Genetic techniques and their application in conservation and management of forest genetic resources

Molecular markers for characterizing diversity in forest trees

G. G. Vendramin & O. K. Hansen

Introduction

It has been widely recognized that loss of genetic diversity is a major threat for the maintenance of the adaptive potential of the species. To preserve and manage this genetic diversity, the ability to identify genetic variation is extremely important. Characterization of diversity has long been based on morphological traits mainly. However, morphological variation is often affected by environmental conditions. Molecular markers avoid many of the complications of environmental effects and for this reason, molecular genetic techniques are nowadays being applied as a complementary strategy to traditional approaches for characterizing genetic resources for *in situ* and *ex situ* conservation.

Due to the recent developments in the field of molecular genetics, a variety of different techniques to analyze genetic variation has emerged (GILLET 1999). These genetic markers may differ with respect to important features, such as, for example, abundance, level of polymorphism detected, reproducibility, technical requirements and costs. Therefore, none of the available techniques is superior to all others for a wide range of applications, but the key-question is rather which marker for which purpose (GILLET 1999). The choice of the most appropriate genetic marker will depend on many factors, *e.g.*, the specific application, the presumed level of polymorphism, the availability of sufficient technical facilities, the efficiency in terms of costs and time.

Molecular markers are not genes in the classical sense; they may have, but often do not have, biological function. Molecular markers arise through mutations, which originates variation in the DNA, often deleterious for the organism. Therefore, markers are mostly found in DNA with no coding/biological function. It follows that molecular markers are better thought of as constant landmarks in the genome. In addition, in order for markers to be useful for conservation genetics, there must be some degree of variation, or polymorphism, at the marker locus. It is just important to keep in mind that the variation at molecular level is not necessarily linked to variation with biologi-

Th. Geburek & J. Turok (eds.)
Conservation and Management of Forest Genetic Resources in Europe
Arbora Publishers, Zvolen, 2005, 337–368, ISBN 80-967088-1-3

cal function.

Reliable information on the distribution of genetic variation is a prerequisite for selection, breeding and conservation programmes for forest trees. Genetic variation of a species is assessed either by measuring morphological and metric characters in the field or by studying molecular markers in the laboratory. Laboratory techniques have, until recently, relied on estimates of genetic diversity and mating system parameters from population surveys using isoenzymes. Isoenzymes continue to provide a relatively simple and inexpensive method of obtaining genetic information. However, their application is limited by the number of enzyme loci (in general low), their relatively low levels of variability, and the fact that they only reveal variation in protein-coding genes and that only a minor proportion of modifications of the quaternary structure of the proteins can be detected by electrophoresis on starch gel (the technique generally used to screen isoenzyme variation). The development of DNA markers, including RFLPs (Restriction Fragment Length Polymorphism), VNTR (Variable Number of Tandem Repeats or minisatellites), SSRs (Simple Sequence Repeats or microsatellites), RAPDs (Random Amplified Polymorphic DNA), AFLPs (Amplified Fragment Length Polymorphism), SSCP (Single Stranded Conformation Polymorphism) and SNPs (Single Nucleotide Polymorphism), has overcome limitations on the number of variable loci and provided the tools to study variation in coding, non-coding, and highly variable regions of both nuclear or organelle (chloroplast and mitochondrial) genomes.

Since the advent of recombinant DNA technology in population genetics in the mid-1980's, the number of genetic markers available for population genetic studies of forest tree species has increased enormously. In order to appreciate this fact, it is interesting to consider the history of genetic marker development in forest genetics. Until the beginning of the 1970's, the only genetic markers available in tree species were the rare morphological traits that could be shown to be controlled by alleles at a single gene locus, such as, for example, the aurea phenotype in Norway spruce (LANGNER 1953). Early attempts to interpret the relative or absolute quantities of the different monoterpenes in the resin of conifer trees (measured using gas chromatography) as genetic markers remained inconclusive due to difficulties in determining mode of inheritance and the probable dependence of their expression on environmental conditions (especially pathogen stress) (HANNOVER 1992, BARADAT *et al.* 1995). It was not until the early 1970's, when BARTELS (1971a and b) and BERGMANN (1973, 1974a and b, 1975a and b, 1978) developed enzyme electrophoresis for Norway spruce (*Picea abies* (L.) Karst.), which direct products of tree DNA were made accessible also in forest trees. Inheritance analysis of the resulting banding patterns enabled inference of their mode of inheritance and, consequently, allowed them to be used as genetic markers.

Genetic markers: characteristics, advantages and disadvantages

Morphological genetic markers were the first markers to be used in studies of inheritance and phylogenetic relationships, simply because it is possible to observe the marker directly expressed in the organism. The molecular markers traditionally used

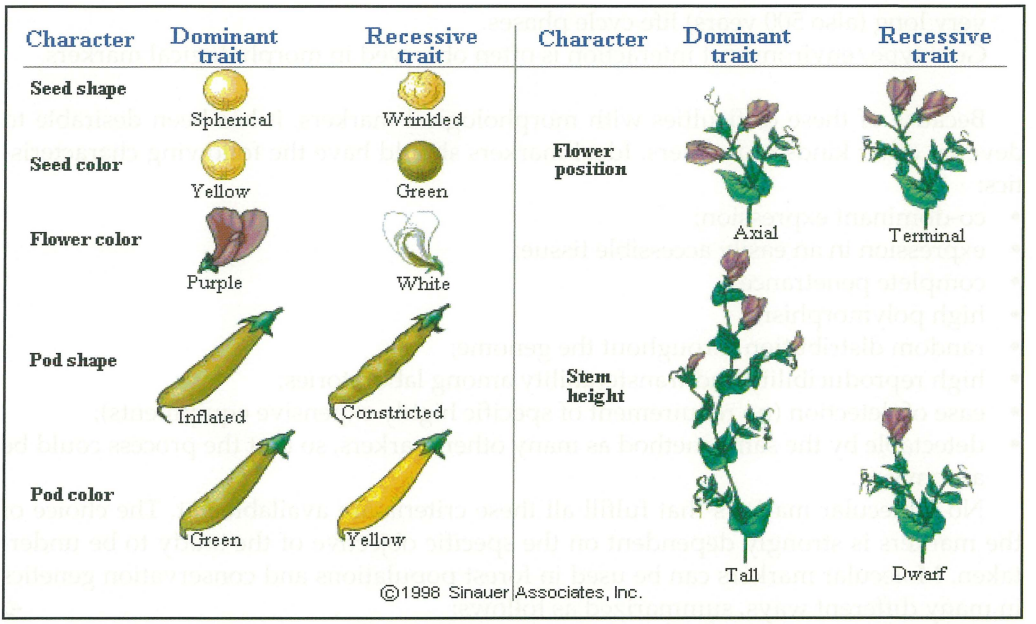

Figure 1. Summary of the basic studies on inheritance of morphological traits of *Pisum sativum* performed by Gregor Mendel [from PURVES *et al.* (1998), © Sinauer Associates, reproduced with kind permission].

in genetic studies vary in a discrete manner: an example was the basic studies of inheritance performed by Gregor Mendel made on different easily and certainly recognizable traits in *Pisum sativum* L.

Examples of these traits are seed shape, seed colour, pod shape, flower colour, pod colour, flower position and stem height (Fig. 1).

Another well-known example of a morphological genetic marker is eye color in humans. However, most morphological traits in organisms are influenced by the environment and controlled by many genes, and therefore not suitable as genetic markers. For example, height and diameter of trees are dependent on many genes and much influenced by the growing site. In general, morphological markers that display Mendelian inheritance are rare in trees.

Morphological markers are difficult to use in forest genetics for the following reasons:

- they are often recessive in nature, therefore heterozygotes are not identifiable;
- they are generally mutations and they often confer a deleterious phenotype to the organism, therefore decreasing the fitness;
- they may also exhibit epistatic effects (interaction between genes), or pleiotrophy, or incomplete penetrance, or their expression may be affected by environmental conditions;
- they may confer a phenotype that is only apparent at one stage of an organism's development, *e.g.*, at maturity; a relevant problem in forest trees characterized by

very long (also 500 years) life cycle phases.
Genotype/environment interaction is often observed in morphological markers.

Because of these difficulties with morphological markers, it has been desirable to develop other kinds of markers. Ideal markers should have the following characteristics:

- co-dominant expression;
- expression in an easily accessible tissue;
- complete penetrance;
- high polymorphism;
- random distribution throughout the genome;
- high reproducibility and transferability among laboratories;
- ease of detection (no requirement of specific highly expensive equipments);
- detectable by the same method as many other markers, so that the process could be automated.

No molecular markers that fulfill all these criteria are available yet. The choice of the markers is strongly dependent on the specific objective of the study to be undertaken. Molecular markers can be used in forest populations and conservation genetics in many different ways, summarized as follows:

- Description and monitoring of genetic diversity: the inventory of diversity is a necessary step before approaching the dynamics of diversity. The monitoring can be done at different geographic and spatial scales and address both neutral and adaptive traits.
- Identification of the key factors shaping the genetic diversity. This can be achieved following two main approaches: the study of past factors by addressing post-glacial history, and the analysis of current and future factors by studying mating system, gene flow and colonization dynamics.
- Genetic linkage mapping, quantitative trait loci analysis (QTL) and marker assisted selection (MAS) (more important in the field of genetic improvement, but useful also in the specific field of population and conservation genetics).
- Certification purposes (origin of provenances, wood and propagation material).

Isoenzymes

Isoenzymes were first identified about 1950 and were then a great advance for population genetic studies of forest trees due to the very limited availability of morphological traits. Isoenzymes are multiple molecular forms of an enzyme with similar or identical catalytic activities. Different individuals and/or different populations of the same species can have different molecular structures, which can be detected, at least partly, by gel electrophoresis. Isoenzymes are efficient molecular markers, with the main exception that the number of isoenzyme loci that can be identified is very limited, partly because of the detection system (enzyme assays generally performed on a starch gel, a low-resolution system). Further, assay methods are different from those used for other DNA markers. In addition, there is lower allelic variability as there is

for DNA markers. An advantage of isoenzymes is that they are co-dominant in nature. Furthermore, they can be easily transferred and used in all different species (therefore, they can be considered 'universal' markers).

Isoenzymes were used to produce the first inventories of the distribution of genetic diversity in many European forest tree species. They showed their usefulness to describe migration processes in the post-glacial period, as for example in Norway spruce (*Picea abies*) (LAGERCRANTZ & RYMAN 1990, MORGANTE & VENDRAMIN 1991), silver fir (*Abies alba* Mill.) (KONNERT & BERGMAN 1995), beech (*Fagus sylvatica* L.) (GÖMÖRY *et al.* 1998, COMPS *et al.* 2001). The first studies aiming at the estimation of the mating system parameters were also performed using isoenzyme and revealed that generally forest tree species have high outcrossing rates (higher than 90 %, *e.g.*, in Norway spruce, BOSCHERINI *et al.* 1993, and in beech, ROSSI *et al.* 1996). Inheritance of all the studied isoenzyme loci was previously determined by analyzing the haploid, maternal megagametophyte tissue of the conifers or in full-sib and half sib families of angiosperm species.

HAMRICK and GODT (1990) produced the most comprehensive review of plant isoenzyme literature, which provided many insights into population structure and breeding systems. They used data from about 322 woody taxa to measure genetic diversity within species, and within and among populations, of plants. Life history and ecological traits explain a significant proportion of the total variation among and within species. Geographic range accounted for the larger proportion of variation at the species level. At the population level (variation within species), the plant breeding system in combination with the geographic range is the major factor contributing to the total variation. Woody plants with large distribution ranges, outcrossing breeding system and wind or animal ingested seed dispersal mechanisms are generally characterised by higher level of diversity compared to species having other combination of traits.

RFLPs (Restriction Fragment Length Polymorphism)

When it was first shown that DNA was the genetic material, and that it functions by encoding proteins, it was assumed that genes were packed next to each other, without spacing, and that every single base pair was critical. However, as shown later, only a small fraction of eukaryotic DNA encodes proteins. Large regions of DNA within and between genes can undergo mutation with no phenotypic effect. Mutations in such regions that occurred in ancestors are transmitted to the progeny just as reliably as are mutations in coding regions. Such detailed knowledge of gene structure became possible only with the development of a series of molecular techniques, beginning with the 1970 discovery of a group of bacterial proteins called restriction enzymes (REs). Evolved by bacteria as protection against DNA-containing viruses, different REs bind to different specific DNA sequences and cleave both strands of the double helix within these sequences. In a fundamental paper, BOTSTEIN *et al.* (1980) took this idea further. They proposed that it ought to be possible to make a human genetic linkage map based on DNA polymorphisms. The initial type of DNA marker examined for this

purpose, which involved the use of REs to cut DNA into fragments, was called restriction fragment length polymorphism (RFLP). RFLPs are homologous pieces of DNA that vary in length after being cleaved with restriction enzymes (REs).

The technique originally used to detect an RFLP is Southern blot analysis (Fig. 1). The main steps of the technique can be summarized as follows:

- Isolation of DNA. This can be a laborious step because the DNA must be of high purity and of relative large quantity;
- Digestion of the DNA with one or more restriction enzymes. A restriction enzyme binds to DNA wherever it finds a specific 4 to 7 base pair sequence and cleaves the molecule at a specific site in the sequence;
- Electrophoresis of the DNA on an agarose or polyacrylamide gel to separate fragments according to size. An electrical charge is applied to the gel, with the positive charge at the bottom and the negative charge at the top. Because DNA has a slightly negative charge, the pieces of DNA migrate toward the positive electrode. The smaller pieces of DNA moves more quickly through the gel, so that a range of pieces of DNA will be distributed throughout the lane;
- Denaturation of DNA (break of hydrogen bonds, production of single stranded DNA);
- Transfer of the single-stranded DNA fragments from the gel to a solid surface, *i.e.* a nitrocellulose or nylon membrane;
- Preparation of a probe. A piece of DNA (clone) is selected to be a probe. Nicks are introduced into one strand of the clone DNA. Nucleotides and DNA polymerase are added in solution to the nicked DNA. The polymerase repairs the clone DNA starting at the nicked sites, using the added nucleotides, one type of which is radioactive. Other labelling systems are also available. The probe is denaturated by heating: it hybridizes, or binds, only to fragments that bear a complementary sequence of bases. The probe can be any piece of DNA. It can be used efficiently in genetic analysis only if it detects single copy sequences. Frequently probes have been chosen from complementary (c)DNA libraries. cDNAs are made from mRNA and in some cases code for single copy genes;
- Soak the membrane in solution with the labelled probe (since few years ago, radioactively labelled probes: now, alternative approaches are available). The solution is maintained at a temperature where the probe will only bind (hydrogen bonds) to sequences on the membrane that have a high degree of complementarity. The specificity of binding can be manipulated using temperature and salt concentration;
- Wash off the no hybridized material;
- Autoradiography to detect bands through exposition of the membrane to X-ray film (in the case of radioactive labelling). Only fragments complementary to the labelled probe will be observable, since they are radioactive (detectable bands on the film).

The observed variation is due to heritable differences in the DNA. For example, point mutations can create or abolish endonuclease sites, while DNA rearrangements, insertions, or deletions can alter their relative positions.

Another possible approach to detect RFLP variation consists in the amplification via Polymerase Chain Reaction (PCR) (MULLIS & FALOONA 1987) of some specific re

gions of the genome and then in cutting them with restriction enzymes (RFLP). In this case, probes do not need to be used: few fragments are generated, easily detectable on an agarose or polyacrylamide gel (higher resolution). This approach (PCR-RFLP) can be easily applied to the organelle genomes (chloroplast, cp-DNA, and mitochondrial, mt-DNA) that, because of their high copy number and their small size in combination with the availability of 'consensus' primers (*e.g.*, GRIVET *et al.* 2001), are very suitable for PCR-RFLP analysis. Chloroplast and mitochondrial genomes are both generally uniparentally inherited in trees (cp-genome paternally inherited in most conifers and maternally inherited in angiosperms, while mt-genome is maternally inherited in both conifer and angiosperm species). The contrasting mode of inheritance provides a unique opportunity to trace both maternal and paternal lineages. Population genetic studies in conifers based on the use of mitochondrial markers revealed a more pronounced differentiation among populations than using chloroplast and nuclear markers, as expected considering that seed dispersal is much more limited than pollen dispersal.

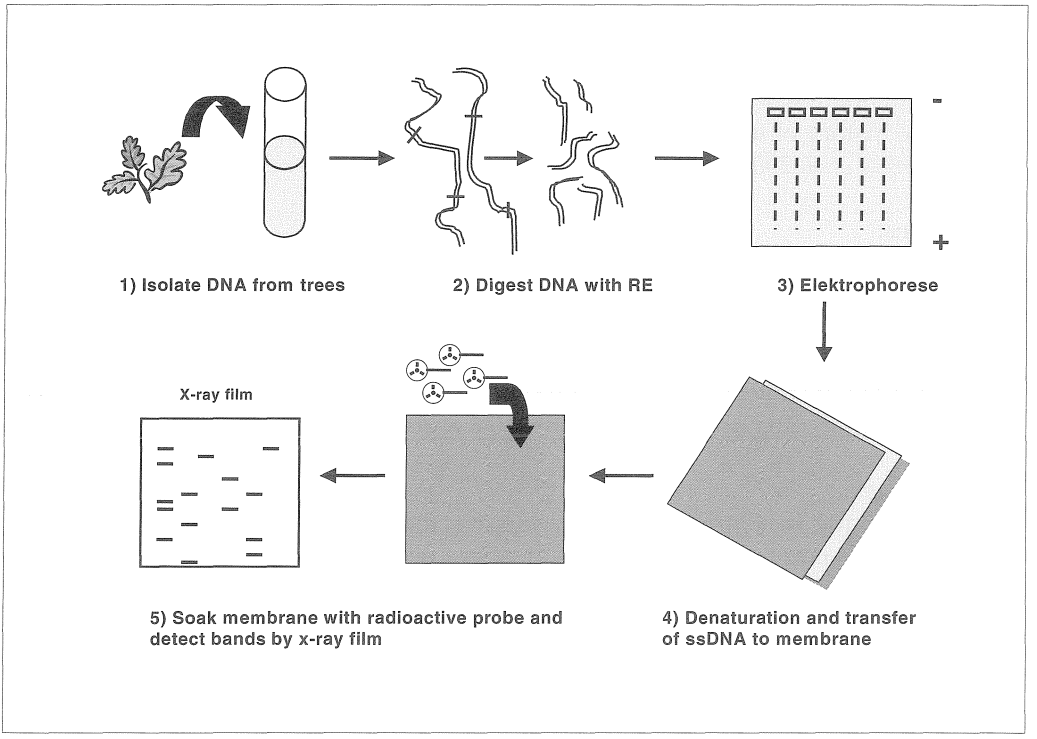

Figure 2. Schematic presentation of the principles used in RFLP markers. (1) Isolation of DNA from individuals to be tested. (2) Digestion of the DNA with one or more restriction enzymes (RE). (3) Electrophoresis of the DNA on an agarose gel to separate fragments. The DNA is loaded with each sample in a separate lane., (4) Denaturation of the DNA to single-stranded DNA by heat and transfer of the single-stranded DNA fragments from the gel to a membrane. (5) Soak the membrane in solution with a radioactively (note that other labelling systems are now available, *e.g.*, DIG oxigenine) labelled probe and detect bands by use of autoradiography.

Based on a large set of data, relevant factors determining the actual distribution of diversity within and among species were identified: i) the location of glacial refugia in the Italian, Iberian and Balkan peninsula and the impact of the resulting migration processes and migration routes; ii) the role of life history traits (in particular the mechanisms of seed and pollen dispersion), and iii) the influence of human activities on the different species. A common migration history of different forest tree species was observed and common phylogeographic patterns among species were identified (PETIT *et al.* 2003).

RFLPs were also used for the identification of individuals and clones (RFLP in repetitive nuclear DNA) and for linkage mapping purpose (RFLP in single or low copy DNA). In general, the number of studies dealing with conservation and population genetics of trees using nuclear single or low copy RFLPs is quite limited, probably because of the advent of the less time consuming and more efficient PCR technique. On the other hand, application of the PCR-RFLP technique in regions of the cp- and mt-genomes, mainly for phylogeographic studies, is still commonly used also because of the difficulty of detecting variation in these genomes via PCR only. One of the few exceptions is represented by the study of SPERISEN *et al.* (2001) in Norway spruce, in which length variation by PCR due to the presence of minisatellite regions in the mitochondrial genome was detected. The geographic distribution of the different lineages reflects the history of this species during the last glaciation and in the post-glacial period (SPERISEN *et al.* 2001).

Advantages:

- highly polymorphic – many alleles may be present in a population for a single locus;
- co-dominant inheritance (RFLP of the nuclear genome);
- many loci can be established;
- highly reproducible between laboratories;
- high discriminating power at the individual and population level.

Disadvantages:

- The technique is laborious, time-consuming and relative expensive, also because generally it requires Southern hybridization, even if not so much in routine any more for detecting and screening polymorphism throughout entire populations or ranges. For this reason it has been almost all completely replaced by PCR based approaches: nevertheless, Southern hybridization has still some importance *e.g.*, for detecting microsatellite loci from genomic libraries, for detecting gene families or for an unambiguous identification or proof of the presence of a transgene;
- usually, but not necessarily, uses isotopes (see comments above).

VNTRs (Variable Number of Tandem Repeats) – Minisatellites

A major form of genetic polymorphism that can occur in non-coding regions of DNA is the variable tandem repetition of DNA sequences. The DNA sequence varies at each

individual locus, and there can be thousands of repeats at a locus. The repeats tend to be GC-rich and ~9–40 base pairs in length. There can be many alleles (alleles are differing by the numbers of repeats) at each of these loci because the number of repeats present can vary widely, due to loss or gain of repeats over evolutionary time. These loci are called 'minisatellites' and are defined hypervariable because they generally display very high polymorphism. These loci are also called Variable Number of Tandem Repeats (VNTRs). Because of their size, they are usually detected by digestion of the genomic DNA by restriction enzymes followed by hybridization with a locus-specific radioactive probe (other labelling systems are now available, see comment above), just as with an RFLP (Fig. 3).

JEFFREYS *et al.* (1985) discovered that a 33 bp repetitive minisatellite had a great deal of sequence homology with minisatellite sequences located elsewhere in the genome. It turns out that many repeated sequences have evolved from a common ancestral core sequence or set of core sequences. Thus, each hypervariable locus has some sequence homology with many other loci. This can be demonstrated by using any such locus as a probe under low-stringency conditions such that the probe act 'generically', detecting any sequence with sufficient sequence similarity to allow hybridization; the closer the match in sequence between probe and core sequence, the more likely the detection of many other hypervariable loci in the genome. Such probes are called multi-locus

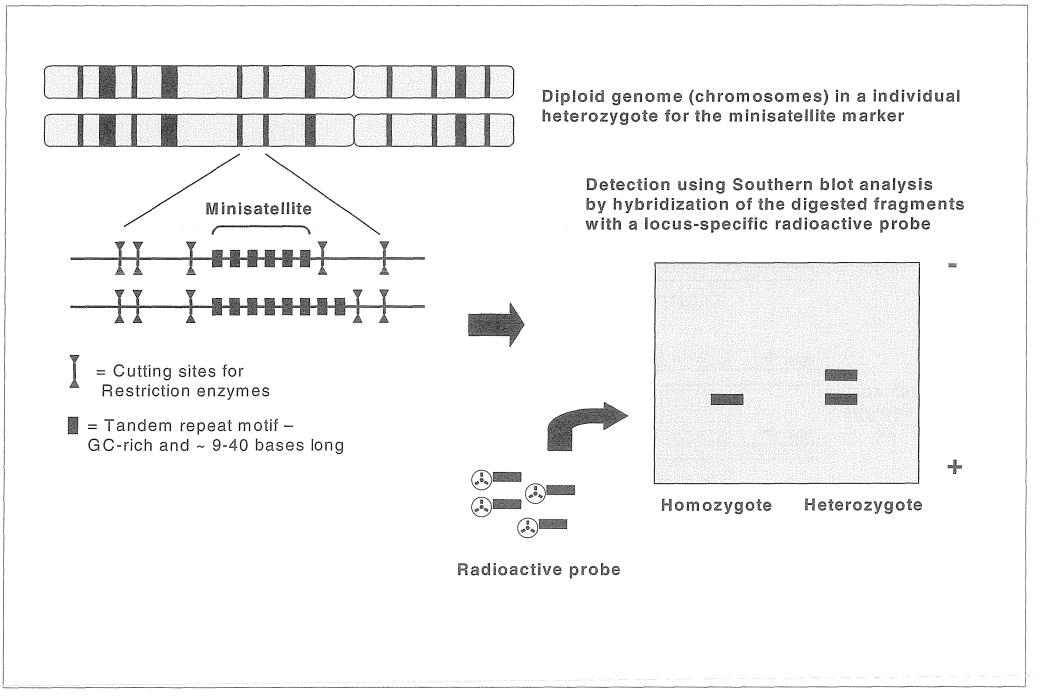

Figure 3. The variation in minisatellites is usually detected by Southern blot analysis with digestion of the genomic DNA by restriction enzymes followed by hybridization with a locus-specific radioactive (other labelling systems are now available, see the text) probe.

probes (MLPs), because they detect many hypervariable regions simultaneously. These probes typically detect between 15 and 20 variable DNA fragments larger than 3.5 kb per individual, plus many smaller bands. DNA fingerprinting can be based on the use of these MLPs.

VNTR markers have been widely used in humans for mapping and fingerprinting. They have been used to a limited degree in plants for fingerprinting. However, their homology to plant sequences is not high and they have not been successfully developed for use in mapping.

These markers can also be detected by Southern blot analysis: the main advantage is that they display co-dominant inheritance. Of course, in this case all the disadvantages of RFLP are present. On the other hand, minisatellites tend to be much more polymorphic than RFLPs but for this specific reason, these markers may be more difficult to analyze (higher resolution required). This marker type has been widely used in forensic analysis.

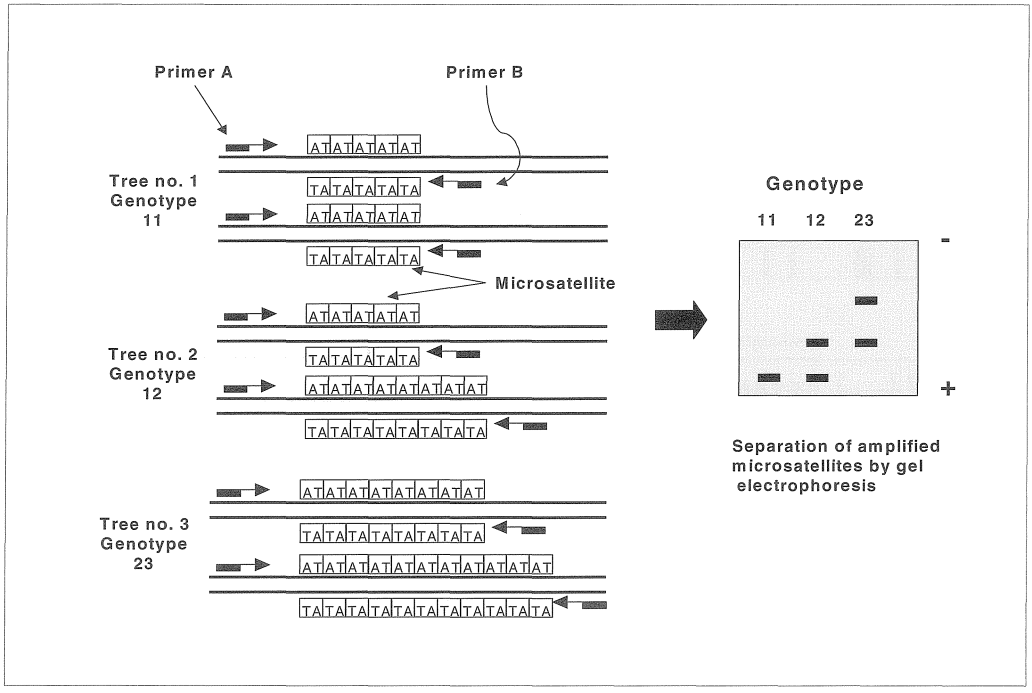

Figure 4. Microsatellites (SSRs) consist of segments of DNA containing tandem repeats of a short 'motif' sequence – here the motif is AT. The microsatellite is amplified through the polymerase chain reaction (PCR), using primers designed to match unique sequences flanking the tandem repeat. The PCR makes a large number of copies of the microsatellite sequence, which afterwards can be separated on a gel.

Microsatellites

Microsatellites are a type of DNA marker, which are considered ideal for individual genotyping and studies of gene flow in forest trees. Microsatellites, or simple sequence repeats (SSRs), consist of segments of DNA containing numerous tandem repeats of a short 'motif' sequence, usually of one to six bases (*e.g.*, ATATATAT... in Fig. 4). They are detected by polymerase chain reaction (PCR) amplification, using primers designed to match unique sequences flanking the tandem repeat. The PCR process makes a large number of copies of the target DNA segment containing the microsatellite sequence (Fig. 4).

The utility of microsatellites is due to their high variability, together with the ability to semi-automate their analysis and scoring. Microsatellites are co-dominant markers (therefore heterozygotes can be distinguished from homozygotes) and consequently they are more informative for genotyping individuals and for linkage mapping than dominant markers such as RAPDs and AFLPs (see below). However, the number of microsatellite sequences in the genome is relatively limited (at least, it is not easy to identify and characterize microsatellite markers), restricting their use for mapping when compared with the potentially unlimited number of RFLP, RAPD and AFLP loci. It should be stressed that the use of different kind of markers is extremely useful to obtain highly saturated maps. Other limitations to their use are mainly the effort and expense required for their development and secondly, evidence that their flanking sequences may not be highly conserved across species in some genera so that markers are not transferable across more distantly related species (ECHT *et al.* 1996, KARHU *et al.* 1999). In population studies, the high variability detected within populations using microsatellites may in fact reduce the power to detect differences among populations. Due to the extremely high level of variation, caution must therefore be taken on the application of statistical procedures developed for less variable markers (*i.e,.* isoenzymes) to microsatellite data (HEDRICK 1999).

Microsatellites were first developed for use in the human genome (WEBER & MAY 1989) and were then widely investigated in plants later on (MORGANTE & OLIVIERI 1993). Microsatellites are generally identified by using two different approaches. The first approach is used in species where a great deal of sequencing data is available. Database searches of sequences can be performed and primers in the flanking regions of the repeat stretch can be designed. In the second approach, library construction, hybridization, and sequencing work is required. Different methods have been used for an efficient isolation of microsatellites from the genome of forest tree species. These species have generally (in particular, conifer species) complex and large genomes, in which the fraction represented by repeated elements is high and SSR stretches are likely to fall into this portion of the chromosomes. Therefore there was the need for methods that allow an efficient recovery of microsatellite markers from DNA libraries. After the first attempts with non-enriched libraries, that resulted in a low percent of positive clones and of single-locus markers (*e.g.*, 19 % of single-locus SSRs in Norway spruce; PFEIFFER *et al.* 1997) the requirement of the enrichment of libraries for positive, microsatellite-containing clones, was addressed (EDWARDS *et al.* 1996, TENZER *et al.* 1999, PASTORELLI *et al.* 2003, SEBASTIANI *et al.* 2004).

Dinucleotide repeats (*e.g.*, CA or TA) are the most important in terms of information content and genome prevalence. A significant relationship between length of the repeat size and level of variation was observed: the probability that a DNA polymerase slips and adds an additional dinucleotide when replicating a longer sequence is higher. As a result, the probability that one individual will be heterozygous for that marker is high. On the other hand, this characteristic also makes it more likely that the polymerase in the PCR reactions will slip as well, resulting in 'stutter bands', making therefore more difficult the interpretation of SSRs amplification pattern. Trinucleotide and tetranucleotide repeats in plants are less frequent than the dinucleotides, but now many are available: these types of SSRs tend to have fewer problems associated with stutter. In general, longer stretches tend to display higher level of variation: however, the observation that most microsatellite arrays are shorter than a few tens of repeats units strongly suggests that there must be size constraints restricting the expansion of repeat arrays (NAUTA & WEISING 1996). SCOTTI *et al.* (2002) reported a significant correlation between the level of polymorphism and the length of the microsatellites in Norway spruce: moreover, they observed that the level of variation for ATC trinucleotide markers in this species was lower than those for dinucleotides. Similarly, a positive and significant relationship between length of chloroplast microsatellite stretches and level of polymorphism was observed by DEGUILLOUX *et al.* (2004).

Originally microsatellite markers were used for genetic mapping and as a diagnostic tool to detect human disease (MURRAY *et al.* 1992). Also in forest trees, they have been used as co-dominant anchor points in genome mapping (*e.g.*, oak: BARRENECHE *et al.* 1998; Norway spruce: PAGLIA *et al.* 1998; beech: SCALFI *et al.* 2004). At the individual level, SSRs are furthermore reported to have potential to identify single individuals or clones, respectively (LEFORT *et al.* 2000, GOMEZ *et al.* 2001). Most frequently, microsatellites are used in population and ecological studies. Microsatellites are excellent markers for studying gene flow, effective population size, migration and dispersal processes, and parentage and relatedness (*e.g.*, DOW & ASHLEY 1996, STREIFF *et al.* 1999, GODOY & JORDANO 2001). The usually great numbers of alleles per locus allows parentage analysis with extremely high exclusion percentages. For example, using only six SSR loci STREIFF *et al.* (1999) reached an exclusion percentage of about 99.9 in oak: similarly in beech using only four SSRs (VENDRAMIN *et al.*, in preparation).

Microsatellites can provide insights into the genetic structure of natural populations in species with little or no allozyme variation. Using microsatellites from the chloroplast genome ECHT *et al.* (1998) were able to detect variation among populations of *Pinus resinosa* Ait., a forest tree species with little morphological variation, no allozyme and very limited RAPD variation.

Nuclear microsatellite analysis, in combination with the development and application of a new statistical test, allowed the identification of the origin of some *Pinus pinaster* Ait. populations (DERORY *et al.* 2002). The test can theoretically be applied to all other forest species, provided that the microsatellite information is available and that the distribution curves of haplotype frequencies do not overlap.

A combination of microsatellites with other markers aimed at a differential display of the distribution of genetic diversity among and within populations assuming different evolutionary drivers beyond the variability of the different markers. Comparing

the distribution of diversity using allozymes and various DNA markers including microsatellites with that of an adaptive morphological trait in Scots pine populations, KARHU *et al.* (1996), however, did not find any correlation. The conclusion is that microsatellites significantly contributed to increase knowledge about stochastic aspects but so far not about population processes governed by selection or adaptation. At the species level, MUIR *et al.* (2000) demonstrated a surprising result with microsatellites indicating the species status of commonly hybridizing *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl.

Although microsatellites are assumed to be evolutionary neutral markers, some evidences of possible functional significance of a part of SSRs were obtained (LI *et al.* 2002). SSRs may influence: (1) chromatin organisation, in particular *a)* chromosomal organisation *b)* DNA structure (SSRs may form a variety of unusual DNA structures and therefore affect gene expression and replication) *c)* centromere organisation (repeated sequences are frequent in the centromeric region of the chromosomes); (2) regulation of DNA metabolic processes, in particular *a)* recombination (SSRs represents also hot spot for recombination) *b)* DNA replication and enzyme controlling cell cycle; (3) regulation of gene activity, in particular *a)* transcription (SSRs located in promoters may influence gene activity *b)* gene expression (for example, some genes can only be expressed at a specific repeat number of SSRs) *c)* genetic disorders (for example, 14 neurological disorders result from the expansion of unstable tri-SSRs) *d)* translation and *e)* protein banding (SSRs may serve as binding sites for some regulatory proteins. It is therefore clear that neutral nature of microsatellite markers cannot be assumed *a priori*. For more details, see LI *et al.* (2002)

Comparative studies of microsatellites and other nuclear markers were recently performed which indicate that microsatellites are more sensitive indicators of fine scale genetic structure. In a study performed with the main aim to verify the possible effects of different harvesting practices on genetic diversity in *Eucalyptus sieberi* L. Johnson forest in southeastern Australia, GLAUBITZ *et al.* (1999) reported genetic distances between nearby stands estimated from microsatellites were 9 times higher than from RFLP data. The microsatellites displayed twice as much variability as the RFLPs. The higher variability of microsatellites, due to their higher mutation rate, has led to the development of more powerful analytical procedures for assigning genotypes to source populations and identifying recent immigrants into populations (important topic related to possible contamination of local gene pools with exotic species). It should be stressed that correct applications of microsatellites in evolutionary studies require the identification of realistic model of the molecular evolution of these loci (see reviews by JARNE & LAGODA 1996, GOLDSTEIN & POLLACK 1997, NIELSEN & PALSBOLL 1999, BALLOUX & LOGON-MOULIN 2002, LI *et al.* 2002).

Microsatellites have also been identified in the chloroplast genome. The discovery of a polymorphic SSR stretch in the chloroplast DNA of *Pinus* species (POWELL *et al.* 1995) enhanced the development of a set of universal primers amplifying 20 cpSSR loci in *Pinus* (VENDRAMIN *et al.* 1996). The sequences of the 20 PCR-amplified loci revealed a characteristic structural trait: they consist of variable numbers of repeat units which are exclusively mononucleotides (VENDRAMIN *et al.* 1996). Chloroplast microsatellites with the same characteristics have been identified also in angiosperms

(WEISING & GARDNER 1999). Recently SEBASTIANI *et al.* (2004) identified also di-nucleotide SSRs in the chloroplast genome of some species belonging to the Fagaceae family. Chloroplast microsatellites are in general maternally and paternally inherited in angiosperms and conifers, respectively. When haploid uniparentally inherited markers are considered in the monoecious conifers, the effective population is supposed to be half the size of that when nuclear markers are used (BIRKY *et al.* 1989). Thus, variation at pollen-mediated chloroplast microsatellite loci is believed to be indicative for genetic drift or for isolation in combination with bottleneck effects. Due to the high degree of conservation of the chloroplast genome, primers for the amplification of cpSSRs can be easily transferred among species. The characteristics of organelle microsatellites described above open up a wide range of applications for the analysis of genetic diversity of conifers at different spatial and temporal scales. At the individual level, the potential usefulness of chloroplast microsatellites for individual identification or paternity analysis has been demonstrated (*e.g.*, in *Abies alba*, ZIEGENHAGEN *et al.* 1998). In three *Abies* species, PARDUCCI *et al.* (2001) attributed 74 % of the total gene diversity to the within-population variation. Numerous studies have been devoted to the analysis of the distribution of cp microsatellite diversity among and within populations. The interpretations mainly address effects of stochastic gene flow via pollen or seeds and thus rely on the assumption of neutrality of the cpSSR markers. Since pollen is the main agent of gene flow in wind pollinated species (LATTA & MITTON 1997, LIEPELT *et al.* 2002), the power of chloroplast DNA markers for differentiating among populations of conifers is generally weak. This holds also true while comparing to the results obtained from maternally inherited markers particularly when gene flow via gravity-dispersed seeds is traced (*e.g.*, SPERISEN *et al.* 2001 in Norway spruce; LIEPELT *et al.* 2002 in silver fir). Though harbouring some problems of putative homoplasy as well as difficulties in distinguishing between historical and contemporary gene flow, cpSSR markers could be used in range wide studies to shed light on certain aspects of population history in conifer species (ECHT *et al.* 1998, MORGANTE *et al.* 1998, VENDRAMIN *et al.* 1999, PARDUCCI *et al.* 2001, RIBEIRO *et al.* 2001, RICHARDSON *et al.* 2002). Several studies revealed the potential usefulness of cpSSRs for differentiating among closely related *Pinus*, *Abies*, *Fraxinus*, *Fagus* species (BUCCI *et al.* 1998, CLARK *et al.* 2000, PARDUCCI *et al.* 2001).

Advantages:

- microsatellites are easy to detect via polymerase chain reaction (PCR);
- they generally display high polymorphism;
- they are co-dominant (nuclear SSRs);
- they are highly reproducible;
- they show a very high degree of 'universality' (cp-SSRs).

Disadvantages:

- Initial identification requires construction and laborious screening of libraries or some other method of obtaining sequence information in order to design the primers. This information is generally not transferable between species (with the exception of chloroplast SSRs);

- A putative reduction or complete loss of amplification of some alleles due to base substitutions or indels within the priming site ('null alleles') may occur. A heterozygote carrying one null allele cannot be distinguished, on a gel, from a homozygote for the only DNA fragment. The frequency of 'null alleles' is relatively high (*e.g.*, in beech, PASTORELLI *et al.* 2003);
- Allelic 'drop-out' is a phenomenon that may occur when the template DNA is of extreme low quantity. In this case, PCR was shown to just amplify the shorter of two alleles (TABERLET & LUIKART 1999); also this phenomenon can lead to an overestimation of homozygosity;
- Hypervariability of markers may not always represent an advantage. Microsatellite variation is based on length variation in base pairs of the amplified fragments. Fragments of the same length (identical in state) may not derived from the same ancestral sequence (and therefore may not be identical by descent), thus introducing the possibility of size homoplasy. Size homoplasy results in an underestimation of population subdivision and genetic divergence between population and species (VIARD *et al.* 1998, TAYLOR *et al.* 1999).

RAPD Markers

In 1990, a new kind of genetic marker was invented – called RAPD (random amplified polymorphic DNA) (WILLIAMS *et al.* 1990).

RAPDs are detected by performing a PCR assay with a single short oligonucleotide primer, usually 10 bases long, of arbitrary sequence. Low annealing temperatures (generally 36–37 °C) are used in RAPD reactions in order to allow binding of these short sequences to matching sequences. Such short nucleotides are likely to be complementary to many sites within a eukaryotic genome. If genomic sequences are present such that the primer can bind in inverted orientation to 2 different sites on opposite strands of the DNA template, and if the sites are close enough to each other (generally less than 3,000 base pairs), the DNA is amplified during the cycles of the PCR, leading to production of a DNA fragment that can be visualized on an agarose gel (generally 3 to 20 detectable fragments in large genomes, *e.g.*, conifers).

Polymorphism between genotypes is due to either a nucleotide base change that alters the ability of the primer to anneal to the DNA template (loss of amplification), or an insertion or deletion within the amplified fragment. Amplification can also be gained from substitution or insertions that create new primer binding sites. Therefore, the polymorphism is generally detected as presence or absence of a particular RAPD product. These polymorphisms can be mapped in the same fashion as other molecular markers; the level of polymorphism detected by RAPDs appears to be similar to the amount detected by RFLPs, and RAPD loci are distributed as randomly throughout the genome.

Because polymorphisms are visualized as the presence or absence of a band, RAPDs are dominant markers, opposite to RFLPs and SSRs (co-dominance of RAPD loci is rarely reported). The technical simplicity of RAPD markers and their application to any species favoured their use in forest tree genetic studies, in particular in

genetic linkage mapping (BINELLI & BUCCI 1994, GRATTAPAGLIA & SEDEROFF 1994, BUCCI *et al.* 1997a, PAGLIA *et al.* 1998). Due to their dominant nature and the low reproducibility of the assay, the use of RAPD markers in population and conservation genetics has decreased dramatically during the last few years; nevertheless some interesting studies using this technique for screening distribution of genetic diversity within and among populations were published (MOSSELER *et al.* 1992, ISABEL *et al.* 1995, VICARIO *et al.* 1995, BUCCI *et al.* 1997b, SCHIERENBECK *et al.* 1997). In conifers the problem of the dominance can be overcome by analyzing the haploid tissue of the megagametophytes that are of maternal origin. This solution, however, reduces the efficiency of the assay considering that at least 6 megagametophytes per maternal tree must be analyzed for a correct classification of an heterozygous (probability of less 3 % to misclassify an heterozygous as homozygous). The identification of efficient methods, in terms of costs and time, for accurate analyses of patterns of diversity in forest trees is extremely relevant, because these organisms typically show a high level of variability, so that sampling of a large number of populations and individuals is required for each study.

Another problem with RAPD markers is the low number of alleles at a locus. Within a species, there are often multiple RFLP and SSR alleles at a locus. With RAPDs, frequently there are only 2 alleles, indicated by the presence or absence of a band. This characteristic may also represent a limitation in studies dealing with mapping: loci that are informative within one cross may not be informative in a different cross; both parents may possess the null allele.

RAPDs became important genetic markers because the assay is very simple and fast; and many loci can be identified, often with a single reaction; sample preparation is much less laborious because only very small amounts of not very pure DNA are required. On the other hand, RAPDs have been strongly criticized as markers because doing PCR with short arbitrary primers can lead to quite different patterns depending on the PCR conditions (lack of reproducibility). The amplification reactions have to be carefully standardized to obtain consistent results. Therefore, results can vary from lab to lab, where different thermocyclers and amplification conditions are used. In general, the problem of reproducibility of the assay was never completely solved, and this is the main reason of the almost all complete abandonment of this technique.

Informative RAPD fragments can be transformed into a new type of markers, having fewer drawbacks than the original RAPD marker: they are called Sequence Characterized Amplified Regions (SCARs). RAPD fragments of interest are isolated, cloned and sequenced and longer primers (generally 24 bp) complementary to the ends of the original RAPDs are designed. Using SCARs only single loci are generally amplified. In the majority of the cases SCARs are still dominant markers, but much more reproducible than RAPDs. SCAR markers allowed refining the postglacial migration history of Norway spruce in the Italian Alpine region (SCOTTI *et al.* 2000).

Advantages:

- rapid, simple, relatively inexpensive assay;
- no prior sequence information is required;
- many loci can be identified quickly;

- the assay can be automated.

Disadvantages:
- Polymorphism is typically dominant;
- Low allelic polymorphism;
- Inconsistency of results;
- Problems of co-migration: the presence of a band of identical molecular weight in different individuals is not evidence that the individuals share the same homologous DNA fragment;
- Being based on random amplification, marker bands can originate from any position within cellular DNA, including nuclear, chloroplast and mitochondrial DNA. To ensure correct data analysis, inheritance of all RAPD fragments must be confirmed before performing population genetic studies. Strong genetic differentiation among populations may derive from the high proportion of RAPD fragments of mitochondrial origin (the mitochondrial genome is generally maternally inherited in both conifers and angiosperm trees), as observed for example in *Pseudotsuga menziesii* by AAGAARD *et al.* (1995).

AFLPs (Amplified Restriction Fragment Polymorphism)

In this method (VOS *et al.* 1995), genomic DNA is digested with 2 restriction enzymes, a rare-cutter (*e.g., EcoRI*) and a frequent-cutter (*e.g., MseI*), creating 2 different overhanging ends. Subsequently, oligonucleotide adapters are ligated to the ends of these restriction fragments. Using primers corresponding to *EcoRI* and *MseI*, only those fragments with an *EcoRI* and *MseI* site at either end are amplified by PCR. This initial PCR amplifies a subset of the fragments originated by the restriction enzymes. The fragments are then subjected to a selective amplification using primers corresponding to synthetic adaptors as well as the first few (one to five) nucleotides of the restriction fragment itself. The amplification is selective because one selective nucleotide on each primer will result in the amplification of only one in sixteen fragments while two selective bases will decrease the number by 1/256. The selective nucleotides can be varied. The number of resulting amplified fragments is determined by the number and composition of the selective nucleotides used as well as the complexity of the genomic DNA. A typical plant genome like soybean amplified with three selective bases will produce about 120 fragments. These fragments are detected on sequencing-type polyacrylamide gels through radioactive labelling or through silver staining and DIG-labelled procedure (ZIEGENHAGEN *et al.* 1999). AFLP analysis performed on species having higher complex genome, such as that of conifers, requires an higher degree of selectivity (generally three selective bases) in order to reduce the number of fragments produced (generally not more than 150 fragments are easily detectable by electrophoretic separation and analysis).

To summarize, the basic AFLP assay consists of four successive steps (Fig. 5):
(1) digestion with restriction enzymes;
(2) ligation with adapters;

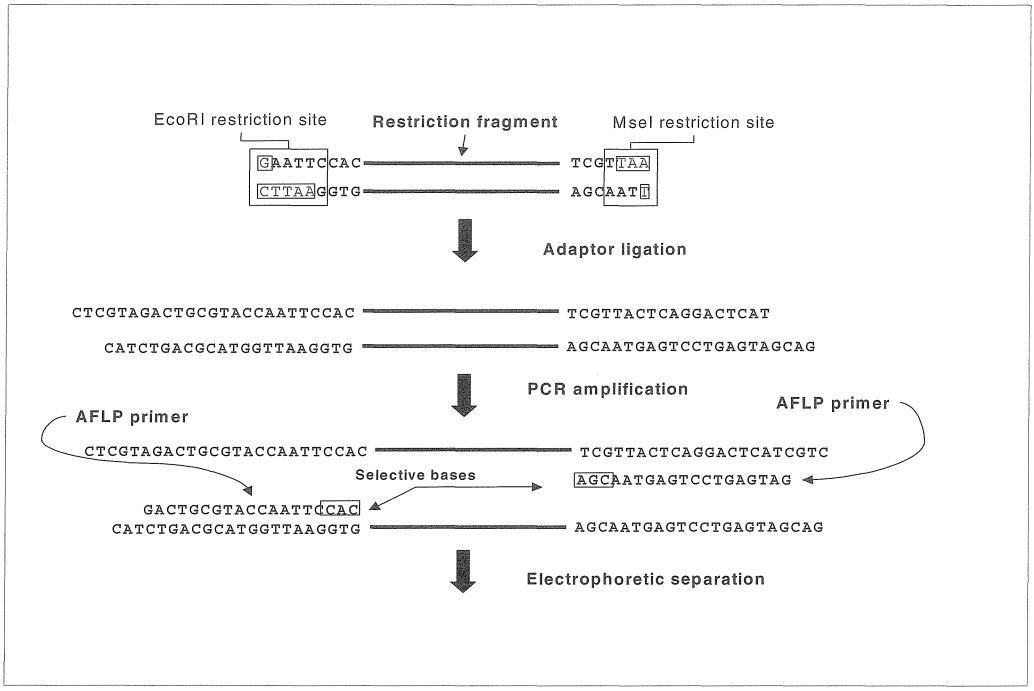

Figure 5. The basic AFLP assay consists of four successive steps: (1) digestion with restriction enzymes; (2) ligation with adaptors; (3) PCR amplification – one initial and one with selective bases. Only the selective amplification is show in the figure; (4) Electrophoresis and detection. AFLP involves the detection of 'presence' or 'absence' of restriction fragments rather than differences in their lengths, and is a dominant marker [from Vos *et al.* (1995), modified].

(3) PCR amplification (always one initial – sometimes also a subsequent selective one); (4) electrophoresis and detection.

Advantages:

- Many more bands, and therefore potentially many more polymorphisms, are identified than with RFLPs or even RAPDs. Routinely about 50–100 bands are observed in each lane of a gel and this enables a rapid construction of highly saturated genetic maps;
- As with RAPDs, only a small amount of template DNA is needed and no probe hybridization is needed;
- No prior sequence information is needed;
- Amplification patterns are more reproducible than with RAPDs.

Disadvantages:

- The method is technically challenging and labour intensive;
- The technology is proprietary;
- Bands are still scored as present or absent (*i.e.*, dominant or recessive). Some

densitometric softwares were developed that may discriminate between heterozygotes and homozygotes based on allelic density, but this approach requires specialized equipment. It should be mentioned that fragment analysis software for automated sequencers were recently developed and are now available;
- Specialized equipment (*e.g.*, automatic sequencer) is necessary to increase the efficiency of the technique.

Although AFLP technique was recently developed, there are already some examples of the application of these markers in forest genetics (CERVERA *et al.* 1996, MARQUES *et al.* 1998, CHAGNÉ *et al.* 2002). RIBEIRO *et al.* (2002) analyzed genetic variation in *Pinus pinaster* populations using AFLPs and chloroplast SSRs. Population level of diversity within French and Portuguese populations were found to be similar with AFLPs, but not with cpSSRs. Although AFLPs revealed much lower genetic diversity than cpSSRs, the level of among-population differentiation with the two types of markers was similar. The concordance found between the two markers suggests that gene flow was the predominant force shaping genetic variation of the populations of France and Portugal (RIBERIO *et al.* 2002). The possibility to screen a large number of loci in a single assay allowed developing highly saturated linkage genetic maps in *Pinus pinaster* (CHAGNÉ *et al.* 2001), in *Picea abies* (PAGLIA *et al.* 1998) and in *Fagus sylvatica* (SCALFI *et al.* 2004). The most important applications of the AFLP technique in breeding include genetic distance analysis, variety identification, isolation of markers tightly linked to specific genes, and marker assisted backcrossing.

SSCPs (Single Stranded Conformation Polymorphism)
SNPs (Single Nucleotide Polymorphism)

The mobility in gel electrophoresis of double-stranded DNA's of a given length is relatively independent of nucleotide sequence. In contrast, the mobility of single strands can vary considerably as a result of only small changes in nucleotide sequence. This fact led to the development of Single-Stranded Conformation Polymorphism (SSCP) techniques.

SSCP markers are also analyzed with PCR techniques. Most SSCP protocols are designed to analyze polymorphism at a single locus at a time, frequently a locus of known biological function. Therefore, in this technique, a specific pair of PCR primers flanking the target region is used to amplify DNA from different individuals. Single-stranded DNA is produced by asymmetric PCR, in which one primer is present in excess over the other. After the low concentration primer is used up, the reaction continues, producing only the target strand. Then the mobility of single strands is compared by gel electrophoresis (Fig. 6).

Single-stranded DNA mobility is sequence dependent because of varying degrees of intra-strand base pairing and the resulting looping and compaction. Single strands may also be generated by denaturing double-stranded DNA. In this case the mobility of each of the two complementary strands is observed. Strands can be distinguished by attaching different fluorescent labels to the two primers.

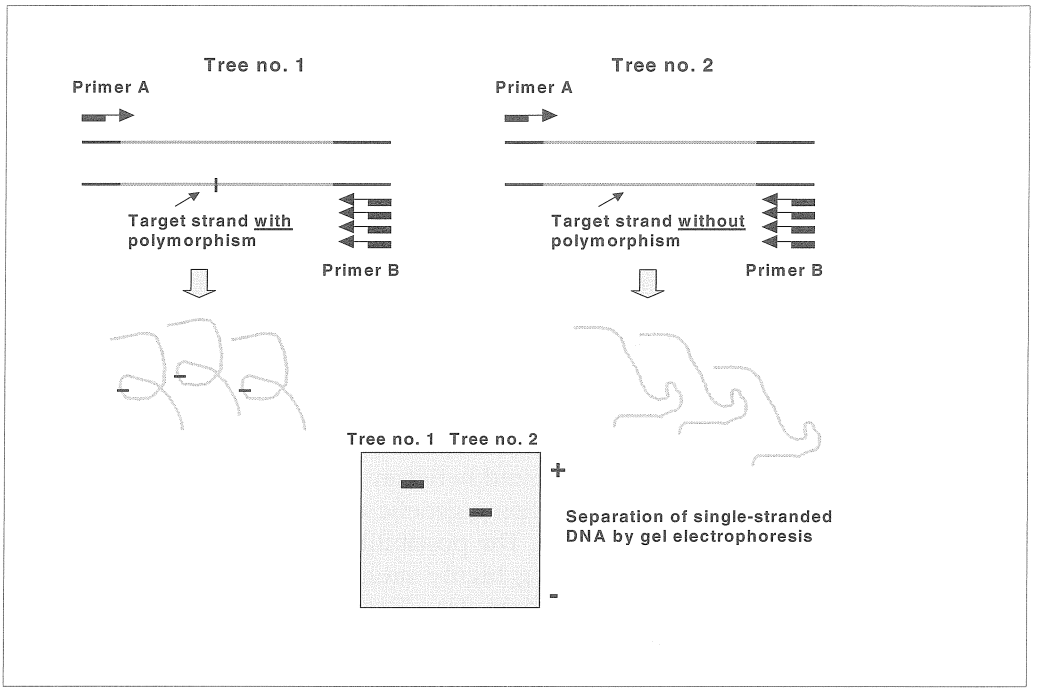

Figure 6. Outline of the principle in the single-stranded conformation polymorphism (SSCP) techniques. The mobility of single strands can vary considerably as a result of only small changes in nucleotide sequence. The single-stranded DNA is produced by asymmetric PCR, in which one primer (B) is present in excess over the other (A). After the low concentration primer (A) is used up, the reaction continues, producing only the target strand. Afterwards the mobility of single strands is compared by gel electrophoresis.

The type of polymorphism observed with these techniques is often referred to as SNP (single nucleotide polymorphism) if a single nucleotide is responsible for the mobility difference. SNPs can therefore be defined as single base changes in a genomic DNA sequence (SNPs generally refer only to substitutions). SNPs represent a natural genetic variability at high density in the human and at least other mammalian genomes. SNPs can of course be detected by direct sequencing, which represents the most comprehensive approach to analyze genetic differences, homologous fragments of different genotypes: sequences are then aligned so that position with a variable nucleotide can be easily identified. Direct sequencing is now routine, but not easily applicable to population and conservation genetics of forest trees because of the still too high costs. This is particularly true for diversity studies of forest tree species that, because of their high level of polymorphism, that often require testing large sample of individuals and populations to obtain accurate estimates. A SNP represents an alternate nucleotide in a given and defined genetic location that occurs at a frequency exceeding 1 % in a given population. This definition does not include other types of genetic variability like insertions and deletions, and variability in copy number of re-

peated sequences. SNPs are considered the major genetic source of phenotypic variability that differentiates individuals within a given species.

First studies performed on plants, including forest trees, revealed a high frequency of SNPs also in these organisms: preliminary evidences indicate that the frequency of SNPs in *Picea abies* is more than order of magnitude greater than that reported in humans and even higher that observed in maize, a species considered with high level of variability (DEGLI IVANISSEVIC & MORGANTE 2001). These authors, analyzing 13 ESTs in Norway spruce, detected an overall frequency of 1 SNP/88 bases, and in detail 1 SNP/30 bases in introns. Lower frequencies of SNPs were detected in *Pinus halepensis* Mill. (VENDRAMIN *et al.* 2003) and *Pinus pinaster* (PLOMION *et al.* 2003). The abundance of SNPs compensates for the fact that these markers generally display only two alleles considering that the probability that a third nucleotide is present at the same position is very low.

SNPs may occur in non-coding regions (SNPs) as well as in coding regions (cSNPs). cSNPs often generate polymorphic variants of expressed proteins that sometimes affect their functional properties. The chance to detect SNPs is generally higher in non coding and intergenic regions of the genome, because they are expected to be under less severe selection than coding regions. SNP frequency in *Pinus taeda* L. ranged from 1/196 in exons to 1/37 in non-coding regions (PLOMION *et al.* 2003).

As mentioned before, direct sequencing of DNA segments (amplified by PCR) from several individuals is the most direct way to identify SNPs (Figure 7). PCR primers are designed to generate fragments of 400–700 bp, frequently derived from genes of interest or ESTs (Expressed Sequence Tag; for more details see p. 380 ff., this volume). PCR can be performed on a set of highly divergent populations, each represented by different individuals. The PCR products are sequenced in both directions. The resulting sequences are aligned and, taking care to distinguish true polymorphism from sequencing errors, polymorphism is identified. In the specific case of conifers, the analysis of DNA isolated from the megagametophytes (haploid tissue of maternal origin) eliminates the problem of the identification of heterozygous positions. A marker tagging a gene under selection is expected to display a different pattern of variation compared with that obtained using neutral markers. An approach to test this hypothesis consists in the comparison of the level of divergence among populations using, for example, Wright's F-statistics, separately for neutral (F_N) and gene-tagged markers (F_T.). Values of F_T higher than F_N indicates divergent selection and local adaptation for the analyzed genes while similar values of the two parameters is an indication that selection doesn't play any role. Markers showing values of F_T higher than F_N are the most valuable for conservation purposes (VAN TIENDEREN *et al.* 2002). First results obtained in *Pinus halepensis* analyzing the CAD region showed a level of genetic differentiation among populations about 5–10 times greater than using neutral markers (isoenzymes, cp- and nuclear SSRs) (VENDRAMIN *et al.* 2003).

In species characterized by outcrossing mode of reproduction and very large population size, such as almost all conifers, the rate of SNP polymorphisms is likely to be high. In other species, pre-screening to detect polymorphism may be necessary: SSCP and/or enzyme cleavage methods may be extremely useful.

Single nucleotide polymorphisms are an extremely reach source of polymorphic mark-

ers for use in the high resolution genetic mapping of traits and for association studies that are based on candidate genes as well as for plant diversity studies based on the analysis of sites that may be undergoing selection (for more details see p. 380 ff., this volume). The recent development in genotyping capabilities should allow the rapid increase of nucleotide surveys based on larger sample sizes for a more robust analysis essential for population genetic studies. The availability of a huge amount of data requires the development of appropriate databases and bioinformatic tools.

Recently GILL *et al.* (2003) published an interesting paper in which evidences for the molecular basis of a null allele of CAD have been described in *Pinus taeda*. The mutation consists in a tandem adenosine insertion in the null allele that causes a frameshift that result in a premature termination of the protein. The CAD downregulation results in a significant increase in stem growth: in general the mutation has an effect in pulping efficiency. Chemical pulping could be increased in efficiency by selecting the trees having the 2-bp adenosine insertion. GILL *et al.* (2003) also developed an effcient assay for the detection of the mutation. On the basis of these results, the importance of collecting information about the nucleotide variation in genes and of developing efficient detection systems became evident.

Advantages:
- SNPs occur with high frequency. Many marker loci could be developed to generate a highly saturated map and for association studies;
- SNPs are co-dominant in nature;
- Detection methods for SNPs are potentially more suitable for genetic screening in automated and large-scale (*e.g.*, DNA chips technology).

Disadvantages:
- SNPs are generally biallelic: there are usually only 2 alleles present in a population. This makes them less informative than other types of markers, for example microsatellites;
- SNPs may be population specific, *i.e.* a marker useful in one population may not be present in another. In contrast, microsatellite markers tend to be universal;
- SNP detection is still a laborious and time-consuming step.

Choice of marker

When a marker needs to be chosen to examine variation at Mendelian loci, the following aspects should be taken into consideration (Table 1):
- *Number of available loci*: limited in the case of isoenzymes, theoretically unlimited in the case of all DNA markers, in particular of AFLPs, RAPDs and SNPs;
- *Level of variation*: low for isoenzymes, very high for microsatellites, intermediate for RFLP;
- *Dominance*: RAPD and AFLP generally dominant, therefore less efficient for population genetic studies and mapping purpose than co-dominant markers as isoenzymes, RFLPs, SSRs;

Figure 7. Schematic representation of Single Nucleotide Polymorphisms (SNPs) detection.

- *Presence of null alleles*: high proportion of null alleles detected at SSR loci, which may cause an underestimation of heterozygosity. Null alleles were also detected at isoenzyme and RFLP loci but at lower frequency;
- *Reproducibility*: low reproducibility for RAPD, higher for AFLPs, highest for isoenzymes SSRs and SNPs;
- *Transferability*: very high for isoenzyme, lower for nuclear SSRs, generally only to closely related species; very high for chloroplast SSRs, also to species of different families. Co-migrating RAPD and AFLP fragments may not necessarily derive from the same loci, even in closely related species: sequencing is necessary to verify the homology of RAPD, AFLP and SSR fragments;
- *Easy of assay*: AFLP, RFLP and SNPs are the most technically demanding markers. Highly efficient systems for SNPs detection were recently developed;
- *Efficiency of the assay* (*e.g.*, multiplexing, *i.e.*, the simultaneous analysis of multiple loci in a single assay): lowest with RFLP, highest with AFLP, intermediate with RAPD and SSRs;
- *Automation* (*e.g.*, using automatic sequencers and robotic workstations): all PCR based markers are amenable to automation. On the contrary, isoenzymes and RFLP are predominantly manual techniques. SNP detection can be automated to a large extend;
- *Costs*: isoenzymes are the cheapest, RFLPs and SNPs the most expensive. In the specific case of SSR, the high costs are mainly due to their identification and char-

Table 1. Characteristics of the different markers (for the labels in the first column, see the text; for more details about the characteristic of the molecular markers, see p. 369 ff., this volume).

Marker	Abundance	Levels of polymorphism	Co-dominance	Reproducibility	Easy of assay	Operational costs	Development costs	Automation
Izoenzymes	low	low	yes	high	high	low	low	no
RFLP	high	medium	yes	high	low	high	medium	no
VNTR	medium	high	no/yes	high	low	high	medium	no
SSR	high	high	yes	high	high	low	high	yes
RAPD	high	medium	no	low	high	low	low	yes
AFLP	high	medium	no	high	low	high	low	yes
SSCP	low	low	yes	medium	medium	medium	high	no
SNP	high	low	yes	high	low	high	high	yes

specific case of SSR, the high costs are mainly due to their identification and characterization; when SSRs are available, the cost of their screening is limited.

A simple conclusion can be drawn from the information reported above. All molecular markers are not equal, none is ideal and some are better for some purposes than others. In a very general way, it can be stated that:

- For applications that require the screening of a large number of loci (estimation of genetic variation and differentiation within and among populations), the markers of choice are AFLPs, RFLP, SNPs and SSRs.
- For applications that require a high discriminant power (mating system studies, paternity and parentage analysis, gene flow studies, DNA fingerprinting), the markers of choice are SSRs.

In addition to the study objectives, the selection of the different markers is also strongly dependent on the availability of specific equipment and of skilled person nel as well as on specific financial support.

Some important software packages for the analysis of population and conservation genetic data obtained using molecular markers were designed during the last years. In the following table, the most commonly used software is compiled:

Conclusions

In the specific field of the conservation genetics, molecular markers can be extremely useful in both *ex situ* and *in situ* conservation. *Ex situ* conservation of genetic resources includes the storage of samples in gene banks, which is intended

Table 2. Most commonly used software for a population genetic analysis (parentage analysis included) of the data obtained using molecular markers.

Software	References	Web-address
Arlequin	EXCOFFIER *et al.* (1992)	http://anthro.unige.ch/arlequin
GeneAlEx (Genetic Analysis in Excel)	PEAKALL & SMOUSE (2001)	http://www.anu.edu.au/BoZo/GenAlEx
GenePop	RAYMOND & ROUSSET	http://www.cefe.cnrs-mop.fr/wbiomed/curtin.edu.au/genepop
PopGen	YEH *et al.* (1998)	http://www.ualberta.ca/~fyeh/
MSA (Microsatellite Analyzer)	DIERLINGER & SCHLTTERER (2002)	http://i122server.vu-wien.ac.at
SPAGpageDdi (Spatial Pattern Analysis of Genetic Diversity)	HARDY & VEKEMANS (2002)	http://www.ulb.ac.be/sciences/lagev/
TPFGA (Tools for Population Genetic Analysis)	MILLER (1997)	http://herb/bio/nau/edu/miller/ ~tfpga. htm http://bioweb.usu.edu/mpmbio/ tfpga. htm
SGD Spatial Genetic Distance)	DEGEN (2000)	http://kourou.cirad.fr/genetique/software.html
GDA (Genetic Data Anlaysis)	LEWIS & ZAYTIN (1996)	http:// lewis.eeb.uconn.edu/lewishome/software.html
FSTAT (F-statistics)	GOUDET (1995)	http://www.unil.ch/izea/softwares/fstat.html
STRUCTURE	PRITCHARD *et al.* (2000)	http://pritch.bsd.uchicago.edu/
BAPS (Bayesian Analysis of Population Structure)	CORANDER *et al.* (2003)	http://www.mi.helsinki.fi/~mjs
SAMOVA (Spatial Analysis of Molecular variance)	DUPANLOUP *et al.* (2002)	http://web.unife.it/progetti/genetica/Isabelle/Isabelle.html

to represent the genetic diversity of the species as much as possible. Gene bank management comprises four major categories of activities, all of which may benefit from the application of molecular genetic markers. These activities include 'acquisition' (development of sampling strategies, identification of populations that need to be preserved), 'maintenance' (quantification of genetic drift, identification of genetic contamination), 'characterization' (genetic evaluation of germplasm) and 'utilization' (molecular markers and functional diversity, molecular markers and genetic improvement) (BRETTING & WIDRLECHNER 1995, BROWN & KRESOVICH 1996, KARP *et al.* 1997).

In situ conservation requires an assessment of the genetic variation within and between populations in order to identify the populations that need to be conserved (inventory of genetic diversity): moreover, it requires the development of management activities to monitor genetic changes over time (study of dynamic of diversity). Therefore, *in situ* conservation studies may also benefit from the implementation of molecular genetic markers (KARP *et al.* 1997).

Many studies performed during the last years demonstrated the usefulness of neutral molecular markers in the field of conservation and population genetics of forest trees, in particular to understand the importance of migration patterns in shaping current genetic and geographic diversity and to measure important parameters such as effective population size, past bottleneck and gene flow. This is relevant to design conservation strategies but is of little value to understand adaptability patterns. Phenotype assessment are time consuming and generally very expensive, and are not useful to gather information about variation in the genes controlling adaptive variation. This is why new types of molecular markers are being developed for some forest trees (for example, single nucleotide polymorphisms (SNPs) in candidate genes of *Picea abies, Pinus pinaster* and *Pinus halepensis*). They may have a great potential in the study of adaptive traits. The use of only neutral genetic markers for the conservation of genetic resources may be questionable. These markers have been and are still useful for characterising common origins of populations and their post-glacial migration routes and for mating system studies. They have less potential for characterising genetic diversity in adaptive traits and the adaptive potential of populations. It should be therefore stressed the need to bring the two approaches together so that genetic markers can also be used to study the adaptability of populations and the effects of selection. The new type of SNP markers detected in expressed regions of the genome may be a first step in this direction. In any case, there is the need to use both neutral and adaptive approaches. In the face of environmental changes, populations will migrate and/or adapt (or eventually become extinct!). Knowledge of both mechanisms is necessary for gene conservation. The establishment of Conservation Management Units (MORITZ 1994), intended as units containing significant adaptive variation, should be based on the combined identification of Evolutionary Significant Units (ESU's aimed at the conservation of the historic, evolutionary diversity of a species than the current distribution of populations and alleles) and of Functionally Significant Units (FSUs, based on differences in allelic frequencies for genes with important ecological functions, as suggested by TIENDEREN *et al.* 2002).

During the next years, a large amount of data at marker loci or at sequence level is expected to be collected: to complement the high statistical power of these data, an

evolutionary perspective is required to evaluate their biological importance, as stressed by HEDRICK (2002).

References

AAGAARD, J.E., VOLLMER, S.S., SORENSEN, F.C. & STRAUSS, S.H. 1995: Mitochondrial DNA products among RAPD profiles are frequent and strongly differentiated between races of Douglas-fir. *Mol. Ecol.* **4**: 441–447.

BALLOUX, F. & LUGON-MOULIN, N. 2002: The estimation of population differentiation with microsatellite markers. *Mol. Ecol.* **11**: 155–166.

BARADAT, PH., MAILLART, M., MARPEAU, A., SLAK, M.F., YANI, A. & PASTUSZKA, P. 1995: Utility of terpenes to assess population structure and mating patterns in conifers. Pp. 5–27 *in*: Baradat, P., Adams, W.T. & Müller-Starck, G. (eds.): Population Genetics and Genetic Conservation of Forest Trees. SPB Academic Publishing, Amsterdam, The Netherlands.

BARRENECHE, T., BODENES, C., LEXER, C., TRONTIN, J.F., FLUCH, S., STREIFF, R., PLOMION, C., ROUSSEL, G., STEINKELLNER, H., BURG, K., FAVRE, J.M., GLOSSL, J. & KREMER, A. 1998: A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. *Theor. Appl. Genet.* **97**: 1090 –1103.

BARTELS, H. 1971a: Genetic control of multiple esterases from needles and macro-gametophytes of *Picea abies*. *Planta* **99**: 283–289.

BARTELS, H. 1971b: Isoenzymes and their significance for forest tree breeding and genetics. *Allg. Forstzeitschr.* **3**: 50–52.

BERGMANN, F. 1973: Genetische Untersuchungen bei *Picea abies* mit Hilfe der Isoenzym-Identifizierung. III. Geographische Variation an 2 Esterase und 2 Leucin-aminopeptidase-Loci in der schwedischen Fichtenpopulation. *Silvae Genet.* **22**: 63–6 6.

BERGMANN, F. 1974a: Genetischer Abstand zwischen Populationen. II. Die Bestimmung des genetischen Abstands zwischen europäischen Fichtenpopulationen (*Picea abies*) auf der Basis von Isoenzym-Genhäufigkeiten. *Silvae Genet.* **23**: 28–32.

BERGMANN, F. 1974b: The genetics of some isoenzyme systems in spruce endosperm (*Picea abies*). *Genetica* **6**: 353–360.

BERGMANN, F. 1975a: Adaptive acid phosphatase polymorphism in conifer seeds. *Silvae Genet.* **24**: 175–177.

BERGMANN, F. 1975b: Herkunfts-Identifizierung von Forstsaatgut auf der Basis von Isoenzym-Genhäufigkeiten. *Allg. Forst- u. J. Ztg.* **146**: 191–195.

BERGMANN, F. 1978: The allelic distribution at an acid phosphatase locus in Norway Spruce (*Picea abies*) along similar climatic gradients. *Theor. Appl. Genet.* **52**: 57–64.

BINELLI, G. & BUCCI, G. 1994: A genetic linkage map of *Picea abies* Karst., based on RAPD markers, as a tool in population genetics. *Theor. Appl. Genet.* **88**: 283–288.

BIRKY, C.W.J.R., FUERST, P. & MARUYAMA, T. 1989: Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparisons to nuclear genes. *Genetics* **121**: 6 13–627.

BOSCHERINI, G., VENDRAMIN, G.G. & GIANNINI, R. 1993: Mating system analysis in two Italian populations of Norway spruce. *J. Genet. Breed.* **47**: 45–48.

BOTSTEIN, D., WHITE, R.L., SKOLNICK, M. & DAVIS, R.W. 1980: Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer. J. Hum. Genet.* **32**: 314 –331.

BRETTING, P.K. & WIDRLECHNER, M.P. 1995: Genetic markers and plant genetic resources. *Pl. Breed. Rev.* **13**: 11–86.

BROWN, S.M. & KRESOVICH, S. 1996: Molecular characterization for plant genetic resources con-

servation. Pp. 85–93 *in:* Paterson, H. (ed.): Genome Mapping of Plants. Academic Press, San Diego, U.S.A.

BUCCI, G., ANZIDEI, M., MADAGHIELE, A. & VENDRAMIN, G.G. 1998: Detection of haplotyp ic variation and natural hybridization in halepensis-complex pine species using chloroplast SSR markers. *Mol. Ecol.* **7**: 1633–1643.

BUCCI, G., KUBISIAK, T.L., NANCE, W.L. & MENOZZI, P. 1997a: A population 'consensus', partial linkage map of *Picea abies* Karst. based on RAPD markers. *Theor. Appl. Genet.* **95**: 643–654.

BUCCI, G., VENDRAMIN, G.G., LELLI, L. & VICARIO, F. 1997b: Assessing genetic divergence of *Pinus leucodermis* Ant. endangered populations: use of biochemical and molecular markers for conservation purposes. *Theor. Appl. Genet.* **95**: 1138–1146.

CERVERA, M.T., GUSMAO, J., STEENACKERS, M., PELEMAN, J., STORME, V., VANDEN BROECK, A., VAN MONTAGU, M. & BOERJAN, W. 1996: Identification of AFLP molecular markers for resistance against *Melampsora larici-populina* in *Populus*. *Theor. Appl. Genet.* **93**: 733–737.

CHAGNÉ, D., LALANNE, C., MADUR, D., KUMAR, S., FRIGERIO, J.M., KRIER, C., DECROOCQ, S., SAVOURÉ, A., BOU DAGHER-KHARRAT, M., BERTOCCHI, E., BRACH, J. & PLOMION, C. 2002: A high density linkage map of *Pinus pinaster* based on AFLPs. *Ann. For. Sci.* **59**: 627–636.

CLARK, C.M., WENTWORTH, T.R. & O'MALLEY, D.M. 2000: Genetic discontinuity revealed by chloroplast microsatellites in eastern North American *Abies* (Pinaceae). *Am. J. Bot.* **87**: 774 –782.

COMPS, B., GÖMÖRY, D., LETOUZEY, J., THIEBAUT, B. & PETIT, R.J. 2001: Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics* **157**: 389–397.

CORANDER, J., WALDMANN P. & SILLANPÄÄ M.J. 2003: Bayesian analysis of genetic differentiation between populations. *Genetics* **163**: 367–374.

DEGEN, B. 2000: SGS: Spatial Genetic Software. Computer program and user's manual, available from http://kourou.cirad.fr/genetique/software.html.

DEGLI IVANISSEVICH, S. & MORGANTE, M. 2001: Sequence diversity and SNP marker development in Norway spruce. Proceedings of the XLV Congress of the Italian Society of Agricultural Genetics – SIGA. 26-29 September 2001, Salsomaggiore Terme, Italy.

DEGUILLOUX, M.-F., PEMONGE, M.-H. & PETIT, R.J. 2004: Use of chloroplast microsatellites to differentiate oak populations. *Ann. For. Sci.* **61**: 825–830.

DERORY, J., MARIETTE, S., GONZALÉZ-MARTÍNEZ, S.C., CHAGNÉ, D., MADUR, D., GERBER, S., BRACH, J., PERSYN, F., RIBEIRO, M.M. & PLOMION, C. 2002: What can nuclear microsatellites tell us about maritime pine genetic resources conservation and provenance certification strategies? *Ann. For. Sci.* **59**: 699–708.

DIERINGER, D. & SCHLÖTTERER, C. 2003: Microsatellite analyzer (MSA) - a platform independent analysis tool for large microsatellite data sets. *Mol. Ecol. Notes* **3**:167–169.

DOW, B.D. & ASHLEY, M.V. 1996: Microsatellite analysis of seed dispersal and parentage of saplings in bur oak *Quercus macrocarpa*. *Mol. Ecol.* **5**: 615–627.

ECHT, C.S., DEVERNO, L.L., ANZIDEI, M. & VENDRAMIN, G.G. 1998: Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Mol. Ecol.* **7**: 307–316.

ECHT, C.S., MAY-MARQUARDT, P., HSEIH, M. & ZAHORCHAK, R. 1996: Characterisation of microsatellite markers in eastern white pine. *Genome* **39**: 1102–1108.

EDWARDS, K.J., BARKER, J.H.A., DALY, A., JONES, C. & KARP, A. 1996: Microsatellite libraries enriched for several microsatellite sequences in plants. *BioTechniques* **20**: 758–760.

EXCOFFIER, L., SMOUSE, P.E. & QUATTRO, J.M. 1992: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.

GILL, G.P., BROWN, G.R. & NEALE, D.B. 2003: A sequence mutation in the cynnamyl alcohol dehydrogenase gene associated with altered lignification in loblolly pine. *Pl. Biotech. J.* **1**:

253–258.

GILLET, E.M. (ed.) 1999: Which DNA marker for which purpose? Final Compendium of the Forest Project in the EU Biotechnology FW IV Programme 'Molecular Tools for Biodiversity'. http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm.

GLAUBITZ, J. & MORAN, G.F. 2000: Genetic tools: the use of biochemical and molecular markers. Pp. 39–59 *in:* Young, A., Boshier, D. & Boyle, T. (eds.): Forest Conservation Genetics – Principles and Practice. CABI, Oxon, United Kingdom.

GLAUBITZ, J.C., STRK, J. & MORAN, G.F. 1999: Genetic impacts of different silvicultural practices in native eucalypt forests. Pp. 183–195 *in:* Matyas, C. (ed.): Forest Genetics and Sustainability. Kluwer Academic Publishers, Dordrecht, The Netherlands.

GODOY, J.A. & JORDANO, P. 2001: Seed dispersal by animals: exact tracking of the source trees with endocarp DNA microsatellites. *Mol. Ecol.* **10**: 2275–2283.

GOLDSTEIN, D.B. & POLLOCK, D.D. 1997: Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *J. Hered.* **88**: 335–342.

GOMEZ, A., PINTOS, B., AGUIRIANO, E., MANZANERA , J.A. & BUENO, M.A. 2001: SSR markers for *Quercus suber* tree identification and embryo analysis. *J. Hered.* **92**: 292–295.

GÖMÖRY, D., HYNEK, V. & PAULE, L. 1998: Delineation of seed zones for European beech (*Fagus sylvatica* L.) in the Czech Republic based on isozyme gene markers. *Ann. Sci. For.* **55**: 425 –436.

GOUDET, J. 1995: FSTAT (vers.1.2), a computer program to calculate F-statistics. *J. Hered.* **86**: 485 –486.

GRATTAPAGLIA, D. & SEDEROFF, R. 1994: Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Genetics* **137**: 1121 –1137

GRIVET, D., HEINZE, B., VENDRAMIN, G.G. & PETIT, R.J. 2001: Genome walking with consensus primers: application to the large single copy region of chloroplast DNA. *Mol. Ecol. Notes* **1**: 345–349.

HAMRICK, J.L. & GODT, M.J.W. 1990: Allozyme diversity in plant species. Pp 43–63 *in:* Brown, A.H.D., Kahler, A.L. (eds.): Plant Population Genetics, Breeding and Genetic Resources. Sinauer, Sunderland, U.S.A.

HANNOVER, J.W. 1992: Applications of terpene analysis in forest genetics. *New For.* **6**: 159–172.

HARDY, O.J. & VEKEMANS, X. 2002: SPAGeDi: a versatile compute program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* **2:** 618.

HEDRICK, P.W. 1999: Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**: 313–318.

HEDRICK, P.W. 2002: Conservation genetics: where are we now? *Trends Ecol. Evol.* **16**: 629–636.

ISABEL, N., BEAULIEU, J.L. & BOUSQUET, J. 1995: Complete congruence between gene diversity estimates derived from genotypic date at enzyme and random amplified polymorphic DNA loci in black spruce. *Proc. Natl. Acad. Sci. U.S.A.* **92**: 6369–6373.

JARNE, P. & LAGODA, J.L. 1996: Microsatellies, from molecules to populations and back. *Trends Ecol. Evol.* **11**: 424–429.

JEFFREYS, A.J., WILSON, V. & THEIN, S.L. 1985: Hypervariable minisatellite regions in human DNA. *Nature* **314**: 67–73.

KARHU, A., DIETERICH, J-H. & SAVOLAINEN, O. 1999: Rapid expansion of microsatellite loci in pines. *Mol. Biol. Evol.* **17**: 259–265.

KARHU, A., HURME, P., KARJALAINEN, M., KARVONEN, P., KÄRKKÄINEN, K., NEALE, D.B. & SAVOLAINEN, O. 1996: Do molecular markers predict patterns of adaptive variation? *Theor. Appl. Genet.* **93**: 215–221.

KARP, A., KRESOVICH, S., BHAT, K.V., AYAD, W.G. & HODGKIN, T. 1997: Molecular Tools in Plant Genetic Resources Conservation: a Guide to the Technologies. IPGRI technical bulletin No.

2. International Plant Genetic Resources Institute, Rome, Italy.
KONNERT, M. & BERGMANN, F. 1995: The geographical distribution of genetic variation of silver fir (*Abies alba*, Pinaceae) in relation to its migration history. *Plant Syst. Evol.* **196**: 19–30.
LAGERCRANTZ, U. & RYMAN, N. 1990: Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation. *Evolution* **44**: 38–53.
LANGNER, W. 1953: Eine Mendelspaltung bei Aurea Formen von *Picea abies* (L.) Karst. als Mittel zur Klärung der Befruchtungsverhältnisse im Walde. *Z. Forstgenet.* **2**: 49–51.
LATTA, R.G. & MITTON, J.B. 1997: A comparison of population differentiation across four classes of gene marker in limber pine (*Pinus flexilis* James). *Genetics* **146**: 1153–1163.
LEFORT, F., DOUGLAS, G.C. & THOMPSON, D. 2000: Microsatellite DNA profiling of phenotypically selected clones of Irish oak (*Quercus* spp.) and ash (*Fraxinus excelsior* L.). *Silvae Genet.* **49**: 21–28.
LI, Y.-C., KOROL, A.B., FAHIMA, T., BEILES, A. & NEVO, E. 2002: Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Ecol.* **11**: 2453–2465.
LIEPELT, S., BIALOZYT, R. & ZIEGENHAGEN, B. 2002: Wind-dispersed pollen mediates postglacial gene flow among refugia. *Proc. Natl. Acad. Sci. U.S.A.* **99**: 14590–14594.
MARQUES, C.M., ARAUJO, J.A., FERRIERA, J.G., WHETTEN, R., O'MALLEY, D.M., LIU, B.H. & SEDEROFF, R. 1998: AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. *Theor. Appl. Genet.* **96**: 727–737.
MORGANTE, M., FELICE, N. & VENDRAMIN, G.G. 1998: Analysis of hypervariable chloroplast microsatellites in *Pinus halepensis* reveals a dramatic genetic bottleneck: a case study. Pp. 407–412 *in:* Karp, A., Isaac, P.G. & Ingram, D.S. (eds.): Molecular Tools for Screening Biodiversity: Plants and Animals. Chapman and Hall, London, United Kingdom.
MORGANTE, M. & OLIVERI, A.M. 1993: PCR-amplified microsatellites as markers in plant genetics. *Pl. J.* **3**: 175–182.
MORGANTE, M. & VENDRAMIN, G.G. 1991: Genetic variation in Italian populations of *Picea abies* L. Karst. and *Pinus leucodermis* Ant. Pp. 205–227 *in:* Müller-Starck, G. & Ziehe, M. (eds.): Genetic Variation in European Populations of Forest Trees. J.D. Sauerländer's Verlag, Frankfurt a. M., Germany.
MORITZ, C. 1994: Applications of mitochondrial DNA analysis in the conservation. A critical review. *Mol. Ecol.* **3**: 401-411.
MOSSELER, A., EGGER, K.N. & HUGHES, G.A. 1992: Low levels of genetic diversity in red pine confirmed by random amplified polymorphic DNA markers. *Can. J. For. Res.* **22**: 1332 –1337.
MUIR, G., FLEMING, C.C. & SCHLÖTTERER, C. 2000: Species status of hybridizing oaks. *Nature* **405**: 1016.
MULLIS, K.B. & FALOONA, F.A. 1987: Specific synthesis of DNA *in vitro* by a polymerase catalysed chain reaction. *Meth. Enzymol.* **155**: 335–350.
MURRAY, J.C., BENNET, S.R., KWITEK, A.E., SMALL, K.W., SCHINZEL, A., ALWARD, W.L., WEBER, J.L., BELL, G.I. & BUETOW, K.H. 1992: Linkage of Reiger syndrome to the region of the epidermal growth factor gene on chromosome 4. *Nat. Genet.* **143**: 1021–1032.
NAUTA, M.J. & WEISSING, F.J. 1996: Constraints on allele size at microsatellites loci: implications for genetic differenciation. *Genetics* **143**: 1021–1032.
NIELSEN, R. & PALSBOLL, P.J. 1999: Single-locus tests of microsatellite evolution: multi-step mutations and constraints on allele size. *Mol. Phylogenet. Evol.* **11**: 477–484.
PAGLIA, G.P., OLIVIERI, A.M. & MORGANTE, M. 1998: Towards second generation STS (sequence-tagged sites) linkage maps in conifers: a genetic map of Norway spruce (*Picea abies* K.). *Mol. Gen. Genet.* **258**: 466–478.
PARDUCCI, L., SZMIDT, A.E., MADAGHIELE, A., ANZIDEI, M. & VENDRAMIN, G.G. 2001: Genetic variation at chloroplast microsatellites (cpSSR) in *Abies nebrodensis* (Lojac) Mattei and three related *Abies* species. *Theor. Appl. Genet.* **102**: 733–740.

PASTORELLI, R., SMULDERS, MJM., VAN'T WESTENDE, WPC., VOSMAN, B., GIANNINI, R., VETTORI, C. & VENDRAMIN, G.G. 2003: Characterisation of microsatellite markers in *Fagus sylvatica* L. and *Fagus orientalis* Lipsky. *Mol. Ecol. Notes* **3**: 76–78.

PEAKALL, R. & SMOUSE, P.E. 2001: GenAlEx V5.04: Genetic Analysis in Excel. Population genetic software for teaching and research. Australian National University, Canberra, Australia. http://www.anu.edu.au/ BoZo/ GenAlEx/.

PETIT, R.J., AGUINAGALDE, I., DE BEAULIEU, J.-L., BITTKAU, C., BREWER, S., CHEDDADI, R., ENNOS, R., FINESCHI, S., GRIVET, D., LASCOUX, M., MOHANTY, A., MÜLLER-STARK, G., DEMESURE-MUSCH, B., PALMÉ, A., MARTÍN, J.P., RENDELL, S. & VENDRAMIN, G.G. 2003: Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**: 1563–1565.

PFEIFFER, A., OLIVIERI, A.M. & MORGANTE, M. 1997: Identification and characterization of microsatellites in Norway spruce (*Picea abies*). *Genome* **40**: 411–419.

PLOMION, C., COOKE, J., RICHARDSON, T., MACKAY, J. & TIKSAN, G. 2003: Report of the forest tree workshop at the plant and animal genome conference. *Comp. Funct. Genom.* **4**: 229 238.

POWELL, W., MORGANTE, M., MCDEVITT, R., VENDRAMIN, G.G. & RAFALSKI, J.A. 1995: Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. *Proc. Natl. Acad. Sci. U.S.A.* **92**: 7759–7763.

PRITCHARD, J.K., STEPHENS, M. C. & DONNELLY, P. 2000: Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

PURVES, W.K., ORUANS, G.H., HELLER, H.C. & SADAVA, D. 1998: Life: The Science of Biology. Fifth edn. Sinauer, Sunderland, U.S.A.

RAYMOND, M. & ROUSSET, F. 1995: GENEPOP (version 1.2) population genetic software for exact tests and ecumenicism. *J. Hered.* **86**: 248–249. http://www.cefe.cnrs-mop.fr/.

RIBEIRO, M.M., MARIETTE, S., VENDRAMIN, G.G., SZMIDT, A.E. & PLOMION, C. 2002: Comparison of genetic diversity estimates within and among populations of maritime pine using chloroplast simple-sequence repeat and amplified fragment length polymorphism data. *Mol. Ecol.* **11**: 869 –877.

RIBEIRO, M.M., PLOMION, C., PETIT, R., VENDRAMIN, G.G. & SZMIDT, A.E. 2001: Chloroplast single-sequence repeats variation in Portuguese maritime pine (*Pinus pinaster* Ait.). *Theor. Appl. Genet.* **102**: 97–103.

RICHARDSON, B.A., BRUNSFELD, S.J. & KLOPFENSTEIN, N.B. 2002: DNA from bird-dispersed seed and wind-disseminated pollen provides insights into postglacial colonization and population genetic structure of whitebark pine (*Pinus albicaulis*). *Mol. Ecol.* **11**: 215–227.

ROSSI, P., VENDRAMIN, G.G. & GIANNINI, R. 1996: Estimation of mating system parameters in two Italian natural populations of *Fagus sylvatica*. *Can. J. For. Res.* **26**: 1187–1192.

SCALFI, M., TROGGIO, M., PIOVANI, P., LEONARDI, S., MAGNASCHI, G., VENDRAMIN, G.G. & MENOZZI, P. 2004: A RAPD, AFLP and SSR linkage map and QTL analysis in European beech (*Fagus sylvatica* L.). *Theor. Appl. Genet.* **108**: 433–441.

SCHIERENBECK, K.A., SKUPSKI, M., LIEBERMAN, D. & LIEBERMAN, M. 1997: Population structure and genetic diversity in four tropical tree species in Costa Rica. *Mol. Ecol.* **6**: 137–144.

SCOTTI, I., MAGNI, F., PAGLIA, G.P. & MORGANTE, M. 2002: Trinucleotide microsatellites in Norway spruce (*Picea abies*): their features and the development of molecular markers. *Theor. Appl. Genet.* **106**: 40–50.

SCOTTI, I., VENDRAMIN, G.G., MATTEOTTI, L.S., SCARPONI, C., SARI-GORLA, M. & BINELLI, G. 2000: Postglacial recolonization routes for *Picea abies* K. in Italy as suggested by the analysis of sequence-characterized amplified region (SCAR) markers. *Mol. Ecol.* **9**: 699–708.

SEBASTIANI, F., CARNEVALE, S. & VENDRAMIN, G.G. 2004: A new set of mono- and dinucleotide chloroplast microsatellites in Fagaceae. *Mol. Ecol. Notes* **4**: 259-261.

SPERISEN, C., BÜCHLER, U., GUGERLI, F., MÁTYÁS, G., GEBUREK, TH. & VENDRAMIN, G.G. 2001: Tandem repeats in plant mitochondrial genomes: application to the analysis of population

differentiation in the conifer Norway spruce. *Mol. Ecol.* **10**: 257–263.

STREIFF, R., DUCOUSSO, A., LEXER, C., STEINKELLNER, H., GLÖSSL, J. & KREMER, A. 1999: Pollen dis–persal inferred from paternity analysis in mixed oak stands of *Quercus robur* L and *Q. petraea* (Matt.) Liebl. *Mol. Ecol.* **8**: 831–841.

TABERLET, P. & LIUKART, G. 1999: Non-invasive genetic sampling and individual identification. *Mol. Genet. Ani. Ecol.* **68**: 41–55.

TAYLOR, J.S., SANNY, P. & BREDEN, F. 1999: Microsatellite alleles size homoplasy in the guppy (*Poecilia reticulate*). *J. Mol. Evol.* **48**: 245–247.

TENZER, I., DEGLI IVANISSEVICH, S., MORGANTE M. & GESSLER, C. 1999: Identification of microsatellite markers and their application to population genetics of *Venturia inaequalis*. *Phytopathology* **89**: 748–753.

TIENDEREN, P.H. VAN, HANN, A.A. DE, LINDEN, C.G. VAN DER & VOSMAN, B. 2002: Biodiversity assessment using markers for ecologically important traits. *Trends Ecol. Evol.* **17**: 577– 582.

VENDRAMIN, G.G., ANZIDEI, M., BAGNOLI, F., PLOMION, C., SEBASTIANI, F. & RACCHI, M. 2003: Sequence diversity and SNP marker development in Aleppo pine (*Pinus halepensis* Mill). *In*: Proceedings of the Tree Biotechnology 2003 Congress, June 7–12 2003, Umeå, Sweden.

VENDRAMIN, G.G., DEGEN, B., PETIT, R.J., ANZIDEI, M., MADAGHIELE, A. & ZIEGENHAGEN, B. 1999: High level of variation at *Abies alba* chloroplast microsatellite loci in Europe. *Mol. Ecol.* **8**: 1117 –1126.

VENDRAMIN, G.G., LELLI, L., ROSSI, P. & MORGANTE, M. 1996: A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. *Mol. Ecol.* **5**: 595–598.

VIARD, F., FRANCK, P., DUPOIS, M.P., ESTOUP, A. & JARNE, P. 1998: Variation of microsatellite size homoplasy across electromorphs, loci, and populations in threee invertebrate species. *J. Mol. Evol.* **47**: 42–51.

VICARIO, F., VENDRAMIN, G.G., ROSSI, P., LIÒ, P. & GIANNINI, R. 1995: Allozyme, chloroplast DNA and RAPD markers for determining genetic relationships between *Abies alba* and the relic population of *Abies nebrodensis*. *Theor. Appl. Genet.* **90**: 1012–1018.

VOS, P., HOGERS, R., BLEEKER, M., REIJANS, M., VAN DE LEE, T., HORNES, M., FRIFITERS, A., POT, J., PELEMAN, J., KUIPER, M. & XABEAU, M. 1995: AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**: 4407–4414.

WEBER, J.L. & MAY, P.E. 1989: Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**: 388–396.

WEISING, K. & GARDNER, R.C. 1999: A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome* **42**: 9–19.

WILLIAMS, J.G.K., KUBELIC, A.R., LIVAK, K.J., RAFALSKI, J.A. & TINGEY S.V. 1990: DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**: 6531 –6535.

ZIEGENHAGEN, B., KUHLENKAMP, V., BRETTSCHNEIDER, R., SCHOLZ, F., STEPHAN, B.R. & DEGEN, B. 1999: DIG-labelled AFLPs in oaks - A DNA marker for reconstruction of full- or half-sib family relationships? *In*: Gillet, E.M. (ed.): Which DNA Marker for which Purpose? Final Compendium of the Forest Project in the EU Biotechnology FW IV Programme 'Molecular Tools for Biodiversity'.
http:/webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm.

ZIEGENHAGEN, B., SCHOLZ, F., MADAGHIELE, A. & VENDRAMIN, G.G. 1998: Chloroplast microsatellites as markers for paternity analysis in Abies alba. *Can. J. For. Res.* **28:** 317–321.

Forest genomics and new molecular genetic approaches to measuring and conserving adaptive genetic diversity in forest trees

K. V. Krutovsky & D. B. Neale

Introduction

Genetic diversity is the basis of the ability of organisms to adapt to changes in their environment through natural selection. Populations with little genetic variation are more vulnerable to the arrival of new pests or diseases, pollution, changes in climate and habitat destruction due to human activities or other catastrophic events. The inability to adapt to changing conditions greatly increases the risk of extinction. Genetic conservation and management aimed to save adaptive genetic diversity should be based on the knowledge of the genetic basis of adaptation. The goal of this paper is to describe how adaptive genetic diversity can be measured using new molecular genetic approaches and achievements in forest genomics.

Traditional methods to measure adaptive genetic diversity

Field experiments

Field experiments (common-garden tests) have been used traditionally to measure adaptive genetic diversity in trees. These tests continue to be used extensively in tree breeding and are very effective in identification of families and clones that are specifically adapted to particular environments or to a broad variety of environments. However, field experiments are very time consuming and relatively expensive, and more importantly, they are based solely on the phenotypes. They can estimate genetic parameters but only on measurable traits, not on individual genes. This method can neither provide information on what particular genes and how many of them are involved in adaptation nor how much of phenotypic variation can be explained by genetic variation in these genes. More details can be found in (see p. 275 ff., this volume).

Th. Geburek & J. Turok (eds.)
Conservation and Management of Forest Genetic Resources in Europe
Arbora Publishers, Zvolen, 2005, 369–390, ISBN 80-967088-1-3

Molecular genetic markers

Another, generally complementary, approach for estimating adaptive genetic diversity is to measure genetic variation using molecular genetic markers. However, DNA variation that resides in the non-coding genomic regions or does not lead to a change in the amino acid sequence (for example, so-called synonymous nucleotide substitutions in the second or the third positions in a codon encoding an amino acid) is unlikely to have any significant contribution to adaptation. Many modern genetic markers belong to so-called anonymous DNA marker type such as microsatellites or simple sequence repeats (SSRs), restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPDs), and amplified fragment length polymorphisms (AFLPs). These marker types generally measure apparently neutral DNA variation, and are very useful (with different efficiency, of course) in the analysis of phylogenetic relationships, population structure, mating system, gene flow, parental assignment, introgressive hybridization, marker-aided selection and genetic linkage. They are not useful for measuring adaptive genetic diversity.

Isoenzymes are another class of genetic markers widely used in forest genetics in the last several decades. Although variation revealed by these markers is caused by amino acid variation, it is unclear whether this variation is selectively neutral or has any adaptive significance. There are many studies showing great adaptive differences (in morphological or phenological characteristics) among populations of forest tree species, but no accompanying differences for the isoenzyme markers (see references in BOSHIER & YOUNG 2000).

Markers of all kinds are used now in forest genetics – both anonymous and genic, dominant and co-dominant, highly and less polymorphic, expensive and inexpensive, supposedly selective and apparently neutral, abundant and less numerous. A classification of genetic markers is offered in Table 1, which takes into account their most important features. Details on the nature of these markers, their advantages and disadvantages and use in different applications are available elsewhere (see the most recent reviews by CERVERA *et al.* 2000, LINHART 2000, GLAUBITZ & MORAN 2000, SAVOLAINEN & KARHU 2000, MANDAL & GIBSON 1998, *loc. cit.* chapters 12–14).

The ideal marker for estimating adaptive variation should meet the following criteria: (1) be directly involved in the genetic control of adaptive traits; (2) have identified DNA sequence and known function; (3) be readily available for genetic analysis, and (4) have easily identifiable allelic variation. No marker fully satisfies all these criteria. However, a promising new marker, expressed sequence tag polymorphisms (ESTPs), seems to satisfy most or all of these criteria, emerged recently as a result of genomic studies.

Table 1. Comparison of commonly used genetic markers.

Feature	RFLP	SSR	RAPD	AFLP	Isoenzymes	STS-EST
Origin	Anonymous genic	Anonymous	Anonymous	Anonymous	Genic	Genic
Maximum theoretical number of possible loci	Limited by the restriction site (nucleotide) polymorphism (tens of thousands)	Limited by the size of genome and number of simple repeats in a genome (tens of thousands)	Limited by the size of genome, and by nucleotide polymorphism (tens of thousands)	Limited by the restriction site (nucleotide) polymorphism (tens of thousands)	Limited by the number of enzyme genes and histochemical enzyme assays available (30–50)	Limited by the number of expressed genes (10,000–30,000)
Dominance	Co-dominant	Co-dominant	Dominant	Dominant	Co-dominant	Co-dominant
Null alleles	Rarely to extremely rare	Occasional to common	Not applicable (presence/absence type of detection)	Not applicable (presence/absence type of detection)	Rare	Rare
Transferability	Across genera	Within genus or species	Within species	Within species	Across families and genera	Across related species
Reproducibility	High to very high	Medium to high	Low to medium	Medium to high	Very high	High

Table 1. (continued).

Feature	RFLP	SSR	RAPD	AFLP	Isoenzymes	STS-EST
Amount of sample required per sample	2–10 mg DNA	10–20ng DNA	2–10 ng DNA	0.2–1 μg DNA	Several mg of tissue	10–20 ng DNA
Ease of development	Difficult	Difficult	Easy	Moderate	Moderate	Moderate
Ease of assay	Difficult	Easy to moderate	Easy to moderate	Moderate to difficult	Easy to moderate	Easy to moderate
Automation / multiplexing	Difficult	Possible	Possible	Possible	Difficult	Possible
Genome and QTL mapping potential	Good	Good	Very good	Very good	Limited	Good
Comparative mapping potential	Good	Limited	Very limited	Very limited	Excellent	Good to very limited
Candidate gene mapping potential	Limited	Useless	Useless	Useless	Limited	Excellent

Table 1. (continued).

Feature	RFLP	SSR	RAPD	AFLP	Isoenzymes	STS-EST
Potential for studying adaptive genetic variation	Limited	Limited	Limited	Limited	Good	Excellent
Cost						
Development	Moderate	Expensive	Inexpensive	Moderate	Inexpensive	Expensive
Assay	Moderate	Moderate	Inexpensive	Moderate to expensive	Inexpensive	Moderate
Equipment	Moderate	Moderate to expensive	Moderate	Moderate to expensive	Inexpensive	Moderate to expensive

How forest genetic conservation can benefit from new achievements in genomics

Introduction to genomics

Genomics has arisen as a new science that studies the whole genome by integrating traditional genetic disciplines such as population, quantitative and molecular genetics with new technologies in molecular biology, DNA analysis, bioinformatics and automated robotic systems (Fig. 1).

A number of subdisciplines of genomics can be combined to provide a powerful approach to studying adaptive genetic variation: structural, functional, comparative, statistical and associative genomics. A brief description of these subdisciplines might be useful in helping those new to the field to understand how modern genomics can affect genetic conservation.

Structural genomics

Structural genomics attempts to identify all the genes in the genome, sometimes called gene discovery (Fig. 2), and to determine their locations on the chromosomes. This goal is achieved by sequencing individual genes, gene segments or entire genomes. The individual genes are identified from the DNA sequence using sophisticated computer algorithms. The biochemical function of a gene is deduced via comparison of the DNA sequence with the sequences of genes of known function in the databases. When complete sequence of an entire genome is not

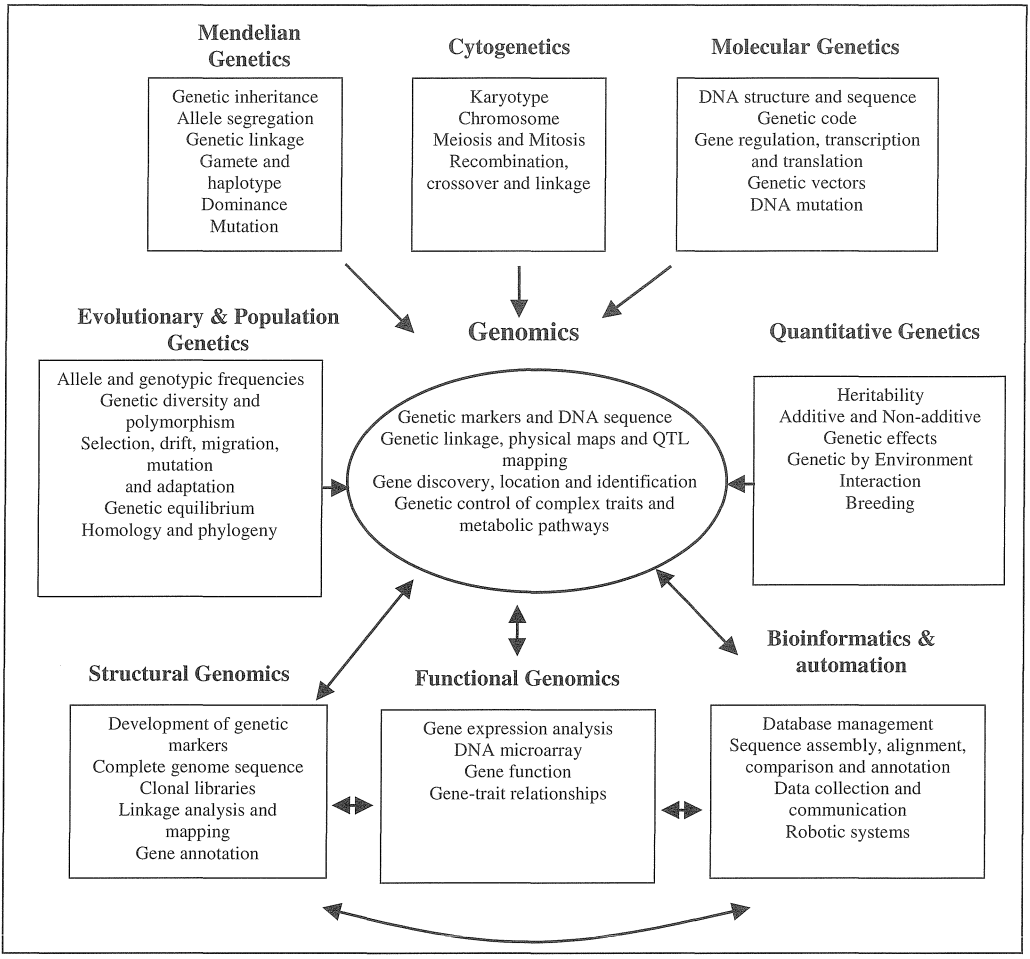

Figure 1. Genomics is a broad discipline that integrates traditional areas of genetics [adapted from Figures 1.1 and 1.2 in LIU (1998)].

available the location of genes can be determined either by direct physical mapping or by genetic mapping of the entire genome using numerous genetic markers. One of the most prominent applications of structural genomics for the study of adaptive genetic variation is quantitative trait loci (QTL) analysis via genome mapping. However, this approach aims to explain genomic structure and gene interaction at the genomic rather than functional level, unlike functional genomics.

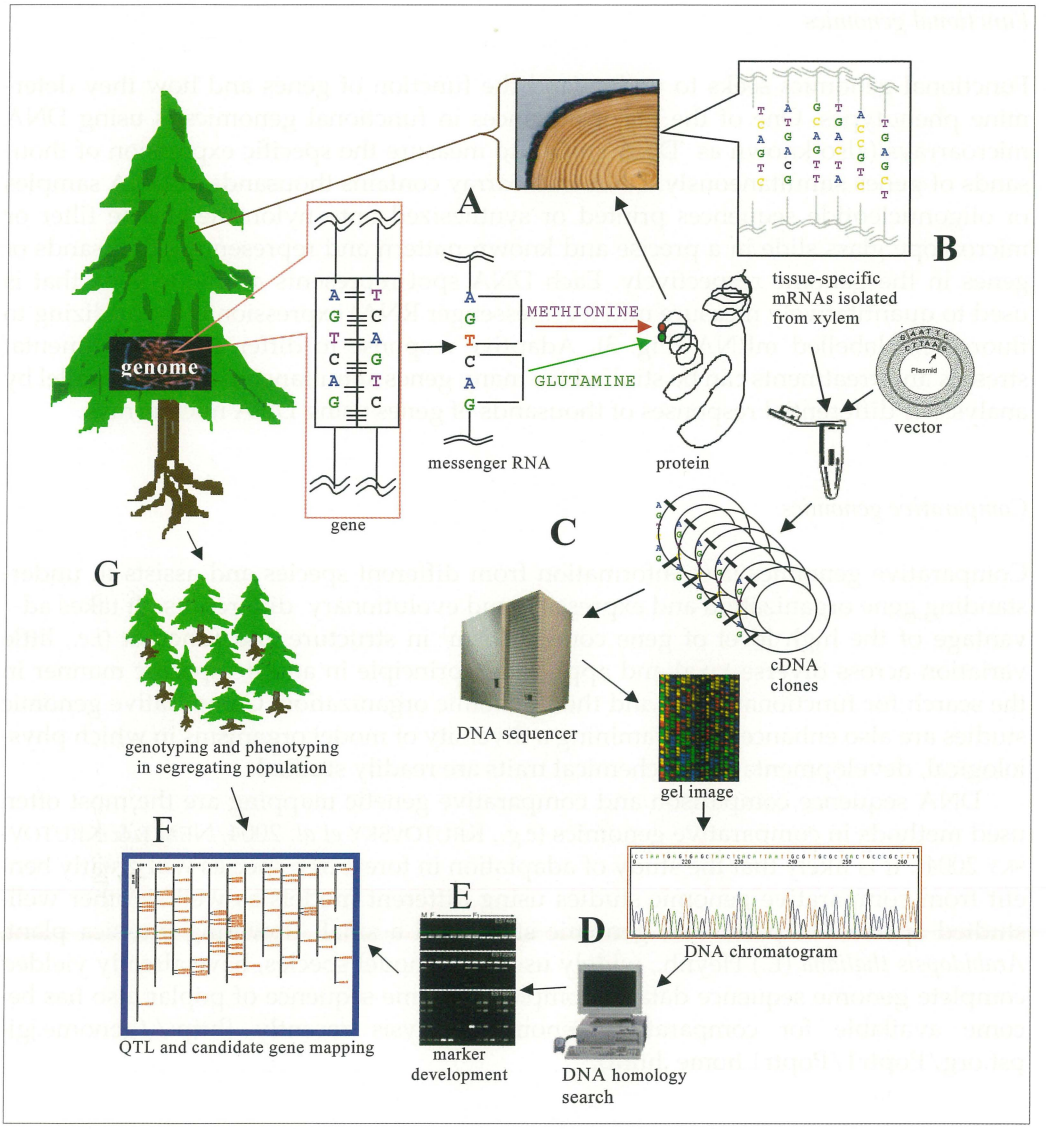

Figure 2. Discovery of adaptive genes. DNA in genes encode different proteins, for example, those with specific cellular functions related to the growth, wood quality and other adaptive traits (**A**); Sets of these genes are expressed according to unique patterns in time and location via messenger RNA (mRNA) that can be isolated, for instance, from xylem tissue and captured in vitro as complimentary DNA (cDNA) (**B**); cDNA are sequenced (**C**), and cDNA are compared to database sequences to suggest gene functions. Large-scale partial cDNA sequencing can identify many genes within a genome. Specific sets of genes expressed to produce specific structures (*e.g.*, wood) or specific physiological responses (*e.g.*, disease resistance) can be identified (**D**); expressed sequence tag (EST) genetic markers can be developed from these gene sequences (**E**); ESTs can be genotyped and mapped in the experimental population or progeny segregating for both genetic marker and adaptive trait (**F, G**).

Functional genomics

Functional genomics seeks to understand the function of genes and how they determine phenotypes. One of the major advances in functional genomics is using DNA microarrays (also known as 'DNA chips') to measure the specific expression of thousands of genes simultaneously. DNA microarray contains thousands of DNA samples or oligonucleotide sequences printed or synthesized onto nylon membrane filter or microscope glass slide in a precise and known pattern and representing thousands of genes in the genome respectively. Each DNA spot represents a unique gene that is used to quantitatively measure mRNA (messenger RNA) expression by hybridizing to fluorescent labelled mRNA (Fig. 3). Adaptive response to differ ent environmental stresses and treatments can be studied for many genes simultaneously or in parallel by analysis of differential responses of thousands of genes using DNA microarrays.

Comparative genomics

Comparative genomics uses information from different species and assists in understanding gene organization and expression and evolutionary differences. It takes advantage of the high level of gene conservatism[1] in structure and function (*i.e.*, little variation across diverse taxa) and applies this principle in an interspecific manner in the search for functional genes and their genomic organization. Comparative genomic studies are also enhanced by examining a diversity of model organisms in which physiological, developmental or biochemical traits are readily studied.

DNA sequence comparison and comparative genetic mapping are the most often used methods in comparative genomics (*e.g.*, KRUTOVSKY *et al.* 2004, NEALE & KRUTOVSKY 2004). It is likely that the study of adaptation in forest tree species will greatly benefit from comparative genomic studies using different models as well as other well-studied species. In particular, genomic studies in a small flowering brassica plant, *Arabidopsis thaliana* (L.) Heynh., widely used as a model species, have already yielded complete genome sequence data. A complete genome sequence of poplar also has become available for comparative genomic analysis recently (http://genome.jgi-psf.org/Poptr1/Poptr1.home .html).

Associative genomics

Associative genomics searches for mutations in populations via linkage disequilibrium analysis and via direct assessment of association between alleles and phenotypes. This approach can be effectively used in the search for adaptive mutations such as disease resistance, drought tolerance, cold hardiness, *etc.* DNA variants or mutations

[1] A nucleotide base sequence in a DNA molecule (or an amino acid sequence in a protein) that has remained essentially unchanged throughout evolution.

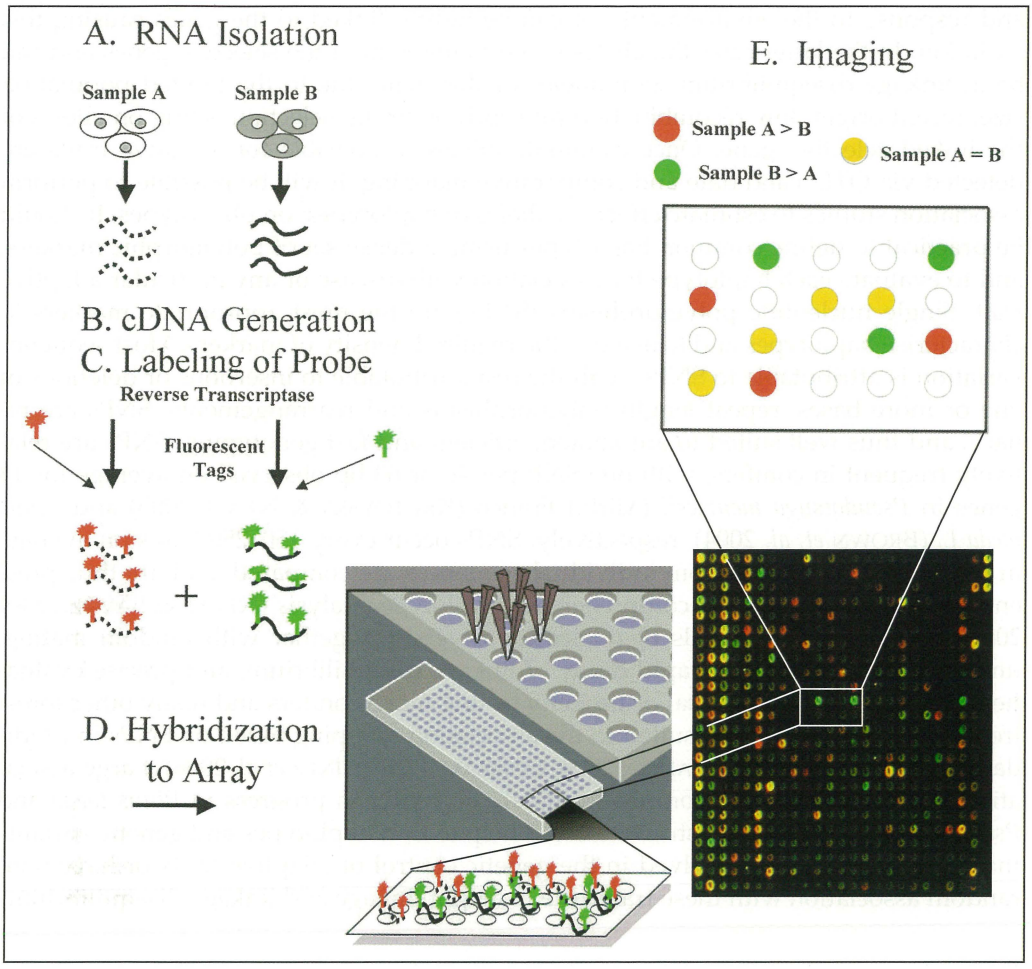

Figure 3. The use of DNA microarrays in differential gene expression analysis (adapted from ALBELDA and SHEPPARD 2000). Comparative hybridization experiment involves isolation of messenger RNA (mRNA) from two separate samples (**A**). The mRNA from each sample is treated with revers transcriptase (**B**) and labelled with a distinct fluorescent tag (**C**). The two pools of labelled RNA are mixed, hybridised to the DNA microarray containing a full set of thousands or tens of thousands of DNA sequences based on genome or complimentary DNA (cDNA) sequences, and washed (**D**). The microarray array is scanned using a specialised fluorimager, and the colour of each spot is determined (**E**). In this example, genes expressed only in Sample A would be red in colour, genes expressed only in Sample B would be green and those genes expressed equally in both samples would be yellow. This allows researchers to determine genes that are specifically expressed in response to the specific treatment or disease, or tissue-specific genes that are expressed in one tissue, but not in other.

(inherited differences in DNA sequence) can either directly contribute to phenotypic variation, influencing an individual's phenotypic characteristics (*e.g.*, risk of disease

and response to the environment), or can be tightly linked to the genes causing this variation. In the latter case, the alleles serve as markers of the selective genes and can be in linkage disequilibrium with alleles of this gene due to the limited population size, recent origin, low recombination rate and/or strong selection acting on alleles of the linked selective gene. Once candidate alleles responsible for adaptive traits are detected via QTL, candidate and comparative mapping, it will be possible to perform association studies to estimate effects of alleles or haplotypes[2] on phenotypes. It should be practical to define common haplotypes using a dense set of polymorphic markers, and to evaluate each haplotype for association with disease or any particular adaptive trait. Single nucleotide polymorphisms (SNPs) are the most appropriate markers to characterise haplotypes and to achieve the required density of markers. Most sequence variation is attributable to SNPs, with the rest attributable to insertions or deletions of one or more bases, repeat length polymorphisms and rearrangements. SNPs are binary, and thus well suited to automated, efficient and fast genotyping. SNPs are relatively frequent in conifers, with one SNP per 46 or 63 bp observed on average for 18 genes in *Pseudotsuga menziesii* (Mirb.) Franco (KRUTOVSKY & NEALE 2005) and *Pinus taeda L.* (BROWN *et. al.* 2004), respectively. SNPs occur every 150–250 bases on average in conifers when two random individual sequences are compared, and are thus present at sufficient density for comprehensive haplotype analysis (NEALE & SAVOLAINEN 2004). These adequate levels of nucleotide diversity together with random mating, unstructured populations, rapid decay of linkage disequilibrium, and precise evaluation of phenotype from clonal or progeny testing make conifers and many other forest tree species especially well suited to the association mapping based on SNPs in candidate genes (NEALE & SAVOLAINEN 2004, GONZÁLEZ-MARTÍNEZ *et al.* 2006). Large association mapping studies of complex traits are currently in progress in *Pinus taeda* and *Pseudotsuga menziesii*. Such studies should help to find haplotypes and genetic variants that are either directly involved in the genetic control of adaptive traits or have non-random association with these traits due to a tight linkage and linkage disequilibrium.

Statistical genomics

Statistical genomics is an integrative sub-discipline and serves all other areas of genomics. It provides statistical tools for genome and QTL mapping in structural genomics, bioinformatics tools for gene search, comparison and annotation in functional genomics, and statistical population genetic methods in associative genomics. Statistical genomics is also very important in developing computerised comprehensive interactive biological databases. New computer tools are required to compose genetic data at all levels of biological organization – from gene to population, species and ecosystems – for multiple purposes, including genetic conservation.

[2] A particular combination of alleles or sequence variations that are closely linked – that is, are likely to be inherited together – on the same chromosome.

Certainly, the division of genomics into these subdisciplines is rather arbitrary. Often the distinctions are vague or overlapping, but may be useful in helping those new to the field to understand modern genomics. In fact, genomics is a synthetic discipline that combines many methods and approaches of molecular biology, population and evolutionary genetics and bioinformatics (Fig. 1). The purpose of genomics is to study the structure, function and evolution of genome as a whole via complete genome sequencing, creating functional genetic maps for entire genomes and simultaneous analysis of patterns of differential expression of all or thousands of genes in the genome representing different cells and tissues and / or different treatments and conditions. It facilitates understanding genomes at both a molecular and a phenotypic level. It is likely that soon we will have a catalogue of all or most of genes expressed in plant and animal genomes and those that play essential roles in species- and population-level adaptation. Identifying and understanding the function of these genes, we can associate genetic variation with phenotypes and study adaptive genetic variation in different populations.

DNA sequencing of entire genomes

Complete sequencing of genomes of several important and model species is a significant achievement of genomics, which provides the basis for comparative and functional analysis. Answers to questions such as (1) the number, location and distribution of genes in genome; (2) gene organization and their function; (3) what genes are the same or highly conserved across different species; and (4) what genes are responsible for species adaptation and evolution can now be obtained. Complete genome sequences have been determined for the yeast *Saccharomyces cerevisiae* (May 1997), the nematode *Caenorhabditis elegans* (December 1998), the fruit fly *Drosophila melanogaster* Meigen (March 2000), the annual plant *Arabidopsis* (December 2000), the human (February 2001), rice (April 2002), mouse (December 2002), maize (April 2004) and, finally, poplar (September 2004).

The number of genes in a genome is limited and turns out to be not as high as expected earlier (for instance, only ~26,000 in plants and animals vs. ~6,000 in baker's or budding yeast, Table 2). Moreover, many genes are common across different species and are practically unchanged from the distant evolutionary past. For instance, only 94 of 1278 protein families in the human genome appear to be specific to vertebrates. The most elementary of cellular functions – basic metabolism, transcription of DNA into RNA, translation of RNA into protein, DNA replication and the like – evolved just once and have remained almost unchanged since the evolution of single-celled yeast and bacteria.

Comparative genomics explores such gene conservatism, which helps to understand and infer the function of a particular gene from the data obtained for similar homologous genes studied in other organisms. Much about forest tree gene functions can be learned from the data obtained in poplar. However, complete genome sequences are not yet available for any conifers, although advances in sequencing tech

Table 2. Genome size of several species for comparison.

Taxonomic rank	Latin name	Common name	Haploid chromosome	Nucleotide base (x 10^6)	Genes (x 10^3)
		Prokaryotae			
Archae	12[1]	archael microrganisms	–	1.6–3.0	1.5–2.7
Bacteria	40[1]	bacterial microorganisms	–	0.6–7.0	0.5–6.6
Bacteria	*Escherichia coli*[2]	no common name	–	4.6	4.3
		Eukaryotae			
Yeast	*Saccharomyces cerevisiae*[2]	baker´s or budding yeast	16	12	6
Worm	*Caenorhabditis elegans*[2]	nematode	5/6	97	19.5
Insect	*Drosophila melanogaster*[2]	fruit fly	4	180	14
Annual plant/angiosperm	*Arabidopsis thaliana*[2]	arabidopsis	5	125	30.7
Annual plant/angiosperm	*Oryza sativa*[2]	rice	12	400	38.9
Annual plant/angiosperm	*Zea mays*[2]	maize	10	2,400–3,200	31–58
Perennial plant/angiosperm	*Lycopersicon esculentum*	tomato	12	900	16–32
Forest tree/angiosperm	*Eucalyptus*[3]	eucalyptus	11	340–580	?
Forest tree/angiosperm	*Populus*[2]	poplars	19	500	24–30
Forest tree/gymnosperm	*Pinus*[3]	pines	12	20,000–30,000	35
Mammals/rodent	*Mus musculus*[2]	mouse	20	3500	27–30
Mammals/primate	*Homo sapiens*[2]	human	23	3400	27–33

[1] Number of species with completely sequenced genomes.
[2] Species with completely or almost completely sequenced genome.
[3] Data are based on several species.

nology should make it possible in the near future. Among various challenges are the complexity and large size of conifer genomes. The size of the pine genome (20,000–30,000 million bp), for example, is 6 to 8 times larger than the human genome (3,400 million bp), and 150 to 200 times larger than the genome of *Arabidopsis* (125 million bp; Table 2).

Gene discovery and expressed sequence tag polymorphisms (ESTPs)

An alternative to complete genome sequencing for discovering genes is being used in trees and many other organisms, which is based on identifying only the DNA that code for genes that are expressed in the genome. These sequences are called expressed sequence tags (ESTs). They are partial or complete sequences of complementary DNA (cDNA) obtained from mRNA isolated from different tissues and therefore represent genes expressed in these tissues with often known or suggested function (Fig. 4). EST sequences are compared to all other sequences in gene databases to identify matches likely representing highly homologous genes. If there is a high similarity (homology) to some other gene sequence whose identity has been determined, then the identity of the EST can be immediately inferred. ESTs can be used as a source for identifying candidate genes for QTLs involved in genetic control of adaptive traits. Large libraries of partial or complete sequences of thousands of expressed genes have already been obtained, and large databases of EST sequences are available for many animal and plant species, including several forest tree species, such as Monterey or radiata pine (*Pinus radiata* D. Don.), loblolly pine[3] (*P. taeda*), Norway spruce (*Picea abies* [L.] Karst.), *Eucalyptus* and *Populus*.

Expressed sequence tag polymorphisms (ESTPs) are derived from ESTs. Using EST sequences polymerase chain reaction (PCR) primers are designed to amplify ESTs from individual genomic DNA (HARRY *et al.* 1998). Allelic polymorphism in the amplification product (ESTPs) can be revealed using different modern methods for detection and visualisation of DNA alterations (KRISTENSEN *et al.* 2001).

ESTPs mostly reveal genetic variation within genes, although variation can be found in both coding and non-coding regions of genes. Thus, ESTPs are the most informative markers in terms of gene function among the most recently developed one and are the first genetic markers that offer real potential for detecting adaptive genetic diversity broadly.

Physical and genetic mapping of the whole genome using numerous genetic markers

Genetic linkage mapping is central to genomics. It allows the positioning of genes and genetic markers on a specific chromosome. There are two kinds of maps: physical and genetic. Physical maps provide the exact location of genes or genetic markers on chromosomes. These maps are either assembled from the complete genome sequences,

[3] See http://www.cbc.med.umn.edu/ResearchProjects/Pine/

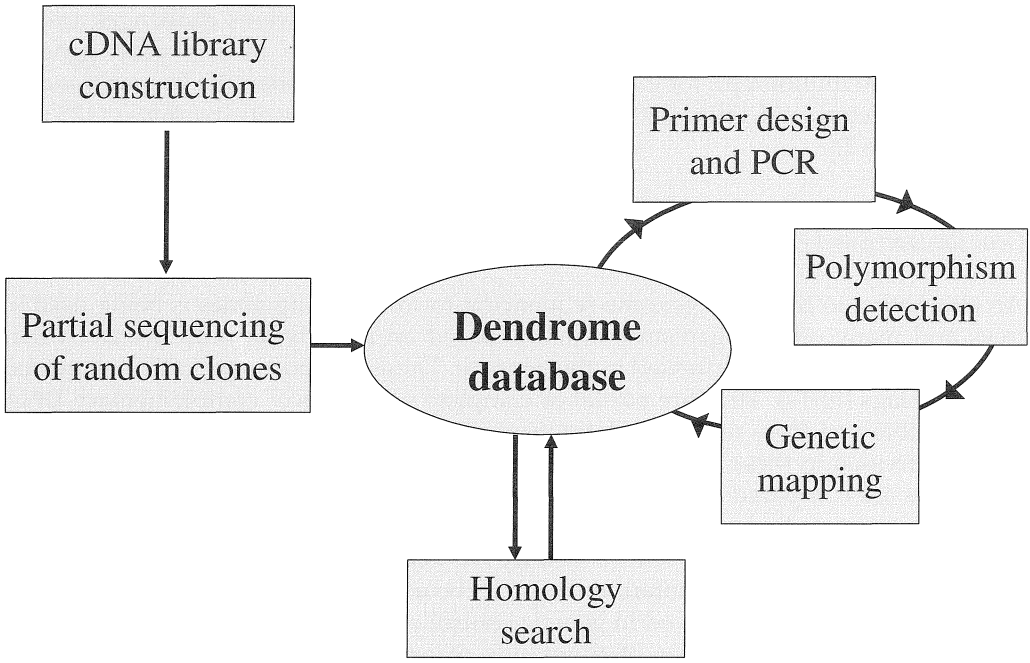

Figure 4. The development of expressed sequence tag (EST) markers in forest trees that can be used in comparative and candidate gene mapping. EST markers are derived from partial or complete sequences of complimentary DNA (cDNA) libraries that obtained from messenger RNA (mRNA) isolated from different tissues (for instance, xylem). EST sequences are submitted to gene databases and compared to all other sequences in the databases to identify matches likely representing highly homologous genes. Polymerase chain reaction (PCR) primers based on the EST sequences are designed to amplify these genes. If these genes are polymorphic and segregate in the experimental population or progeny, they can be used in the genome and quantitative trait loci (QTL) mapping. They are good candidate genes for QTLs.

BAC[4] contigs[5], or based on *in situ* hybridization or other methods. However, as long as the complete genome sequences of forest tree species are not available the alternative approach is to develop genetic linkage maps by segregation and linkage analysis. Genetic maps identify the distance and order between markers based on the number of recombination events between them. Genetic maps have been already constructed for many different forest tree species using a variety of genetic marker types (see Table 1; NEALE & SEDEROFF 1996, KRUTOVSKII *et al.* 1998, CERVERA *et al.* 2000 for review). A complete sequence alone is not sufficient to understand the genetic control of adaptive

[4] Bacterial artificial chromosome (BAC): A chromosome-like structure, constructed by genetic engineering that carries large segments of DNA – 100,000 to 200,000 bases – from another species cloned into bacteria. Once the foreign DNA has been cloned into the host bacteria, many copies of it can be made.

[5] A group of clones representing overlapping regions of a genome.

traits. These traits are usually very complex, have quantitative inheritance and are controlled by many genes each with relatively small effects, which are called quantitative trait loci (QTL). Genetic maps can be used to study the number, location and distribution of QTLs in a genome via their genetic linkage mapping with molecular markers. Following this approach, a new genomic technique called QTL mapping has been relatively recently developed.

Analysis of genetic control of complex adaptive traits via quantitative trait loci (QTL) mapping

The method for finding and locating QTLs is called QTL mapping. The conceptual basis of this method is comparatively simple but it requires relatively dense genetic maps with evenly distributed markers covering the entire genome, appropriate statistical tools, and an experimental population of sufficient size segregating for both genetic markers and phenotypic traits (*e.g.*, PATERSON 1998). First, multi-locus genotypes (molecular markers) and phenotypes (quantitative traits) are measured on all individuals of segregating population. Then, phenotypic values are statistically associated with genotypic values, usually using multiple regression or maximum likelihood methods to identify markers that have a strong association (joint segregation) with the quantitative trait. Such association can be the result of either tight linkage of the genetic marker and QTL (*i.e.*, because they reside in the same region of the chromosome) or direct involvement of this particular marker(s) in genetic control of the trait. QTLs have been already detected and mapped in forest trees for such adaptive traits as growth rhythm, phenology, form, wood quality, disease resistance, cold hardiness, drought tolerance, and others (see NEALE 1998, SEWELL & NEALE 2000 and NEALE *et al.* 2002 for review). Once a QTL controlling an adaptive trait has been precisely mapped, it then may become possible to clone the gene underlying the QTL based solely on the knowledge of its genetic map position and without knowing its function or DNA sequence. This is known as positional or map-based cloning.

Numerous recently developed PCR-based markers (*e.g.*, RAPD, AFLP, SSR, STS, *etc.*) are used in QTL mapping (*e.g.*, SEWELL & NEALE 2000 and NEALE *et al.* 2002 for review). However, many of these markers are either dominant or anonymous, and their functions are unknown. There are three important aspects to consider when choosing a genetic marker system for QTL mapping: the outbred nature of forest tree pedigrees (1), the potential for comparative (2) and candidate gene (3) mapping. First, each parent of an outbreed pedigree is typically a different, highly heterozygous individual, where the transmission of up to four different alleles must be followed from the parents to progeny. Therefore, multiallelic co-dominant markers are best suited to detect the maximum number of polymorphisms found in the heterozygous parents. Second, comparative mapping, both within and among species, is an important tool for relating results from different mapping experiments. Therefore a subset of the markers used in a mapping experiment should be orthologous[6] across pedigrees and species.

[6] Loci in two species that have arisen from the same locus of their common ancestor.

Third, to identify actual genes controlling a quantitative trait, genes with known or suggested function should be used in QTL mapping. Complete or partial cDNA sequences (ESTs) allow now researchers to design ESTP markers that take into account all these aspects and can be used for genetic mapping of the entire genome and for measuring adaptive genetic diversity via QTL mapping analysis (*e.g.*, HARRY *et al.* 1998, TEMESGEN *et al.* 2001, NEALE *et al.* 2002). These are the most informative markers for adaptive trait candidate gene mapping that is now used in animal and plant species, mostly agriculture stocks and crop species, to identify genes for different yield and quality traits including also adaptive traits such as biomass, growth rate, fecundity and other reproductive traits, disease resistance, *etc.*

Candidate gene mapping of adaptive genes

Candidate gene mapping is based on the assumption that a gene with known or assumed function that may affect genetic control of a trait can be considered a 'candidate gene' for this trait (*e.g.*, GION *et al.* 2000, NEALE *et al.* 2002, WHEELER *et al.* 2005). Furthermore, it is assumed that if this gene is also mapped to the same genomic region as a QTL for this trait, then this gene is very likely to be this QTL that directly controls the trait, although the likelihood depends on marker density, precision of QTL map and genome size.

Large forest tree EST projects will identify and provide DNA sequences that give researchers enough material to develop genetic markers for an unlimited number of genes that can be used as a source of possible candidate genes to target particular adaptive traits (TEMESGEN *et al.* 2001, NEALE *et al.* 2002, WHEELER *et al.* 2005). Different subsets of specific EST markers can be used in mapping adaptive gene. For instance, EST markers derived from genes that are supposedly related to the cell defence mechanism can be used to map QTLs controlling disease resistance; EST markers derived from genes that are involved in the wood formation can be efficiently used in QTL mapping of wood related traits, *etc.* If function of genes used to derive ESTs is unknown but they represent cDNA isolated from a specific tissue or obtained from the cells that undergone a specific treatment, they still can be used as candidate genes in QTL mapping. For instance, heat shock genes expressed during experimental heat stress can be used to map genes related to drought resistance. The use of such meaningful markers as ESTs directly in genetic mapping makes analysis of adaptive variation more efficient and focused. In addition, highly efficient and sensitive methods are now being developed to detect allelic differences between these genes that can be used for mapping (*e.g.*, SNP detection).

Identifying candidate genes for QTLs controlling adaptive traits in trees would ultimately provide the diagnostic tools to screen large amounts of wild germplasm for individuals carrying alleles worthy of conserving. The challenge is to identify DNA polymorphisms within candidate genes that will distinguish alleles and then associate alleles with differences among phenotypes. This can be accomplished through SNP discovery and association studies. The approach is to identify SNPs within regions of candidate genes involved in the control of a trait, to genotype a large number of individuals from the natural population at these SNPs, and to test for associations between

SNPs and phenotypes. This approach will soon be available for application in forest tree conservation programs because of the intensity and progress of research and development activities.

Comparative mapping of adaptive genes

Comparative mapping is one aspect of comparative genomics and another very promising genomic approach for discovering adaptive genes. It takes advantage of high similarity in gene location in chromosomes of closely related species and applies it across different species to search for functional genes and their genomic organization. Comparative mapping in various species has shown that gene content and gene order are conserved over long chromosomal regions among related species of animals or angiosperm plants. These results strongly suggest that similar studies can be effectively done in the forest trees. The genetic maps of closely related species can be directly compared due to synteny (*i.e.*, co-occurrence of two or more genes on the same chromosome) among the genomes of these species. Indeed, the high levels of co-linearity among, for instance, pine species (*e.g.*, BROWN *et al.* 2001, CHAGNÉ *et al.* 2003, KOMULAINEN *et al.* 2003) means that genetic information from one species can be applied to others (Fig. 5).

The most valuable alleles of adaptive genes can be identified from the pool of all species and possibly incorporated into breeding and conservation strategies. Furthermore, the controls and interactions affecting adaptive trait expression can be studied. Further studies should show whether comparative mapping between distantly related forest trees, for example between *Populus* or *Eucalyptus* and *Arabidopsis*, is also possible.

The development of genetic resources for comparative genomic analysis in forest trees would have significant impacts in many areas of forest genetic conservation research. Comparative mapping would facilitate: (1) verification of QTLs controlling adaptive traits, (2) identification of candidate genes and (3) the understanding of evolutionary relationships. The emphasis in forest genetic conservation is not on a single species, but on many,

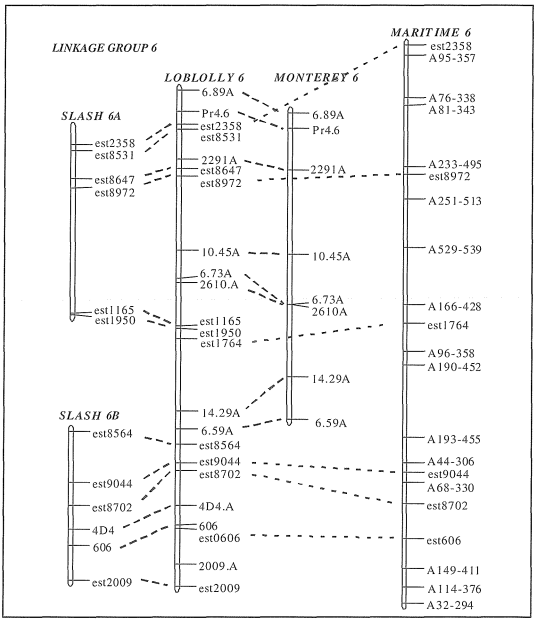

Figure 5. Comparative genetic linkage maps of linkage group 6 of loblolly, slash, Monterey and maritime pines aligned using common expressed sequence tag (EST) markers and illustrating the potential utility of loblolly pine ESTs as anchored reference loci. Loci connected by a dotted line were detected by the same EST marker (BROWN *et al.* 2001).

each with its own regional economic and ecological distinctions. Comparative genetic mapping in pines and other conifers follows this paradigm, focussing not only on the creation of individual species maps but also on the consensus maps to identify the genomic locations of genes affecting quantitatively inherited adaptive phenotypes, resistance to pathogens, and other biological and physiological characteristics.

Comparative mapping is possible if orthologous[7] genetic markers have been mapped to each of the species maps to be compared. Orthologs are genes that have descended from a common ancestral locus, whereas paralogs are loci that have originated by gene duplications within an individual species.

Most of the anonymous markers (*e.g.*, RAPD, AFLP, and SSR) cannot be used for comparative mapping because they are not orthologus among species. Genetic markers that are based on genic DNA sequences, such as RFLPs and ESTPs, are more suited for comparative mapping. For example, RFLP loci from both *Pinus taeda* and *Pinus radiata* have been used to construct comparative maps between these species (DEVEY *et al.* 1999). However, because RFLP markers do not easily distinguish between orthologs and paralogs and because they are difficult to apply, they are unlikely to be used widely for comparative mapping. ESTPs are the most useful markers for comparative mapping and have been already used in genetic mapping in conifers (TSUMURA *et al.* 1997, PERRY & BOUSQUET 1998, CATO *et al.* 2001, TEMESGEN *et al.* 2001). ESTPs reveal orthologs among species and only occasionally paralogs. EST markers from *Pinus taeda* have been used to construct comparative maps for this species and slash pine, *Pinus elliottii* Engelm. (BROWN *et al.* 2001), maritime pine, *Pinus pinaster* Ait. (CHAGNÉ *et al.* 2003), Scots pine, *Pinus sylvestris* L. (KOMULAINEN *et al.* 2003) and Douglas-fir, *Pseudotsuga menziesii* (KRUTOVSKY *et al.* 2004; see also NEALE & KRUTOVSKY 2004 for review).

Bioinformatics and genomic databases

The highly efficient, fast and productive technologies of genomic studies enable the collection of overwhelming amounts of data. The primary genomic data types are DNA and protein sequences, genetic mapping data and data resulting from functional analysis. Much of the data are freely available to the public via the Internet and World-Wide-Web (WWW). Everybody benefits from public access to the genomic databases, but especially researchers with a small research budget who can still do efficient computer data analysis and gene discovery. The National Centre for Biotechnology Information (NCBI) in the USA and European Molecular Biology Laboratory (EMBL) in Europe are the primary sites for DNA sequence databases and DNA sequence analysis tools. The primary databases are called GenBank and EMBL. They also provide on-line access to the BLAST (Basic Local Alignment Search Tool) programs, which are the primary tools used to search the databases and identify matches among sequences.

[7] Similarity in DNA or protein sequences between different species due to common ancestry. Describes the evolutionary origin of a locus. Loci in two species are said to be orthologous when they have arisen from the same locus of their common ancestor.

The primary repository of forest tree genomic data is the TreeGenes Database that is maintained by the Dendrome Project at the Institute of Forest Genetics, Davis, California (http://dendrome.ucdavis.edu). TreeGenes contains a variety of data-types and is an object-oriented database that allows complex queries and searches. Through the use of databases and bioinformatic tools, it is possible to perform experiments *in silico* and begin to understand all the complex relationships among genes and how they work together to determine adaptive phenotype.

Conclusions

The study of adaptation is fundamental to forestry and forest genetic conservation. Forest geneticists have long used common-garden experiments and, to a lesser extent, molecular markers to study patterns of adaptation in forest trees. Phenotypic assessments are time consuming and expensive, and provide no information about variation in the genes controlling adaptive variations. There are numerous molecular marker technologies available, but most measure either neutral or highly conservative genetic variation of limited adaptive value. There is a need for developing rapid and informative diagnostic techniques for evaluating large numbers of adaptive genes and prospective trees for *in situ* conservation. Genomics provides new tools to study adaptation in trees. Forest geneticists can use automated, highly efficient, fast and productive technologies to determine DNA sequences and to genotype large numbers of individuals. They can ultimately identify genes responsible for forest tree adaptation via EST sequencing, QTL and candidate gene mapping. Then, using modern genotyping technologies and association studies they can determine allelic diversity for these candidate genes in forest tree populations and directly measure adaptive allelic diversity for thousands of genes simultaneously.

Despite remarkable progress much work remains to be done to understand the nature of genetic variation that underlies adaptive forest tree phenotypes. Comprehensive understanding will first require discovering, annotating and cataloging all genes in the forest tree genome. One approach towards achieving this goal is to determine the DNA sequence of the entire genome and infer the genes from the DNA sequence. This approach is currently not feasible in all forest trees because of their large genome size, but *Populus* – with a relatively small genome size – can serve as a model species. An alternative (or parallel) approach is to determine the DNA sequences for the gene-coding regions only. This can be accomplished by isolating mRNA, preparing cDNA libraries from this mRNA and sequencing cDNA. These EST sequences are submitted to databases and compared to all other sequences in the databases to see if they match to genes whose function has been determined. EST databases of tens of thousands of ESTs have been already produced and are publicly available for *Pinus*, *Picea*, *Populus*, and *Eucalyptus*.

The second step towards understanding adaptation involves construction of genome, QTL, comparative and consensus linkage maps for most forest tree species (*e.g.*, SEWELL *et al.* 1999). Genetic maps show the position of genes and are valuable for understanding genome organization and evolution. Maps are extremely useful tools for identifying genes controlling interesting phenotypes. Loci controlling quantitatively

inherited traits, so-called QTLs, have been already identified in many forest trees for a variety of growth, wood quality, and other economic and adaptive traits. These data are immediately useful for tree improvement and genetic conservation.

Next, DNA microarray analysis can be used to study the expression patterns of genes, and to understand the function of all genes and their interactions. The relationship between the vast amount of allelic diversity in genes and the array of different phenotypes found in forest tree populations can be studied. A catalogue of common coding–sequence variants in forest tree genes can be created and tested for association with a phenotype. Genome-wide high-resolution maps of known polymorphisms can be used to scan the genome for marker-adaptive trait associations.

The analysis need not be limited to coding sequences. It may be that the majority of relevant mutations reside in regulatory regions. Thus, it is important to identify variants in at least the proximal and distal regulatory sequences as our poor understanding of 'regulatory' elements dictates the need for a more global approach. An approach in which marker-trait associations are sought will require the construction of a high-resolution map of genetic variants. SNPs are the natural candidates for this map because they are abundant, have a smaller mutation rate than microsatellites and can be genotyped en masse using microarray technology.

A map-based association search for multiple adaptive loci, each contributing to the total phenotype in a small yet measurable way, is feasible via haplotype analysis. The alleles of these loci can be indirectly recognized by their historical associations with other genetic variants (*e.g.*, SNPs) in their neighbourhood. The non-random association of variants with one another (linkage disequilibrium) is a well-known feature of the plant and animal genomes. DNA microarrays will have a major role in genotyping thousands of genes simultaneously, in the creation of fine maps and in mapping the components of complex adaptive phenotypes. Forest genomics has a bright future and awaits exiting applications in forest tree management and genetic conservation.

Acknowledgements

We would like to thank Garth Brown (Institute of Forest Genetics), Deborah Rogers (University of California at Davis) and Christel Palmberg-Lerche (Forest Resources Development Service, FAO) for providing useful comments on the manuscript.

References

ALBELDA, S.M. & SHEPPARD, D. 2000: Functional genomics and expression profiling: be there or be square. *Amer. J. Resp. Cell Mol. Biol.* **23**: 265–269.

BOSHIER, D.H. & YOUNG, A.G. 2000: Forest conservation genetics: limitations and future directions. Pp. 289–297 *in*: Young, A., Boshier, D & Boyle, T. (eds.): Forest Conservation Genetics: Principles and Practice. CABI, Oxon, United Kingdom.

BROWN, G.R., GILL, G. P., KUNTZ, R.J., LANGLEY, C.H. & NEALE, D.B. 2004: Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc. Natl. Acad. Sci. U.S.A.* **101**: 15255–15260.

BROWN, G.R., KADEL, E.E., BASSONI, D.A., TEMESGEN, B., VAN BUIJTENEN, J.P., MARSHALL, K.A. & NEALE, D.B. 2001: Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. *Genetics* **159**: 799–809.

CATO, S.A., GARDNER, R.C., KENT, J. & RICHARDSON, T.E. 2001: A rapid PCR-based method for genetically mapping ESTs. *Theor. Appl. Genet.* **102**: 296–306.

CERVERA, M.T., PLOMION, C. & MALPICA, C. 2000: Molecular markers and genome mapping in woody plants. Pp. 375–394 *in*: Jain, S.M. & Minocha, S.C. (eds.): Molecular Biology of Woody Plants. Forestry Sciences, Volume 64. Kluwer Academic Publishers, Dordrecht, The Netherlands.

CHAGNÉ, D., BROWN, G., LALANNE, C., MADUR, D., POT, D., NEALE, D. & PLOMION, C. 2003: Comparative genome and QTL mapping between maritime and loblolly pines. *Mol. Breed.* **12**: 185–195.

DEVEY, M.E., SEWELL, M.M., UREN, T.L. & NEALE, D.B. 1999: Comparative mapping in loblolly pine and radiata pine using RFLP and microsatellite markers. *Theor. Appl. Genet* **93**: 656 –662.

GION, J.-M., RECH, PH., GRIMA-PETTENATI, J., VERHAEGEN, D. & PLOMION, C. 2000: Mapping candidate genes in *Eucalyptus* with emphasis on lignification genes. *Mol. Breed.* **6**: 441–449.

GLAUBITZ, J. & MORAN, G.F. 2000: Genetic tools: the use of biochemical and molecular markers. Pp. 39–59 *in*: Young, A., Boshier, D & Boyle, T. (eds.): Forest Conservation Genetics: Principles and Practice. CABI, Oxon, United Kingdom.

GONZÁLEZ-MARTÍNEZ, S.C., KRUTOVSKY, K.V. & NEALE, D.B. 2006: Forest tree population genomics and adaptive evolution. *New Phytologist* (in press).

HARRY, D.E., TEMESGEN, B. & NEALE, D.B. 1998: Codominant PCR-based markers for *Pinus taeda* developed from mapped cDNA clones. *Theor. Appl. Genet.* **97**: 327–336.

HUMAN GENOME PROJECT: *Nature* **409** (6822), February 15, 2001 and *Science* **291** (5507), February 16, 2001.

JAIN, S.M. & MINOCHA, S.C. (eds.) 2000: Molecular Biology of Woody Plants. Forestry Sciences, Volume 64. Kluwer Academic Publishers, Dordrecht, The Netherlands.

KOMULAINEN, P., BROWN, G. R., MIKKONEN, M., KARHU, A., GARCIA-GIL, M. R., O` MALLEY, D., LEE, B., NEALE, D.B. & SAVOLAINEN, O. 2003: Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda. Theor. Appl. Genet.* **107**: 667–678.

KRISTENSEN, V.N., KELEFIOTIS, D., KRISTENSEN, T. & BØRRESEN-DALE, A.-L. 2001: High-throughput methods for detection of genetic variation. *BioTechniques* **30**: 318–332.

KRUTOVSKII, K.V., VOLLMER, S.S., SORENSEN, F.C., ADAMS, W.T., KNAPP, S.J. & STRAUSS, S.H. 1998: RAPD genome maps of Douglas-fir. *J. Hered.* **89**: 197–205.

KRUTOVSKY, K.V. & NEALE, D.B. 2005: Nucleotide diversity and linkage disequilibrium in cold-hardiness and wood quality-related candidate genes in Douglas-fir. *Genetics* **171**: 2029–2041.

KRUTOVSKY, K.V., TROGGIO, M., BROWN, G.R., JERMSTAD, K.D. & NEALE, D.B. 2004: Comparative mapping in the Pinaceae. *Genetics* **168**: 447–461.

LINHART, Y. B. 2000: Variation in woody plants: molecular markers, evolutionary processes and conservation biology. Pp. 341–374 *in*: Jain, S.M. & Minocha, S.C. (eds.): Molecular Biology of Woody Plants. Forestry Sciences, Volume 64. Kluwer Academic Publishers, Dordrecht, The Netherlands.

LIU, B.-H. 1998: Statistical Genomics: Linkage, Mapping and QTL Analysis. CRC Press, New York, U.S.A.

MANDAL, A.K. & GIBSON, G.L. (eds.) 1998: Forest Genetics and Tree Breeding. CBS Publishers, New Delhi, India.

NEALE, D.B. 1998: Molecular genetic approaches to measuring and conserving adaptive genetic diversity. Pp. 385–390 *in*: The Proceedings of International Symposium on *In Situ* Conservation of Plant Genetic Diversity, November 4–8, 1996, Antalya, Turkey. CRIFC, Turkey.

NEALE, D.B. & KRUTOVSKY, K.V. 2004: Comparative genetic mapping in trees: the group of conifers. Pp. 267–277 *in*: Lörz, H. & WENZEL, G. (eds.): Biotechnology in Agriculture and Forestry: Molecular Marker Systems. Springer Verlag, Berlin, Germany.

NEALE, D. B. & SAVOLAINEN, O., 2004: Association genetics of complex traits in conifers. *Tr. Pl. Sci.* **9**: 325–330.

NEALE, D.B. & SEDEROFF, R.R. 1996: Genome mapping in gymnosperms: A case study in loblolly pine (*Pinus taeda* L.). Pp. 407–423 *in*: Paterson, A.H. (ed.): Genome Mapping in Plants. Academic Press, New York, U. S.A.

NEALE, D.B., SEWELL, M.M. & BROWN, G.R. 2002: Molecular dissection of the quantitative inheritance of wood property traits in loblolly pine. *Ann. For. Sci.* **59**: 595-605.

PATERSON, A. H. (ed.) 1998: Molecular Dissection of Complex Traits. CRC Press, New York, U.S.A.

PERRY, D.J. & BOUSQUET, J. 1998: Sequence-tagged-site (STS) markers of arbitrary genes: development, characterization and analysis of linkage in black spruce. *Genetics* **149**: 1089–1098.

SAVOLAINEN, O. & KARHU, A. 2000: Assessment of biodiversity with molecular tools in forest trees. Pp. 395–406 *in*: Jain, S.M & Minocha, S.C. (eds.): Molecular Biology of Woody Plants. Forestry Sciences, Volume 64. Kluwer Academic Publishers, Dordrecht, The Netherlands.

SEWELL, M.M. & NEALE, D.B. 2000: Mapping quantitative traits in forest trees. Pp. 407 –423 *in*: Jain, S.M. & Minocha, S.C. (eds.). Molecular Biology of Woody Plants. Forestry Sciences, Volume 64. Kluwer Academic Publishers, Dordrecht, The Netherlands.

SEWELL, M.M., SHERMAN, B.K. & NEALE, D.B. 1999: A consensus map for loblolly pine (*Pinus taeda* L.). I. Construction and integration of individual linkage maps from two outbred three-generation pedigrees. *Genetics* **151**: 321–330.

TEMESGEN, B., BROWN, G.R., HARRY, D.E., KINLAW, C.S. & NEALE, D.B. 2001: Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). *Theor. Appl. Genet.* **102**: 664–675.

TSUMURA, Y., SUYAMA, Y., YOSHIMURA, K., SHIRATO, N. & MUKAI, Y. 1997: Sequence-tagged-sites (STSs) of cDNA clones in *Cryptomeria japonica* and their evaluation as molecular markers in conifers. *Theor. Appl. Genet.* **94**: 764–772.

WHEELER, N.C., JERMSTAD, K.D., KRUTOVSKY, K.V., AITKEN, S.N., HOWE, G.T., KRAKOWSKI, J. & NEALE, D.B. 2005: Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. IV. Cold-hardiness QTL verification and candidate gene mapping. *Mol. Breed.* **15**: 145 –156.

YOUNG, A., BOSHIER, D. & BOYLE, T. (eds.) 2000: Forest Conservation Genetics: Principles and Practice. CABI , Oxon, United Kingdom.

Selection of target species and sampling for genetic resources in absence of genetic knowledge

G. Eriksson

Introduction

Separate design and implementation of genetic conservation programmes for individual species are unlikely owing to the huge number of tree species. For this reason it may be useful to examine ways to select species. Selected species are usually referred to as *target species*. Another concept that is frequently used in discussions of genetic conservation is associated with *keystone species*. This type of species is of importance for the continued survival and well-being of many other species, which are referred to as *associated species*. There is no clear-cut border between keystone species and associated species. Even if target species are selected, too many species may still remain to make individual genetic conservation programmes feasible. Strategies for grouping of species may be required such as was done for noble hardwoods within the European Forest Genetic Resources Programme (EUFORGEN) network (ERIKSSON 2001).

Many analyses of the optimum strategy for sampling of populations for genetic conservation of crop plants were presented even before genetic conservation in forestry had been discussed (*e.g.*, FRANKEL & HAWKES 1974). In conservation of crop genetic resources the focus is on saving alleles that may be of value for future use in breeding whereas conservation of forest genetic resources has safeguarding of the potential for adaptation as its prime objective. Besides, the long-generation time of most forest tree species requires another sampling strategy than for mostly annual crop species. The absence of genetic knowledge about adaptive population differentiation for most noble hardwoods called for an attempt to develop a sampling strategy for these species (ERIKSSON 1998).

Target species

Selection of target species is often based on charismatic, salvage, scientific, or socio-economic grounds.

Th. Geburek & J. Turok (eds.)
Conservation and Management of Forest Genetic Resources in Europe
Arbora Publishers, Zvolen, 2005, 391–411, ISBN 80-967088-1-3

The conflict between foresters and conservation biologists around the conservation of the spotted owl in the north-western Americas involves one of the most spectacular, charismatic species. *A priori,* it may be assumed that the cost of saving individual species would drain resources from conservation of a larger number of species. Since the last decade of the previous century ecosystem conservation rather than conservation of individual species, has been widely promoted. However, charismatic species have probably contributed to fund raising for conservation, so they may serve as a means to broaden conservation activities (SOULÉ 1991). Conservation of many charismatic species may frequently be a salvage of the species as well. It is understandable that many conservationists focussed on endangered or threatened species in their conservation efforts. The American chestnut (*Castanea dentata* [Marsh.] Borkh.) is a good example of a seriously threatened species, which urgently needs to be saved since it suffers severely from the chestnut blight disease (*Cryphonectria parasitica* [Murrill] Barr.).

It is quite natural that geneticists favour selection of target species on scientific grounds (*e.g.*, FRANKEL 1983). With respect to speciation, genera with large speciation during recent time constitute one extreme while genera with just one species constitute the other extreme. The latter type of species would be of value to conserve since such a species may have unique genetic constitution. However, ERWIN (1991) dismissed the idea of conservation efforts on 'living fossils', which he regards as doomed to extinction. Instead, he suggested concentration of the efforts on actively evolving species, for which he saw a higher probability of success in the conservation efforts. The reason for this would be that species in such a genus should be equipped with a large adaptability. This is not easy to verify (see p. 199 ff., this volume). Selection of target species based on ecological characteristics might also be regarded as a 'scientific' selection. Pollination pattern (wind or animal vectors), type of distribution (wide or limited; continuous or scattered), and behaviour in the ecosystem as climax or pioneer species, are examples of the characteristics that can also be used for selection of species. VARELA and ERIKSSON (1995) argued that ecological keystone species should be targeted for conservation since a proper genetic conservation of such species would probably also guarantee conservation of the species associated with them. The designation of a species as a target species based on scientific grounds is not a simple task since we lack so many genetic facts and data.

Since the pressure on forests is expected to increase considerably, especially in the developing countries, there is an urgent need to concentrate the conservation efforts on tree species of present and future economic value. NAMKOONG (1986) presented one sensible classification on economic grounds, in which he separated between species of commercial, potentially commercial and non-commercial use.

A combination of socio-economic and ecological reasons for selection of target tree species is likely the most sensible. In countries with net growth of their forests the socio-economic grounds for selection of target species may be alleviated to the benefit of other grounds for selection.

Grouping of species

Whether a species is included in long-term breeding activities or not is a useful first grouping criterion (ERIKSSON 2000). In the Nordic countries, for instance, breeding of *Picea abies* (L.) Karst. and *Pinus sylvestris* L., is both intensive and long-term, comprising most of their distribution in these countries. This means that the genetic conservation is guaranteed within the breeding programme of these species.

The EUFORGEN Noble Hardwoods network has focussed on grouping of species since there are 25 tree species listed as mandate species of this network (ERIKSSON 2001). The first criterion for grouping of the Noble Hardwoods is abundance of the species. It is based on the assumption that rarely occurring species will harbour less additive variance and thereby having lower potential for adaptation than commonly occurring species. As a corollary of this assumption, the methods for conservation of rarely occurring species will differ from commonly occurring species. The network has come to the conclusion that measures should be taken to increase the effective population size, N_e, of rarely occurring species. Species with recent rarity should be given priority among the rarely occurring tree species. Far-reaching fragmentation caused by human activities is one example of recent rarity.

The next grouping level is the mating system, *i.e.* whether the species is wind-pollinated or whether insects or other animals are pollen vectors. A large number of isoenzyme studies on among- and within-population variation in tree and other plant species were carried out. These studies show that the ratio between within-population and among-population genetic variation is on an average larger in wind-pollinated species than in species with animals as pollen vectors (*e.g.*, GOVINDARAJU 1988). Behaviour in the ecosystem – pioneer, intermediate and climax – is still another level in the grouping of species. Pioneer species typically invade newly opened areas, which are comparatively uniform. A pioneer species would benefit from a more limited within-population variation than a climax species, which faces less uniformity during its growth. If a genotype of a pioneer species has a high adaptedness to a specific site condition then it would be advantageous for this species to rely on this genotype. Therefore, vegetative propagation of this genotype such as is the case for the weed species *Taraxacum vulgare* Schrank would be advantageous. It should be noted that a species might behave like a pioneer under certain conditions and like a climax or intermediate species under different environmental conditions.

Capturing adaptedness

Sampling of the existing adaptedness is a central element of any forest genetic conservation programme. The sampling of the existing adaptedness may be simple if all variation is included within every population and not between them. Then, conservation would only have to ensure that sufficient population sizes are included and any one population or any mixture of populations would be as good as any other for those purposes. However, if variation exists among populations, the apportionment and design of population sampling, as well as sample sizes within them, becomes an issue.

Therefore, the existing variation in adaptedness should be looked upon as a means to capture maximum variance rather than as a final goal of conservation. The sampling of the present state does not necessarily oblige a conservation programme to freeze the available structure but should instead be regarded as a starting point for continued evolution (ERIKSSON *et al.* 1993).

For the majority of tree species there are limited or no hard facts on among-population genetic variation in adaptedness. This means that sampling at best will be informed guesses about existing structure. The main purpose of this paper is to discuss how such guesses might be derived from expected among-population genetic differentiation, with the main emphasis on random mating populations.

Evolutionary factors

The effects of the evolutionary factors: mutation, genetic drift, natural selection, and gene flow are discussed in more detail in the (see p. 199 ff., this volume). It is widely known and documented that gene flow is a strong factor to prevent fixation of neutral alleles. Similarly, genetic drift is a strong evolutionary factor in small populations, though unlike gene flow it causes fixation of alleles. According to a survey on the strength of natural selection in the wild, ENDLER (1986) found that natural selection could vary from weak to as strong as in breeding. Mutation frequencies are mostly regarded as being low at individual loci amounting to approximately 10^{-5} per generation. However, the pooled mutation frequency at loci influencing a trait may be considerably higher 10^{-2}–10^{-3} (LANDE & BARROWCLOUGH 1987). KÄRKKÄINEN *et al.* (1994) reported a mutation frequency for chlorophyll mutants in *Pinus sylvestris* of 10^{-2} based on data published by EICHE (1955). Phenotypic plasticity is another evolutionary factor, which might be regarded as a disguise of the genotype. It has not been discussed much as an opposing factor to natural selection.

The factors promoting and constraining among-population differentiation are visualized in Fig. 1. In most cases, natural selection within a population is stabilizing, which will result in a sharpening of differences among populations at the species level. Therefore, natural selection at the species level will be experienced as disruptive. The reason for this is that the phenotypes with highest fitness will vary among populations.

The action of these factors on within-population variation is illustrated in Fig. 2. The gene flow, the genetic drift and inbreeding constitute the factors making up the mating pattern. The latter is defined as the matings realized in the population. The mating pattern creates the raw material that the stabilizing natural selection can act upon. Gene flow brings in migrants from other populations with other allele frequencies and is the only factor that causes an increase of the genetic variation within a population. Similarly, new mutations will increase the within-population variation. Note that in Fig. 1 it is assumed that mutations increase the differentiation among populations. The contradiction that might appear from a comparison of Fig. 1 and 2 is

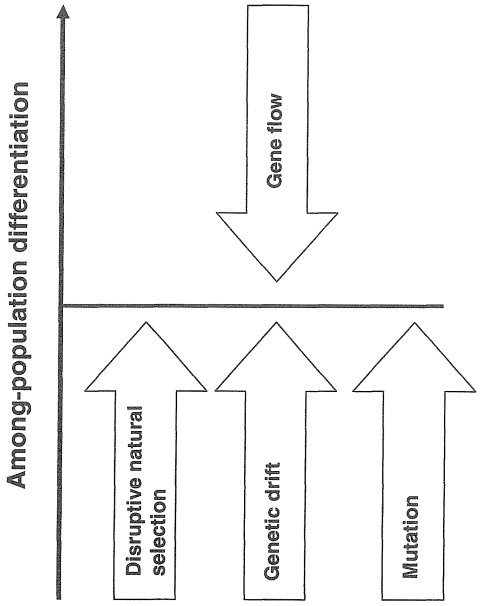

Figure 1. Schematic illustration of the evolutionary factors that tend to increase and decrease the differentiation among populations. Natural selection should be understood as disruptive selection among populations.

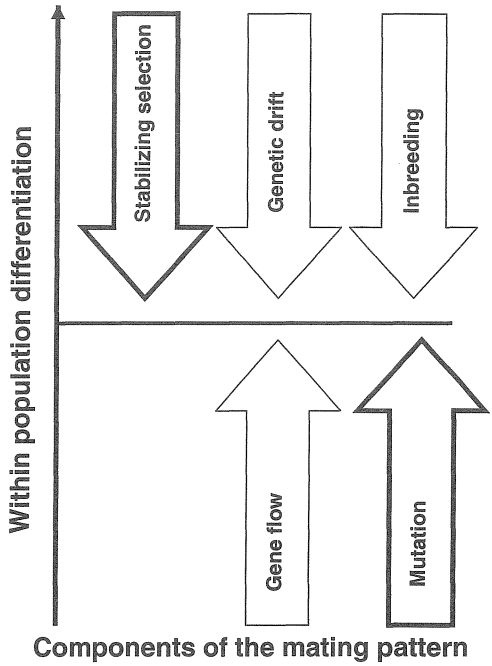

Figure 2. Schematic illustration of the evolutionary factors influencing the genetic variation within a population. The mating pattern, defined as the zygotes formed in a particular population, is a consequence of gene flow, genetic drift, and inbreeding in that population. The mating pattern creates the raw material that may be changed by stabilizing natural selection among populations.

due to the assumption that individual mutations are of such a low frequency that different populations will get different mutations. In this way they will contribute to a very small extent to population differentiation, simultaneously each new mutation will increase the genetic variation within a population.

Random mating populations

In populations with a large effective population size the effect of genetic drift is negligible. Nor will mutations influence the structure to any great extent in large populations. Therefore, the focus of the discussion is on the interaction between disruptive natural selection and gene flow in development of among-population genetic differentiation. Some comments on phenotypic plasticity are given.

Selective environmental neighbourhoods

A prime prerequisite for among-population differentiation is that the environment occupied by a species is experienced as heterogeneous by this species. Without heterogeneity there will be no difference in optimum phenotypes within this area. A second prerequisite is that there is some consistency in this heterogeneity over generations. Otherwise, natural selection would favour different phenotypes in different generations. The concept of selective environmental neighbourhood (SEN) introduced by BRANDON (1990) is useful for a discussion of this prerequisite. Within a SEN there are no ranking changes of the genotypes with respect to fitness; conferring a degree of homogeneity to a SEN. Thus several SENs are required for among-population differentiation. Stabilizing selection within each of several populations in different SENs will be experienced as disruptive selection among populations (ERIKSSON & EKBERG 2001). Besides, the geographic extension of SENs over generations must not vary too much otherwise natural selection will cause shifts in different directions during different generations.

A SEN should be regarded as a genetic delimitation rather than a geographic one, which means that a SEN may contain non-contiguously growing populations. As one example of this, populations at each of the two slopes of a valley may physically not grow adjacent to each other but still be parts of one SEN. As the environment changes the physical shape of a SEN may change or even merge with another SEN. One physical population constituting one single SEN remains so until environmental changes cause ranking shifts of fitness in this population. It should be noted that SENs are trait specific. Within a certain geographic area, one trait may contribute to fitness whereas another does not. In another area, both traits may contribute to fitness and the SENs for the two traits may in the extreme case be identical. Thus, different traits may have different numbers of SENs within a species. Differences in the physical area of different SENs among traits are dependent upon the contribution to fitness of the trait and its correlation with other fitness contributing traits. The trait specificity of SEN is a great advantage of this concept compared to niche, which is understood as a geographic area with some unique characteristics and valid for all traits. Below some examples will be given, in which SENs within a region vary dependent on the trait.

With high probability, a wide climatic distribution of a species will give rise to numerous SENs and adaptation to different climatic zones (= among-population differentiation) may take place. Other abiotic and biotic elements of site conditions may also be suspected to give rise to different SENs, although much less is known about adaptation to such environmental variables (*e.g.*, JONSSON & ERIKSSON 1989, WEISER 1995). In summary, a prerequisite for a species to consist of one single SEN is that it grows under homogeneous conditions both in space and time.

As long as there is additive variance for fitness of a trait, disruptive natural selection may cause a trait differentiation among populations inhabiting different SENs. A large difference between two SENs will give rise to a stronger disruptive selection than when the difference is small. Even if there are large differences and ample additive variance is present, among-population differentiation may be absent owing to substantial gene flow among individuals of different SENs.

Gene flow among different selective environmental neighbourhoods

If there is no mating contact among selectively divergent populations, evolution can be expected to eventually lead to speciation, but with complete inter-mating, every generation would re-establish the original Hardy-Weinberg frequencies regardless of the intensity of selection. For most forest trees, however, populations are neither completely isolated nor completely inter-mating, and in these cases, the outcome of the balance between selection and migration is not clear. When the population structure is one in which a very large central population (a continent) exists and small islands are attached by isolated migrants only to that continent, it does not take more than one migrant per generation to prevent strong divergence from evolving. However, if populations are only partially isolated, and particularly if migration is limited to only a few adjacent populations of similar size (a stepping stone model) then the same exchange rate may allow a wide distribution of allele frequencies to evolve among the populations even if no speciation occurs (HEDRICK 1988). Even if selection was not divergent among the populations, local mutations might still lead to the existence of genetic variations if migration was at a low rate (PHILLIPS 1996). Forest trees might also differ among the sexes in their rates of migration, with sometimes, pollen or seed being less mobile than the other, and the effect of the less mobile sex may allow for more variations to evolve than if an average migration rate were computed (GREGORIUS & NAMKOONG 1983, NAMKOONG & GREGORIUS 1985).

The consequences of inter-pollination among different SENs are discussed with the aid of the schematic illustrations in Figs. 3–5. An example for a species with a continuous distribution is given in Fig. 3. Pollen of wind-pollinated tree species may be transported over large distances (> 10 km according to KOSKI 1970). The share of such distant pollen in the formation of a new generation in a SEN might be considerably less. Since inter-mating among different SENs in many cases is a strong and preventive factor operating against genetic differentiation, the mere occurrence of swamping of pollen from other SENs may prevent or strongly reduce the among-population differentiation even if there are several SENs. Therefore, the degree to which there is inter-pollination among different SENs will be important for the possibility to develop among-population differences. Transfers of seeds and other propagules will also reduce the possibilities for differentiation. It is therefore relevant to introduce the concept of a zone of shared gene pool (ZSGP). ZSGP will mostly be different from N_e and may vary from one extreme of total random mating within the ZSGP (ZSGP = N_e) to exchange of one single migrant among different subzones of a ZSGP. Therefore, the strength of the gene flow (SGF) among SENs will influence the amount of differentiation that may take place. The degree of genetic differentiation will depend on the strength of the two counteracting factors, disruptive natural selection and gene flow.

If the trees in the seven SENs of Fig. 3 can be regarded as members of one randomly mating population (*i.e.* ZSGP = N_e), then natural selection must be extremely strong to give rise to among-population differentiation. A contrasting case is given for a species with a scattered distribution in Fig. 4. In this case the two populations (SEN1 and SEN2) are growing at some distance from each other and there is no exchange of pollen or propagules between the two SENs. Such a situation is a good starting condi-

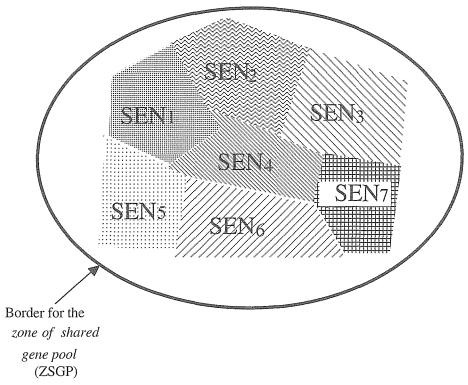

Figure 3. Schematic illustration of 7 selective environmental neighbourhoods, among which there is free intermating.

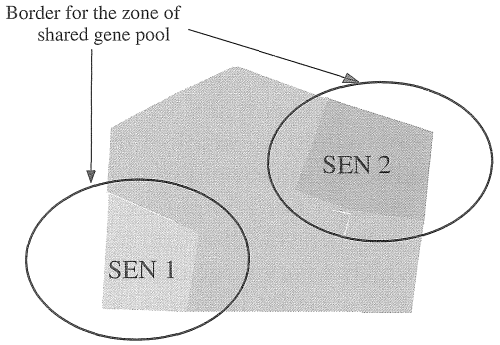

Figure 4. Schematic illustration of 2 selective environmental neighbourhoods, between which there is no intermating.

tion for specific adaptation to varying site conditions. A situation intermediate to the two described above is shown in Fig. 5. This is the situation for many wind-pollinated and continuously distributed species where there is an overlap of the zones of shared gene pool along an environmental gradient. This would in itself favour the development of clinal variation and isolation by distance (WRIGHT 1943).

In lowland tropical rain forests it is very likely that a SEN covers a huge area. Within this area, species that have a few sexually mature individuals per hectare, may be characterized by a meta-population structure with most of the pollinations within each subpopulation. If each subpopulation is small, $N_e < 20$, this may lead to pronounced allele fixation by genetic drift.

Phenotypic plasticity

A species with a short generation turnover can respond to changed directions of selection by adaptation via natural selection, presupposed that the species contains enough additive genetic variance. On the other hand a species with a long generation turnover cannot during its lifetime respond to the multitude of directions of selection that the environment may provoke. Rather, it is probable that a certain degree of phenotypic plasticity would be evolutionary advantageous (Fig. 6). If the situation for a tree is that of Fig. 3, the tree might spread its genes over all seven SENs. Then it would be an advantage for that tree, teleologically speaking, to give rise to a progeny that performs well in all SENs, *i.e.* to have a large phenotypic plasticity. In other words phenotypic plasticity confers fitness to its carrier. The larger the gene flow the higher the probability that phenotypic plasticity would confer fitness in species with long generation turnover.

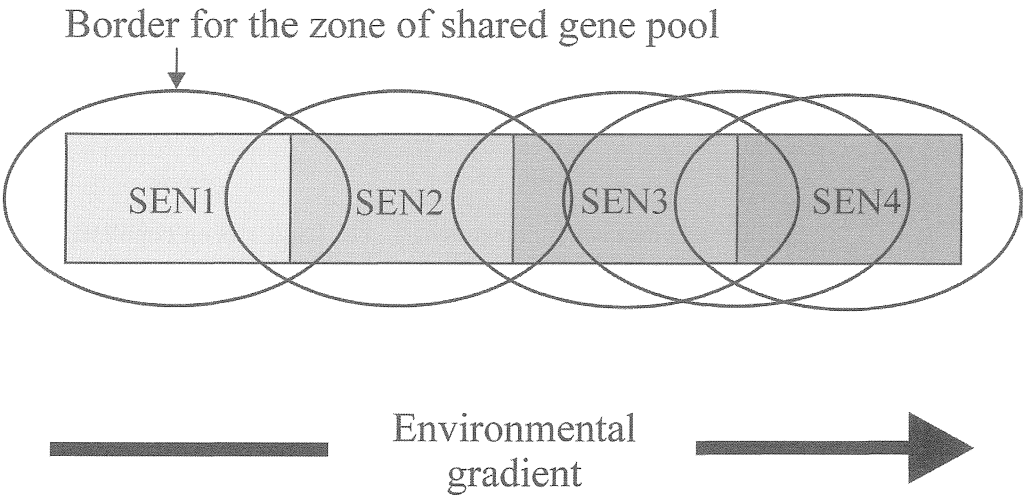

Figure 5. Schematic illustration of 4 selective environmental neighbourhoods along an ecological gradient with intermating among adjacent selective environmental neighbourhoods.

Differentiation in species with large random mating populations

In Table 1 the outcome with respect to among-population differentiation and within-population variation at different combinations of gene flow and among-population disruptive selection is schematically summarised. At limited gene flow and weak selection there may be some among-population differentiation for random reasons (cell 1). A strong among-population selection will occur if limited gene flow is combined with large differences among the SENs occupied by the species (cell 2). This is the case for the situation depicted in Fig. 4 and to some extent also for Fig. 5. The condition for this is merely that there are large differences among the SENs. The result in cell 3 is obtained if there is a substantial gene flow among populations without an accompanying disruptive among-population selection, *i.e.* there will be just one SEN. When both the gene flow is sub-

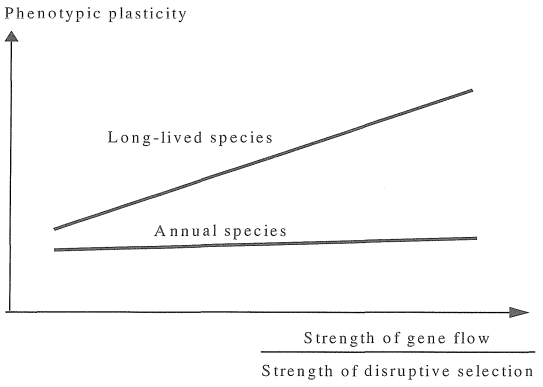

Figure 6. Schematic illustration of the expected phenotypic plasticity in long-time and short-time generation species plotted against the ratio between strength of the gene flow and strength of disruptive among-population selection.

Table 1. Schematic presentation of the among-population differentiation and within-population variation in large random-mating populations at different combinations of strong and weak disruptive among-population selection at limited or substantial gene flow, respectively.

		Disruptive among-population selection	
		weak	strong
Gene flow	limited	1. Some among-population variation due to random reasons	2. Large among-population differentiation
	substantial	3. Very limited among-population differentiation and a large within-population variation	4. Some among-population differentiation and a large within-population variation

Box 1. Ecocline and ecotype.

Ecocline, a continuous change of population means along an ecological gradient attributed to changes in allele frequency
Ecotype, group of individuals within a species with a certain adaptedness to the conditions at a specific site.

stantial and the disruptive selection is strong we may obtain some among-population differentiation accompanied with a large within-population variation caused by the strong gene flow (cell 4). The table depicts extreme situations in this 2 × 2 factorial combination. In nature the situation may be less extreme and many species will take intermediate positions.

The counteracting factors gene flow and disruptive among-population selection (Fig. 1) are of great significance for the development of clinal or ecotypic variation. G. Turesson was probably the first to use the term ecotype for plants with specific adaptedness to certain site conditions (TURESSON 1922). The combination in cell 2 of Table 1 is a good condition for development of ecotypes within a species while the combination in cell 4 is a good condition for ecoclinal variation.

G. Turesson studied many perennial plants growing under strongly different site conditions. Since there was no or very limited gene flow among the sites, he found certain 'rock ecotypes' and 'meadow ecotypes'. These ecotypes kept their characteristics when cultivated outside their natural habitats. This suggests that adaptation to the different site conditions had taken place. The studies in California by Clausen and his co-workers are classical (CLAUSEN *et al.* 1940), in which it was shown that the plants kept their performance after transplantation at other site conditions than at their origins, are other good examples of ecotypic differentiation. The most thorough discussion of the controversy between ecotypes and ecoclines was carried out by O. Langlet

in his classical paper *A cline or not a cline – the question of Scots pine* (LANGLET 1958). Already in his dissertation on variation in dry matter among populations of *Pinus sylvestris* during autumn (LANGLET 1936) he showed a clinal variation. Therefore, it is not appropriate to talk about ecotypes in species such as *Pinus sylvestris, Picea abies* and many other wind-pollinated species in situations where they occur with a wide and continuous distribution.

Informed guesses about the relationship between the strength of disruptive selection and gene flow in a species can be obtained from its type of distribution and type of pollination. A wide distribution as compared to a limited distribution mostly means that there are several SENs, especially when the distribution covers a broad span of climatic conditions. The gene flow is also dependent on the type of pollen vector. Wind-pollinated species will on an average disperse their pollen over wider areas than species with other types of pollen vectors since the pollen dispersal by insects, bats and birds is assumed to be over shorter ranges. If there are physical obstacles between disjunct populations, then it will result in reduced gene flow among populations.

Species with assumed large random mating populations

Some examples from tree species are given to illustrate observations about among- and within-population variation. Pine and spruce species are wind-pollinated with pollen flights over wide areas (KOSKI 1970, SAVOLAINEN 1991). Many species also occupy wide, continuous, and climatically variable areas, which would indicate firstly that they have a large number of SENs and secondly that the gene flow is considerable among different SENs. In agreement with this, estimates of N_e are high amounting to several thousand trees (SCHOEN & BROWN 1991, YANG & YEH 1995).

Of particular interest is the investigation of among- and within-population variation in *Pinus sylvestris* in Finland by KARHU *et al.* (1996) since it treats one adaptive trait, bud set, and several molecular markers in the same populations. There was a 21 days difference in bud set between the most southern (lat. 60°) and the most northern population (lat. 70°). All the molecular markers showed limited among-population variation while the within-population was large. REED and FRANKHAM (2001) analyzed 71 data sets, in which markers and quantitative were studied simultaneously. They concluded that '*... molecular measures of genetic diversity have only a very limited ability to predict quantitative genetic variability*'. The absence of strong relationships between molecular and quantitative traits was foremost attributed to the selection that influences quantitative traits.

The results presented clearly demonstrate that the number of SENs varies from trait to trait. Isoenzymes, RAPDs, and RFLPs and other markers are assumed to be neutral (KIMURA 1983). It should be noted that there are several reports suggesting that isoenzymes not always are neutral (*e.g.*, HATTEMER 1994). For neutral traits there is just one SEN and thereby no disruptive selection for them. Since *Picea abies* and *Pinus sylvestris* have a substantial gene flow among populations, one expects that the neutral traits will occupy a position close to the line 1–3 in Fig. 7. For random reasons there will be some genetic variation among populations and the species position will

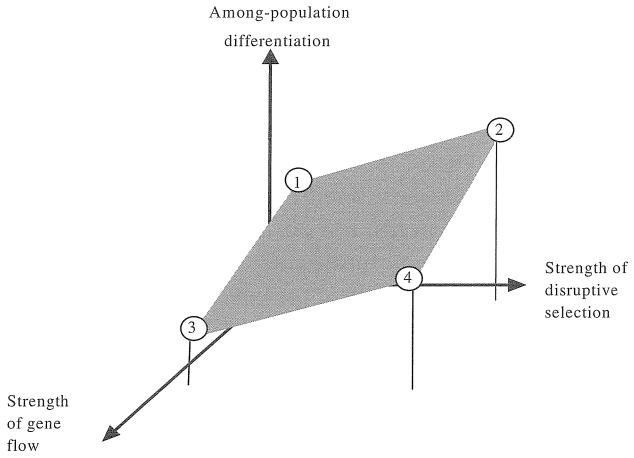

Figure 7. Schematic illustration of the expected among-population differentiation as a result of interaction between gene flow and disruptive natural selection. Random mating and a certain stability of the selective environmental neighbourhoods over generations are assumed.

not be at position 1, which is the first expectation.

Pinus sylvestris in the northern part of its distribution in Sweden (latitude 61–68°) shows a steep clinal variation in tree survival (Fig. 8, EICHE 1966). For trials in this part of Scandinavia transfers in northern direction and to a higher altitude cause an increased plant mortality compared to the mortality of the local population. Contrary to this, transfers southwards and downwards increase the survival with approximately 10 percentage units per degree of transfer or a transfer of 300 meters in elevation (ERIKSSON *et al.* 1980). This simple relationship between geographic origin and survival in *Pinus sylvestris* must be attributed to the topography of Sweden with a mountain change mainly in north-south direction, which means that the climate gradually changes from south to north and from low elevation to higher elevation.

There is also a large within-population variation in survival (ERIKSSON *et al.* 1980). At different latitudes an adaptation to a growth period that matches the climate at this latitude is taking place. This matching is to the largest extent exerted by the photoperiodic conditions at different latitudes. The night length triggers the onset of hardening. With a too late onset of hardening the trees suffer from severe frost damage. The observation on survival in *Pinus sylvestris* fits the anticipation that the among-population variation in survival would be large since a strong disruptive selection counteracts the swamping of pollen into several different SENs (position 4 in Fig. 7). The disruptive selection thus causes a clinal variation of the means of the populations. The large variation within the populations could mainly be explained by swamping of pollen among populations (position 4 in Fig. 9). This figure is a mirror image of Fig. 7. Thus, when there is a high value in Fig. 7, there is a low value in Fig. 9, and *vice versa*.

In Sweden south of latitude 59°, there is good survival among all *Pinus sylvestris* populations (JOHNSSON 1971). With respect to survival, there is probably just one SEN in this part of the country, which explains the absence of variation in this trait (below point 3 in Fig. 7). In contrast, there is a clinal variation in stem volume all over Sweden (JOHNSSON 1971, ERIKSSON *et al.* 1980). Again, there is an adaptation going on of the duration of the growth period to different latitudes. The night length triggers the growth cessation. Less total growth is accomplished during a short growth period

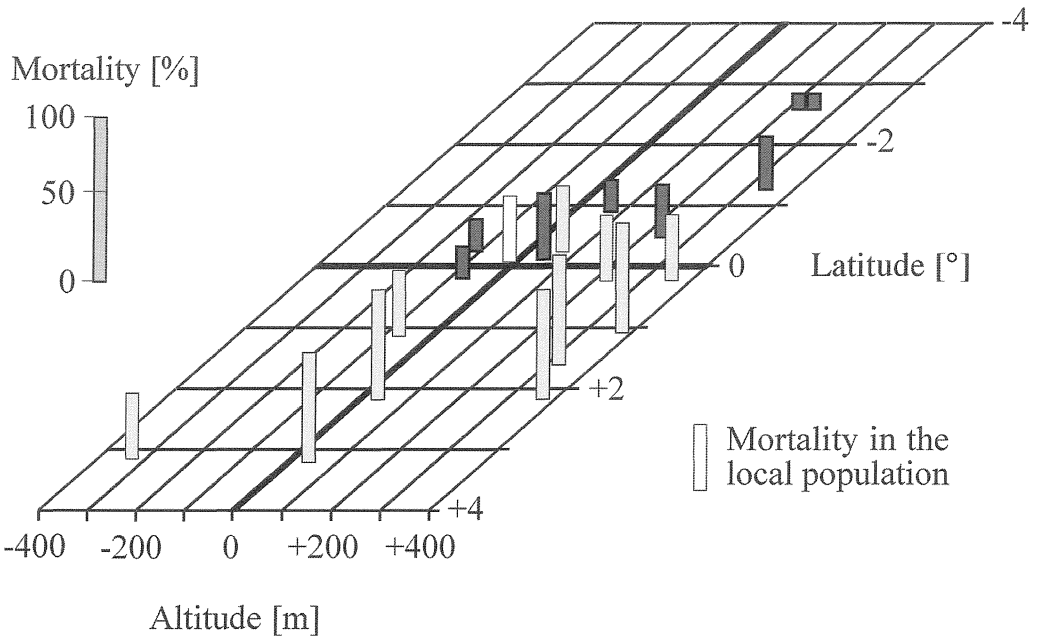

Figure 8. The percentage of plant mortality between the years 1956–1973 of the populations tested in a field trial at Nordanås Lat. 64° and 400 m asl. The populations showing a lower mortality than the local population are shown by black columns. The locations of the populations in the grid indicates the transfer made in latitudinal and altitudinal direction. Columns below the bold horizontal line were moved in northern direction while those above this line were transferred in southern direction. Columns to the right of the bold diagonal axis were transferred to a higher elevation and those to the left to a lower elevation.

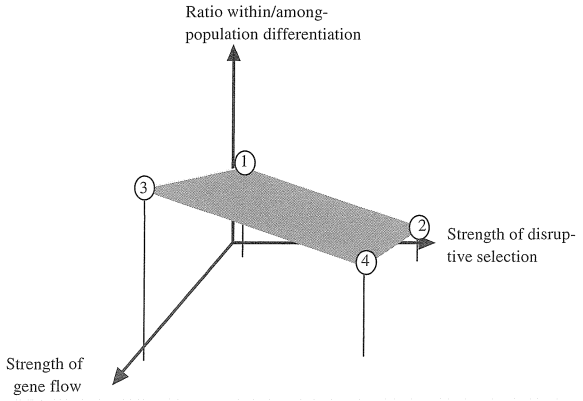

Figure 9. Schematic illustration of the expected within- to among-population variation. Random mating and a certain stability of the selective environmental neighbourhoods over generations are assumed.

than during a long growth pe riod. This means that for growth we have different SENs not only in northern Sweden but also in the southern.

In *Picea abies* steep clines have been reported for bud flushing (ERIKSSON *et al.* 1974), bud set (HOLZER 1966, DORMLING 1979) and stem volume (PERSSON & PERSSON 1992), as well as large within-population variation for growth and growth rhythm traits (EKBERG *et al.* 1985). It is probable that there are several SENs for each of

these traits. In spite of a large gene flow among SENs the disruptive selection has caused a population differentiation (point 4 in Fig. 7) accompanied by a large within-population variation (point 4 in Fig. 9). The latter must be attributed to pollen swamping.

Several conifer species from North America show similar results to those found for *Picea abies* and *Pinus sylvestris* from the old world (CONKLE 1973, REHFELDT 1970, 1980, 1982, CAMPBELL 1979). There are a few exceptions, *Pinus monticola* Dougl. ex D. Don and *P. resinosa* Ait. being two of them. In *P. monticola* most genetic variation in growth and phenology traits remains within populations and only small differences are found among populations, though with a sharp boarder between south and north in central Oregon (CAMPBELL & SUGANO 1989 and literature cited). Similarly, there is some variation among populations of *P. resinosa* without a clear clinal pattern of variation (OVERTON & JOHNSON 1985 and literature cited). These species are expected to have a position close to point 4 in Fig. 7 but *P. resinosa* is close to point 3 and *P. monticola* is located close to point 3 on the line 3–4 in Fig. 7, and point 3 in Fig. 9.

As regards *P. monticola*, CAMPBELL and SUGANO (1989) discussed a few evolutionary explanations for the limited among-population variation for a large area of its distribution in north-west U.S.A. They rule out phenotypic plasticity as the major reason for this since the large genetic variation among trees should contradict such an explanation. Another possible explanation is that the species does not experience the environment as heterogeneous; with the terminology of BRANDON (1990) there is just one single SEN for the species north of central Oregon (position 3 in Fig. 7). In support of this it seems to occupy a specific habitat and is evidently out-competed by other species in other habitats. Still another explanation would be that the environment changes strongly over time preventing any stabilizing selection within certain areas. SENs thus should vary strongly over time. However, it remains to be proven why *P. monticola* experiences the environment as very variable while other conifers experience the same environment as stable. SENs thus should vary strongly over time. However, it remains to be proven why *P. monticola* experiences the environment as very variable while other conifers experience the same environment as stable.

P. resinosa grows pre-dominantly on xeric sites in a band approximately 700 kilometers wide from south-eastern Manitoba and Minnesota to the Atlantic coast. The east-west distribution of the species covers different climatic zones which presumably should give rise to several SENs. As a consequence of that one would expect adaptation to these different SENs as long as large random mating populations exist. It has repeatedly been reported that it has low genetic diversity (see MOSSELER 1995) and the species does not suffer from much inbreeding depression (FOWLER 1964, 1965). This has given rise to speculations that the species has passed through several bottlenecks after the last glaciation and in this way lost genetic diversity. However, as stated by OVERTON and JOHNSON (1985) *P. resinosa* is not devoid of genetic variation in quantitative traits but the variation is less than in other comparable pine species. One possible explanation for the limited genetic variation is that SENs of this species do not differ as much as for other species. The fact that *P. resinosa* predominantly grows on xeric sites (MOSSELER 1995) supports that view. The larger genetic variation in *P. strobus* L. from the same area may be attributed to its occupation of a wider range of site conditions

Table 2. Schematic presentation of the among-population differentiation and within-population variation in non-random-mating populations at different combinations of strong and weak disruptive among-population selection at limited or substantial gene flow, respectively. The more * signs the more differentiation or variation are expected.

	Weak disruptive selection		Strong disruptive selection	
	Limited gene flow	Substantial gene flow	Limited gene flow	Substantial gene flow
	Among-population variation			
Genetic drift	**	_	***	**
	Within-population variation			
	**	***	**	**

and accompanying larger differences in its SENs.

In agreement with the observations by KARHU *et al.* (1996), many analyses of isoenzyme variation revealed a much larger within-population variation than among-population variation in *Betula pendula* Roth., *Picea abies, Pinus sylvestris* and *Pseudotsuga menziesii* [Mirb.] Franco (*e.g.*, BERGMANN 1973, GULLBERG *et al.* 1985, WHEELER & GURIES 1982, RUSANEN pers. comm.). Sometimes clines are observed (BERGMANN 1978, LAGERCRANTZ & RYMAN 1990) but never as steep as for metric traits. The clinal variation observed for some isoenzyme markers is most likely a consequence of isolation by distance (WRIGHT 1943) and not a consequence of disruptive selection since it takes long time to reach the same frequency of a new mutant allele in all subpopulations. The absence of marker differences among *Pinus resinosa* populations summarized by MOSSELER (1995) must also be attributed to absence of different SENs for these markers. The lack of within-population variation of biochemical markers (FOWLER & MORRIS 1977, MOSSELER 1995) contrasts with the observations in most other conifer species. This lack of genetic variation lends support to the bottleneck hypothesis put forward by FOWLER and MORRIS (1977).

Species with non-random mating populations

Many Noble Hardwoods (TUROK *et al.* 1996) and other non-climax tree species are rarely random mating owing to their scattered distribution. Some of them have pollen vectors that fly over short distances only, which would promote adaptation to local conditions. However, pollinating insects may mediate effective gene flow over surprisingly large distances (> 1,000 m) as was the case for *Tilia cordata* Mill. (FROMM 2001). Studies of gene flow in *Sorbus torminalis* (L.) Crantz have indicated that gene flow is not as restricted as may be expected from the scattered distribution of this species (ODDOU-MURATORIO *et al.* 2001). The observation of an estimated gene flow at 2.3

migrants per generation in *Acer platanoides* L. (ERIKSSON *et al.* 2003) lends further support to a rejection of the *a priori* assumption that gene flow in scattered and rarely occurring species is restricted. Loss of additive variance in a population occurs at a rate of $1/(2N_e)$ per generation. Since there are many unlinked loci, a random fixation of alleles over the whole genome will lead to a substantial differentiation among populations. It must be remembered that N_e in many cases is much less than the census number of trees in species with small populations. Populations with less than 10 mating trees for several generations will have a high degree of allele fixation. If an originally large random mating population is split into many small populations, then the frequency of fixation of allele a_1 after many generations of exposure to genetic drift will be proportional to the original frequency of this allele in the original large population. The predictions of the among-population differentiation will in absence of other evolutionary factors be a reflected image of the curve for loss of additive variance due to drift. The joint effects of genetic drift, gene flow and natural selection in presence of genetic drift on among-population and within-population variation are schematically illustrated in Table 2.

The among-population variation will be largest when the disruptive natural selection is strong and the gene flow is absent (cell 3) whereas the smallest among-population variation is expected in the combination strong gene flow and weak disruptive selection. The within-population variation will be highest for the same combination, strong gene flow + weak disruptive selection (cell 6) while the within-population will be smallest in cells 5 and 7, which both have a limited gene flow.

To capture the existing genetic variation in situations when genetic drift is thought to play a major role, the number of populations sampled has to be larger than when random mating prevails. Even if the existing tree population is far from maximum fitness it has proven to be competitive in its present environment and can advantageously be included in the genetic conservation population (ERIKSSON *et al.* 1993). Although methods of genetic conservation are not the subject of this paper, it must be noted that for the long-time survival of the species, putting trees together from different populations into a seed orchard would be the best measure to improve the additive variance in the offspring.

Species with assumed non-random mating populations

MATTILA and VAKKARI (1997) reported on F_{ST} values for isoenzyme variation at 18 loci in *Ulmus laevis* Pall. populations in Finland. Many of the populations are small and the species has its northern margin in Finland. The estimated F_{ST} value amounted to 0.33, which must be regarded as a high value for a wind-pollinated species suggesting that random genetic drift might be a cause of this substantial differentiation.

Like the elm species referred to above, *Acer platanoides* has its northern margin in southern Finland. RUSANEN *et al.* (1996) reported a F_{ST} value of 0.126 for this species based on a study of 14 isoenzyme loci in 14 populations. In a later study of the same species RUSANEN *et al.* (2003.) the F_{ST} value was estimated at 0.097 for populations in the latitudinal range of 49–62°. Compared with the values for *Ulmus laevis*, the impor-

tance of genetic drift seems to be much less. The results suggest that the pollen vector can transport vital pollen over considerable distances.

HÅBJØRG (1978) reported on growth rhythm in *Acer platanoides* and showed that there was a clinal variation for growth cessation. Similarly WESTERGAARD and ERIKSEN (1997) reported on clinal variation in the same species. It is evident that there is some adaptedness to the climatic conditions of this trait in Scandinavia. In the latter report there was one population from latitude 56° that did not differ from a population with its origin more than 4 degrees further north. One possible explanation for the deviating performance of the mentioned populations is genetic drift.

BALIUCKAS *et al.* (2000 and 2001) reported on within-population variation (= family variance components) of juvenile growth and growth rhytm traits in *Fraxinus excelsior* L., *Prunus avium* L., *Fagus sylvatica* L. and *Quercus robur* L. According to expectation, there should be a larger within-population variation in the wind-pollinated *Fraxinus excelsior* (cell 8 or 6 in Table 2) than in the insect-pollinated species *Prunus avium* (cell 7 or 5 in Table 2). For tree height there was a slightly higher variation in *Fraxinus excelsior* than in *Prunus avium* while there was no variation for bud flushing between these two species. It should be noted that there was a large variation among the four populations as regards their within-population variation. Of the typically climax and wind-pollinated species, *Quercus robur* had higher within-population variation than all other species both for height and growth rhythm traits. In an earlier study, BALIUCKAS *et al.* (1999) reported a higher among-population variation for some juvenile traits in *Acer platanoides* than in *Fraxinus excelsior* and *Quercus robur* supporting the hypothesis that insect-pollinated species such as *Acer platanoides* should show larger among-population differentiation than wind-pollinated species.

Conclusions

Since there are so many forest tree species, genetic conservation of individual species is unrealistic, which calls for identification of target species. Their selection may be based on different criteria, of them scientific and socio-economic considerations ought to be favoured. It is useful to consider grouping of species, which share characteristics, and thereby may have common principles for genetic conservation. Before this is done it is important to examine if breeding takes care of the genetic conservation of species included in breeding programmes. Rarity, distribution, behaviour in ecosystem, and pollination vectors are examples of characteristics that can be used for grouping of species for genetic conservation.

The concept of selective environmental neighbourhood (SEN) introduced by Brandon (1990) is useful for an understanding of population differentiation and thereby differences in adaptedness. Within a SEN there are no ranking changes of the genotypes with respect to fitness, conferring a degree of homogeneity to a SEN. A species will have different extensions of the SENs for different traits. For neutral traits there will be just one SEN since the neutral marker does not confer fitness to its carriers. For among-population differentiation there are two requirements: (1) several SENs are required and (2) some stability of SENs over generations is needed. In random mating

populations the among-population differentiation is mainly dependent on the strength of the two counteracting factors, gene flow and disruptive natural selection. In absence of any gene flow from other SENs, the strength of the disruptive selection will be responsible for the among-population differentiation. With a large gene flow among populations in absence of disruptive selection, the expected among-population differentiation is low. When both factors are strong there may be among-population differentiation accompanied with a high ratio of within-/among-population genetic variation. The degree of phenotypic plasticity ought to be larger in long-lived, widely distributed and wind-pollinated species than in annual species. In widely occurring, continuously distributed, and wind-pollinated species the anticipated among- and within-population variation in adaptive and neutral traits was confirmed except for the two pine species, *Pinus monticola* and *P. resinosa.*

At low effective population sizes (< 15) over many generations, a large differentiation may take place, which is partly non-adaptive. Some of the data observed support the expected among-population variation. Some of the studies including isoenzymes suggest that genetic drift has taken place in small isolated marginal populations.

When genetic information is totally absent and sampling for genetic conservation needs to be carried out, the below procedure may be followed:

(1) is it likely that there is random mating? If this is the case then consider;
(2) how many SENs there may be for the target species and how strong the disruptive selection may be among different SENs for traits of adaptive significance;
(3) how strong the gene flow among different SENs may be;
(4) how these two evolutionary factors interact;
(5) if genetic drift is assumed to have taken place, then relatively more samples have to be included to capture the existing genetic variation which is partly non-adaptive.

Acknowledgement

Useful comments to the EUFORGEN version of this paper by Gene Namkoong is gratefully acknowledged. As usual valuable criticism was given by Inger Ekberg. Thanks are due to David Clapham for revision of the English of this paper.

References

BALIUCKAS, V., EKBERG, I., ERIKSSON, G. & NORELL, L. 1999: Genetic variation among and within populations of four Swedish hardwood species assessed in a nursery trial. *Silvae Genet.* **48**: 17–25.

BALIUCKAS, V., LAGERSTRÖM, T. & ERIKSSON, G. 2000: Within-population variation in juvenile growth rhythm and growth in *Fraxinus excelsior* L. and *Prunus avium* L. *For. Genet.* **7**: 183–192.

BALIUCKAS, V., LAGERSTRÖM, T. & ERIKSSON, G. 2001: Within-population variation in juvenile growth rhythm and growth in *Quercus robur* L. and *Fagus sylvatica* L. *For. Genet.* **8**: 259–260.

BERGMANN, F. 1973: Genetische Untersuchungen bei *Picea abies* mit Hilfe der Isoenzym-Identifizierung III. Geographische Variation an 2 Esterase- und Leucine-Aminopeptidase-

Loci in der schwedischen Fichtenpopulation. *Silvae Genet.* **22**: 63–86.
BERGMANN, F. 1978: The allelic distribution at an acid phosphatase locus in Norway spruce (*Picea abies*) along similar climatic gradients. *Theor. Appl. Genet.* **52**: 57–84.
BRANDON, R.N. 1990: Adaptation and Environment. Princeton Univ. Press, Princeton, U.S.A.
CAMPBELL, R.K. 1979: Genecology of Douglas-fir in a watershed in the Oregon cascades. *Ecology* **60**: 1036–1050.
CAMPBELL, R.K. & SUGANO, A.I. 1987: Seed zones and breeding zones for white pine in southwestern Oregon. Res. Pap. PNW–RP–407. Dept. of Agr., For. Serv., Pac. Northwest Res. Sta., Portland, U.S.A.
CLAUSEN, J., KECK, D.D. & HIESEY, W.M. 1940: Experimental studies on the nature of species. I. Effect of varied environments on western North American plants. Carnegie Inst. Publ. **520.**
CONKLE, M. T. 1973: Growth data for 29 years from the California elevational transect study of ponderosa pine. *For. Sci.* **19**: 31–39.
DORMLING, I. 1979: Influence of light intensity and temperature on photoperiodic response of Norway spruce provenances. Pp. 398–408 *in*: Proc. IUFRO Joint Meet. Work Part. Norw. Spruce Prov. & Norw. Spruce Breed., Bucharest, Romania.
EICHE, V. 1955: Spontaneous chlorophyll mutants in Scots pine. *Medd. Stat. Skogsforskningsinstitut* **45**.
EICHE, V. 1966: Cold damage and plant mortality in experimental provenance plantations with Scots pine in northern Sweden. *Stud. For. Suec.* **36**.
EKBERG, I. ERIKSSON, G. & WENG, Y. 1985: Between- and within-population variation in growth rhythm and plant height in four *Picea abies* populations. *Stud. For. Suec.* **167**.
ENDLER, J.A. 1986: Natural Selection in the Wild. Princeton Univ. Press, Princeton, U.S.A.
ERIKSSON, G. 1998: Sampling for genetic resource populations in the absence of genetic knowledge. Pp. 61–75 *in*: Turok, J., Collin, E., Demesure, B., Eriksson, G., Kleinschmit, J., Rusanen, M. & Stephan, B.R. (compilers): EUFORGEN Noble Hardwoods Network, Report of the second meeting, 23–25 March 1997, Sagadi, Estonia. International Plant Genetic Resources Institute, Rome, Italy.
ERIKSSON, G. 2000: Gene conservation of broadleaved tree species – a Scandinavian-Baltic perspective. *Baltic For.* **6**: 48–55.
ERIKSSON, G. 2001: Conservation of noble hardwoods in Europe. *Can. J. For. Res.* **31**: 577–587.
ERIKSSON, G., ANDERSSON, S., EICHE, V., IFVER, J., PERSSON, A. 1980: Severity index and transfer effects on survival and volume production of *Pinus sylvestris* in northern Sweden. *Stud. For. Suec.* **156**.
ERIKSSON, G. & EKBERG, I. 2001: An Introduction to Forest Genetics. Dept. For. Genet., Swedish Agricultural University, Uppsala, Sweden.
ERIKSSON, G., KRUTZSCH, P. & KIELLANDER, K-L. 1974: Reforestation from a genetic point of view. *Sver. Skogsv. Förb. Tidskr.* **72**: 143–152. [In Swedish with English summary].
ERIKSSON, G., NAMKOONG, G. & ROBERDS, J. 1993: Dynamic gene conservation for uncertain futures. *For. Ecol. Manage.* **62**: 15–37.
ERIKSSON, G., BLACK-SAMUELSSON, S., JENSEN, M., MYKING, T., RUSANEN, M., SKRØPPA, T., VAKKARI, P. & WESTERGAARD, L. 2003: Genetic variability in two tree species – *Acer platanoides* L. and *Betula pendula* Roth. with contrasting life history traits. *Scand. J. For. Res.* **18**: 1-12.
ERWIN, T.L. 1991: An evolutionary basis for conservation strategies. *Science* **253**: 750–752.
FOWLER, D.P. 1964: Effects of inbreeding in red pine, *Pinus resinosa* Ait. *Silvae Genet.* **13**: 170 –177.
FOWLER, D.P. 1965: Effects of inbreeding in red pine, *Pinus resinosa* Ait. II Pollination studies. *Silvae Genet.* **14**: 12–23.
FOWLER, D.P. & MORRIS, R.W. 1977: Genetic diversity in red pine: evidence for low genic diversity. *Can. J. For. Res.* **7**: 343–347.
FRANKEL, O.H. 1983: The place of management in conservation. Pp. 1–14 *in*: Schonewald-Fox, F.M., MacBryde, B. & Thomas, W.L. (eds.): Genetics and Conservation. Benjamin Cum-

mings, Menlo Park, U.S.A.
FRANKEL, O.H. & HAWKES, J.G. (eds.) 1974: Crop Genetic Resources for Today and Tommorrow. Sinauer, Sunderland, U.S.A.
FROMM, M. 2001: Reproduktion einer entomophilen Baumart in geringer Populationsdichte – Das Beispiel der Winterlinde (*Tilia cordata* Mill.). Doctoral dissertation, University of Göttingen, Germany.
GOVINDARAJU, D.R. 1988: Relationship between dispersal ability and levels of gene flow in plants. *Oikos* **52**: 31–35.
GREGORIUS, H.-R. & NAMKOONG, G. 1983: Conditions for protected polymorphism in subdivided plant populations. *Theor. Pop. Biol.* **24**: 252–267.
HÅBJØRG, A. 1978: Photoperiodic ecotypes in Scandinavian trees and shrubs. *Meld Norges Landbruksh.* **57**: 33.
HATTEMER, H.H. 1994: Die genetische Variation und ihre Bedeutung für Wald und Waldbäume. *Schweiz. Zeitschr. Forstw.* **145**: 953–975.
HEDRICK, P.W. 1988: Genetic polymorhpism in heterogenous environments. *Ann. Rev. Ecol. Sust.* **17**: 535–566.
HOLZER, K. 1966: Die Vererbung von physiologischen und morphologischen Eigenschaften bei der Fichte I. Sämlingsuntersuchungen. *Mitt. Forstl. Bundesversuchsanstalt, Mariabrunn* **71**.
JOHNSSON, H. 1971: Proveniensförsök med tall i Sydsverige. *För. Skogsträdsför. Inst. Skogsförb. Årsbok* **1970**. 87–122. [in Swedish].
JONSSON, A. & ERIKSSON, G. 1989: A review of genetic studies of some important traits in the genera *Acer, Fagus, Fraxinus, Prunus, Quercus, Tilia and Ulmus.* Res. Not. 44. Dept. For Genet., Swedish Agricultural University, Uppsala, Sweden.
KARHU, A., HURME, P., KARJALAINEN, M., KARVONEN, P., KÄRKKÄINEN, K., NEALE, D. & SAVOLAINEN, O. 1996: Do molecular markers reflect patterns of differentiation of adaptive traits of conifers? *Theor. Appl. Genet.* **93**: 215–221.
KÄRKKÄINEN, K., KOSKI, V. & SAVOLAINEN, O. 1996: Geographical variation in the inbreeding depression of Scots pine. *Evolution* **50**: 111–119.
KIMURA, M. 1983: The Neutral Theory of Molecular Evolution. Cambridge Univ. Press, Cambridge, New York, New Rochelle, Melbourne, Sydney.
KOSKI, V. 1970: A study of pollen dispersal as a mechanism of gene flow in conifers. *Comm. Inst. For. Fenniae* **70**/4.
LAGERCRANTZ, U. & RYMAN, N. 1990: Genetic structure of Norway spruce (*Picea abies*: concordance of morphological and allozyme variation. *Evolution* **44**: 38–53.
LANDE, R. AND BARROWCLOUGH, G. F. 1987: Effective population size, genetic variation, and their use in population management. Pp. 87–123 *in*: Soulé, M. (ed.): Viable Populations for Conservation. Cambridge Univ. Press, Cambridge, New York, New Rochelle, Melbourne, Sydney.
LANGLET, O. 1936: Studier över tallens fysiologiska variabilitet och dess samband med klimatet. *Medd. Statens Skogsförsöksanstalt* **29**. [in Swedish].
LANGLET, O. 1958: A cline or not a cline – the question of Scots pine. *Silvae Genet.* **8**: 1–36.
MATTILA, A. & VAKKARI, P. 1997: Genetic variation of *Quercus robur* and *Ulmus laevis* in Finland. *In*: Proc. Nord. Meet. For. Genet. Tree Breeders, Estonia, June 3–7, 1996. *Metsanduslikud uurimused* **28**: 63–68.
MOSSELER, A. 1995: Red pine: A model for the loss of genetic diversity in trees. Pp. 359–370 *in*: Baradat, Ph., Adams, W.T. and Müller-Starck, G. (eds.): Population Genetics and Genetic Conservation of Forest Trees. SPB Academic Publishing, Amsterdam, The Netherlands.
NAMKOONG, G. 1986: Genetics and the forest of the future. *Unasylva* **38**: 2–18.
NAMKOONG, G. & GREGORIUS, H.-R. 1985: Conditions for protected polymorphism in subdivided plant populations. 2. Seeds versus pollen migration. *Am. Nat.* **125**: 521–534.
ODDOU-MURATORIO, S., PETIT, R.J., LLE GUERROUE, B. & DEMESURE, B. 2001: Pollen – versus seed-

mediated gene flow in a scattered forest tree species. *Evolution* **55**: 1123–1135.
OVERTON, R.P. & JOHNSON, L. C. 1985: Red pine – potential for genetic improvement and observations on cone and seed production. Pp. 177-188 *in*: Proc. Northeastern forest tree improvement conference, Durnham, U.S.A.
PERSSON, A. & PERSSON, B. 1992: Survival, growth, and quality of Norway spruce (*Picea abies* L. Karst) provenances at the three sites of the IUFRO 1964/68 provenance experiment. *Dept. For. Yield, Swed. Univ. Agr. Sci. Res. Notes* **29**.
PHILLIPS, P.C. 1996: Maintenance of polygenic variation via migration-selection balance under uniform selection. *Evolution* **50**: 1334–1339.
REED, D.H. & FRANKHAM, R. 2001: How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* **55**: 1095-1103.
REHFELDT, G.E. 1970: Genecology of *Larix laricina* (DuRoi) K. Koch in Wisconsin. *Silvae Genet.* **19**: 9–16.
REHFELDT, G.E. 1980: Cold acclimation in populations of *Pinus contorta* from the northern Rocky mountains. *Bot. Gaz.* **14**: 458–463.
REHFELDT, G.E. 1982: Differentiation of *Larix occidentalis* populations from the northern Rocky Mountains. *Silvae Genet.* **31**: 13–19.
RUSANEN, M., MATTILA, A. & VAKKARI, P. 1996: Jalojen lehtipuden geneettinen monimuotoisuussäyliytä ja käytä. *Metsänt. Tied.* **605**: 45–52.
RUSANEN, M., VAKKARI, P. & BLOM, A. 2003: Genetic structure of *Acer platanoides* and *Betula pendula* in northern Europe. *Can. J. For. Res.* **33**: 1110-1115.
SAVOLAINEN, O. 1991: Background pollination in seed orchards. Pp. 6-13 *in*: Proc. Meet. Nord. Group Tree Breed., Sweden, August 1991. Report 10, Dept For. Gen. Plant Phys., Swedish Agricultural University, Umeå, Sweden.
SCHOEN, D.J. & BROWN, A.H.D. 1991: Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 4494–4497.
SOULÉ, M.E. 1991: Conservation: Tactics for a constant crisis. *Science* **253**: 744–750.
TURESSON, G. 1922: The genotypical response of the plant species to the habitat. *Hereditas* **3**: 211–350.
TUROK, J., ERIKSSON, G., KLEINSCHMIT, J. & CANGER, S. (compilers) 1996: EUFORGEN Noble Hardwoods Network, Report of the first meeting, 24–27 March 1996, Escherode, Germany. International Plant Genetic Resources Institute, Rome, Italy.
VARELA, M.C. & ERIKSSON, G. 1995: Multipurpose gene conservation in *Quercus suber* – a Portuguese example. *Silvae Genet.* **44**: 28–37.
WEISER, F. 1995: Beitrag zur Existenz von Ökotypen bei gemeiner Esche (*Fraxinus excelsior*). *Forstarchiv* **66**: 251–257.
WESTERGAARD, L. & NYMANN ERIKSEN, E. 1997: Autumn temperature affects the induction of dormancy in first-year seedlings of *Acer platanoides* L. *Scand. J. For. Res.* **12**: 11–16.
WHEELER, N.C. & GURRIES, R.P. 1982: Population structure, genetic diversity and morphological variation in *Pinus contorta* Dougl. *Can. J. For. Res.* **12**: 595–606.
WRIGHT, S. 1943: Isolation by distance. *Genetics* **28**: 114–138.
YANG, R.-C. & YEH, F. 1995: Patterns of gene flow and geographic structure in *Pinus contorta* Dougl. *For. Genet.* **2**: 65–75.

On the appropriate size of forest genetic resources

H. H. Hattemer

Introduction

The appropriate size of forest genetic resources has very different aspects. For instance, conservation *ex situ* may start out from a sample of the target population. Hence, the quality and size of this sample drawn for the establishment of the first generation sets the frame for the genetic structure of the resource to be conserved *ex situ* during the subsequent generations. The sample represents also the population to be conserved in its initial phase. It is similar to a founder population. In the context of conservation *in situ* it has to be decided, whether it is possible to maintain a population at its given size, whether it should be allowed to grow or whether it should be merged with another appropriate population that is likewise too small for long-term persistence. In the former example, a decision has to be made on the appropriate size of the seed sample drawn from the seed produced by the target population. When making this decision, the expected mortality has to be accounted for. In the latter example the decision has even more long-term implications. It has to consider not only the mortality between the seed and the reproductive stage. The size of the population is appropriate only, if the population is likely to survive one, several or even many cycles of natural regeneration in a given environment. Population size is still an important criterion both in the selection and in the management of populations as forest genetic resources to be conserved dynamically. Almost needless to say, the situation considered in this latter example is far more complex.

Establishment of a resource to be conserved *ex situ*

In this section the individuals that make up the population are in fact seeds. Since not all seeds germinate, it has to be pointed out that an excess of viable seed is required in order to account for incomplete germination and nursery survival. Sampling trees in a population would require a more sophisticated design, because trees eventually possess a spatially inhomogeneous genetic structure.

Th. Geburek & J. Turok (eds.)
Conservation and Management of Forest Genetic Resources in Europe
Arbora Publishers, Zvolen, 2005, 413–434, ISBN 80-967088-1-3

The risk of loss of alleles at a single gene locus

The size of the resource has to be carefully planned in order to minimize the risk of not catching genetic variants in the sample. This type of planning is done under the assumption of random drawings from the homogeneous total seed produced by a population. Since also the seed produced by a forest tree population may possess a spatial structure, all seeds are assumed to be collected, subsequently filled into a big vessel and thoroughly mixed. A sample is then drawn by using a special device.

Catching all alleles at a locus in our sample would be impossible if these alleles would occur only once in heterozygous condition. Any sampling would then imply loss of alleles. We had better considering the case of allele frequencies $\alpha > 1/(2N)$ which comply more with our experience (N denotes population size). Even then we could be sure to catch an allele only if our sample size exceeded the proportion of non-carrier individuals of that allele in the population. Since these seeds are impossible to identify, we run a risk of losing alleles occurring with frequencies α even if our sample size amounts to more than $\alpha \cdot N$. This risk of loss implied by sampling we call L.

Consequently, we want to catch certain alleles at low L or with high confidence $1-L$. It can be concluded from the above that in drawing small samples we are faced with larger L, if we want to catch also rare alleles, *i.e.* if α is smaller.

GREGORIUS (1980) derived critical sample sizes N_c with which we may expect to catch with confidence $1-L$ at least once allelic variants that occur with frequency α in the population. Some pertinent data are compiled in Table 1. The conventional significance levels of 0.05 and 0.01 are taken as levels of tolerated risk of loss.

It can be seen from Table 1 that by drawing larger samples we may expect to catch also less frequent alleles. However, this decrease in α is fast only under smaller sample sizes. For instance, if we want to catch an allele having 10 % frequency with 99 % confidence, it is sufficient to sample 50 individuals. If we desire to catch an allele possessing only half of this frequency with same confidence, we have to sample 150 or three times as many individuals. Finally, if we want to halve α once more, we have to boost N_c to far more than 300. When relaxing the confidence criterion to 95 %, we have to choose a sample size of only 125 individuals in order to expect to catch an allele possessing frequency α = .05. The advantages offered by planning the

Table 1. Frequency α of an allelic variant in the population that we expect to catch with confidence $1-L$ at least once in the sample if the sample size is equal to or larger than a critical size N_c. The data were computed by use of a program written by H.-R. Gregorius.

N_c	Least allele frequency α	
	$1-L = 0.99$	$1-L = 0.95$
50	0.1	0.1
100	0.07	0.06
125	0.06	0.05
150	0.05	0.04
175	0.045	0.035
200	0.039	0.031
225	0.034	0.028
250	0.031	0.026
300	0.027	0.022

sample size in this way are clearly attractive.

Fig. 1 shows the risk of loss of alleles of various frequencies under various sample sizes. It can clearly be seen that for both very small and very large sample sizes the risks of loss are more or less constant, *i.e.* they are close to unity if N_c is too small and close to zero if it is unnecessarily large. In the latter case too much effort is made on a certain resource at the cost of others. However, in the interval of medium sample sizes there is much room for careful planning. For sample sizes between 25 and 300 we observe that $L(N_c, \alpha)$ responds drastically to reductions or increases in N_c. If N_c is determined by the available funds rather than genetic reasoning, these curves are still helpful for judging the quality of the samples drawn in terms of complete conservation of less frequent variants.

The data presented in Table 1 need some further clarification. It is easy to conceive that under given N_c a heterozygous individual added to the sample adds two different alleles, while a homozygous individual adds only one. In order to be on the safe side, Table 1 was, therefore, computed for the somewhat adverse conditions of the genotypic structure encountered under predominant or even complete self-fertilization, *i.e.* complete homozygosity. However, under Hardy-Weinberg (HW) conditions (see p. 150 ff., this volume) the proportion of individuals carrying an allele with frequency α is

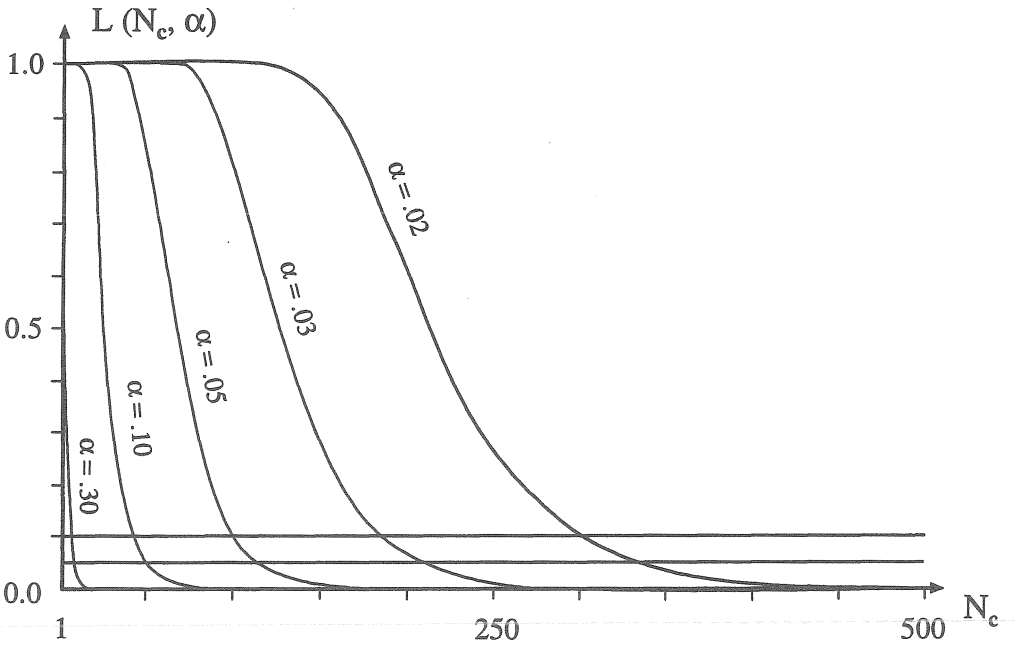

Figure 1. The risk of allele loss $L(N_c, \alpha)$ as a function of N_c for different values of a [from GREGORIUS (1983)].

$$\alpha + 2\alpha(1 - \alpha) = \alpha(2 - \alpha) \quad [1]$$

For small α this expression approaches to 2α, while under complete self-fertilization the proportion of carrier individuals would be only α. This reduces the risk of loss by a factor of roughly ½, which implies lower N_c. Various special cases such as various numbers of equally frequent alleles at a gene locus have been analyzed by GREGORIUS (1983). This publication shows both risks of loss and minimum samples sizes for various allelic structures under panmictic conditions and those under complete self-fertilization. In a paper published in a regular journal, HATTEMER *et al.* (1982) elaborated on this subject in the context of genetic multiplicity in clonal seed orchards.

The considerations of the least frequent allelic variant to be conserved and the required minimum sample sizes derived therefrom are based on mathematical necessities and are largely unspecific under a genetic aspect. The results are valid for just any allele but due to our limited information, no attention is paid to the individual alleles or their relevance for adaptation processes. Neither is any specific attention paid to the conservation of genetic structures as they were. It may only be expected that by appropriate sampling procedures and by keeping sample size large no undue depauperation of genetic structures is induced through sampling. In this situation it is only meaningful to concentrate on the frequency of the least frequent allele, *i.e.* assuming constant adaptive equivalence of all alleles. FINKELDEY and GREGORIUS (1994) have shown that different types of allelic profiles are maintained by different evolutionary processes and, consequently, have an important bearing on adaptability of a resource.

So far only the probability of allele loss was dealt with. However, a risk has one more component besides the probability of an event, *i.e.* the amount of resulting damage implied by this event. It can readily be seen from Fig. 1 that the definition of α largely determines the minimum sample size N_c. Therefore, the desire to catch even rare alleles implies the amount of effort we have to take. However, besides the adaptive value of an allele its rarity as such has some influence on the fitness of a population and its dynamics under processes of adaptation. By disregarding alleles with frequency $< \alpha$ in sampling we neglect their eventual adaptive value. This appears to be justified by two lines of reasoning: Extremely rare variants may represent deleterious mutants with negative adaptive value. Secondly, a variant must be frequent enough to guarantee persistence of the population if all other variants fail to survive under extreme conditions. Since such a variant determines the reproduction-effective size of the population, its absolute number rather than its relative frequency determines the ability of the population to cope with adverse environmental conditions (FINKELDEY 1992). Under less severe conditions it would take the rare variant several generations to attain larger frequency and to become integrated into the adaptive potential of the population. The required number of generations is difficult to specify. Under the simplest model of viability selection, the frequency change of a viability-determining allele depends on selection pressure, its present frequency, its selective value, and its degree of dominance. FINKELDEY (1992) argued a frequency of 2 % as a possible threshold between adaptively significant and deleterious alleles.

Nevertheless, we may conclude that natural reproduction systems close to panmixia represent a protection against the loss of rare alleles in finite populations. However,

some forest trees have been reported to possess an excess of homozygosity which is to be accounted for in the establishment of genetic resources. Eventually there exists a relationship between the fixation coefficient of the seeds and their germination percent; there is clearly no such relationship between the fixation coefficient and the proportion of empty grains.

It may be added that planning the sample size in this way helps to adequately survey and compare populations for making a decision on which one to consider as a genetic resource.

Loss of alleles at one of several gene loci

In the foregoing part of this section we have considered risks 1–L of allele loss at just one arbitrary gene locus. We shall continue this reasoning while looking at several gene loci at a time (NAMKOONG *et al.* 1980, HATTEMER *et al.* 1993, *loc. cit.* chapter 12). We then have to envisage risks of allele loss at any of those loci. If we denote these risks by L_i (with $i = 1, 2, ..., k$), the probability of conserving all alleles at the k gene loci under the assumption of their independence amounts to

$$\prod_{i=1}^{k} (1 - L_i) \qquad [2]$$

and, consequently, the probability of loosing at least one allele of at least one of the gene loci is

$$W = 1 - \prod_{i=1}^{k} (1 - L_i) \qquad [3]$$

The L_i are obviously different, since it can be read from Table 1 that under a given sample size the probability of an allele getting lost increases with its decreasing frequency. Nevertheless, we assume that all L_i are equal to a common L in order to simplify expression [3]:

$$W = 1 - (1 - L)^k \qquad [3a]$$

Some pertinent quantities are shown in Table 2. This table tells us that the combined risk W of loss at any gene locus is, of course, much larger if compared to the data presented in Table 1. For instance, if the risk of loss at one locus $L(1)$ is 5 %, it rises to an $L(20)$ of 64 % if we admit that the loss may happen at any of 20 gene loci. If a risk $L(1)$ amounts to 20 %, it is almost certain with $L(20) = 0.99$ that at least one allele gets lost at some gene locus or gene loci under the given sample size. The conditions at one gene locus serve as the basis of reasoning. The magnitude of the rise in minimum sample size when looking at several gene loci is impressing. Table 2 tells us that, when considering the totality of the genome, losses of genetic variants are as good as inevitable under sampling. The only satisfaction we have in this situation is our ability to evaluate the scale of these losses. Unlike the situation involved in genetic inventories of marker gene loci, in sampling a population for resource conservation *ex situ* we

Table 2. Risks *W* of losing at least one allele at any of *k* gene loci, if the risk *L* of losing allelic variants is constant at these gene loci.

L	Number of independent gene loci				
	2	3	5	10	20
0.001	0.002	0.003	0.005	0.010	0.020
0.01	0.020	0.030	0.049	0.096	0.182
0.02	0.040	0.059	0.096	0.183	0.332
0.05	0.098	0.143	0.226	0.401	0.642
0.1	0.002	0.271	0.410	0.651	0.878
0.2	0.360	0.488	0.672	0.893	0.988

must not confine our thinking to just one gene locus but must focus on the totality of the genome. The long-term effect of catching or losing a genetic variant in a resource is much more important than the mere knowledge, whether it is encountered in a population or not.

Again, these computations disregard the adaptive relevance of gene loci, not to speak of the adaptive value of co-adapted multilocus genotypes.

As has been mentioned above, the statements made on risks of loss and minimum sample sizes refer to strictly random, representative sampling. Any departure of the sampling procedure from randomness leads to additional genetic change and eventually to additional losses of variants beyond those supposedly tolerated under idealized conditions.

KRUSCHE and GEBUREK (1991) developed a procedure for easy computation of lower bounds of risks of loss of rare genetic variants at several independent loci assuming equal frequency of those variants. When assuming complete homozygosity but discriminating between common and rare alleles, these authors found a combined risk of loss

$$W = M \cdot e^{-\alpha N_e} \qquad [4]$$

and

$$N_c = (\ln M - \ln W)/\alpha \qquad [5]$$

when maintaining the notion used in this paper. M is the total number of rare alleles and α is their frequency. The authors also stated that these approximations are conservative, *i.e.* the minimum sample sizes may in fact be smaller.

KRUSCHE and GEBUREK (1991) also presented some minimum sample sizes which deviate only very little from those derived by GREGORIUS (1980). For instance, a total number of M = 1,000 rare alleles with common frequency α = 0.005 and confidence $1-W$ = 0.995, they estimated N_c = 2,435. With reference to adult reproducing trees in a pure stand at a density of 200 trees per hectare, this implies an area of 12 ha. When establishing a forest genetic resource *ex situ*, the number of individuals has to be mul-

tiplied by a factor accounting for the expected reduction of population size during tree life.

It may be added that the risks of loss described in this section apply in every generation. When considering the time scale of conservation efforts it is recommended to account for this fact and apply conservative excess sample sizes.

A comparison of the risks of losing alleles at a single locus and at several loci, respectively, teaches us that sampling for purposes of genetic inventory requires smaller numbers of individuals than those required in designing population size for genetic resource conservation. There exist fundamental differences both in probabilities of merely not detecting or losing genetic variants, respectively, and in damages implied by losses of genetic variants.

Population size and the decay of heterozygosity

According to CROW and KIMURA (1970, *loc. cit.* chapter 3.11) the inbreeding coefficient F of an isolated monoecious random-mating population of finite size increases from generation t–1 to

$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1} \qquad [6a]$$

in generation t in the absence of selection. According to many standard population genetics texts (see also HATTEMER *et al.* 1993, *loc. cit.* chapter 12.4), [6a] can be transformed into the recursion formula

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t \qquad [6b]$$

if $F_0 = 0$. This denotes the inbreeding coefficient t generations after a non-inbred population of unrelated individuals became finite with size N. This increase of F_t with increasing time t is related to the chance-induced shift in gene frequencies towards fixation. This shift implies a reduced potential for the formation of heterozygous genotypes.

CROW and KIMURA (*ibid.*) have also derived the relation of the inbreeding coefficient to the potential for heterozygosity. If we substitute the heterozygosity at time t, *i.e.*

$$H_t = H_0 \cdot (1 - F_t),$$

into the above result, we obtain the corresponding expression for both the decrease of the average heterozygosity from generation t–1 to generation t and the corresponding recursion formula

$$H_t = \left(1 - \frac{1}{2N}\right) \cdot H_{t-1} = \left(1 - \frac{1}{2N}\right)^t \cdot H_0 \qquad [6c]$$

It has to be kept in mind that F measures the probability of the two alleles at an arbitrary gene locus in any individual being identical by descent. This probability is generally less than the proportion of homozygotes at an arbitrary gene locus. Likewise, H measures the potential for heterozygosity, which may exceed the average proportion of heterozygotes at the respective gene locus. F is not due to a departure from HW conditions because of a mating preference among individuals of like genotype. The increase in homozygosity and the implied decay of heterozygosity are rather permanent properties of the population, which could be restored only by either mutation or by immigration, and/or the influx of effective pollen from other populations (CROW & KIMURA *ibid.*). The population would then no longer be isolated. This F and this H have nothing to do with the realized mating system deviating from random mating and cannot be estimated by analyzing genotypic structures at marker gene loci. If not all individuals contribute the same amount of effective gametes to the subsequent generation, then N must be replaced by N_e, the effective population size. Unfortunately, the conditions described in [6] have led to much misunderstanding.

It can be seen from [6c] that in a population of size $N_e = 50$ the decay of heterozygosity amounts to no more than just one percent per generation. Therefore, 50 is the minimum effective population size that keeps the speed of this decay equal to or below 1 %. As has been mentioned above, this quantity is valid only under strong assumptions such as the absence of selection and the randomness of mating between the individuals all of which have to be in their reproductive phase. Nevertheless, a minimum effective size of 50 individuals is often referred to as the basic rule of conservation genetics.

The number 50 (short-term rule) is to prevent a severe reduction of the genetic variation involved in the decay of heterozygosity. It is simplistic for several reasons and might not even be the right order of magnitude in tree populations. For the reasons explained above it is hardly justified to expect that in a population of this size selection may be able to counteract the increase in homozygosity. The respective allele with a selective advantage must be assumed to be rare. Even if it became more frequent due to its positive selective value, the diversity could hardly be raised much. If the selection were directed against homozygotes under over-dominance, its homozygous carriers would be even scarcer so that selection could become effective only in very few members of the population. It may be needless to say that, although selection may retard the decay of heterozygosity, the concept of the basic rule was originally based on the assumed absence of selection.

A population size of 500 (long-term rule) is hoped to represent a safety device and to provide a better chance for the restoration of genetic variation by mutation, a rare event. The decay of heterozygosity would then, of course, be expected to be only one tenth of a percent. However, the expectation of a balance between the expected loss of genetic variants and the origination of new variants through mutation is rather vague. The positive effect of a newly arisen and, therefore, rare allele on diversity and, consequently, on the potential for heterozygosity can be only low even if the allele is not deleterious and has the tendency to increase in frequency.

In any event, the numbers according to the basic rule might represent the absolute minimum of effective population size. The actual number of individuals may be sev-

eral times larger for the various reasons discussed in the context of keeping the risk of chance-induced loss of genetic variants low (see also FRANKEL & SOULÉ 1981, *loc. cit.* chapter 5).

If population size varies among the subsequent generations, the harmonic mean of population size is crucial, *i.e.* generations with smaller size induce an increase in homozygosity that can only partly be compensated by a period of larger population size.

The basic rule of conservation is based on the assumption that there is no selection pressure. It does not account for the fact that environmental change eventually requires adaptation to a new environmental factor. Various genetic aspects of population size were discussed by FRANKEL and SOULÉ (1981, *loc. cit.* chapter 4). It has also to be kept in mind that in plants, particularly in trees, the relationship between population size and inbreeding is not as simple as in animals. This relationship is blurred by various elements of the complex system of reproduction in plant populations.

Concept of the minimum viable population

Definition and applications

The concept of the minimum viable population (MVP) was first explained by SHAFFER (1981). It refers to two aspects of populations (GILPIN & SOULÉ 1986):

- A size of an isolated population of a given species in a given habitat above a certain threshold below which its extinction would become imminent. It is clear that in addition its age structure must be close to equilibrium.
- Critical factors of genetic structure of a population above a level, below which inbreeding depression and the loss of adaptive capacity could become a problem for its continued survival.

This concept was developed in wildlife biology but is still useful also for designing conservation strategies, ecological reserves, and for implementing species recovery plans for the restoration of rare, threatened, and endangered plant species. As SCHWARTZ and BRIGHAM (2003) have demonstrated, plants possess more complex life history attributes than animals such as vertebrates. This brings about special conservation challenges.

Genetic conservation is cost-intensive and therefore requires careful planning. BRIGHAM and THOMSON (2003) and ELDERD *et al.* (2003) review various approaches to a judgment of the viability status of populations and thus species. ELITH and BURGMAN (2003) concentrate on modeling the appropriate habitat for a species in support of population viability analysis, an extension of demographic analysis that focuses on determining whether a species or population has the ability to persist in an environment (MENGES 1991).

The MVP is to ensure at some acceptable level of probability that the isolated population persists in a viable state for a given interval of time. Usually an acceptance level of 95 or 99 % probability is chosen. Some authors (SHAFFER 1981) have specified the respective time span as 1,000 years. This period is long in view of human history

but somewhat short in view of the generation interval of many tree species. Depending on authors, the time period was either specified explicitly excluding stochastic events or despite existing uncertainties (MOSSELER 1998). The definitions stress the time factor but refer to a single target population. Accounting for disturbances and for environmental heterogeneity greatly impedes modeling of population viability (MENGES & QUINTANA-ASCENCIO 2003) but it is just these complex situations where modeling may help to make the appropriate conservation decision.

Even when complying with the various MVP criteria, in view of many unforeseeable events we may only expect to conserve a population by maintaining innate population dynamics that is able to endure a variety of external calamities. In other words, we may keep the probability of extinction at a low level. Making crisis-oriented immediate decisions on long-term viability and conservation of populations, we must tolerate uncertainty and be prepared to make best guesses on available data (MENGES 1991). However, the risk that a given species or population gets lost can easily be defined but hardly ever be estimated or a predicted loss be verified, since it is a very complex task to design models of population dynamics and population genetics possessing sufficient predictive power which account for demographic and environmental change, their complex interactions and the population genetic processes involved in long-time persistence. The analysis of population viability, therefore, includes the study of several scenarios and their comparison. The damage involved in losing a species or population is up to our evaluation (ecological, ethical, aesthetical, economic, *etc.*).

It is true that smaller populations are more likely to get extinct than larger ones. However, designing MVPs is complex, because the risk of extinction still cannot simply be pegged to population size alone. In addition, each situation will have a set of 'minima', depending on the life history of the species or population, the temporal and spatial distribution of its resources, and its level of genetic variation. There is no 'magic number', no single MVP which is universally applicable to all species and populations. According to GILPIN and SOULÉ (1986) the MVP concept is, nevertheless, useful for three reasons:

- The term 'population' reminds us that the single population is the object of study. Furthermore, outlooks on metapopulations are helpful (BRIGHAM & THOMSON 2003);
- The term 'viability' stresses that we are concerned with the persistence of the population over some long time interval;
- The term 'minimum' suggests that there are critical aspects of total size, spatial distribution, and genetic structure, which govern its transition from existence to extinction.

This alone helps to substantially improve the biological basis of conservation decision-making. The MVP concept is important to a number of concerns. For instance, captive populations must be kept above the MVP in order to assure retention of genetic variation and fitness. Ecological reserves must be large enough to provide for the minimum viability at least of their essential species for a long time. In their review of modeling approaches, BRIGHAM and THOMSON (2003) suggest that population viability analysis is less a tool for predicting the fate of a population or species. It

represents rather a tool for asking questions about relative viability such as a ranking of extinction probabilities among populations of a species.

Remaining shortcomings of modelling will certainly be cured in future in order to identify critical conditions of the persistence of populations. Models must be based on a large body of data collected in repeated censuses. If one accepts the results of long-term modelling only with inner reservations, repeated monitoring of the conditions used as input information is advised. This will help to guide decisions on the size of forest genetic resources and their adequate management.

Dynamics of species viability

In this section some sort of a checklist of indicators for assessing the viability or vulnerability of a tree population is presented (GILPIN & SOULÉ 1986). It is not meant to be comprehensive. Most of the relevant biological information is easy to get.

The population phenotype comprises the physical, chemical and general biological properties of the population:

- Physiology (including phenology), morphology and disease resistance.
- Mode of reproduction. Trees are iteroparous perennials, mostly monoecious. Their pollen fertility and seed production vary greatly.
- Adaptedness to the given environment (ability to survive and reproduce).
- Spatial distribution of the trees of the target population. This condition matters greatly for pollen dispersal and mating. Particularly the distribution of the sexual types in dioecious populations is of vital importance for the proportion of pollinated flowers.
- Geographic variation.
- Ability to occupy the given habitat and to migrate into other habitats (seed dispersal).

The environment including 'worst-case' eventualities that affect the viability of the population:

- Habitat quality including herbivore pressure (game browsing, *etc.*), insect gradations and fungal epidemics (BYRNE 2000). As far as phytophagous insects are concerned, biological control possesses particular importance in forest genetic resources (PIMENTEL & HUNTER 1992).
- Habitat quantity available for the target species.
- Both of the above conditions steer the carrying capacity of the environment and, hence, the potential number of individuals. The deterioration of either can extinguish a population.
- Disturbance regime (in view of forest trees of the moderate and boreal zones the frequency of windfall and snow break).
- Population size, structure and fitness are the field of dynamic interactions between a population and its environment.
- Population size (see below).
- The age structure of the individual population determines the fluctuations of population size. In addition, the totality of genetic resources of a species should have

an age structure in order to have reproductive material continuously available for use, and for safety considerations.

- Intrinsic rate of increase r and its spatial variation.
- Sex ratio. In dioecious species the sex ratio is among the determinants of the completeness of pollination and the evenness of seed distribution. It varies also within species (see HATTEMER *et al.* 1993, *loc. cit.* chapter 9).
- Dynamics of spatial distribution (size and distribution of patches).
- Genetic variation (proportion and number of polymorphic gene loci and the numbers of their alleles). Pertinent information exists mainly about marker gene loci, only some of which are of adaptive significance (ZIEHE *et al.* 2000).
- Heterozygosity.
- The two latter terms are often used as synonyms. It is true that without genetic variation there is no heterozygosity. However, heterozygosity is a parameter of the genotypic structure and does not directly measure genic variation. If used as a variation measure for characterizing populations, heterozygosity is meaningless in predominantly self-fertilizing species. Furthermore, the proportion of heterozygotes that we expect under the assumption of random mating has hardly any biological meaning for comparing genic variation of populations, if these differ in the proportion of self-fertilization (KAYA *et al.* 2001).
- Adaptability. Genetic variation is considered to be the sole basis of adaptability. Environmental degradation challenges adaptational processes in tree populations (GEBUREK 2000). However, recent experiments in Norway spruce (*Picea abies* [L.] Karst.) have shown that certain changes at the DNA level during embryo development are another source of adaptability (SKRØPPA & KOHMANN 1997, SKRØPPA & JOHNSEN 2000). This does not require generations of selective frequency change of genetic variants and becomes effective immediately.
- Spatial genetic structure. Restrictions on the transport distances of effective pollen and viable seed imply the development of spatial genetic structures. This is eventually enhanced or blurred by viability selection. The spatial pattern displayed by genetic structures is also reflected by population genetic parameters such as diversity and heterozygosity (GEBUREK *et al.* 1998, DOUNAVI 2000, DOUNAVI *et al.* submitted).

In order to evaluate the relevance of these items for the persistence of a population, we have to rely also on assumptions about unpredictable events, which are only difficult to foresee.

Extinctions

Extinction, *i.e.* the reproductive failure or death of the last individuals of a population or species, is a common phenomenon in evolution. It is only more difficult to study than speciation. However, the speed of species extinction has recently been suspected to exceed the speed of speciation. The gross number of living species is not known. Even more so, the rate of extinction is difficult to assess. Instead, we are able to explore the situation only in a given ecosystem or only in given ecological or taxonomic

groups of living organisms. Most of the extinction events that we have witnessed and chronicled during historical time, particularly on islands, were in conspicuous animals such as birds and all of them were due to human activities (SOULÉ 1983). The reasons were human population growth and the attitude of humans towards other biological species. The same is true with the extinction of tree species. Less conspicuous processes such as the at least local extinction of mycorrhizal symbionts of trees have rarely been monitored.

However, the fragmentation of habitats is easy to observe almost everywhere. It is one of the most important threats to population viability and is hazardous to species persistence. In a comprehensive review, OOSTERMEIJER (2003) addresses the various types of threats to rare plants and advises ways to counteract their genetic effects by management.

The decline of species is due to usually widespread factors. This is much in contrast to the loss of populations, which is often due to stochastic ecological events. Therefore, attempts to rescue a single local population may largely mean waste of time and effort either if the species is abundant elsewhere or if the species as such loses abundance. However, such attempts do make sense if a population is unique in terms of genetics because of its morphology, its adaptedness, or its particular role in the ecosystem, *etc.*

Deterministic extinctions result from relentless, inexorable forces from which there is no hope of escape. An example in animals is the introduction of rats or domestic cats and dogs into islands homing wingless birds. An example in plants is the invasion of a foreign species thrusting aside indigenous species. Examples in trees are large-scale deforestation or desertification. In Mediterranean cypress (*Cupressus sempervirens* L.) it is the prevention of regeneration by the release of domestic goats. The loss of pollinators and seed predation have been made responsible for a threat to rare plants. BRIGHAM (2003) could not find any consistent evidence in support of this hypothesis. The decline of bats and flying dogs as important mutualists of tropical trees due to the destruction of mangrove forests is still expected to have a bearing on pollinator service and seed dispersal in tropical forests.

Stochastic extinctions result from 'normal' random changes or perturbations. They thin a population rather than destroy it. Yet they render it vulnerable. There is no clear-cut difference between the deterministic and stochastic types. For instance, deterministic events such as overharvesting or habitat destruction may reduce a tree population to the point where stochastic extinction is sooner or later inevitable. Stochasticity is largely a problem of small populations. SOULÉ (1983) has listed possible factors contributing to the extinction of local populations.

Subsequently in this section, several aspects of stochasticity as identified by SHAFFER (1981) are briefly described. They impede predictions and make the design of MVPs complicated. It is clear that any stochasticity can only increase the extinction risk. SOULÉ (1983) and MOSSELER (1998) have elaborated on this subject. The biological factors contributing to extinction have been discussed by FRANKEL and SOULÉ (1981, *loc. cit.* chapter 2).

Demographic stochasticity

Non-genetic demographic chance events are considered responsible for the most severe threats to the persistence of small populations such as abnormal mortality rates or reproductive failure.

Demographic stochasticity or random fluctuation of population parameters such as the distribution of age classes or the sex ratio has been observed in animals. Individuals of any age have specific rates of survival and reproduction. The theory of population dynamics tells us that imbalanced age structure is responsible for fluctuations in population size. In many animals and plants the population size as such is subject to fluctuations. In annual plants their amplitude is said to be $\sqrt{K}$, if K denotes the carrying capacity of the environment. In rare long-lived forest trees, fluctuations in age structure are less easy to assay. They have still to be documented over long periods in order to derive appropriate risk warnings. Demographic stochasticity has a direct and strong impact on population size and biological conservation. For this reason, LANDE (1988) advised against planning MVP size on genic grounds alone. Nevertheless, demographic stochasticity does not generally act alone in causing extinction (MENGES 1991).

Environmental stochasticity including catastrophes

Environmental stochasticity is induced by temporal changes of rates of survival and reproduction. In forests of all climatic zones it is due to fires, damage by wind and snow, drought periods, or large-scale cutting of forests. Insect gradations and outbreaks of parasites contribute to this stochasticity. In trees, environmental stochasticity is buffered by life span or in some species by the ability to regenerate from root sprouts. The same is true for the potential for self-fertilization. The latter is, of course a two-edged sword with tradeoffs for population viability considered over short versus long time periods (MENGES 1991). In some species-rich tropical forests many key species such as gap species occupy narrow niches, which come and go. Consequently, their mobile mutualists such as pollinators or seed dispersers are subject to patchy distribution and environmental stochasticity. The seed dispersers of several European tree species such as the yew (*Taxus baccata* L.) and members of the rose family are common birds. However, the establishment of saplings is strongly affected by the local density of the deer population.

Genetic stochasticity

The main source of genetic stochasticity is finite population size. Chance-induced changes in gene frequencies are related to drift processes in finite populations, eventually due to the bottleneck or founder effects. The smaller the population size, the more powerful are these processes. In the extreme case they lead to the decay of genetic variation. If this decay is only local, lost genetic variants may be replaced by influx of

external effective pollen or seed migration from genetically differentiated conspecific neighbour populations. The degree of fragmentation is, therefore, important.

Various forms of genetic deterioration are part of the ubiquitous genetic dynamics of populations of predominantly outbreeding species. They are usually considered separately as inbreeding depression and the decay of genetic variation. The two topics are in fact closely related as was pointed out in the contribution on genetic equilibria and population size (see p. 150 ff., this volume). Increased probability of genetic identity by descent leads to increased homozygosity or loss of heterozygosity. Increased homozygosity increases the frequency of deleterious recessive alleles becoming expressed. Unfortunately, the genetic load is population-specific. In any event, selection cannot weed out those genes because of the overriding effect of small population size. This holds particularly true in the case of rapid reduction of population size. The inbreeding, which is enforced by small population size, is due to the lack of unrelated mating partners rather than mating preferences for relatives leading to homozygote excess relative to HW structure. The variation in this type of inbreeding to be expected in patchy tree populations adds to the genetic stochasticity if inbreeding depression occurs. Fine-scale analysis of spatial genetic structures will clarify this in the near future. In any event, the problem is to keep inbreeding below a threshold, above which fitness relentlessly declines and below which fitness can be maintained. Relevant tools are the maintenance of large size and density of the target population.

Genetic stochasticity has been proven experimentally in caged insect populations. MORAN and HOPPER (1983), MCCLENAGHAN and BEAUCHAMP (1986), and BILLINGTON (1991) have presented data on the relationship between the size of natural tree populations and their amount of genetic variation. However, the establishment of a clear relationship of this type takes many generations until the results of multiple bottlenecks become manifest in variation, because in every generation with finite size a renewed size-dependent risk exists that genetic variants get lost (SAVOLAINEN & KUITTINEN 2000). This relationship is only expected to exist under the assumption of equal initial variation and continuously small size of the reproduction-efficient population in the past. Last but not least, the risk of loss of alleles is determined not only by the size of the isolated population but also the frequency structure of alleles. BARRETT and KOHN (1991) have pointed out that the relationship becomes clearer if variation is measured by parameters that are sensitive to the number of alleles. The rare alleles get lost first but this is not reflected by all population genetic parameters. BRIGHAM (2003) discussed the complex relationship between the size of populations and their genetic variation on the basis of data collected in both trees and annuals.

In central European yew (*T. baccata*), for instance, some small extant stands have turned out to be genetically more variable than stands of tenfold size (CAO 2002). This is presumably due to the few generations since the 16th century when wood of this species was widely used for the manufacture of archers' bows. It has also turned out in this species (CAO 2002) that some small populations among a moderate number of populations surveyed were polymorphic at more gene loci and they were more representative of the whole set of populations studied. This may have made them eligible for the selection as genetic resources. For their management as genetic resources some augmentation or their merger with other small populations *ex situ* is advised.

Finally, we must not forget about the fact that few populations are completely isolated (see p. 150 ff., this volume). The influx of external effective pollen may have a twofold effect. Firstly, it eventually greatly enlarges the reproduction-effective population size. Secondly, depending on the location of donor populations and their genetic structure, the changing direction of the winds, which prevail in given reproductive periods, might have a stochastic effect on the genetic structure of the progeny generation. LEINEMANN & HATTEMER (submitted) have found a markedly high proportion of external effective pollen in a small and supposedly isolated population of European yew.

Although empirical results in forest trees do not always confirm a negative correlation between population size on the one hand and genetic stochasticity and variation on the other, there is no doubt about the basic effect of random genetic drift. This does not only represent a mathematical necessity but has also been verified in caged insect populations of prescribed size and by the monomorphic structure of notoriously rare endemic plant species (see BARRETT and KOHN 1991 for a review).

GILPIN and SOULÉ (1986) have analyzed how the decay of one of the factors addressed in this section may aggravate the effect of one or several others and thus enlarge the risk of extinction (see also SOULÉ 1983). For instance, low population density may reduce both mating success and, consequently, population size due to obstacles to cross pollination as well as genotypic population structure due to a raised proportion of self-fertilization. Small population size implies the loss of genetic variants and this may cause greater uniformity of the nutrient substrate of phytophagous insects. This raises the risk of insect gradations and reduces the viability of the population.

In plants possessing a system of incompatibility the loss of *S*-alleles represents a problem. Active *S*-loci prevent self-fertilization and part of the matings among other relatives and, hence, counteract the chances for inbreeding. In small or fragmented populations also *S*-alleles get at least locally lost despite their frequency-dependent advantage in mating. We must expect an increasing amount of pollen eliminated with decreasing number of *S*-alleles. For instance, if n denotes the number of alleles at a gametophytically active *S*-locus as it exists in the rose family, under the assumption of random pollination in a population with equal frequencies of the *S*-alleles, a proportion of $(2/n)$ of the produced pollen does not become effective. It can be derived that the minimum number of *S*-alleles is three. If 50 alleles exist, the negligible amount of 4% of the pollen will be eliminated. This proportion increases to 20 % if the number of alleles has gone down to only 10. It is true that several woody plants are able to mitigate this adverse effect of inbreeding-avoidance, because some genetically incompatible pollen becomes able to slowly penetrate stigma and style tissue if normally compatible pollen is absent (STERN 1963, HUMMEL *et al.* 1982). The level of this pseudocompatibility in a species is crucial so that generalization is hardly possible.

BRIGHAM and THOMSON (2003) discuss several ways of integrating genetics into population viability analysis. They consider the importance of population size mainly for inbreeding depression and genetic diversity and presume that in some species it may not be necessary to include these aspects of genetics in models of population viability and its development. It may be hoped that in future it will be possible to clarify this question on the basis of experimental evidence on the presence and frequency

structure of certain adaptive genetic variants.

It must be added that the extinction event itself is of less interest than the preceding process of decline of a species or population. Some time before the disappearance of its last representatives the population must have reached the status of being more or less bound to become extinct. This instant is difficult to determine. Active measures of conservation must be taken up much before that time.

Approaching adequate size and structure of genetic resources

After having considered various threats to the persistence of populations and species we ought to find a very general guideline for adequate population size. This is difficult to accomplish by mathematical methods alone. Therefore, the problem of determining MVP size has been tackled by simulation studies aimed at minimizing the probability of extinction of populations of given size during certain time spans. These studies integrate ecological parameters, age structure, life history traits, information on reproductive biology, symbionts, pests and diseases, and catastrophes. Much unlike the more widespread commercial species, the problem in rare species is the general lack of knowledge on these conditions. We must not forget that the predictions derived therefrom in actual fact are the results of empirical studies of artificial data based on numerous assumptions about essentially stochastic events. It can hardly ever been tested, to which degree the totality of these assumptions is valid.

Also predictions based solely on standard population genetic models rather than ecological parameters and chance events imply a certain body of assumptions about unknown facts. In order to overcome these shortcomings, population genetic models must be improved by implementing the results of experimental studies on genetic variation, on genetic load and on reproduction systems of the respective species and populations under various conditions of human-made stress.

Note that it is the size of the population rather than area size that matters. In receding species occurring in mixed forests not only the two measures differ widely. In these species the population density becomes important in view of the ease of pollen transport between reproducing trees. Any excess of self-fertilization due to restricted mating contact between individual trees would increase homozygosity at least at the seed stage. A question in nature reserves is whether the given area is large enough to support sufficiently many trees of a certain receding species, which occurs with notoriously low population density. For receding species in a neotropical most forest, HUBBELL and FOSTER (1986) recommended that the large areas, which carry sufficiently many trees of those species, possess minimal perimeter-to-area ratios. In order to maintain mating contact, this advice is worth considering also in temperate forests. It warrants greater efficiency of conservation of notoriously restricted numbers of tree individuals.

The reliable design of the size and the structure of genetic resources must, therefore, fully integrate ecology, demography, genetics, and silviculture. Options for developing solid indicators for appropriate size of forest genetic resources must be based on both theory and experimental studies in populations of varying size. Theory helps

us to understand the problem and to identify gaps in our knowledge. Experimental studies help to fill those gaps and guide our model-building.

Determining the appropriate size of forest genetic resources is hardly separable from the question of design. For instance, a certain aspect of planning the size and structure of genetic resource populations has become known as the SLOSS issue: conserving a single large or several small resources? The latter concept accounts for the special adaptation of plant species to microsites and provides for the support of at least as many species as fewer large reserves of the same total size (JARVINEN 1982). Conservation of populations *ex situ* is certainly safer and genetically advantageous by maintaining metapopulation structure allowing for selective differentiation of several small resources connected by gene flow. More isolated populations are expected to have lower survival probabilities than partially connected populations in a metapopulation. Gene flow is an effective means of buffering genetic stochasticity in the member populations. FINKELDEY and GREGORIUS (1994) have advocated a differentiated deme structure of a forest genetic resource to be conserved *ex situ*. On the basis of ideas developed by NAMKOONG (1983) these authors discussed the advantages of establishing small populations that are genetically differentiated and exposed to heterogeneous environmental conditions. In the context of conservation *in situ*, the problem of artificial subdivision of resources hardly arises, because nobody would ever intentionally disrupt a large population. In forest genetic resources *ex situ* subdivision reduces the risk that the population gets lost due to a natural catastrophe or building operation.

Considering the totality of eventual genetic resources of three pine species occurring in Korea, KIM *et al.* (1994) studied the increase in the encountered multilocus multiplicity at 18 enzyme gene loci by increasing the number of sampled populations and numbers of individuals per population. Under sampling an adequate number of individual trees, the captured multiplicity did no longer react to further increasing the number of sampled populations beyond a certain point. The flattening-out of these curves depends on the genetic differentiation of those populations and their variation. It is shaped by variability of the markers employed. However, the approach taken in this study is very helpful in making decisions on how to allocate conservation efforts on various populations making up the total genetic resources of the respective species.

It is clear that the maintenance of a resource population must comply with the twin goals of sustainable forestry. Ecological sustainability may be attained through maintaining population viability. Economic sustainability requires silvicultural treatment. This has a variety of genetic implications in connection with management operations such as logging (WICKNESWARI & BOYLE 2000) and regeneration. Conservation *ex situ* comprises silvicultural operations such as seed collection, seed storage, planting stock production, and outplanting (FINKELDEY & ZIEHE 2004). Various types and degrees of genetic change due to these operations have been reported (GEBUREK & THURNER 1993, SAVOLAINEN & KÄRKKÄINEN 1992).

It is also clear that the MVP debate must focus on a given target population. If the conservation goal is to protect a maximum number of species and communities in a nature reserve, the MVP of the species requiring most space governs the size of genetic reserves. Conversely, populations occurring with less than 500 individuals do hardly

justify efforts for their long-term conservation except in conjunction with supporting measures. Viability analysis can be used for both estimating the MVP of a certain species but also for estimating the extinction time for given populations of varying size (cf. MENGES 1991 for examples). It is most urgently needed for species that are subject to environmental stochasticity and natural catastrophes.

MOSSELER (1998) reports on an empirical approach, *i.e.* the long-term observation of replicated populations of varied size (1 to 500 trees) kept in isolation. Monitoring relevant demographic and population genetic parameters might help to find out something on widespread tree species such as *Picea glauca* (Moench) Voss and *Pinus strobus* L. in Canada. This approach may not be practicable in countries with higher density of the human population and in species with a more complicated genetic system but, nevertheless, represents an interesting approach.

Last but not least, large population size may be looked at as a provision against the effect of too much effective foreign pollen. As was pointed out in the contribution on population size, a small population in the close vicinity of several large ones must be expected to be flooded by their pollen. A large population produces more pollen so that its own pollen is expected to prevail in local fertilization processes. Short distance to large allochthonous populations is, therefore, all but ideal. The possible danger through pollen contamination from domestic varieties such as in the rose family has raised concern among some forest geneticists.

Conclusions

As a first approximation to the appropriate size of forest genetic resources, numbers between 500 and several thousand trees may serve the purpose, if the resources are not completely isolated from influx of external effective pollen. They should preferably be conserved in a network of partially isolated populations.
Further recommendations include the following:

- The populations must be large enough so that demographic population structure is not likely to develop stochasticities implying the risk of strong reduction of fertility or the risk of extinction due to demographic accidents. These risks depend very much on the biology of the respective species.
- The risk of chance-induced losses of genetic variants must not exceed a tolerable limit, *i.e.* a 5 % risk of losing variants with more than 2 % frequency.
- The minimum population size derived from considerations of avoiding loss of genetic variants exceeds those required when following the basic rule. Consequently, a rise of more than one per cent of the inbreeding coefficient need not be expected during foreseeable time, if the tolerated risk given above is not exceeded. Keeping a large minimum population in order to avoid losses of genetic variants directly counteracts the increase of the inbreeding coefficient but implies fewer assumptions and does presumably serve the purpose better.
- Some extra size should be added as a safety measure against destruction through conversion of forest to other types of land use (building operations, food production).

- It is strongly recommended to initiate empirical studies on the development of populations of varied size in representative tree species simultaneously with provisional steps taken in resource conservation.
- Still today, not all managers consider genetics an important factor in planning conservation measures. This is partly due to ignorance in a field of biology felt to be difficult. Regrettably enough, it is also due to prejudice (FRANKEL & SOULÉ 1981, *loc. cit.* chapter 5). In accordance with LANDE (1988) it remains, therefore, to be seen whether conservation efforts will increase fast enough in relation to the rate of destruction to preserve much of the natural diversity that existed up to the 19th century.

Acknowledgements

The author is heavily indebted to Dr. M. Ziehe for highly valuable discussions. The technical help of Dr. Heide Glock and Mrs. Sabine Fehrmann is gratefully acknowledged.

References

BARRETT, S.C.H. & KOHN, J.R 1991: Genetic and evolutionary consequences of small population size in plants: implications for conservation. Pp. 3–30 *in:* Falk, D.A. & Holsinger, K.E. (eds.): Genetics and Conservation of Rare Plants. Oxford Univ. Press, New York, U.S.A.

BILLINGTON, H.L. 1991: Effect of population size on genetic variation in a dioecious conifer. *Conserv. Biol.* **5**: 115–119.

BRIGHAM, C.A. 2003: Factors affecting persistence in formerly common and historically rare plants. Pp. 59–97 *in:* Brigham, C.A. & Schwartz, M.W. (eds.): Population Viability in Plants. Springer Verlag, Berlin, Germany.

BRIGHAM, C.A. & THOMSON, D.M. 2003: Approaches to modeling population viability in plants: an overview. Pp. 145–171 *in:* Brigham, C.A. & Schwartz, M.W. (eds.): Population Viability in Plants. Springer Verlag, Berlin, Germany.

BYRNE, M. 2000: Disease threats and the conservation genetics of forest trees. Pp. 159 –166 *in:* Young, A., Boshier, D. & Boyle, T. (eds.): Forest Conservation Genetics. Principles and Practice. CABI, Oxon, United Kingdom.

CAO, C. 2002: Untersuchungen zur genetischen Variation und zum Genfluß bei der Eibe (*Taxus baccata* L.). Master's thesis, Faculty of Forestry and Forest Ecology, University of Göttingen, Germany.

CROW, J.F. & KIMURA, M. 1970: An Introduction to Population Genetics Theory. Harper and Row, New York, U.S.A.

DOUNAVI, K.D. 2000: Familienstrukturen in Buchenbeständen (*Fagus sylvatica* L.). Doctoral dissertation, University of Göttingen, Germany. http://wedoc.sub.gwdg.de/diss/2000/dounavi.

DOUNAVI, K.D., ZIEHE, M. & HATTEMER, H.H. Spatial patterns of genetic variation within three beech stands (*Fagus sylvatica* L.) of high morphological variability (Submitted).

ELDERD, B.D., SHAHANI, P. & DOAK, D.F. 2003: The problems and potential of count-based population viability analysis. Pp. 173–202 *in:* Brigham, C.A. & Schwartz, M.W. (eds.): Population Viability in Plants. Springer Verlag, Berlin, Germany.

ELITH, J. & BURGMAN, M.A. 2003: Habitat models for population viability analysis. Pp. 203–235 *in:* Brigham, C.A. & Schwartz, M.W. (eds.): Population Viability in Plants. Springer Verlag,

Berlin, Germany.
FINKELDEY, R. 1992: Auswahlkriterien und Anlage genetischer Ressourcen bei der Fichte (*Picea abies* [L.] Karst.). *Forstarchiv* **63**: 25–32.
FINKELDEY, R. & GREGORIUS, H.-R. 1994: Genetic resources: selection criteria and design. Pp. 322–347 *in:* Kim, Z.S. & Hattemer, H.H. (eds.): Conservation and Manipulation of Genetic Resources in Forestry. Kwang Moon Kag, Seoul, Korea.
FINKELDEY, R. & ZIEHE, M. 2003: Genetic implications of silvicultural regimes. DYGEN Conference, Strasbourg. *For. Ecol. Manage.* **197**: 231–244.
FRANKEL, O.H. & SOULÉ, M.E. 1981: Conservation and Evolution. Cambridge University Press, Cambridge, United Kingdom.
GEBUREK, TH. 2000: Effects of environmental pollution on the genetics of forest trees. Pp. 135–157 *in:* Young, A., Boshier, D. & Boyle, T. (eds.): Forest Conservation Genetics. Principles and Practice. CABI, Oxon, United Kingdom.
GEBUREK, TH., MOTTINGER-KROUPA, S., MORGANTE, M. & BURG, K. 1998: Genetic variation of Norway spruce (*Picea abies* [L.] Karst.) populations in Austria. II. Microspatial patterns derived from nuclear sequence tagged microsatellite sites. *For. Genet.* **5**: 231–237.
GEBUREK, TH. & THURNER, G. 1993: Verändert sich der Genpool von Waldbeständen durch forstwirtschaftliche Maßnahmen? *Cbl. ges. Forstw.* **110**: 49–62.
GILPIN, M.E. & SOULÉ, M.E. 1986: Minimum viable populations: processes of species extinction. Pp. 19–34 *in:* Soulé, M.E. (ed.): Conservation Biology. The Science of Scarcity and Diversity. Sinauer, Sunderland, U.S.A.
GREGORIUS, H.-R. 1980: The probability of loosing an allele when diploid genotypes are sampled. *Biometrics* **36**: 643–652.
GREGORIUS, H.-R. 1983: Klonanzahl in Samenplantagen und genetische Vielfalt. Pp. 58–62 *in*: Verh. 2. Arbeitstagung Forum Genetik – Wald – Forstwirtschaft, Göttingen, Germany.
HATTEMER, H.H., BERGMANN, F. & ZIEHE, M. 1993: Einführung in die Genetik für Studierende der Forstwissenschaft. J.D. Sauerländer's Verlag, Frankfurt a. M., Germany.
HATTEMER, H.H., GREGORIUS, H.R., ZIEHE, M. & MÜLLER-STARCK, G. 1982: Klonanzahl forstlicher Samenplantagen und genetische Vielfalt. *Allg. Forst- u. J. Ztg.* **153**: 183–191.
HUBBELL, S.P. & FOSTER, R.B. 1986: Commonness and rarity in a neotropical forest: implications for tropical tree conservation. Pp. 205–231 *in:* Soulé, M.E. (ed.): Conservation Biology. The Science of Scarcity and Diversity. Sinauer, Sunderland, U.S.A.
HUMMEL, R.L., ASCHER, P.D. & PELLETT, H.M. 1982: Genetics of self-incompatibility in redosier dogwood. *J. Hered.* **73:** 308–309.
JARVINEN, O. 1982: Conservation of endangered plant populations: single large or several small reserves? *Oikos* **38**: 301–307.
KAYA, Z., TOLUN, A.A., ÇENGEL, B.N., VELIOĞLU, E. & TOLUN, G.G. 2001: The pattern of genetic variation in *Pinus nigra* subspecies *pallasiana* natural populations from the Kazdağı and Bolkar Mountains, Turkey: Implications for *in situ* gene conservation. Pp. 293–305 *in*: Müller-Starck, G. & Schubert, R. (eds.): Genetic Response of Forest Systems to Changing Environmental Conditions. Kluwer Academic Publishers, Dordrecht, The Netherlands.
KIM, Z.S., YI, C.H. & LEE, S.W. 1994: Genetic variation and sampling strategy for conservation in *Pinus* species. Pp. 294–321 *in:* Kim, Z.S. & Hattemer, H.H. (eds.): Conservation and Manipulation of Genetic Resources in Forestry. Kwang Moon Kag, Seoul, Korea.
KRUSCHE, D. & GEBUREK, TH. 1991: Conservation of forest gene resources as related to sample size. *For. Ecol. Manage.* **40**: 145–150.
LANDE, R. 1988: Genetics and demography in biological conservation. *Science* **241**: 1455–1460.
LEINEMANN, L. & HATTEMER, H.H.: Genetic variation and mating pattern in a stand of yew (*Taxus baccata* L.) (Submitted).
MCCLENAGHAN, L.R. & BEAUCHAMP, A.C. 1986: Low genic differentiation among isolated populations of the Californian fan palm *(Washingtonia filifera). Evolution* **40**: 315–322.

MENGES, E.S. 1991: The application of minimum viable population theory to plants. Pp. 45–61 in: Falk, D.A. & Holsinger, K.E. (eds.): Genetics and Conservation of Rare Plants. Oxford Univ. Press, New York, U.S.A.

MENGES, E.S. & QUINTANA-ASCENCIO, P.F. 2003: Modeling the effects of disturbance, spatial variation, and environmental heterogeneity on population viability in plants. Pp. 289–311 *in:* Brigham, C.A. & Schwartz, M.W. (eds.): Population Viability in Plants. Springer Verlag, Berlin, Germany.

MORAN, G.F. & HOPPER, S.D. 1983: Genetic diversity and the insular population structure of the rare granite rock species, *Eucalyptus caesia* Benth. *Austral. J. Bot.* **31**: 161–172.

MOSSELER, A. 1998: Minimum viable population size and the conservation of forest genetic resources. Pp. 191–205 *in:* Puri, S. (ed.): Tree Improvement. Applied Research and Technology Transfer. Science Publishers, Enfield, U.S.A.

NAMKOONG, G. 1983: Preserving natural diversity. Pp. 317–334 *in:* Schonewald-Cox, C.M., Chambers, S.M., MacBryde, B. & Thomas, L. (eds.): Genetics and Conservation. A Reference for Managing Wild Animal and Plant Populations. Benjamin Cummings Publ., London, United Kingdom.

NAMKOONG, G., BARNES, R.D. & BURLEY, J. 1980: A Philosophy of Breeding Strategy for Tropical Forest Trees. Commonwealth Forestry Institute, Oxford, United Kingdom.

OOSTERMEIJER, J.G.B. 2003: Threats to rare plant persistence. Pp. 17–58 *in*: Brigham, C.A. & Schwartz, M.W. (eds.): Population Viability in Plants. Springer Verlag, Berlin, Germany.

PIMENTEL, D. & HUNTER, M.S. 1992: Status of the biological control of pests. *Evol. Tr. Pl.* **6**: 11 –15.

SAVOLAINEN, O. & KÄRKKÄINEN, K. 1992: Effect of forest management on gene pools. Pp. 329 –345 *in:* Adams, W.T., Strauss, S.H., Copes, D.L. & Griffin, A.R. (eds.): Population Genetics of Forest Trees. Kluwer Academic Publishers, Dordrecht, The Netherlands.

SAVOLAINEN, O. & KUITTINEN, H. 2000: Small population processes. Pp. 91–100 *in*: Young, A., Boshier, D. & Boyle, T. (eds.): Forest Conservation Genetics. Principles and Practice. CABI, Oxon, United Kingdom.

SCHWARTZ, M.W. & BRIGHAM, C.A. 2003: Why plant population viability assessment? Pp. 3–15 *in:* Brigham, C.A. & Schwartz, M.W. (eds.): Population Viability in Plants. Springer Verlag, Berlin, Germany.

SHAFFER, M.L. 1981: Minimum population sizes for species conservation. *BioScience* **31**: 131 –134.

SKRØPPA, T. & JOHNSEN, Ø. 2000: Patterns of adaptive genetic variation in forest tree species: the reproductive environment as an evolutionary force in *Picea abies*. Pp. 49–58 *in*: Mátyás, C. (ed.): Genetics and Sustainability. Kluwer Academic Publishers, Dordrecht, The Netherlands.

SKRØPPA, T. & KOHMANN, K. 1997: Adaptation to local conditions after one generation in Norway spruce. *For. Genet.* **4**: 165–171.

SOULÉ, M.E. 1983: What do we really know about extinction? Pp. 111–124 *in*: Schonewald-Cox, C.M., Chambers, S.M., MacBryde, B. & Thomas, L. (eds.): Genetics and Conservation. A Reference for Managing Wild Animal and Plant Populations. Benjamin Cummings Publ., London, United Kingdom.

STERN, K. 1963: Versuche über die Selbststerilität bei der Sandbirke. *Silvae Genet.* **12**: 80–82.

WICKNESWARI, R. & BOYLE, T.J. 2000: Effects of logging and other forms of harvesting on genetic diversity in humid tropical forests. Pp. 115–122 *in:* Young, A., Boshier, D. & Boyle, T. (eds.): Forest Conservation Genetics. Principles and Practice. CABI, Oxon, United Kingdom.

ZIEHE, M., HATTEMER, H.H., MÜLLER-STARCK, R. & MÜLLER-STARCK, G. 2000: Genetic structures as indicators for adaptation and adaptational potentials. Pp. 75–89 *in*: Mátyás, C. (ed.): Forest Genetics and Sustainability. Kluwer Academic Publishers, Dordrecht, The Netherlands.