial ecosystems and the sustainability of their resource. It is generally accepted that maintaining or restoring biodiversity is a basic precondition to give a long-term, evolutionary answer to these challenges.

Diversity can be studied at different levels, but DNA sequence data represent the highest level of genetic resolution (Järvinen *et al*. 2003). Genetic markers have been extensively used during the past two to three decades to unravel patterns of distribution of genetic diversity and infer possible mechanisms of plant evolution. The use of neutral molecular markers (e.g. simple sequence repeats, SSRs) has allowed historical patterns and the role of evolutionary forces—such as genetic drift—to be depicted because these phenomena affect all markers in similar ways. However, they are not well suited to providing information about variation in functional regions, where selection operates. Many clines for adaptive variation have been described in forest trees, e.g. bud set in *Pinus sylvestris* L. (García-Gil *et al*. 2003) and they are likely to be the products of variation in a limited set of genes and biosynthetic pathways that might not be detected by correlation with random neutral genetic markers. Thus, studies of genetic diversity could benefit from targeting genes that exhibit relevant variation on adaptive traits, rather than targeting random markers (Brumfield *et al*. 2003; Morin *et a*l. 2004; van Tienderen *et al*. 2002; see review for forest trees in Krutovsky and Neale 2001).

Dissection of complex adaptive traits in plants, including forest trees, was traditionally undertaken through genetic linkage analysis (quantitative trait loci [QTL] mapping) based on DNA polymorphisms in highly structured populations with known pedigrees. The study of the pattern of variation of adaptive traits benefited from classical tree breeding experiments, which provided information about families particularly adapted to specific environmental conditions. The opportunity to bring together population genetic and functional genomic studies by identifying candidate genes controlling target traits or underlying QTLs has been made possible by: (1) the development of molecular markers in functional regions (genes, promoters, etc.), such as SNPs (single nucleotide polymorphisms), (2) the discovery of candidate genes via transcript profiling, and (3) an extraordinary effort in EST (expressed sequence tags) sequencing (e.g. more than 250 000 ESTs are already available in pine). Subsequently, polymorphisms that are in strong linkage disequilibrium (LD) with phenotypic traits (i.e. phenotype–genotype genetic associations) have been revealed by using natural populations; this has been termed 'association mapping', e.g. see Plomion *et al*. (2003) and Neale and Savolainen (2004). Association mapping was first developed in humans, where classical approaches are not feasible, and was recently extended to plants.

Natural populations can be used to map traits by means of association analysis. Association mapping takes advantage of LD created after many generations of recombination and random mating. In these conditions, only tightly linked loci will show statistical association, allowing finer mapping than standard QTL approaches. To avoid false association, it is extremely important to have detailed information on basic population parameters, such as the extent of LD, the level of genetic variation and the degree of population structure; and also how these parameters vary across the genome of the target species. In principle, association studies can identify variation down to the single-nucleotide substitutions that are responsible for variation in phenotypes (QTNs, quantitative trait nucleotides) (Ingvarsson 2005).

Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) are Mendelian, co-dominant markers that can be analysed by any statistical method that is based on genotype frequencies. They are the result of single point mutations that produce base-pair alternative sequences (alleles) in genomic DNA. SNPs are abundant and widespread in the genomes of the species studied so far; for example, they represent 90% of the genetic variation detected in the human genome. Moreover, they can potentially be associated with adaptive traits. They are more frequent in non-coding regions than in coding ones but the mean frequency varies greatly among species (Table 1). Usually, SNPs have only two alleles, but SNPs with three allelic variants appear at low frequency (~1–2% in *Pinus taeda* L.).

The use of SNPs as molecular markers became possible only recently thanks to sequencing projects of model species (e.g. *Arabidopsis* and *Populus*) that produced redundant databases highlighting the prevalence of nucleotide polymorphism in the genomes. Moreover, rapid progress in sequencing technology has made it easier to collect many sequences by automating the processes and reducing the costs.

There are two main steps that need to be followed to use SNP markers: SNP discovery and SNP genotyping.

Table 1. Frequency of single nucleotide polymorphisms (SNPs) in different plant species, including several forest trees

Plant species	SNPs/kb†	Reference(s)
Beta vulgaris L.	8	Schneider et al. 2001
Picea abies (L.) H. Karst.	13	De Paoli and Morgante 2004
Pseudotsuga menziesii (Mirb.) Franco	26	Krutovsky and Neale 2005
Pinus taeda L.	16	Brown et al. 2004
Pinus pinaster Ait.	6–10	Le Dantec et al. 2004; Pot et al. 2005
Pinus radiata D. Don	3	Pot et al. 2005
Pinus halepensis Mill.	10	Sebastiani et al. unpublished data
Populus tremula L.	60	Ingvarsson 2004
Populus nigra L.	10	Zaina and Morgante 2004
Glycine max Merr.	3	Zhu et al. 2003
Zea mays L.	16	Ching et al. 2002

† single nucleotide polymorphisms per kilo base pair

SNP discovery

SNP discovery is the process of finding the polymorphic sites in the genome of the species and populations of interest. There is not a single way to discover SNPs, and different approaches may be adopted depending on the availability of DNA sequence information. These include the re-sequencing of polymerase chain reaction (PCR) amplicons (fragments) with or without pre-screening, electronic SNP (eSNP) discovery in shotgun genomic libraries and eSNP discovery in EST libraries (Rafalski 2002a). The strategies adopted to develop SNP markers differ between model and non-model species (a model species is one that is extensively studied to understand particular biological phenomena, with the expectation that discoveries made will provide insight into the workings of other organisms, i.e. the non-model species).

For model species, redundant overlapping databases exist, derived from sequencing, ESTs or large-scale SNP identification projects, making it possible to directly retrieve SNPs for the genes of interest. The increasing availability of software and databases is helping to facilitate SNP discovery enormously through the implementation of automatic platforms (Le Dantec *et al*. 2004). Some of these are able to provide transcript profiling information; for instance, the MAGIC Gene Discovery tool (Cordonnier-Pratt *et al*. 2004), available at http://fungen.org/genediscovery/.

For non-model species, the most direct way to identify SNPs is to sequence a genome fragment from multiple individuals. Candidate fragments for different genes can be obtained from model species or expressional studies. In order to avoid ascertainment bias in allele frequencies, it is advisable to obtain high-quality sequence from a relatively large sample of individuals representing all the populations in the study. To avoid sequencing errors and low-quality sequence due to the presence of repetitive

regions, sequencing from both ends is also advisable. An inexpensive pre-screening, e.g. single strand conformation polymorphism (SSCP) of several samples can facilitate the choice of the most informative loci and provide a preliminary estimate of the level of polymorphism. Unfortunately, these pre-screening methods for SNP detection are often labour-intensive and not very sensitive. Pre-screening may be necessary for species like soybean, where the rate of SNP is low, but more rarely in forest trees which generally show high levels of standing nucleotide variation.

The high level of conservation of gene sequences across species facilitates the design of primers to amplify orthologous gene regions in related species, starting from information available in model species. PCR primers (a nucleic acid strand, or a related molecule that serves as a starting point for DNA replication) are carefully designed to amplify the loci of interest, excluding any other member of the same gene family. The PCR products are then sequenced in both directions and the resulting sequences are aligned. Taking care to distinguish true polymorphisms from sequencing errors, polymorphisms are identified (Figure 1).

For those non-model species where it is not possible to amplify orthologous genes, a random sequence approach can be followed. This approach involves sequencing anonymous nuclear loci

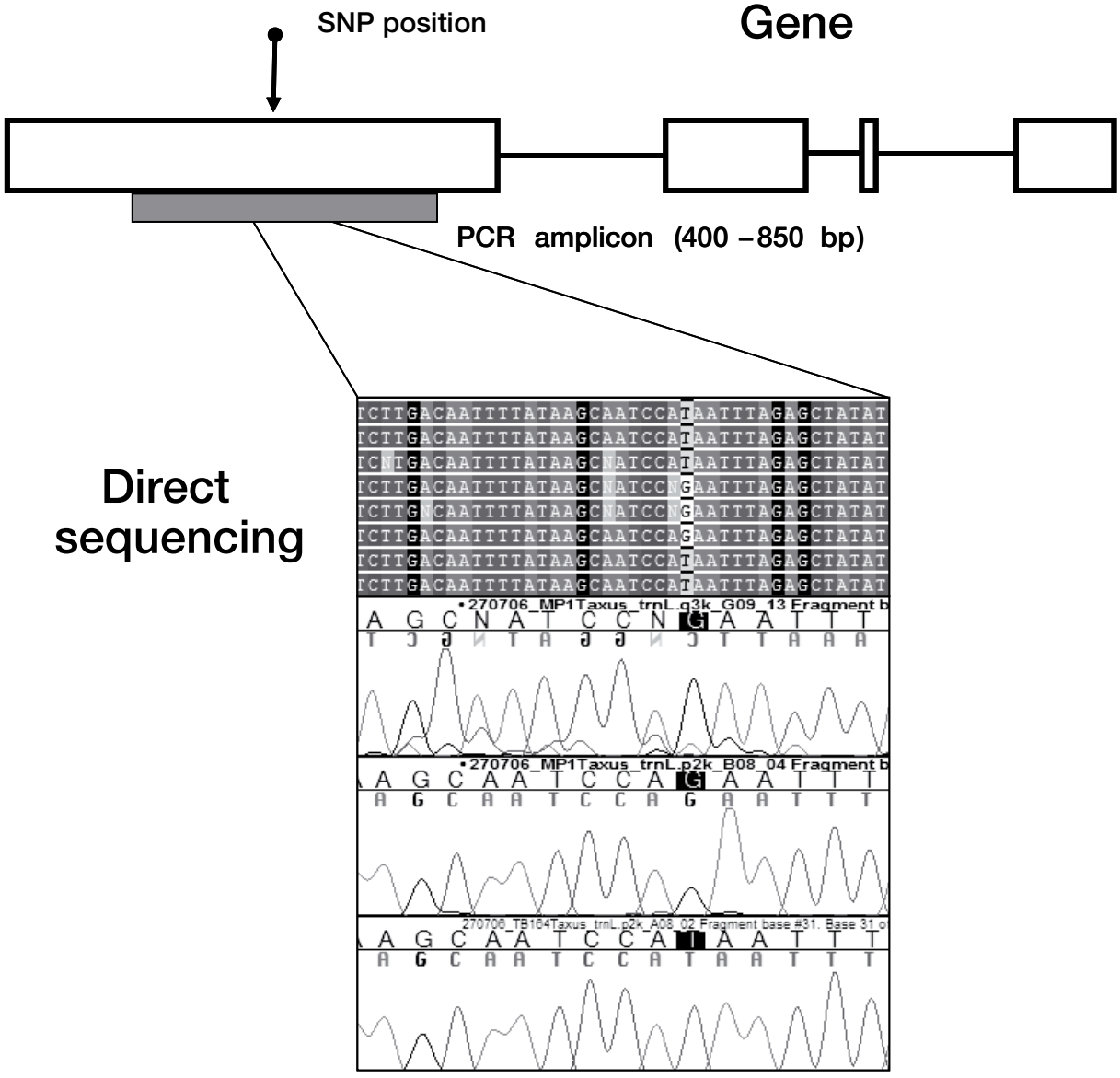

Figure 1. DNA sequence analysis for SNP discovery. After amplification from different individuals, PCR fragments (i.e. amplicons) are purified and sequenced from both ends. Sequences are then aligned and SNPs (and insertions/deletions) are identified. In this example, the SNP (G/T) is indicated in black.

derived from a genomic or cDNA (cytoplasmic DNA) library or produced through amplified fragment length polymorphisms (AFLPs) and in designing PCR primers that can be used to screen multiple individuals and find polymorphic SNP loci.

SNP genotyping

Multiple technologies are available for SNP genotyping. The choice of the method depends on the study to be performed and other criteria, such as cost, throughput level and equipment available.

PCR-RFLP (restriction fragment length polymorphism) and PCR-SSCP (single strand conformation polymorphism) are the traditional techniques used to gather information at the sequence level. The main advantage of these methods is that they only require standard laboratory equipment, and can therefore be used in every laboratory; the main drawback is that they are not well suited to high throughput genotyping.

Another category of methods relies on the direct acquisition of sequence information. One way is the traditional sequencing approach based on the standard method with dideoxy fluorescent label nucleotides. This method is useful for SNP discovery but can be expensive for genotyping, with the exception of those cases where SNP frequency is high, and therefore sequencing allows the genotyping of multiple loci at the same time. An alternative, recent method to genotype multiple loci at the same time is eco-tilling, a low-cost technique for rapid identification of haplotypes (Comai *et al*. 2004).

Fluorescent-based sequencing has an increasing role in SNP analysis because of the development and availability of programs that automate the base calling, assembly and finishing of sequences; such as: Phred, Phrap and Consed (see www.phrap.org/phredphrapconsed.html). Polyphred is another program that operates together with Phred, Phrap and Consed to identify SNPs as high-quality base mismatches in assembled sequences. Importantly, Polyphred can also detect SNPs as heterozygotes (two bases at a single position in the sequence) in diploid sequences amplified by PCR (Brumfield *et al*. 2003).

Newer methods for allelic discrimination are based on primer extension. There are numerous variations in the primer extension approach that are based on the ability of DNA polymerase to incorporate specific deoxyribonucleosides complementary to the sequence of the template DNA. However, all these methods can be grouped into two categories. The first one is a mini-sequencing technology, a single base extension (SBE) where the identity of the polymorphic base in the target DNA is determined. In this case, only dideoxy nucleotides are used, causing the addition of only the complementary nucleotide. The second one, called allele specific primer extension (ASPE), is an allele-specific PCR approach where the DNA polymerase is used with deoxy nucleotides to amplify the target DNA; the PCR product will be obtained only if the primers are perfectly complementary to the target DNA sequence.

Several ingenious methods have been devised which differ in the way they monitor the reaction. Most of these approaches combine novel nucleic acid analogues and new methods of monitoring differences in physical properties between starting reagents and primer extension products. Alleles can be sorted and detected using various methods; including gel electrophoresis, macro and microarrays and fluorescence polarization. These genotyping methods are suited to automatic machines, such as automatic sequencers, and can ensure medium to high throughput results. Finally, several commercial high-throughput genotyping platforms that can handle as many as 100 000 assays simultaneously have been developed (Hirschhorn and Daly 2005). A more comprehensive survey of SNP genotyping methods can be found in Kwok (2001).

Functional vs neutral genetic markers

The main difference between a random marker and a functional marker is the distance of the mutation causing the phenotypic effect in the trait of interest (van Tienderen *et al*. 2002). During the last 20 years, the majority of studies aiming to monitor the level and distribution of genetic diversity in natural populations were based on the use of neutral markers. Indeed, molecular markers, such as allozymes or microsatellites, provided useful information on historical demography and population evolution. However, neutral markers do not generally reflect selective processes or are related to fitness which is an indicator of the level of adaptive variation within populations and therefore of the adaptive potential of populations to changing environments (Morin *et al*. 2004).

SNPs are particularly useful markers for finding genes under selection and studying the dynamics of these genes in natural populations. SNPs are robust markers, easy to score and widespread in the

genome. The availability of high-density markers, such as SNPs, opens the possibility of studying, by association genetics, the molecular basis of complex quantitative traits in natural populations of plants, taking advantage of the fact that genetic markers in close proximity to mutant genes may be in LD to them. Association studies can be carried out using a genome-wide approach (without assuming one region of the genome to be more likely to harbour the associated genetic factor than another) or with a candidate gene approach (using biological knowledge to prioritize some fragments of the genome for the study). The magnitude and distribution of LD determine the choice of association mapping methodology. When LD declines slowly with increasing distance from the mutation or gene responsible of the phenotype even a low density of markers is sufficient to identify associated markers. When LD declines rapidly around the causative gene, a much greater density of markers is required to identify an associated marker (Rafalski 2002b). Extension and distribution of LD depend on many factors including population history (e.g. the presence of population bottlenecks or admixture) and the frequency of recombination. First studies on forest tree species revealed a rapid decay in LD with distance. LD declines to negligible levels in <500 bp (base pairs) in *Populus tremula* L., although in some cases LD extends in local populations up to 1 kb (Ingvarsson 2005). Similarly, LD declines very rapidly within 200 bp in *Picea abies* (L.) H. Karst. (De Paoli and Morgante 2004) as well as in *Pseudotsuga menziesii* (Mirb.) Franco (Krutovsky and Neale 2005), *Pinus halepensis* Mill. (Sebastiani *et al*. unpublished data) and *Pinus taeda* (Brown *et al*. 2004). A rapid decay of LD in forest trees is consistent with what is expected from outcrossing species with large effective population size, and is in strong contrast with what is observed in self-fertilizing plant species. For example, in *Arabidopsis thaliana* (L.) Heynh. selfing dramatically reduces the effective recombination and LD extends up to 250 kb (Nordborg and Tavaré 2002). Some variation among genes in LD has been observed in barley (Lin *et al*. 2002), pines (Brown *et al*. 2004) and maize (Remington *et al*. 2001), suggesting that, in addition to mating systems, other factors such as demographic history or variable recombination rates across the genome, can play an important role.

The association mapping approach is very promising for long-lived, relatively undomesticated forest trees where the general high level of genetic variability of natural stands can be successfully used to identify markers linked to economically and ecologically relevant traits. Moreover, as forest trees are predominantly outcrossing organisms characterised by large effective population size, they generally show LD extending only a few hundred base pairs. For these reasons and considering that in some species (e.g. some conifers) the genomes are extremely large (> $1x10^{10}$ bp), the whole-genome scan approach is not feasible because of the too high number of SNPs required for adequate genome coverage. On the other hand, very fine-scale mapping is possible if candidate gene approaches are used and it might even be more advisable given the high variation found in tree genomes (Ingvarsson 2005; Neale and Savolainen 2004).

A limiting step in forest trees is the choice of candidate genes. The choice of appropriate candidate genes can be facilitated by the availability of information about the biochemical and/or physiological pathways related to the trait of interest, i.e. by selecting genes involved in these pathways. Unfortunately, this information is rarely available for forest trees; therefore, sequences of genes identified in model species, such as *Arabidopsis thaliana,* are used to design consensus or degenerated primers for the amplification of orthologous loci. The recent sequence of the complete poplar (*Populus trichocarpa* Torr. & Gray) genome (see http://genome.jgi-psf.org/Poptr1/Poptr1.home.html), which is four times larger than the *Arabidopsis* genome, opens exciting new possibilities to identify novel genes in forest trees. Other methods to identify candidate genes exist, e.g. via transcript profiling through cDNA and oligonucleotide microarrays (for more details see Pflieger *et al*. 2001). However, it should be stressed that substantial work is still needed to elucidate the functional role of genes and for the successful transfer to non-model species.

Both association mapping approaches (genome-scan and candidate genes) require abundant SNPs in the studied species and populations. For this reason, preliminary analyses to assess nucleotide diversity in different species and association populations are required. First estimates indicate that nucleotide diversity varies considerably between plant species, from the highest (maize) to the lowest (the highly-domesticated soybean). Interestingly, some conifers, such as *Cryptomeria japonica* (L.f.) Don (Kado *et al*. 2003) and pines, e.g. *Pinus sylvestris* (Dvornik *et al*. 2002) and *Pinus taeda* (Brown *et al*. 2004) are not amongst the most variable species, contradicting expectations from the results obtained using neutral markers and their life history characteristics. First evidences showed that broadleaved genera, e.g. *Populus* (Ingvarsson 2005) and *Quercus* (Pot *et al*. 2005) might display higher nucleotide diversity than pines or *Cryptomeria*. In fact, *Populus* displays about 2- to 10-fold higher nucleotide polymorphism than *Pinus* or *Cryptomeria*. However, other conifers, such as *Pseudotsuga menziesii*, showed levels of variation

comparable to broadleaved species (Krutovsky and Neale 2005). The high level of variation detected in *Populus* and *Quercus* is in agreement with earlier studies based on allozyme analysis (Jelinski and Cheliak 1992; Petit *et al*. 1995).

Markers in specific functional regions of the genome need to be statistically analyzed in order to test for the possibility that these regions might have experienced different selective pressures. In unstructured populations, standard neutrality tests might be applied. When variation is structured in populations, an interesting and relatively easy approach is the comparison of genetic differentiation estimates, such as Wright's *F*-statistics among markers tagging a putative gene under selection and neutral markers, or expected distributions computed using coalescence theory (see reviews in van Tienderen *et al*. (2002) and Luikart *et al*. (2003)). If population divergence (F_{st}) is higher for the gene-targeted marker with respect to divergence estimates obtained from random markers, this might indicate divergent selection and local adaptation for the tagged gene (van Tienderen *et al*. 2002). Pot *et al*. (2005) found a higher differentiation among populations at the *Pp1* (glycine-rich protein homologue) gene in *Pinus pinaster* Aiton than in neutral markers. This result is consistent with diversifying selection acting at this locus in this species, which would have lead to the presence of different haplotypes; possibly adapted to local environmental conditions. On the other hand, the absence of differentiation observed for the gene *CesA3* (cellulose synthase) compared with the significant level observed at neutral markers may indicate balancing selection acting on this gene. Note that the presence of significant differentiation among populations may produce spurious associations; therefore, care has to be taken when sampling for association studies.

Currently, SNPs are used primarily in association studies; but their ubiquity, tractable levels of variation and readiness in screening suggests that they will increasingly dominate as markers for elucidating the evolutionary history of populations. Unlike microsatellites, SNPs have relatively low mutation rates. Multiple mutations at a single site are rare, thus facilitating high-throughput genotyping and minimizing recurrent substitutions at a single site (i.e. homoplasy) that would confound the population history (Brumfield *et al*. 2003). Moreover, in conservation genetics, the availability of markers able to detect functional variation could help to define functionally significant units (FSUs), based on differences in allelic frequencies for genes with important ecological functions (van Tienderen *et al*. 2002). FSUs might help managers in conservation biology to identify those conservation units that contain adaptive genetic variation that is worthwhile protecting.

Conclusions

In conclusion, SNPs are becoming the marker of choice in population genetics, ecology and evolution studies because of ease of modelling, genotyping efficiency and genome-wide coverage. Forest tree species, which comprise undomesticated and unstructured large populations where linkage disequilibrium is expected to be limited represent ideal organisms to efficiently apply a candidate-gene based approach to detect association between markers and ecologically and economically important traits.

References

Brown GR, Gill GP, Kuntz R, Langley CH, Neale DB. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. Proceedings of the National Academy of Sciences of the USA 42:15255–15260.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV. 2003. The utility of single nucleotide polymorphisms in inference of population history. Trends in Ecology and Evolution 18:249–256.

Ching A, Caldwell KS, Young M, Dolan M, Smith OSH, Tingey S, *et al*. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3:19.

Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA *et al*. 2004. Efficient discovery of DNA polymorphisms in natural populations by ecotilling. The Plant Journal 37:778–786.

Cordonnier-Pratt MM, Liang C, Wang H, Kolychev DS, Sun F, Freeman R, Sullivan R, Pratt LH. 2004. MAGIC Database and interfaces: an integrated package for gene discovery and expression. Comparative and Functional Genomics 5:268–275.

De Paoli E, Morgante M. 2004. Association study for timing of bud set in *Picea abies*. Atti del Congresso Società Italiana Genetica Agraria, Lecce, p.70.

Dvornik V, Sirviö A, Mikkonen M, Savolainen O. 2002. Low nucleotide diversity at the *palI* locus in the widely distributed *Pinus sylvestris*. Molecular Biology and Evolution 19:179-188.

García-Gil MR, Mikkonen M, Savolainen O. 2003. Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. Molecular Ecology 12:1195–1206.

Hirschhorn JN, Daly MJ. 2005. Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics 6:95–108.

Ingvarsson PK. 2005. Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European Aspen (*Populus tremula* L., Salicaceae). Genetics 169:945–953.

Järvinen P, Lemmetyinen J, Savolainen O, Sopanen T. 2003. DNA sequence variation in BpMADS2 gene in two populations of *Betula pendula*. Molecular Ecology 12:369–384.

Jelinski DE, Cheliak WM. 1992. Genetic diversity and spatial subdivision of *Populus tremuloides* (Salicaceae) in a heterogeneous landscape. American Journal of Botany 79:728–736.

Kado T, Yoshimaru H, Tsumura Y, Tachida H. 2003. DNA variation in a conifer, *Cryptomeria japonica* (Cupressaceae sensu lato). Genetics 164:1547–1559.

Krutovsky KV, Neale DB. 2001. Forest genomics for conserving adaptive genetic diversity. Forest Genetic Resources Working Papers FGR/3. FAO, Rome. Available from: http://www.fao.org/DOCREP/003/X6884E/X6884E00.HTM. Date accessed: 03 October 2006.

Krutovsky KV, Neale DB. 2005. Nucleotide diversity and linkage disequilibrium in cold hardiness and wood quality related candidate genes in Douglas-fir. Genetics 171:2029–2041.

Kwok PY. 2001. Methods for genotyping single nucleotide polymorphism. Annual Review Genomics Human Genetics 2:235–258.

Le Dantec L, Chagné D, Pot D, Cantin O, Garnier-Géré P, Bedon F, *et al*. 2004. Automated SNP detection in expressed sequence tags: statistical considerations and application to maritime pine sequences. Plant Molecular Biology 54:461–470.

Lin JZ, Morrell PL, Clegg MT. 2002. The influence of linkage and inbreeding on patterns of nucleotide sequence diversity at duplicate alcohol deydrogenase loci in wild barley (*Hordeum vulgare* subsp. *spontaneum*). Genetics 162:2007–2015.

Luikart G, England P, Tallmon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: From genotyping to genome typing. Nature Reviews Genetics 4:981–994.

Morin PA, Luikart G, Wayne RK and the SNP workshop group. 2004. SNPs in ecology, evolution and conservation. Trends in Ecology and Evolution 19:208–2169.

Neale D, Savolainen O. 2004. Association genetics of complex traits in conifers. Trends in Plant Science 9:325–330.

Nordborg M, Tavaré S. 2002. Linkage disequilibrium: what history has to tell us. Trends in Genetics 18:83–90.

Petit RJ, Bahrman N, Baradat P. 1995. Comparison of genetic differentiation in maritime pine (*Pinus pinaster* Ait.) estimated using isozyme, total protein and terpenic loci. Heredity 75:382–389.

Pflieger S, Lefebvre V, Causse M. 2001. The candidate gene approach in plant genetics: a review. Molecular Breeding 7:275–291.

Plomion C, Cooke J, Richardson T, MacKay J, Tuskan G. 2003. Report on the forest trees workshop at the plant and animal genome conference. Comparative and Functional Genomics 4:229–238.

Pot D, McMillan L, Echt C, Le Provost G, Garnier-Géré P, Sheree C, *et al*. 2005. Nucleotide variation in genes involved in wood formation in two pine species. New Phytologist 167:101–112.

Rafalski AJ. 2002a. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 5:94–100.

Rafalski AJ. 2002b. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Science 162:329–333.

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, *et al*. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proceedings of the National Academy of Sciences of the USA 98:11479–11484.

Schneider K, Weisshaar B, Borchardt DC, Salamini F. 2001. SNP frequency and allelic haplotype structure of *Beta vulgaris* expressed genes. Molecular Breeding 8:63–74.

van Tienderen PH, de Haan AA, van der Linden CG, Vosman B. 2002. Biodiversity assessment using markers for ecologically important traits. Trends in Ecology and Evolution 17:577–582.

Zaina G, Morgante M. 2004. Nucleotide diversity and linkage disequilibrium in *Populus nigra*. Atti del Congresso Società Italiana Genetica Agraria, Lecce, p.67.

Zhu YL, Song QJ, Hyten DL, Van Tassel CP, Matukumalli LK, Grimm DR, *et al*. 2003. Single-nucleotide polymorphisms in soybean. Genetics 163:1123–1134.

Darius Danusevicius
Department of Forest Genetics and Reforestation
Lithuanian Forest Research Institute
Girionys 1
4312 Kaunas region
Lithuania
Tel: (370-7) 547426
Fax: (370-7) 547446
Email: darius. takas.lt

Vlatko Andonovski
Faculty of Forestry
P.O. Box 235
1000 Skopje R.
Macedonia FYR
Tel: (389-70) 332113
Fax: (389-2) 3079748
Email: unet.com.mk

Eman Calleja
Ministry of Agriculture and Fisheries
Department of Agriculture
Notabile Road
Attard
Malta
Tel: (356) 21415488
Fax: (356) 21493176
Email: eman. magnet.mt

Tore Skrøppa
Norwegian Forest Research Institute
Høgskoleveien 12
1432 Ås
Norway
Tel: (47) 64 949067
Fax: (47) 64942980
Email: tore. skogforsk.no

Jan Matras
Forest Research Institute
Dept. of Genetics & Physiology of Woody Plants
ul. Bitwy Warszawszkiej 1920 roku, 3
00 973 Warsaw
Poland
Tel: (48-22) 7150478
Fax: (48-22) 7150313
Email: bles.waw.pl

Mikhail Pridnya
NIIGORLESECOL
74 Kurortnyi prospect
354002 Sochi-2
Russian Federation
Tel: (7-8622) 621842
Fax: (7-8622) 621842
Email: sochi.ru

Vasilije Isajev
Šumarski Fakultet – Beograd
Kneza Višeslava 1
11030 Belgrade
Serbia and Montenegro
Tel: (381-11) 553 122
Fax: (381-11) 545 485
Email: hotmail.com

Rudolf Bruchánik
State Forest of Slovak Republic
Námestie SNP 8
975 66 Branská Bystrica
Slovakia
Tel: (421-48) 4125903
Fax: (421-48) 4125904
Email: lesy.sk

Gregor Bozic
Slovenian Forestry Institute
Vecna pot 2
1000 Ljubljana
Slovenia
Tel: (386-1) 2007821
Fax: (386-61) 273589
Email: gregor. gozdis.si

Sonia Martín Albertos
Servicio de Material Genético
Dirección General de Conservación de la Naturaleza
Gran Vía de San Francisco, 4
28005 Madrid
Spain
Tel: (34-91) 5964639
Fax: (34-91) 5964905
Email: mma.es

Jonas Bergquist
National Board of Forestry
Skogstyrelsen
55183 Jönköping
Sweden
Tel: (46-36) 155720
Fax: (46-36) 166170
Email: jonas. svo.se

Marcus Ulber
Eidgen. Forschungsanstalt WSL
Zürcherstrasse 111
8903 Birmensdorf
Switzerland
Tel: (41-1) 7392493
Fax: (41-1) 7392215
Email: marcus. wsl.ch

Hacer Semerci
Forest Tree Seeds & Tree Breeding
Research Directorate
P.O. Box 11
06560 Gazi-Ankara
Turkey
Tel: (90-312) 2126519
Fax: (90-312) 2123960
Email: yahoo.com AND
hotmail.com

C.J.A. Samuel
Forest Research
Northern Research Station
EH25 9SY Roslin, Midlothian
United Kingdom
Tel: (44-131) 4456927
Fax: (44-131) 4455124
Email: c.j.a. orestry.gsi.gov.uk

Observers

Jon Loftsson
Iceland Forestry Service
Box 98, IS 700 Egilsstaðir
Iceland
Tel: (354-471) 2100
Fax: (354-471) 2172
Email: skogur.is

Adalstein Sigurgeirsson
Iceland Forestry Service
Box 98, IS 700 Egilsstaðir
Iceland
Tel: (354-471) 2100
Fax: (354-471) 2172
Email: skogur.is

Throstur Eysteinsson
Iceland Forestry Service
Box 98, IS 700 Egilsstaðir
Iceland
Tel: (354-471) 2100
Fax: (354-471) 2172
Email: skogur.is

Cathleen Baldwin
Forest Research
Northern Research Station
EH25 9SY Roslin, Midlothian
United Kingdom
Tel: (44-131) 4456927
Fax: (44-131) 4455124
Email: cathleen. forestry.gsi.gov.uk

Esther Ker
Forest Research
Northern Research Station
EH25 9SY Roslin, Midlothian
United Kingdom
Tel: (44-131) 4456912
Fax: (44-131) 4455124
Email: esther. forestry.gsi.gov.uk

Steve Lee
Forest Research
Northern Research Station
EH25 9SY Roslin, Midlothian
United Kingdom
Tel: (44-131) 4456926
Fax: (44-131) 4455124
Email: steve. orestry.gsi.gov.uk

Rob Sykes
Forest Research
Northern Research Station
EH25 9SY Roslin, Midlothian
United Kingdom
Tel: (44-131) 4456927
Fax: (44-131) 4455124
Email: rob. orestry.gsi.gov.uk

EUFORGEN Secretariat

Jarkko Koskela
EUFORGEN Coordinator
Regional Office for Europe
Bioversity International
Via dei Tre Denari 472/a
00057 Maccarese (Fiumicino)
Italy
Tel: (39) 06 6118223
Fax: (39) 06 61979661
Email: j. cgiar.org

Unable to attend

Lennart Ackzell
National Board of Forestry
Skogstyrelsen
Vallgatan 8
55183 Jönköping
Sweden
Tel: (46-36) 155706
Fax: (46-36) 166170
Email: lennart. svo.se

Bioversity International

ISBN: 978-92-9043-768-0

Cover photo: A Sitka spruce (*Picea sitchensis*) stand, J. Koskela/Bioversity International