Molecular Markers In Conservation Genetics:
Chloroplast DNA Variation In Natural Scottish
*Pinus sylvestris* L.

Thorunn Helgason

PhD. Thesis
University of Edinburgh
1993.
### TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>v</td>
</tr>
<tr>
<td>Declaration</td>
<td>vi</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>vii</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>viii</td>
</tr>
<tr>
<td>Chapter 1</td>
<td></td>
</tr>
<tr>
<td>Introduction to Molecular Genetics</td>
<td>1</td>
</tr>
<tr>
<td>Prologue</td>
<td>1</td>
</tr>
<tr>
<td>Genes and Genome Organisation</td>
<td>2</td>
</tr>
<tr>
<td>DNA Organisation Within The Cell.</td>
<td>3</td>
</tr>
<tr>
<td>DNA variation at the level of the individual</td>
<td>6</td>
</tr>
<tr>
<td>Variation in DNA Sequences.</td>
<td>6</td>
</tr>
<tr>
<td>Rates of DNA Change</td>
<td>7</td>
</tr>
<tr>
<td>Population Genetic Structure.</td>
<td>9</td>
</tr>
<tr>
<td>Genetic Differentiation Among Populations.</td>
<td>12</td>
</tr>
<tr>
<td>The Detection Of Genetic Variation</td>
<td>16</td>
</tr>
<tr>
<td>Phenotypic Variation</td>
<td>16</td>
</tr>
<tr>
<td>Isozyme Variation</td>
<td>18</td>
</tr>
<tr>
<td>Direct Detection of DNA Variation</td>
<td>20</td>
</tr>
<tr>
<td>Restriction Fragment Length Polymorphisms</td>
<td>20</td>
</tr>
<tr>
<td>Polymerase Chain Reaction (PCR) based techniques.</td>
<td>24</td>
</tr>
<tr>
<td>&quot;The Right Tool For The Job&quot;</td>
<td>27</td>
</tr>
<tr>
<td>Conclusions</td>
<td>30</td>
</tr>
<tr>
<td>Chapter 2</td>
<td></td>
</tr>
<tr>
<td>Native Scottish <em>Pinus sylvestris</em> L.: a case study</td>
<td>31</td>
</tr>
<tr>
<td>Prologue</td>
<td>31</td>
</tr>
<tr>
<td>The Natural History of <em>P. sylvestris</em> in Scotland</td>
<td>32</td>
</tr>
<tr>
<td>Taxonomy of Scots pine</td>
<td>32</td>
</tr>
<tr>
<td>The Ecology of Scots Pine.</td>
<td>35</td>
</tr>
<tr>
<td>The Forests of Scotland in a Regional Context.</td>
<td>36</td>
</tr>
<tr>
<td>What Happened to the Forests?</td>
<td>38</td>
</tr>
<tr>
<td>The Pinewood Resource Today</td>
<td>39</td>
</tr>
<tr>
<td>Conservation of the Native pinewoods</td>
<td>40</td>
</tr>
<tr>
<td>The Protection of the Existing Native Pinewoods.</td>
<td>40</td>
</tr>
<tr>
<td>The Extension of Native Pinewoods</td>
<td>42</td>
</tr>
<tr>
<td>Population Structure and Genetics of <em>P. sylvestris</em> in Scotland</td>
<td>44</td>
</tr>
<tr>
<td>The Reconstruction of The Vegetation History of Scottish Pine Populations.</td>
<td>47</td>
</tr>
<tr>
<td>The Traditional Story</td>
<td>47</td>
</tr>
<tr>
<td>Biochemical Markers of Population Differentiation</td>
<td>48</td>
</tr>
<tr>
<td>A Review of the Evidence</td>
<td>49</td>
</tr>
<tr>
<td>Critique of Palynological evidence</td>
<td>49</td>
</tr>
<tr>
<td>Critique of Genetic analyses</td>
<td>52</td>
</tr>
</tbody>
</table>
The Invasion of *P. sylvestris* into Scotland: A Review of the Hypotheses.  
Summary.  

The Alternative Approach: Molecular Markers  
Genetic Conservation in Scots pine.

Chapter 3  
RFLP Analysis of The Chloroplast DNA Of *Pinus sylvestris*.  
Introduction  
RFLP analysis in plant populations.  

DNA Extraction And Manipulation  
Sample Collection and Maintenance  
DNA Extraction Techniques  
Restriction Digests  
Gel Electrophoresis and Southern Blotting.  

Probes  
Probes DNA:Bacteria, probes and vectors.  
Excising probe fragments from plasmid vectors  
Selection and use of Probes.  

Labelling, Hybridisation and Detection  
Labelling  
Hybridisation  
Detection  

Evaluation and Conclusions  
Comparison of DNA Extraction Methods  
Comparison of Labelling and Detection methods  
Conclusions  

Chapter 4  
Chloroplast DNA variation in Scottish populations of *P. sylvestris*.  
Introduction.  

Materials and Methods  
Sites  
Methods  

Results  
Chloroplast genome survey  
Length and Sequence variation in Scottish *P. sylvestris*  
Population survey  

Discussion  
Chloroplast DNA variation in *Pinus*  
RFLP Analysis of *P. sylvestris*.  
CpDNA as a marker in *P. sylvestris*.  
Implications for conservation of Scots pine.  

Conclusions  

Chapter 5  
Population Genetics in Conservation: Conclusions and Recommendations  
Introduction
ABSTRACT

Underlying the diversity of form and function present in the living world is the genetic diversity within and among taxa. The measurement of such diversity using genetic markers is increasingly being used by conservation biologists to improve the evaluation and management of genetic resources. The native Scottish populations of Pinus sylvestris L. were selected as a case study to evaluate the usefulness of molecular markers as a way to distinguish between populations of differing origin. There is some evidence that there are two gene pools present in these populations, possibly of refugial origin and not therefore closely related to European populations. Pollen contamination from the extensive plantations of non-native P. sylvestris therefore presents a potentially serious threat to the integrity of the native gene pool. Population specific genetic markers would allow the origin of the Scottish populations to be identified, and the extent of pollen contamination from non-native sources to be measured. The implementation and monitoring of genetic conservation management would therefore be more effective.

Scottish P. sylvestris populations have been extensively researched, using monoterpane and isozyme markers, but differentiation among populations is too low to identify origins or gene flow. Protocols were developed for the analysis of chloroplast DNA (cpDNA) Restriction Fragment Length Polymorphisms (RFLP) in P. sylvestris to determine whether cpDNA markers could be used for these purposes. Nine populations from throughout the range of pine in Scotland were sampled. Two individuals from each of these populations were surveyed using seven restriction enzymes and 13 probes from P. contorta, a total of 91 probe/enzyme combinations. The cpDNA genome of P. sylvestris was found to be about 119 kilobase pairs in length. 100% of the length of the genome was sampled, and 0.52% of the sequence length. No variation was found in any individual. These results were compared with a survey of P. sylvestris from China, Sweden and Turkey. There is no evidence that the cpDNA genotype of Scottish pine differs in any way from these varieties, suggesting that the cpDNA genome of P. sylvestris is homogeneous over a large part of the species' range. A survey of 191 individuals for 1 probe/enzyme combination revealed one variant individual, which appeared to be heteroplasmic for two cpDNA haplotypes. It was not possible to determine whether this was due to biparental inheritance or somatic mutation within that individual. The implications of these results for further research on the structure and inheritance of cpDNA in gymnosperms, the use of cpDNA as a genetic marker in P. sylvestris, and the conservation of Scottish populations of this species are discussed. Finally, it is suggested that an increased collaboration between molecular biologists and ecologists is the best way to approach studies using sophisticated molecular techniques to measure genetic diversity in natural populations. In this way, the potential of this approach to improve the management of genetic resources can be fully realised.
DECLARATION

I hereby declare that this thesis was composed entirely by myself, except where otherwise acknowledged.

Thorunn Helgason,
November 1993.
ACKNOWLEDGEMENTS

I would like to thank the Science Faculty of The University of Edinburgh for awarding me the Colin and Ethel Gordon Scholarship, and the Royal Society for additional financial support.

I would like to acknowledge the help of the following people, without whom this project would never have been completed:

My supervisor, Richard Ennos for his help, advice and encouragement over the years.

All the landowners, agents, Scottish Natural Heritage and Forestry Authority staff, who gave permission to sample from their native pinewoods, and often provided a great deal of welcome extra help and advice.

Robbie Waugh and his colleagues at the Scottish Crop Research Institute, Dundee, who initially taught me how to do RFLP analysis in 1989.

John Morman and Tang Qian for assistance in the laboratory.

Joan Davidson and Aileen Greig for an endless supply of sparkling glassware and cough sweets.

Fiona Stewart, Dave Curnow and Mike Groves for helpful comments and ideas on parts of the manuscript.

Sheila Wilson for help with typing the references and figure legends, and Dave Haswell and John Anthony for printing the photographs on pages 109-110.

Fiona, Pam, Teresa and all the other staff of the Darwin Library for their efficiency and tolerance while I was writing up.

My parents, for all their support, moral and financial, and for innumerable Sunday dinners.

And finally, all the friends and colleagues that have helped to make my time at the Darwin Building such an entertaining experience.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>βME</td>
<td>2-mercaptoethanol. Sigma M-3148</td>
</tr>
<tr>
<td>(k)bp</td>
<td>(kilo)base pairs.</td>
</tr>
<tr>
<td>BP</td>
<td>Years before present.</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin. Sigma A-7906.</td>
</tr>
<tr>
<td>cpDNA</td>
<td>Chloroplast DNA.</td>
</tr>
<tr>
<td>CTAB</td>
<td>Mixed alkyltri-methylammoniumbromide. Sigma M-7635.</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid.</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediamine-tetracetic acid. Sigma ED2SS.</td>
</tr>
<tr>
<td>EtBr</td>
<td>Ethidium Bromide (2,7-Diamino-10-ethyl-9-phenyl-phenanthridium bromide). Sigma E-8751.</td>
</tr>
<tr>
<td>IAA</td>
<td>Pentan-2-ol. BDH (Merck) 29456.</td>
</tr>
<tr>
<td>mtDNA</td>
<td>Mitochondrial DNA.</td>
</tr>
<tr>
<td>NNR</td>
<td>National Nature Reserve.</td>
</tr>
<tr>
<td>NPGS</td>
<td>Native Pinewood Grant Scheme.</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction.</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene Glycol. Sigma P2139.</td>
</tr>
<tr>
<td>Polyclar®-AT</td>
<td>Polyvinylpyrrolidone insoluble. BDH (Merck) 44201.</td>
</tr>
<tr>
<td>PTE</td>
<td>50mM Tris, 10mM EDTA.</td>
</tr>
<tr>
<td>RAPD</td>
<td>Random Amplified Polymorphic DNA.</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction Fragment Length Polymorphism.</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid.</td>
</tr>
<tr>
<td>SDS</td>
<td>Lauryl Sulphate (Sodium dodecyl sulphate). Sigma L-4509.</td>
</tr>
<tr>
<td>SSC (20×)</td>
<td>Standard Saline Citrate, (3M NaCl, 0.3M Tri-Sodium Citrate).</td>
</tr>
<tr>
<td>SSSI</td>
<td>Site of Special Scientific Interest.</td>
</tr>
<tr>
<td>TAE</td>
<td>0.04M Tris Acetate, 0.001M EDTA.</td>
</tr>
<tr>
<td>TE</td>
<td>10mM Tris, 1mM EDTA.</td>
</tr>
<tr>
<td>Tris</td>
<td>2-amino-2(hydroxymethyl)-1,3,-propandiol. Boehringer 708 976.</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION TO MOLECULAR GENETICS

Prologue

The central aim of conservation biology is to maintain diversity in ecosystems at all levels. Underlying the evolution of diversity of form and function we see in the living world is the genetic diversity that exists within and among taxa. Maintenance of the existing genetic structure and diversity of species has long been acknowledged to be a cornerstone of conservation management (Frankel 1974, Simberloff 1988 and references therein). In addition to representing past evolution, the level of genetic variability influences the evolutionary potential of a species, that is the ability to respond to a changing environment and thus avoid extinction (Frankel and Soulé 1981).

The pattern of variation that exists in a species can be called the population structure. This is defined by Jain (1975) as

"...the totality of ecological and genetic relationships among the member individuals as well as the sub-divisions of a species." (Jain 1975 p16.)

The genetic diversity of a population is an integral part of the population structure of a species, and reflects both evolutionary processes and historical accidents (Avise 1992, Slatkin 1987). Genetic variation arises in populations, and evolutionary (in the loosest sense of the word) processes occur that modify the distribution of that variation within the population as a whole. In order to achieve effective conservation of genetic diversity, a knowledge of the existing genetic structure, and an understanding of how it arose in the population is essential.

The aim of this thesis is to evaluate the use of molecular markers that directly measure variation in DNA sequences, for analysing the population genetic structure of plant populations. Many techniques have been developed to measure variation in natural populations. There is an increasing movement towards the direct measurement of DNA variation, because of the availability of molecular biology techniques. The examples used will therefore come from studies of plant species, with data included from the animal kingdom where comparable data is not available from plants.
Population genetics and the measurement of diversity using genetic markers are now being used by conservation biologists, whether for conservation of genetic resources (e.g. Brown et al. 1989, 1990) or for the monitoring of endangered species (e.g. Falk and Holsinger 1991). The development of molecular markers for the analysis of populations has provided a much larger array of techniques for the analysis of populations (e.g. Hoelzel 1992). In order to use this technology effectively, the processes that create the patterns of diversity that we observe in populations and measure using these markers have to be understood. The behaviour of any genetic marker depends ultimately on the DNA sequence that it represents. Although the genome of an organism consists of this one biochemical compound, it is by no means homogeneous and different kinds of DNA behave in different ways. To fully understand the nature of any genetic marker it is essential to review the structure, organisation and evolution of DNA.

Genes and Genome Organisation

The structure and function of DNA underpins all evolutionary processes. Without the ability of DNA to replicate and its potential for varying, evolution would be impossible. The kinds of variation that can arise in DNA and the rates at which that change occurs influences the mode and tempo of evolution that can occur.

The structure of DNA

All living organisms, except some viruses, carry their hereditary information as deoxyribonucleic acid (DNA). DNA consists of two chains of nucleotides twisted into a double helix. Each nucleotide in the chain is made up of a phosphate group, a deoxyribose sugar and either a purine or a pyrimidine base. The backbone of the double helix is formed by the phosphate and sugar groups, and the two strands are held together by hydrogen bonds between the bases. There are four bases in DNA, two purines, adenine and guanine, and two pyrimidines, thymine and cytosine. Adenine always pairs with thymine, and guanine always pairs with cytosine (Li and Graur 1991). This structure is the key to the ability of DNA to replicate. If the double helix separates, and one strand is used as a template for synthesising a new strand, then the new strand will be an exact complementary copy of the template. Thus, information can be passed unchanged between generations of cells (Goodenough 1984).

Functional and Non-Functional DNA.

To convert the hereditary information carried in the DNA sequence, the DNA is arranged to form genes. Genes are defined as sequences of DNA that have a
specific function. Three types of genes are recognised: protein coding genes, RNA (ribonucleic acid) genes and regulatory genes. Protein coding genes carry the information necessary to build chains of amino acids that form proteins. Converting this information from DNA sequence to protein chain is called translation and transcription. Each of the 20 primary amino acids is coded for by one or more nucleotide triplets called codons. The codons for each amino acid in a polypeptide chain occur in the correct order on the DNA strand and are non-overlapping. All codons code for amino acids except for the "stop" codons which end transcription. RNA genes code for the gene products necessary for protein transcription, the ribosomal and transfer RNAs. Less is known about regulatory genes, but they are thought to be of four types: "switches", or replicator genes, that control whether a gene is transcribed or not, recombinator genes that control site specific recombination during eukaryotic meiosis, centromere genes that control segregation of the chromosome, again during meiosis and finally, attachment sites for other molecules such as hormones or proteins. Many other regulatory gene families may exist that have not yet been identified (Li and Graur 1991, Cavalier-Smith 1985).

The functional genes described above form only a part of the genome of an individual. Because they code for gene products that control physiological processes in the individual, or in some way regulate their operation, they are part of the genotype of the organism that is expressed in its phenotype. The genotype is defined as the genetic or hereditary composition of the organism, whereas the phenotype is the observable properties of the individual, produced by the expression of the genotype under the influence of environmental conditions. The DNA that contains no genetic information will not have any influence on the phenotype but constitutes anything from 30% to nearly 100% of the genome of eukaryotic organisms (Cavalier-Smith 1985).

DNA Organisation Within The Cell.

Most of the genome of plants is located in the nucleus in the form of chromosomes. In addition to this, two organelles, the mitochondria and the chloroplasts, contain DNA that replicates and is inherited independently of the nuclear genome. While only a single copy of the nuclear genome is present in each cell, there are many chloroplasts and mitochondria. The organelle genomes are therefore present in far higher copy numbers than nuclear genes. Each genome carries characteristic gene families, and may have different modes of inheritance, which are described next.
Nuclear DNA

The nucleus of a eukaryotic cell contains the chromosomes that carry most of an organism's functional genes. Nuclear DNA exists as either single copy sequences, where only one copy of that sequence is present in the cell, or as repeat families. Functional eukaryotic genes are mainly present as single copy sequences, but much of the single copy DNA may be non-functional. In humans, for example, an estimated 64% of the genome is single copy, and only 3% of that is transcribed (Hartl 1980, Li and Graur 1991). Most of the chromosomal or nuclear DNA is not, therefore, apparently functional or expressed in the phenotype. The DNA is arranged as chromosomes, which are linear strands of DNA wrapped around protein molecules (mainly histones) to form the DNA-protein complex called chromatin. This is thought to enable long DNA strands to be more easily packed into the nucleus of the cell (Goodenough 1984). The chromosomes of eukaryotes are usually present in homologous pairs i.e. they are diploid, but many plant species are polyploid. In the germ cells formed by meiosis only one copy of each chromosome is present and the cell is haploid.

The inheritance of the diploid nuclear genome is biparental, or Mendelian. An organism receives one haploid set of chromosomes from each parent. When subsequent daughter cells undergo meiosis to form further haploid germ cells, the pairs of homologous chromosomes recombine and exchange segments of chromatin. Thus an individual of the next generation will receive half of its chromosomes from its grandparents, but it will be a random combination of genes from the two.

Mitochondrial DNA

Mitochondrial DNA (mtDNA) is present in all eukaryotes and shows fundamental differences in properties among the major kingdoms of organisms (see reviews by Sederoff 1987 and Gray 1989). Plant mtDNA genes code for polypeptides involved in cell respiration, but up to 95% of mitochondrial polypeptides are coded for by the nuclear genome (Leaver and Gray 1982, Palmer 1990). The structure of plant mtDNA is complex and shows a striking lack of homology between closely related species. The mtDNA of plants has not been as extensively studied as the chloroplast DNA, but large amounts of variation have been reported, even within the same genus. The mtDNA of the genus Brassica, for example, has been characterised in both B. hirta (white mustard) and B. campestris (turnip), and while the former has a unicircular mtDNA genome, B. campestris was found to
consist of three circular mtDNA molecules, one master chromosome, and two smaller ones, created by recombination across a repeat unit (Palmer and Sheilds 1984, Palmer and Herbon 1987). The size of the mtDNA genome has been shown to vary significantly between species within one family, the Cucurbitaceae, with values ranging from 220-1600 Md for the four species examined (Ward et al. 1981). High levels of inter-specific variation in mtDNA has been found in the genus Vaccinium (Haghighi and Hancock 1992). Plant mtDNA shows a 10-fold range in size from 200-2500 kbp, and the bulk of this variation in size appears to be due to increases in the amount of non-coding DNA, which may account for as much as 95-99% of the sequence complexity (that is the "coding potential" of the genome, considered independently of any repeated sequences) (Ward et al. 1981, Palmer 1990).

The inheritance of mitochondrial DNA in the higher plants is predominantly maternal (Conde et al. 1979, Palmer 1985a, Neale and Sederoff 1988, 1989). As the number of species investigated increases, some cases of paternal inheritance and paternal leakage have been reported (Holwerda et al. 1986, Erickson and Kemble 1990, Wagner et al. 1991a).

Chloroplast DNA

The chloroplast DNA (cpDNA) of plants is a circular DNA molecule, ranging from 100-217 kbp in length (Palmer 1985b, 1990). The cpDNA of angiosperms has been extensively researched and the complete sequences of two species, tobacco and rice, have been determined, in addition to the liverwort Marchantia (Palmer 1987, 1990). The cpDNA genome has very little non-coding DNA (Shinozaki et al. 1986, Clegg 1990). Unlike mtDNA, the structure of the cpDNA genome has been investigated in many species of plant, and has been found to be remarkably similar throughout the plant kingdom, with some significant exceptions. In angiosperms, the cpDNA molecule contains an inverted repeat sequence coding for a number of rRNA genes and ranging from 10 to 76kb in length. The variation in size of cpDNA throughout the plant kingdom can largely be explained by variation in the length of this repeat sequence (Palmer 1985b, 1990). The subfamily Papilonoideae of the Fabaceae, including the tribes Viciae (Broad bean and pea), Trifoliæ (Alfalfa and subclover), and Tephrosieae (Wisteria) (Palmer and Thompson 1981, 1982, Palmer et al. 1987) is the only group of angiosperms so far found to have lost the inverted repeat altogether, and has the smallest cpDNA length of 120kbp. The first gymnosperm to be investigated was Ginkgo biloba, which also proved to possess the inverted repeat
(Palmer and Stein 1986). This species is not closely related to other gymnosperms, and is the only genus in the class Ginkgoopsida. Subsequent mapping studies of gymnosperms in the Coniferopsida showed that the inverted repeat is lacking, and this appears to be universal in this class (Strauss et al. 1988, Lidholm et al., 1988, White 1990a).

The inheritance of angiosperm cpDNA is predominantly maternal, that is an individual inherits the cpDNA genome solely from the seed bearing parent (see review by Palmer 1985a and e.g. Conde et al. 1979, Soltis et al. 1990, Corriveau et al. 1990). Biparental inheritance in the angiosperms is estimated to occur in about 14% of genera, and 19% of families (Corriveau and Coleman 1988 and e.g. Smith et al. 1986, Forsthoefel et al. 1992, Derepas and Dulieu 1992). Gymnosperms have, by contrast paternally inherited cpDNA, resulting in a genome that is transported via pollen (Neale et al. 1986, 1988, Dong et al. 1992, Szmidt et al. 1987, Stine and Keathley 1990, Neale and Sederoff, 1989 Wagner et al. 1989).

DNA VARIATION AT THE LEVEL OF THE INDIVIDUAL

The genome organisation described above is the basic model that occurs throughout the plant kingdom. All individuals, however, are not alike and it is this variation between the genotype of individual organisms that enables evolution to occur (Endler 1986). DNA can vary in two qualitatively different ways: by substitution of one base for another in the DNA sequence, and by rearrangement, that is changes in the position and length of DNA sequences. The rates at which these processes occur are not uniform, and in order to predict the behaviour of a type of DNA marker it is necessary to be aware of the rates of change and how they vary among the three different genomes. A description of the modes of change and the rates at which they occur in each of the three genomes follows.

Variation in DNA Sequences.

Substitution

A substitution occurs when a mutation causes one nucleotide to be substituted for another in the DNA sequence. Nucleotide substitutions that result in a codon that codes for the same amino acid as before are termed "silent" or synonymous substitutions. Non-synonymous substitutions can either be "missense", in which case the codon now codes for a different amino acid, or "non-sense" where a stop codon occurs, thus prematurely ending transcription. Only non-synonymous
substitutions will change the phenotype of the individual because gene products are affected (Li and Graur 1991).

**Structural Mutations**

Structural mutations can be of three types: Translocations/inversions, insertions/deletions, or duplications. Translocations and inversions do not cause any change in the size or composition of the genome, but change the relative position of genes. Insertions and deletions change the size of the genome, and may cause changes in both genotype and phenotype because of the addition or removal of a gene. Copy number changes or duplications, increase the size of the genome, but do not affect the phenotype or sequence complexity. However, duplication in transcribed genes, either in whole or in part, may be an important mechanism in the evolution of new genes (see Li and Graur 1991, chap6).

**Non-mutational Variation**

Variation can also be generated that does not require mutation in the DNA within an individual, but occurs as a result of hybridisation between individuals of different species. If the hybrids are fertile, this process results in introgression of genes from one species into another. In plants, polyploidy is common, that is, where the haploid complement of chromosomes is present in more than two copies. Polyploidy is thought to be associated with speciation, as divergence among some closely related plant species is correlated with its occurrence. This phenomenon is common throughout the plant kingdom, with many examples from angiosperms, bryophytes and pteridophytes, but there are no known polyploid gymnosperms.

**Rates of DNA Change**

**Substitution Rates**

The most important factor governing observed rates of nucleotide substitution is whether or not that change is non-synonymous. Substitution rates at sites where a change is more likely to result in the transcription of a different amino acid are significantly lower than the rates for non functional DNA, non-coding regions within a gene, and sites where substitutions are more likely to be synonymous. Data from mammalian nuclear gene sequences have given values for non-synonymous substitution rates averaging $0.85 \times 10^{-9}$ substitutions per site per year, whereas the synonymous rates are five times higher at an average of $4.6 \times 10^{-9}$ (Li and Graur 1991). These rates are similar to those calculated for plants (Wolfe et al. 1989). These rates are variable within genes, with for example histone genes showing
almost no non-synonymous substitution, compared to some interferon genes which have rates nearly three times the average.

The above estimates of nucleotide substitution are based on nuclear DNA sequences. Organelle DNA substitution rates have been found to be significantly different from those of nuclear DNA. Non-synonymous substitution rates in plant mitochondrial and chloroplast DNA are broadly similar, and are slightly less than those of nuclear sequences. Synonymous substitution rates are much less, with the rates in mitochondrial, chloroplast and nuclear sequences in the ratio of 1:3:12 (Wolfe et al. 1987, 1989).

**Structural Rearrangements.**

There are types of DNA that have an intrinsic capacity to either move or multiply. These include transposable or mobile elements that move between chromosomes, and the classes of highly repetitive DNA such as satellite DNA. Rates of structural rearrangement vary as much as substitution rates do (see Li and Graur 1991 for review). In particular, the multicopy families of genes and hypervariable sequences commonly used in population studies evolve at rates orders of magnitude higher than the substitution rates of single copy genes. Mutation rates as high as 5% per gamete have been measured for human minisatellite loci (Jeffreys et al. 1988).

The plant mtDNA genome undergoes rapid structural rearrangement. The arrangement of the genes is highly variable in mtDNA and this genome also has a high rate of uptake of nuclear and chloroplast DNA sequences into its structure. The numerous repeat sequences present in the mtDNA genome are thought to be partly responsible for the variable genome conformation, because this may allow recombination to occur (Palmer 1985a). The large amount of variation between closely related plant species in the size of the mtDNA genome has already been noted (see above). More recent research has shown that variation within several species of pine is high, with for example, eight mtDNA haplotypes being found among two closely related species, *P. contorta*, and *P. banksiana* (Dong and Wagner 1993, Strauss et al. 1993).

Structural rearrangements are rare in chloroplast genomes that contain the inverted repeat. The absence of the inverted repeat in the legumes and gymnosperms is significant because it is correlated with a higher level of structural rearrangement. Although the absence of the repeat does not necessarily induce rearrangements, it may remove constraints on recombination. The homologous pairing permitted by
the presence of the repeat sequence, for example, means that the cpDNA molecule may exist in a conformation that makes recombination impossible (see Palmer 1985a, Day and Ellis 1985 and Strauss et al. 1988 for discussion). Studies have shown that the inverted repeat has been lost independently in the legumes and the conifers, and that rearrangements have subsequently occurred (Palmer and Thompson 1982, Palmer et al. 1987, Strauss et al. 1988). The variation that occurs as a result of this is mainly insertion and deletion of short sequences concentrated in the non-coding regions (e.g. Teeri et al. 1985, Palmer et al. 1988). The first documented evidence for this in the cpDNA of higher plants was in three closely related species of *Trifolium*: *T. subterraneum*, *T. globosum*, and *T. batmanicum* (Milligan et al. 1989), closely followed by similar findings in the gymnosperm, Douglas fir (*Pseudotsuga menziesii*, Tsai and Strauss 1989). Both papers suggest that these are transposable elements. Indeed, one family of the Douglas fir repeats was sequenced, and found to have transposon-like sequences at both ends. The evolution of these repeat families in distantly related organisms suggests that this is a recent phenomenon that is strongly related to the absence of the inverted repeat.

**POPULATION GENETIC STRUCTURE.**

So far, genetic variation has been considered at the level of the individual. Individuals exist within populations that are subdivided into finite groups showing non-random spatial distribution of individuals and genotypes. The extent to which populations become genetically differentiated depends on the balance between forces that tend to produce local differentiation, and those that increase genetic homogeneity (Slatkin 1987). The definition of population structure given above (Jain 1975) explicitly includes interactions at the individual and the population level. The following section reviews the factors that influence the population genetic structure at both levels. The genetic structure of a single population is considered first.

*Life History Strategy*

The life history strategy of a species determines how far genes travel within and between populations. The terms "dispersal" and "migration" have been fairly loosely used, and I use Endler's (1977) distinction between migration, dispersal and gene flow. In plant populations, migration is not a relevant parameter because they are sessile. Dispersal in the form of seeds, or plantlets, may or may not lead to gene flow. The extent of this depends on the "vagility" of the organism and the
availability of suitable regeneration niches in the recipient population. More important is the breeding system, that is, the level of inbreeding and the dispersal of pollen. Breeding systems in plants are diverse, ranging from obligate inbreeding and apomixis at one extreme to obligate outcrossing through self incompatibility or dioecy at the other (Brown 1979 and references therein). Analyses of both these parameters have shown significant correlation with the existence of population sub-structuring. In continuously distributed plant populations predominantly selfing plant species showed higher levels of differentiation among sample sites than outcrossing species. This is consistent with the expectation that gene flow is low in selfed plants because of the lack of pollen flow among genotypes. Outcrossing species nonetheless exhibit substructuring in continuous populations, suggesting that limited gene flow as a result of low dispersal creates genetic differentiation within populations (Heywood 1991).

Population Size and Random Genetic Drift.

The number of breeding individuals in a population has a profound influence on the level of genetic diversity among populations of a species. In large populations, genotype frequencies remain stable in the absence of perturbing factors (selection, migration and mutation). In small populations, however, random sampling of genes may lead to certain alleles becoming fixed and others becoming extinct purely by chance, an effect known as random genetic drift (Falconer 1989, Hartl 1980). Populations may therefore show significantly different gene frequencies over time: indeed, this is considered to be an important step in the evolution of reproductive isolation and speciation (Slatkin 1987).

The effect of random genetic drift within a population is to lower overall genetic variability. A special case of random genetic drift, the "founder effect" occurs in populations that are descended from a small initial population. Individuals that form a founder population are a random and limited sample of the donor population and the genetic base of the new population is correspondingly low, even if that population subsequently expands. A large population, despite high absolute numbers may also have a high proportion of fixed alleles if in the past it has undergone a period of low numbers or "bottleneck". In both of these cases, the effective population size may be much smaller than the actual population size because the probability is high that two genes sampled from the population are identical by descent (Hartl 1980).
Natural selection is defined as a process where if there is: (a) Variation in a trait, (b) Differential fitness among individuals for that trait, and (c) Inheritance of variation for the trait, then the frequency distribution of that trait will differ among life history stages beyond expectations from normal development (i.e. as "fit" individuals reproduce more than "unfit" ones), and if the population is not at equilibrium, then generations will differ significantly in trait frequency distribution (after Endler 1986). Measurements of natural selection in the wild have shown that selection pressures can be high, and that there is significant geographical and temporal variation in selection, leading to variation among populations for those adaptive traits. The fundamental difference between selection and drift is the existence of a fitness differential, that is, where one genotype is more likely than another to survive into the next generation. For this to occur in any living organism, variation in the trait, however weakly inherited, must have some effect on the way that individual functions in everyday life. Fitness is therefore by definition part of the phenotype of an individual. It follows that parts of the genotype of an organism that are not expressed in the phenotype cannot contribute to fitness, and cannot therefore be selected for. They may however be linked to an adaptive trait if they are physically close to one another on a chromosome, and correlate with adaptive traits, a process called "hitch-hiking" selection. Similarly, any gene that is expressed in the phenotype can potentially affect fitness, and may be or has been selected for.

For many years, selection was thought to be the most important process in evolution, with random drift only being important when population size is small. In this view, mutation occurs at random and variation is maintained in the population through balancing, directional or geographically varying selection (Mayr 1963, Nei 1983, Endler 1986). Mutation and recombination are the sources of the variation, but the evolution of change in the frequency of those mutations occurs because of selection. There are different modes of selection, (e.g. frequency dependant, disruptive and directional summarised by Endler 1986), but the balance theory of selection states that genetic variation is maintained at an equilibrium level by natural selection (Hartl 1980). This view has been criticised for "adaptive story telling", where most traits in organisms are thought to be adaptive, and any failure to resolve the reason for differences between groups of organisms is due to our inability to define what the selective pressure actually is (Gould and Lewontin 1979). This criticism notwithstanding, selective pressures
measured in natural populations can be high, and spatial and temporal variation in
selective pressures may be enough to maintain polymorphism at a high level.

Random drift of Neutral Alleles.

The "Neutral Mutation-Random Drift" hypothesis proposes that many mutations
are deleterious, but that only a few are advantageous, and that natural selection acts
mainly to remove deleterious mutations. It also assumes that mutations can occur
in structural genes that are effectively neutral. If a neutral mutation arises in an
individual, the probability that it will survive into the next generation is 0.5,
because there is no fitness difference between the mutant and the normal allele.
This sampling continues every generation until the mutant is either lost or fixed in
the population- the pattern is often described as a "random walk". In principle this
process is similar to that of random drift in small populations. The rate of the
allele's fixation is approximately equal to the mutation rate, and the average time
an allele takes to reach fixation is the inverse of the mutation rate. Given that
mutation rates average $10^{-5}$ per site per generation, the locus may be polymorphic
for long periods (Goodenough 1984). This process has the potential to produce
high levels of heterozygosity because of the large number of possible alleles (the
infinite alleles model). Evolution occurs predominantly by random fixation of the
mutations which are neutral or nearly so. Selective forces that bring about
evolutionary change are, in contrast to the previous view, very weak (Kimura
1983, Crow 1986). This theory was proposed to explain the unexpectedly high
levels of protein polymorphism found in plants and animals as enzyme
electrophoresis data accumulated after Lewontin and Hubby (1966) first used the
technique on natural populations. Even in large populations therefore, where
adaptive genes are not subject to drift because of restricted population size,
differentiation of neutral alleles can occur within populations due solely to the
balance between drift and migration.

GENETIC DIFFERENTIATION AMONG POPULATIONS.

Genetic differentiation is hierarchical in that groups of populations interact and the
structuring of genetic variation within and among populations can be considered at
any level, from local populations, to varieties within a species, to species within a
genus, and so on. The processes described above determine the overall genetic
variability within a population, and how it is structured. How these processes
affect the apportioning of genetic variation among populations is considered next.
In addition to this, larger scale geographic and historical influences on populations are described.

**Breeding System**

The extent of dispersal and gene flow between geographically separated populations depends on the vagility of the organism. Outcrossing species that are adapted to long distance dispersal of pollen, for example, are more likely to cross-pollinate individuals from another population and among population differentiation will be correspondingly low. A self-pollinating species or one with low vagility, by contrast, may have highly differentiated populations because the interaction between them is relatively low even though it may show similar overall levels of variation when the populations are considered together. This has been demonstrated in an extensive review of allozyme variation in plant species, where inbreeding species were shown to apportion a significantly larger proportion of genetic variation among populations than species with mixed mating or outcrossing systems, and that species with wind dispersed pollen showed significantly less population sub-division than species using animal vectors to disperse pollen. There are also significant differences in population structuring among different modes of seed dispersal (Hamrick and Godt 1990).

**Random Drift**

The effect of random drift leads to differentiation among populations because alleles become fixed at random. When groups of populations are compared, therefore, each one may be fixed for a different allele. This effect is only maintained, however, if the population is isolated, either reproductively, or as is more likely, by distance. Comparatively small amounts of gene flow are required to offset the effects of genetic drift: on average, one breeding individual per generation exchanged between populations is enough to prevent different neutral alleles at a locus being fixed (Slatkin 1987). This applies both to random drift of adaptive variation due to small population size and to random drift of neutral alleles in large populations. Some species, particularly those that are adapted to utilising ephemeral habitats may be constantly undergoing this process of random drift, as they colonise a habitat and disperse (Barrett and Husband 1990).

**Selection**

Natural selection may lead to significant genetic differentiation among populations because the environment is not uniform, and the selective pressures experienced by a species in one region may be entirely different from those in another. In certain
circumstances this can lead to rapid differentiation. Perhaps the best known example of this is industrial melanism in the peppered moth *Biston betularia*, where melanic forms increased in frequency in industrial areas from the middle of the last century, due in part to enhanced crypsis on sooty tree trunks. In rural areas where pollution was less, the melanic form remained at lower frequencies (Kettlewell 1958). The evolution of heavy metal tolerant plant populations in the vicinity of industrial sites has been demonstrated in many plant species. *Agrostis tenuis*, for example is a wind pollinated, self-incompatible species, that might therefore be expected to show low levels of among-population differentiation. Selection against genotypes that are less able to tolerate high concentrations of lead and copper in the soil in contaminated areas in Wales has, however, led to sub-structuring of genotypes over short distances, despite there being no barrier to gene flow (Bradshaw *et al.* 1965).

**Biogeography**

On a global scale, the geographic distribution of a species depends in part on geology and historical accident. As the climate and geology of a region changes, barriers such as bodies of water and mountain ranges arise and disappear, so too may the distribution of a species change (Pielou 1979, Slatkin 1987). At the community level, this is demonstrated by different, and usually impoverished, species composition on islands. Britain for example has only three native conifers, not because others that are common in the rest of Europe cannot grow in Britain but because they failed to colonise this land area before it became isolated from mainland Europe after the last glacial episode. Within a plant species, such separation may lead to genetic differentiation, and measured variation among populations may reflect a historical event and not adaptation.

Major geographic disturbances can be a significant factor influencing population genetic structure, and may affect entire communities of organisms. The South Eastern United States, where the Florida peninsula separates the Gulf of Mexico from the Atlantic Ocean has undergone many climatic and sea-level changes during the Pleistocene glaciations. The Gulf region was effectively cut off from the Atlantic region for much of this time. A comprehensive survey of population genetic structure in marine organisms on the coasts of the Florida region shows that most can be divided into an eastern and a western group, and those that do not are those species that show higher levels of dispersal between the two regions (Avise 1992). Many of these species are continuous in distribution along the coast, and the extent to which they remain differentiated seems to depend on the dispersal
ability, and what we are seeing is a major disruption of range causing a random divergence, which is gradually returning to an equilibrium level depending on the rate of gene flow. This differentiation is therefore not being maintained by an equilibrium between dispersal and selection or drift. There is therefore a growing recognition that population geneticists should know the history of a population in order to interpret population structure (Birky et al. 1989, Avise and Nelson 1989, Avise 1992, Slatkin 1987).

The Relative Importance of These Processes.

The relative importance of selection and random drift in evolutionary change has been the subject of heated debate among biologists (often called "Selectionists" and "Neutralists" respectively). The main demarcation between these views has less to do with the details of the theories than the type of variation being looked at. Those who study selection are primarily investigating morphological traits in natural populations, and the neutral theory has evolved from studies of biochemical and molecular traits observed from laboratory studies. Studies of natural selection have shown that selective pressures can be high and environmental variation can be sufficient to explain maintenance of genetic polymorphism by selection. On the other hand, predictions based on the random drift hypothesis have been shown to fit patterns of molecular variation comparatively well; but not enough is known about the function of molecular variants to assess what selective constraints and pressures they might be subject to (Endler 1986). One of the criticisms of the neutral theory is the difficulty of proving whether or not an allele is in fact truly neutral. The evidence for codon preferences, where organisms show non-random usage of codons for amino acids coded for by more than one codon, tends to suggest that even synonymous substitutions may not be truly neutral (Goodenough 1984). Even if neutrality could be proved, an allele that is neutral now may have been selected for in the past, since selection pressures can vary significantly over time. Ultimately, both processes are probably happening in natural populations, and the question is not which is more important, but rather what are the circumstances that cause one to predominate. There are arguments that support both theories, but there is still much work to be done to determine how they interact. Attempts to reconcile the two theories centre on the fact that selection cannot explain the origin of variants only their replacement, except in exceptional circumstances (Endler 1986) and that individual alleles may be effectively selectively neutral, while their additive effect as a component of a polygenic trait means they are not phenotypically neutral (Milkman 1982).
THE DETECTION OF GENETIC VARIATION

The nature of genetic variation and the processes that cause non-random distribution of variation have been considered in some detail in the preceding paragraphs, because an understanding of the components of evolutionary change is essential to appropriate choice and interpretation of genetic markers. Without being able to identify and quantify that variation, it is not possible to measure the processes that are occurring in populations. There is now an array of techniques available to the conservation biologist for investigating the genetic structure of populations. With the advent of molecular genetic techniques to supplement more traditional methods, acquisition of genetic data, although often costly, is nevertheless easier than ever before. The modern methods are, however, technically difficult, and rely on a considerable knowledge of molecular genetics for interpretation. Despite this these methods are being used ever more frequently to evaluate, for example, threatened populations (Peters et al. 1990), estimate levels of diversity (Terachi et al. 1990, Clegg et al. 1984, Banks and Birky 1985), and quantify the interactions between widespread and rare congeners (Ellstrand 1992, Crawford et al. 1992). Indeed the potential for rapid assessment of some previously intractable questions is enormous (Clegg 1990, Schaal et al. 1991). There has never been a greater imperative therefore, for conservation biologists to familiarise themselves with a subject that in the first instance may seem irrelevant to the everyday problems of protecting endangered species.

The following section describes the most commonly used techniques for quantifying and characterising genetic variation. For each technique, I will describe the portion of the genome being sampled, and the processes that it is a marker for. Examples of their use, and the sorts of question each method can and perhaps more importantly, cannot answer, are discussed.

Phenotypic Variation

Morphological Variation

 Quantitative characters are those that do not show sharply demarcated types, but where each individual's expression of that trait is part of a continuum between extreme values. The study of quantitative characters depends upon measurement and not counting, and the value of the character expressed by a single individual is not informative. Quantitative genetics necessarily involves the measurement of populations of individuals. Familiar examples are size and weight of humans, ear length in maize and litter size in mice (Falconer 1989). The phenotypic value of a
quantitative character (as defined by Falconer 1989) is made up of a genetic and an environmental component. The effect of the genes that cause the character to be expressed is the genetic value and the modification of the genetic value is the environmental deviation. Low weight and height in malnourished humans are examples of environmental modification of genetic value. The proportion of the variation in phenotype that is due to the genetic value is expressed in that character's heritability, or the extent to which the character is genetically determined ("broad sense" heritability).

Measuring morphological variation is often the simplest way of assessing genetic variation because it can be done in the field and requires no breeding or laboratory studies (Schaal et al. 1991). Plant species often have morphological characters that are indicators of underlying genetic differences, for example flower colour and leaf shape. These are phenotypic characters, and are likely to be ecologically adaptive. Studying morphological variation without any further analysis, although simple, is only an estimate of variation, and very broad assumptions are made about the extent of genetic determination. The environmental variation or phenotypic plasticity of character traits in plants can be very high so it is advisable to determine the underlying genetic basis as well.

Quantitative genetics is a formalised experimental approach to quantifying genetic variation from the phenotypes of populations of individuals. The analysis uses data from generations whose relatedness is known (e.g. the progeny from one seed parent are assumed to be half-sibs because the pollen donor is unknown) to estimate the heritability of character traits, and hence, the environmental variance. This approach is used extensively in breeding experiments such as forest tree provenance trials, to estimate the genetic differentiation between provenances and performance under similar environmental conditions (e.g. Wright and Bull 1963). In Ecology, measurements of quantitative morphological traits have been used to demonstrate natural selection for heavy metal tolerance in Agrostis tenuis (Bradshaw et al. 1965). Clearly, these are adaptive traits, that have been and may be selected for.

**Phenotypic Markers**

Phenotypic characters whose variation is controlled by small numbers of genes have also been used to estimate the genetic variation in populations. These are generally traits that are controlled by simple genetic systems, such as melanic morphs in the peppered moth Biston betularia (Kettlewell 1958), banding patterns
in the snail *Cepea nemoralis* (Bantock and Ratsey 1980) and cyanogenesis in many plant species (e.g. Crawford-Sidebotham 1972, Ennos 1981). The simple genetic determination of these markers means that environmental variation is at a minimum, and genotype can therefore be inferred from the phenotype of the individual. The above examples have all been shown to be subject to natural selection, often of high magnitude.

**Isozyme Variation**

Isozymes are enzyme molecules that have the same function and are therefore phenotypically indistinguishable, but that are structurally different. Isozyme electrophoresis has been used to quantify genetic variation in large numbers of species, and most of what is known about the genetic structure of non-cultivated species has come from studies of isozymes. Electrophoresis of allozymes samples a small portion of the genotype of an individual. Specifically, it detects loci in nuclear DNA coding for soluble enzymes. Polymorphisms detectable at these sites are the result of non-synonymous substitution that give the polypeptide a net change in electrostatic charge. Starch gel electrophoresis can separate these allozymes (so called because they are alleles at the same locus), by running samples of tissue extract through a starch gel with a voltage differential across it. Allozymes of different charge migrate at different rates, and an enzyme specific stain applied to the gel will reveal the allozymes as bands at different distances from the origin. This method cannot detect silent or non-synonymous substitutions that result in no net change in charge, and it is estimated that about 67% of alleles per locus are not detectable by electrophoresis (Ayala 1982).

Isozyme analysis is the most widely used technique to estimate overall genetic variation in natural populations, and for this reason I will consider the important characteristics of isozyme variation in some detail. The vast majority of the eukaryotic genome, as we have seen, does not code for structural soluble enzymes so only a tiny, and very specific proportion of the genome is being sampled by this method. There is some evidence to suggest that the level of variability in enzyme sequences is higher than in the genome overall (Schaal et al. 1991). This has been used as a criticism of isozyme analysis, but I would suggest that from the preceding evidence about variability in DNA, this is not necessarily a reason for dismissing the methodology out of hand. Rather, I would contend that this is something the worker must be aware of, and that while isozyme variability is a good measure of the relative variation among populations, it has no more value as an absolute indicator than any other.
Another factor to be considered when comparing published values of heterozygosity and polymorphism is that laboratories have different criteria for assigning allozyme bands to loci. This is a particular problem for enzymes where many loci code for gene products that perform a similar function, and also in polyploid organisms. Complex banding patterns result, and if they are monomorphic, without detailed analysis it is not known whether the system represents one locus or many. How this is interpreted will clearly influence the calculated levels of heterozygosity and polymorphism (R. Ennos, P. Hall, pers. comm.).

The assumption that isozyme variation is neutral has been questioned. The neutral theory predicts that selection coefficients will be so low for allozymes that when a population, or its effective population size is not very large, then they will be effectively neutral. However, the frequencies of allozyme variants have been shown to be influenced by selection (Prus-Glowacki and Nowack-Bzowy 1989, Prus-Glowacki and Godzik 1991, Table 5.1 in Endler 1986). Related to this is the assumption that allozyme variation is a good indicator of overall genetic variation in a population, including quantitative variation. This may not be the case, particularly if the population has been bottlenecked in the past. Lande and Barrowclough (1987) have shown that if this has happened, quantitative variation, which arises by selection, recombination and migration, recovers much more quickly than allozyme variation, which can only arise by mutation or migration. Low levels of allozyme variation may therefore be an indicator of a historical event and therefore be a poor measure of a species' ability to respond to selection. For conservation purposes therefore, allozyme variation may have a limited use in isolation, and should be combined with a wider analysis of morphology and other adaptive variation.

There are, however, questions for which isozyme analysis is a particularly useful method. Isozyme analysis is second to none for providing quick and accurate estimates of rates of inbreeding, knowledge of which is essential for predicting levels of gene flow and demographic processes. The effect on genotypic versus phenotypic differentiation of known historical events can be assessed, and relative comparisons of diversity and differentiation among species can be made. Isozyme analysis also has the merit of being relatively inexpensive and simple to perform, and the protocols for many species have been published, making development of the systems much easier than for the more complex methods described below. It is for this very reason, however, that isozyme variation should be interpreted
cautiously. It is all too tempting to measure the isozyme variability before considering why and even if the information thus collected is relevant.

**Direct detection of DNA Variation**

When the routine analysis of DNA sequences was made possible by developments in molecular biology, a whole array of new techniques was made available to population biologists. Direct analysis of DNA sequences was possible and thus it is now possible to look at variation that is genuinely neutral because it is not transcribed. Ideally, DNA variation would be determined by sequencing the genome of interest allowing the number of nucleotide substitutions and rearrangements to be directly determined. However, this is rarely practical, so methods of sampling DNA variation have been developed. In theory, any kind of DNA can be examined provided a probe or primer has been identified that is homologous to the DNA sequence of interest. The first of these techniques uses homologous sequences of DNA to act as probes to detect Restriction Fragment Length Polymorphism (RFLP) in DNA sequences that have been cut at specific base pair sequences by restriction enzymes. The second uses the Polymerase Chain Reaction (PCR) to amplify sequences using random or specific primers to initiate replication of particular sequences. These are both standard methods in molecular biology, and are used for many applications. Population biology has focussed on particular DNA sequence types, especially those that are present in multiple copies. The following sections describe the basic principles behind RFLP analysis and PCR, and describe how these techniques have been applied to problems in population biology.

**Restriction Fragment Length Polymorphisms**

RFLP analysis has been widely used in the analysis of populations. The protocols for the laboratory methods required to do this analysis are described in detail both here (Chap 3) and elsewhere (e.g. Hoelzel 1992, Sambrook et al. 1989 Kricka 1992, Palmer 1986). The most important elements of the process are:

- **Restriction enzymes**, which cleave DNA at particular nucleotide sequences. They generate a series of fragments of different length, and because the enzyme always cuts at the same nucleotide sequence, the fragment patterns are repeatable, and can be compared between samples.

- **Probes**. These are sequences of DNA that are used to detect the fragments containing the sequence of interest. For example, a copy of a ribosomal DNA gene can be isolated from an individual barley plant. This piece of DNA can
be cloned and replicated and then hybridised to homologous fragments (i.e. fragments containing similar nucleotide sequences to the probe). The probe is labelled, allowing the fragments to be detected and recorded. Again, this is a repeatable procedure, allowing populations to be scored and compared.

Restriction Fragment Length Polymorphism (RFLP) analysis detects two different kinds of variation in the DNA sequence. Firstly, the presence or absence of a particular restriction fragment indicates that a point mutation has changed the DNA sequence either creating a new or removing an existing restriction site. Secondly, variations in the length of fragments can be detected indicating that there have been rearrangements of the DNA of any of the kinds described above. The detection methods described below all use restriction digestion to generate arrays of fragments of different lengths. In addition, they all detect multicopy DNA sequences. Each, however, uses probes from a different region of the genome. The data set thus generated resolves relationships between organisms that have arisen over varying timescales, and the interpretation and appropriate use of each is different.

Chloroplast DNA

In plants, analysis of organelle DNA has centred on the the chloroplast genome. The mtDNA of plants has only recently begun to be used for population studies, so the following examples come from investigations of cpDNA. CpDNA has been used mainly as a method for generating characters that can be used to reconstruct the phylogenetic relationships among groups of plant species, for example the genera *Hordeum* (Clegg et al. 1984, Holwerda et al. 1986, Baum and Bailey 1989), *Rubus* (Waugh et al. 1990), *Gossypium* (Wendel and Albert 1992), *Pisum* (Palmer et al. 1985), *Salix* (Brunsfield et al. 1992), *Glycine* (Doyle et al. 1990), *Pinus* (Wang and Szmidt 1992a, Szmidt et al. 1988a, Strauss and Doerksen 1990), and others (see reviews by Palmer 1987 and Palmer et al. 1988). The nature of evolution in the cpDNA makes it ideal for this purpose, because slow nucleotide substitution rates and infrequent rearrangements are ideal characters for identifying major phyletic divergence events (Palmer et al. 1988).

Chloroplast DNA variation has mainly been used as a marker at the inter-specific level because of its slow rate of sequence change. For phylogenetic analysis, many workers have analysed only a few individuals of each species, making the assumption that cpDNA is invariant. As more information on the levels of cpDNA variation is assembled, however, and larger populations of species are studied,
there are increasing numbers of reports of intra-specific variation in cpDNA, and this assumption is being questioned by some authors (Baum and Bailey 1989, Milligan 1991). The chloroplast genome undergoes three specific kinds of variation, nucleotide substitution, major inversions and small insertions/deletions in non-coding regions, and the point is clearly made by Palmer et al. (1988) that these occur at different rates, and that they therefore resolve evolution at different timescales. To use cpDNA markers for evolutionary studies, it is important to determine what the detected variation represents, and to sample the populations adequately.

Although cpDNA variation has been found to be low or absent in many species (e.g. Ecke and Michaelis 1990, Strauss and Doerksen 1990, Wagner et al. 1992, Haghigi and Hancock 1992), certain populations have revealed sufficient variation to allow the study of population differentiation for example Heuchera micrantha (Soltis et al. 1989), Tiarella trifoliata (Soltis et al. 1992), Plantago spp (Wolff and Schaal 1992), Trifolium pratense (Milligan 1991) and Lupinus texensis (Banks and Birky 1985). The study of population differentiation in organelle genomes is potentially useful in plant breeding: comparisons of cpDNA diversity in natural populations of Hordeum (barley) have shown that in the process of domestication cultivars have significantly less cpDNA variation than their wild relatives (Clegg et al. 1984).

**Ribosomal DNA**

Ribosomal DNA (rDNA) genes have a particular construction that has made them useful in population studies. They occur in tandemly repeated arrays or blocks at loci throughout the nuclear genome. Each repeat consists of five separate elements: three regions coding for rRNAs, internal and external transcribed spacers, and a non-transcribed Inter Genic Spacer (IGS) that itself is composed of a number of repeats, of the order of 130 kbp long. Ribosomal DNA variation in plant populations has been found to be due to variable numbers of the IGS repeat units (Clegg 1990, Saghai-Maroof et al. 1984, Schaal and Learn 1988).

The amount of variation present within and among plant populations varies considerably among species, from no variation in some isolated endemics, to up to 20 different length variants in a single Vicia faba plant (Schaal and Learn 1988). Levels of population differentiation are often low in natural populations, but rDNA variants have been found to vary significantly between cultivated barley and its
wild relative (Saghai-Maroof et al. 1984), and significant differentiation has been found among natural populations of *Phlox divaricata* (Schaal et al. 1987).

Though it has been suggested that the coding regions could be useful for phylogenetic studies on distantly related groups of plants because of the low sequence divergence (Govindaraju et al. 1992), most population genetics work on rDNA has centred on the non-transcribed IGS because the variability in this region is higher. Despite the fact that the IGS is presumed to code for no gene products, there is some evidence that it may nonetheless be subject to selective forces. One class of subrepeat elements has been found to enhance transcription in *Xenopus*, which would put them into the class of modifier or regulatory genes (see above) (Reeder 1984, Reeder et al. 1983). Indirect evidence for directional selection of rDNA variants comes from barley, where breeding experiments and comparisons of cultivars and wild populations have shown levels of differentiation that are too high to be explained by random genetic drift (Saghai-Maroof et al. 1984, 1990), and that specific rDNA variants are favoured under specialized growth conditions (Zhang et al. 1990, Allard et al. 1990). Similar results are reported by Rocheford et al. (1990) in a single maize cultivar mass-selected for high grain yield. In a haploid fungal pathogen *Rhyncosporium secalis*, rDNA variation and among clone differentiation was found to be high, a factor attributed to selection for virulence. This pathogen is imperfect (i.e. has no known sexual form) and very low rates of recombination, so even neutral variants become linked to selected genes (McDermott et al. 1989).

The mechanism governing evolution of rDNA repeats is thought to be unequal crossing over within the subrepeat region of the IGS, and the evolution of varying copy numbers appears to be rapid. Ribosomal DNA is therefore potentially useful for identifying variation within and among populations. There is strong circumstantial evidence that IGS subrepeat variation is at least closely linked to adaptive genotypes and may be under selection for increased efficiency of rDNA transcription (Jorgensen et al. 1987, Schaal and Learn 1988, Rocheford et al. 1990). It is probably not, therefore appropriate for determining historical relationships because relationships may have been masked by subsequent adaptive evolution, though this approach has revealed relationships between species similar to that of classical morphological studies in the genus *Secale* (Reddy et al. 1990). However rDNA markers have potential as markers of useful variants in plant breeding (Rocheford et al. 1990, Saghai-Maroof et al. 1984).
Minisatellite DNA is a sequence of up to 20 kbp that is made up of multiple copies of a short sequence, a tandemly repeating unit that is usually 65 base pairs in length (Li and Graur 1991). This type of DNA is mostly non-coding, and has been shown to evolve rapidly- the most variable locus ever detected is a human minisatellite locus (Jeffreys et al. 1988). There are families of minisatellite DNA in which the tandem repeat units contain the same core DNA sequence. The technique popularly known as "DNA fingerprinting" uses a core sequence as the probe, which hybridises to all the minisatellites containing that core sequence. Many codominant loci showing Mendelian inheritance can therefore be detected, and this results in the complex DNA "fingerprints" which are now familiar (Jeffreys et al. 1985, Burke 1989).

The core sequences at the heart of minisatellite DNA appear to be universal in eukaryotes. The human minisatellite sequences originally isolated by Jeffreys et al. (1985) and similar M13 phage sequences have proved to be homologous to sequences in plants (Dallas 1988, Rogstad et al. 1988). These probes have been used to determine pollen donor identity in apples (Nybom and Schaal 1990b) and genotypic distributions in apomictic and outcrossing species of Rubus (Nybom and Schaal 1990a). The hypervariability of these sequences precludes them from use as markers to investigate evolutionary relationships, but the full potential of these markers in monitoring processes occurring on a short timescale such as maternity and paternity testing and selective breeding has yet to be fully realised. DNA fingerprints are potentially useful in marker assisted selection programs because minisatellites are dispersed randomly throughout the genome, and therefore can be used as markers for overall genomic qualities, something that a single locus marker cannot do (Hillel et al. 1990).

**Polymerase Chain Reaction (PCR) Based Techniques.**

The development of PCR as a population genetics tool has allowed new approaches in the molecular genetic analysis of populations. PCR uses the enzyme Taq polymerase to synthesis DNA from a template, using oligonucleotides (i.e. short single stranded DNA sequences <50 bp long) as primers. Under the correct conditions, the template DNA is denatured so that it is single stranded. The primers then anneal to the complementary sequences on the single stranded template. The primers act as starting points for synthesis of new strands of DNA between two primer sites. The resulting double stranded DNA is denatured, and
the cycle continues. This is carried out for many cycles thus rapidly increasing the concentration of the synthesised fragment. This sample can then be used for further analysis.

This is a relatively recent technique, and as yet has not been extensively used in population studies. However, two approaches have emerged based on this technique, differing in the selection of primers for the PCR reaction. The use of random primers to generate random amplified polymorphic DNA (RAPD) markers was first described by Williams et al. (1990). This technique enables large numbers of polymorphic sites to be identified relatively quickly, because of the enormous number of primers available. This technique has been used to assess levels of genetic variation within and among populations of economically important tree species (Chalmers et al. 1992), parasitic protozoa (Tibayrenc et al. 1993) and rare disjunct plant populations (Brauner et al. 1992). The first criticism of this technique as a method to assess variation relates to the technique itself. Unless conditions are strictly controlled and maintained at a constant level, artifactual variation may be obtained. Ellsworth et al. (1993) demonstrate the potential extent of this problem, and recommend a rigorous testing regime before deciding which fragments are genuine variants. Secondly, given all that has been discussed so far about variation in rates of evolution and resolving power of types of DNA, the use of RAPDs should be restricted to linkage mapping and marking of known traits or clonal structure in populations. The reason for this is that what each fragment represents is not known unless the fragment is isolated and sequenced. It has been estimated, for example, that in plants, 1% and 5% of fragments will originate from the chloroplast and mitochondrial DNA respectively (P.E. Smouse pers. comm.). The mode of inheritance of these markers is different from nuclear DNA, and using such an unknown mixture of fragments seriously compromises any attempt to interpret observed patterns in terms of ecological or historical processes. In addition, the PCR reaction is such that DNA is synthesised from only one strand of the DNA resulting in effectively dominant markers. Levels of polymorphism and other parameters that depend on the knowledge of allele frequencies cannot therefore be used with RAPD data.

From published sequence data it is possible to design primers complementary to sequences within known genes. Using such primers generates fragments that come from known regions of the genome. This allows fragments to be bulked up from very small samples, or from single copy regions, where conventional RFLP analysis is impractical. PCR amplified microsatellite markers have been used to
analyse populations in social insects, where the sample available from, and the genetic variation in each individual is generally low (Queller et al. 1993). Alternatively, primers from known genes can be used to generate species specific probes, for regions of the genome that show little inter-specific homology. MtDNA probes for pines have been generated in this way because the commonly used mtDNA probes have proved to be difficult to use with the Pinaceae (Dong and Wagner 1993, Strauss et al. 1993). Another approach to population analysis using PCR is to use primers from known genes to generate a larger fragment from a particular region of interest which can subsequently be used for RFLP analysis (Soltis et al. 1992).

**Congruence Between Morphological, Isozyme and DNA Variation.**

The final point to make about the analysis of genetic variation within and between populations is that data from different methods may sometimes, but not always, show the same pattern. The many instances of varying rates of evolution, and the influence of random historical events discussed in the preceding paragraphs should make it clear that the most important thing to consider when embarking upon any measurement of genetic variation is what that measurement represents. A measurement of genetic variation *per se* is of no value to a conservation biologist unless the meaning of that measurement is clear. The increasing amount of genetic data from so many parts of the genome made this problem more obvious. Where two measurements are congruent, there is less of a problem. However, instances of non-congruence do occur, and this needs to be explained.

An example of this problem on the microevolutionary scale has come from diverse groups of organisms on Hawaii. For example, *Bidens* has undergone rapid divergence since migration to the archipelago, and this genus comprises large numbers of endemic species. This genus, however, also shows low levels of DNA and allozyme polymorphism (Schaal et al. 1991). This is an example of the situation modelled by Lande and Barrowclough (1987), where the recovery of neutral variation after a bottleneck or in this case, a founder event, is much slower than the increase in morphological and other adaptive variation. Higher levels of allozyme variation than morphological variation have been documented among the two species *Datisca cannabina* and *D. glomerata*. Divergence times estimated from allozymes of these two species range from 10-49Myrs, but morphologically, they are very similar, suggesting that morphological stasis has occurred in these species (Crawford et al. 1992).
The frequent instances of congruence between phylogenetic trees constructed using molecular and morphological data has led Szmidt et al. (1988a) to suggest that they may be "old wine in new bottles", though it is clearly corroborative evidence in favour of the classical taxonomic approach. In a phylogenetic study of Salix, Brunsfield et al. (1992) found high levels of cpDNA variation and introgression between species, and a genetic structure that was poorly correlated with morphological characters. They strongly recommended that combining data from morphological and genetic studies would provide a more robust picture of evolution in that group. Recent work on molecular evolution in mammals has proved to be more controversial. As an order, the Rodents are unusually variable, and have frequently been cited as an example of a rapidly evolving group. The Guinea pig family are particularly divergent. Comparisons of DNA and amino acid sequence data led Li et al. (1992) and Graur (1993) to suggest that the Guinea pig family is not closely related to the rest of the rodents, and that their placement in that order should be reconsidered. The sceptical response from palaeontologists to this idea (see pp339-342, Trends Ecol. Evol. vol. 8 1993) suggests that this controversy will not be quickly resolved.

"The Right Tool For The Job"

Each of the methods for detecting genetic variation described above samples a different portion of the genome, and each marks processes occurring at varying timescales. The population biologist, before embarking on an investigation must therefore be aware of these differences if s/he is to select the appropriate marker for the purpose. The following section gives examples of the use of many kinds of genetic markers and the reasons for selecting the particular technique in that instance.

The most common technique used to measure overall diversity within and among natural plant populations has been isozyme analysis (Hamrick and Godt 1990, Hamrick et al. 1990). The Gst statistic (Nei 1975), because it is calculated using only polymorphic loci, overcomes many of the problems associated with assignation of monomorphic isozyme bands to loci, and is a valuable measure of the distribution of variation among sub-populations. Other measures, for the reasons described earlier, are more problematic. It is becoming increasingly clear, however, that non-congruence between genetic structure measured by isozymes and by other methods is not uncommon (e.g. Crawford et al. 1992, Chechowitz et al. 1990, Schaal et al. 1991). Any effort to measure variability present in populations for genetic conservation must include data from as many kinds of variation as
possible. If one type of variation is monitored to the exclusion of others, selection may inadvertently result in loss of the unmeasured variation, a phenomenon that has been recorded in barley, where selection for variability in the nuclear genome in the form of isozymes appears to have coincided with a loss of cpDNA diversity (Clegg et al. 1984). For this reason, selection for isozyme variation should be avoided (Lande and Barrowclough 1987).

Another concern of the conservation geneticist is the ability of a plant species to respond to the changing environment. Plant species cannot respond to changes in habitat quality by migration (sensu Endler 1977), and are therefore vulnerable to isolation and fragmentation. The effect on the population of such a change may be important for the conservation of that species, and to assess this, natural selection is clearly the most important parameter. In Phlox, for example, Lepidopteran pollinators were shown to be able to discriminate between colour variants and flower outlines, and that among two species, P. glaberiana and P. drummondii, the phenotypes that were most similar showed the highest levels of inter-specific pollen flow (Levin 1969). In order to quantify such a change adaptive traits must be used as markers.

Demographic parameters such as gene flow and breeding system can be investigated using any marker that has the appropriate level of variation. Measurement of these parameters are usually made by comparing one generation to another, and because the timescale is short, strict neutrality is not as important, though traits that are under strong selection pressure should be avoided. Mendelian markers such as isozymes have been used for determining the breeding system in such diverse species as for example, Pinus sylvestris (Helgason and Ennos 1991), Arabidopsis thaliana (Abbott and Gomes 1989), Larix laricina (Knowles et al. 1987) and two species of Acacia (Moran et al. 1989). DNA fingerprints have also been used in paternity analysis in domestic apples (Malus), and because the fingerprints are clone-specific, individual paternal contribution can be identified (Nybom and Schaal 1990b). Using genetic markers allows direct measurements of realised gene flow to be made, which overcomes many of the problems of studying pollen vectors and pollen capture by the plant. This has been done in the musk-thistle, Carduus nutans (Smyth and Hamrick 1987) and Pinus flexilis (Schuster et al. 1989) using allozyme markers. The uniparental inheritance of organelle genomes allows the contribution of seed and pollen producing parent to be identified. In angiosperms, where both organelle genomes are maternally inherited, pollen flow has to be inferred from a combination of nuclear and
organelle genomes, an approach that has been used to investigate hybridisation among Lousiana irises using allozymes and cpDNA markers (Arnold et al. 1992). In conifers, where cpDNA is largely paternally inherited, the potential exists for direct measurements of pollen and seed flow, an interesting opportunity to dissect the relative contribution of each to the population genetic structure (Clegg 1990).

The slow rate of change of cpDNA means that variation is more likely to be interspecific, making it a particularly useful marker for the study of introgression. This has been analysed in the *Picea sitchensis*, *P. engelmannii* and *P. glauca* complex known as "Interior spruce" using species specific cpDNA markers. This has enabled the species composition and proportion of hybrids in seedlots and plantations to be determined (El-Kassaby *et al.* 1988, Szmidt *et al.* 1988b, Sigurgeirsson *et al.* 1991, Sutton *et al.* 1991). This has a practical use in forestry, because each of the three species has different phenological patterns, and knowledge of seedlot composition allows the growth conditions in tree nurseries to be set at the optimum (El-Kassaby *et al.* 1988, Sutton *et al.* 1991). Species-specific markers have also shown that the origin of *Pinus densata*, a Chinese pine of long evolutionary history, was hybridisation between *P. yunnanensis* and *P. tabulaeformis*, two species which are now reproductively isolated (Wang and Szmidt 1990).

The analysis of longer term processes require the use of a marker that is truly neutral. The estimation of divergence times between morphologically similar species requires a marker whose rate of change is not influenced by selection, and will not undergo stasis (Crawford *et al.* 1992). Similarly, phylogenetic studies require slowly evolving markers that are unlikely to show convergence due to selection (Palmer *et al.* 1988). Golenberg *et al.* (1990) published a DNA sequence from a Miocene era fossil *Magnolia* (15-20 Myrs old), which offers the exciting possibility that molecular clocks in plants may be calibrated against fossils of known age, allowing better phylogenetic inferences to be made. Phylogenetic analysis of molecular markers can also be applied within a species, where sufficient intra-specific variation exists. The relationship between genetic variation and its geographic distribution, termed "Intraspecific Phylogeography" by Avise *et al.* (1987) can be used to determine the effect of historical events on populations. Although this has so far concentrated on mtDNA in animals (e.g. Avise 1992), the study by Soltis *et al.* (1992) on three species in the family *Saxifragaceae* show that plant populations may also show significant population sub-structuring of this kind.
To measure this, however, only a relatively slowly evolving and completely neutral marker will detect such a pattern.

CONCLUSIONS

The processes underlying the genetic variation between individuals and the structuring of that variation among populations have been described, as have the methods for detecting that variation. Population genetics is an important component of conservation biology, and recent developments have provided the population geneticist with a much wider array of techniques with which to analyse populations. However, it is one thing to acquire the data on populations using these methods and another to convert that information into effective management strategies. Each of the techniques described above enables a different aspect of the evolution and development of genetic differentiation to be resolved. The challenge for conservation biologists is to use this information in such a way that conservation of genetic diversity becomes less haphazard, and in particular, to ensure that the often limited resources available for such efforts are efficiently allocated. The key to this is an adequate understanding of the techniques, and what they represent, so that the conservationist asks the right question, and uses the best marker for the purpose. This study set out to determine whether it was possible to use molecular markers to resolve outstanding questions about the origins and genetic structure of natural populations important for conservation, where other methods have failed. The next chapter describes the selected case study, and what methodology was finally decided upon.
CHAPTER 2

NATIVE SCOTTISH PINUS SYLVESTRIS L.: A CASE STUDY

Prologue

The native Scottish woodlands dominated by Pinus sylvestris are remnants of an ecosystem that was formerly widespread, and has been reduced to its current size primarily due to exploitation by man. This process is continuing, despite the cessation of timber exploitation, because management practices in the highlands prevent many of these woodlands from regenerating. Despite the small area of pine woodland left, they still represent a significant proportion of the remaining ancient woodlands of Scotland, and moreover, the populations in the east of Scotland are the largest continuous tracts of natural and semi-natural woodland left anywhere in the British Isles, and the importance of conserving these remnants has often been emphasised (e.g. Steven and Carlisle 1959, McVean 1966, Faulkner 1977, Ratcliffe 1977, Worrell 1992).

Public awareness of the need to conserve these remnants has greatly increased in the last ten years, and many organisations, government and voluntary, and landowners are actively involved in efforts to restore derelict pinewoods, and establish new ones. The amount of attention that has been focussed on genetic aspects of this conservation effort is unusually high, and this makes Scottish pine populations a particularly interesting case study for the investigation of the potential uses of molecular markers in conservation (Faulkner 1977, Kinloch et al. 1986, Soutar and Spencer 1991).

The recent increase in the amount of molecular data available has enabled conservation biologists to look at the genetic structure of populations in more detail. The evidence suggests that most species are not monotypic, and possess a genetic structure that not only reflects adaptive differentiation, but also historical events, especially in species which have been influenced by physiographic and climate changes caused by glaciations (Avise 1992). Scottish pine populations are also interesting in this respect because the influence of glaciations on the distribution of pine populations and genotypes in Scotland has not been resolved, because evidence from various sources, including palaeoecology and population genetics is contradictory.
The methods outlined in the previous chapter offer the opportunity to look at the genetic structure of populations in a very detailed way. The aim of this project was to investigate whether this potential could be realised in a real-life situation, in an ecosystem that has particular problems which could be addressed using molecular techniques. There is an extensive body of literature already available on Scots pine, and Scottish populations have already been investigated using quantitative genetics, monterpene markers, and isozyme markers. These however, have as yet, failed to provide unambiguous evidence for the origin of Scottish pine populations, nor can they be used as practical tools for identifying populations, for the reasons described below. The following chapter describes the ecology and history of Scots pine in Scotland, the measures being undertaken to conserve it, how the available genetic data has influenced this policy, and finally, how molecular markers could potentially resolve many of the outstanding problems.

THE NATURAL HISTORY OF P. SYLVESTRIS IN SCOTLAND

Taxonomy of Scots Pine

The Genus Pinus

The genus *Pinus* comprises nearly 100 species, and is one of the most widely distributed genera in the Northern hemisphere. Basically the pines are divided into two subgenera, the *Haploxylon* or *Strobus* pines and the *Diploxylon* or *Pinus* pines. This division was originally based upon leaf morphology, for example, the patterning of the ray tracheid walls. This important subdivision has subsequently been confirmed using cpDNA RFLPs (Szmidt *et al.* 1988a, Strauss and Doerksen 1990, Wang and Szmidt 1992a), *rbcL* (large subunit of ribulose bis-phosphate carboxylase) gene sequences (Bousquet *et al.* 1992), ribosomal DNA sequences (Govindaraju *et al.* 1992), seed storage proteins and other biochemical characters (Schirone *et al.* 1991 and references therein) Each of the subgenera is subdivided into subsections and groups of species. The classification of pine is reviewed by Mirov (1967).

*Pinus sylvestris* is a *Diploxylon* pine. Shaw (1914) (cited in Mirov 1967) classifies it in Subsection *Pinaster*, and groups it with, among others, the species *P. nigra* and *P. montana*, both of which are geographically close to *P. sylvestris*. Subsequent revisions have tended to preserve this group. Chloroplast DNA variation among species has been used to construct a phylogenetic tree of a selection of *Pinus* species, and has also retained the close grouping of *P. sylvestris*.
with *P. nigra* and *P. mugo* (Szmidt et al. 1988a). Analysis of the seed proteins of 14 species of *Pinus* has also preserved this grouping (Schirone et al. 1991).

**Pinus sylvestris in Europe**

The western European pines are made up of two groups of species: the "Mountain" pines including *Pinus sylvestris*, and *P. nigra* and the Mediterranean shore or island pines (Mirov 1967, Shirone et al. 1991). The present distributions of the mountain pine species are mostly post-glacial in origin whereas the Mediterranean pines show relatively greater affinities to the pine species of the Carribean and are considered to be more ancient (Klaus 1989).

*Pinus sylvestris* itself is the world's most widely distributed conifer. It is found through most of the Eurasian sub-continent from Mongolia to the Atlantic coast of Europe in the west and from the Arctic circle to the Mediterranean. Although it is generally regarded as a single species, it is a complex taxon differentiated into many different types, as would be expected for a species with such a wide distribution (Mirov 1967). Carlisle (1958) describes 26 geographic varieties based upon phenotypic characters. Pravdin (1969) divides the species into 5 subspecies comprising a total of 31 climatic ecotypes, again based on phenotypic characters and geographic distribution. A revised classification (Ruby and Wright 1976) was based on a long term provenance trial of 180 seed samples planted in the USA and divided the species into 22 varieties, based on genetic characters and geographic origin, though the authors acknowledged that other valid varieties may exist in the less well sampled regions. The distribution of Scots pine is shown in Fig 2.1, showing all the varieties mentioned in the text (After Critchfield and Little 1966, Ruby and Wright 1976).

**Pinus sylvestris in Scotland**

Scottish pine populations are remnants of a previously more extensive population of a distinct ecotype within a very heterogeneous species complex. They are at the north western extreme of the distribution and are usually named as a distinct geographic variety, var. *scotica* (Schott) Beissner, based on phenotypic characters and the separation of these populations from the rest of Europe (Steven and Carlisle 1959, Carlisle and Brown 1968, Pravdin 1969). This has subsequently been shown to have a genetic basis (Ruby and Wright 1976). In provenance trials comparing samples from throughout the range of Scots pine var. *scotica* was shown to have affinities to southern European varieties for some characters and the
Fig. 2.1 Distribution of *P. sylvestris*, showing location of varieties mentioned in the text. After Critchfield and Little (1966) and Ruby and Wright (1976); AQUtaina, ARMenA, IBERica LAPponica, MONgolica, PYReneica, SCOtica, SYLvestriformis.
northern post-glacial varieties for others. Ruby and Wright (1976) suggest that Scots pine is a Pleistocene relict population.

The Ecology of Scots Pine.

Habitat

Throughout its distribution Scots pine occupies a wide variety of habitats. In arid environments it behaves as a mesophyte and it can be found in the dry sandy areas occupying a xerophytic niche in wetter climatic zones, but it is not a competitive species and tends to form the "para-climax" vegetation in habitats where more demanding competitors cannot thrive (Mirov 1967). In the west of Scotland, pine encounters a more oceanic climate than is found elsewhere in its range, with a mean annual precipitation of 2794 mmyr⁻¹ in Barisdale (Highland Region, Steven and Carlisle 1959). While there are still significant populations remaining in the west, the largest populations are in the areas of Scotland, that experience a more continental climate, for example, Strathspey where the precipitation is much less at 762-889 mmyr⁻¹ (Steven and Carlisle 1959). They are found mainly on freely draining, non-calcareous and relatively nutrient poor soils, such as glacial or fluvio-glacial deposits overlying granite (Carlisle and Brown 1968). Pine does not grow well on waterlogged soils, but it will colonise peat hollows in some areas. The populations which occupy such habitats are often in poor condition, however, and often show a higher susceptibility to pests and pathogens.

Phenology

*Pinus sylvestris* in its natural state has a maximum life span of 300 years, and attains a height of between 15-30m and up to 46m (Dallimore and Jackson 1966). It is a pioneer species and germinates most freely on open, disturbed sites (McNeill 1955, Steven and Carlisle 1959, Carlisle and Brown 1968). Regeneration occurs most frequently after catastrophic disturbances due to, for example, windthrow or fire, that create those conditions most favourable for regeneration. As a result, Scots pine tends to grow in groups of similar age, although it occurs in mixtures with other species (Carlisle and Brown 1968). Seed is first produced between 5-10 years of age. Flowering occurs during May and June and the cones take about 18 months to mature, with ripe seeds dispersing between December and March (Krugman and Jenkinson 1974).
Breeding System

Most conifers, including Scots pine are monoecious, that is separate male and female inflorescences are borne on the same plant. The evolution of monoecy in conifers may be an adaptation to avoid inbreeding (Smith 1976). In natural Scottish populations of Scots pine the outcrossing rate has been measured in two ways: by counting the number of empty seed on the assumption that these fail to germinate because they are self fertilised (Boyle and Malcolm 1985), and also by estimating the number of outcrossed progeny in germinated seedlings using isozyme analysis (Helgason and Ennos 1991). The former estimated the outcrossing rate to be 63% of fertilisations. The latter found the outcrossing rate to be not significantly different from 100%, suggesting that self-fertilisations are eliminated very early in the life cycle. Similar results have been found in Swedish pine populations (Muona and Harju 1988, Yazdani et al. 1985a). So although there is no barrier to self fertilisation in pine there is clearly a very strong selection pressure against selfed individuals, preventing them from being recruited into the population.

The Forests of Scotland in a Regional Context.

The forests of Scotland formed a transition between temperate and boreal European forests (McVean and Ratcliffe 1962). The pinewoods of Scotland are the southern and western outlier of the European boreal coniferous forest, and the deciduous woodlands reach their northern limits in Scotland, for example the Rassal ash woodland near Kishorn in Wester Ross (Ratcliffe 1977). In historical times there has been no obvious latitudinal zonation of forest types in Scotland, as oak and pine woodland occur in similar areas (McVean and Ratcliffe 1962). However there is evidence that pine did occur further north at one time (Gear and Huntley 1991), and the distributions of forest types considered at the regional level suggest that Scotland is a transitional zone.

The zonation of woodland types within Scotland can usually be attributed to edaphic factors. In areas with rich soils pine is excluded by deciduous woodland of more competitive species, usually oak (Quercus petrea and to a lesser extent Q. robur) and elm (Ulmus spp). The western oak woods for example, are often found in low lying areas on south facing slopes. The majority of the pinewoods remaining in Scotland are found on north facing slopes. Although the reason for this is not certain, it is likely to be an effect of competition between oak and pine.
for the best sites, and may also be due to pine being relatively more shade tolerant than other tree species, enabling it to survive on darker northern slopes (Steven and Carlisle 1959, Anderson 1967). The pine woodlands remain on these unfavourable sites, whereas the oak has since largely been removed by human populations.

There is a well-defined altitudinal transition on the north side of Loch Maree, where the pine cannot compete with the oak on the richer soils further down the
slope (McVean and Ratcliffe 1962). In the east, particularly on the Cairngorm massif, the soils are poor and acidic. For this reason, pine and birch woodland was always more common than oak, which was rare in this region especially at higher altitude (Anderson 1967). The theoretical distribution of pine forest in Scotland is described by McVean and Ratcliffe (1962), but it is likely that pine rarely formed pure stands, and the fact that the ancient woodlands that we call native pinewoods extant today are dominated by pine may reflect that the sites on which these woodlands grow are the poorest and most inaccessible, and were therefore not cleared by man for agricultural or other purposes (Fig 2.2). As such, the native woodlands that remain in Scotland today may not be representative of what occurred naturally after the Ice Age (Anderson 1967, Gimingham 1977).

What Happened to the Forests?

The following brief summary of the effect of man on the forests of Scotland is based mainly on Anderson's exhaustive work, "A History of Scottish Forestry" (1967). Neolithic man arrived in Scotland around 6000 years BP and began the process of deforestation in Scotland probably clearing much of the woodland on the coastal fringe for agriculture and grazing. Forest clearance of this type would have been concentrated in the lowlands, and the mountain and highland forest where pine would have prevailed was probably not affected. The Bronze Age was characterised by larger and more permanent settlements, and increasing numbers of grazing animals that would put pressure on the upland forests. It is at this time that the first depression of the tree line due to grazing pressure impeding regeneration may have occurred, though Carlisle (1977) suggests that the influence of man was probably minimal at that time. The Iron Age people continued this process, and added to this the use of wood for charcoal to smelt iron. However, the Romans, who provided the first written evidence about Scottish forests suggested that the Caledonian forest was large and abundant, necessitating large scale clearance for the network of Roman roads. Anderson (1967) sums up the effect of the Romans thus

"...the Romans found Scotland 'fearful with woods' (horrida sylvis) and left it somewhat depleted of, but to a very great extent clad with, forests" (Anderson 1967 Vol I p75).

This reference should be taken to mean Scotland north of the Forth, as Central Scotland was not forested at this time, and locating the exact location and extent of the forest the Romans called silvae Calidoniae from their writings is fraught with difficulty (Breeze 1992). More extensive forest clearance must have occurred
during the period of the Celtic Kings (446-1097AD), but large areas of forest remained, and a list of Scottish forests of that time includes many of the present native pinewoods, e.g. Rannoch, Glentanner (Glen Tanar), Abernethy, Rothiemurchus, Glengarry, Glenmoriston, Strathglass, Strathfarrar and Glenmore. It was the transition to feudalism however, that marked the beginning of the real decline in forest area. Forests were still managed for hunting, which was controlled by extensive legislation, but there was as yet no concern about forest decline, nor a perception of any real need to manage the forests. It was not until the 16th Century that the Scottish parliament passed laws in an attempt to halt forest destruction and encourage replanting. In the 17th Century increased attempts were made to manage the remaining forests of Scotland. Forestry became an established form of management in some areas, especially estates like Atholl, Seafield and Glen Tanar in the eastern Highlands, with large scale planting beginning in the early 17th Century. This is the region with the largest areas of forest today considered to be natural, but it is likely that although the trees are native in origin, the stands have been extensively managed or at least influenced by man in historical times. In areas outside of the estates managed for forestry however, the forests were under great pressure as a source of charcoal and building material, and it is during the 400 or so years up to the present day that most of Scotland's woodlands have been removed. To give an idea of the scale of the exploitation, Anderson (1967) describes a sale in 1808 of thinnings totalling 5000 stems from Glen Falloch, a pinewood consisting now of less than 300 scattered trees. In 1809 the sale of 16-20,000 stems was advertised from the Mar estate in the eastern Cairngorms, a woodland that has shown practically no regeneration since that time (Steven and Carlisle 1959) and is considered to be one of the most derelict and threatened pinewoods in the region (E. Cameron, pers. comm.) It is primarily those which were too inaccessible that have survived. Deforestation has continued into this century, with many of the pinewoods being felled during the two world wars (Anderson 1967, Steven and Carlisle 1959).

The Pinewood Resource Today

The native pinewoods of Scotland are currently restricted to 35 populations, and some scattered remnants covering a total of 11,000 ha (Goodier and Bunce 1977) (Fig 2.2). Individually, they range in size from less than 50 scattered trees (e.g. 29 at Glen Avon (Forrest 1980)) to large areas of forest such as Abernethy, with 160 hectares of dense woodland (Forster and Morris 1977). The total area of high forest in Scotland in 1988-89 was estimated as 1.1 million ha (Forestry
Commission 1989), of which 145,000 ha, or 14% is Scots pine (Forestry Commission 1984). The native pinewoods therefore represent 1% of the forest area and only 8% of the total area of Scots pine. Many of the remaining pinewoods are now designated as Sites of Special Scientific Interest (SSSIs), and many are designated as nationally important areas for conservation (Ratcliffe 1977).

CONSERVATION OF THE NATIVE PINEWOODS

The Protection of the Existing Native Pinewoods.

The future of the existing native pinewoods, despite the various statutory protection mechanisms, is by no means secure. Although the pinewoods are largely protected from clearfelling, regeneration in many of the pinewoods is scarce. The age-class distribution is often skewed towards the older age-classes, and the average lifespan of Scots pine is only 150-200 years (Goodier and Bunce 1977). Clearly regeneration is a priority in these stands and the major barriers to regeneration need to be identified. In National Nature Reserves (NNR) such as Beinn Eighe and Glen Tanar and the larger forestry estates such as Glen Affric there have been efforts to promote regeneration using fencing and direct sowing (Mc Vean 1961a, 1961b, 1963a, 1963b, Innes and Seal 1971, Fenton 1985, Gong et al. 1992). This has had varying degrees of success, but it appears that there is generally no shortage of available seed, and that given the right ground conditions, Scots pine can regenerate freely.

Deer

The major single factor preventing regeneration in native pinewoods today is the enormous population of red deer (Cervus elaphus) in Highland Scotland, currently estimated to number around 300,000. Many plant communities are under severe grazing pressure from red deer. As few of the native pinewoods are actively managed, it is uneconomic for most landowners to deer fence them, and as a result, regeneration is severely limited. In actively managed sites, competing ground vegetation is often removed to encourage seedling establishment, but this practice may make seedlings more obvious to browsing animals. Any increase in seedling numbers may therefore be offset by increased grazing, if the site is not also deer fenced (Gong et al. 1992). The Red Deer Commission (RDC) announced recently that statutory minimum culls would be imposed this year (1993) totalling 100K animals- double the previous year's total cull. If successful, this will go some way to minimising the problem, but it may require years of high
culling levels to significantly reduce population size. However, the Royal Society for the Protection of Birds (RSPB) began a programme of extensive red deer hind culling in 1990 on the Abernethy forest reserve in Strathspey. This is already resulting in a higher level of tree regeneration, suggesting that reducing the red deer population is an effective management measure (D. Dugan pers. comm.). Sheep and goats have been a problem in the past, but their impact is not now as great as that of red deer, though they may present a problem in some areas.

**Pollen Contamination**

A less tangible threat to the pinewoods, but no less real is the presence of a large amount of non-native pine planted in Scotland. The native pine populations of Scotland account for only 8% of the total pine population of Scotland (see above), and it is known that Scots pine seed was imported into this area before the end of the 18th Century (Lines 1964). Pine pollen is adapted for long distance dispersal and many of the seeds produced by native pinewoods may have been fertilised by pollen from non-native sources. This has been investigated in Swedish seed orchards and figures as high as 37.7% contamination have been estimated from a pollen source 800m distant (Nagasaka and Szmidt 1985). The only work done in native pinewoods in Scotland analysed progeny arrays from the native pinewood at Loch Maree. This analysis provides an estimate of the maternal genotype, allowing the gene frequency within the pollen cloud to be inferred, and found that the gene frequency of the pollen cloud was significantly different from that found in the seed bearing trees (Helgason and Ennos 1991). This is very indirect evidence and more work would need to be done to demonstrate if this was due to pollination from different provenances. Pollen contamination is a potential problem in conservation genetics for one main reason. Locally adapted populations mixing with a common, but less well adapted provenance, may result in the break up of co-adapted gene complexes, and result in the loss of specifically adapted traits. This has been identified as a particular problem in obligate outcrossers because as the species becomes rarer, the selection against inbreeding will encourage mating with the neighbouring, less well adapted populations (Ledig 1986, Ellstrand 1992). Pine is vulnerable to this because although it is not an obligate outcrosser, the selection against selfed progeny is high (see below).

**Inbreeding Depression**

Even if the pinewoods are successfully isolated from contaminating pollen sources, there is another problem which may arise due to the breeding system of Scots pine.
Increased levels of selfing in populations with reduced effective population size may seriously reduce their viability because of inbreeding depression (Ellstrand 1992). This has been measured in gymnosperms, including Scots pine, and characters such as size in selfed progeny can be reduced to as little as 27% of the mean value for outcrossed progeny (Charlesworth and Charlesworth 1987). However, the effect of increased selfing levels in native pine populations is more likely to result in a decrease in the availability of viable seed. Polyembryony (the presence of >1 embryo per ovule) is unusually common in gymnosperms and provides considerable opportunity for selection against selfed embryos after fertilisation, but before the maturation of the seed (Sorensen 1982). The high outcrossing rates measured in germinated seedlings of a native pine seed lot demonstrate that the number of selfed individuals recruited into the population will be small. However, if self-fertilisation levels are high, then the number of viable seed produced by a population may be greatly reduced, and regeneration therefore correspondingly difficult (Helgason and Ennos 1991).

The Extension of Native Pinewoods

There is now increasing recognition of the importance of the native pinewood community. The forestry industry, in the face of widespread public criticism of forestry practice and a subsequent change in the financing of forestry projects in 1988, has been seeking a more "sympathetic" kind of forestry, that satisfies ecological and aesthetic objectives in addition to the production of timber. One of the responses to this was the New Native Pinewoods Grant Scheme (NPGS) launched by the UK Forestry Commission in 1989. This marked a change in forestry policy, because the stated aim of the scheme was not simply to plant trees of a particular type, but to


This grant scheme aims to encourage growth of tree and shrub species typical of pinewoods as well as the Scots pine, producing an ecological community that can still be maintained and exploited. This change was partly motivated by the accumulated evidence that Scottish *Pinus sylvestris* was not only distinct but also better adapted to the Scottish environment (see below). Production forestry could therefore also benefit from an ecological approach. This approach has continued to gain momentum, with for example, the Cairngorms Working Party stating
"...we strongly believe that the re-establishment of the native Caledonian Forest should be a major priority." (Cairngorms Working Party 1992 para 3.2.4. p 12).

This initiative is enabling not only traditional foresters, but Non-Governmental Organisations (NGOs) such as the Royal Society for the Protection of Birds (RSPB) to establish forests in an ecological way with the level of financial support that has been available to traditional foresters for many years.

Genetic Conservation in Native Pinewoods

The NPGS addresses the next stage of native pinewood conservation, that is to promote regeneration of new pinewoods. One of the interesting features of the scheme is the emphasis that is placed on the conservation of the genetic base of the pinewoods. Central to any planting scheme aiming to promote natural regeneration is some degree of certainty that the regeneration is Scottish in origin. The NPGS explicitly recognises the existence of genetic structure and differentiation among populations of native and non-native pine. The need to separate native and non-native populations of pine is recognised, because there is the potential both for pollen contamination as described above, and also foreign seed being transported in. Secondly, the scheme recognises the importance of separating populations within Scotland. There are exclusion zones in the north-west and south-west where "...while modest transfer in the vicinity of boundaries may be acceptable, in general planting stock must derive from within the region." (Forestry Commission 1991).

This management strategy is, however, based upon the information currently available on the genetic structure of the native pinewoods. The first step towards effective genetic conservation is a formal evaluation of the existing populations (Ledig 1986, Melchior et al. 1986). Ideally, the pattern of geographic variation of genotypes and the genetic variability within and among populations should be known to enable informed decisions to be made. From the discussion in the previous chapter, it should be clear that the value of population genetic structure information depends upon being able to relate it to known ecological and historical parameters and the timescale in which that variation has arisen. Otherwise, conservation genetic management, particularly when it is pro-active as the NPGS is, may be influencing the populations in unforeseen ways. An extreme example of this was the attempt to conserve at least some of the genes of the dusky seaside sparrow, by mating the 6 remaining individuals, all males, to another sub-species that was considered to be the most closely related. However, subsequent
examination of the mtDNA of the genus showed that this conservation effort may have created an entirely new genome, as the two parental types were not in fact all that similar. Pro-active gene conservation efforts must therefore be based on as much information as is available, and where the gaps are, and the potential influence of those gaps must also be considered (Avise and Nelson 1989).

POPULATION STRUCTURE AND GENETICS OF P. SYLVESTRIS IN SCOTLAND

The information on native pinewoods has been built up from many different disciplines in an unstructured way. The generally accepted picture of the origins and genetic structure of native pinewoods is in no way conclusive, and it is important at this stage to review the evidence supporting it. On the one hand, there is evidence that Scottish pine is genetically distinct from other varieties, but more recently, evidence has emerged that suggests that Scottish pine is not a homogenous entity, and that there may be two distinct gene pools present in Scotland. While what governs peoples' perception of conservation value is outside the scope of this thesis, I want to emphasise that this perception is what has led to the interest in conserving native pinewoods, but that subsequent conservation policy can only be based upon what scientific information is available. The whole conservation policy of this species and the ecological community that it has come to represent is based on how it is thought to function and what its structure, ecological and genetic, is perceived to be. It is critical at this stage to review this evidence and assess how accurate it really is, because pro-active management could potentially make a big impact on these populations, and we have to be able to predict the consequences of management practices. This information is reviewed next, beginning with the existing data on genetic variation in Scottish pine populations from provenance trials and moving through palaeoecology, genetics and vegetation history, arriving at a synthesis of the current state of knowledge.

Provenance Trials of Scottish P. sylvestris.

Provenance studies that were set up between 1929 to 1951 by the UK Forestry Commission have investigated the performance of Scots pine of Scottish provenance collected from both estate plantings and native pine woodlands relative to pine of foreign provenance. The results suggest that Scottish pine of even partly native origin, such as those from older estate plantations that may be partly European in origin (Lines 1964) performs better overall in Scotland than foreign provenances. An assessment of the older pine provenance trials (Lines and
Mitchell 1964), showed that the Scottish provenances were superior in height increment to both central European and Fenno-Scandinavian provenances. The most striking differences were in survival rate, with many European provenances failing completely and Scottish provenances always showing at least adequate survival. These data, other Scots pine provenance trials and trials of other species of tree have since been reviewed (Worrell 1992), and a similar picture emerges, that native provenances show better overall adaptation to the Scottish environment.

Comparisons between Scottish provenances of Scots pine are more scarce, but patterns have emerged. The Inchnacardoch 92 (Highland Region) experiment planted between 1931-35 is on a site of moderate rainfall, and provenances from high rainfall sites planted there showed a marginally better height increment than those originating from low rainfall sites. Experimental design was poor on this plot, however, so this result should be treated with caution (Lines and Mitchell 1964). Provenances from high rainfall sites in Scotland including Loch Maree and Glen Garry showed poor survival in a provenance trial planted in Thetford, southeast England, an area with a more continental climate, when compared to plants from the more continental regions of Scotland (Lines and Mitchell 1964). On an exposed site in the north east of Scotland (Glenlivet in Grampian Region, planted 1954), trees from the Loch Rannoch area performed less well than more easterly provenances, although the difference was less marked at low elevation sites (Lines 1964). Disease resistance also varies significantly between Scottish provenances. Pine from Loch Maree, a high rainfall site, showed a significantly greater incidence of the pine stem rust *Peridermium pini* when compared to two eastern provenances when all three were grown at Teindland (Grampian, planted 1928-31), a low rainfall site (Greig and Sharpe 1991). Finally, on a very exposed, high elevation site at Glentrool, Dumfries and Galloway (planted 1953), assessments at six and ten years showed the west coast provenances to be generally superior to those from the central or eastern Highlands (Forestry Commission 1965).

*Genetic structure of P. sylvestris Populations*

Both monoterpene and isozyme markers have been investigated in native pinewood populations. The monoterpene composition of the Scottish populations has been determined by Forrest (1980, 1982). Forrest (1980) took 41 sets of samples from 32 native woodland sites representing the 8 regional groups throughout Scotland defined by Steven and Carlisle (1959). The percentage frequency of monoterpene types for each set of samples was determined, and analysed using a principal components analysis (PCA). Within population variation was found to be much
higher than among population variation, and the first three co-ordinates of the PCA analysis accounted for only 26% of the observed variation. The only consistent result from this analysis is that the north western populations in Scotland cluster into a distinct group. The distinguishing feature of this group is its low level of 3-carene (mean 23.8%). The western most population in the group, Sheildaig, has much the lowest 3-carene value, of 11%.

Genetic variability measured using isozyme markers gives similar results. As with monoterpenene data, the overall genetic variability is high. Values of among population differentiation (Gst) for Scandinavian and Eurasian populations of Scots pine are low, suggesting that over much of its range, Scots pine is relatively homogeneous. The small amount of data from central European provenances show values that are higher, and the Scottish populations are most similar to these (Table 2.1).

<table>
<thead>
<tr>
<th>Country of origin</th>
<th>No of Populations</th>
<th>No. of Loci</th>
<th>$H_T$</th>
<th>$G_{ST}$</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>9</td>
<td>11</td>
<td>0.31</td>
<td>0.02</td>
<td>Gullberg et al. (1985)*</td>
</tr>
<tr>
<td>Sweden</td>
<td>3</td>
<td>3</td>
<td>0.39</td>
<td>0.02</td>
<td>Rudin et al. (1975)</td>
</tr>
<tr>
<td>C. Europe</td>
<td>19</td>
<td>3</td>
<td>0.36</td>
<td>0.09</td>
<td>Mejnartowicz (1979)*</td>
</tr>
<tr>
<td>Poland seed orchard</td>
<td>12</td>
<td></td>
<td>0.36</td>
<td></td>
<td>Krzakowa et al. (1977)*</td>
</tr>
<tr>
<td>Scotland</td>
<td>14</td>
<td>16</td>
<td>0.31</td>
<td>0.03</td>
<td>Kinloch et al. (1986)</td>
</tr>
<tr>
<td>Scotland</td>
<td>40</td>
<td>5</td>
<td>0.35</td>
<td>0.05</td>
<td>Kinloch et al. (1986)</td>
</tr>
<tr>
<td>Sweden and China</td>
<td>3 each</td>
<td>12</td>
<td>0.247</td>
<td>0.003</td>
<td>Wang et al. (1991)</td>
</tr>
<tr>
<td>Finland</td>
<td>3</td>
<td>13</td>
<td>0.283</td>
<td>0.017</td>
<td>Muona and Harju (1989)</td>
</tr>
<tr>
<td>Sweden</td>
<td>3</td>
<td>14</td>
<td>0.277</td>
<td>0.006</td>
<td>Muona and Szmidt (1985)</td>
</tr>
<tr>
<td>Poland</td>
<td>2</td>
<td>7</td>
<td>0.422</td>
<td>0.069</td>
<td>Prus-Glowacki and Nowak-Bzowy (1989)</td>
</tr>
<tr>
<td>Sweden</td>
<td>3</td>
<td>11</td>
<td>0.269</td>
<td>0.004</td>
<td>Szmidt and Wang (1993)</td>
</tr>
<tr>
<td>Turkey</td>
<td>4</td>
<td>11</td>
<td>0.304</td>
<td>0.007</td>
<td>Szmidt and Wang (1993)</td>
</tr>
<tr>
<td>China, var mongolica</td>
<td>4</td>
<td>11</td>
<td>0.237</td>
<td>0.008</td>
<td>Szmidt and Wang (1993)</td>
</tr>
</tbody>
</table>

Table 2.1. Published genetic diversity statistics for populations of Scots pine. *cited in Kinloch et al. (1986)
Within Scotland, the North Western population at Sheildaig is significantly different from all the rest, but this appears to be due to relatively few differences in the isozyme frequencies. Only one locus shows a strikingly different allele frequency from the rest. One allele of ADH2 has a frequency of 0.711 in the Sheildaig population where the mean is 0.428. Sample sizes are small in this study (13-37 alleles sampled per locus per population) compared to many other studies which limits the significance we can place on these data (Kinloch et al. 1986).

THE RECONSTRUCTION OF THE VEGETATION HISTORY OF SCOTTISH PINE POPULATIONS.

The Traditional Story

The reconstruction of the Holocene (Flandrian) post-glacial history of Scots pine in Scotland was originally based largely on pollen evidence, because Scots pine pollen is so well preserved and easy to recognise. The consensus is that described by Birks (1989). This description is based on the information from 135 published pollen diagrams, which have been combined to form "isochrone" maps of pollen concentrations (Fig 2.3). After the last ice receded, Scots pine spread into the southern British Isles at two different points. Southern England was colonised before 9000BP, and pine was well established in south west Ireland by 8800BP. These populations continued to spread northwards, spreading at rates of 100-700myr\(^{-1}\). The Irish population was probably the source of the pine populations of south west Scotland, with migration over the Irish Sea the most likely route. The populations migrating from the south east on the other hand never reached Scotland. There is a conspicuous gap in the pollen record of Scots pine between Northumberland and the Forth-Clyde valley, suggesting that Scots pine was never dominant in this region (Godwin 1975b, Huntley and Birks 1983).

Neither of these populations appears to have been connected to the Scottish populations which expanded independently at 8500BP in the Wester Ross area of Scotland. This expansion occurred 1500 years before pine appears as a significant component of the pollen record in south west Scotland. Pine expanded very rapidly to become dominant in the east of Scotland at about 7500BP, where the bulk of the populations persist today. The distribution of pine reached the maximum extent of its range in the Holocene at 5000BP, reaching from the north of Scotland to the Forth-Clyde valley. From this maximum, pine underwent a severe decline just before 4000BP. The climate of Scotland became more oceanic
Fig 2.3 Isochrone map of the rational limits of *Pinus sylvestris* pollen in the British Isles. Dates are shown as radiocarbon yrsBP. Arrows show the possible entry points of pine into the British Isles. After Birks (1989).

At this time, the range of pine contracted in the south and west, probably due to blanket peat formation (Birks 1975, Gear and Huntley 1991). Birks (1975) suggests that the activity of Neolithic man, who was now present in the region may have exacerbated this process, by impeding regeneration. Pine however, has persisted as the dominant species in the eastern Highlands, where the climate has remained more continental.
Biochemical Markers of Population Differentiation

A different approach to the reconstruction of vegetation history is to look at the relationship between extant populations to see how closely related they are and if there are any patterns within and among populations. It has emerged from the measurement of cortical monoterpane composition in pines that high 3-carene levels are characteristic of the Eurasian "mountain" species of Pinus, and that in the Mediterranean pines, it is low or absent (Mirov 1967). Further study on provenances of P. sylvestris have shown that this pattern exists also within the species, with Mediterranean populations of P. sylvestris having much lower percentages of 3-carene (Tobolski and Hanover 1971). Swedish populations of P. sylvestris have shown mean values of 3-carene between 30-34% (Yazdani 1986). In another study, Yazdani and Nilsson (1986) found strong clinal variation in 3-carene values, with southern populations ranging in value from 49.4-55.7%, whereas the two populations located in the far north of Sweden (68°N), had values of 21.3-22.4%. Yazdani et al. (1985b) also found that northern populations in Sweden had lower 3-carene values on average than southern populations. The study by Tobolski and Hanover (1971) investigated 4 Scottish samples, 2 each from the west and east. It is unfortunately not specified where these came from, as all samples had a high 3-carene value, which led the authors to suggest that Scottish populations were more closely related to northern European varieties. This is at odds with Ruby and Wrights' (1976) classification in which var. scotica was placed with the southern Pleistocene relict provenances. This apparent contradiction is all the more interesting as both conclusions were arrived at using data from the same provenance trial. The extensive survey of monoterpane composition of Scottish pine populations suggests that the north western populations of pine are more closely related to pine populations of south west Europe, and the rest have closer affinities to central European provenances of Scots pine (Forrest 1982). Unfortunately, no direct comparison of isozyme composition has been made between Scottish P. sylvestris and European varieties, so it is not possible to make any inferences about relatedness based on isozymes.

A REVIEW OF THE EVIDENCE

Critique of Palynological evidence

Methods for analysing pollen grains from sediment and peat layers have been described in detail elsewhere (Faegri and Iversen 1964, Lowe and Walker 1984). The technique has proved to be useful, but there are a number of specific problems
associated with it which must be considered when interpreting published pollen data.

**Methodological Considerations**

Counting absolute concentrations of pollen is difficult because the volume from which the concentration is calculated depends upon the rate of deposition and compression of the sediment. It is more common to find pollen counts expressed as percentages either of arboreal (AP) or land (LP) pollen. The percentage value is based on a predetermined number of pollen grains counted per sample, the pollen sum. This is usually between 300-500, but is sometimes less.

Percentage pollen diagrams may have the effect of distorting patterns of species abundance. If, for example, there is a sudden increase in the pollen frequency of one species, the percentage abundance value of another species will be reduced, even though the absolute pollen concentration of that species has not changed (Lowe and Walker 1984).

In addition to the representation of the data, the sample sizes used are often very small. Pollen analysis is an enormously time consuming procedure, and this prevents many projects from considering larger numbers of cores. Many published pollen diagrams of regional pollen assemblages, are based on data from one or two cores (e.g. Pennington et al. 1972, Birks 1970). However, spatial variation in pollen frequency over relatively short distances, and hence any standard deviation placed upon those values can be very high (J. McConnell pers. comm.). Turner and Hodgson (1991) published data from 46 cores taken within a 100km by 60km area. These showed pine pollen frequencies ranging from 0.26-31.27%.

**Interpretation of Pollen Diagrams**

It is important to consider the following properties of pollen before interpreting pollen diagrams:

- All plants do not produce the same amount of pollen. This leads to the over representation of high pollen producers, most of which are trees. Faegri and Iversen (1964) give the average pollen production characteristics of most of the common pollen types.

- Pollen can be transported in different ways. Although tree pollen is usually adapted to wind transport, this not always the case. In cores taken from lake sediments, much of the pollen may be washed into the lake from upstream (Birks 1970). Pine pollen in particular is adapted to wind transport. The
pollen grain has two air sacs which enable these grains to be transported immense distances. In moss polsters taken from tundra in North America, pine pollen, (probably *P. contorta.*) accounted for 2% of AP, when the nearest population of that species was at least 150km distant (Birks 1977). A value of at least 5% AP is usually considered to be the minimum value indicating local presence of pine.

- The pollen grains of different species have different characteristics, and settle in the sediment and peat at different rates. Lacustrine sediments may therefore have different pollen assemblage characteristics than peat (Birks 1970), and in soils there is the additional problem of mixing, for example by earthworms.

- Differential representation of species occurs because some are more vulnerable to decay than others. For example juniper is under-represented in Scottish pollen profiles (J. McConnell pers. comm.), and plants that have delicate pollen types like *Urtica* and *Populus*, may be absent altogether (Lowe and Walker 1984).

Pollen diagrams are often used, particularly in the context of vegetation history, to reconstruct what the vegetation was like at that time. The most important question about pollen diagrams therefore, is how closely do they reflect the climate and vegetation that was actually present? There are many species that, for the reasons outlined above do not appear in pollen diagrams, or are not adequately represented. The environment reflected by the pollen diagram may for this reason be localised, and extrapolations will often be unreliable indicators at a regional level.

The problems with extrapolating pollen data are illustrated by the pollen and macrofossil profile of Loch Kinord, Aberdeenshire (Vasari and Vasari 1968). These data consist of two sediment cores that are not included in Birks' (1989) review. There are pine needles present in one of the two cores from this site that date to the Allerød late-glacial interstadial, or 11-12000BP. Birks (1970), in a pollen core from Abernethy forest (about 55km or 35 miles from Loch Kinord) records a 55%AP value of pine pollen from the same pollen assemblage zone. Despite this, the author is not prepared to suggest a local presence of pine in the region at that time. The pollen diagrams for Loch Kinord also show a marked increase in pine pollen to 40-60% AP at about 8000BP, again a much earlier presence of pine in the region than is suggested by Birks (1989) or Birks (1970). The increase in pine pollen at Loch Kinord is concurrent with a marked increase in pine macrofossils, including large quantities of bark, which is highly unlikely to be
brought in by long distance transport from overseas. This suggests at the very least a local presence of pine in the east of Scotland more than 1000 years before it is usually thought to have arrived and possibly 3000 years before that.

**Critique of Genetic Analyses**

Many of the technical difficulties associated with monoterpane and isozyme analysis have been dealt with previously, and need not be reiterated here, except where it specifically influences the analysis of population structure in Scots pine. The monoterpane analysis of the Scots pine populations clearly indicates that there is a low 3-carene type present in the north west, and the isozyme data, apart from confirming the atypical nature of the Sheildaig population, is inconclusive in determining whether there is any evidence for a refugial origin of Scots pine. However, doubt has been cast upon the ability of this method to distinguish between populations or even groups of populations because of the high variation within populations (Kinloch et al. 1986). In addition, there is also circumstantial evidence in pine that variation in monoterpane composition may be under selection, and therefore an unreliable marker for population movement. For the purpose of this analysis, the assumption has to be made that the pattern of genotype frequencies is the result of random drift and migration, and not due to differentiation that has occurred due to geographically varying selection differentials. This may be the case in some populations, but on balance it is probably not a valid assumption. It is perhaps intuitively obvious that resins in pines may act as part of the individual's defence against pests and pathogens. Isolates of *Armillaria mellea* taken from Scots pine were found to have substantially reduced growth rates when grown in the presence of oleoresin vapour, circumstantial evidence that monoterpane composition may be changed by natural selection in Scots pine (Rishbeth 1972). In ponderosa pine (*P. ponderosa*), a history of predation by the bark beetle *Dendroctonus brevicomis* is highly correlated with the absence of limonene which is toxic to the beetle, and with high concentrations of α-pinene which is a possible precursor of the beetle's pheromone compounds (Sturgeon 1979). The level of α-pinene in ponderosa pine has also been found to be the single best predictor of food tree selection by tassel-eared squirrels (*Sciurus aberti*), and although this data is based on absolute amounts of monoterpenes rather than relative composition, subsequent research found that food trees had significantly lower percentages of β-phellandrene and β-pinene suggesting a potential selection differential for monoterpane composition (Farentinos et al. 1981, Snyder 1992). In ponderosa pine at least, the distribution
of monoterpane genotypes is unlikely to be a result of historical patterns of migration, isolation or genetic drift, and interpretations based on this assumption would be highly misleading.

Monoterpane and isozyme markers fail to discriminate between populations because differentiation among populations is low. The high dispersal ability and outcrossing rates typical of Scots pine are responsible for this (see above). The selection against selfing is so strong that there is an advantage to be gained by an individual in dispersing its genes as far afield as possible, to minimise the likelihood of fertilising a close relative. Gene flow via pollen is significant in Scots pine because the pollen is so well adapted to long distance dispersal. Effective population sizes are probably relatively high, suggesting that in natural populations of Scots pine, the potential for significant population differentiation in the nuclear genome is low, especially for neutral variation.

The Invasion of *P. sylvestris* into Scotland: A Review of the Hypotheses.

The accumulation of data, genetic and palynological, suggests that there is something atypical about the history and genetic structure of the Scottish pine populations. On the basis of what was then known about the post-glacial history of Scotland, particularly with respect to land bridges between Britain and the continent, Anderson (1967) suggested that the earliest invasion of forest into Scotland must have been along the east coast. The morphological data collated by Ruby and Wright (1971) suggest that Scottish *P. sylvestris* is a Pleistocene relict population, related to the populations of southern Europe. The pollen evidence however, suggested initially that pine expanded in Scotland from a source in the north-west, possibly a glacial refugium. This hypothesis was suggested by Godwin as early as 1975 when referring to the gap between England and Scotland in the pollen record he stated,

"...if this gap can be shown to be real it will strengthen the case for considering that the tree reinvaded from the northwest as well as from the south, a fact possibly to be associated with the separate taxonomic status of the Scottish native pine as *P. sylvestris var. scotica*" (Godwin 1975a, p110).

This has lead to a great deal of discussion about the origin of Scots pine in Scotland, but both pollen and genetic studies have only really confirmed that there are indeed real questions to be answered. Huntley and Birks (1983) also suggested the possibility of a glacial refugium in the north west. Kinloch *et al.* (1986) discuss this at length, and acknowledge that although the glacial refugium
hypothesis is improbable, they consider it less unlikely than long distance dispersal. Birks (1989) states that there were at least two origins of pine in the British Isles, and possibly three if there is an endemic origin of pine in Scotland. He also agrees that while the glacial refugium hypothesis is improbable, the Holocene history of pine was clearly different in the north-west and east but concludes that there are no firm insights into the reasons for this.

The difficulty of actually detecting population differentiation in pine using monoterpane and isozyme genetic markers makes the distinctiveness of the north-west populations all the more striking, particularly in view of the monoterpane affinity of these populations with the relict populations of south western Europe. This could indeed be suggesting a relict refugium origin of the north-west populations if these markers could be demonstrated to be selectively neutral. Among population differentiation within the north-west group is also much higher than that of other groups in Scotland, which might indicate that it is a relict of a larger population rather than a founder population. What the genetic data does not suggest, however, is that this refuge population expanded to form the entire population in Scotland. If this was the case, this group of populations would not be so distinct. This still leaves the question of the origin of the remaining Scots pine unanswered.

The consensus is that the evidence suggests that Scots pine in Scotland originated from at least two sources, and that if there was an endemic origin, it would have come from a refugium in the north west. The real migration routes of pine into Scotland would explain the current distribution of genetic variability, so it is important at this stage to look at the available evidence from a biogeographical viewpoint, and see if any of the possibilities match the observed patterns. The possible routes into Scotland of Scots pine are reviewed below.

Migration by Diffusion

Migration by diffusion is defined as the the gradual movement over hospitable terrain of individuals of a species (Pielou 1979). Although this is the intuitively obvious route for the migration of pine into Scotland, it is simply not supported by the data (see above), because of the early appearance of pine in the north west, and possibly also in the east. If this is real and not a misrepresentation, then migration by diffusion cannot have occurred because by definition, the pine could not have reached Scotland so soon after the last ice receded.
The dispersal of Scots pine in this way would be influenced primarily by seed dispersal and the availability of suitable habitat. The latter would be unlikely to be a problem. The soils left behind after glaciation are ideal for pine, as it performs best on freely draining gravelly soils in the absence of competition. Scots pine seeds are adapted for wind dispersal, but most seedlings are found within 100m of the parent tree (McVean 1963c). Individual saplings have been recorded at much greater distances than this and distances of 2km (McVean 1963c), about 800m (Smith 1900) and 500-600m (Phillips 1989) from the nearest living tree have been recorded. These figures suggest that the rates of spread of pine in Britain of 100-700myr\(^{-1}\) given by Birks (1989) may be an overestimate, given that a pine seedling will not produce seed until it is 5-10yrs old (Krugman and Jenkinson 1974). Since the evidence suggests that pine appeared in Scotland within a few hundred years of its appearance in the south of the British Isles, migration would require diffusion rates an order of magnitude higher than those suggested by Birks (1989). The alternative is that the pine moved very rapidly up the north west coast, with populations remaining localised and small, to expand once it reached a suitable habitat. However, extensive searches for pine macrofossils in the western islands failed to reveal any evidence for pine in these locations, suggesting this is an unlikely route (Bennet 1984, Bennet and Lamb 1984). Forrest (1982) suggests that the north western pines may have migrated northwards along this route, but at a much earlier time, during the late-glacial interstadial (see below), that is before the Loch Lomond readvance. It is possible that any evidence, at least on mainland sites, of such a migration would have been eradicated by the subsequent glaciation. If this is the case it certainly suggests a dual origin of pine in Scotland, and would be consistent with the apparent similarity of north west pines to the southern varieties.

"Jump" Dispersal

The alternative to diffusion is jump dispersal, defined by Pielou (1979) as the movement of individual organisms over great distances within a very short space of time, usually over inhospitable terrain. The example given is spiders being transported over oceans. For Scots pine, we have to consider the likelihood of long-distance seed transport. Although this theory has been criticised for its improbability, the point made by Simpson (1952) is that the probability of one event succeeding increases with each attempt. So, given the right conditions, you could expect a number of seeds or cones to travel a long distance and a few of them to actually establish. In the case of Scots pine, the possibility of rapid long-
distance dispersal over snow has been suggested (Phillips 1989). The critical factor is time. It is an interesting but unresolved (unresolvable?) question how this might have occurred, and if there was enough time between the retreat of the ice and the increase in pine pollen frequency for enough attempts to be made. There is some indication that plants, including pine, travelled across from Northern Ireland to the Galloway region this way (Birks 1989) and perhaps pine seeds are able to spread further. Unfortunately it is impossible to test this hypothesis.

Glacial Refugia

The final possibility is that pine survived the glaciation in one or more refugia, which enabled it to very quickly exploit the available niches when the ice receded. To assess the likelihood of this scenario, the late Weichselian (late Devensian) and Holocene (Flandrian) history of Scotland must be reviewed. Briefly, the late Weichselian ice sheet reached its maximum in the south of England around 18000BP, with little land remaining ice-free. In Scotland the main ice-caps were centred upon the Rannoch Basin, the eastern Highlands and the Southern Uplands. There were ice caps on the Western Isles, though whether they were part of or separate from the main Scottish ice cap is disputed (Sutherland 1984). After 14000BP the ice began to recede, leaving most of Britain ice-free during the late glacial interstadial 14-11000BP. The climate cooled again at this point, causing the Loch Lomond readvance, a period of 1000 years when the main Scottish ice cap, which was centred upon the Rannoch moor area expanded again to cover most of north west Scotland from Wester Ross to the Clyde valley (see fig in Lowe and Walker p341).

Those authors that have invoked the glacial refugium theory have focussed upon the north west as the most likely region (Godwin 1975a, Kinloch et al. 1986, Huntley and Birks 1983). Likely sites of a refugium in the north west are the islands off the mainland and at the margin of the ice-sheet. There are sites which are considered to have been ice-free during the Devensian period. Pollen analyses from three interstadial sites in the Western Isles, Tolsta head (27,300BP) (von Weymarn and Edwards 1973) and Toa Galston on the Hebrides (>40,000BP) (Sutherland and Walker 1984), and St. Kilda (Sutherland et al. 1982), suggest that at these times the vegetation was largely treeless. The %AP values for pine are less than 5% in all cases, suggesting that pine was not a component of the vegetation in these regions during the last glaciation. Holocene (Flandrian) pollen profiles from St. Kilda show no evidence of local presence of pine at any time since the last glaciation, which again is inconsistent with the idea that this was the
region from which pine spread (McVean 1961c, Walker 1984). If the north west had been a refugium, there is not as yet any firm evidence supporting it.

All this speculation from workers in vegetation history neglects one fundamental confounding factor, and that is that there is as little consensus about the margins of the Scottish ice-sheet as there is about the origins of Scots pine. Ten years ago, the consensus was that the Scottish ice-sheet was deflected to the north-west by the Scandinavian ice-sheet over Shetland, and took some time to de-glaciate. In addition, the ice on the Hebrides was thought to be part of the mainland Scottish ice-sheet, an idea that has been questioned by Sutherland (1984). Sissons (1984) in a deliberately speculative paper suggests that because of climatic factors, the maximum northern and southern limits of the ice sheets may have occurred at different times. The northern margin would have been less active than the southern one because it was not being nourished by precipitation, while the southern one was nourished by precipitation from the polar front. He further suggests that at the time the southern parts of the ice-sheet were at their maximum, the northern margin may even have been in retreat. Consequently, the deglaciation in this region may have been very rapid. In the conclusions, he suggests that anomalies in the deglaciation patterns can be resolved to a large extent by assuming that Shetland nourished its own ice cap, the Scandinavian and Scottish ice sheets did not meet, and thus a lot of Caithness, and the north east of Scotland may have been covered by "cold-based" ice that was inactive.

On the mainland, the movement of ice is thought to have changed from southerly to northerly as the ice sheet centred on the Southern Uplands became more dominant than the more northerly ice sheet centred on the Rannoch moor. Sutherland (1984) implies that this, and the presence of ice free areas in northern Scotland is evidence that the northern ice-caps were relatively inactive during the latter stages of glaciation. This ice would have downwasted relatively quickly, and this area may have become ice-free early in the deglacial cycle (Stewart 1991), suggesting that either a north east refugium or an early migration of pine into the east is more likely than a north westerly one. I consider this in some detail, because it demonstrates that interpretation of vegetation patterns based on the pattern of deglaciation depends on which account is considered.

That the north east may have deglaciated earlier than previously thought is particularly interesting in view of the very early appearance (c. 11000BP) of pine macrofossils at Loch Kinord (Vasari and Vasari 1968). It suggests that the north
west, if it was a refugium, may not have been the only one and that further investigation of the eastern Highlands would be worthwhile.

Summary.

Evidence for the genetic differentiation of var. *scotica* compared to other varieties of Scots pine comes from three different sources.

- **Provenance trials** have shown that var. *scotica* is distinct, and there is a suggestion that there is adaptive differentiation across Scotland's rainfall gradient. In addition, var. *scotica* shows mixed affinities with other varieties, with some traits most like Southern varieties, and others resembling European and Scandinavian ones.

- **Palaeoecological studies** suggest that var. *scotica* has been separated from all the other populations of pine for at least 6000 years. A mode of migration other than diffusion from Southern glacial refugia is inferred from the pattern of *Pinus* expansion in Scotland, with the existence of a north west refugium the most frequent hypothesis. However, the evidence is contradictory, and glacial geomorphology studies tend to suggest that refugia could also have existed in the north and east of Scotland.

- **Population genetic studies** using monoterpene and isozyme markers also provide evidence for the presence of two gene pools in var. *scotica*. Monoterpene compositions within populations also suggest an affinity between north-west Scottish populations and south European varieties because of the very low values of 3-carene found in these populations, and the presence of a north-west refugium or of very early post-glacial expansion from the south west has been suggested. The remaining populations show affinities to central European and Scandinavian varieties. North western populations also have distinctive isozyme allele frequencies. Overall genetic variability has not been reduced despite severe depletion and fragmentation of populations in historical times. Differentiation among populations, while not the highest recorded, is still high.

The data from all these sources, considered together, point to the existence of two gene pools within var. *scotica*. Despite the mass of data however, the origin of the two gene pools is obscure, as is the distribution of the two types within var. *scotica*, except at a very basic level. The information available in the genetic structure of native pine populations comes from disparate sources and is often contradictory. A reliable marker for discriminating between non-Scottish and
Scottish pines is not available, nor can the Scottish ecotypes be distinguished without long-term investigation because of the high phenotypic and genetic variability still present in these populations. Despite this, the NPGS management guidelines are based on the available evidence, with potentially profound effects on the real genetic structure of the pinewoods.

This situation presented an ideal opportunity to determine whether it was possible to find a reliable set of markers that could on the one hand be used as a practical tool for monitoring and managing the genetic composition of var. scotica, and on the other to resolve the remaining ambiguities about the origin of the populations in Scotland.

THE ALTERNATIVE APPROACH: MOLECULAR MARKERS

Genetic Conservation in Scots Pine.

There is no single ideal method for determining genetic variation and structure in populations. Each method available has particular technical advantages and disadvantage associated with it and in addition each method has a different resolution, both in terms of the type of variation and the timescale in which it arises (Schaal et al. 1991, chapter 1). In Scots pine, phenotypic variability within populations is so high that it is not easy to use quantitative characters as field markers for populations. It is clear that pine shows genetic differentiation at the population level but monoterpenes and allozyme markers have also proved to apportion most variation within populations. Very large sample sizes are therefore required to distinguish groups of individuals, and in any case, this information may not be reflecting historical relationships but differentiation due to selection. Assigning a particular sample to a region depends upon frequencies of types because all types are present in all populations. It is not possible therefore to assign any individual to a particular group or population. The logical step forward was to investigate the possibility that molecular markers could be developed that would perform these tasks.

The development of genetic markers that (a) are specific to Scottish pine, and (b) enable individuals to be assigned to a particular region in Scotland would greatly aid the management of gene conservation, particularly where there has been such a large introduction of non-native genetic stock. For practical identification purposes, the marker would at least have to distinguish Scottish from non-Scottish pine. In addition, a marker that could discriminate between the two founder gene pools that are thought to be present in Scotland, would enable the distribution of
the two types and the way in which they are mixing to be determined. This would allow

- seed lots to be monitored for genetic origin,
- reliable separation distances between different populations to be established,
- seed stands to be certified as genuine- which would put an "added value" onto existing stands and therefore encourage conservation measures.

Properties of the Marker

The DNA marker described above that would be most likely to fulfil the above criteria would have to be highly conserved. Hypervariable sequences were rejected because they are specific to an individual, and evolve so rapidly that any historical difference in genotype would be obscured. Ribosomal DNA was rejected on the grounds that determining the maternal and paternal contribution to the genome would be difficult because of Mendelian inheritance and recombination. Mitochondrial DNA markers, would have been ideal because they are maternally inherited and seed transport distances are not as great as pollen dispersal. There is considerable opportunity for population sub-structuring with such a marker. However, the protocols were not well developed for investigating mtDNA in conifers, and though it would be possible to detect individuals descended from non-native seed bearing trees, the amount of pollen contamination of native seed bearing trees cannot be determined. Recent reports of population surveys of mtDNA variation in several species of North American Pinus species suggest also that the rapid rate of structural rearrangement causes a high degree of convergence, or homoplasy in mtDNA haplotypes, making such markers unsuitable for phylogenetic studies, and possibly also for a study such as this, where differentiation of ancient origin is to be detected (Strauss et al. 1993, Dong and Wagner 1993).

The cpDNA of conifers is subject to a higher level of rearrangement than is found in angiosperms, and the possibility existed therefore, that there might be a highly conserved, but intra-specific marker that would discriminate between varieties of P. sylvestris. Although cpDNA has mainly been used for phylogenetic studies, the point has been made that cpDNA variation might be expected in any taxon with a wide distribution, if enough samples are screened (Neale et al. 1988, Brunsfield et al. 1992). If intra-specific markers existed, the paternal inheritance of the cpDNA genome would allow for simple monitoring of pollen contamination levels in regeneration and seedlots. In order to quantify gene flow by pollen using a maternally inherited organelle genome, the identity of the pollen parent has to be
inferred from a comparison with nuclear markers (Arnold et al. 1992). CpDNA thus has a considerable advantage over mtDNA markers. In addition the protocols for investigating cpDNA in conifers were well developed by other workers by 1988 (Wagner et al. 1987, White 1986) potentially making it easier to begin such work in a new laboratory. This point is illustrated by Palmer (1990), who notes that thousands of cpDNAs have been investigated, but only 14 mtDNAs.

For the above reasons, and especially the fact that cpDNA extraction methods were the best developed, the investigation of chloroplast DNA variation was decided upon for this project. The following chapter describes the protocols that were used, and how they were adapted for this particular situation.
CHAPTER 3

RFLP ANALYSIS OF THE CHLOROPLAST DNA OF PINUS SYLVESTRIS.

Introduction

To resolve the ambiguities about the origin and population structure of native Scottish pine populations, chloroplast DNA was selected as the marker most likely to have the appropriate properties for this purpose. The technique used for detecting variation in cpDNA was the analysis of Restriction Fragment Length Polymorphisms (RFLPs). This is a well developed technique, and has been successfully carried out on many angiosperm species, and an increasing number of gymnosperms. RFLP analysis requires many laboratory protocols, and the methods developed in this laboratory for the study of Scots pine are the subject of this chapter. An overview of RFLP analysis is given, followed by detailed descriptions of the protocols tested and those which were finally selected for the population survey.

RFLP Analysis in Plant Populations.

Basically, DNA is extracted from the organism of interest, and cut into lengths using restriction endonucleases. The resulting fragments are separated by running the digested sample through an agarose gel with a voltage across it. Pure extracts of organelle DNA can be detected at this stage, but the separated fragments are usually transferred to a membrane so that they can be manipulated more easily. A previously cloned piece of DNA from the target region of the genome called a "probe", is labelled with an isotopic or immunological marker, and this is then hybridised to the DNA immobilised on the membrane. The marker produces a signal which is detected on X-ray film. When this analysis is being tested on a plant species for the first time, many decisions on the best approach to make have to be made. The basic procedure is outlined in Fig 3.1, showing the critical points, which are described in more detail below.

DNA Extraction

There are two basic approaches to the extraction of cpDNA from plant tissue (A). Once a crude extract of cell contents has been made, total DNA can be extracted from the sample, and you then rely upon the lack of homology between nuclear
Fig 3.1  Restriction Fragment Length Polymorphism analysis of DNA. See text for discussion. Labels correspond to descriptions in text.
and cpDNA genomes to allow you to probe the DNA with a cpDNA specific probe (see below). Alternatively, the chloroplasts can be extracted from the tissue and then lysed to produce a pure extract of cpDNA that can then be run on a gel and detected directly, or by using Southern hybridisation and probing with labelled probes.

**Restriction Enzymes**

Restriction endonucleases are one of the key discoveries that have made DNA analysis possible. Three types of Restriction enzymes are known, but only the typeII endonucleases, those that always cut the DNA in the same place need concern us here. These are endonucleases extracted from bacteria that recognise specific base pair sequences and cleave the DNA at that point. For example, *HindIII* recognises the sequence

\[ 5' \text{A}\text{AGCTT}3' \]

The enzyme cuts the complementary sequences at the arrow, producing staggered ends called "sticky ends". Some enzymes cut at the same base pair producing "blunt ends". This particular example is a palindromic sequence, reading the same sequence in either direction. This is common, but not universal in restriction enzymes. These enzymes allow two types of variation to be detected:

- **Nucleotide substitution** If the base sequence of the DNA changes so as to create a new restriction site or to remove an old one, this will change the number of fragments

- **Length Variation** The insertion or deletion of a sequence of DNA will cause a change in the length of that restriction fragment.

Purified DNA extracts are incubated with enzyme in the presence of the appropriate ions to produce DNA fragments. The fragment patterns produced are repeatable, allowing comparisons between samples (B).

**Detection of RFLP Variation**

The fragments generated by restriction digestion are run through an agarose gel to separate them according to their length (C). If the fragments from a purified organelle extract are pure enough then they can be directly measured from an agarose gel, but it is more common to use DNA probes. Probes are fragments of DNA that have been inserted into vectors, cloned and then extracted and purified (D). The DNA target fragments are fixed onto a solid support as denatured single strands (E), and under the right conditions, single stranded probe DNA will
hybridise to the strands of DNA with the most similar or "homologous" sequences. The degree of hybridisation depends on the homology of the DNA (F). The probe DNA is labelled using radioisotopes or immunological labels (in this case Digoxigenin, DIG in Fig. 3.1) so that once it has hybridised it can be detected, and the patterns recorded (G). The method used for labelling determines which detection method is employed (H).

This chapter describes the methods I used to estimate RFLP variation in the chloroplast genome of natural populations of Scots pine. Some of them proved to be inadequate, for reasons which I have tried to identify. Firstly, I describe the techniques in some detail, and then describe how they performed in this laboratory. All modified protocols used in this study are listed in Appendix 1. They are then compared, and a summing up is given of how to approach this in other populations.

DNA EXTRACTION AND MANIPULATION

Sample Collection and Maintenance

Many workers have stated that the key to successful DNA extraction is the condition of the material. Palmer (1986) states that it is essential that the material should be the freshest and healthiest available and Robinson and Barnett (1988) state that for the successful isolation of chloroplasts, plant growth must be carefully regulated. Much of the earliest work on the molecular genetic analysis of populations was done on seedlots collected from the wild (e.g. Banks and Birky 1985) or crop species (eg Timothy et al. 1979) which can be easily grown up in controlled conditions and at optimal nutrient regimes, thus ensuring the best quality samples. Those who work on natural populations however, often do not have this luxury. Despite this, protocols have been developed that overcome this and the extraction of DNA from samples collected in the field is possible. However, there are two main problems:

- Collection and Storage. If, as is often the case, the populations of interest are not easily accessed from the laboratory, expeditions or field trips must be organised to collect the samples. The decision whether to collect seeds or tissue samples must be made at this stage. Seeds have the advantage that their growth can be controlled in the lab, but the amount of tissue that the seedlings yield is often small, and with many plants, seed may either be difficult to find and germinate, or too slow to grow. Tissue samples can be collected in the quantity required, but these then have to be transported to the lab without the
material deteriorating significantly. Because all the samples required have to be collected at once, they then have to be stored prior to DNA extraction in a form that does not damage the DNA. Fresh leaf samples can be kept at 4°C, they can be frozen and kept at -20°C or -70°C, or they can be freeze dried. The method of storage may influence the subsequent performance of a particular DNA extraction method.

- Sample Quality  The condition of material collected from natural populations is often not of the best quality. The selection of plants from such populations is usually a compromise between the quality and the ease of obtaining it. This is a particular problem with tree species because they are often tall, and self-thinning, so that the tissue you can reach will be from the lowest part of the crown and not the healthiest part of the plant. The time of year that the samples are collected can be important. I found that Scots pine samples collected during the winter months when they are hardened off worked best for the CTAB extraction (see below), and other workers have reported similar observations (D. Curnow, A. Sigurgeirsson, pers. comm.). In particular, they reported excessive amounts of starch being present in actively expanding leaves and needles, causing DNA purification to be particularly difficult.

DNA Extraction Techniques

There are many protocols for the extraction of DNA from plant tissue (e.g. Murray and Thompson 1980, Kreike 1990, Robinson and Barnett 1988, Milligan 1989). The extraction of DNA must be adapted for use on each new species. Plant species differ widely in properties such as the pH of the cytoplasm, amount and type of secondary metabolites, and metabolic activity resulting in the production of starch, all of which will affect the DNA extraction. For investigation of populations, large sample sizes are often required, and it is important in the first instance to determine the value of a particular system. The worker investigating a species for the first time must be prepared to modify a protocol to improve its performance. To do this effectively, the purpose of each step in the procedure must be understood so that the modification process is less haphazard. Three methods of DNA extraction that were tried in this laboratory are described below. Each is based upon a published protocol, but all have been modified, either by myself or another worker.
**CTAB Extraction**

This method was adapted from Murray and Thompson (1980) (R. Waugh, pers. comm.). Tissue ground after being flash frozen in liquid nitrogen was incubated at 65°C in a Tris/ETDA buffer containing 1% w/v CTAB detergent (Mixed Alkyl-tri-methylammonium-Bromide, Sigma). This detergent ruptures cell walls, and total DNA comes out in solution. The solution was purified using Chloroform:IAA extraction, and the DNA precipitated using Isopropanol or Ethanol. RNA is also present in these preparations which can be removed by incubation with RNase A.

**Chloroplast Extraction using Buffer of High Ionic Strength**

This method was first described by Bookjans et al. (1984). To reduce the number of differential centrifugation steps required to remove all the nuclear DNA from the preparation, this method uses high ionic strength buffer (1.5M NaCl) to rupture the membrane surrounding the nucleus, thus concentrating the amount of chloroplasts in the extraction. I tried Milligan’s (1992) adaptation, and made a few additional modifications to the protocol. White (1986) found that many published protocols do not work on conifer tissue and suggested that the reason for this might be because the cytoplasm of conifers is more acid than that of angiosperms. Reducing the pH of the extraction buffers from 8.0 to 6.7 significantly improved the yield of DNA from this extraction method. This may be because the acidity minimises the activity of the phenolic compounds present in conifer tissue. This may also be due to the presence of endonucleases in the tissue that work best at higher pH, although the extent of this phenomenon is not known (Jofuku and Goldberg 1981). The chloroplast preparation was then treated with a detergent, both with CTAB as described above, which was not successful and with Triton-X-100, which worked for small amounts of tissue.

**Organelle Extraction Using High Sucrose Buffers**

Many variations of this method have been described (Palmer 1986). The tissue is ground in buffer that is isotonic to the cell contents (e.g. 0.3M sucrose), which prevents the organelles from rupturing. The organelles, once removed and washed are lysed, and the DNA extracted from them. The organelles can be separated and purified using density gradient centrifugation using for example Percoll or sucrose (Palmer 1986). The method presented here is an adaptation of White’s (1986) method (A.E. Szmidt pers. comm.). To increase the relative amount of chloroplast DNA, the resulting extract was subject to a differential centrifugation step to remove some of the nuclei (which are denser than chloroplasts), but total
purification of chloroplasts was not attempted. The extract was then treated with proteinase K and purified in a CsCl gradient.

**DNA Purification.**

When the DNA has been extracted it must be purified, and there are many different protocols that can be followed to remove contaminants from the DNA sample. After the lysis treatments, the contaminating substances may interfere with digestion, and these must be removed. The following are the methods most commonly used to purify DNA solutions.

- **RNaseA** DNA extractions also remove the RNA from the cells. Although the presence of RNA in the preparation should not affect subsequent reactions, it is common to remove the RNA from the sample. The RNA is removed by incubating the sample in the presence of RNase A, a nuclease extracted from bovine pancreas. This enzyme breaks down the RNA into oligonucleotide residues as described in Sambrook *et al.* (1989). These residues do not precipitate because they are too small, and are thus removed. Alternatively, centrifuging the sample in a CsCl gradient separates the RNA and DNA because they have different densities (see below).

- **Proteinase K** This enzyme breaks down protein, and can be used to remove enzymes, especially nucleases that may degrade the DNA or inhibit reactions.

- **Solvents** Proteins are often extracted along with the DNA, and these can be removed by extracting the DNA solution with organic solvents, usually Phenol or Chloroform. These solvents precipitate protein molecules, but leave the DNA in solution. Chloroform is always mixed 24:1 with an alcohol, usually Pentan-2-ol (Iso-amyl-alcohol or IAA) as an anti-foaming agent. Phenol must be distilled and equilibrated to >pH8.0 with Tris, because the DNA separates in the organic phase if the pH is lower than that. Sometimes a mixture of both solvents is used Phenol:Chloroform:IAA in the ratio 25:24:1.

- **Ethanol Precipitation** DNA in aqueous solution precipitates when mixed with alcohols. Propan-2-ol (Isopropanol) can be used but ethanol is more efficient (Sambrook *et al.* 1989). Ethanol requires two volumes to precipitate the sample, so isopropanol, which requires only 0.6-1 volume, is used for large volumes. The alcohol is usually mixed with monovalent cations, such as Sodium Acetate, or Lithium Chloride, which helps the DNA to precipitate (see
Sambrook et al. (1989) pE.15 for pros and cons). This is also used to concentrate the nucleic acid solution.

- **Caesium Chloride Density Gradients** Caesium Chloride (CsCl) is a dense compound, which when centrifuged at very high speed forms a continuous density gradient. The molecules present in the solution band in the gradient at the point that has the same density. Because this separates out the compounds in this way, it is a very effective method for removing such contaminants as polysaccharides and proteins and also for separating different types of DNA e.g. RNA and plasmid DNA. The only limitation is how accurately you can remove the band of DNA.

- **Centrifugation** Extracting DNA samples from CsCl gradients sometimes results in samples containing a significant quantity of insoluble debris, probably from the cell wall. This can be removed by pelleting the debris by centrifuging, and decanting the sample into a clean tube.

**Restriction Digests**

**Digestion**

Complete digestion of DNA is the most important process to get right, since incompletely digested DNA produces unrepeatable results, and these are much more difficult to interpret. Partial digestion is sometimes used for restriction mapping, but not for this sort of project. Digestion of DNA is affected by a number of different factors:

- **Ionic concentration and pH** These are controlled by the buffer in which the reaction takes place. Most major manufacturers of restriction enzymes provide reaction buffers, and the worker is strongly advised to use these wherever possible, as they are batch tested for each specific enzyme, and generally work a good deal better than anything one can make in the laboratory.

- **Concentration of DNA** The efficiency of the digestion depends upon the availability of the DNA to the enzyme. If the DNA is too concentrated or not mixed sufficiently well, then the active sites of the enzyme may not be available to the DNA and digestion is likely to be incomplete.

- **Concentration of the enzyme** Restriction enzymes are stored in glycerol to prevent them from denaturing. If the concentration of the stock solution is too high, then the excess glycerol may cause "star" activity where the enzyme
recognises sequences and cuts the DNA at sites other than its proper one. A maximum concentration of 5% glycerol in the final reaction volume is recommended (Aquadro et al. 1992).

- **Purity of DNA** Obviously, if the DNA is not sufficiently pure, and is contaminated by any inhibitors or enzymes, digestion may be inhibited. This problem if encountered is solved by purifying the DNA using one or more of the processes described above.

- **Reaction time** One unit of enzyme is defined as the amount of enzyme required to digest 1μg of DNA in 1 hour. To reduce the amount of enzyme used, the length of the digestion can be extended. This requires pure samples as prolonged digestion reactions using impure DNA may result in degradation of the sample.

- **Choice of enzyme** The purpose of the experiment should be considered when selecting enzymes. Some enzymes seem to be more "tolerant" than others, and generally more reliable. Some enzymes are also methylation sensitive, that is are not able to recognise DNA to which methyl groups are attached. Some restriction enzymes cut chloroplast DNA much more frequently than others (Palmer 1986). This is useful for population variation studies because more fragments are generated and more sites sampled. In restriction mapping studies however, generating too many fragments is not useful for determining the order of sites in a sequence, so infrequent cutters should be selected (Aquadro et al. 1992)

**Gel Electrophoresis and Southern Blotting.**

**Electrophoresis**

Once the DNA samples have been satisfactorily digested, the next step is to separate the resulting fragments so that they can be compared. The standard way of separating DNA fragments is through agarose gels. DNA strands have an overall negative charge at neutral pH, and migrate towards the anode. The movement of the DNA through the gel is affected by the following factors.

- **DNA fragment size.** Longer DNA fragments migrate more slowly because of the greater frictional drag they experience as they move through the pores of the gel.

- **Agarose concentration** The DNA migrates through the interstices of the agarose matrix. Different concentrations of agarose therefore resolve different
ranges of fragment size efficiently. Another limiting factor is the handling of the gel, as low concentration gels tend to be very flimsy.

- **Conformation of the DNA.** DNA occurs in three different forms: linear, circular, or nicked circular. Clearly, because RFLP analysis deals with restriction digests of DNA, it is linear DNA that is dealt with most of the time. However, it is important to bear in mind that circular DNA migrates more slowly than single stranded DNA of the same size, because this helps to explain the bands shown by incomplete digests of circular, e.g. plasmid DNA.

- **The applied voltage.** At low voltage, the resolution of fragments is best, because the smearing of samples is minimised. Also at voltages below 5Vcm\(^{-1}\), the migration of all fragments is proportional to the voltage. As the voltage applied increases, the migration of long fragments increases differentially.

- **Temperature** The behaviour of DNA in the gel is not affected by the temperature, and most gels are run at room temperature. However, agarose gels of low concentration and Low melting point agarose gels should be run at 4°C, because they are stronger.

- **Addition of dye** Ethidium bromide (EtBr) can be added to the gel before it sets. Although this reduces the mobility of the DNA by about 15%, it enables the progress of the gel to be monitored, which is an advantage in some cases, particularly preparative gels where a particular fragment is to be cut from the gel.

- **Electrophoresis buffer** The buffer used for gel electrophoresis is largely a matter of personal preference, and the commonest is Tris-Acetate (TAE) buffer. Its buffering capacity is quite low, and if electrophoresis is being carried out for extended periods, recirculation of the buffer is recommended.

**Electrophoresis Equipment**

In this project, gel electrophoresis was carried out using horizontal slab gels in a gel tank that allows the gel to be submerged. This is a versatile system as it allows the worker to adapt the size of the gel and the wells in it easily. Handling the gel is simple because the weight of the gel is supported by the gel former. Gel tanks are expensive items of equipment, and there are a number of design features that should be considered before purchasing them. The handling of large gels is much easier if the gel former is made from a UV transparent material such as acrylic. This means that the gel can be stained and examined on a transilluminator without
being removed from the gel former, thus reducing the likelihood of accidental damage of the gel. The design of the electrical connections should also be carefully considered.

The exact methods for setting up gel electrophoresis have been described in detail elsewhere (e.g. Sambrook et al. 1989). The systems used in this lab were as follows:

- **Minigels** 0.8% agarose, run at 2.5 Vcm\(^{-1}\) for 2-3 hours.
- **Sample gels** 1% agarose, run at 0.8-1Vcm\(^{-1}\) for 16 hours (overnight)

**Staining and Recording Gels**

Ethidium bromide (EtBr) is the most common stain used for detecting DNA fragments. It is a fluorescent dye that intercalates between the stacked DNA bases. The gel can either be stained after electrophoresis, or the dye can be added to the gel just before it sets. It is a powerful mutagen, and should be handled with care. See Sambrook et al. (1989) for storage, use and decontamination procedures. Ultraviolet radiation is absorbed by the DNA and the dye and transmitted at 590nm in the visible spectrum. This can be recorded on Polaroid film, which is sensitive enough to detect as little as 1ng of DNA. It is however expensive, and for simple recording of gels, many laboratories are now using TV camera images transferred on to heat sensitive paper as the running costs are less, and it is faster to use. The sensitivity of this method is significantly less, however, and for gels which are to be analysed at this stage, Polaroid photography is recommended. The alternative to EtBr is Silver staining, described in Milligan (1992). This is a complex method used only for detecting fragments directly from the gel. For record keeping it is too complicated and costly.

**Transfer of DNA to Solid Supports: Overview**

Another key development in DNA analysis was Southern's (1975) capillary transfer method for transferring DNA in gels to a solid support that would then enable further analysis to be done. For many years, the support used was nitrocellulose membrane. In recent years this has been superseded by membranes made from nylon. These are now produced by many different manufacturers, and are of two basic types, charged or un-charged. Only the basic principles are dealt with here, as manufacturers supply specific instructions for the use of their product.
The membranes used in this project were both charged nylon membranes, Hybond-N+ (Amersham) and Positively charged membranes from Boehringer UK. These two brands were recommended by the manufacturers for the specific non-isotopic detection protocols they produce. The DNA must be bound onto the membrane in single stranded form so that probe DNA can hybridise to it. If the DNA is transferred using a neutral transfer buffer (e.g. 10×SSC), the gel has to be treated in alkali to denature the DNA in the gel. Alternatively, an alkaline transfer buffer (0.4M NaOH) can be used, which denatures the DNA as the transfer proceeds, so no preparation of the gel is necessary. Once the transfer is complete, the DNA is immobilised onto the membrane by baking, alkaline fixation or UV fixation. The method used depends partly on the membrane and transfer buffer used, and partly on what hybridisation and detection method will be used subsequently. Detailed instructions for the use of the membranes are provided by the manufacturers and these should be followed for the optimum results.

PROBES

Probe DNA: Bacteria, Probes and Vectors.

Bacterial plasmids are closed circular DNA molecules, that occur naturally in a wide variety of bacterial species. They replicate and are inherited independently of the host, though they require some of the host's enzymes and proteins for replication and gene transcription. In nature, they are often the genetic elements responsible for phenotypes advantageous to the host, for example antibiotic resistance, degradation of complex organic compounds and production of modification enzymes (Sambrook et al. 1989).

In genetic engineering the cloning of DNA fragments would be impossible without vectors capable of being ligated to foreign DNA, and subsequently being able to replicate in a bacterial host. To do this, restriction fragments of the DNA to be cloned are generated. The plasmid is digested with the same enzyme. The two linear fragments are mixed with a modifying enzyme that joins the DNA back together, and forms a circular piece of DNA containing the plasmid and the fragment. There is now a large number of plasmid vectors and bacterial host strains available, each having particular properties that make them suitable for particular tasks. The essential features of a plasmid are that they replicate, and that once transformed into a new host, the hosts can be screened for successful transformation, because the plasmids have marker genes that can be selected for. These are usually antibiotic resistance genes, or genes coding for hydrolysing
enzymes such as β-galactosidase. Most plasmids will carry two such markers. If the foreign DNA is inserted into the plasmid at a site within one of these genes, the gene is inactivated, and the bacteria can be tested for their response to one, but not both markers. See Sambrook et al. (1989) for a comprehensive review.

Transformation

To bulk up probes in sufficient quantity for use in Southern hybridisation, the probe, consisting of the DNA fragment and a plasmid vector, must be inserted into a live bacterium, which then proliferates, replicating the plasmid along with it. The culture can then be treated to remove all the bacterial components, including the bacterial and damaged plasmid DNA, leaving a pure extract of closed circular plasmid DNA.

Mandel and Higa (1970) showed that the "competence" of the bacteria, that is the ability to accept a plasmid, and be able to recover and to proliferate, depends upon the presence of calcium ions. Protocols for the transformation of bacteria have been developed based upon this property, one of which is listed in Appendix 1 (D. Finnegan, pers. comm.). It is not clear how this works, but the critical steps in this procedure are the concentration of CaCl₂ and the heating to 42°C of the DNA/Competent cell mixture. There are other methods for producing competent cells, but this is the simplest, and is also very reliable.

Maintenance

For short term storage of plasmids, the bacteria are grown on L broth Agar plates containing the appropriate antibiotic. In the long term, bacteria from liquid cultures also containing the appropriate antibiotic, are mixed with a solution of sterile 60% glycerol and kept at -70°C. Cultures kept in these conditions will be viable for 1-3 years. The glycerol inhibits bacterial respiration, inducing the equivalent of hibernation in the bacteria, and at the same time protects them from the effects of being frozen in an aqueous medium, which would otherwise rupture the cell walls.

Extraction and Purification of Plasmid DNA

It is essential to be able to extract sufficient quantities of pure plasmid DNA from transformed bacterial cultures. The method chosen depends upon the size of the plasmid being extracted, and the purpose for which the resulting preparation will be used. For the sort of probes described below, in the size region 4-15kbp, alkaline lysis was tried, and found to work well. The bacteria are first treated with
EDTA in an isotonic (50mM Glucose) buffer to break down the cell wall, and then with alkali and SDS, which denatures the linear bacterial DNA, but leaves the more tightly bound plasmid DNA intact. For many purposes, the "mini-prep" protocol described in Sambrook et al. (1989) p1.25, produced sufficient quantities of plasmid DNA. The plasmid DNA produced by this method was used for testing the transformed cultures to make sure that the fragments produced were correct, and for labelling. In some cases, particularly when excised probe DNA was required, the method was adapted to cope with quantities of up to 1 litre of bacterial culture (A.E. Szmidt, pers. comm.). This produces a much greater amount of plasmid DNA allowing the inserted probe sequence to be separated from the plasmid vector. The approach is basically the same, but uses lysozyme to help break up the cells, and the plasmid DNA is purified using CsCl gradients. Cultures can be treated with Chloramphenicol which increases the replication of the plasmid and at the same time inhibits the replication of the bacteria. Most modern plasmid vectors in use today, however, replicate in such high copy numbers, that it is not essential.

Excising Probe Fragments From Plasmid Vectors

Large plasmid preparations are usually used where subsequent procedures require that the probe DNA sequence be separated from the plasmid vector. To do this, the plasmid DNA is completely digested with the appropriate enzyme and the resulting fragments separated on an agarose gel to which EtBr has been added so that the progress of the fragments can be monitored. There are many methods of extracting the DNA from the gel, and two of them are described here. The full protocols are described in the Appendix 1.

Electrophoresis onto DEAE-Cellulose

After the fragments are separated, a slit is cut in the gel just ahead of the probe fragment to be excised, and a piece of DEAE-Cellulose (e.g. Whatman DE81) membrane inserted. Electrophoresis is then continued until all the DNA, shown by staining with EtBr, is bound to the DEAE membrane. This is then removed and treated with a high salt concentration buffer (0.7M NaCl) to remove the DNA from the paper.

Low Melting Point Agarose Gels

Low melting point agarose has hydroxyethyl groups added to the polysaccharide chains. This causes the agarose to gel at 30°C and to melt at 65°C. The latter is much lower than the melting point of most double stranded DNA. This allows the
extraction of DNA from the gel by cutting the desired fragment from the gel, melting and diluting the agarose and purifying the resulting DNA solution.

Two methods of purifying the DNA were tested, and eventually a combination of the two found to be simplest. The method usually used (e.g. Sambrook et al. 1989, A.E. Szmidt pers. comm.) requires a large number of organic solvent extractions to remove the agarose. This involves a lot of phenol which is not a very user friendly substance. It has recently been found, however, that much of the agarose can be removed by freezing the diluted agarose mixture at -70°C (X-R. Wang pers. comm.). This appears to break up the agarose molecules, making it possible to centrifuge them out, thus removing the need for any solvent extractions. I found however that this is a bit messy, and that the freezing combined with one Phenol:Chloroform:IAA extraction worked well. In addition, the trouble shooting guide for the Digoxigenin labelling reaction (see below) suggested that the probe should be phenol extracted before labelling, so this combination of treatments is probably the best compromise.

Selection and use of Probes.

The information obtained from Southern hybridisation is indirect because the assumption is made that the hybridisation pattern is an accurate reflection of what is actually present in the target organism. Ideally therefore, the probe should be a homologous, same species probe. These are not always available, in fact for ecological work of any kind it is most unlikely that a same species probe would be available, so a compromise has to be made. In this situation, the probe available that is most closely related to the species in question should be used.

Chloroplast DNA probes

When this project began, conifer gene libraries were not available at all, and most of the original work on conifer RFLPS used the Petunia probes (e.g. Strauss et al. 1988, Wagner et al. 1987). I began by using the Barley probes pHvCpPl-10 described by Day and Ellis (1985), kindly supplied by N. Ellis, John Innes Institute. These and the rDNA and mtDNA probes described below, were supplied in plasmid form, and were transformed into a Rec+, Lac- derivative of E.coli strain RRI supplied by D. Finnegan, ICMB, Edinburgh.

Southern hybridisations to *P. sylvestris* of the Barley probes produced good signals, but there remained an ambiguity about the relationship of fragments to one another in the cpDNA genome, because the structure of angiosperm cpDNA differs in several respects from gymnosperm cpDNA. Gymnosperms lack the rRNA
encoding inverted repeat (Lidholm et al. 1988), possess some genes that have been found in liverworts and algae but not angiosperms (Lidholm and Gustafsson 1991, Karpinska and Karpinski 1993) and gene order is significantly rearranged (Strauss et al. 1988). When the *P. contorta* probes described by Lidholm and Gustafsson (1991) (kindly supplied by A.E. Szmidt, Umeå) became available, I used those in preference, because the homology is such that it is much easier to interpret the patterns obtained. In addition, much more stringent hybridisation conditions could be used.

**Ribosomal DNA probes**

One ribosomal DNA probe was experimented with, a repeat fragment from flax pBG35 (Goldsbrough and Cullis 1981) It gave a very strong signal, but this was not pursued as a line of study for the reasons described in Chapter 2. There was some indication of a difference between two Swedish individuals and the remaining Scottish ones, which is a pointer to future work.

**Mitochondrial DNA probes**

A mitochondrial DNA probe was supplied by J. Ross, Oxford IPS. This is a fragment of the Cytochrome oxidase II gene from maize, pZem1 (Fox and Leaver, 1981). This was successfully hybridised, but the signal was poor. This is probably because the structure of plant mitochondrial DNA allows for much greater recombination, and the homology between the mtDNA of these species is likely to be much less than that of the cpDNA genome. There is still a lack of adequate mitochondrial probes, but this approach has been superseded by Polymerase Chain Reaction (PCR) techniques (Williams et al. 1990), which allow DNA sequences to be amplified using known or random primers, which are then detected directly, thus removing the Southern hybridisation steps and the need for labelled probes. Alternatively, the construction of species-specific probes is possible, an approach that has been used in pines (Dong and Wagner 1993, Strauss et al. 1993).

**LABELLING, HYBRIDISATION AND DETECTION**

The techniques dealt with in this section are well documented elsewhere (e.g. Sambrook et al. 1989). Here, I will describe the principles behind the different methods tried, and refer the reader to the manufacturers for more detail. All the methods described here are supplied as kits by the manufacturers, and they supply detailed protocols for their use.
Labelling

Labelling Using Random Primers

This method uses random hexanucleotides that is, random sequences of DNA 6 base pairs in length, to initiate replication of the template DNA. This method, developed by Feinberg and Vogelstein (1983) was originally used for radioisotopic labelling, but has been used successfully for a number of non-radioactive labels. The template DNA is denatured by boiling and the hexanucleotides are annealed to the template. Free nucleotides, a proportion of them incorporating the hapten or isotope label, are then added along with Klenow enzyme (the Klenow fragment of E coli DNA polymerase I). This is incubated for at least an hour, and up to 24hrs. The result is a mixture of the template DNA and short sequences of synthesised DNA incorporating the label. Unincorporated nucleotides should be removed after labelling to minimise non-specific hybridisation. With radio-labelled probes this is done using Sephadex column chromatography, and for DIG-labelled probes, the synthesised and template DNA is Ethanol/LiCl precipitated. Both complementary strands of DNA are used as templates, and the labelled sequences are therefore complementary, and must be denatured before use, either by boiling in the case of DIG labelled probes or with NaOH for \(^{32}\)P labelled probes.

Covalent Labelling with Horseradish Peroxidase (HRP)

This method does not use DNA replication. Instead, the denatured probe is covalently bound to a modified Horseradish Peroxidase (HRP) molecule using the cross-linking reagent Glutaraldehyde. The amount of label incorporated depends upon the size of the probe, as the HRP is found to bind to the DNA strand approximately once every 50-100 base pairs. Labelling prevents the reannealing of DNA strands and denaturation before use is therefore unnecessary. The labelled probe is added directly to the hybridisation reaction, or it can be stored in glycerol at -20°C before use.

Hybridisation

General

There are many available protocols for hybridising labelled probes to DNA immobilised on nylon membranes. The method used depends mainly on personal preference, and what works best for the system being used. When using labelling and detection kits, the supplier’s recommendations should be followed in the first instance, and adjustments made to suit your own working conditions.
Factors Affecting Hybridisation

• **Solvent and temperature.** The hybridisation solution can be aqueous, in which case the usual hybridisation temperature is 68°C (but this is not the case with the ECL system) or 50% Formamide can be used, in which case the temperature can be lowered to 42°C. Formamide is a bit gentler on the filters, as the temperature used is not so high, but it is toxic, and workers may choose not to work with it if possible. 50% Formamide also increases the reaction time of the hybridisation by about 10%.

• **Volume of hybridisation buffer and length of reaction time.** Although it may seem counter-intuitive, the time required for hybridisation is least when the volume of hybridisation buffer is kept to a minimum. This is because the smaller volume increases the contact of the probe with the target DNA. Longer hybridisation times will increase the signal, but it may also increase background signal.

• **Agitation.** For hybridisation with random primer labelled probes hybridised in small buffer volumes, agitation is not necessary, as long as there is sufficient volume present that the filter is always covered by a film of hybridisation buffer. The filter must not be allowed to dry out. For HRP probes, which are hybridised in larger volumes, incubation in a shaking water-bath is recommended.

• **Use of blocking reagent.** Pre-hybridisation of the filter in a buffer containing a blocking reagent such as BLOTTO or Denhardt's reagent (Sambrook *et al.* 1989), greatly reduces the background signal at the detection stage. For both the non-isotopic kit systems, this blocking step is recommended and the reagent supplied. The blocking reagent is usually also added to the hybridisation buffer, even though Sambrook *et al.* (1989) suggest this is only necessary for single copy Southern blots.

• **Concentration of probe.** The probe concentration used is a matter of optimising the signal:noise ratio of the result. Using too little probe obviously gives a fainter signal, but if too much is added, the signal:noise ratio will be low, particularly if you are using a probe that has not been separated from the vector sequence. The optimum concentration for the system you are using is usually determined empirically.

• **Stringency.** The washes performed after hybridisation and before detection are an essential step for removing any probe that has not bound properly to the
target DNA. The more severe the washing step, the higher is the stringency. Stringency is controlled by the temperature of the wash, and by the concentration of Salt (in the form of SSC buffer). The stringency required can be estimated by the homology of the probe, because a species-specific homologous probe will be bound to the target DNA at more bases, and a more stringent wash may be used because the probe is less likely to be washed off.

Detection

Using Radio-labelled Probes

With a radio-labelled probe, the simplest detection method is possible. The hybridised blot is simply placed in an autoradiography cassette along with a piece of X-ray film, and the radioactivity of the probe results in a signal on the film. To increase the signal, intensifying screens can be used. When these screens are hit by $\beta$ particles they emit photons, thus increasing the strength of the signal on the film. Autoradiographs are usually exposed at -70°C. Exposure times vary from a few hours to two weeks, and strict containment procedures must be followed.

Enhanced Chemiluminescence (ECL) with HRP probes (Amersham)

In this method, the hybridised membrane is exposed to detection reagents, one of which is the peracid salt substrate for the Peroxidase enzyme, and the other, the enhancer and luminol mixture. Durrant (1992) gives a detailed description of the mechanics of the reaction. The membrane is then exposed to X-ray film as before. Exposure times are much shorter for this system. The activity of the enzyme gradually decays over a period of four hours or so, and hard copy results can be produced in as little as 15 minutes if the signal is strong.

AMPPD* or Lumigen® Mediated Chemiluminescence of Alkaline Phosphatase.

This is the most complex of the detection systems, because of the number of intermediate steps required. The probe has been labelled with a hapten, in this case Digoxigenin. The hybridised blot, after being washed in blocking reagent, is incubated in a solution containing the antibody Anti-digoxigenin, that has been conjugated to Alkaline Phosphatase. It is then washed to remove unbound antibody, and finally incubated in the detection substrate AMPPD* (Tropix inc.) or Lumigen® (Lumigen inc.). This is the substrate for the enzyme, and is dephosphorylated and simultaneously emits light at a wavelength of 477nm. Exposure times are slightly longer than with ECL, but hard copy results are still obtained within 24hrs. Despite the complexity of the method, it is simple to use,
and is the most sensitive of the non-isotopic methods on the market today (Kessler 1992).

Re-probing Southern Blots

The advantage of using Southern blots for RFLP analysis is that because the target DNA is irreversibly bound to the membrane, the probe can be stripped off, and the blot re-probed with a different probe. Random primer labelled probes must be removed by incubating the blot in a solution containing NaOH which denatures the hybridised DNA strands, and a detergent, usually SDS, to wash the stripped probe off. Again the degree of homology determines the temperature at which this reaction should be carried out. Covalently labelled probes do not need to be stripped, but the reason for this is not clear. The probe moves off naturally, or is displaced by subsequent hybridisations (Durrant 1992).

EVALUATION AND CONCLUSIONS

The following section compares the performance of all the methods tested in this study, and the reasons for selecting the methods ultimately adopted described. A final summary of the optimum set of protocols for determining the level of RFLP variation in *P. sylvestris* is given.

Comparison of DNA Extraction Methods

The factors that most influence the choice of DNA extraction method for studies into populations are the ease and cost of processing the large numbers of samples necessary for adequate sample sizes. In this particular study, the problem that all methods had in common was that of adapting a method to work successfully with conifer tissue collected in the field, which differs in many respects from angiosperm tissue, and may be more intractable than tissue grown in controlled conditions. A summary of the important features of each method tested in this study is given in Table 3.1.

CTAB Extraction

The main advantages of this method are that it is cheap and quick to perform. It is sufficiently fast to allow sizeable sample sizes to be processed. It extracts total DNA, which for Southern analysis is the most versatile option, because all genomes can be examined, nuclear, mitochondrial and chloroplast. However, its performance is erratic. On material collected in the field in January 1990, the samples were not frozen, and the batches of samples that were processed at the end
<table>
<thead>
<tr>
<th></th>
<th>CTAB</th>
<th>Organelle Extraction</th>
<th>High Ionic strength buffer.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of samples per day</td>
<td>8-12 (limited by space in centrifuge)</td>
<td>8</td>
<td>at least 8 (limited by space in centrifuge)</td>
</tr>
<tr>
<td>Material required (g)</td>
<td>1-5g</td>
<td>20-30g</td>
<td>5g</td>
</tr>
<tr>
<td>DNA extracted</td>
<td>Total</td>
<td>Cp and nuclear</td>
<td>Cp and nuclear</td>
</tr>
<tr>
<td>Yield of DNA (µg g⁻¹)</td>
<td>500-1000</td>
<td>1-5</td>
<td>2</td>
</tr>
<tr>
<td>Success rate (%)</td>
<td>63.7%ᵃ</td>
<td>100%</td>
<td>(100%)ᵇ</td>
</tr>
<tr>
<td>Cost</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

ᵃ See discussion for breakdown of this figure.
ᵇ Based on v. low sample size, therefore not representative.

Table 3.1 Comparison of DNA extraction methods.

produced far fewer DNA samples than the first batches. The number of successful extractions was only 45% overall. I then decided to collect samples from one population at a time, and extract all of those before going on to the next population. This gave much better results, with 72% of extractions working. The last population to be treated in this way was collected in late August 1991 from Glen Tanar, and none of the samples extracted from this population worked at all. Extraction from entire seedlings were the most successful, despite the small amount of material available. After substantial effort trying unsuccessfully to discover what element of the method was wrong, it was abandoned in favour of a more reliable one. Other workers have reported similar problems with this method (R.A. Ennos, A.E. Szmidt, D. Curnow, pers.comm.), which seems to be mostly encountered with woody perennials, or material collected from the field. It is not clear why this is the case, but this problem has not been reported so often with, for example, greenhouse grown material, or from angiosperms. I found that seedlings and material obtained from tree nurseries caused fewest problems, but that the samples from older trees (and many of the trees I sampled were deliberately selected to be as old as possible) gave the poorest results. Angiosperms may have less secondary compounds, or cell walls that are easily ruptured. Another factor seems to be the time of year that the samples are collected. Samples collected during the winter generally produce better results than any others. It may be important whether or not the samples are frozen prior to analysis. Many workers use this technique on frozen or lyophilised material, but it is perhaps noteworthy that no frozen sample that I worked with ever yielded DNA using this extraction
method. However, whatever the reason, both its unpredictability and the difficulty of identifying the reason for this precludes it from being the method of choice, unless the material proves to respond well to this method.

**High Ionic Strength Buffers**

Again this is a simple method, though it requires more centrifugation steps than CTAB extraction. This method produced DNA using small samples of tissue (<5g), but it was not in sufficient quantity to allow many restriction digests to be carried out, and it was not successful using larger amounts of tissue. As described above and in Appendix 1, I made a number of modifications to this method to get it to work on conifer tissue, and it may be that with more work, it could be improved. However, it is limiting in that it only allows cpDNA to be extracted. If it could be refined to produce cpDNA of the purity shown in Bookjans et al. (1984), then it could be used to provide total cpDNA for labelling, which is a much faster way of screening for RFLPs. If the material was going to be used for procedures that require small amounts of DNA such as PCR, however, the small extracts could be useful as templates.

**Organelle Extraction**

This was the only method to produce DNA from every sample. The cost of this method is the highest of any described here, in both time and money, and requires the greatest amount of equipment, particularly specialist items such as an ultracentrifuge. This method does not allow for the easy extraction of mtDNA. Mitochondria are much smaller than nuclei or chloroplasts, and are in all probability discarded at the differential centrifugation stage. To recover these, a whole series of additional steps are required. In this study, restriction digestion was erratic with some enzymes, especially EcoRI and XbaI, but this may be partly a property of the enzyme rather than the sample per se. However, despite the problems it is the method of choice for difficult tissue types, because of its reliability.

**Recommendations**

It is worth beginning with CTAB extraction because it is fast and cheap, always bearing in mind that it may become unpredictable. If it does, very little troubleshooting should be attempted. The main problems are likely to be related to the quality of the tissue, the pH of the buffers, or the presence of contaminating substances that can be removed for example by proteinase K digestion. If the problem is not solved quickly, then this method should be abandoned. The other
two methods are limited to extracting cpDNA, but this is not a problem unless other genomes are going to be analysed. If the number of digestions to be carried out and thus the quantity of DNA required is small, then the high ionic strength buffer reaction would be worth refining, particularly if it could be refined such that CsCl gradients were not required. Smaller tissue samples are required, and more samples could be processed in a day.

Comparison of Labelling and Detection Methods

As nucleic acid hybridisation assays have become an established method of genetic analysis, technical improvements have focused on developing labelling systems that overcome the weaknesses of $^{32}$P labelling. The major limitation is its short half-life of only 14.2 days. This means it cannot be easily incorporated into commercial kits, nor can its use be automated. The logistics of maximising the use

<table>
<thead>
<tr>
<th></th>
<th>$^{32}$P</th>
<th>HRP</th>
<th>Digoxigenin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shelf life of components</td>
<td>1 month</td>
<td>12 months</td>
<td>18 months</td>
</tr>
<tr>
<td>Determining labelling activity</td>
<td>Yes, scintillation count</td>
<td>No</td>
<td>Yes, dot blotting and detection</td>
</tr>
<tr>
<td>Shelf life of labelled probe</td>
<td>1 week</td>
<td>6 months</td>
<td>12 months</td>
</tr>
<tr>
<td>Re-use of probe</td>
<td>yes, within time</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Time to hard copy results</td>
<td>o/n to 15 days</td>
<td>1min to 1 hour</td>
<td>1-24 hours</td>
</tr>
<tr>
<td>Blot stripping method</td>
<td>SSC/SDS</td>
<td>none required</td>
<td>SDS/ NaOH</td>
</tr>
<tr>
<td>Re-use of blot</td>
<td>Yes</td>
<td>up to 3 (DNA washes off)</td>
<td>At least 12</td>
</tr>
</tbody>
</table>

Table 3.2 Comparison between labelling and detection methods.

obtained from one batch are particularly difficult for a small research group, which is a major disincentive to anyone considering setting up these protocols in a new laboratory. There is an added penalty because the wastage may be quite high, especially in small groups, thus greatly increasing an already high cost. Because radioactivity is being used, and $^{32}$P, a high energy $\beta$-radiation emitter, is a relatively hazardous radioisotope despite the short half-life, the containment, monitoring and disposal procedures required are an added cost, and again for a
group considering setting up such work, the infrastructure required may not be available (Table 3.2).

As a response to this, many non-isotopic labelling systems have been developed, (Reviewed by Kessler 1992), two of which I have described above. These represent two very different approaches to the same problem. The DIG system is very similar in principle to radio-isotope methods, because the label behaves in a very similar way in response to hybridisation, but simply requires a different detection method. The general principle of the ECL system by contrast, is totally different, which results in protocols that are fundamentally unlike the other system. The merit of the ECL system is that it is very simple. There are few steps involved, and the results produced when successful, are very good, at least for multicopy sequences. The nature of the project, however, meant that the advantages of the ECL system were outweighed by the lack of flexibility of the system. The DIG system by contrast is very similar to and as effective as $^{32}$P labelling, and in addition the increased shelf-life of the system makes it much more efficient. Both protocols are described in detail in Durrant (1992) (ECL) and Tumolo et al. (1992) (AMPPD detection). The latter describes the method sold by Tropix inc., but apart from using biotin label instead of digoxigenin, the two are virtually the same.

**Labelling and Probes**

The DIG system proved to be more useful due to the handling of the probe being much more flexible, outweighing any of the disadvantages of the system. It is not possible to re-use the hybridisation buffer containing ECL probes, whereas the DIG labelled probes can be recovered and stored at -20°C and reused at least 6 times. Since minipreps of the plasmids usually yielded enough DNA for 2-3 labelling reactions, many fewer of these procedures were required for the population survey. The ECL protocols also strongly recommend the use of probes excised from the vector, as the increase in background signal caused by labelled vector sequences is unacceptably high. Although the use of unexcised vectors does produce increased background with the DIG labelled probes, it is not sufficiently bad to require the use of excised probes. This significantly decreases the amount of work involved in an initial survey. Once the best probes have been identified, these can be bulked up and excised as appropriate.
Labelling method

The labelling of probes using random primers with a hapten label has certain key advantages. The probe is as stable during the reaction steps as the isotope probe because the labelled probe is DNA synthesised using modified nucleotides, rather than a modified probe sequence. The full range of adjustments that can be made to the hybridisation reactions are available. Covalently bound labels do not have this advantage. The hybridisation, because the enzyme is bound to the probe, cannot be carried out at more than 42°C, because enzyme activity is affected. Stringency can therefore only be controlled by salt concentration. In addition, formamide cannot be used as a buffer component, as this also degrades the enzyme bound to the probe. Another advantage of a DIG labelled probe is that it is possible to test the activity of the label before using it. An aliquot of the probe is dot blotted onto a piece of nylon membrane and chemiluminscent detection carried out as normal. This is not possible with the ECL system.

The problem with the DIG system is that the label is too stable. The target DNA is bound very strongly to the membrane, and the homologous probe DNA binds very tightly to the target. As a result, it is difficult to strip the blots completely. The hapten is very stable, and unlike the HRP probe the activity of the label itself does not decrease with time (Durrant 1992). If the blot is not completely stripped, the antibody binds during subsequent hybridisations to the labelled probe remaining from the previous one, and extra bands appear. The regime for stripping the blot has to be stringent. For radiolabelled probes, SDS/SSC buffers are used to strip the blots. For DIG labelled probes, SDS/NaOH buffers are used, and are carried out at high temperature for longer. The problem with this is that if the temperature is too high (in my case I found 68°C to be too high), then the target DNA also gets washed off the blot. This is not a problem with 32P probes because even if the stripping is not complete, the activity of the labelled probe will wear off eventually due to radioactive decay, and the repeated signal is minimal. Similarly, if it were an enzyme label, the activity could be depleted easily by denaturing the enzyme. Although I used a rigorous stripping regime, repeat bands still appeared on some blots. The only way round this was to make sure that the probe concentration and hybridisation conditions were such that the signal was strong enough to appear very quickly and thus to distinguish them from repeated signals which are weaker.
Detection Method

The detection method for the ECL label is very simple because the probe, once hybridised does not have to be modified in any way. The blot is simply washed and mixed with the detection substrates, and then exposed to X-ray film. This contrasts with the DIG system, which requires a series of washes in blocking reagent and antibody buffers to activate the label. Although time consuming however, it is simple and reliable.

Conclusions

The protocols ultimately adopted for this study are outlined in Fig 3.2. The DNA extraction method employed for most of the samples was organelle extraction in high sucrose buffers, chosen mainly for its reliability. The probes used for the main study of *P. sylvestris* were from *P. contorta*, a species of pine from the same subgenus, *Diploxylon*. These fortunately became available in 1992. The methods adopted for the other elements of the procedure depended on the selection of the labelling and detection system.

The ECL system was rejected in favour of the DIG system for three main reasons. Firstly, the ECL blots could only be probed about three times. Alkaline blotting was inefficient, but baking and UV fixation are not the recommended fixation methods for this combination of nylon membrane and detection system. Secondly, the use of probe by the system is inefficient. All probes must be excised before labelling which adds significantly to the effort and cost of surveying large numbers of individuals for variation throughout the chloroplast genome. Finally, the DIG system proved to have the greater overall flexibility. To obtain consistent results with any labelling and detection system, it is important that as many stages in the process as possible can be cross-checked before proceeding to the next step. A wider range of adjustments to the method and more cross checks are possible, making this a better system for developing markers for new species, where modifications are likely to be necessary.
Fig 3.2 Flow diagram showing protocols adopted for cpDNA analysis of Scots pine.

The development of the necessary protocols for the molecular analysis of natural populations is likely to be an integral part of any project initiating research on a species previously unsurveyed for genetic variation. Anyone embarking upon this for the first time will therefore have to familiarise themselves with an array of techniques they may not have encountered before. An essential part of the technical development is the ability to pinpoint what element of the technique should be adjusted to improve a particular aspect of the results, and it may be that this is best done in collaboration with a molecular biologist from whom advice can be sought. During this project, the development of a reliable DNA extraction method was found to be the single most time-consuming problem, which can be attributed to the difficulty of working with field-collected material. This is likely to be the biggest problem with the analysis of a new species. Once pure DNA can be reliably extracted, however, the analytical methods are relatively
straightforward, and it is a matter of selecting the methods most suitable for a particular laboratory. Workers in the fields of ecology or conservation who feel that these techniques may be useful to them should not feel that it is beyond the scope of what they do for technical reasons, but seriously consider whether to get involved in it. This study has shown that it is possible for workers inexperienced in this area of research with little prior knowledge of molecular techniques to set up, with the appropriate guidance, a molecular genetics laboratory for the analysis of populations.
CHAPTER 4

CHLOROPLAST DNA VARIATION IN SCOTTISH POPULATIONS OF P. SYLVESTRIS.

Introduction.

Traditional techniques for the analysis of genetic variation have so far failed to resolve the origin of the genetic structure of natural populations of P. sylvestris in Scotland. In Chapter 2, molecular markers were proposed as an alternative way to do this. The chloroplast genome was selected for analysis because it is a highly conserved genome, and was therefore more likely to reflect an ancient divergence of gene-pools. In addition, if such markers were population-specific, they could greatly aid the management and monitoring of conservation programmes. Protocols for the analysis of cpDNA in conifers were already well established when this project was initiated, and the further development of the techniques used in this experiment are outlined in the previous chapter. This chapter describes the experimental work subsequently undertaken to evaluate the use of cpDNA polymorphisms as a way of distinguishing populations. The objectives of the experimental work were as follows: To sample a representative group of the native Scottish pinewoods, to extract cpDNA from all those samples, and to survey them for RFLP variation.

MATERIALS AND METHODS

Sites

Fig 4.1. shows the 8 regions defined by Steven and Carlisle (1959), and the sample sites selected. Nine sites were selected for sampling, at least 1 from each region to represent the range of native pinewoods throughout Scotland. Two sites, Sheildaig and Loch Maree were sampled from the Wester Ross group because of the high heterogeneity within this group. Steps were taken to ensure that the material collected was of Scottish origin. Tissue samples from each woodland were taken from the oldest trees. The paternal, pollen mediated mode of inheritance of chloroplast DNA in gymnosperms ruled out seed collection, because the risk of pollen contamination in many of these sites is considerable. The only way therefore to be reasonably confident that the material collected was Scottish
Fig 4.1. Sites sampled for cpDNA analysis showing which regional groups each originates from. Site names as in Table 4.1.

was to sample individuals resulting from pollinations that occurred prior to any significant importation of non-native stock. As this began late in the 18th Century (Lines 1964), trees of 180 years or older are very unlikely to have a non-native cpDNA genotype. The sites sampled are described below.
**Amat**

OS Ref. NH46-89-
Owner Mrs A.E.R. Shaw.
Ann Rainfall 1270mm.
Elevation 90-375m.
Designations SSSI, grade 2 (Ratcliffe 1977)

This woodland is characterised by drumlins, with deep peat in the hollows. The pine tends to grow on the shoulders of the knolls, where the soils are thinner, but more freely draining. Birch and other broadleaved species are predominant in the hollows. This is a very diverse woodland, and in the past has had good regeneration (Steven and Carlisle 1959). Interestingly, it is classed as a south west type woodland in terms of vegetation and soils by Bunce (1977).

The woodland is part of an estate managed primarily for sport shooting. It has been deer fenced in the past.

**Black Mount  Doire Darach**

OS Ref. NN 28-41-
Owner P. Fleming.
Ann Rainfall 2032mm.
Elevation 180-285m.
Designation SSSI.

This woodland is a typical south west pinewood with high rainfall, and vegetation types with increased representation of species such as *Molinia* and *Eriophorum*. It is on a gentle undulating slope, with damp hollows and morainic knolls that are less pronounced than those in Amat (Fig 4.2). When Steven and Carlisle (1959) described this site the age class distribution was very skewed towards the older trees, and regeneration was poor. The Gaelic name of this site means "thicket of oak", which suggests that pine was not always the dominant species here, although there is none present now, and the first written record from this area is of a pine woodland, from 1771 (Pennant, cited in Steven and Carlisle 1959).

This site is actively being managed for conservation by agreement with SNH, and regeneration is being encouraged by the establishment of exclosures.
Fig. 4.2 Black Mount pinewood (Doire Darach) and Loch Tulla
**Glen Affric  Pollan Bhuidhe**

- **OS Ref.** NH 19-23-.  
- **Owner** Forestry Commission.  
- **Ann Rainfall** 1778mm.  
- **Elevation** 180-450m.  
- **Designation** SSSI, nationally important.

This is one of the larger areas of native pinewood and is situated on the south side of the Glen with a northerly aspect. The best stands of pine are growing on morainic knolls, and most of the pine is over 100 years old. Few tree species are represented other than pine and birch, and in the 1950s regeneration was poor.

This site was acquired in 1951 by the Forestry Commission, and management since then has focussed on amenity and conservation (Booth 1977). Substantial areas of the pinewood have been ring fenced, and restoration has been attempted through deer control and planting, with some degree of success (Innes and Seal 1971, Fenton 1985).

**Glen Loy  Choille Phuiteachan**

- **OS Ref.** NN 09-84-  
- **Owner** Forestry Commission  
- **Ann Rainfall** 2032-2286mm.  
- **Elevation** 90-300m.  
- **Designation** SSSI.

The vegetation of this pinewood is again characteristic of the western pinewoods. Broadleaved species such as oak and alder are present, and the *Molinia* dominated understory is very luxuriant. The stocking density is high in places, which adds to the rather "rank" feel of this woodland. Again, the age class distribution is skewed, with low numbers of young trees present.

A deer fence was erected around this woodland as early as 1930, and conservation management has continued since the Forestry Commission acquired the site in 1931, but attempts at encouraging regeneration have not been successful (Booth 1977). This stand has been used as a seed source for many years and most of the pine planted in Glen Loy originates from here.
Glen Tanar  Water of Tanar

OS Ref.  NO 47-94-
Owner  Mrs Bruce.
Ann Rainfall  889mm.
Elevation  180-420m.
Designation  NNR, Grade 1.

Glen Tanar is one of the biggest native pinewoods (Forster and Morris 1977). It is the most easterly of the pinewoods, and has the low rainfall and thin podsolised soils typical of the eastern group (Bunce 1977). Pine is the dominant species, with trees of other species restricted to open sites and stream banks. There are many dense stands of pine, and regeneration is abundant within the deer fence, but regeneration was almost non-existent from the middle of the last Century until the erection of the deer fence in 1936, as this age class is virtually absent (Steven and Carlisle 1959).

Management for timber has been a primary activity here for many years, and records suggest that pine planted in the region may primarily be of native origin. It is quite isolated, and pollen contamination may be relatively low (Phillips 1989). Natural regeneration is encouraged, and silvicultural methods have been tested to determine the optimum strategy (Gong et al. 1991 and references therein). The pinewood is managed as a NNR under a joint agreement between the owners and SNH, and a pro-active approach is taken to integrating conservation, sporting and forestry objectives within the reserve (E. Cameron pers. comm.).

Loch Maree  Coille na Glas-Leitire

OS Ref.  NG 99-65-.
Owner  Scottish Natural Heritage.
Ann Rainfall  1778-2032mm.
Elevation  10-375m.
Designation  NNR Grade 1.

The boundaries of this pinewood are probably much as they were prior to extensive timber extraction in this region. The primary cause of deforestation in this area was the production of charcoal for iron-smelting. The pinewoods are situated on the north facing slope of Loch Maree, and the opposite and now largely treeless slopes were probably dominated by oak and birch, both of which were more sought after than pine for charcoal. This woodland is on a moderately steep slope, and
pine is the dominant species, though there is some oak, aspen and birch. There has been continuous regeneration in this woodland this century, resulting in a more mixed age class structure than is found in any other western pinewood (Steven and Carlisle 1959).

SNH owns and manages this pinewood as part of the Beinn Eighe NNR, and amenity, recreation and conservation objectives are integrated. A great deal of work on the establishment of pine was done on this reserve by McVean (1961a, 1961b, 1963a, 1963b, 1966), and it remains one of the most important pinewood sites for conservation.

**Rannoch Black Wood**

OS Ref.  NN 57-56-
Owner Forestry Commission.
Ann Rainfall 1270mm.
Elevation 200-360m.
Designation SSSI Grade 1.

This pinewood lies on a moderate slope on the south side of Loch Rannoch, on the eastern edge of the Rannoch moor. This is basically a south western site, but the elevation is higher, and the rainfall lower. There is more birch in the Black Wood than in many other sites, and more juniper than is usually found in western pinewoods.

The eastern portion of the Black Wood was acquired by the Forestry Commission in 1947, and was designated as an SSSI in 1972. Conservation is the primary objective in this woodland, and it has been a focus for research into pinewood dynamics (Booth 1977).

**Glenmore Ryvoan Pass**

OS Ref.  NH 99-10-
Owner Forestry Commission.
Ann Rainfall 1143mm.
Elevation 320-480m.
Designation Grade 1, National Forest Park.

The steep slopes of the Ryvoan Pass are covered in trees where the soils are stable enough to support the vegetation. Pine is the dominant species, but juniper is abundant, and birch, aspen and holly are all present in low numbers. Most of the trees date from the beginning of the last century, and in common with many
pinewoods, grazing pressure has been sufficient to suppress regeneration since the middle of the last century. This woodland was heavily exploited during World War I (Steven and Carlisle 1959).

This site has been owned by the Forestry Commission since 1923, and has been actively managed for timber production since that time. The area of native pinewood that remains is managed as a nature reserve under an agreement with the Scottish Wildlife Trust (SWT). Extensive underplanting of exotic species and pine of uncertain origin has been carried out throughout Glenmore, and the Ryvoan pinewood has been highlighted as one of the pinewoods most at risk from pollen contamination (Phillips 1989, Booth 1977). Concern about pollen contamination and the lack of regeneration led the Forestry Commission and SWT to begin a felling programme in 1990 of a large area to the south west of the native pinewood, allowing natural regeneration to establish. The site is within the area covered by the Cairngorms Working Party consultation paper (1992), and despite being considered by Ratcliffe (1977) as nationally important, it is not statutorily protected in any way.

**Sheildaig Coille Creag-loch**

<table>
<thead>
<tr>
<th>OS Ref.</th>
<th>NG 82-52-</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Owner</td>
<td>Hon. Hugh Tollemache.</td>
<td></td>
</tr>
<tr>
<td>Ann Rainfall</td>
<td>1524mm.</td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>0-300m.</td>
<td></td>
</tr>
<tr>
<td>Designation</td>
<td>SSSI, Grade 1</td>
<td></td>
</tr>
</tbody>
</table>

Sheildaig is the most westerly of the native pinewoods, and has all the features typical of the Western pinewoods. Holly is abundant, a few oak, birch and rowan are scattered throughout, and there is no juniper present in this woodland. The slope of this woodland is steep and many of trees are growing on rocky outcrops (Fig 4.3). *Molinia, Calluna* and bracken (*Pteridium aquilinum*) are all present as dominants in the ground vegetation. The irregular topography of this site has contributed to the patchy stocking density in this woodland, but the age structure is even, with many relatively young trees, and successful regeneration (Steven and Carlisle 1959, Phillips 1989).

This pinewood is of significant conservation importance, because it is one of the least "derelict" of the pinewoods, despite its relatively small size. There have been several major fires in this woodland, which, combined with low levels of grazing
Fig. 4.3 Sheildaig pinewood
pressure, may have contributed to the success of regeneration (Phillips 1989). One potential problem is the presence of bracken, which expands vegetatively into the bare patches created by fire, and may inhibit further regeneration (Sykes and Horrill 1981), though no long term monitoring of the spread of bracken has been carried out to my knowledge. Sheildaig is isolated and pollen contamination is probably at a low level (Phillips 1989).

Methods

Collection and storage of Scots pine material.

The samples were collected during the months of January to March 1990-1992. Needle bearing shoots were cut from the crown of each tree to fill a 9"x12" plastic bag. This was sealed, and transferred to a cold room at 4°C as soon as possible. If the seeds were required, the cones were removed from the sample and the needles stored at -20°C. Before DNA extraction, the needle samples were ground into fine powder, by freezing in liquid nitrogen and grinding in a coffee grinder (Moulinex Junior model). Conifer tissue can be kept for at least 12 months in either form.

<table>
<thead>
<tr>
<th>Population</th>
<th>Organelle extraction</th>
<th>CTAB</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amat (AM)</td>
<td>20</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Black Mount (BM)</td>
<td>20</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>Glen Affric (GA)</td>
<td>19</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>Glen Loy (GL)</td>
<td>19</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>Glen Tanar (GT)</td>
<td>20</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Loch Maree (LM)</td>
<td>19</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>Rannoch (RA)</td>
<td>18</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>Ryvoan Pass (RY)</td>
<td>20</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Sheildaig (SG)</td>
<td>13</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>168</strong></td>
<td><strong>15</strong></td>
<td><strong>183</strong></td>
</tr>
</tbody>
</table>

Table 4.1 The origin and number of samples from which DNA was extracted and the method used.

Extraction, Digestion and Blotting.

Between 19 and 21 DNA samples from each Scottish population were extracted using organelle extraction and purification in CsCl gradients and CTAB extraction.
In addition, DNA was obtained from 7 Swedish samples (kindly supplied as DNA samples by X-R. Wang, Umeå) (Table 4.2), and one individual from Speymouth, a stand thought to be of native origin (kindly supplied by C. Fleming, Forestry Commission, Elgin). 5 μl aliquots of each sample were digested with restriction enzymes according to the manufacturer's instructions. The digestions were carried out overnight at 37°C. The samples were run on a 1% agarose gel on a 20×20 cm gel plate in 1×TAE buffer for 18 hours at 0.8 V cm⁻¹. 18 Samples were run on each gel, with 2 lanes, one at each end, containing λ HindIII fragments as size markers. Gels were run until the bromophenol blue marker dye was 12-13 cm from the wells. The gel was cut to a length of 15 cm, and capillary blotted (Southern 1975) onto a piece of positively charged nylon membrane (Boehringer) 15×20 cm in size, using alkaline (0.4 M NaOH) transfer buffer. DNA was fixed onto the blot by baking for 20-30 mins at 120°C. The protocol described here allowed fragments of up to 1.4 kbp to be resolved. As the hybridisation signal from fragments smaller than this had been found to be weak, and larger blots costly to analyse, this was considered to be the best compromise.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Origin</th>
<th>Latitude</th>
<th>Altitude (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bg 419</td>
<td>Flinsbjerget</td>
<td>60°</td>
<td>130</td>
</tr>
<tr>
<td>Z 426</td>
<td>Lillhärdal</td>
<td>62°</td>
<td>575</td>
</tr>
<tr>
<td>Ac 560</td>
<td>Brån</td>
<td>64°</td>
<td>300</td>
</tr>
<tr>
<td>749</td>
<td>Umeå</td>
<td>64°</td>
<td>30</td>
</tr>
<tr>
<td>750</td>
<td>Umeå</td>
<td>64°</td>
<td>30</td>
</tr>
<tr>
<td>Bd 404</td>
<td>Ljuså</td>
<td>66°04′</td>
<td>90</td>
</tr>
<tr>
<td>Bd 425</td>
<td>Linå</td>
<td>67°12′</td>
<td>300</td>
</tr>
</tbody>
</table>

Table 4.2 The origin of Swedish samples of *P. sylvestris* used in this study. Samples 749 and 750 were extracted using CTAB, the remaining samples using Organelle extraction.

**Labelling and Hybridisation.**

The probes used for hybridisation were labelled using DIG-labelled dUTP and random hexanucleotide primer extension. Blots were pre-hybridised at 42°C in hybridisation buffer (50% v/v Formamide, 5×SSC, 2% w/v Blocking Reagent (Boehringer), 0.1% N-lauroylsarcosine and 0.02% w/v SDS) for at least one hour. The blot was incubated overnight at 42°C in hybridisation buffer containing 10-20 ng labelled probe per cm² of blot. Stringency washes were 2×5 mins in 2×SSC,
0.1% SDS and 2×15 mins in 0.1× SSC, 0.1% SDS at 58°C. Detection was carried out according to manufacturer's instructions using LumigenPPD® as the substrate (Boehringer).

Double digests were carried out using three enzymes that had compatible digestion conditions, \textit{EcoRI}, \textit{HindIII} and \textit{SstI}. All six combinations of single and double digests were carried out, and these were run on the gel side by side: \textit{SstI}, \textit{SstI/EcoRI}, \textit{EcoRI}, \textit{EcoRI/HindIII}, \textit{HindIII}, \textit{HindIII/SstI}, so that the single and double patterns for each enzyme could be compared directly. The restriction site map was constructed as described elsewhere (Aquadro \textit{et al.} 1992).

\textit{Probes}

The following 13 \textit{P. contorta} probes were used in the genome survey: pPC-K178, H326, H220, E94, H273, H228, K32, H302, E101, K140, H157, H132, X914 (Lidholm and Gustafsson 1991). These probes are a total of 94.3 kbp in length, and allowing for the overlap between probes, this represents about 75% of the length of the cpDNA genome of \textit{P. contorta}. Interpretation of the cpDNA patterns obtained from the hybridisations depends on the homology between \textit{P. contorta} and \textit{P. sylvestris}. Inferring the order in which fragments occur makes the assumption that there are no significant differences in DNA sequence or gene order between the two species. Restriction maps of both species have been published and are shown in Fig 4.4. Comparison of the two restriction enzymes that each study have in common show that apart from an extra \textit{KpnI} restriction site in \textit{P. contorta} and differences for both enzymes in the region around the \textit{psbA} gene which is duplicated in \textit{P. contorta}, the patterns are very similar, suggesting that it is reasonable to assume that \textit{P. sylvestris} fragments are in the same order as in \textit{P. contorta}. A map of the probes is given in Fig 4.5.

\textit{Restriction Fragment Analysis}

The lengths of DNA fragments revealed by Southern hybridisation were calculated using a published program converted from Fortran to Basic (Schaffer and Sederoff 1981). This is an algorithm based on the migration of a standard set of fragments of known lengths, in this case \textit{λ HindIII} fragments. The measured migration distances of the standards are used to define the constants \( m_0 \) and \( L_0 \) which are then used to derive the lengths of the unknown fragments from their migration distances. The deviation of the measured fragments from the theoretical curve gives a standard deviation (in bp) allowing confidence limits to be calculated for the length of each fragment. (See Appendix 2 for program listing and outputs).
Fig. 4.4 Restriction site and partial gene maps of *P. sylvestris* (Karpinska and Karpinski 1993) and *P. contorta* (Lidholm and Gustafsson 1991). The gene order in both species is identical.
Fig 4.5. A linear representation of the *P. contorta* clone collection used in this study (Lidholm and Gustafsson 1991). The circular genome is opened at the *KpnI* site in the *gidA* gene in Fig 4.4, and drawn so that right is anti clock-wise. The probes used in this study are hatched, and those used in both this study and that of Szmidt and Wang (1993) are marked with an asterisk.
RESULTS

Chloroplast Genome Survey

Unfortunately it was not possible to survey all 183 samples for all probe/enzyme combinations. Instead, two individuals were selected at random from each population and these 18 individuals, and two Swedish samples were digested with the following 6-base pair cutting restriction enzymes: *BamHI, EcoRI, HindIII, PstI, Sall, Smal*, and *SstI* (Gibco-BRL), and were tested for a total of 78 probe enzyme combinations. The first three of these enzymes cut the cpDNA genome frequently and the latter are infrequent cutters, giving a balance between sampling a large number of restriction sites and detecting the whole length of the cpDNA genome (Palmer 1986). *Sall, PstI* and *SstI* appear to cut the nuclear genome very rarely, possibly because of sensitivity to methylated DNA. In addition all these enzymes are easy to obtain and are inexpensive. This sample size gives a 95% confidence limit that variants occurring at a frequency of 0.15 or greater will be detected. Due to overlapping between probes for some of the infrequently cutting enzymes, it was not necessary to analyse all 91 probe/enzyme combinations because the hybridisations that were not done would have provided no additional information (Table 4.3).

<table>
<thead>
<tr>
<th></th>
<th>BamHI</th>
<th>EcoRI</th>
<th>HindIII</th>
<th>PstI</th>
<th>Sall</th>
<th>Smal</th>
<th>SstI</th>
</tr>
</thead>
<tbody>
<tr>
<td>K178</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>H326</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>H220</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>E94</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>H273</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>H228</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>K32</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>H302</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>E101</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K140</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>H157</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>H132</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>X914</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Table 4.3 Probe/enzyme combinations analysed in chloroplast genome survey.
Size of the P. sylvestris cpDNA Genome.

A list of all the fragments detected for each restriction enzyme is given in Table 4.4. Hybridisation to the probes described covered 100% of the cpDNA genome for the four infrequently cutting restriction enzymes, *PstI*, *SalI*, *SmaI* and *SstI*. The sum of each of these sets of fragments gives an estimate of the total size of the cpDNA genome in *P. sylvestris*. The remaining three enzymes, *BamHI*, *EcoRI* and *HindIII*, cut the cpDNA genome more frequently. As a result, the sums of these sets of fragments are slightly lower than the other totals, reflecting the fact that some fragments have not been detected where they occur in the region of the cpDNA genome not covered by the probes used. The number of fragments detected is also limited by size, since the protocols adopted for this study enabled detection of a minimum fragment size of 1.4kbp. The sizes of the fragments were estimated using Schaffer and Sederoff’s (1981) method, calculated on the basis of the known size of the marker DNA. Inevitably, there is an error associated with these estimates, which is estimated by this method to be between 225-695bp per fragment.

The four infrequently cutting enzymes produced a size of the Scots pine cpDNA genome of about 119 kbp, a value in broad agreement with that determined by Szmidt *et al.* (1986), and by Karpinska and Karpinski (1993). The latter also used *PstI* in restriction mapping so their results can be compared directly. Both the sizes and the order of the fragments are similar, apart from the absence of a single 1.0 kbp fragment that was not detected in this analysis. As the smallest fragment detected in this study was 1.4kbp long the absence of this small fragment is almost certainly due to it not being detected rather than because of a genuine difference. The order of the fragments was inferred from the position of the *Pinus contorta* probes that they hybridised to as explained above.
<table>
<thead>
<tr>
<th>Fragment no.</th>
<th><strong>Bam HI</strong></th>
<th><strong>Eco RI</strong></th>
<th><strong>Hind III</strong></th>
<th><strong>Pst I</strong></th>
<th><strong>Sal I</strong></th>
<th><strong>Sma I</strong></th>
<th><strong>Sst I</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.7</td>
<td>20.3</td>
<td>11.4</td>
<td>28.2</td>
<td>33.5</td>
<td>30.0</td>
<td>27.4</td>
</tr>
<tr>
<td>2</td>
<td>7.5</td>
<td>12.1</td>
<td>10.9</td>
<td>20.2</td>
<td>18.9</td>
<td>30.0</td>
<td>24.6</td>
</tr>
<tr>
<td>3</td>
<td>7.3</td>
<td>7.7</td>
<td>10.9</td>
<td>14.6</td>
<td>18.1</td>
<td>17.5</td>
<td>13.6</td>
</tr>
<tr>
<td>4</td>
<td>6.9</td>
<td>6.3</td>
<td>10.4</td>
<td>14.6</td>
<td>18.1</td>
<td>17.5</td>
<td>11.2</td>
</tr>
<tr>
<td>5</td>
<td>6.7</td>
<td>5.7</td>
<td>9.8</td>
<td>13.0</td>
<td>11.6</td>
<td>13.4</td>
<td>9.4</td>
</tr>
<tr>
<td>6</td>
<td>6.2</td>
<td>4.5</td>
<td>9.3</td>
<td>10.6</td>
<td>5.3</td>
<td>5.8</td>
<td>9.4</td>
</tr>
<tr>
<td>7</td>
<td>5.2</td>
<td>4.4</td>
<td>8.7</td>
<td>9.3</td>
<td>4.3</td>
<td>2.1</td>
<td>8.5</td>
</tr>
<tr>
<td>8</td>
<td>4.7</td>
<td>4.2</td>
<td>6.9</td>
<td>6.7</td>
<td>3.8</td>
<td>1.5</td>
<td>8.0</td>
</tr>
<tr>
<td>9</td>
<td>4.7</td>
<td>4.1</td>
<td>5.1</td>
<td>2.1</td>
<td>2.3</td>
<td></td>
<td>4.5</td>
</tr>
<tr>
<td>10</td>
<td>4.2</td>
<td>4.1</td>
<td>5.0</td>
<td>1.7</td>
<td>2.2</td>
<td></td>
<td>2.7</td>
</tr>
<tr>
<td>11</td>
<td>4.1</td>
<td>3.9</td>
<td>4.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3.6</td>
<td>3.7</td>
<td>3.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>3.1</td>
<td>3.5</td>
<td>2.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2.8</td>
<td>3.1</td>
<td>2.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2.7</td>
<td>3.0</td>
<td>2.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>2.6</td>
<td>2.9</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>2.5</td>
<td>2.8</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>2.2</td>
<td>2.7</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>2.1</td>
<td>2.6</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2.0</td>
<td>2.5</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>1.9</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>1.6</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>97.4</strong></td>
<td><strong>108.7</strong></td>
<td><strong>111.3</strong></td>
<td><strong>121.0</strong></td>
<td><strong>118.1</strong></td>
<td><strong>117.8</strong></td>
<td><strong>119.3</strong></td>
</tr>
<tr>
<td><strong>+/- SD</strong></td>
<td><strong>5.4</strong></td>
<td><strong>11.5</strong></td>
<td><strong>6.2</strong></td>
<td><strong>2.2</strong></td>
<td><strong>3.2</strong></td>
<td><strong>1.5</strong></td>
<td><strong>2.0</strong></td>
</tr>
</tbody>
</table>

Table 4.4 Number and size (in kbp) of fragments detected for each of 7 restriction enzymes. Standard deviations are sd values calculated by Schaffer and Sderof's (1981) algorithm for each blot multiplied by the number of fragments. See text for further explanation.

This assumes that the gene order of the probes is equivalent to that of the target, so this approach would not be possible with a less closely related species, even of pine. The published fragment patterns in four species from the subgenus *Diploxyylon, P. contorta* (Lidholm and Gustafsson 1991), *P. thunbergii* (Tsudzuki et al. 1992) *P. sylvestris* (Karpinska and Karpinski 1993, this study) and *P.*
radiata (Strauss et al. 1988) when compared, are broadly similar, with the two most closely related of these four, *P. thunbergii* and *P. sylvestris* being almost identical, but compared to a *Haploxylon* pine, *P. monticola* (White 1990a) there are a number of rearrangements, for example, there are only 9 *PstI* fragments in *P. monticola*, but 11 in *P. sylvestris* (Table 4.5).

<table>
<thead>
<tr>
<th>Fragment No.</th>
<th><em>P. sylvestris</em></th>
<th><em>P. sylvestris</em></th>
<th><em>P. thunbergii</em></th>
<th><em>P. monticola</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28.2</td>
<td>29.1</td>
<td>29.9</td>
<td>27.8</td>
</tr>
<tr>
<td>2</td>
<td>20.2</td>
<td>19.4</td>
<td>20.6</td>
<td>27.8</td>
</tr>
<tr>
<td>3</td>
<td>14.6</td>
<td>14.0</td>
<td>13.8</td>
<td>20.2</td>
</tr>
<tr>
<td>4</td>
<td>14.6</td>
<td>14.0</td>
<td>13.7</td>
<td>15.3</td>
</tr>
<tr>
<td>5</td>
<td>13.0</td>
<td>13.1</td>
<td>12.1</td>
<td>13.5</td>
</tr>
<tr>
<td>6</td>
<td>10.6</td>
<td>9.7</td>
<td>9.7</td>
<td>10.9</td>
</tr>
<tr>
<td>7</td>
<td>9.3</td>
<td>9.0</td>
<td>9.0</td>
<td>~3.5§</td>
</tr>
<tr>
<td>8</td>
<td>6.7</td>
<td>6.9</td>
<td>6.7</td>
<td>2.0</td>
</tr>
<tr>
<td>9</td>
<td>2.1</td>
<td>2.0</td>
<td>2.04</td>
<td>1.8</td>
</tr>
<tr>
<td>10</td>
<td>1.7</td>
<td>1.4</td>
<td>1.55</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1.0</td>
<td>1.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5 *PstI* fragments for *P. sylvestris* (*this study*, *Karpinska and Karpinski 1993) *P. thunbergii* (Tsudzuki 1992) and *P. monticola* (White 1990a ~approximate value: data not given).

**Length and Sequence Variation in Scottish *P. sylvestris***

**Length Variation**

No length variation was revealed by any of the probe/enzyme combinations tested. The earliest blot produced showed apparent length differences for the fragments of certain individuals (Fig 4.6). However, subsequent hybridisations revealed that this displacement occurred for all fragments from each varying individual, irrespective of the probe used. I reasoned that this must be an artefact produced by some contaminant altering migration of that sample through the gel, as the length mutations that would be required to produce this pattern are clearly improbable. The samples used on this blot were all repurified to remove the debris present in some of them, and this phenomenon ceased to occur (Fig 4.7).
Sequence Variation

In total, 104 restriction fragments were detected, using six base pair cutting enzymes (Table 4.4). This represents a total of 624 base pairs sampled, or 0.52% of the sequence length. Of the 18 samples tested in this survey, none showed any additional or missing restriction sites, and therefore no point mutations have occurred among these samples for this portion of the cpDNA genome. Although this is a small sample of the sequence length, this is consistent with the low substitution rate found in the cpDNA genome of other plants.

Population Survey

Since no variation was found in the genome survey, it was decided that all 191 samples should be analysed for one probe/enzyme combination. 183 Scottish, 1 tree nursery and 7 Swedish samples were digested with EcoRI and hybridised to pPCH326. This particular combination initially appeared to have a length variant, shown in Fig 4.8. This was subsequently shown to be due to incomplete digestion, because the pattern produced by the variant individuals appeared to be due to a change in length or arrangement rather than an addition of a restriction site. When compared to digests using other enzymes, however, there was no change in length, and eventually this was attributed to incomplete digestion. Despite this, it was considered to be worthwhile determining whether or not there were any rare variants that might not be detected in a survey of smaller sample size as the H326 probe is close to a region where a number of species specific fragments have been found in *Pinus* species (Wagner *et al.* 1992).

Only one sample, GA18, out of 191 samples scored showed variation for the EcoRI/H326 enzyme/probe combination. This sample and two typical samples (AM4 and RY30) were restriction site mapped to determine the nature of the variant (Fig 4.9). The consensus restriction site map for a typical sample is shown in Fig 4.10. GA18 possesses all the typical fragments, plus extra fragments, some of which are very intense.
Fig 4.6  SstI digest probed with pPCH326. Note the displacement of all 3 bands in lane 4, and the poor resolution of the bands generally. Lanes 1&2 GL, 3&4 GA, 5&6 BM, 7&8 RA, 9&10 AM, λ HindIII fragments (sizes in kbp).

Fig 4.7  HindIII digest probed with pPCK140. Samples are the same as in Fig. 4.6. Note the improved resolution of the bands in this blot.
Fig 4.8  EcoRI digest probed with pPCH326. Lanes 1,2,7 and 9 show a clear double band at ~20kbp (marked by arrow). This proved to be due to incomplete digestion. Lanes 1&2 GT, 3&4 GL, 5&6 GA, 7&8 BM, 9&10 RA, 11&12 AM, λ HindIII fragments (sizes in kbp).

Fig 4.9  Restriction Mapping blot probed with pPCH326. Lanes 1-6, RY30 (typical) Lanes 7-12, GA 18 (variant). Lanes 1&7 HindIII/SstI, 2&8, HindIII, 3&9 HindIII/EcoRI, 4&10 EcoRI, 5&11 EcoRI/SstI, 6&12 SstI. λ HindIII fragments (sizes in kbp). Additional bands marked with arrows. *Band due to incomplete digestion (Lanes 5&9).
There are several possible causes of these extra bands:

*Partial digestion* can be ruled out because the smallest band in some of the digests was one of the extra bands. If the bands were solely the product of partial digestion, they would always be larger than the smallest band.

*Homology to probe vector sequence.* For the H326 hybridisation, the probe was excised from the vector, so hybridisation to a homologous vector sequence cannot explain the presence of extra bands. They must therefore be homologous to a sequence in the probe DNA.

![Diagram of restriction site map](image)

**Fig 4.10** Restriction site map of typical individuals. The *SstI* and *EcoRI* site maps are complete, but there may be additional *HindIII* sites between the probes.

- *SstI*, ○ *EcoRI*, ● *HindIII*.

*Mutation.* Having ruled out methodological artefact as a cause of the extra bands, the next step was to determine what kind of mutation could have caused this pattern. The distinctive feature of this variant is that there is no apparent replacement of fragments; all the typical fragments are present in addition to the different bands. An inversion or insertion would be expected to change the lengths and numbers of at least some of the typical fragments, which is clearly not the case here. Although this cannot be ruled out without further analysis, no combination of fragments could be found that was consistent with a mutation of this kind. The other possibility is that a somatic mutation has occurred in the crown of this individual. If it could be shown that a subset of the typical bands could be combined with the novel bands to produce a restriction fragment map, then the presence of two haplotypes could be inferred. The tissue used for DNA extraction originated from many meristems within one individual, so it is possible that two haplotypes could be detected simultaneously.

*Biparental inheritance.* This individual may have inherited two cpDNA haplotypes, one of which differs from the typical type. Heteroplasmy has been recorded in one species of pine, *P. monticola* (White 1990b), but this does not explain the existence of the variant itself. If heteroplasmy is the cause, presumably
there is, or more likely was, since sample trees were selected to be as old as possible, an individual present in that population with only the variant haplotype. It remains to be proved, however, again because of the difficulty in ascertaining which of the typical bands also occur along with the novel bands.

**Extra chromosomal DNA** There have been many instances of extra chromosomal DNA recorded in mitochondrial DNA but none in chloroplast DNA that I know of (See Palmer 1985a). If this is a plasmid DNA, linear or circular, then it should be possible to construct a restriction site map from the fragment patterns in Fig 4.9. of the extra fragments alone. However, inspection of the extra fragments reveals a pattern inconsistent with such a hypothesis. No restriction site pattern can be constructed from the extra fragments that corresponds to a single additional DNA fragment.

The same double digests were also hybridised to K178 in order to determine whether hybridisation to an adjacent probe would reveal a different pattern. The probe hybridised to the same extra fragments except for the SstI/4.0kbp fragment, again with no apparent fragment replacement in the typical bands, which were the same as those obtained in the genome survey. A BamHI digest of GA18 was also hybridised to H132, which is about 12kbp distant from K178, and showed no hybridisation to any extra bands, proving that the extra bands are homologous to a sequence in the cpDNA in the region of the K178 and H326 fragments. Inspection of the fragments strongly suggest to me that a mutation has occurred, where a sequence from the K178 region has been duplicated in the region homologous to the H326 probe, causing a series of variant fragments to be created, all of which are homologous to H326, but only some of which are homologous to K178. Unfortunately, however, due to the difficulty of determining which fragments belong to each haplotype, and the fact that the measurement of the fragments is not sufficiently accurate, the construction of a reliable restriction site maps is difficult without further information. In addition, it is not possible to determine from the data whether the presence of both haplotypes is due to inheritance or somatic mutation.

**DISCUSSION**

**Chloroplast DNA Variation in Pinus**

This study has shown that Scottish populations of *P. sylvestris*, for the probe/enzyme combinations analysed here at least, have the same cpDNA haplotype. There was, with the sample size employed, a 95% confidence limit of
detecting a variant with a frequency of >0.15, which is the sort of pattern expected if two founding cpDNA haplotypes existed in the population, and although a larger sample size is desirable, the conclusion is that there is no variation within or between populations.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Probe</th>
<th>Origin</th>
<th>Fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bam HI</td>
<td>H132</td>
<td>Eur</td>
<td>1.5 4.5 5.0</td>
</tr>
<tr>
<td></td>
<td>Sco</td>
<td></td>
<td>1.7 4.7 5.2</td>
</tr>
<tr>
<td></td>
<td>H157</td>
<td>Eur</td>
<td>(0.3) 1.9 2.7</td>
</tr>
<tr>
<td></td>
<td>Sco</td>
<td></td>
<td>2.1 2.7</td>
</tr>
<tr>
<td></td>
<td>K32</td>
<td>Eur</td>
<td>(0.5) 2.5 9.0</td>
</tr>
<tr>
<td></td>
<td>Sco</td>
<td></td>
<td>2.7 9.7</td>
</tr>
<tr>
<td></td>
<td>K140</td>
<td>Eur</td>
<td>1.3 1.6 [2.8] 4.5</td>
</tr>
<tr>
<td></td>
<td>Sco</td>
<td></td>
<td>1.6 1.9 2.8 [4.1]</td>
</tr>
<tr>
<td>HindIII</td>
<td>H132</td>
<td>Eur</td>
<td>(1.3) 9.5</td>
</tr>
<tr>
<td></td>
<td>Sco</td>
<td></td>
<td>10.8</td>
</tr>
<tr>
<td></td>
<td>H157</td>
<td>Eur</td>
<td>(1.5) 10.5</td>
</tr>
<tr>
<td></td>
<td>Sco</td>
<td></td>
<td>10.4</td>
</tr>
<tr>
<td></td>
<td>K32</td>
<td>Eur</td>
<td>4.6 7.0 8.2</td>
</tr>
<tr>
<td></td>
<td>Sco</td>
<td></td>
<td>4.0 6.9 8.7</td>
</tr>
<tr>
<td></td>
<td>K140</td>
<td>Eur</td>
<td>(0.4) (0.5) 1.0 1.2 10.0</td>
</tr>
<tr>
<td></td>
<td>Sco</td>
<td></td>
<td>1.4 1.5 10.4</td>
</tr>
</tbody>
</table>

Table 4.6 Comparison of fragments detected in Scottish (Sco) and Eurasian (Eur) P. sylvestris. (Data kindly supplied by X.-R. Wang). Figures in brackets are too small to be resolved by methods used in this study. Figures in square brackets are uncertain fragments. See text for discussion of differences.

Although this suggests that all Scottish samples possess the same cpDNA haplotype, it is possible that they differ from populations found elsewhere. The lack of intra-varietal variation is consistent with results from surveys of cpDNA from three other varieties of Scots pine, vars. lapponica, armena and mongolica (Szmidt and Wang 1993). This study used eight of the same probe/enzyme combinations. Comparison of the two data sets shows that the fragment patterns are the same, confirming that vars. scotica, lapponica, armena and mongolica have the same cpDNA haplotype for these probe/enzyme combinations (Table 4.6). The figures for the sizes of the fragments are not identical, but this is likely to be due to differences in the method used to calculate fragment size. The values gained for
A fifth variety, var. *sylvestriformis* was found to contain cpDNA haplotypes typical of both *P. sylvestris* and a sympatric congener, *P. densiflora*. This variety has long been considered intermediate between the two species, and illustrates the importance of hybridisation as a process generating new genetic combinations in this genus (Szmidt and Wang 1993 and references therein).

*P. sylvestris* has spread relatively recently over the vast range that it now covers. The bulk of the populations, with the exception of the Mediterranean relicts, are post-glacial in origin, and have expanded into these areas since the last ice age i.e. within the last 10,000 years (Mirov 1967, Szmidt and Wang 1993). The generation time of Scots pine is comparatively long, so few generations have passed in that time. From a limited genetic base and slow evolution rate, limited differentiation might be expected. However, the fact that the cpDNA of the Scottish populations is from the data presented here, apparently identical to that of vars. *mongolica* from China and *armena* from Turkey, suggests that the glacial refugia of *P. sylvestris* also possessed undifferentiated cpDNA haplotypes.

The remaining uncertainty is the cpDNA of the southern and central European populations of Scots pine, i.e. vars *aquitana, pyreneica* and *iberica* (sensu Ruby and Wright 1976, see Fig 2.1). The cpDNA haplotype of these varieties has not, to my knowledge, been determined. The European glacial refugia of *P. sylvestris* are thought to have been concentrated in two regions: south western Europe where the Iberian varieties are now concentrated, and the eastern Mediterranean where var. *armena* is located. Tobloski and Hanover (1971) measured the monoterpene composition in a wide range of pine populations, and found very low levels of 3-carene in both these possible refugia, contrasting with the higher levels found throughout central and northern Europe. They suggest that this reflects limited gene exchange between southern and northern populations, and that the latter were therefore not originally derived from southern populations, providing there has been no significant selection pressure affecting 3-carene synthesis in either group. If the Iberian varieties of *P. sylvestris* were shown to have the same cpDNA haplotype as the varieties presented here, then the migration pattern suggested by
monoterpene composition assumed to be selectively neutral remains a valid hypothesis. If, on the other hand, these varieties were shown to have different cpDNA haplotypes from those presented here, then it could be said with reasonable certainty that there has been no post-glacial link between Iberian and Scottish or Eurasian populations. The low level of 3-carene typically found in some Scottish populations would therefore be more likely to be due to convergent evolution, possibly related to similarity in the environment, e.g. a more oceanic climate, experienced by the western populations generally. This would also be consistent with the lower level of 3-carene found in the Mediterranean species generally (Mirov 1967). This would suggest that Scottish *P. sylvestris* originated endemically, migrated from central and eastern Europe, or from a combination of the two, but not from south west Europe. Until, however, the cpDNA of the Iberian pine populations is investigated, this remains speculative.

Surveys of other species of pines show that levels of cpDNA polymorphism vary considerably among species. *P. densiflora* (Szmidt and Wang 1993), *P. tabulaeformis*, *P. yunnanensis*, *P. massoniana* (Wang and Szmidt 1992b), *P. taeda* (Ali *et al.* 1991, Wagner *et al.* 1992), *P. echinata*, *P. palustris* (Wagner *et al.* 1992) and *P. torreyana* (Waters and Schaal 1991) have all, like *P. sylvestris* been shown to have no cpDNA variation, or have variants that occur at low frequency. *P. monticola* has a polymorphic restriction site due to a point mutation that is more frequent in eastern than western populations in British Columbia, and occurs at sufficiently high frequencies in eastern populations to discriminate between the two regional groups (White 1990b). *P. elliottii* was the only one of four species of pine surveyed to have clear intra-specific polymorphisms, though these were not shown to be correlated with geographic origin (Wagner *et al.* 1992). *P. densata* possesses three distinct cpDNA haplotypes that are polymorphic at several restriction sites, a phenomenon attributed to the putative hybrid origin of this species as two of the three haplotypes are identical to the invariant haplotypes of two of the possible parental species (Wang and Szmidt 1992b). The Californian closed cone pines, *P. muricata*, *P. attenuata* and *P. radiata* were also found to possess intra-specific restriction site polymorphisms. *P. muricata* showed the highest number of polymorphisms, most of which were fixed within populations resulting in a very high degree of differentiation that is attributed to the almost complete reproductive isolation of the northern and southern populations (Hong 1991). Studies of allozyme variation in the northern populations suggests that the
post-glacial history of this species is complex, and that these populations may only recently have rejoined after a long separation (Millar 1983).

The most extensive survey of intra-specific cpDNA variation is in the two species *P. contorta* and *P. banksiana*, which both have high levels of intra-specific polymorphism due to small insertions and deletions, but especially in the hybrid zone between the two species (Wagner *et al.* 1987). It has been suggested that the increase in cpDNA variation in the hybrid zone is due to destabilisation of the cpDNA of the two species and hence a large amount of mutation. Variation in cpDNA type has been detected within single individuals, suggesting a high rate of evolution, and geographic sub-structuring of variants has also been detected despite high levels of gene flow, suggesting a degree of incompatibility between hybrids (Govindaraju *et al.* 1989, Wagner *et al.* 1991b). Recombination is not thought to be a factor because despite the large number of variants, none are present in both species (Wagner *et al.* 1987, Govindaraju *et al.* 1988, 1989). In addition, they have at least one significant rearrangement, a duplication of the *psbA* gene, that distinguish them from other pines so far studied (Lidholm *et al.* 1991). This gene duplication is also associated with a polymorphic repeat sequence, and it is thought that this is the site at which recombination occurred to create the duplication. A comparison of *P. contorta* and *P. sylvestris* shows that a similar sequence is also present in *P. sylvestris*, but although such a process is possible, no rearrangement of this kind has become fixed in this species (Lidholm and Gustafsson 1991b).

It appears, then, that structural rearrangement is not a simple process, and its occurrence may depend upon the existence of specific sequences and configurations within the cpDNA molecule. In the genus *Pinus*, population studies have been carried out in seventeen species originating from throughout the Northern hemisphere, and nine of these have proved, despite high sample sizes in some cases, to have little or no intra-specific variation. Other species, particularly the *P. contorta/P. banksiana* hybrid zone and *P. muricata*, show high levels of cpDNA variability, in some cases allowing populations to be distinguished from one another. The heterogeneity of cpDNA variation within a single genus suggests that the rate of structural rearrangement is not constant, and that some species may have an inherently more stable cpDNA genome than others (Govindaraju *et al.* 1989, Hong 1991, Tsuzuki *et al.* 1992, White *et al.* 1993). Significant differences have also been found in synonymous and non-synonymous substitution rates in a cpDNA gene, suggesting that this result is not altogether surprising (Bousquet *et al.* 1992). This has implications for anyone attempting to calculate
divergence times on the basis of molecular phylogenies for example, and again emphasises the need to combine data from many sources (White et al. 1993).

**The Occurrence of Heteroplasmy in Pinus.**

The single variant individual found in this survey is most likely to be the result of a duplication or translocation event, and the presence of two haplotypes in the single individual is either due to somatic mutation resulting in meristems with different haplotypes, or to biparental inheritance. This kind of variation has been found before, in a single individual of *Betula pubescens* collected from a natural population (D. Curnow pers. comm.). Such variation has not, to my knowledge, been reported elsewhere, but this may be due to the smaller sample sizes investigated in other surveys failing to detect low frequency variants. A sample of 902 individuals of *P. contorta* and *P. banksiana* was examined for intra-specific variation by Govindaraju et al. (1989), and though 24 variants were detected, only 3 were present in frequencies greater than 2.5%, and 14 were detected in only one or two individuals, a frequency of 0.2% or less. No samples in this survey apparently possessed two haplotypes simultaneously. Similarly, an extensive survey of *P. densata* failed to reveal any evidence for the occurrence of heteroplasmy in that species (Wang and Szmidt 1992b). Heteroplasmy has been documented in *P. monticola* and in four out of six individuals from the *P. contorta/P. banksiana* hybrid zone (White 1990b, Govindaraju et al. 1988). In the latter sample, two of the heteroplasmic individuals contained more than two cpDNA haplotypes, suggesting that the parental trees were also heteroplasmic.

The fact that in this study the only varying individual detected may possess two cpDNA haplotypes suggests that either somatic mutation or biparental inheritance is occurring. It also suggests that at least one of these processes must be more common than is usually thought, as the likelihood of detecting two rare events in one individual seems rather small. If somatic mutation was common, however, a higher level of intraspecific variation should have been detected. The possibility exists, however, that biparental inheritance is more common than is generally thought, and that it is simply not detected because of the lack of intra-specific variation. Most of the studies of cpDNA inheritance in gymnosperms have been done using inter-specific crosses, mainly because of the lack of intra-specific polymorphism (e.g. Wagner et al. 1991, Szmidt et al. 1987, Stine and Keathley 1990, Neale and Sederoff 1989). More detailed studies of cpDNA inheritance have been made using intra-specific crosses in *P. contorta* (Dong et al. 1992), *P. banksiana*, (Wagner et al. 1989) and Douglas fir (Neale et al. 1986). All three
studies found progeny containing non-parental and maternal cpDNA patterns, a result in most cases attributed to pollen contamination during the crossing experiments. However, a low level of non-paternal cpDNA inheritance cannot be ruled out, and this may be what was detected here.

The two possible explanations for the origin of the single variant individual found in this survey would each have different effects on the genetic structure of the population. If it is due to biparental inheritance, given that the sample came from an old tree, the donor of the variant haplotype has almost certainly died, but a larger survey of that population might detect more individuals of that genotype. If however, a somatic mutation has arisen in this individual, it may now be releasing pollen from strobili arising from the meristem that has the variant haplotype. It would be interesting to screen seedlots from that population to see if there was any evidence for spread of that variant through the population.

**RFLP Analysis of *P. sylvestris***

There were three main difficulties encountered in this study with the interpretation of the hybridisation data. Incomplete digestion regularly occurred with some enzymes, especially *EcoRI*, and some samples proved to be particularly susceptible to this. The presence of weakly hybridising bands of intermediate size in some but not all individuals is a clue that incomplete digestion has occurred, but it is not always obvious, as was the case with the *EcoRI/H326* digest described above. In this case, incomplete digestion was eventually inferred from the patterns produced by hybridisation to other probes, and digestion with different enzymes. If a variant is detected, attempts to corroborate it in some way should therefore be made.

Difficulties with the interpretation of banding patterns can be met with where differences in hybridisation signal occur. A particular problem in this study was the difficulty of completely stripping the blots described in the previous chapter. This led to the appearance of weak bands in subsequent hybridisations, and in cases where the probe was poorly labelled, it was difficult to determine whether or not such a fragment appearing on the blot was due to the current or the previous hybridisation. It is important in this case to record the order in which hybridisations are carried out, so that bands from the previous hybridisation can be eliminated. In addition, fragments of nearly equal size occur, and it is often difficult to determine whether a band represents one or more bands.

The estimates of fragment size are subject to several sources of error. The accuracy with which the migration of a band can actually be measured is limited,
because the point at which the band should be measured is difficult to standardise, particularly where the intensity of the signal varies. Larger fragments are especially prone to error because the separation is less, and an inaccurate measurement produces a much greater difference in fragment size. In Schaffer and Sederoff's (1981) method, the marker lanes are measured to produce a regression line of migration distance with fragment length. If the blot being analysed has a "smile" in it, where the migration rate in the gel has varied across the width of the gel, then clearly the regression for the marker lane will bias the estimate of the sample fragments. The number of fragments in the marker lanes also influences the accuracy of the regression. The marker used in this study consisted of only 6 bands. If a marker producing more bands had been used, the measurements may have been more accurate, because the slope of the regression can be more accurately calculated.

CpDNA as a Marker in *P. sylvestris*

It is important to emphasise that the results reported here do not *disprove* the existence of two genomes in Scottish pine populations. It can only be concluded that the cpDNA of Scots pine has not evolved rapidly enough to discriminate between groups, despite the loss of the inverted repeat. With hindsight, therefore, the value of cpDNA as a marker can be evaluated. This result illustrates the difficulty of selecting markers based on the information available from a limited number of taxa. When this project began, the only published population survey of cpDNA variation in *Pinus*, was in the most variable group, *P. contorta* and *P. banksiana* (Wagner *et al.* 1987). CpDNA variability appeared to be significantly higher in the conifers, as had already been predicted from the absence of the rRNA encoding inverted repeat in many conifers (Lidholm *et al.* 1988, Strauss *et al.* 1988). It has, as we have seen, however, transpired that this level of variability may be the exception rather than the rule.

The strategy adopted for this study, that of a wide-ranging search for RFLPs, although the commonest approach at the time, is relatively inefficient. It is now more widely understood that the types of variants most likely to be found are small length variants concentrated in the non-coding regions of the cpDNA molecule (e.g.Ali *et al.* 1991, Lidholm and Gustafsson 1991a). RFLP analysis, especially using infrequently cutting restriction enzymes, produces fragments where the polymorphic region is attached to long and uninformative pieces of DNA, which may obscure the polymorphism itself because of the reduced resolution. Our knowledge of the structure of the cpDNA genome is now much better, and the
development of PCR protocols and improved primer design will allow the screening of samples for polymorphisms of this nature to be better targeted. For example, in population studies carried out on three families of conifers, Douglas fir (Pseudotsuga menziesii), coast redwood (Sequoia sempervirens) and incense cedar (Calocedrus decurrens) all showed polymorphisms due to short length mutations of the type described above, and two of these, coast redwood and incense cedar had polymorphic fragments revealed by the same probe (P9 from Petunia: Palmer et al. 1983) as that used for the study of the P. contorta and P. banksiana hybrid zone (Wagner et al. 1987, Ali et al. 1991). A recent study on the inter-specific relationships in Picea also found that restriction site differences were limited to two regions of the genome, suggesting that nucleotide substitution rates are not uniform throughout the cpDNA genome (White et al. 1993). As the information available increases, such non-coding regions and potentially polymorphic sites become easier to identify, and a large portion of the genome likely to be uninformative could be eliminated from the screening at the outset. Primers could even be designed to produce species-specific probes and screening could then concentrate on a particular sequence that could be analysed for a large number of restriction enzymes, thus greatly increasing the likelihood of uncovering point mutations as well as length variants.

Implications for Conservation of Scots Pine.

The results presented here do not disprove the existence of two distinct gene pools of differing post-glacial origin within the Scottish populations of Scots pine. There is, however, evidence albeit limited, that there are genetic differences among populations in certain phenotypic markers and quantitative traits measured in provenance trials. In particular, the evidence that there is adaptive differentiation for survival among eastern and western populations suggests that these genetic traits should be conserved as part of the diversity within this species that has evolved at the extreme of the distribution, and in an essentially atypical habitat for the species as a whole. Irrespective of the origin of that variation, transfer of genetic stock between east and west should be discouraged, a measure which is currently stipulated by the Native Pinewood Grant Scheme. In the context of genetic conservation, pollen contamination is potentially the more serious problem, and I feel that an unambiguous marker for distinguishing native and non-native pine would be a valuable contribution to the management of pinewoods for conservation at all levels.
CONCLUSIONS

This experiment has shown that cpDNA variability detectable by RFLP analysis is limited in *Pinus sylvestris*, both within and among populations of var. *scotica*. Comparison with data from three other varieties shows that var. *scotica* is the same and that the cpDNA of *P. sylvestris* is not significantly differentiated among populations or distant geographic locations. This negative result does not disprove the existence of two gene pools in Scotland, but that cpDNA has been found to be insufficiently differentiated in this species to be useful as a population marker. Work is now in progress to determine the level of variation in mtDNA in Scots pine which may prove to be more variable, and is more likely to be differentiated because the mitochondrial genome is dispersed via seed which travels significantly shorter distances than pollen (Ennos 1994). These markers may prove to be useful as population markers within Scotland, but pollen contamination will be much more difficult to prove because the maternal inheritance of mtDNA prevents direct determination of the pollen donor.

This result demonstrates the importance of appropriate marker selection, and the necessity of considering all the data available from closely related species in order to do this. Unfortunately, there is a random element in evolution which means, of course, that however good the data on which marker selection is based, the usefulness of the marker cannot be predicted. Despite the lack of differentiation present in the cpDNA genome of this organism, there continue to be reports from diverse groups of plants, of higher levels of cpDNA variation than was initially expected, and the results described here should not deter anyone from dismissing cpDNA as a population genetic marker out of hand. Each case must be taken on its own merit, and allowance made for the possibility that variation may occur in unexpected places.
CHAPTER 5

POPULATION GENETICS IN CONSERVATION: CONCLUSIONS AND RECOMMENDATIONS

Introduction

The central aim of this thesis was to show whether or not it is possible, and indeed, useful, to develop molecular markers as an aid to the conservation and management of the genetic resources present in natural populations. The use of genetic markers was addressed in this thesis from first principles: what are genetic markers, how do they behave, and what can they detect? It is important for those who are not familiar with these techniques to understand the behaviour of genetic markers in order to interpret the data that they provide. Genetic markers of the kinds described in this thesis have different properties and mark different processes at the population level. They have a great deal of potential for improving gene conservation management at all levels, provided that the marker used is appropriate for the purpose. To evaluate these possibilities, the native pinewoods of Scotland were analysed using cpDNA markers. The following chapter summarises what the work presented here has shown, suggests further avenues of study, both general and specific, and finally, suggests a more philosophical conclusion about the nature of the work presented here.

Project Planning: The Initial Stages.

Selecting an Appropriate Marker.

The first thing to be done at the beginning of any project involving population genetics, is that the purpose of the project must be clearly defined. If this is sufficiently strongly articulated at the outset, then deciding which marker to use will be simpler, because certain basic requirements of the project will eliminate some markers from consideration. Chapter 1 describes the techniques available to the population geneticist, and the properties of the variation thus detected, which in turn, determines the questions that can be answered by each one. By presenting the range of techniques available in detail, it has emerged that there are three key questions that should be asked when deciding upon the marker that is to be studied.

What is the influence upon the marker of selection? Knowing the potential selection pressure that the marker is, or may have been, subject to is an important
step to deciding upon the appropriate one. Studies of natural selection require adaptive traits, studies of long-term historical relationships require neutral traits, and for shorter term studies, the adaptive differential may not be important.

**What is the rate of evolution of the marker?** The timescale over which the observed genetic structure of a marker is likely to have evolved determines which part of the genome or what traits it can be used to investigate. Traits evolve at different rates which in turn determines the level of variability of the marker within and among species and populations. Slowly evolving markers are therefore not appropriate for the study of short-term changes in populations, because there is likely to be insufficient variation. By the same token, rapidly evolving markers will rarely be informative about the historical relationships between populations because subsequent evolutionary change will have obscured any ancestral similarity between the populations.

**How is the marker inherited and expressed?** The inheritance of the marker determines the demographic processes that can be detected. In some cases the mode of inheritance of a marker prevents it from detecting the process of interest, limiting its use in that context. For example, if the extent of gene flow via pollen is to be measured, this cannot be done using a maternally inherited and thus seed transported genome. The expression of some genetic markers is limited, especially those that are gene products. Detection may therefore be limited to some of the alleles in the case of dominant expression, and to particular stages of the life-cycle in, for example, some isozymes.

The answers to these questions should make it clear what kind of marker should be used to analyse the process under consideration.

**Genetic Analysis of Natural Populations.**

The development of the necessary protocols for DNA analysis in *P. sylvestris* represents a major component of the work undertaken during this project, and this work is described in detail in Chapter 3. Again, elements of the protocols proved to be of particular importance if the population genetic structure of natural populations is to be studied, and these are described below.

**Sampling Strategies.**

Above and beyond the usual requirements of adequate sampling, the use of genetic markers, especially DNA markers, requires that the tissue collected provides the appropriate biochemical compounds in an active and uncontaminated state. Many
methods used for preserving samples in the field will not, therefore be appropriate. Initial testing of any available tissue is strongly recommended, as it is essential to determine what tissue it is necessary to collect, particularly if collecting trips to isolated areas for extended periods are required. These practical considerations alone may rule out the analysis of some genetic markers.

Protocol Development

This study has shown that it is possible for a research group with no previous experience of molecular biological techniques to establish a laboratory capable of carrying out molecular genetic analysis of populations. However, this experience has highlighted some of the major problems with this approach. In studies concerning natural plant populations, it will often be the case that the protocols for working on that species have not been perfected. There is therefore, no alternative but to find the protocols used for the most similar species, and to adapt them until the results obtained are satisfactory. For the worker who is not initially familiar with the techniques being employed it is essential that they obtain as much information about the existing protocols as possible, in particular those elements of the protocol that were adopted specifically for that taxon. For example, there may be properties of, (e.g. cytoplasm pH), or compounds in (e.g. starch or resins) the tissue used that require special treatment, that are likely to be common to all species in a genus.

The time necessary to perfect protocols is unpredictable, but must be allowed for in this type of research. In particular, it needs to be accepted that even if the protocols employed work well at first, problems may still arise subsequently. Troubleshooting is especially difficult for someone who is not yet sufficiently familiar with the behaviour of these techniques to know where the best place is to begin looking for a solution to the problem.

Molecular Genetic Analysis

Perfecting a sufficiently reliable DNA extraction method was undoubtedly the biggest single problem that had to be overcome in this project. However, by using protocols that had been adapted specifically for use with conifer tissue, which differs in several respects from angiosperm tissue, it was possible to extract DNA of the quality required for RFLP analysis. The method adopted was, however, the most time consuming and costly, and for these reasons was not ideal. Had time permitted, a more efficient technique for extracting DNA from large numbers of conifer tissue samples could have been developed.
Selection and use of DNA probes was limited by what clones were available, as the construction of a genome library was considered to be outwith the scope of this project. Fortunately, a set of probes from *P. contorta*, a species in the same subgenus as *P. sylvestris* became available in 1992, which greatly increased the information that could be obtained from the Southern hybridisations. From the comparison between the results obtained from the barley probes initially used and the *P. contorta* probes, it is clear that probes should be used from as closely related a species as possible, which, with the developments in PCR technology is fortunately becoming an easier requirement to fulfil.

There are many protocols and kits now available for the labelling and detection of DNA probes in Southern hybridisations. Three of these were tested in this study, and the one finally selected for the cpDNA survey of *P. sylvestris*, Digoxigenin labelling and LumigenPPD detection (Boehringer), combines the flexibility of $^{32}$P labelling with the economies of scale allowed when using a non-isotopic label.

**The Origins and Genetic Structure of Native Scottish *P. sylvestris* Populations.**

The populations selected to evaluate the development and use of molecular genetic markers were natural Scottish populations of *P. sylvestris*. The previous research carried out on the population genetic structure of these populations is described in Chapter 2. Evidence from palynological studies and from analysis of monoterpenic composition has led to the widespread belief that some of the native Scottish populations of *P. sylvestris* have been isolated from European populations for many generations, surviving the ice-age in a glacial refugium to the north west of Scotland. Two sources of evidence, however, contradict this hypothesis. Geomorphological evidence and pollen analysis from inter-stadial periods suggest that the presence of pine populations in this area was unlikely, and secondly, though monoterpenes may be good indicators of relatedness as distinct from similarity, they are not necessarily so, as there is evidence from other species of pine for selection acting upon monoterpenic composition. The small amount of data available from provenance trials nonetheless suggests adaptive differentiation among Scottish populations. The availability of data from these differing sources provided a situation where the information gained from the analysis of molecular markers could be compared directly with the results from the more commonly used techniques. On the basis of the strategy for selecting markers presented in Chapter 1, the analysis of chloroplast DNA was the method chosen, and the protocols described in Chapter 3 were developed. The survey of cpDNA variation in Scottish populations of *P. sylvestris* subsequently undertaken, and the results of
that survey are described in Chapter 4. As is often the case in scientific research, the analysis posed as many questions as it answered, and the most interesting of these in the context of population genetics and the use of molecular markers are described below.

**Recommendations for Further Research on P. sylvestris.**

This project involved using a technique developed by molecular biologists, and applying that to a practical problem of genetic conservation. Chloroplast DNA was selected as the genome most likely to provide a marker for both pollen contamination measurements and ancestral relationships between populations of native Scottish *P. sylvestris*. Variation in cpDNA, however, was found to be very low, a result consistent with the findings of recently published population surveys of other varieties of *P. sylvestris*. Despite these negative results, there are two interesting features of cpDNA at the molecular level that are suggested by the data, the presence of a more stable conformation of cpDNA than was thought to be the case, and the possibility that strict paternal inheritance is not universal in the gymnosperms.

The analysis of chloroplast DNA of Scottish *P. sylvestris* has not succeeded in identifying the historical relationships between populations, nor can it discriminate between varieties for the purpose of quantifying pollen contamination. The properties of cpDNA would have enabled it to fulfil both these purposes, had it been sufficiently variable. Given the results of this study, however, it will be necessary in future to investigate these two questions separately, as in the absence of cpDNA variation, two markers with different properties are required.

**Genetic Origins and Population Structure.**

The level of cpDNA variation within five dispersed varieties of *P. sylvestris* has been investigated, and apart from var. *sylvestriformis* where cpDNA from *P. densiflora* has introgressed into the population, they have all been found to be monotypic, and the suggestion is also that there is no inter-varietal variation (this study, Szmidt and Wang 1993). What remains to be determined is the pattern of cpDNA variation in the central European populations and in particular, the south western populations of southern Europe and the Iberian peninsula. These populations are highly differentiated from other varieties of *P. sylvestris*, and it has been suggested that these populations are Pleistocene relicts, that may have been isolated from the populations of the north and east for long periods. Clearly the investigation of these populations is required if an accurate picture of the cpDNA
variation in the species as a whole is to be constructed (Szmidt and Wang 1993). In addition, comparison of the cpDNA haplotypes with the monoterpene compositions of *P. sylvestris* may finally resolve whether or not similarity of monoterpene composition is a good indicator of genetic relatedness in this species.

Work is currently underway to investigate mitochondrial DNA in Scots pine. The higher rate of structural rearrangement in this genome, the fact that significant intra-specific variation has already been identified in several species of pines (Strauss *et al.* 1993, Dong and Wagner 1993) and the greater likelihood of population differentiation in seed dispersed genomes (Ennos 1994) makes this genome more likely to provide information on population structure and genetic origin on this timescale.

**Genetic Conservation and Pollen Contamination**

The major drawback with mtDNA is that, due to the maternal mode of inheritance, it is not possible to identify the pollen parent of any one individual. If this is considered to be a primary objective in Scots pine conservation, then another strategy will need to be adopted. A nuclear marker that is less variable than allozymes could be used so that the paternal and maternal parents could be identified, at least in seed lots taken from known maternal trees. In this case, the marker has to be co-dominantly inherited so that both parental contributions can be identified. Since all that is required for this is a marker that is fixed within varieties so that pollen contamination that is currently occurring can be identified, the origin of that difference is less important than the ability to reliably discriminate between populations. Ribosomal DNA markers could for example be used, despite the evidence that they are under the influence of selection. Alternatively, mtDNA variation can be combined with a Mendelian marker to infer the pollen donor by subtraction, the approach used by Arnold *et al.* (1992). Since paternal cpDNA inheritance has rarely been reported in angiosperms, this is likely to be a consideration in many studies of plant populations.

**Chloroplast DNA Structure and Inheritance in Gymnosperms.**

The discovery of one individual with an atypical banding pattern has interesting implications for work on cpDNA in gymnosperms. The atypical banding pattern seems to be due to the presence of two different cpDNA haplotypes within one individual. Biparental inheritance of two different variants, or somatic mutation within the variant individual could be the cause of this, but the latter is thought to be rare (Govindaraju *et al.* 1988). If, as is often implied, biparental inheritance is
also rare, then the probability of detecting this particular individual must be extremely low. The possibility exists however, that biparental inheritance is in fact more common than is usually thought, and that the extent of it has been difficult to estimate because of the lack of intra-specific variation in cpDNA. This clearly has implications for the way in which cpDNA data is analysed in population studies, as organelle DNA is usually assumed to be strictly uniparentally inherited. This is a possibility that merits further investigation to determine whether or not this is in fact a valid assumption.

In some respects, the homogeneity of such a widespread and ancient taxon is unexpected, especially given that cpDNA genomes lacking the rRNA inverted repeat have been found to be more variable than those that possess it. Recent data from the closely related species *P. thunbergii* showing that a residue of the inverted repeat remains, suggests the possibility that such a repeat could also exist in *P. sylvestris* (Tsudzuki et al. 1992). The conserved conformation of cpDNA observed in angiosperms may therefore also be a factor reducing the variability in the DNA of some, though not all, pine species. It would be interesting to know whether or not there were any sequences present in *P. sylvestris* that, by conferring greater stability on the cpDNA genome would explain the lack of variability that has so far been found in this species. The presence or absence of such a feature in two closely related species could also shed light on the evolutionary relationships within the genus as a whole.

**Palynology studies.**

The review of the Holocene vegetation and glacial history of Scotland presented here, has revealed that there are a number of important conflicts within the data. Geomorphological studies are increasingly suggesting that the northern and eastern margins of the ice-sheet covering Scotland were weak, and that this area may have been ice-free very early in this period. This contrasts strongly with the idea from the pollen data in Scotland that glacial refugia of pine would have existed in the north west. There are however, indications of an earlier presence of pine in the eastern Highlands than is usually thought. The Forth-Clyde valley and the north east coast of Scotland appear to be relatively poorly sampled, at least for cores that have been accurately radiocarbon dated (see map in Birks 1989), and it is important that these areas are investigated to see if there is any more conclusive evidence to be found about the timing of pine's arrival in the east. Similarly, the south west, another putative entry point for pine into Scotland is apparently under sampled. Ultimately, however, the inferential nature of vegetation history
constructed from pollen diagrams, particularly for species like pine whose pollen is produced in large quantities and disperses very widely, means that underlying genetic relationships between populations is potentially at least, a more useful approach to this problem, providing that an appropriate marker is used.

**Genetic Markers and Conservation**

Conservation of the genetic resource that native populations of *P. sylvestris* represent is one of the main aims of the NPGS in Scotland. Genetic markers have been used to define the population boundaries and to determine what level of genetic transfer is permissible between them. In Chapter 2, it is described in detail why the inferences upon which these definitions are based may not be correct. Despite this, the regulations that make up the NPGS are practical and should maintain the integrity of the genetically differentiated Scottish populations. What this illustrates is that genetic conservation can in many cases be done without the aid of sophisticated genetic markers. They do, however, enable the manager to monitor the effectiveness of the adopted strategy, and to predict the consequence of changes in management practices. In this particular case, however, cpDNA failed to provide an unambiguous marker for distinguishing populations, this result has recently been confirmed for other varieties of *P. sylvestris*, and we know not to look any further at this genome for the purpose of assessing the level of intra-specific population differentiation.

This study has shown that it is possible to develop protocols for analysing genetic markers in sufficiently large sample sizes to determine the genetic structure of natural populations of *P. sylvestris*. However thorough the process of selecting a marker is at the outset, there is no guarantee that the appropriate level of variation will be found, and it has been clearly demonstrated here that this is a significant risk. All additional information on the level of DNA variation present in natural populations provides subsequent researchers who wish to use genetic markers with better data on which to base the selection of markers. Though the result presented here is, in itself inconclusive, it adds to the data available to future workers in this field.

**Dialogue and Co-operation**

As an ecologist by training, I was attracted to the study of molecular markers because of the promise that they held of providing unambiguous information about the relationships within and among populations, that, using other techniques would take many years of observations and experiments to uncover. At the time I
embarked upon this work, there was comparatively little work published using molecular markers on natural populations, and less on gymnosperms. So with the information from the limited number of papers that had been published, a lot of guidance from friends and colleagues, and a certain amount of blind faith, I undertook a project using the language and technology of a discipline that was unfamiliar to me. I have proved that it is possible for a novice in the field of molecular biology to use the techniques for a purpose other than that for which they were perhaps originally intended. However many of the problems that I encountered with the methods used for this study could have been overcome so much more quickly by someone better versed in molecular biology. A project such as the one presented here, would ultimately have been the better for having a molecular biologist and an ecologist working as a team. This can perhaps be best summed up by the adage "Two heads are better than one", and that what is really needed above all is a more efficient use of people's skills in the pursuit of better evaluation and management of genetic resources.

In addition to the protocols developed for RFLP analysis, and the information obtained about Scottish pine populations, what this project has demonstrated above all, if only by default, is that attempts to combine two disciplines for a new purpose should be able to call upon the support and skills of specialists from both disciplines. Ultimately it will be much more difficult to refine the use of molecular techniques in a conservation context without any input from molecular biologists as well as ecologists. In order to improve the way in which this technology is used in conservation genetics, therefore, the following points should be considered before the start of any project wishing to use genetic makers as part of a conservation effort:

- There should be increased collaboration between the molecular geneticists and the conservation ecologists responsible for the laboratory and field work respectively. This will enable the more technically difficult laboratory procedures to be developed and executed by workers with the appropriate training while at the same time ensuring that the data thus collected is appropriate.

- The worker responsible for the conservation management should be familiar with the basic principles of the molecular techniques that may be used in the study. This will enable the process of selecting a marker to be better targeted at the problem in question, and such projects will ultimately be better designed from the outset, whether or not the laboratory protocols are developed by
somebody else. Equally, molecular biologists should be aware of what the data they are collecting is being used for, and the context within which their work thus lies.

If workers in these two disciplines can be persuaded to learn enough about each other's work to collaborate in this way, efforts to measure and evaluate genetic diversity at all levels in natural populations will produce results more quickly and provide better quality data. Thus, the increasingly urgent problem of genetic conservation will be addressed more effectively.
REFERENCES


Clegg, M.T. (1990) Molecular Diversity in Plant populations. In: *Plant population genetics, breeding, and genetic resources.* (A.H.D. Brown,


APPENDIX 1

PROTOCOLS

DNA Extraction methods.

*CTAB Extraction Method*

1. Weigh out 3-5g of material.
2. Add 1g of *Polyclar AT* (BDH-Merck Ltd).
3. Add CTAB Buffer and mix. For fresh material, add 5mls of 1×CTAB and 5mls 2×CTAB and for freeze-dried material add 10mls 1×CTAB (The concentrated buffer compensates for the water present in the fresh samples).
4. Transfer the mixture, which should have the consistency of a smooth paste, into a tube, and incubate in a water bath for 30 minutes at 60°C.
5. Add an equal volume of Chloroform:IAA, and mix gently to avoid shearing the DNA for 5 minutes.
6. Spin down in a benchtop centrifuge (Wifug 500-E or equivalent) at 3500rpm for 10 minutes.
7. Remove supernatant using a wide bore glass pipette, and decant into an equal volume of Chloroform:IAA, and mix gently.
8. Repeat steps 6 and 7 until no interface appears (usually 2 or 3 extractions required).
9. Remove supernatant into 1 volume of Isopropanol. The DNA should precipitate out in a lump. Using a glass pipette drawn out into a hook, hook out the lump of DNA into a microfuge tube containing 1ml of 70% Ethanol. Still using the hook, remove the lump from the 70% Ethanol and leave to air dry for 10 mins, and then dissolve the pellet in 0.4mls of TE buffer.

*Organelle DNA Extraction.*

1. Weigh out 25g of ground up pine needles into a mortar.
2. Add enough E1 buffer to make a smooth paste, and grind the sample for 10 mins. This method depends on thorough grinding, so this and the following grinding steps are critical.
3. Strain this into a beaker through 2 layers of cheesecloth, with squeezing. Empty the dry grounds back into the mortar.
4. Repeat steps 2-3 a further 2 times.
5. Strain the crude extract of organelles into a 250ml Oak ridge style centrifuge bottle (Nalgene), through one layer of Miracloth (Calbiochem). This should not be squeezed, but left to strain. This greatly improves the quality of the extract later on.


7. Spin down the bottles for 5 mins at 1000rpm in a Beckman JA-14 rotor or equiv. This removes excess nuclear DNA and the larger pieces of debris in the extract.

8. Decant the supernatant into a clean 250ml bottle, and balance them.

9. Spin down for 5 mins at 4000rpm (rotor as before). This step pellets the organelles.

10. Discard the supernatant and resuspend the pellet in a small amount of E2 buffer, using a clean paintbrush.

11. Make up the extract to about 150ml, and balance the tubes, as before.

12. Repeat step 9.

13. Discard the supernatant and resuspend the pellet in a small amount of PTE buffer.

14. Decant this into a 50ml polycarbonate tube, and make up to about 40mls, and balance the tubes.

15. Spin down in a Beckman JA-21 rotor or equivalent fixed angle rotor at 5000rpm for 5mins.

16. Discard the supernatant and wipe the inside of the tube with a clean tissue.

17. Resuspend the pellet in a small amount of PTE buffer, and make up to 40mls and balance as before.

18. Repeat steps 15-16.

19. Make up Proteinase-K solution (1mg/ml). Weigh out 2.5mg per sample and dissolve in 2.5mls of PTE buffer per sample.

20. Resuspend each pellet in 2.5 ml of this solution, and transfer the chloroplasts and Proteinase-K mixture to a clean 10ml sterilin tube (or equivalent).

21. Leave to incubate for 30 mins at room temperature.

22. Add 60μl of 10% Triton-X-100 (v/v) in PTE, and leave to incubate for at least two hours. (This can be left overnight at this stage).

23. Make up the solution to exactly 4mls.

24. Add exactly 4.00g of CsCl to the tube and dissolve it.

25. Add 400μl of EtBr (5mg/ml stock solution) to the tube and swirl it around to mix it thoroughly.
26. Load the gradients into Beckman ultracentrifuge tubes. These amounts are correct for 13×15mm polyallomer ultracentrifuge tubes. If the solution is very gummy then it may be difficult to get the correct amount into the tube and the tube will have to be topped up with a solution of CsCl in PTE (1mg/ml).

27. Balance the ultracentrifuge tubes. This must be done on a fine balance to within 0.001g. The tubes should be full up to the neck, otherwise they deform during the centrifugation.

28. Seal the tubes according to the manufacturer's instructions and spin down using an appropriate program and rotor. We used an NVT-90 (near-vertical) rotor, on a 3 hour sequential program (90-81Krpm). A 16 hour spin at 56Krpm in a vertical rotor can be used as an alternative.

29. After centrifugation, remove the DNA using a wide bore needle (19 gauge) and a 1ml syringe. There is usually a lot of DNA present, and the trick is not to try and recover too much of it, as it is likely that contaminating proteins will also be picked up. Extract the EtBr with CsCl saturated Isopropanol, until the pink colour is gone.

30. Pipette the DNA solution into a piece of prepared dialysis tubing, and dialyse against 1×TE buffer for at least two buffer changes, and at least 4 hours. Decant the DNA into a 1.5ml microfuge tube, and store at -20°C.

**DNA Extraction using High Ionic Strength buffers.**

1. Add 100 mls of cold Isolation buffer to 10g of tissue, and grind in a mortar and pestle to get a smooth paste. Filter through 4 layers of cheesecloth with squeezing. If a lot of particles remain, filter the extract through 1 layer of Miracloth (Calbiochem) without squeezing. Divide the filtrate between two 50ml Sorvall tubes or equivalent.

2. Pellet the chloroplasts in a bench top centrifuge (Wifug 500-E or equivalent) at 3500rpm for 10 minutes.

3. Resuspend the pellets in a total of 10 mls of Isolation buffer, and combine the pellets in one tube.

4. Pellet chloroplasts as before, at 3500 for 5 mins.

The chloroplasts are now lysed to release the DNA. Two methods were tried: for lysis with Triton-X-100, go to step 13 of the Organelle DNA extraction method, and proceed as before. For CTAB extraction, proceed as follows.

5. Resuspend pellets in 9mls of isolation buffer, and add 1ml (1/10th) vol of CTAB stock. Incubate at 60°C for 5 mins.

6. Extract the mixture once with Chloroform:IAA. Decant the supernatant into a clean tube. There is a lot of debris at this stage, and it is quite difficult to pipette.
7. Add 2/3\textsuperscript{rd} vol of cold Isopropanol, and mix gently. The DNA may precipitate immediately, but may need to be left overnight at -20°C.

**Probe preparation and manipulation**

*Transformation of E. coli using Calcium Chloride.*

1. Grow up *E. coli* cells overnight at 37 °C in L-Broth without antibiotic.
2. Dilute the overnight culture 1:20, and incubate for a further 1.5 hours.
3. Transfer the cells to sterile centrifuge tubes, and spin in a benchtop centrifuge, at 3500 rpm for 10 mins.
4. Resuspend the cells in a half volume of 50 mM Calcium chloride (CaCl$_2$). Use a vortex mixer for efficient suspension. Incubate on ice for 30 mins.
5. Spin down as step 3.
6. Resuspend in 1/10th original volume. Incubate on ice for 1/2 to 24 hours. The transformation is more efficient if you leave this incubation for longer, but only up to 24 hours. After that the efficiency decreases.
7. Into a 1.5 ml microfuge tube, put: 1ml of cells
   - 50µl "10mM Solution"
   - 1-10µl plasmid DNA.
   It is best to set up a series of transformations, eg 1, 5 & 10µl. This ensures that you get the best density of transformed colonies. Too little plasmid results in too few colonies, and too much means such dense colonies that you cannot pick out single ones. In addition, a control, with no plasmid DNA should also be set up.
8. Incubate the mixture on ice for 30 mins. Then incubate the tube in a 42°C waterbath for 2mins.
9. Add 0.5 mls of L-broth, and incubate at 37°C for 45 mins. Plate out 0.1mls minimum of transformed cells, onto L-broth agar plates containing the appropriate antibiotic and incubate at 37°C. Sometimes colonies take a while to grow, as they are building up from single individuals. Overnight incubation will often not be sufficient. Note also that you must determine whether the plasmid is relaxed or stringent as the concentration of antibiotic used depends on this property of the plasmid. The control plate, as it contains no transformed DNA should of course show no colonies at all. If colonies appear on the control plates, the methodology should be checked.

**Large Scale Plasmid Preparations.**

Note: Buffers mentioned in this protocol are the same as those required for mini-preps as described in Sambrook *et al.* (1989).

1. Grow a 5ml pre-culture o/n at 37°C, in a shaking incubator.
2. Next day: put 1ml of o/n culture into a 250ml flask containing 50-100mls of L-Broth containing the appropriate antibiotic. Incubate this at 37°C for 4-5 hours.

3. Empty all of this culture into a flask containing 900mls of L-Broth containing the appropriate antibiotic, and incubate o/n as before. You will now have about 11 of culture. The reason for gradually building up the amount of the culture is that the bacteria can go into "shock" if they are decanted into too large an amount of medium.

4. Next day: Take the 11 of culture and divide it equally between 4 sterile 250ml Nalgene centrifuge bottles and balance them.

5. Spin for 5 mins at 5000rpm in a Beckman JA-14 rotor or equivalent.

6. Resuspend the pellets in a total of 25 mls of buffer I containing 100mg of lysozyme, and combine each pair of tubes, leaving two tubes with 12.5mls each.

7. Leave at room temperature for 5mins.

8. Add a total of 40 mls (20mls per tube) of freshly prepated Buffer II (By freshly prepared this means days rather than weeks old).

9. Mix the tube gently by inversion, and leave at room temperature for 10 minutes. It becomes clear and viscous at this stage.

10. Add 30mls (15mls per tube) of Buffer III. Incubate on ice for 10 mins. It should look a lot like egg soup at this stage!

11. Balance the tubes and centrifuge for 10 mins at 10K rpm, in a JA-14 rotor.

12. Decant the contents of each tube into 2 polycarbonate Sorvall tubes, using a wide-bore glass pipette (A total of 4 tubes).

13. Precipitate the DNA with 0.6vols of Isopropanol, usually about 14 mls per tube. This should be left on the bench at room temperature for at least 15mins, but it can be left to precipitate for longer in the fridge or freezer.

14. Balance the tubes and centrifuge them for 10mins at 10K rpm for 30 mins. Drain the tubes and rinse the pellets in ice cold 70% Ethanol.

15. Re-dissolve the pellets in 3mls of TE buffer (0.75mls per tube). Once dissolved, combine two pellets in one 10 ml Sterilin tube (or equivalent) (A total of 2 tubes).

16. Spin down in a benchtop centrifuge (Wifug 500-E or equivalent) at 3500rpm for 6 minutes, and decant the supernatant into two fresh 10ml tubes. Make up the solution to exactly 4mls.

17. Add 320µl of 10mg/ml Ethidium Bromide, and load into polyallomer Beckman tubes as described previously.
18. Proceed as for DNA extraction. Once the centrifugation is complete, the gradient should show two distinct bands; an upper one of bacterial DNA, and a lower one of closed circular plasmid DNA. This band can be extracted without risk of contamination from the upper band by inserting the syringe needle through the side of the centrifuge tube as described in Sambrook *et al.* (1989)

**Excising Probe DNA From Plasmid Vectors**

Gels should be prepared as normal, except that when the gel cools to below 60°C, and before it is poured, Ethidium Bromide should be added to a final concentration of 0.5µg/ml. The digested plasmid is then run on the gel until the fragments are completely separated. This can be monitored by periodically checking the gel in UV light. This is most easily done if the gel former is made of a UV transparent material such as acrylic, as handling of the gel is therefore kept to a minimum. This is of particular importance with LMP gels, as they are more fragile.

**Electrophoresis onto DEAE-Cellulose membrane.**

1. Cut out a piece of Whatman DE81 paper (or equivalent) that is slightly bigger than the size of the well, and soak it in NET buffer.
2. Cut a slit in the gel just ahead of the desired fragment, and insert the DEAE paper. If necessary, place another piece of paper behind the fragment to avoid contamination.
3. Run the gel at a high voltage (70-100V), until the band has completely run into the paper.
4. Elute the DNA from the paper by incubating the paper in HNET buffer, for 2 × 15 mins at 60°C, vortexing every 5 mins.
5. Extract once with phenol, and precipitate with 2 vols ethanol.

**Extraction from Low Melting Point Agarose**

1. Cut the desired fragment from the gel, place in 1.5ml microfuge tube and add 5 vols of 20mM Tris/1mM EDTA buffer. Heat to melt agarose for 5 mins at 60°C.
2. Leave at -70°C for at least 30 mins.
3. Spin down the mixture from frozen in a microfuge at top speed for 10 mins.
4. Extract once with phenol, and precipitate with 2 vols ethanol.
BUFFERS

E1 Extraction Solution
Make up the following solution: 0.3 M sucrose, 10%(w/v) PEG 6000, 25mM HEPES, 2mM EDTA, 1mM CaCl₂, adjusted to pH 6.7 with 10N NaOH.
Autoclave, then add the following: 0.1% (w/v) BSA, 0.6% (w/v) of Polyclar AT and 78μl of βME.
Store at -20°C.

E2 Extraction Solution
Make up the following solution: 0.3 M sucrose, 10%(w/v) PEG 6000, 25mM HEPES, 3mM EDTA, adjusted to pH 6.7 with 10N NaOH.
Autoclave then add the following: 0.1% (w/v) BSA and 78μl of βME.
Store at -20°C.

CTAB Extraction Solution
100mM Tris pH 7.5, 1%(w/v) CTAB, 0.7M NaCl, 10mM EDTA, 1% βME.

Isolation Buffer
Make up the following solution: 1.25M NaCl, 50mM Tris pH 8.0, 5mM EDTA, adjusted to pH 6.7 with HCl.
Autoclave and then add the following: 0.1% (w/v) BSA and 0.1% (v/v) βME.

CTAB Stock
10% CTAB 0.7M NaCl.

"10 mM" Solution
10mM MgCl₂, 10mM CaCl₂, 10mM Tris HCl pH 7.5.

NET Buffer
150mM NaCl, 0.1mM EDTA, 20mM Tris HCl

HNET Buffer
1M NaCl, 0.1mM EDTA, 20mM Tris HCl
LIST OF SUPPLIERS

Amersham International plc,
Northern Europe Region,
Lincoln Place,
Green End,
Aylesbury,
Buckinghamshire, HP20 2TP.

BDH (Merck) Ltd
Head Office and International Sales,
Merck House,
Poole,
Dorset, BH15 1TD.

Beckman Ltd.,
Progress Road,
Sands Industrial Estate,
High Wycombe,
Buckinghamshire, HP12 4JL.

Boehringer Mannheim UK,
(Diagnostics and Biochemicals Ltd.),
Bell Lane,
Lewes,
East Sussex BN7 1LG.

GIBCO-BRL,
P.O. Box 35,
Trident House,
Renfrew Road,
Paisley PA3 4EF.

Sigma Chemical Co. Ltd.
Fancy Road,
Poole,
Dorset BH17 7BR.

Whatman Scientific Ltd.
St. Leonard's Road,
20/20 Maidstone,
Kent ME16 0LS.
APPENDIX 2

ESTIMATION OF DNA FRAGMENT SIZE

Program Listing
LIST
10 DIM WT(50),DIST(50),PROD(50),DWT(50),DDIST(50),DPROD(50),
C(50),D(50)
20 READ N
30 FOR I = 1 TO N
40 READ WT(I), DIST(I)
60 NEXT I
70 SWT = 0
72 SDIST = 0
74 SPROD = 0
80 FOR I = 1 TO N
90 SWT = SWT + WT(I)
100 SDIST = SDIST + DIST(I)
110 PROD(I) = WT(I) * DIST(I)
120 SPROD = SPROD + PROD(I)
130 REM THESE ARE SUM ACCUMULATORS
140 NEXT I
150 MWT = SWT / N
160 MDIST = SDIST / N
170 MPROD = SPROD / N
180 FOR I = 1 TO N
190 DWT(I) = WT(I) - MWT
200 DDIST(I) = DIST(I) - MDIST
210 DPROD(I) = PROD(I) - MPROD
220 NEXT I
230 CSSL = 0
240 CSSM = 0
250 CSCPML = 0
260 CSPMLL = 0
270 CSPMLM = 0
280 FOR I = 1 TO N
290 CSSL = CSSL + DWT(I)^2
300 CSSM = CSSM + DDIST(I)^2
310 CSCPML = CSCPML + DWT(I)*DDIST(I)
320 CSPMLL = CSPMLL + DPROD(I) * DWT(I)
330 CSPMLM = CSPMLM + DPROD(I)*DDIST(I)
335 NEXT I
340 DET = (CSSL * CSSM) - (CSCPML^2)
350 MO = (CSSM * CSPMLL - CSCPML * CSPMLM) / DET
360 LO = (-CSCPML * CSPMLL + CSSL * CSPMLM) / DET
370 SC = 0
380 SSC = 0
390 FOR I = 1 TO N
400 C(I) = (WT(I) - LO) * (DIST(I) - MO)
410 SC = SC + (C(I) - C(1))
420 SSC = SSC + (C(I) - C(1))^2
430 NEXT I
440 CBAR = SC / N + C(1)
450 SDC = SQR((SSC - SC^2 / N) / (N-1))
470 SD = 0
480 SSD = 0
485 PRINT "Stand.d Wt Dist(mm) Pred Wt Deviat.n %Dev.n c(i)
490 FOR I = 1 TO N
500 PREDWT = CBAR / (DIST(I) - MO) + LO
510 WTDEV = WT(I) - PREDWT
520 PERC = 100 * WTDEV / WT(I)
530 SD = SD + WTDEV
540 SSD = SSD + WTDEV^2
545 INTPERC = (INT(PERC*1000))/1000
560 PRINT WT(I);" DIST(I);" PREDWT;" WTDEV;" INTPERC;" C(I)
570 NEXT I
580 SDWT = SQR((SSD - SD^2 / N) / (N-3))
590 PRINT "MO= "; MO; " LO= "; LO; " CBAR = " ; CBAR
600 PRINT "SC= "; SDC; " SD= "; SDWT
610 READ N
620 FOR I = 1 TO N
630 READ D(I)
640 PREDWT = CBAR/(D(I)-MO)+LO
650 PRINT "FOR A DISTANCE OF "D(I)"mm", "PREDICTED LENGTH = ";PREDWT
660 NEXT I
670 DATA 6
680 DATA 23130,9.9
690 DATA 9416,20.6
700 DATA 6557,28.3
710 DATA 4361,39.9
720 DATA 2332,60.5
730 DATA 2027,64.4
780 DATA 2
785 REM MIGRATION DISTANCES
790 DATA 15.8,56.6
800 END

Example Output
RUN
Stand. d Wt Dist(mm) Pred Wt Deviat. n %Dev. n c(i)
23130 9.899999 23023 107.0039 .462  213442.9
9416 20.6 9653.835  -237.835  -2.526    207882
6557 28.3 6554.127   2.873535  .043    212585
4361 39.9 4210.471  150.5288  3.451    218339.4
2332 60.5 2306.683  25.3169  1.085    214009.5
2027 64.4 2085.877 -58.87696  -2.905    208783.3
MO= 1.153818 LO= -1274.12 CBAR= 212507
SC= 3805.6  SD= 177.7356
FOR A DISTANCE OF 15.8 mm PREDICTED LENGTH = 13235.26
FOR A DISTANCE OF 56.6 mm PREDICTED LENGTH = 2558.551
Ok_