STRATEGIES FOR IMPROVEMENT OF FOREST TREES

Edited by G.C. Douglas

STRATEGIES FOR IMPROVEMENT OF FOREST TREES *Edited by G.C. Douglas*

Strategies for Improvement of Forest Tree Species

Edited by

G. C. Douglas

Teagasc, Kinsealy Research Centre Malahide Rd., Dublin 17, Ireland

COFORD National Council for Forest Research and Development

COFORD National Council for Forest Research and Development
Agriculture Building
University College Dublin
Belfield, Dublin 4
Ireland
Tel: + 353 1 7067700
Fax: + 353 1 7061180
email: info@COFORD.ie

First published in 1999 by, COFORD, National Council for Forest Research and Development, University College Dublin, Belfield, Dublin 4, Ireland.

ISBN 1 902696 02 6

Title: Strategies for Improvement of Forest Tree Species

Editor: G. C. Douglas

Publisher: COFORD National Council for Forest Research and Development

Citation: Strategies for Improvement of Forest Tree Species.
Douglas, G. C. (ed.), COFORD, Dublin, Ireland

Design: G. C. Douglas

Strategies for Improvement of Forest Tree Species

Proceedings of the Symposium on Forest Genetics

March 9th, 1998

At Swift Theatre, Arts Building
Trinity College Dublin, Ireland

Scientific Committee and Organisers
Dr. F. Lefort, Dr. G. C. Douglas
Dr. D. Kelly, Dr. T. Hodkinson

Teagasc, Agriculture and Food Development Authority, Kinsealy Research Centre
and the
Department of Botany, Trinity College Dublin.

Use of microsatellite markers in management of conifer forest species

Craig S. Echt

Research Plant Molecular Geneticist, United States Department of Agriculture Forest Service, North Central Forest Experiment Station, Forestry Sciences Laboratory, Rhinelander, Wisconsin, USA 54501, Tel: (715) 362-1114 Fax: (715) 362-1166, e-mail : cecht/nc_rh@fs.fed.us, or cecht@newnorth.net

Gene diversity / Pinus / cp SSR / microsatellite / conifer / molecular markers

WHY MICROSATELLITE MARKERS?

Within the past ten years a new class of genetic marker[1] has risen to prominence as the tool of choice for many geneticists. Microsatellite DNAs, or simple sequence repeats (SSRs), were first characterized as highly informative genetic markers in humans (Weber and May, 1990; Litt and Luty, 1990), and have since been found in practically all organisms. SSRs are short stretches of repeated DNA sequence, where the repeats are composed of a particular 2 to 5 DNA nucleotide motif, and are organized in tandem arrays of 4 to >40 repeat units. Compared to other types of DNA sequence variation, such as single nucleotide substitutions, SSRs tend to be more mutable. The mutations are generally manifested as gain or loss of individual repeat units, referred to as step-wise mutations, thus giving rise to a series of alleles that have discrete size differences (Figure 1). Within most species' genomes, that is, within their complement of chromosomal DNA, SSRs are abundant, widely distributed, and frequently surrounded by non-repetitive, unique DNA sequences. It is this latter property that allows use of the PCR (polymerase chain reaction) technique to amplify a specific single SSR locus from a small sample of DNA. Following PCR amplification, alleles of the locus can be scored as phenotypically codominant 'bands on a gel'. By use of PCR methodology many informative SSR loci can be genotyped very quickly from small amounts of tissue, and for comparatively reasonable costs.

Reasonable, that is, once the all-important oligonucleotide primer pairs needed to drive (or prime) the PCR assay have been designed, synthesized and evaluated for a sufficient number of SSR loci. It is these development costs, combined with the attendant DNA cloning,

[1] A genetic marker is any quantifiable physical character that distinguishes, or marks, underlying genetic differences among individuals. It is often encountered as a particular enzyme or DNA fragment having a defined position on an electrophoretic gel. Any given 'band on a gel' generally does not gain marker status, however, unless two or more forms (alleles) of it exist at a single chromosomal location (locus). At their most basic level, alleles are simply the DNA sequence variants found at a locus.

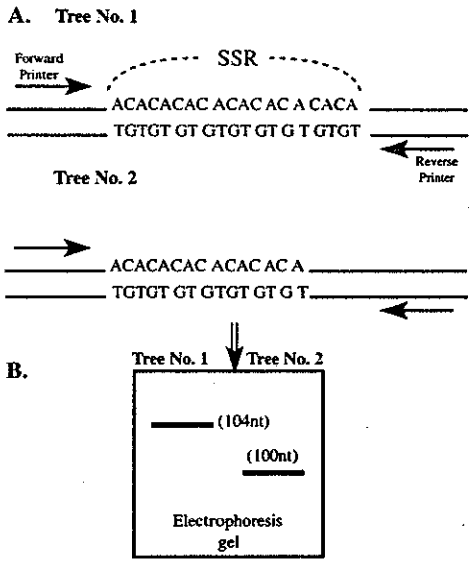

Fig 1. Schematic representation of the genetic basis of phenotypic differences of two Simple Sequence Repeat alleles (SSR) by their length, one has 100 nucleotides (nt) and the other 104 nt, each representing two different trees.
A) The PCR primers are oligonucleotides, shown as arrows indicating the direction of DNA priming, each bind to unique DNA sequences which flank the SSR, here shown as a repeat of the unit (AC)*n* (not to scale). During consecutive cycles of PCR only the locus, defined by the far ends of the two priming sites, is amplified (molecularly cloned) so that the resulting DNA fragments can be scored by electrophoresis on a gel. After electrophoresis, the sequences of amplified DNA are represented by two lines; they are a section of the two strands of DNA on a chromosome.
B) Diagram of the allelic markers from trees 1 and 2 separated by electrophoresis through a polyacrylamide gel.

selection, and sequencing costs, that can make starting an SSR program an expensive and time consuming venture. Yet, after the initial work is done and suitable primer pairs are available, most labs can do SSR genotyping as easily, or often more easily, as any other type of marker genotyping. If just a dozen or so markers are needed, as is the case for many population genetic or DNA fingerprinting applications, then marker development costs are not all that great for most labs with molecular biology experience. When hundreds of markers are needed, however, as for genome mapping or marker aided selection projects, then the marker development phase is best left to those institutions with proportionally higher budgets. Even so, development costs have been declining as improved methodologies and commercial services become available. Another factor that can ameliorate the expense of development is the identification of SSR primers that can be used among related species.

While it is not a one-size-fits-all marker system, the SSR approach is particularly well suited for population genetics, DNA fingerprinting (e.g. paternity analysis), and marker aided selection and breeding applications. Because of their high variability, SSR analysis may be one of the few methods able to reveal quantifiable levels of variation in genetically depleted species. For large-scale genotyping needs, as are frequently found in marker aided tree improvement programs, the ability to automate SSR marker analysis using pipetting robots, DNA sequencing machines and fluorescent marker detection methods, can provide substantial time savings. Automation also minimizes errors in sample handling and genotype scoring. For SSR markers with unit allelic size differences of 3 to 5 nucleotides, high resolution agarose gels can be used for genotyping, provided that suitable methods are employed for accurate allele scoring.

The biological function of microsatellite DNA is unknown. In nearly all studies, SSRs are used as neutral markers, so biological function is irrelevant (and assumed not to exist). However, certain SSR repeats are found in functional genes, and expansion of some trinucleotide

repeats causes a number of neurological disorders in humans (Margolis *et al.*,, 1997; Mitas 1997; O'Donovan and Guy, 1997; Reddy and Housman, 1997). An intriguing possibility it that SSRs may be engines of quantitative variation (Kashi et al., 1997; King et al., 1997), and thus may play a role in adaptive variability in trees.

AVAILABILITY OF SSR MARKERS IN FOREST TREE SPECIES.

SSR sequences are abundant in conifer genomes (Echt and May-Marquardt, 1997; Pfeiffer *et al.*,1997; Smith and Devey, 1994), as they are in plant genomes in general (Morgante and Olivieri, 1993; Wang *et al.*, 1994). For use in forest genetics, SSR markers are available from *Fraxinus excelsior* (Lefort *et al.*, 1998), *Pinus contorta* (Hicks *et al.*, 1998), *Picea abies* (Pfeiffer *et al.*, 1997), *Picea sitchensis* (van de Ven and McNicol, 1996), *Pinus radiata* (Smith and Devey, 1994; Fisher *et al.*, 1996), *Pinus strobus* (Echt *et al.*, 1996), *Pinus sylvestris* (Soranzo *et al.*, 1998; Kostia *et al.*, 1995), and *Quercus* species (Dow *et al.*, 1995; Isagi and Suhandono, 1997; Steinkellner *et al.*, 1997). In addition, there are on-going SSR marker development efforts in various laboratories for *Larix spp, Pinus taeda, Pinus halepensis, Pinus monticola , Pseudotsuga menziezii* and *Thuja plicata.* The greatest number of SSR markers currently available for a pine species is, 37 for *Pinus radiata* (G. Moran, CSIRO, personal communication), and for a spruce species is 122 for *Picea abies* (M. Morgante, Univ. Udine, personal communication). Within the next year or two there should be hundreds of more markers available for *Pinus radiata* and *P. taeda.* The Dendrome web site (http://dendrome.ucdavis.edu/Data/primer.html maintains information about SSR markers in trees. Some of the pine SSR primers are also commercially available.

USING SSR PRIMERS AMONG DIFFERENT SPECIES

The ability to use SSR primer pairs among similar species can reduce marker development costs, and provide useful information on comparative linkage relationships between species. We have recently tested twenty three *Pinus strobus* SSR primer pairs in 11 other conifer species. In general, most $(AC)_n$ SSR loci that are polymorphic in *P. strobus* could be amplified in other soft pines, i.e. *P. lambertiana* and *P. cembra*, but not in hard pines, i.e. *P. brutia, P. halepensis, P. leucodermis, P. pinaster, P. radiata, P. resinosa, P. taeda,* or in other conifers, i.e. *Picea glauca* or *Pseudotsuga menziesii* (Figure 2). Of the few loci that were monomorphic in *P. strobus* (RPS3, RPS61, RPS105, RPS150, RPS152, RPS160), all could be amplified in some portion of the hard pines tested, and only one (RPS150) could be amplified in non-pine conifers. All such markers were also monomorphic in the other species, but four of them (RPS61, RPS105, RPS152, RPS160) showed size variation among species. Markers that are monomorphic within species, but polymorphic between species could be useful as species-specific markers in taxonomic studies, or for identifying inter-species hybrids.

SSR primer pairs for six polymorphic *Pinus radiata* $(AG)_n$ loci (PR4.6, PR9.3, NZPR1, NZPR4, NZPR5, NZPR6) were also tested in other hard pines as well as in some soft pines and other conifers. None of them amplified loci in *Pinus cembra* or *P. strobus,* nor in *Larix laricina, Picea abies, Picea glauca* or *Pseudotsuga menziesii,* but did amplify loci among certain other hard pines i.e. *P. brutia, P. halepensis, P.*

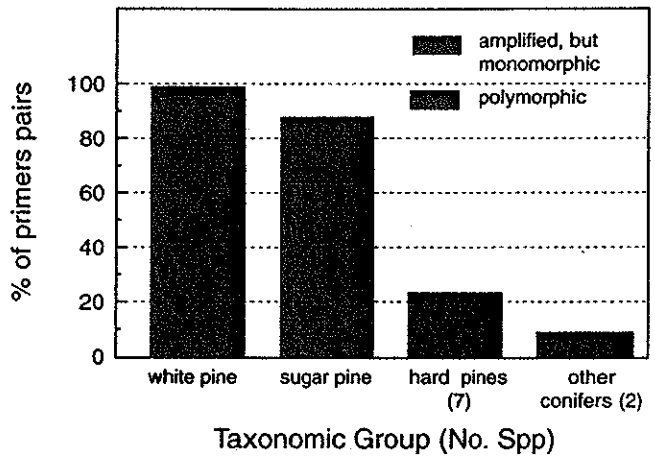

Fig 2. Summary of results using 23 *Pinus strobus* SSR primer pairs on other conifer species. The solid shading represents the proportion of monomorphic markers. The monomorphic markers in *P. strobus* were completely contained within the monomorphic sets of *P. lambertiana* and the 'hard pines'. None of the polymorphic *P. strobus* loci, all of the (AC)n motif, amplified reliably in any 'hard pine' or 'other conifer' species. The 'hard pine' and 'other conifer' species are listed in the text.

Table 1. Gene diversity comparisons of 8 $(AC)_n$ SSR loci between two white pine species, using primers pairs developed from *P. strobus.*

	P. strobus	*P. lambertiana*
n_a*	53	63
H_e	0.61	0.76
F_{it}	0.14	0.40

* n_a = total number of alleles, the number of alleles shared between the species=24
H_e = unbiased estimate of heterozygosity
F_i = total fixation index (the ratio of expected heterozygosity/observed heterozygosity)

leucodermis, P. pinaster, P. resinosa, P. taeda. A generalization can be made that primer pairs for polymorphic dinucleotide microsatellite loci will PCR amplify homologous loci only among members of a closely related taxonomic group, subsection, or in this case, perhaps subgenus.

To test the level of polymorphism of *P. strobus* SSR primer pairs in *P. lambertiana* (sugar pine), another white pine, genotypes were obtained for 8 SSR loci common between the species. For each species, DNA was extracted from bud tissue of 24 individuals distributed throughout the species' natural ranges. *P. lambertiana* grows only in the western coastal states of the USA and Mexico, while *P. strobus* grows only in the eastern part of North America, and the two species do not hybridize. As can be seen from Table 1, *P. lambertiana* had a higher allelic diversity than *P. strobus.* Of the 92 total alleles, only 24 were shared (based on allele sizes) between the species. DNA sequencing may reveal single nucleotide

substitutions (point mutations) in alleles having the same size in the two species, which would reduce the number of shared alleles. The low fixation indices, F_{it}, of both white pine and sugar pine showed that there were more individuals having homozygous loci than would have been expected from random mating throughout the population. Assuming a high level of gene flow and low genetic differentiation among subpopulations this suggests a significant degree of inbreeding. There is additional evidence for this in white pine (see below), but the notably higher value for *P. lambertiana* should be regarded with caution because of the possibility of null alleles at some of the SSR loci.

The phenotype of a null allele is the absence of a PCR amplified fragment. The effect is that loci that are heterozygous for a null allele will be scored as a homozygote for the allele that did amplify, thus inflating the observed proportion of homozygotes. Segregation analysis of megagametophytes from previously genotyped trees can confirm the presence of a null allele(s) at suspect loci.

Use of SSR markers to monitor genetic diversity in managed forests.

The forestry program on the 95,000 ha Menominee Indian Reservation in Wisconsin, USA, is considered by many to be the finest example of sustainable management of eastern white pine on the continent. In collaboration with tribal foresters, genetic diversity was estimated for native, managed, and regenerated local populations to determine the genetic effects of current management strategies. Eight SSR loci were genotyped in 50 individuals from each of 9 populations representing 6 sites distributed within the reservation boundaries. At three sites both the 160 year-old thinned overstory and the naturally regenerated progeny were sampled. At another site only the recently thinned overstory was sampled, as natural regeneration had not yet occurred. The fifth site was representative of old growth, unmanaged, 240 year old trees. The last site was an artificially regenerated plantation established from off-reservation seedlings.

For the eight loci examined, there were an average of 14 alleles per locus. Genetic diversity of all nine populations, when measured as the proportion of heterozygous loci expected based on the observed allele frequencies, was high and did not significantly differ among the populations. For the three sites that were thinned to prepare the stands for shelterwood regeneration, the proportion of alleles that were present in less than 5% of the standing individuals (the low frequency alleles) was not significantly different from the proportion in the sawtimber populations at other sites that were not thinned. These results indicate that current management

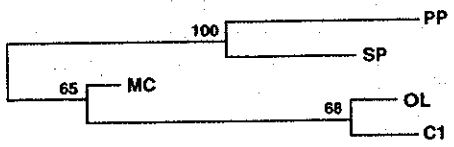

Fig 3. Historical effects of high grading? Genetic distance tree for 5 Menominee white pine sawtimber stands, constructed by neighbor joining analysis using distance values based on a step-wise mutation model for SSR alleles. The numbers at branch nodes are bootstrap values, expressed in percentages, obtained from 1000 replications, and indicate the relative degree of confidence in the topological relationships. Only the MC stand is suspected of being subject to "high grade" timber harvesting during the late 1800's, but was not thinned for recent pine release management. The SP stand is virgin timber, while the C1, OL and PP stands had 25% of sawtimber tree harvested for shelterwood management. Dendrogram branch length is a measure of relative SSR genetic diversity, while branch and node position indicate genetic relatedness.

of the Menominee white pines is preserving native genetic diversity. An earlier study using isozymes came to a similar conclusion for shelterwood management of Douglas fir (Neale, 1985).

There is preliminary evidence, however, that pre-reservation "high-grade" timber harvesting has had genetic consequences at one of the sites (Figure 3). While the practice of harvesting all trees having the highest volume can erode genetic diversity in a stand (see also Buchert *et al.*, 1997), the regenerated progeny at this particular Menominee site appear to have recovered native levels of diversity, perhaps as a result of pollen coming in from neighboring stands.

An unexpected finding was that a common measure of inbreeding, the fixation index, F_{it}, indicated that there were significantly less heterozygous loci than would be expected from random mating among individuals. The high level of gene flow and low level of genetic differentiation observed among populations suggests that the excess of homozygotes resulted from a moderate level of self matings, or matings among relatives. However, an isozyme survey of Quebec *P. strobus* populations (Beaulieu and Simon, 1994; 1995) found the expected frequency of heterozygous loci in mature trees, indicating no excess of inbreeding, while a very recent study of two old-growth *P. strobus* stands in Ontario (Buchert *et al.*, 1997) found a deficit of heterozygotes, similar to the Menominee white pines. The cause of the differences between expected and observed heterozygosities between different *P. strobus* populations is not known.

At one of the 8 SSR loci studied, locus RPS 25b, it appeared that an artificially high number of homozygotes were scored because of the presence of a null allele (or alleles) in most populations. But null alleles cannot account for the excess of homozygotes at most of the other loci, and so the overall heterozygote deficit did not result from genotyping errors.

P. strobus is fairly tolerant to inbreeding more than most other pine species (Fowler 1965). The two other pine species of the American northeast, *P. resinosa* (red pine) and *P. banksiana* (jack pine), are even more tolerant to inbreeding (Fowler, 1965). Perhaps the evolutionary and natural histories of these three pine species has selected against many of the deleterious recessive alleles that are maintained through high outcrossing rates in most other pines. However, Beaulieu and Simon (1995) mentioned the possibility that homozygous loci in two Quebec *P. strobus* populations may be selected against. The important point is that there is no evidence that the levels of inbreeding found in the Menominee white pine forests resulted from genetic mismanagement of the stands, as inbreeding is found in both unmanaged and managed populations, and appears to be a natural characteristic of the species.

Use of chloroplast SSR to study population structure and evolutionary history of conifer forests.

The advantages of using chloroplast microsatellite (cpSSR) markers is well described elsewhere in these proceedings and will not be reiterated here. The special case of cpSSR markers in *P. resinosa* reveal population genetic structuring in a species that is generally devoid of variability for other types of markers. In a study of seven populations of red pine distributed across the species' natural range, cpSSR analysis found genetic differences between all of the populations (Echt *et al.*, 1998) Analysis of the molecular data provided evidence for a population bottleneck, thought to have occurred during the last Ice Age (Fowler and Lester, 1970; Mosseler, 1992). The cpSSR data

also supported the hypothesis made by Mosseler (1992) that red pine exists as a nonequilibrium metapopulation structure, where disjunct populations have limited gene flow, and go through cycles of decline or extinction, and recolonization and establishment. Additional sampling is planned so as to obtain a range-wide pattern of cpSSR diversity, from which lineages of descent, and routes of migration, can be inferred through new methods of computational statistical genetics (B. Epperson, Michigan State U., personal communication). The resulting information will be used for germplasm management, and for increasing genetic diversity, in managed populations.

In conclusion, SSR markers offer great promise for increasing the efficiencies of various forestry programs. For a few key species, sufficient numbers of markers will soon be available to cover most tree improvement or forest genetic management applications. For other species for which SSR markers are not available, sufficient numbers of markers for parentage analysis or population genetic studies (about a dozen) can be obtained with a modest budget and experienced personnel.

ACKNOWLEDGMENTS

Thanks to G.G. Vendramin and C. Dana Nelson for their contributions to the interspecies work, to Dan Pubanz and Dave Congas for their help with the Menominee white pines, and to Paula Marquardt, Julie Hagstrom and Ron Burns of the Rhinelander Forestry Sciences Lab for their expert technical support.

REFERENCES

Beaulieu, J., and Simon, J.P. 1994. Genetic structure and variability in *Pinus strobus* in Quebec. Canadian Journal of Forest Research 24, 1726-1733.

Beaulieu, J., and Simon, J.P. 1995. Mating system in natural populations of eastern white pine in Quebec. Canadian Journal of Forest Research 25, 1697-1703.

Buchert, G.P., Rajora, O.P., Hood, J.V., and Dancik, B.P. 1997. Effects of harvesting on genetic diversity in old growth eastern white pine in Ontario, Canada. Conservation Biology, 11, 747-758.

Dow, B.D., Ashley, M.V., and Howe, H.F. 1995. Characterization of highly variable (GA/CT)n microsatellites in the bur oak, *Quercus macrocarpa*. Theoretical and Applied Genetics 91, 137-141.

Echt, C.S., May-Marquardt, P., Hseih, M., and Zahorchak, R. 1996. Characterization of microsatellite markers in eastern white pine. Genome 39, 1102-1108.

Echt, C.S. and May-Marquardt, P. 1997. Survey of microsatellite DNA in pine. Genome 40, 9-17.

Echt, C.S., L.L. DeVerno, M. Anzidei & G.G. Vendramin. 1998. Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. Molecular Ecology 7, 307-316.

Fisher, P.J., Gardner, R.C., and Richardson, T.E. 1996. Single locus microsatellites isolated using 5' anchored PCR. Nucl. Acids Res. 24, 4369-4372.

Fowler, D.P. 1965. Effects of inbreeding in red pine Pinus resinosa Ait. IV. Comparison with other northeastern *Pinus* species. Silvae Genetica, 14, 76-81.

Fowler, D.P., and Lester, D.T. 1970. The genetics of red pine. U.S.D.A. Forest Service Research Paper, W-8: pp 13.

Hicks, M., Adams, D., O'Keefe, S., Macdonald, E. and Hodgetts, R. 1998. The development of RAPD and microsatellite markers in lodgepole pine (*Pinus contorta* var. latifolia) Genome (in press).

Isagi, Y., and Suhandono, S. 1997. PCR primers amplifying microsatellite loci of *Quercus myrsinifolia* Blume and their conservation between oak species. Molecular Ecology 6, 897-899.

Kashi, Y., King, D., and Soller, M. 1997. Simple sequence repeats as a source of quantitative genetic variation. Trends in Genetics 13, 74-78

King, D.G., Soller, M. and Kashi, Y. 1997. Evolutionary tuning knobs. Endeavour 21, 36-40.

Kostia. S., Varvio, S-., Vakkari, P. and Pulkkinen, P. 1995. Microsatellite sequences in a conifer, *Pinus sylvestris*. Genome 38, 1244-1248.

Lefort, F., Bracket, S. Edwards, K.J. and Douglas, G.C. 1999 Identification and characterisation of microsatellite loci from ash *Fraxinus excelsior* and their conservation in the Oleaceae family. Molec, Ecol. (in press).

Litt, M. and Luty, J.A. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. 44, 397-401.

Margolis, R.L., Abraham, M.R., Gatchell, S.B., Li, S.H., Kidwai, A.S., Breschel, T.S., Stine, O.C., Callahan, C., McInnis, M.G. and Ross, C.A. 1997. cDNAs with long CAG trinucleotide repeats from human brain. Human Genetics 100, 114-122.

Mitas, M. 1997. Trinucleotide repeats associated with human disease. Nucleic Acids Research 25, 2245-2253.

Morgante, M. and Olivieri, A.M. 1993. PCR-amplified microsatellites as markers in plant genetics. The Plant Journal 3, 175-182.

Mosseler, A. 1992. Life history and genetic diversity in red pine: implications for gene conservation in forestry. The Forestry Chronicle, 68, 701-708.

Neale, D.B. 1985. Genetic implications of shelterwood regeneration of Douglas-fir in southwest Oregon. Forest Science 31, 995-1005.

O'Donovan, M.C. and Guy, C. 1997. Brief Research Communication: Bias in the genomic distribution of CAG and CTG trinucleotide repeats. American Journal of Medical Genetics 74, 62-64.

Pfeiffer, A., Olivieri, A.M. and Morgante, M. 1997. Identification and characterization of microsatellites in Norway spruce (Picea abies K). Genome 40, 411-419.

Reddy, P.S. and Housman, D.E. 1997. The complex pathology of trinucleotide repeats. Current Opinion in Cell Biology 9, 364-372.

Smith, D.N., and Devey, M.E. 1994. Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome 37, 977-983.

Soranzo, N., Provan, J., Powell, W. 1998. Characterization of microsatellite loci in *Pinus sylvestris* L. Mol. Ecol. (in press).

Steinkellner, H., Fluch, S., Turetshek, E., Lexer, C., Streiff, R., Kremer, A., Burg, K., and Glössl, J. 1997. Identification and characterization of (GA/CT)n-microsatellite loci from *Quercus petraea*. Plant Molecular Biology 33, 1093-1096.

van de Ven, W.T.G. and McNicol, R.J. 1996. Microsatellites as DNA markers in Sitka spruce. Theoretical and Applied Genetics 93, 613-617.

Wang, Z., J.L. Weber, G. Zhong & S.D. Tanksley, 1994. Survey of plant short tandem repeats. Theor. Appl. Genet. 88, 1-6.

Weber, J.L. and P.E. May 1990. Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. Amer. J. Hum. Genet. 44, 388-396.