Evolution of Hybrid Larch (*Larix x eurolepis* Henry) on the Atholl Estate in Scotland

Anthony Martin Langdon

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at Edinburgh University

Institute of Ecology and Resource Management
University of Edinburgh
October 1998

To Berry

I declare that this thesis, and all work contained therein, is entirely my own.

CONTENTS

Acknowledgements

Thanks are due to the following
The members of my lab, past and present, who helped me through the project in various capacities- Billy, the ever-smiling Eric Easton and of course John, who all deigned to help, occasionally uncomplainingly, a mere Englishman.
Richard, my supervisor, for being so supportive and encouraging, and having a highly developed (and in this business, adaptive) sense of humour.
The Atholl Estate and the Dunkeld House Hotel for their co-operation.
All members of Adrian Bird's lab, who extended much appreciated help and assistance with the use of their facilities.
The stokers of the 'engine-house' of the fourth floor, Joan and Aileen, for not being too busy to talk to me.
Morag, for putting up with sharing a flat with me for nearly 4 years.
My friends doing cushy little numbers up in the I.C.M.B. for providing friendship, and knowing what it means to have an experiment fail: Craig, Jo, Darren and Simone.
The staff at the Junction Bar, West Preston Street.
Theo for scientific and emotional support, helping me through my difficult times with Lara, and being a top drinking partner.
Mum, Dad, Jack and Chris.
And finally Berry, for everything.

Abstract

Plant hybridisation is a complex process that may be an important factor in the evolution of the world's flora. Many studies and much speculation have centred on the phylogenetic consequences of hybridisation. There is, however, a dearth of information on the processes governing the interactions between allopatric species that have only recently been bought into contact.

This study uses molecular techniques to study the consequences of hybridisation between two exotic larch species on the Blair Atholl estate in Scotland. Japanese larch (*Larix leptolepis* Mill.) and European larch (*Larix decidua* Mill.) have been hybridising on the estate since the turn of the century. The hybrid, the well known Dunkeld larch (*Larix x eurolepis* Henry), has been used in forestry around the world and forms a large part of the estate's silvicultural output.

A suite of molecular markers, consisting of isozymes, nuclear DNA markers and cytoplasmic markers were used to analyse the composition of several stands on the estate. All the Japanese ancestors of the hybrids and a selection of the likely European parents were also examined, as were controlled crosses between the two species.

The use of organellar markers allowed inferences regarding the directionality of hybridisation to be made.

Genetic parameters, including cytonuclear disequilibria utilising both nuclear and cytoplasmic data, were calculated, and the use of these parameters is discussed.

One stand was found to contain no hybrids at all, rather a mix of pure parents. Another two stands derived entirely from Japanese seed parents but with a large European pollen contribution. Analysis of these stands in terms of cytonuclear disequilibria and genotypic distributions demonstrated that one was a first generation stand, and the other a second generation.

Regenerants in two of these stands were also examined. Results suggested that the regeneration in the hybrid stands derived largely from the parental stands with limited seed and pollen influx from the neighbouring parental populations.

Larch plantation on the Atholl estate are seen to be a complex mix of backcrosses and first and second generation hybrids, and are fertile and viable. Implications of the results for silviculture on the estate, for the possible evolutionary future of the hybrids in Scotland, and the study of the process of hybridisation in general are discussed.

1. Hybridisation

1.1 Introduction To Hybridisation

Hybridisation is recognised as a process that may be significant in the evolution of many taxa (Rieseberg and Ellstrand 1993). Studies on butterflies (Jiggins *et al.* 1996), birds (Grant and Grant 1992) and fish (Pyle and Randel 1994) show that it is of some importance in the animal kingdom. It is amongst plants, however, that hybridisation has attracted most interest, and it is plants that will be the focus of this discussion. The importance of hybridisation in agriculture, horticulture, and silviculture has long been appreciated, and more recently naturally occurring hybrids are attracting interest. This is not surprising, given that up to 20% of the British flora comprises of hybrids (Stace 1991), a figure which is likely to be mirrored in other regions with a high proportion of alien species such as North America (Mooney and Drake 1986). Studies of natural hybrids to elucidate the evolutionary history of such taxa are also relevant to the question of how related species remain distinct (Barton and Hewitt 1989) and to speciation in general. Furthermore, a full understanding is necessary to cope with the impact it may have on phylogenetic trees (Mcdade 1990; Mcdade 1993).

The literature shows that a variety of methodologies have been used to study hybridisation. Unfortunately, the results have not been unified into one conceptual framework. In this introduction I shall review the literature, with particular emphasis on the possible consequences of genetic contact between two species and how different approaches have attempted to answer this question.

1.1.1 Definition Of Hybridisation

One of the largest contributory factors to the confusion surrounding hybridisation is uncertainty as to its definition. It is clear that hybridisation does not mean the same thing to all biologists (Arnold 1997). Most authorities on hybridisation adopt an extremely liberal definition, such that natural hybrids are the result of crosses between individuals from populations that are distinguishable on the basis of at least one heritable character (Arnold 1997; Harrison 1990; Rieseberg and Wendel 1993). However, the same authors focus exclusively on inter-specific hybridisation and its importance with respect to speciation or the maintenance of species barriers. Thus the usual definitions of hybridisation make no explicit distinctions between taxonomic levels, but do make an implicit one. This is to avoid tackling species concepts, which are all challenged by hybridisation, but leads to confusion when its evolutionary and phylogenetic implications are discussed.

One of the most widely used species concept is the phylogenetic species concept (Cracraft 1989), which suggests species are defined solely by ancestry. This definition is the cause of most of the problems in discussions of hybrid speciation and hybrid relationships- it is unavoidable that an individual that can trace its ancestry through two species presents complications to this view. It would be fruitless to avoid using this concept in a review of the literature. However, it is possible that a classification based, at least in part, on adaptive and ecological factors, would be more useful. Under rigid interpretation of the biological species concept (Mayr 1942), any interspecific cross that results in even partially fertile offspring would force the two parents to be reclassified as subspecies. This has usually been enough for those working on hybrids to reject the biological species concept, or variants of it. A slightly more fluid version might still allow species to be defined on the basis of properties of groups of organisms (Templeton 1989), while not being negated by inter-specific hybridisation. I shall adopt the approach that species be defined on the ecological, morphological and adaptive traits of the populations that appear to make up that species. This (somewhat circular) approach makes it explicit that species definitions are primarily artificial constructs based on utility. The inter-fertility of *individuals* from two species need not lead to them being reclassified, though it may be that the long term consequence of hybridisation will be that only one species need be used to describe all the populations.

If a distinction between species and populations is made on the grounds of utility, it is also the case that hybridisation will, in part, be defined on the basis of whether it is a useful concept. Thus, I shall reject the usual assertion that inter-population gene flow is hybridisation and explicitly state that only mating between species (and sub-species, an unavoidable outcome of the huge variety of artificial taxonomic divides that have been made) be regarded as hybridisation.

1.1.2 Hybridisation In Agriculture, Silviculture, And Horticulture

Ever since agriculture was developed, hybridisation has been used to introduce desirable characteristics into crop species. As an example, the common hexaploid wheat *Triticum aestivum* is thought to have been the result of a cross between a tetraploid wheat and a wild diploid grass (Janich *et al.* 1981). Agricultural research has long been preoccupied with creating new varieties by crossing suitable parents (Janich *et al.* 1981). Tree breeding, while much less advanced in this respect (largely due to the much longer generation time of trees), also utilises the principle to some success (Zobel and Talbert 1981). Another example is the

horticulturist's creation of unusually coloured or patterned flower varieties exemplified by the wide range of primrose hybrids available.

It is clear from these considerations that artificial hybrids can have desirable and, to some extent, predictable properties. One use of artificial crossing is to combine desirable parental traits, and to use the F_1 generation directly. This can be achieved by planting admixtures of the parental types with the aim of producing commercially viable quantities of hybrid seed- a seed orchard (Zobel and Talbert 1981). When a parental trait is expressed in a hybrid, this is often due to simple single or dual locus control of a trait (Gottlieb 1984), making it easy to backcross and introduce the relevant gene into one parent's genetic background if this is what is required. Such introgressive hybridisation, the introduction of a limited amount of genetic material from one taxon into another, is a second, and more common, use of hybridisation. Examples of this are the backcrossing of (inter-specific) Jeffrey x Coulter pine hybrids to introduce the latter's weevil resistance into the better growth form of the Jeffrey pine (Zobel and Talbert 1981).

Finally, many artificial F_1 hybrids show hybrid vigour (Jinks and Jones 1958). Improved performance of hybrids over the parents - characterised by faster growth, better resistance and other commercially important traits - has often been noted. It is of much importance in agriculture and horticulture (Janich *et al.* 1981). It is also seen in trees, for example *Eucalyptus* (Venkatesh and Vakshasya 1977), though until recently it was generally thought to be of limited use in tree breeding (Fowler 1978). Though hybrid vigour is not a universal character of hybrids, or the mechanisms by which it may occur agreed upon, it is of both commercial and theoretical interest.

One final reason that hybridisation is of importance to plant breeders is becoming more and more evident- the escape of trangenes from genetically modified organisms. As disease, pest, or herbicide, resistance plants are increasingly being produced through genetic engineering, fears that such genes might escape and have a negative ecological impact is increasing (Adam and Koehler 1995). Introgression of engineered genes into weedy relatives has been shown in both sunflower in the U.S (Whitton *et al.* 1997), and rape in Europe (Baranger *et al.* 1995). The responsible attitude is that hybridisation *will* lead to the 'escape' of such genes, and it is the selective and ecological impact of such escape that should be assessed.

Despite the importance of hybridisation in plant breeding, the rest of this review will concentrate on the issues pertaining to *natural* hybridisation, to determine its importance in the evolutionary history of the plant kingdom.

1.1.3 Allopolyploid And Diploid Hybrid Speciation

Probably the most important consequence of hybridisation is the formation of new species through allopolyploidy. Allopolyploidy, where the nuclear genome consists of multiple genomes from diploid species, is well documented, and is a frequent cause of plant speciation (Grant 1981). Homoploid taxa that have arisen from hybridisation (such that the chromosome number of the hybrid taxa is the same as both parents), in contrast, is comparatively rare. However, this discussion will focus largely on such taxa, or on the processes giving rise to them, and some justification is required. Allopolyploidy is far easier to detect - morphologically, polyploids are often quite distinct from their diploid ancestors, and molecular studies can easily provide unambiguous evidence of their hybrid nature (Soltis and Soltis 1993). This means the phylogenetic confusion that such hybrids generate is limited. More importantly, their existence does not teach us so much about the interactions between species. As a new species can, in principle, arise in just one generation through such an event, there is no need to analyse the dynamics that allow such hybrid speciation. In the case of diploid hybridisation, hybrid individuals are generally expected to be able to mate with the parent species. Thus, if a diploid taxon has arisen from hybridisation, the question of how it was or is maintained as a discrete unit arises. This may tell us about barriers between species and interactions between them, and how much gene flow occurs between them - for example, does the existence of hybrid individuals lead to homogenisation of the parental gene pools? Allopolyploids, on the other hand, are not expected to interact genetically with their parent species and so have no such implications. Thus, the almost complete neglect of allopolyploids in this discussion is not to deny their importance in speciation and the generation of morphological and genetic novelty - where they are likely to far exceed hybrid diploid species - but arises from a focus on other issues.

1.1.4 What Are The Consequences Of Hybridisation?

Most studies of hybridisation attempt to throw some light onto the question of the consequences of bringing two inter-fertile species into contact. The possibilities, which are not mutually exclusive, and often depend on how taxa are classified, are threefold. Firstly, there may be gene flow between the species, or introgression. This in itself can have a number of possible outcomes (and definitions) that will be enlarged on later. Secondly, it may lead to speciation. Finally, it may result in the production of localised hybrid zones which consist of a mixture of parents and hybrids. The purpose of this review is to examine the different

approaches that have been used and discuss how far they have gone to determining the importance of each of these outcomes.

1.2 Use Of Species Specific Genetic Markers In The Study Of Hybridisation

Whatever the approach to hybridisation, all studies must utilise genetic markers of some form. A genetic marker is any trait that is determined by the genetic constitution of the parents, and can be morphological as well as the molecular markers commonly used. The definition will be restricted somewhat to include only those traits for which the inheritance is understood- that is, traits that can be used to infer details of the parentage of an organism. When applied to hybridisation, a useful genetic marker is a trait which distinguishes between the parental species. If the parental species are not identified, as is often the case with species of ancient hybrid origin, these are traits that are assumed to have distinguished between the putative parents prior to speciation. Such markers are usually referred to as species-specific markers, the assumption being that they are present in the vast majority of one species and in very few of the other. Prior to a discussion of hybridisation, I shall introduce the various markers that are often used in its study.

1.2.1 Morphological Markers

Most hybrids have been initially identified on the basis of their morphology (Knobloch 1971). A species that appears intermediate in many respects to two related species may often be supposed to be the outcome of hybridisation. In this sense, it could be stated that morphological markers are commonly used to detect hybrids.

However, morphological markers have not been widely used in the further confirmation of hybrid origin, or the analysis of hybrid zones. This is because it is very rarely that clear predictions can be made regarding the inheritance of morphological traits. Hybrids may express intermediate, parental, or even novel characters (Wilson 1992), making it impossible to determine parentage by examining these markers unless the inheritance of these traits is rigorously analysed. A telling example is with artificially produced F1 hybrids between several species of *Aphelendra* (Mcdade 1993). The traits analysed were all morphological, and demonstrated clearly that not all hybrids show the same relationship to their parents (on a phylogenetic tree). If it can be difficult to infer relationships from morphological traits with controlled F_1's, it must be supposed to be even more difficult when later generations of hybrids are involved.

Thus, it is therefore not usually possible to put much confidence in the classification of hybrids without utilising molecular methods, and the majority of this discussion will focus on the different classes of molecular markers available. It should be noted, however, that their utility is in no way related to their importance- in fact, those markers that are easiest to use and most informative in population genetics may turn out to have little bearing on the consequences of hybridisation.

1.2.2 Molecular Markers

In a sense, the distinction between phenotype and 'genotype' (as revealed by molecular markers) can be artificial. Unless the molecular marker is a sequenced region of DNA (where genotype can be directly observed), even molecular markers are the products of interaction between DNA, the cell, the environment and the experimenter (i.e. phenotypes), just as morphological traits are. Not all markers that are referred to as molecular markers yield the same level of information regarding genotype. Thus, molecular markers are simply those that have been analysed using the *techniques* of molecular biology, and will include chemical compounds, isozymes, RAPDs, and RFLPs as well as sequence information. This is to avoid the very artificial distinction that has been made between nuclear markers and 'supranuclear' markers (if such a distinction is required, it should be between DNA markers and non-DNA markers).

1.2.2.1 Nuclear Markers

Nuclear markers are those that are inherited through the nucleus, which carries the vast bulk of genetic information. Sexually reproducing organisms are defined by the fact that their nuclear genome is inherited equally from mother and father.

An extremely important facet of almost all studies of nuclear variation is the 'neutral theory'. This theory, originally put forward to explain the large amount of genetic variation present at isozyme loci, suggests that the vast majority of variation -particularly at the DNA sequence level- has no selective effect (Kimura 1983). It has been largely supported by empirical evidence, though is still a matter of some debate (Gillespie 1987). Thus, a species-specific marker may not be expected to cause the differences that lead to the recognition of two separate taxa, or be affected by the differences between them.

The first such molecular markers to be used on a wide scale were isozymes, which are variants of a protein that differ in their shape, size, or charge and can be differentiated on the basis of mobility through a gel in the presence of an electric field. They are usually inherited in

a simple Mendelian fashion. These have been used widely in population genetics since the 1960's (Hillis and Moritz 1996). There are a limited number of isozyme loci, each with a limited amount of variation present. Moreover, it is rarely possible to infer the relationship between two alleles. It is for these reasons, as well as technical ones, that DNA markers are more applicable in many studies (Hillis and Moritz 1996). However, it is also fair to say that isozymes are often rejected for fashion's sake alone (Easton 1997), and certainly the use of species-specific isozyme markers is applicable to the study of hybrid populations.

DNA manipulation such as cloning and the Polymerase Chain Reaction, whereby manageable sequences of DNA are amplified, and the use of restriction enzymes which recognise specific DNA sequences, allow a whole range of techniques for the analysis of genetic variation (Hillis and Moritz 1996). Of these, microsatellites, RFLPs, RAPDs and sequencing are the most commonly used. Microsatellites are fast evolving loci made up of short repeat units and have too high a rate of mutation to be useful for most interspecific comparisons (Hillis and Moritz 1996). Randomly Amplified Polymorphic DNAs (Hadrys *et al.* 1992) are useful as no *a priori* knowledge of the organism's genome is required. However, they are dominant, i.e. it is not possible to distinguish heterozygotes from homozygotes, and they therefore have serious limitations (Lynch and Milligan 1994).

A common technique in the analysis of the nuclear genome of hybrids is Restriction Fragment Length Polymorphism (RFLP) analysis. RFLP analysis is simply the characterisation of a locus, or a set of loci, based on the presence or absence of sequences that are recognised by restriction enzymes (Hillis and Moritz 1996). 'Probes' which recognise the locus under consideration are constructed from clones or PCR, and a variety of restriction enzymes are used to digest the target DNA. This approach, while the mainstay of much molecular biology, has been largely restricted to the ribosomal (rDNA) genes (Waters 1995) in plant population genetics. Recently, sequencing has become feasible even for large surveys (Soltis *et al.* 1997), but again generally only ribosomal sequences are utilised.

Ribosomal genes were originally used because they are multi-copy and thus suitable for RFLPs (Hillis and Moritz 1996). Furthermore, they have regions that evolve slowly (the genes themselves) and relatively rapidly (the ITS regions) making them applicable at a variety of taxonomic scales (Soltis *et al.* 1997). Though they are multi-copy, the paralagous loci interact with each other. This 'concerted evolution' acts to homogenise the various copies of the genes, which is what makes them useful in taxonomy (Baldwin *et al.* 1995). However, it also means they may be particularly error-prone when taxa have hybrid histories (Sang *et al.* 1995). The

contribution of one parent's ribosomal genes may be erased by this process over a short evolutionary time. This has actually been demonstrated in *Gossypium* (Wendel *et al.* 1995).

Thus, despite the revolution in molecular biology that allows the characterisation of so much of the nuclear genome, the analysis of the nuclear genome of hybrids at the DNA level has largely been restricted to the use of one, imperfect, set of loci. A corollary of this is that more, single-copy, nuclear genes must be characterised (Strand *et al.* 1997; Waters 1995) before DNA markers can really fulfil their potential in this field.

1.2.2.2 Use Of Cytoplasmic Markers.

The organelles - the mitochondria and the chloroplast (but excluding the nucleus, which technically is also an organelle)- are essential cellular components of all plants. They both contain their own genomes which have, in recent years, been extensively used in both taxonomy (Olmstead and Sweere 1994) and ecological/population genetics (Bachmann 1994). Two methods for analysing their variation are widely used, RFLP analysis using probes constructed from organellar sequences, or the analysis of organellar sequences that have been amplified using the PCR reaction.

There are several reasons why the variation present in organelle genomes provides additional information beyond that supplied by the nuclear genome. The first is their (generally) uniparental inheritance patterns (Mogensen 1996). Chloroplasts are maternally inherited (i.e. through the seed alone) in almost all angiosperm, while most conifers show paternal inheritance. Maternal inheritance of the mitochondria is also the rule, though again there are exceptions amongst the conifers (Mogensen 1996).

This unidirectional pattern has many applications, such as the construction of maternal (or paternal) lineages (Neale *et al.* 1988), something that has not been possible with nuclear or morphological markers. In recent years, it has become evident that it is possible to use such information in an ecological setting, as pollen and seed are produced at differing times, in differing amounts, and travel different distances. Thus, they can be used to study patterns of seed and pollen flow (Ennos 1994). Similarly, population structures should be easier to detect when looking at maternal markers, as these reflect only seed flow rather than seed and pollen flow. As pollen is more highly dispersed, one expects a tighter population structure with maternal markers than with nuclear markers. This was the case in a study of *Silene alba*, where Fst values were much greater with cpDNA markers (Mccauley 1994). These differences between seed and pollen production and movement will make organellar markers potentially extremely useful in understanding the dynamics of natural hybridisation.

One consequence of their inheritance is that organelles are asexual. This means recombination does not reorganise linked polymorphisms. One way of looking at this is that "the ancestry of individual haplotypes may remain recognisable even after many generations of sexual reproduction" (Wang and Szmidt 1994). In some instances, it is useful to think of it as increasing the number of alleles available at a single genetic locus, which emphasises that examining multiple organellar sequences is not equivalent to examining multiple nuclear loci.

Another feature of organellar genomes that makes them useful is they have different rates of evolution to the nuclear genome (Palmer 1990). The rate of evolution (with respect to point mutations) is lower in both organelles than in nuclear markers, so they are useful for constructing deeper phylogenies. This means that comparisons between species or even genera, where many nuclear sequences may have been saturated with mutations so that resolution of relationships is not possible, are feasible with organellar DNA (Szmidt 1991). In conjunction with their haploidy, which reduces the effective population size for organelles and makes intraspecific variation lower (Birky *et al.* 1983), this makes organellar interspecific variation easier to analyse.

1.2.2.3 Limitations of Organellar Markers

Though these are good theoretical reasons why organellar DNA is suitable for many studies, another reason for its widespread use is that it is easier to work with, at least in the case of the chloroplast. This is due primarily due to its abundance (many copies of each organellar genome to one nuclear genome), which facilitates DNA analysis techniques such as Southern hybridisation ,and its conserved sequence which allows the utilisation of probes and primers over a wide range of taxa (Gielly and Taberlet 1994). Mitochondrial studies in plants are relatively rare - over the last 6 years, there have been nearly 3 times as many studies employing cpDNA variation as mtDNA variation. Mitochondrial genomes in plants exhibit a far higher rate of size variation and reorganisation than either nuclear or chloroplast genomes (Palmer 1992), such that restriction analyses of the whole genomes are difficult and rare. Even when performed, they cannot be readily applied across species in the way cpDNA studies can (Pradhan *et al.* 1992). This problem is not helped by the large size of the mitochondrial genome, the lower abundance of mitochondria relative to chloroplasts, and the difficulty of extracting pure mitochondrial DNA. However, since the introduction of PCR based techniques, it has become technically easier to utilise mitochondrial genes in phylogenetic and population studies, which should lead to mtDNA being used more often to complement chloroplast studies.

The relevance of the difference in usage between the two organelles is that it highlights that technical, rather than theoretical, considerations can play the major part in determining choice of genetic markers. However, there are often theoretical reasons for choosing a different marker. It has been suggested that chloroplasts cross species boundaries more readily than nuclear genes (Rieseberg and Soltis 1991), implying that organellar based phylogenies are more susceptible to distortion from reticulate evolution than nuclear phylogenies. This is an even greater problem given that no amount of sequencing can change the fact that an organelle is essentially a single genetic locus. Thus, results based on, for example, the chloroplast genome can only be replicated by utilising the mitochondrial genome as well (Ennos *et al.* 1998).

Another problem is that possible selective pressures on the organelles are often ignored. As with nuclear markers, the variation in organellar genes is generally assumed to be neutral. There are two consequences of this. The first is that explanations for variation based on selection at the loci under consideration are often ignored. An example in plants is a study of mitochondrial variation in two species of pine (Dong and Wagner 1993). While the probes used were important coding sequences of the mitochondria (COXI and COXII), no mention of the neutrality assumption or possible selective pressures was made. This is particularly likely to be a problem in the plant mitochondria, as its non-conserved organisation means most of the regions analysed have been coding .

A second and more general problem with ignoring selection is due to the non-recombining nature of organelles. Any selectively advantageous mutation that arises will cause any other variants on the same organelle to be subject to exactly the same selective forces, raising their frequency and eventually bringing them to fixation. The net result is that an episode of selection on any organellar variation will cause a 'selective sweep' whereby the overall organellar variation is reduced (Marayuma and Birky 1991). How large an effect this has on overall patterns of organellar diversity depends on how often selectively advantageous variants are created (Ennos *et al.* 1998). Little is known about this, but what examples there are in both plants (Hanson 1991) and animals (Templeton 1996; Wise *et al.* 1998) suggest selection on organellar variation could well have an effect. It is, however, quite rare that neutrality assumptions are discussed with reference to the patterns of variation found.

These problems suggest that there remains much uncertainty as to the behaviour of organellar genomes on an evolutionary time scale, and that at least a partial return to analysing the nuclear genome is essential, particularly for phylogenetic reconstruction.

1.2.3 Cytonuclear Disequilibria

Organellar variation is frequently used on its own for phylogenetic purposes, or to assess the extent of population subdivision. Sometimes it is used in addition to nuclear variation to refine a phylogeny, or provide alternative estimates of gene flow. However, a more recent approach is to use it in conjunction with nuclear variation, meaning that data is collected and analysed simultaneously for both genomes, which allows a novel approach to the study of many of the traditional parameters in population models (such as gene flow, selection, drift etc.). Such 'cytonuclear' data is useful largely because it allows associations between cytotypes and nuclear genotypes to be observed (Basten and Asmussen 1997). Such associations may, theoretically at least, be a result of selection (Paige *et al.* 1991), assortative mating (Cruzan and Arnold 1994) and differential seed and pollen flow (Schnabel and Asmussen 1992). In the same way that studying disequilibria amongst nuclear loci allows locus specific effects to be separated from genome wide effects, analysis of cytonuclear disequilibria should allow the separation of effects on the genome as a whole (e.g. drift) and those that affect only paternal or maternal markers. Such information could be useful to the study of naturally occurring hybrid zones. Coadaptation between organellar and nuclear genes, such as cytoplasmic male sterility, is one phenomenon which could be approached in this way.

The theoretical background behind cytonuclear disequilibria, though relatively new, is really an extension of traditional linkage disequilibria (Weir 1990). See Appendix C for full derivations, taken from Asmussen *et al.* (1987). The cytonuclear allelic disequilibrium measures associations between alleles at a nuclear locus and the cytotype, and is directly analogous to the gametic disequilibrium for a diallelic system. The three cytonuclear genotypic disequilibria, which measure the association between a cytotype and a particular genotype can also be calculated. One of these genotypic disequilibria, the allelic disequilibrium, the Hardy-Weinberg disequilibrium, and the allele frequencies are sufficient to describe, in an intuitive way, all the associations that can occur between one nuclear locus and one cytoplasmic locus. The theory has become sophisticated enough to incorporate such modifications as exact tests for small sample sizes (Basten and Asmussen 1997) and the analysis of dominant markers such as RAPDs (Dean and Arnold 1996), and it appears capable of generating questions and answers in the field of evolutionary biology. Whether it has been applied successfully yet, at least to plants, is another question. It is uncertain what realistic conditions may lead to measurable cytonuclear disequilibria, and whether the measurement of these disequilibria allows the testing of biologically plausible hypotheses.

1.2.4 Genetic Markers: Conclusions

The study of plant hybridisation has relied on the same markers that are used in the rest of plant phylogenetics and population genetics. Morphological studies have been replaced by studies using molecular markers. Of these, isozyme, rDNA and organellar loci are the most widely used. These markers are almost always neutral, or at least assumed to be so, and the behaviour of adaptively important traits is rarely analysed using the powerful molecular techniques available.

While each of these suites of markers is extremely powerful, they each have limitations. A lack of nuclear markers may partly have contributed to the reliance on organellar markers, which are not always suitable. In particular, their suitability for analysing 'population' processes- over a recent time-scale- may not extend to evolutionary time-scales (Ennos *et al.* 1998). rDNA markers require cautious interpretation as a result of concerted evolution, and isozymes also have many limitations. RAPDs have been used extensively, but as their limitations become more evident it is likely that they will fall out of fashion, in all but a few areas (such as classification of clones).

It is possible that the development of new nuclear markers may help ameliorate some of these problems. Furthermore, it is possible that the use of cytonuclear disequilibria could allow the analysis of existing markers to yield more information (though this requires testing). However, the limitations of the markers used must be borne in mind while interpreting the results of any study.

1.3 Approaches to the Study of Plant Hybridisation: Pattern and Process

Armed with the various genetic markers available, students of plant hybridisation have followed two approaches to determine both the importance and mechanism of hybridisation in evolution. These have been termed 'pattern' based approaches and 'process' based processes (Arnold 1997). Pattern based approaches involves systematic studies- often not geared towards or expecting to find hybridisation- and making inferences about past evolutionary events. Process based approaches study the dynamics of natural hybridisation by directly observing the interaction between two inter-fertile taxa, generally in naturally occurring hybrid zones. These distinctions have been criticised on the basis that pattern is just the consequence of process, and that analysis of process in any case is just making inferences from patterns (Nixon and Wheeler 1990). A case in point is probably the best known example of an invasive hybridisation event- the origin of a radiate variety of the common British groundsel, *Senecio*

vulgaris. The radiate variety (var. *hibernicus*) arose following introgressive hybridisation from the radiate alien *Senecio squalidus* into the non-radiate native (var. *vulgaris*) (Abbott *et al.* 1992). The spread of the alien since its introduction in the 18th century has been well documented (Abbott and Lowe 1996), and as it is a recent event it could be debated whether the study here is of 'pattern' or 'process'. This indicates that the divide is by no means absolute.

However, the distinction is useful for two reasons. Firstly, they separate the literature into two groups that whatever their logical equivalence might be, are distinctive as regards the methodology of the study. Secondly, it is by no means certain that the pattern of hybridisation that is seen in systematic studies is a direct consequence of the sorts of processes that occur in those hybrid zones that have been studied.

First of all I shall review the process based studies- those which attempt to directly observe the dynamics of natural hybridisation- and then draw conclusions as to what has been learnt from this approach and what questions remain unanswered. This will be followed by a discussion of the pattern based approaches.

The study of the dynamics of plant hybridisation is still very much in its infancy. As with most evolutionary biology, it is very difficult to observe the processes directly and it may not be strictly correct to label it dynamics. However, there is a definite shift from the identification and classification of hybrids, to trying to understand what factors govern the production and future of hybrid populations. From this, it may be possible to learn about hybrid speciation or, more commonly, how hybridising species remain distinct from each other.

1.3.1 The Process of Natural Hybridisation: Hybrid Zones

The usual approach to studying the dynamics of natural hybridisation is to analyse the population genetic make up of areas characterised by high numbers of hybrids. These areas- hybrid zones- are generally the result of secondary contact (i.e subsequent to the divergence of the parents) between two inter-fertile species.

Mathematical models of hybrid zones tend to focus on how such populations can exist without leading to widespread interbreeding and introgression. The most mathematically tractable of these is the 'Tension Zone Model' (Barton and Hewitt 1985). The approach used in this and related models is to assume selection against hybrids (endogenous selection) with dispersal of parental types into the zone maintaining the equilibria. Clines in different molecular and morphological traits should generally be concordant and linkage disequilibria (both nuclear and cytonuclear) should be high (Arnold 1993a; Barton and Gale 1993). Clines should be

fairly narrow, as environmental selection is assumed not to have an effect within the hybrid zone, and stability of hybrid zones is also an expected outcome of this model.

The predictions of this model are borne out in studies of many animal hybrid zones (Hewitt 1988). However, the existence of these patterns is explicable by other models as well (Arnold 1997), so that the assumption of hybrid unfitness is not tested simply by conforming to these predictions. Furthermore, plant hybrid zones are not adequately described by this model. Alternatives, such as the 'Mosaic' model (Harrison 1986), the 'Bounded Hybrid Superiority' model (Moore 1977) and the 'Evolutionary novelty' model (Arnold 1997) include the effect of environment, or exogenous selection. The Mosaic model simply states that the environment can have a selective effect (in contrast to the entirely endogenous selection acting in a Tension zone), and notes that environments are distributed patchily, explaining the heterogeneous distribution of hybrids in many zones. However, this is such a general approach that, whilst clearly correct in many cases, it lacks predictive power - and the discovery of zones with a mosaic structure says little about the forces maintaining that structure. The Bounded Hybrid Superiority model assumes superiority of hybrids in a hybrid zone and inferiority outside the zone, with dispersal not playing a major role. Like the Tension Zone model, this can account for stable hybrid zones. The Evolutionary Novelty model also assumes superiority of some hybrid genotypes in some environments, but also suggests that selection against early generation hybrids could account for the clinal pattern often seen and predicted by the Tension Zone model. Furthermore, this approach emphasises the long term future for some hybrid genotypes even outside the original contact zone.

1.3.2 Examples Of Hybrid Zones

Here, I will describe those hybrid zones that have proved most useful in understanding the complex nature of natural hybrid populations. Earlier studies tended to be geared towards demonstrating the existence of hybrid populations; more recent studies have attempted to understand their dynamics and it is those that are reviewed here. See Table 1-1 for a brief overview of examples of plant hybrid zones that have been analysed with molecular markers (varying degrees of morphological analysis are part of these studies). It is by no means exhaustive, and avoids studies which do not include geographical, ecological, or temporal data as they are uninformative in this regard.

Table 1-1: Molecular analyses of hybrid zones

TAXA	MARKERS	CONCLUSIONS	SOURCE
Aesculus	Isozymes	Broad hybrid zone Asymmetrical gene flow.	(DePamphilis and Wyatt 1990)
Carduus nutans/acanthoides	Isozymes, rDNA, cpDNA	Stable hybrid zone? No dispersed introgression.	(Warwick *et al.* 1989)
Iris fulva/brevicaulis	RAPDs, cpDNA	Rare formation of F_1 s. Selective and ecological factors important.	(Cruzan and Arnold 1994)
Iris fulva/hexagona	RAPDs,cpDNA	Contact between populations largely pollen mediated.. Asymmetrical gene flow.	(Arnold *et al.* 1991)
Iris tenax/thompsonii *douglasiana/innominata*	cpDNA	Discordance between clines. Possible movement of zone.	(Young 1996b)
Picea engelmannii/ sitchensis	rDNA, cpDNA, mtDNA	Rare formation of F_1 s. Complex backcrosses. Asymmetrical gene flow.	(Sutton *et al.* 1994)
Picea rubens/marians	rDNA, mtDNA, cpDNA	Frequency of hybridisation dependent on ecological factors.	(Bobola *et al.* 1996)
Pinus pumila/parviflora	cpDNA,mtDNA	Differential distribution of markers	(Watano *et al.* 1996)
Pinus taeda/echinata	cpDNA, Isozymes	Asymmetrical hybridisation Recent, no introgression.	(Edwards-Burke *et al.* 1997)
Populus fremontii/ angustifolia	mtDNA, RFLPs	Movement of hybrid zone. Unexplained cytonuclear disequilibria.	(Paige, Capman and Jennetten 1991)
Quercus grisea/gambelii	RAPDs	Low levels of introgression.	(Howard *et al.* 1997)
Quercus petraea/robur	RAPDs	Asymmetrical hybridisation. Hybridisation as dispersal	(Bacilieri *et al.* 1996b)
Silene dioica/latifolia	Isozymes	High levels of hybridisation, but no effect on taxonomic boundaries.	(Goulson and Jerrim 1997)

Notes: Dispersed Introgression refers to evidence for introgression outwith the area characterised by hybrid formation

1.3.2.1 Louisiana Irises

The Louisiana irises provide the most extensively studied examples of hybrid zones. Several studies utilising ecological and genetic data, both nuclear and organellar, have been performed to try and unravel not only the structures of the hybrid zones, but the forces that produce those structures. It has long been recognised that hybrid populations are relatively common in irises, but no evidence that introgression affects the stability of the species was found using morphological data (Randolph *et al.* 1967). Furthermore, despite the fact that hybrid

populations are common, spontaneous hybridisation has been shown to be a rare event (Arnold 1993c).

In the case of *I.fulva* and *I.hexagona* , several allopatric parental populations and one hybrid population were analysed for nuclear (RAPD) markers and chloroplast markers (Arnold, Buckner and Robinson 1991). The allopatric populations contained low levels of nuclear markers from the other species, which had already been indicated by isozymes and rDNA (Arnold *et al.* 1990a; Arnold *et al.* 1990b). Such introgression was not found with the chloroplasts, hinting that contact between populations of the two species may be largely through pollen, as might be expected. Furthermore, the distribution of chloroplast haplotypes amongst the hybrids was somewhat patchy, with one open region containing only the *hexagona* type, suggesting that this subpopulation was the result of pollination of *hexagona* by *fulva.* In a subsequent study of a hybrid zone that consisted of a third species, *I.brevicaulis* along with *fulva* and an allopatric population of *hexagona*, it was found that no hybrids had the *hexagona* haplotype but they often contained its nuclear markers (Arnold *et al.* 1992). These results all demonstrate the differential role of seed and pollen in the formation of hybrid zones. This approach was extended by collecting detailed ecological data in populations consisting of *I.fulva* , *I.brevicaulis* and their hybrids (Cruzan and Arnold 1993). The inclusion of phenological information and cytonuclear data from seeds allowed a detailed examination of the mating system (Cruzan and Arnold 1994). The hybrids were found to be largely advanced generation backcrosses to one parent species or the other with intermediate or F_1 individuals very rare; this suggested that the first hybridisation events occurred some time ago and few F_1s have been formed since. Indeed, intermediate seed aborted more frequently than seed with parental or near parental genotypes. Furthermore, it appeared that *fulva* mated assortatively, whereas *brevicaulis* mothers did not discriminate against *fulva* pollen. These patterns gave rise to significant cytonuclear disequilibria, though it should be added that the formal analysis in terms of the disequilibria parameters did not add qualitatively to the studies. The distribution of hybrids was also interesting, with the *brevicaulis* backcrosses competing effectively with their parents, but backcrosses to *fulva* largely confined to a habitat distinct and distant from either parent's, suggesting a positive role for selection on some hybrid genotypes. It is notable that there is also a recognised species, *I.nelsonii,* with some similar ecological and morphological traits (but more distinct from the parents) that has been shown to be of the same hybrid origin (Arnold 1993b). It has been suggested that this species is the outcome of the same kind of hybridisation event.

1.3.2.2 Pacific Coast Irises

Pacific coast Iris hybrid zones have also been studied with cpDNA markers. Limited by utilising only morphological and not nuclear markers to allow separation of pollen and seed contributions, there are still interesting conclusions to be drawn. In the case of the *I.douglasiana/I.innominata* zone, three transects were taken which all showed a cline in both haplotype and morphological traits (Young 1996b). Furthermore, the morphological cline was displaced from the chloroplast cline. This could be a result of the encroachment of *douglasiana* territory by *innominata* (for example, if ecological factors have shifted). If morphology accurately reflects the nuclear make up of the hybrid zone, then one would expect the higher pollen flow relative to seed flow to lead to residual haplotypes of the displaced species within the encroaching one. Alternatively, this displacement could be caused by asymmetric mating success, a condition which is not unlikely considering the studies of the Louisiana irises (though field observations suggest this is not the case here). Interestingly, in the other Pacific coast hybrid zone, between *I.tenax* and *I.chrysophylla* , the two clines were concordant. That there appeared to be stronger barriers to cross-fertilisation (Young 1996a) between these two species, but stronger habitat selection in *douglasiana* and *innominata* suggests the two zones, even in such closely related and geographically close species, may be maintained by different historical, ecological and genetic factors not provided for in existing hybrid zone models.

1.3.2.3 Cottonwood

Even in the absence of ecological data, the patterns of genetic markers alone can lead to some information about the behaviour of hybrid zones. For example, studies on a hybrid zone between cottonwood populations (*Populus angustifolia* and *P.fremontii)* with both nuclear markers (Keim *et al.* 1989) and maternal markers (Paige, Capman and Jennetten 1991)- in this case mitochondrial - suggested movement of the hybrid zone. In an analogous situation to the Pacific coast irises, there were residual *fremontii* haplotypes in a largely *angustifolia* nuclear background at one end of the hybrid zone, suggesting encroachment on the former's territory by the latter. However, there were also cytonuclear disequilibria that could not be explained easily on the basis of differential seed and pollen flow. A model based on cytoplasmic male sterility could account (partly) for these observations (Cellino and Arnold 1993), but it is also possible that more detailed ecological and mating system data would have yielded an alternative explanation. Either way, none of the existing hybrid zone models prove to be useful in describing this situation.

1.3.2.4 Oaks

The white oaks provide particularly interesting opportunities to study hybridisation. Though many species have been shown to be interfertile, they often exist sympatrically with only occasional formation of hybrid populations (Whittemore and Schaal 1991). Where hybrid populations are seen, they have generally been analysed morphologically. However, the well documented discord between their large morphological distances and their close genetic relationships (Howard *et al.* 1997) suggest that such analyses may not provide a reliable estimate of hybrid nature (Aas 1993). Using molecular markers, a hybrid zone between two North American white oaks (*Quercus grisea* and *Quercus gambelii*) was shown to be patchy in its makeup (Howard *et al.* 1997), rather than clinal as in the cases of cottonwood and the Pacific coast irises. However, it was not possible to infer the causes of this structure. Despite this, it is clear that models based on clinal distributions are inadequate for such populations.

1.3.3 The Utility Of The Hybrid Zone Models For Describing Natural Hybridisation

In summary, hybrid zones often exist over larger geographical ranges than predicted by the Tension Zone Models (DePamphilis and Wyatt 1990), they have a patchy distribution with the frequency of hybrids varying from area to area as with the irises and also spruce (Bobola *et al.* 1996; Perron and Bousqet 1997), and often consist of advanced generation backcrosses rather than F_1s and F_2s (Cruzan and Arnold 1993; Paige, Capman and Jennetten 1991; Perron and Bousqet 1997). There are no examples of taxonomic boundaries breaking down even in areas characterised by a high degree of hybrid formation, and inferences regarding such processes can only be regarded as highly speculative.

The Tension Zone model (TZM) makes the very clear assertion that selection against hybrids can cause stable, clinally varying hybrid zones. However, the structure of contact zones in plants is often patchy and does not conform to the pattern predicted by this model. Even where it does, for example the Pacific Coast Irises, this does not necessarily provide evidence for this model over other explanations. Importantly, that hybrid zones are generally stable equilibria is not suggested by available data either. A variety of other models have been put forward, none of which have been able to make useful generalisations about the structure of hybrid zones as they take into account exogenous effects which are different in each case.

However, the limited use of the TZM (and mosaic model) as formulated does not imply that their basic assumption - that hybrids are generally unfit - is wrong. Rather than analyse the genetic structure of a hybrid zone and hope to obtain evidence for or against mathematical models such as the TZM, the following approach testing the premises of the models directly is

suggested. First, the assumption that the zone is at equilibrium must be questioned. This is by no means easy to test, but discordance of maternal and nuclear markers (as in, for example, the cottonwood) may suggest that the structure seen is a result of recent invasion of one species' territory by the other. Also, most hybrid zones consist of very few F_1 hybrids, suggesting that the initial hybridisation events are rare, also arguing against equilibrium based approaches. If possible, analysis of more than one generation, as with the European oaks (Bacilieri *et al.*,1996b), should provide a valuable insight into the temporal dimension of hybridisation. The results of this approach strongly suggest that equilibrium is not the norm which does not disprove the basic premises of the models, but makes detailed mathematical treatments of questionable value. Secondly, in cases where the zone is patchy, there should be more emphasis on environmental associations- this approach worked well in Louisiana irises where one class of hybrids was shown to exhibit greater fitness than its parents in some environments. Where the structure of the zone is a clinal change that fits with the TZM, this could still be a result of selection for hybrids in this zone rather than selection against hybrids and parental migration (Moore 1977). In this case, the only way to differentiate between the two very different hypothesis is to test experimentally for hybrid fitness.

Where the fitness of naturally occurring hybrids has been tested, it has been found that reduced fitness is by no means the norm (Hodges *et al.* 1995). In light of this, a completely different approach to hybrid zones is suggested by Arnold (Arnold 1997). This 'Evolutionary Novelty model', which treats hybridisation as adaptive rather than maladaptive is useful in that it focuses on the important ecological and environmental factors that are ignored by the TZM. It is exemplified by the Louisiana irises. However, this example may prove to be the exception rather than the rule. It would perhaps be better to rename this the 'Ecological determinants' model to remove the implication of an *a priori* assumption that hybridisation must be adaptive for the methodology suggested by this approach to be a useful one.

1.3.4 Other Approaches To Studying The Processes Of Hybridisation

Naturally occurring hybrid zones have been studied in order to obtain data about the dynamics of hybridisation. However, just determining the spatial distribution of genetic markers has failed to generate robust hypotheses concerning the future of the hybrids. The "natural laboratory" that hybrid zones provide (Hewitt 1988) do not allow the investigator to separate the various factors (selection, migration, drift e.t.c.) that cause their structure. One way of rectifying this is by doing transplant experiments to test the fitnesses of parents and various classes of hybrids under different ecological conditions (Emms and Arnold 1997).

Another approach that is likely to become more common is to emulate naturally occurring hybridisation under more controlled conditions. The *de novo* bringing together of inter-fertile species allows observation of the processes governing the early evolution of hybrid populations, particularly when the tendency of sympatric parents to swamp the hybrids is reduced. This situation has arisen in Britain between two species of *Larix,* and it is this that will be studied in the remainder of this investigation. The value of this approach is obvious from a study of white oaks (Bacilieri *et al.* 1996b). In a planted hybrid stand of the European white oaks *Q.petraea* and *Q.robur, petraea* pollinated *robur* far more than the reverse, and the population was gradually becoming more *petraea* dominated (Bacilieri *et al.* 1996b). This agrees with ecological data which suggest that *robur* is a pioneer species, but is succeeded by *petraea*- and suggests that pollen flow and inter-specific hybridisation may play a part in the succession. This has also been seen in *Eucalyptus* (Potts and Reid 1988). It should be pointed out that the close genetic relationship and the existence of F_1 oak hybrids of almost pure parental form (Bacilieri *et al.* 1996a) may suggest that the significant morphological and ecological differences in these species are caused by only a few genes (Stebbins 1950). Though this does not preclude them being treated as separate species, it means the dynamics of interspecific hybridisation in oaks may be considerably different from genetically more distinct species.

A variation of this approach was the artifical synthesis of *Helianthus annus* and *H.petiolaris* hybrids (Rieseberg *et al.* 1996). Mapped genetic markers were used to analyse in detail the genomic composition of three independently produced lineages. The interesting conclusion from this study was that all three lineages evolved to have very similar compositions. Furthermore, this compostion was similar to a naturally occurring hybrid, *H.anomalus.* This demonstrates the importance of selection in the evolution of hybrids, and that hybrids that become established do so because they are not just random combinations of genes from each parent.

Finally, mathematical and computer based models of such situations may make it possible to determine the conditions that may be important and the parameters that are worth measuring. This is not a widely used approach- the models that have been used are designed to explain the supposed stability of hybrid zones. In order to provide a theoretical basis for the use of various genetic markers in examining the early evolution of newly created hybrid populations, a computer model will be applied to *Larix* in Britain.

1.3.5 The Process of Natural Hybridisation:Conclusion

Despite the increasing body of work on plant hybrid zones, no model has provided a useful paradigm for examining their structure. It is still unknown whether selection for or against hybrids plays a large role, and no conclusion as regards the maintenance of distinct species in the face of hybridisation is easily drawn. The view that hybrid zones are natural laboratories for testing evolutionary hypotheses (Hewitt 1988) has perhaps led to the neglect of the important experimental tests on hybrid fitnesses and fertilities that would help to answer these questions. Hybrid zones that are stable are likely to represent local and evolutionarily unimportant situations. Hybrid zones that are unstable are not described adequately by the TZM and mosaic zone models, and these should no longer be used to guide investigations or be the focus of debate. The majority of studies that purport to examine the processes of natural hybridisation have actually managed to avoid observing those processes directly. An inclusive approach, whereby ecological factors and selective forces are measured is likely to be far more instructive. This approach, adopted by Arnold in the study of irises is labour intensive but facilitates the testing of hypotheses regarding the eventual fate of hybrid zones, whereas without such information one appears to be limited to the conclusion that hybrid zones do, indeed, exist. However, it is entirely plausible -even probable- that this approach will also demonstrate that most hybrid zones do, after all, reflect selection against hybrids. Rather than being "natural laboratories", they may represent evolutionary dead-ends that are decidedly inappropriate for the study of evolutionary forces.

1.4 Approaches to the Study of Hybridisation(2): Inferring Hybridisation from Pattern

An alternative approach is to interpret the results of systematic (usually phylogenetic) surveys in such a way that hybridisation is included in the analysis as a factor that may influence tree topology. This is a more common approach, if only because the data-set includes the very many studies that uncovered evidence of reticulate evolution without the study of hybridisation being the investigator's main concern. While this approach should never ignore the dynamics that may have led to the observed pattern, it may not be necessary to have a detailed understanding of the interaction between two species to draw conclusions. In fact, the process orientated approach has been so limited in its application that it is this pattern approach that has been seen as the more informative of the two.

Studying hybridisation based on systematic patterns involves answering two linked questions. Firstly, to what extent can the ability to test for reticulate evolution be improved (either

through experimental method or analytical procedure)? Secondly, what does the observed pattern tell us about the significance of hybridisation in plant evolution?

As hybridisation is often used to explain discrepancies in phylogenetic trees (Bain and Jansen 1996; Soltis and Kuzoff 1995; Wendel *et al.* 1991) the first question must be answered before the second can be tackled. This requires an understanding of the phylogenetic effect of hybridisation.

1.4.1 The Position Of Hybrids In Phylogenetic Trees

Traditional phylogenetic methods are designed to represent hierarchical relationships, and are not able to adequately pick up or represent reticulate evolution (Rieseberg and Ellstrand 1993). It has further been suggested that hybrids can affect the relationships between non-hybrids in the same tree (Cronquist 1987; Funk 1985). Even where the existence of hybrid taxa is already known it can still be problematic to predict their position and effect on a phylogenetic tree (Mcdade 1990).

There are several approaches to dealing with hybrids but a common method is to include all data in the phylogenetic analysis and try to identify hybrids from diagnostic patterns in the final tree. The problem is that there are no clear predictions as to where hybrids will appear in phylogenetic trees. This is particularly the case with morphological analyses, as shown by the artificially produced F1 *Aphelendra* (Mcdade 1993), though the problems of representation and interference also apply to molecular phylogenies (Rieseberg and Ellstrand 1993). Thus, though many of the putative cases of hybridisation have been suggested by morphology, it is to the patterns of hybridisation as suggested by molecular phylogenies that I shall turn.

1.4.2 Using Molecular Phylogenies to Infer Hybridisation

The issue of hybridisation is one that has emphasised the differences between phylogenies based on morphological and molecular data sets. It has often been the case that molecular data sets have rewritten traditional morphological phylogenies (Hillis and Moritz 1996). This has been largely accepted as an indication of the greater resolving power of molecular methods, particularly with neutral markers. There is no doubt that it is much easier to determine an exact mechanism of inheritance and mutation for molecular than for morphological variation, and it is therefore the case that molecular data should ultimately be more suitable for providing detailed information about evolutionary history. This is particularly the case in recent years, when widespread sequencing has made it possible to obtain data that allows the exact interrelationship of polymorphisms to be inferred. The tracing back of genes to their

common ancestors based on sequence (the coalescent process) should allow the phylogenetic history of neutral markers to be inferred to a high degree of confidence (Hudson 1990). However, throughout the following discussion, it should be borne in mind that the extension of these to biologically important relationships between taxa is a matter of debate, particularly if hybridisation has occurred.

1.4.2.1 The Effect Of Hybridisation On Nuclear Genome Phylogenies

A good example of a phylogeny based on nuclear markers that has provided evidence of hybrid speciation is in *Helianthus* section *Helianthus*, North American relatives of the sunflower (including *Helianthus annus* mentioned earlier). RFLP mapping of ribosomal genes was performed on a large number of individuals from over 20 taxa (Rieseberg *et al.* 1991b). The resulting tree was largely congruent with a morphologically based phylogeny, the main difference being the treatment of three putative hybrid species. That these were hybrid was convincingly shown from both the ribosomal data, where the hybrid individuals contained RFLP patterns diagnostic of both parents, and previous isozyme studies (Rieseberg *et al.* 1990b). This study demonstrates that it is possible to achieve a nuclear based phylogenetic reconstruction, without *a priori* knowledge of hybridisation, that detects the occurrence of reticulate evolution.

However, if hybridisation has occurred but not been detected, the phylogeny will have erroneous implications for the evolutionary history of the taxa. Similarly, diagnosing a hybridisation event that has not occurred will lead to misconceptions about not only the taxa in question, but the nature of evolution and speciation itself. It is important, therefore, to maximise the chance that hybridisation is detected, without inferring it where it has not happened. Alternative approaches to analysing the data are possible (Rieseberg and Ellstrand 1993) but no overall protocol for phylogenetic reconstruction that takes into account possible reticulate evolution has been taken up. Instead, hybridisation is usually cited as the probable cause whenever a phylogenetic uncertainty arises, without an analysis of the statistical significance of such an uncertainty. However, for any individual locus, there are many other explanations for how it may lead to a false phylogeny - insufficient data (Saitou and Nei 1986), selective effects, high variance of evolutionary rates (Wilson *et al.* 1990), and phylogenetic sorting (Maddison 1997). Thus, inferences regarding the hybrid origin of taxa based on the extant pattern of variation are necessarily inexact- particularly if the hypothetical hybridisation happened long in the past. Examples such as the sunflowers are robust- there are relatively large numbers of species-specific markers and continued work has confirmed the

hybrid origin of a least one of the species (Rieseberg *et al.* 1995). However, many examples are based on far fewer markers, and some of these may be explicable by chance events alone. Even were an improved method of analysing the data that specifically deals with reticulate evolution taken up, it is unlikely that this would greatly increase either the certainty with which hybrids, particularly ancient hybrids, are detected, or the ability to reject hybridisation. The only approaches likely to achieve such a goal are to increase the number of samples included in a phylogenetic analysis and to replicate the data by increasing the number of loci sampled. The first point is important as while the taxa in question may not be of hybrid origin, there may be individuals that contain introgressed markers if hybridisation has occurred since cladogenesis. It is now well known that sampling of such individuals without sampling non-introgressants is a possible source of error (Soltis *et al.* 1993). The second point is at least as important. The need to increase genome coverage for good phylogenetic reconstruction is well known (Doyle 1992). This is particularly important where hybridisation is an issue, as not only may individuals exist that are not representative of the taxa as a whole, loci may exist that are not represent of the genome as a whole. Unfortunately, the nuclear genome is rarely sampled at multiple loci, except for techniques such as RAPDs and isozymes which are not, as mentioned, generally applicable. Even if ribosomal genes weren't so problematic *per se*, it is clearly inadequate to rely on one source alone. However, there is a lack of other loci known to have sufficiently slow rates of evolution to be useful (Waters 1995). While some effort is being made to find regions which can be used across many species (Strand, Leebens-Mack and Milligan 1997; Waters 1995), the shortage only serves to increase the reliance on rDNA sequences, or organellar sequences.

1.4.2.2 The Effect On Organismal Phylogenies Inferred From Organellar Phylogenies

An example of a phylogeny based on chloroplasts that was subsequently re-evaluated using nuclear markers is *Heucheura*. Originally, a cpDNA phylogeny of 20 members of this group was constructed (Soltis *et al.* 1993). When multiple ribosomal sequences of the same taxa were analysed, there were large, well-supported differences in the tree topology (Soltis and Kuzoff 1995) also supported by isozyme and morphological data. Another clear example is in *Gossypium*, where species that clearly belong to one section on morphological grounds contain chloroplast genomes closely related to those from other sections (Wendel and Albert 1992). This phenomenon has been referred to as chloroplast capture, though its exact meaning is not always clear.

Furthermore, it has been suggested that organelles cross species barriers more easily than nuclear genes (Rieseberg and Soltis 1991), which would result in a greater frequency of chloroplast capture after hybridisation than might be expected. If so (and this will be expanded on later), it would clearly be the case that organellar phylogenies will react differently to hybridisation events. However, even in the absence of easier cytoplasmic transfer, much has been made of the consequences of chloroplast capture for chloroplast based phylogenies.

The real issue is the extent to which a single-locus phylogeny can provide information about the history of a taxon in which hybridisation has taken place. Extensive analysis of the chloroplast will give a robust single locus phylogeny, which can be extended to organismal phylogeny only if hybridisation has never occurred. On the other hand, if hybridisation has occurred- and the evidence is that it can rarely be ruled out- a more informative phylogeny will use more loci, even if the phylogeny of each locus is less robust. This is because it then becomes possible to not only detect hybridisation, but to quantify the contribution of each parent to the hybrid taxon. The difficulty of inferring either parentage, or evolutionary importance of hybridisation, from one molecular marker alone is seen in a study of the arctic grass *Poa hartzii* (Gillespie *et al.* 1997). In this example, the putative hybrid species is clustered with one parent on the basis of morphology, but some populations include a chloroplast from another species. However, the lack of nuclear markers made it impossible to conclude that the chloroplast genome was behaving any differently from other loci. Certainly, the inferences that "Chloroplast DNA transfer can occur with little or no apparent nuclear transfer" and that hybridisation "played an important role...in the evolution and speciation of arctic grasses" were largely unsupported. Rather than questioning the validity of chloroplast based phylogenies (Soltis and Kuzoff 1995), one should question the validity of using any single locus phylogeny to make inferences about past hybridisation events.

1.4.3 Conclusions: Inferring Hybridisation from Pattern

Molecular markers are clearly suited to detecting past hybridisation events. However, the limitations place on the investigator who has no way of testing the alternative hypotheses- which may have been the results of random drift and other unknown contingencies- should not be overlooked. As with all other evolutionary studies, the further in the past the event, the harder it is to provide other evidence to test the hypothesis. These caveats in mind, we can now ask what this approach, as opposed to the study of process in hybrid zones, has taught us about hybridisation.

1.5 Patterns Of Reticulate Evolution(1): Introgression Outside Hybrid Zones

The study of naturally occurring hybrid zones, while inconclusive with regards to generalisations about the long term consequences and implications of hybridisation, has clearly demonstrated plausible mechanisms for interspecific gene flow, or introgression (Anderson 1949). Whether hybrids are unfit or fit, whether they establish themselves as new evolutionary lineages or not, and whether they have any ecological significance, it is clear that genes have the potential to cross from one taxon to another through hybridisation and backcrossing. The radiate form of groundsel is the classic example of a morphological trait - now believed to be controlled by a single gene (Abbott 1998) - that has crossed from one taxon to another. However, breeding programmes aside, introgression of morphological and selective traits has rarely been convincingly detected (Rieseberg and Wendel 1993). It is the discovery of high levels of introgression of *molecular* markers, however, that has been one of the most surprising discoveries of molecular systematics (Arnold 1997). This suggests extreme caution for molecular phylogeneticists using low sample sizes and limited loci and indicates that hybridisation may have played a role in the evolution of almost any taxon. Here, I will focus on what studies outside localised or ephemeral hybrid zones- those that have adopted the pattern based approach- have concluded about inter-specific gene flow.

1.5.1 Definitions Of Introgression

A distinction is made between 'dispersed introgression' and 'localised introgression' (Arnold, Buckner and Robinson 1991). Dispersed introgression refers to the occurrence of markers from one taxon in another taxon outside areas of contact, and it can occur even when no hybrids or hybrid populations have previously been identified (Wolfe and Elisens 1994). Localised introgression refers to the existence of populations or subpopulations with such markers within a hybrid zone. Studies on hybrid zones find much localised introgression - in areas where hybridisation is occurring, many, if not most of the hybrids will have involved backcrossing and therefore be termed 'introgressants' (Arnold *et al.* 1992; Keim *et al.* 1989; Watano, Imazu and Shimizu 1996). Strictly, this usage is correct. However, examining localised introgression involves a splitting of the parental taxa into taxa that are taking part in hybridisation and those that are not. In this case, introgression clearly affects parental taxa- but this is a somewhat artificial divide. Thus, the following discussion of introgression will focus on the possibility that introgression leads to evolutionary change or cladogenesis in taxa

that have not been defined by their involvement in hybridisation. This equates to examining introgression outwith contact zones.

Similarly, stabilised introgression is the production of intraspecific taxa that are largely related to one species but contain markers from another, and have arisen by repeated backcrossing of individuals of the 'donor' species to individuals of the 'parental' species (Anderson 1949). Hybrid speciation is the production of taxa that are roughly equidistant between each parent and is caused by descent just from the F_1s. Both are extremes of the same continuum, and it has been argued that they be treated as the same phenomenon (Abbott 1992). This is logically consistent. However, it leads to the conclusion that the spread of a gene from one taxon throughout another is a speciation event.

These two treatments of introgression are logically and mechanistically correct. However, they lead to much confusion in the literature. Introgression is here treated as an evolutionary force akin to mutation- it can lead to changes in a taxon without necessitating the reclassification of that taxon (Harrison 1990). This requires that the basis of classification of taxa is independent of the presence of introgressant markers. A taxon that has already been recognised and is then found to contain introgressive markers can reasonably be classified as a hybrid species (Soltis and Kuzoff 1995). On the other hand, what about a population that has been grouped with other populations, and is then found to contain an introgressive marker (Rieseberg *et al.* 1990a)? Under the phylogenetic approach usually adopted, this would require that such a population is given a new taxonomic status. Other classification systems may be more forgiving.

1.5.2 Examples Of Introgression

Even if introgression within hybrid zones is discounted, there are still very many examples of apparent gene transfer across species barriers (Table 1-2). One such example has already been mentioned, where nuclear markers from one species of iris were found in allopatric populations of another species (Arnold *et al.* 1992). This is an interesting example, as the same marker applied to sympatric populations and hybrid zones demonstrated entirely plausible mechanisms for such gene flow.

Another clear example is found in the new world cottons, *Gossypium barbadense* and *G.hirsutum*, where allozyme and RFLP markers found at high frequency or fixation in one species were found at low frequencies in some accessions of the other species (Brubaker *et al.* 1993; Wendel and Albert 1992). Furthermore, levels of introgression varied between accessions and depended on whether the accession was a modern cultivar or a feral population

- but in a way that could be accounted for by the known history of that accession. While much of the introgression was clearly a result of breeding practices, there were also significant levels of introgression from wild populations of one species into wild populations of the other.

Evidence of introgression can also be found even when a significant level of hybridisation is not obvious. In *Peltanthera,* rDNA markers diagnostic of *Penstemon centranthifiolus* were seen in some populations of *P. grinellii* (Wolfe and Elisens 1994) throughout its range, including areas of allopatry between the two species. No other evidence of widespread hybridisation had been found. This demonstrates that introgression may be even more widespread than the apparent frequency of hybridisation.

Table 1-2: Introgression outside contact zones

TAXA	INTROGRESSED MARKERS	SOURCE
Aesculus	Morphological, Isozymes	(DePamphilis and Wyatt 1990)
Gliricidia sepium	Chloroplasts	(Alavin *et al.* 1991)
Gossypium barbadense/hirsutum	Isozymes (+ low levels of organellar)	(Brubaker, Koontz and Wendel 1993)
Helianthus annus/debilis	Chloroplasts, Morphological, Ribosomal	(Rieseberg, Beckstrom-Sternberg and Doan 1990a)
Helianthus annus/petiolaris	Chloroplasts	(Dorado *et al.* 1992)
Heuchera hallii	Chloroplasts	(Rieseberg and Soltis 1991)
Iris fulva/hexagona	RAPDs, Morphological	(Arnold, Buckner and Robinson 1991)
Juniperus virginianum/scopularum	Morphological, Secondary compounds	(Flake *et al.* 1978)
Pinus densata	Chloroplasts	(Wang and Szmidt 1994)
Quercus	Chloroplasts	(Whittemore and Schaal 1991)

NOTE: This table does not include the very large number of taxa where past hybrid speciation appears responsible for introgressive patterns of variation. Rather, it is intended as a guide to those species which may still be incorporating genes from other extant species.

1.5.3 Detecting Introgression

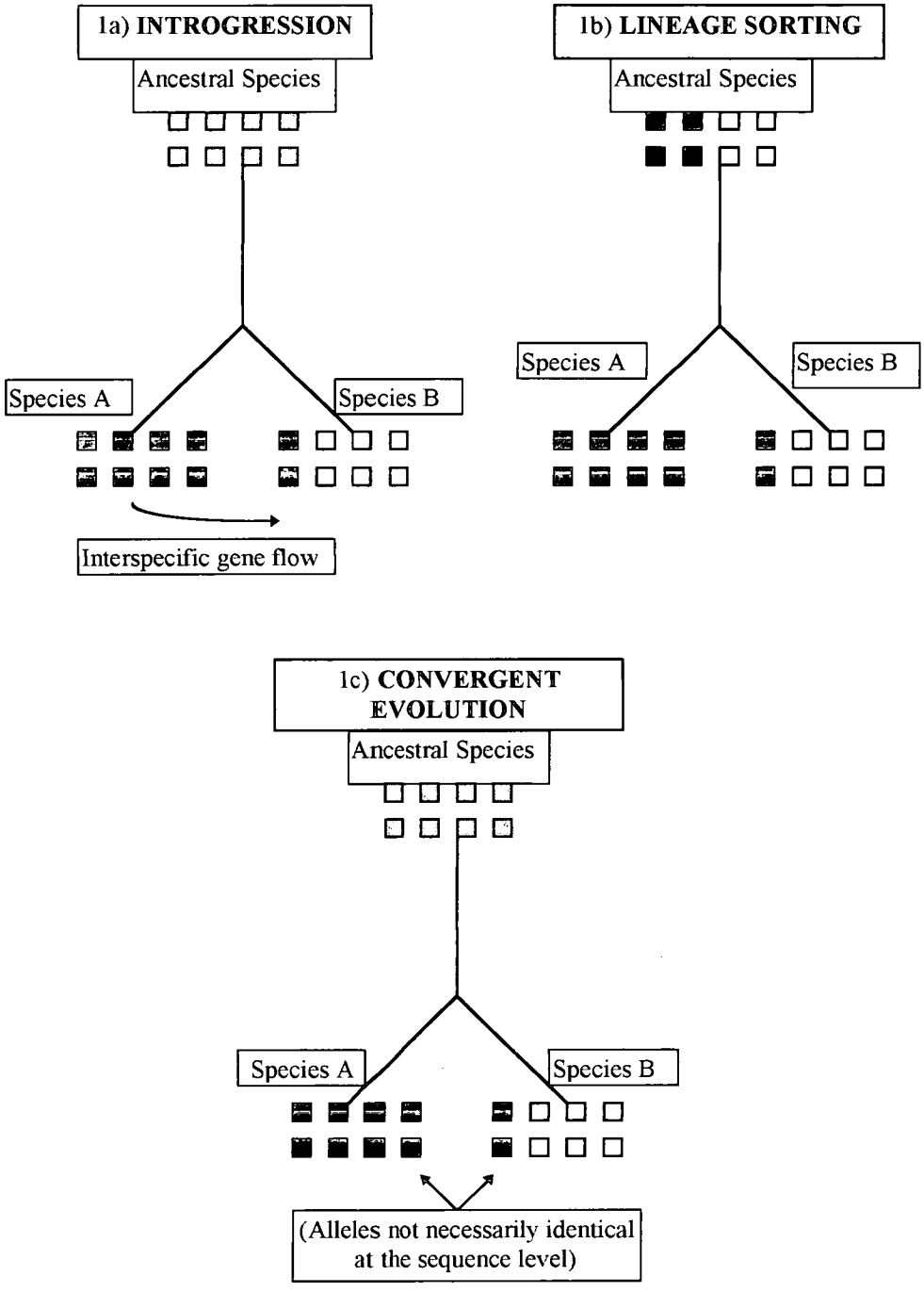
These studies often have other data to support the hypothesis of introgression (in *Gossypium,* for example, the patterns make sense based on the known history of the accessions) and are fairly robust. However, they also highlight the difficulty with detecting introgression reliably. To quote from the study on cottons (Brubaker, Koontz and Wendel 1993) "Our inferences on introgression are based on the assumption that molecular markers that are fixed or nearly fixed in one species and occur in low frequencies in the other do so as a consequence of historical interspecific gene flow" (Figure 1-1). While this is not generally stated explicitly,

this is indeed the assumption that must be made where only allele frequencies are known. Other explanations exist, such as the existence of variation in the ancestral species (Nei 1987) (Figure 1-1), and convergent evolution - including the misidentification of non homologous alleles as homologous as can happen easily in RAPDs (Rieseberg 1996) and isozymes (Figure 1-1).

This problem of identifying alleles of introgressive origin could lead to large ascertainment biases. It is obvious that lineage sorting may give rise to overestimates of the levels of introgression. Without *a priori* expectations about the level of variation within a taxon for each marker, it may not be possible to distinguish the two hypothesis. If enough loci are examined, and the ancestral species had high levels of variation, it would be expected to find some that displayed the 'introgressive' pattern due to drift alone. However statistical tests to confirm introgression are rarely, if ever, performed.

Another ascertainment bias can lead to underestimates of the frequency of introgression. If inter-specific gene flow has been common, it will be difficult to find species diagnostic *neutral* markers. This should not lead to the conclusion that introgression has not occurred, as appears to have happened in North American white oaks (see pg 41). Furthermore, the behaviour of a few markers that have been chosen for their species-specificity should not be taken as indicating absolute levels of introgression. An obvious error of this kind is seen in the North American oaks *Q.gambelii* and *Q.grisea (Howard et al. 1997)*, where allopatric populations were screened for RAPD markers. Six species diagnostic markers were found. That these markers "demonstrated geographical consistency" was used to argue "that each species is distinct". However, this is circular reasoning, and nothing can be inferred about levels of introgression between the species from this study (though these markers were then used successfully to study a hybrid zone).

Figure 1-1: Different evolutionary histories can give rise to the same 'introgressive' pattern of variation



1.5.4 Asymmetrical Introgression

An asymmetry of introgression between two taxa is often found. This suggests that it is often easier for genes to pass from one species to the other than vice-versa (Table 1-2). This is not a surprising result. If ecological and environmental effects play as large a role in hybridisation as the process approach appears to suggest, there is no reason why one should expect symmetrical gene flow. Thus, such patterns are unlikely to reflect the different selective regimes in each species as is sometimes suggested (Howard *et al.* 1997). However, that asymmetrical introgression is so common can be taken as evidence for two hypotheses. Firstly, models of hybridisation that ignore ecology do not account for the patterns of introgression that are found, as has already been argued. Secondly, it provides an additional line of evidence against lineage sorting being responsible for 'introgressive' patterns of variation. This is because though drift could lead to apparent introgression at multiple loci, it would be expected to occur symmetrically.

1.5.5 Cytoplasmic Introgression

Many of the examples of introgression in Table 1-2 involve cytoplasmic markers. This is uncontroversial - organelle markers are the most widely used plant genetic markers, and are therefore likely to have uncovered many of the examples of introgression. However, another conclusion that has been drawn from these studies is that cytoplasmic introgression occurs more frequently than introgression of nuclear genes- in other words, that "nuclear genes may be exchanged less freely between species than organellar genes" (Rieseberg and Wendel 1993).

This conclusion has been drawn because of the large number of 'chloroplast capture' events seen. Chloroplast capture is not usually adequately defined. I shall use it to refer to cases when a taxon's most common chloroplast haplotype and its nuclear genome can be shown to have derived from different taxa. Thus defined, it refers simply to a phylogenetic pattern and nothing is implied about its mechanism or significance.

If cases of chloroplast capture are removed from the analysis, the supposed ease of cytoplasmic gene flow is not so strongly supported. In fact, in the few studies where cytoplasmic markers and nuclear markers have been used in conjunction, there are few, if any, convincing examples of increased cytoplasmic gene flow in plants.

1.5.5.1 Differential Cytoplasmic And Nuclear Gene Flow

One example that has been widely interpreted as showing an excess of cytoplasmic over nuclear introgression is in the North American white oaks. Chloroplast markers showed that variation in five species was between geographical regions rather than between species (Whittemore and Schaal 1991). The same pattern of a shared chloroplast pool is found in European white oaks (Dumolin-Lapegue *et al.* 1997). Thus, it is possible to rule out convergent evolution and lineage sorting. However, the data suggesting that (neutral) *nuclear* gene flow is low is not convincing. As mentioned, differences between oak species may be due to a few differences in major genes. Thus high levels of hybridisation could lead to high levels of neutral nuclear gene flow without the production of many morphological hybrids (Stebbins 1950). Therefore, reinterpretation of this data does not strongly support the assertion that "nuclear genes may be exchanged less freely between species" (Rieseberg and Soltis 1991).

If cases of 'chloroplast capture'(see below) are removed, the level of nuclear introgression is higher than cytoplasmic introgression in plant taxa. Certainly, between fully interfertile taxa, such as populations, the amounts of cytoplasmic and nuclear gene flow are in line with expectations based on pollen and seed flow (Ennos 1994). In the interfertile new world cottons, *Gossypium barbadense* and *G.hirsutum*, species diagnostic chloroplast markers found fewer cases of introgression than did the nuclear markers (Brubaker, Koontz and Wendel 1993). Similarly, most studies of hybrid zones find *individuals* with introgressive chloroplasts and parental nuclear markers (i.e. backcross individuals) but, at the population level, chloroplasts are strongly associated with conspecific nuclear markers (Cruzan and Arnold 1993).

1.5.5.2 Chloroplast Capture

Sometimes, it appears that an introgressing cytoplasm entirely, or nearly entirely, replaces the native cytoplasm in some populations. *Helianthus annus* and *H.petiolaris* which are recent introductions into Southern California demonstrate this. Chloroplast and rDNA markers which distinguished the two species throughout the rest of their range were found. In Southern California, it was found that over 99% of the *H.petiolaris* individuals contained the typical rDNA type (Dorado, Rieseberg and Arias 1992). However, almost the same proportion contained the *H.annus* chloroplast, the remainder having the haplotype of *H.petiolaris*. Similarly, a study on the *Heucheura* group of *Saxifragaceae* found a species (*Tellima*) that had one population that contained only the chloroplast of another species, but ITS sequencing found no evidence of the other's nuclear genome (Soltis and Kuzoff 1995). From these and

many other examples, it is clear that a chloroplast from one species can become associated with the nuclear genome from another (not just within one individual, but within an entire population or species).

That chloroplast capture requires a special explanation depends on the inference that chloroplast phylogenies more frequently show a discrepancy with the true (usually unknown) species phylogeny than phylogenies based on single nuclear genes. Any individual case of chloroplast capture could simply represent fixation of the 'alien' cytoplasm through drift alone, following its introduction into a population. Inferring a lack of 'nuclear marker capture' from the examples of 'chloroplast capture' is a simple ascertainment bias- drift alone would be expected to generate such patterns. Only if it can be demonstrated that nuclear markers show a significantly lower number of such events is there any case for chloroplast capture occurring through any selective explanation. Unfortunately, the data is inconclusive in this regard. Until more single copy nuclear genes are used routinely as markers, it will not be possible to say for certain whether chloroplast phylogenies really are more prone to distortion from hybridisation than nuclear genes.

Even if this was demonstrated, an explanation based purely on drift has been suggested (Ennos 1998). Under this extension of the null hypothesis of drift, the lower effective population size for organelles (Birky, Marayuma and Fuerst 1983) would lead to more rapid fixation of organellar genes than nuclear genes. Thus, more chloroplast capture events would be diagnosed as they are more likely to have occurred within the limited time-span since initial hybridisation. However, it would not be correct to conclude from this that cytoplasmic genes cross species barriers more easily than nuclear genes. Thus, under the hypothesis of neutrality and drift, the perceived ease of cytoplasmic transfer is an ascertainment bias.

1.5.5.3 Selective Mechanisms Of Cytoplasmic Introgression Without Nuclear Introgression

Though I have argued that there is not strong evidence in plants that cytoplasmic genes cross species barriers more easily than nuclear genes, there are plausible mechanisms for how it might occur. Population studies of nuclear and mitochondrial genes in animals have also suggested that cytoplasmic genes may be able to introgress (Marchant 1988) from one species to another more easily. Furthermore, in conifers which have paternally inherited chloroplasts, no examples of chloroplast capture have been documented, albeit in a much smaller sample set. If drift alone was the explanation, one might expect at least as many examples (as the initial influx of alien cytotypes might be expected to be higher for pollen inherited markers).

It has been suggested that such patterns reflect 'pollen swamping' (Ennos 1998). What is meant by this is not clear. In any case of cytoplasmic introgression, individuals with the 'invading' cytoplasm have arisen by repeated backcrossing to the native as paternal parent, and this could be reasonably termed pollen swamping. However, this term has little explanatory power as it is clear that the issue is not that repeated backcrossing took place, but how it took place at the population or species level.

Higher selective effects on the nuclear genome than on the cytoplasmic genome may account for a reduction in nuclear gene flow (Powell 1983). Strong heterozygote disadvantage at multiple loci could reduce introgression of even neutral alleles (Barton and Hewitt 1989) due to genetic linkage and hitchhiking. Similarly, coadaptation amongst nuclear genes could lead to strong concerted selection throughout the genome with the same effects (Barton and Hewitt 1989). Under this hypothesis, 'chloroplast capture' is still a chance event accounted for by drift alone, but nuclear genes introgress less frequently than they otherwise would- hybrid zones may act as 'semi-permeable membranes' (Harrison 1990) through which organellar markers diffuse more rapidly than nuclear markers. Hybridisation events could be relatively common, but if the nuclear genome of each taxa is buffered against change only the organellar genomes retain evidence of the genetic contact.

If the neutrality of cytoplasmic variation is not assumed, there are several other selective explanations that could give rise to chloroplast capture. One of these is superiority of the introgressing cytoplasm. This could give rise to a complete replacement of the native cytoplasm, even with only a small selective advantage (Frank 1989). Alternatively, interaction between cytoplasmic and nuclear genomes may lead to increased fitness for individuals with a mismatched cytoplasm (Wendel, Stewart and Rettig 1991). Patterns of variation in hybrid zones, and analyses of inter-population diversity, do not suggest that non-neutrality is common. However, it is known that organellar variants are not always neutral (Hanson 1991), and tests of these hypotheses would require detailed fitness analyses.

An important aspect of these explanations is that the non-recombinational nature of cytoplasmic genomes means that neutral variation is also brought to fixation in a 'selective sweep'. Thus, the spread of a selectively advantageous variant will be exactly reflected in the pattern of neutral polymorphisms. Selectively advantageous *nuclear* genes have a much lower effect on neutral variation (only causing a noticeable rise in frequency of very closely linked genes). Even if selectively advantageous variants introgress more commonly than in the

cytoplasmic genome, analysis of neutral genetic markers would not be expected to strongly reflect this.

A common explanation for how an alien organelle can spread throughout the population without its nuclear genes invokes hybrid male sterility. This has generally been assumed to be through cytoplasmic male sterility (Dorado *et al.* 1991), but hybrid male sterility in the first few generations of backcrosses will also achieve the same result (Gyllensten and Wilson 1987). The cytoplasm from an immigrant female (in a dioecious species) or an immigrant hybrid (in a monoecious species) can displace the native under these conditions, without significant nuclear gene introgression. What is not often stated is that this mechanism also requires selection. In the case of a selectively advantageous organelle, hybrid male sterility will simply speed up its introgression and reduce nuclear introgression (Frank 1989). Hybrid fitness and hybrid sterility have the same effect. This has been demonstrated experimentally in *Drosophila* (Aubert and Solignac 1990), where the introduction of a single alien female caused replacement of the entire population's native cytoplasm. This resulted from the hybrids having extremely high selective advantages (between 5 and 10 in the first generations). Thus while male sterility can enhance the spread of an organelle relative to the nuclear genes, a selective advantage either to hybrids or the organelle itself must exist. However, this explanation is plausible as a general mechanism given that reduced male sterility is a common outcome of hybridisation (Frank 1989).

1.5.5.4 'Nuclear' Capture

All these mechanisms suggest some degree of seed contact without pollen flow. For example, for the sunflowers it was postulated that "a single *H.annus* propagule was introduced into the founder population of *H.petiolaris*" and that the rate of cytoplasmic introgression was "enhanced by the maternal inheritance and vegetative segregation of organellar genes" (Dorado, Rieseberg and Arias 1992). This is unlikely to represent a common situation.

An alternative explanation is that asymmetrical pollen flow could occur from one population to another, for ecological reasons such as pollinator behaviour, prevailing wind, or simply relative population sizes. At the same time, seed flow must be restricted. Under this hypothesis, the maternal population, over time, builds up the nuclear genes of the paternal population without its cytoplasmic genes changing. This would mean that chloroplast capture, as defined by its phylogenetic effect, cannot be used to infer that inter-specific cytoplasmic introgression is easier than nuclear introgression. On the contrary, it could remain true that introgression is more likely to occur for nuclear genes if two taxa are brought into genetic

contact. Furthermore, subsequent full genetic contact between the two populations could lead to the introduction of 'alien' cytoplasm with no 'alien' nuclear genes, followed by possible fixation by drift. This mechanism is qualitatively different from simple drift, as it explains the supposed lack of nuclear introgression by an initial period of asymmetric pollen-only contact between two populations. It has the same phylogenetic implications as other mechanisms, in that it would make discordance between molecular and morphological phylogenies greater for organellar genes.

1.5.5.5 Cytoplasmic Introgression: Conclusions

If neutral cytoplasmic markers are more prone to cross-species transfer than nuclear genes, it has phylogenetic implications. For example, if nuclear genes are prevented from introgressing from one species to another they are likely to provide phylogenies more representative of the whole genome than cytoplasmic markers. Conversely, they are less likely to detect past hybridisation events. It has even been argued that the relationship of organelles between taxa may suggest how frequently these taxa have hybridised in the past and how 'crossable' they might currently be- as homogenisation of the organellar gene pool could be uncoupled from homogenisation of the nuclear gene pool (Van Raamsdonk 1995).

Thus, there are reasons for testing the hypothesis of substantially different behaviour for cytoplasmic and nuclear markers after hybridisation. However, the data- biased as it is towards organellar gene sequences- does not allow conclusive testing of this and must be regarded as tentative. Furthermore, the tendency to use 'chloroplast capture' as an explanation for phylogenetic discordance must be avoided. Chloroplast capture is the name given to a subset of such discordances, and does not offer an explanation for how they arose.

1.5.6 Adaptive Consequences Of Introgression

The evidence is clear that both neutral cytoplasmic and nuclear genes have crossed naturally from one species into another. The pattern of variation present within and between taxa often suggests such transfer, and studies of process in extant contact zones demonstrate the ease with which it may occur. Thus, it should also be possible for adaptive (or maladaptive) genes to cross species barriers, with several different possible outcomes.

1.5.6.1 Transfer Of Adaptive Traits

One of the most commonly suggested outcome is that adaptively advantageous traits that have arisen in one taxon could be transferred into a taxa without these traits after a hybridisation event. Such traits should be even more likely to cross species barriers than neutral genes

(Barton and Hewitt 1989). However, there is surprisingly little strong evidence, given that hybridisation itself is fairly widespread, that this plays a role in the evolution of plant taxa. This may be partly due to the experimental difficulties with inferring introgression from morphological traits (Rieseberg and Wendel 1993). Similar selective regimes may well produce the same morphological response, particularly in closely related species (Mayr 1942), and this convergent evolution is often indistinguishable from introgressive origin of adaptive traits (Klier *et al.* 1991). To test this would require the molecular characterisation of the genes responsible for putative introgressed selective traits.

It may be that the apparent rarity of morphological introgression, outside hybrid zones, reflects the rarity of traits and genes that have the same effects in two genetic and ecological backgrounds. The more distantly related two taxa are, the less likely it is that a gene will function advantageously in both. In this case, hybridisation as a mechanism for transferring adaptive traits between species may well have been overrated. The most likely exception would be with disease resistance genes, as these can function in a very simple, almost modular, fashion (Hughes 1996). It might be expected that it is just this sort of adaptive trait that would most easily spread between species. However, this remains untested, and the view held by some botanists that 'genetic assimilation' allows selectively advantageous traits to be filtered out of one taxa (Harlan and De Wet 1963) must be regarded as untested.

1.5.6.2 Creation Of New Adaptive Traits

It is perhaps more likely that introgression could result in new traits that are novel with respect to either parent. Through creation of novel gene combinations, or even alleles (Woodruff 1989), introgressed individuals may be able to occupy new niches. There is much evidence for this. It has long been suggested that hybridisation and introgression are of particular importance in disturbed environments such as farmland or volcanic areas (Anderson 1948), which would fit with this suggestion. The Louisiana irises provide at least one case where the introgressed population outcompetes either parent in a new environment. Similarly, there are at least two stabilised introgressant species of sunflower (Rieseberg 1991) that have more extreme ecological niches than either parent.

The analogy with mutation is useful here. It is not adaptive, in that any selective forces would act to prevent it, but it could increase the rate of adaptive evolution under certain conditions (Lewontin and Birch 1966). However, focusing on neutral variation has led to a neglect of the molecular basis of selective variation, and the role of introgression in the formation of new adaptations, even if introgression is known to have occurred, is unproved.

1.5.6.3 The Effect Of Introgression On Mating Barriers

Introgressive hybridisation might be expected to reduce the extent of mating barriers between two taxa. Either the mating systems themselves might become more similar (pre-zygotic mating barriers reduced) or the genetic incompatibilities (post zygotic mating barriers) lost. However, there is little evidence that this has occurred -though how it could be tested for is an open question. In fact, the examples used to support a breakdown in mating barriers demonstrate simply that the establishment of hybrid zones offer the opportunity for gene flow between species, without demonstrating such gene flow. For example, the existence of novel chromosome arrangements in a hybrid zone of *Clarkia* facilitates the transfer of genetic material between the chromosomally differentiated parents (Hauber and Bloom 1983), with the hybrid individuals providing a 'stepping stone'. However, it is not clear that this is an example of the mating barriers between two species breaking down. Unless this situation results in an overall shift in the inter-fertility of each species, it represents yet another example where the interesting dynamics of a localised contact zone are assumed to have greater implications for the evolution of the parents than is warranted.

It is also possible for hybridisation to reinforce mating barriers (Howard 1993). While not strictly linked to introgression, it is possible that direct selection against individuals carrying slightly deleterious introgressant genes could lead to indirect selection for individuals incapable of hybridising. There is evidence for this in soldier beetles (Howard 1993). However, no convincing evidence for it exists in plants.

Despite the logic of these two possibilities, it remains the case that the frequency with which hybridisation has occurred is not reflected by any detectable impact it has had on mating barriers in plants.

1.5.6.4 Genetic Assimilation, Extinction By Introgression And Dispersal By Hybridisation

Amongst the more dramatic claims made for the evolutionary impact of introgressive hybridisation are that some species are 'compilospecies' (Rieseberg and Wendel 1993) which expand their geographical and ecological ranges by 'plundering' locally adapted gene pools. Though it hasn't been demonstrated, it is certainly possible that adaptations could pass from locally adapted species to more generalised relatives as has been proposed for *Bothriochloa intermedia*, a widespread American grass (Harlan and De Wet 1963). Implying that some species may be particularly adept at this is, or that it is an adaptation itself is, however, without justification.

That some species are in danger of extinction from introgression has been suggested for many animal species (Echelle and Conner 1989), and several plant species (Levin *et al.* 1990). This is most likely after the introduction and spread of an alien species (Abbott 1992). It is not always clear, however, what exactly is meant. In the case of the now extinct British *Spartina stricta*, hybridisation with the invading *S.alterniflora* led to a polyploid that may have outcompeted and eventually replaced the native (Raybould *et al.* 1991). In such cases the threat comes not from hybridisation as such, but by competition from a hybrid variety (which may only have arisen once). In contrast, the rare island endemic *Argyranthemum cornopifolium* is under threat from the invasive weed *A.frutescens* with which it is highly fertile (Brochmann 1984) and forms hybrid swarms. However, caution is needed to eliminate the possibility that the species is under threat simply from competition, with hybrid swarm formation a coincidental by product. In fact, in such cases, hybridisation may (through chance alone) be the only way the endemic's gene pool is not completely eliminated.

An example where genetic differences were expected to be lost as a result of hybridisation is *Silene dioica* and *S.latifolia* (Goulson and Jerrim 1997). These two species are morphologically and ecologically distinct, but they cannot be differentiated on the basis of any neutral markers- suggesting a high degree of neutral gene flow between them. Furthermore, they are interfertile and produce identifiable, viable and fertile hybrids (Clapham *et al.* 1987). However, despite worries about the threat to one of the species 50 years ago (Baker 1947) from introgression, no change has occurred in its distribution. This shows that even when inter-specific mating is common, the homogenisation of the gene pools is by no means inevitable.

Finally, it has been suggested for both Oak (Bacilieri *et al.* 1996b) and Eucalyptus (Potts and Reid 1988) that hybridisation is a dispersal mechanism that facilitates the spread of a successor species into the range of a pioneer species. This treatment would depend on the succession requiring genetic contact between the two rather than genetic contact being a contingency of dispersal, which is doubtful. A treatment of hybridisation as an adaptive mechanism or a necessary part of a species life history is here, as elsewhere, suspect.

1.5.7 Patterns Of Introgression: Conclusions

There is little doubt that neutral genes, at least, can cross species barriers. Thus, at the very least, it is a useful indicator that there has been, or is, reproductive contact between two species. Beyond this, introgressive hybridisation can have two major consequences. One of these is phylogenetic. This is particularly the case with molecular markers, where neutral

markers suggest different phylogenies from morphology, ecology or adaptive traits. Furthermore, it is possible (but unproven) that organelle genes are more likely to introgress. It should be pointed out that the significance of this effect is related to the importance given to phylogenetic relationships that are based on neutral markers- it could still be argued that introgression has had little effect on the true relationship of the species.

Secondly, introgression could have had consequences for the adaptive evolution of the taxa involved. This is still a very open subject. However, convincing examples are few and far between. One must not conflate the fact that introgression has occurred with the idea that it is important. The real question is whether a taxon would have a different distribution, ecology, or morphology if it was not able to hybridise with other taxa. There is little direct evidence to suggest this is the case except where it has led to speciation. There are several examples of purported stabilised introgressants in both *Gilia* (Grant 1981), *Helianthus* (Rieseberg *et al.* 1991a) amongst others. There are good reasons for believing introgression to have played a role in the formation of these, but it is never possible to rule out that introgression occurred after speciation. Even if it could be shown that introgression played a causative role in the speciation of these new taxa, this must not be confused with the hypothesis that introgression is likely to have an impact on hybridising parental taxa. Introgression may have played a role in some speciation events, but there is little evidence that it is important in the normal adaptive evolution of plant taxa.

1.6 Patterns Of Reticulate Evolution(2): Homoploid Hybrid Speciation

Systematic studies have also uncovered a number of examples of species of homoploid hybrid origin (Table 1-3), whereby the hybrid species has the same chromosome number as each of its parents. However, the majority of these are suggested by limited molecular data (often one locus), and may represent the occasional fixation of an introgressive neutral marker subsequent to speciation. This highlights the biggest difficulty with this approach. It is often impossible to distinguish between species that have arisen because of hybridisation, and species that happen to contain a few markers derived from other taxa.

Table 1-3 shows some species of putative hybrid origin (by no means all of them), regardless of the relative contribution of each parent. All studies, implicitly or explicitly, have used morphology to infer hybrid origin, so only the molecular markers have been included. It does not include populations which contain introgessive markers but are not, at least partially, reproductively isolated from the parental taxa (either ecologically or genetically). This distinction is important as it is possible for hybrid speciation to occur in the absence of

introgression, and *vice-versa*. Allopolyploidy, for example, provides a mechanism for the production of new species without implying gene flow between the parental species.

Species for which the evidence is strong that homoploid hybridisation played a causative role in their genesis are marked by an asterix. Though they are somewhat limited in number, much discussion has centred on the mechanism by which such homoploid hybrid speciation might occur.

Table 1-3: Species of homoploid hybrid origin

Taxon	Consequences Of Hybridisation	Molecular Evidence For Hybrid Orgin	Reference
Helianthus anomalus, deserticola, and paradoxus	Speciation*	Many molecular markers	(Rieseberg 1991)
Iris nelsonii	Speciation*	Many molecular markers	(Arnold 1993b)
Pinus densata	Speciation*	Many molecular markers	(Wang *et al.* 1990)
Stephanomeria diagensis	Speciation*	Many molecular markers	(Gallez and Gottlieb 1982)
Peaonia	Speciation	ITS sequences	(Sang, Crawford and Stuessy 1995)
Gossypium bickii	Introgression of chloroplast genes	Chloroplast genome and rDNA genes	(Wendel, Stewart and Rettig 1991)
Helianthus debilis, neglectus	Introgression of chloroplast genes	Chloroplast genome and rDNA genes	(Rieseberg 1991)
Heuchera hallii, micrantha, nivalis, parviflora	Introgression of chloroplast genes	Chloroplast genome and ITS sequences	(Soltis and Kuzoff 1995)
S. cariophyllum subsp. ehrenbergi	Introgression of chloroplast genes	Chloroplast	(Rodriguez and Spooner 1997)
Populus nigra	Introgression of chloroplast genes	Chloroplast	(Smith and Sytsma 1990)
Encelia virginensis	Speciation?	RAPD markers	(Allan *et al.* 1997)

1.6.1 Mechanisms Of Homoploid Hybrid Speciation

It has been suggested that such speciation is problematic and mechanistically difficult (Rieseberg 1997), as it may require the evolution of reproductive isolation between the emerging hybrid species and its parents. For homoploid speciation, one mechanism by which reproductive isolation might arise rapidly is genomic reorganisation (Templeton 1981). This has been demonstrated in *H.anomalus* (Rieseberg, Van and Desrochers 1995), where the hybrids' chromosomes contain inversions and translocations, which presumably contributes to reproductive isolation from its parents. However, the conditions necessary for such

reorganisations to become fixed also lower the fertility of hybrids in their early evolution and thus lower the chance that such hybrids will result in long-lived lineages.

In actuality, it is only a small subset of possible hybrid speciation events that *logically* require such explanations- only those where the hybrid species has arisen in sympatry with both parents and not displaced either. An alternative situation would be the contact of two heterospecific populations that are geographically isolated from their respective conspecific populations. The hybrid population so produced would be somewhat isolated from its parents and could possibly evolve into a distinct species. This is equivalent to the view that speciation often occurs at the population level (Jonseil 1984). Similarly, hybrid speciation could occur without the evolution of further reproductive isolation if the hybrid outcompetes either, or both, of its parents in some niches. For example, the evolution of *Pinus densata* (Wang and Szmidt 1990; Wang *et al.* 1990; Wang and Szmidt 1994) seems not to have required anything other than the same kinds of selective and ecological factors that keep its two parents distinct.

However, as there are so few homoploid hybrid taxa, these situations must be somewhat uncommon. Such taxa should not be seen as presenting difficulties for speciation theories, but as the rare examples of hybrids that have, through a variety of possible mechanisms (including chance) avoided extinction.

1.6.2 Homoploid Hybrid Speciation: Conclusions

Only situations that lead to assortative mating of hybrids- with the parents contributing less to the production of the next generation- can result in hybrid speciation. Allopolyploidy is an extreme example of this. Chromosomal reorganisation is another, less extreme, mechanism that may enhance genetic isolation from parents. Ecological and geographic factors could also present hybrid populations with the opportunity to evolve without 'dilution' from the parents. In the same way that introgression is sometimes suggested to assist the expansion of a species' range, hybrid speciation is thought to cause the colonisation of novel or extreme niches.

However, the paucity of homoploid hybrid species does not justify the view that many "novel evolutionary lineages, adapted to a variety of habitats, have arisen through natural hybridisation" (Arnold 1997) which is still speculative when applied to homoploid hybrids.

1.7 What Has Been Learnt About The Consequences Of Hybridisation?

We are now in a position to attempt an answer to the questions posed at the outset of this discussion. The two disparate lines of investigation- process and pattern- have lead to the same conclusion as regards introgression of neutral markers. Such markers can, and do, cross

species barriers. The implications of this depend purely on how species are defined. The phylogenetic consequences of hybridisation is to distort relationships based on molecular markers, and there is no way to simplify this. However, it may be seen as less important if molecular markers are seen as an indication of relationships between taxa, rather than defining those taxa. Of course, a greater genetic distance as demonstrated by molecular markers makes genetic incompatibility, which must be part of any species definition, more likely- but the genetic distance itself is not what defines the species. The confusion generated by hybridisation is a result of the uneasy interface between biologically useful classifications and well-understood population genetic processes.

There is sparse evidence about the role of introgression in adaptive evolution. Studies of pattern have not attempted to identify the genes responsible for selective differences between taxa, and as such are uninformative in this regard. One exception to this is the Louisiana irises, where it appeared that early generation hybrids were unfit but the backcrosses were competitive in certain ecological niches. Similar approaches in other species should be able to detect whether this adaptive introgression is common or not. It is noteworthy that in this case, the putative introgression of adaptive genes is considered to be the first step in a hybrid speciation episode.

It is rare for hybridisation to lead to homoploid speciation. Very little is understood about this process, as studies of the dynamics of natural hybridisation have rarely addressed this issue (with the aforementioned exception). Certainly, there is no evidence that it is an important part of evolutionary history in that the plant kingdom may not be significantly different had inter-specific hybridisation not been possible.

Though I have suggested that the usual outcome of homoploid hybridisation is the production of hybrid populations with little evolutionary importance, this does not mean it will never be a significant event. Thus, it is important to have a consistent framework for viewing hybridisation. I suggest that introgression, rather than hybridisation itself, be treated as the central concept. It appears that hybridisation is more important as a source of variation for existing species than the creation of new ones, and it is the effect of this variation that needs to be analysed. In this way, introgression is analogous to mutation in population genetics. It is impossible to fully answer the question "What is the role of hybridisation in speciation?" without an answer to the (largely pointless) analogue "What is the role of mutation in speciation?"

1.8 Further Work

Throughout this discussion I have emphasised the limitations of past approaches. One consistent criticism has been the complete reliance on neutral markers in *systematic* studies (neutral markers are far more useful in population genetic studies of the processes of hybridisation). If adaptive traits are regarded as the relevant indicators of relationships then molecular phylogeneticists may be greatly overrating the tendency of hybridisation to affect systematic relationships. Generally, the adaptive traits studied have been morphological, the behaviour of which in hybrids is rarely simple. However, there is no reason why adaptive molecular traits cannot be analysed. I suggest that the use of neutral markers, while essential in population genetics and short term evolutionary studies, should be replaced by such selective genes for systematic studies. The corollary of this is that much more work needs to be done on the molecular basis of morphological and ecological differences before the importance of hybridisation in evolution can be properly assessed. It has been suggested that incongruence between molecular and morphological phylogenies can be due to the differences being due to a relatively small number of genes having large morphological effects (Kadereit 1994). If true, it might be supposed that the patterns of hybridisation and introgression revealed by such genes will be qualitatively different from those shown by neutral genes.

Another criticism is directed at studies of process in hybrid zones, which have often failed to make use of the additional information that makes such an approach potentially so useful. This may stem from a view that nature has already provided the relevant study materials under appropriate conditions. However, extending this approach to include the temporal aspects, as will be done in *Larix*, should eventually allow taxonomic studies to take into account homoploid hybridisation, to determine what impact it might have on ecological and conservation issues, and to determine if it is justified to view it as an important evolutionary process.

2. The Genus Larix and the Dunkeld Hybrid

"When the North wind forays
among the larches-
greener than all the greens assembled here
I hear the sound of water boiling
but really it's only the wind
stirring surfaces as the sun travels North"
A.Martin in "The Larch Plantation"

2.1 Introduction

European and Japanese larch (*Larix decidua and L.leptolepis*) are inter-fertile species that have very recently been introduced into Scotland. The consequences of hybridisation between them was studied on the estate in Scotland where hybrids were first discovered. Here, I shall introduce both the larch, its history in Scotland, and the study sites.

Figure 2-1: The larch (adapted from *Python,M* 1974)

2.2 The Genus Larix

The larches are deciduous, needle-leafed gymnosperms in the pine family (Figure 2-1). They are unusual in sharing many structural traits with other conifers (such as canopy architecture, leaf morphology, and water conductance anatomy), while having the deciduous habit (Gower and Richards 1990). They are distributed throughout the colder regions of the Northern hemisphere. There are generally considered to be ten species (Ostenfeld and Larsen 1930), three of which are endemic to North America, and the remaining seven to Asia and Europe (Table 2-1).

Larix occupies a wide range of ecological habitats (Gower and Richards 1990), but is most notable for its pioneering tendency and often forms the dominant species in the treeline of mixed-conifer forests (LePage and Basinger 1992). It is well adapted to colder climes (Dallimore and Jackson 1948), and is able to thrive on poor soils. Its deciduous habit and efficient nutrient use also contribute to its pioneering qualities (Gower and Richards 1990). It is, however, highly shade intolerant and does not do well in competition with other species in densely packed forests (Dallimore and Jackson 1948).

Many of the larch species are valued in forestry. They are light tolerant, have fast juvenile growth, and high-quality timber (Paques 1989). Furthermore, their deciduous habit makes them particularly attractive in autumn and winter, especially when used to break up the

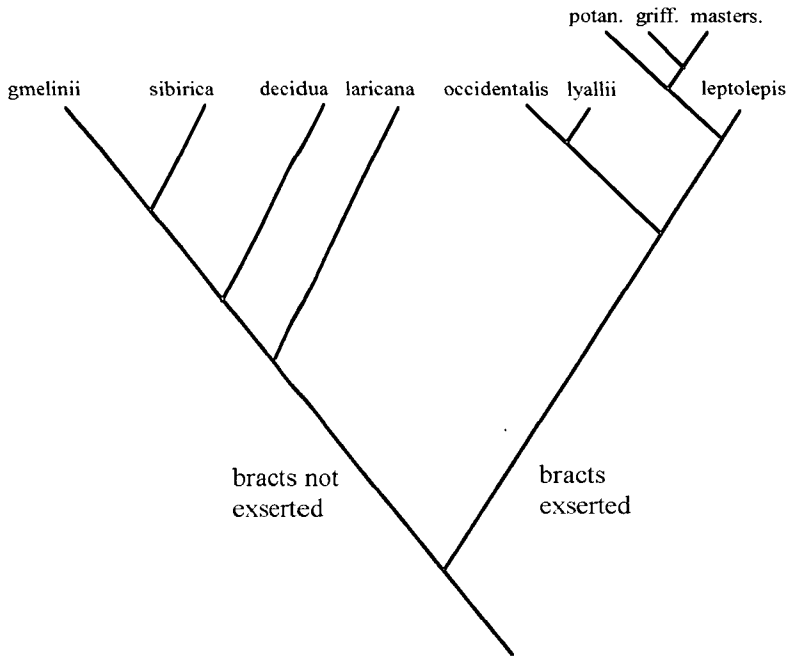
monotony of evergreen plantations, and are valued for aesthetic as well as commercial reasons (Blocker 1992; Forestry Commission 1988).

The proposed phylogeny of the genus is shown in Figure 2-2. This is based largely on the anatomy of the female cone (Qian *et al.* 1995), the meagre fossil record (LePage and Basinger 1992), and current geographic distribution. In particular, note the splitting into two major lineages based on whether the bracts are visible beyond the cone scales (exserted) or not. This is confusing, as the bracts of *L.leptolepis* are not exserted and other phylogenies (Vidakovic 1991) have placed it in the same group as *L.decidua* (Section *Multiseriales* Patschke). It appears that geographical considerations have given rise to this placement, and it may well be incorrect.

Table 2-1: Larix species and their distribution (adapted from Schmidt 1992)

Latin name	Common name	Distribution	Habitat
L.occidentalis	Western larch	Rockies and Cascades (U.S., Canada)	Upper montane to lower subalpine
L.lyalli	Alpine larch	Rockies and Cascades	Upper subalpine to timberline
L.laricana	Tamarack	NE states, lake states, Alaska (U.S), Canada	Mainly boreal
L.russica	Siberian larch	N Russia, Mongolia	Boreal to timberline
L.gmelinii	Asian larch	Eurasia (E of Siberian larch)	Subalpine to timberline
L.mastersiana	Masters larch	south China	Upper montane to lower subalpine
L.griffithiana	Sikkim larch	Himalayas in Nepal, Bhuten, Tibet, S.China	High subalpine
L. potaninii	Chinese larch	Western China	High subalpine
L. leptolepis	Japanese larch	Honshu, Japan	Subalpine
L. decidua	European larch	Alps throughout Europe	Subalpine

Figure 2-2: A Phylogeny for *Larix* (adapted from LePage and Basinger 1992)



*2.2.1 European Larch (*L.decidua *Mill.)*

European larch's (Table 2-2, Figure 2-3) natural range is the subalpine zone of the Alps area in France, Switzerland, Austria, Italy, Yugoslavia, Germany, with scattered patches in Poland, Romania, and Czechoslovakia (Schmidt 1992). Four varieties are recognised, Alpine, Polish, Sudetic and Carpathian (Holtmeier 1992).

Since the 18th century, European larch has been widely planted outside it's natural range. A serious problem in Europe has been it's susceptibility to diseases such as larch canker (*Lachnellula willkommii*), which has contributed to a decline in it's usage over the last century (Yde-Andersen 1979). However, it is considered a promising exotic in parts of the U.S. (Schabel *et al.* 1992), often outperforming the native tamarack as well as spruce, pine and fir. It is now part of a breeding programme (along with Japanese and hybrid larch) responding to the increase in demand for softwood in the U.S. (Li and Wyckoff 1994).

*2.2.2 Japanese Larch (*L.leptolepis *Gord.)*

The natural range of Japanese larch (Table 2-2, Figure 2-4) covers only about 120 square miles: the highland region of central Japan, between 35 to 37° latitude and 137 to 140° longitude (Takei 1992). It is often called *Larix kaempferi* Carr. Since the late 19th century it has been transplanted to Europe and America (Takei 1992). It is less frost tolerant than European larch (Schabel, Lee and Wyckoff 1992) and it's timber is of lower quality

(Dallimore and Jackson 1948). However, it outperforms European larch in milder climates (Li and Wyckoff 1994), is relatively resistant to larch canker (Yde-Andersen 1979), and is a valued ornamental tree (Dallimore and Jackson 1948).

2.2.3 *The Hybrid Larch (*L.leptolepis x decidua*)*

The hybrid larch (Table 2-2, Figure 2-5) was first noted in 1904 (Henry and Flood 1919), the result of pollination of Japanese trees by European trees at Dunkeld, Perthshire. It's commonly referred to as the Dunkeld hybrid. It's Latin name is usually given as *Larix x eurolepis* Henry, though there are doubts about the validity of this name (Nelson 1980). Possible hybrid vigour, and the combination of the frost tolerance of European larch with the Japanese parent's disease resistance has made the hybrid a valuable forest crop (Paques 1989), often outperforming either parent (Li and Wyckoff 1994; Schabel, Lee and Wyckoff 1992). The question of hybrid vigour has not been proven beyond all doubt, as much of the perceived effect may be due to the use of 'plus' (i.e. superior) European parents (Paques 1989). Seed orchards are used to produce the F1 hybrid (Ennos and Qian 1994; Matthews 1954), though micropropagation techniques are being developed to circumscribe their limitations (Benkrima and Aderkas 1992).

Table 2-2:Comparison of European, Japanese and hybrid larch

adapted from Laing (1947),Henry and Flood (1919), Paques (1989)

	L.decidua	*L.leptolepis*	*L.eurolepis*
1st year shoots	glabrous straw coloured	often pubescent red	slightly pubescent intermediate
Pulvini	ending at base of bud	overlapping bud	slightly overlapping
Cones	conical straight scale (but variable)	globose reflexed scales	conical slightly reflexed scales
Bracts	exserted	concealed	some exserted
Resin canals of pulvini	circular	oval	oval
Leaf of long shoot papillate cells endodermal cells position of bundle	thicker, raised midrib absent 13-16 central	broad, flat absent 20-22 asymmetrical	variable variable 17-19 variable
Bark of stem	thick, with visible white striations	thin, without striations	thin, sometimes with striations
Male flowering time	late March (1 wk)	late March to early April (approx. 2 wk.)	Variable
Female flowering	mid to late March	mid March	Variable
Canker resistance	low	high	high
Juvenile growth rate	medium	fast	fast
Frost tolerance	high	low	high

Note: Many other characters are given in the references. In the case of the morphological, particularly microscopic characters, only those which I was able to detect are included.

Figure 2-3: European larch at Kennel bank, Dunkeld

Figure 2-4: Japanese larch at The Avenue, Dunkeld

Figure 2-5: Hybrid larch (from outside the castle at Blair Atholl)

2.3 Larch in Scotland

Table 2-3 gives the results of the most recent Forestry Commission woodland census (1983) in Scotland. Japanese and Hybrid larch were counted as one class. This reflects the fact that much of the 'hybrid' larch planted in private forestry is of unknown character possibly referring to seed grown from Japanese mothers which may or may not be hybrid. Thus, the opportunity for the evolution of a new taxon is obvious. Two exotic species, and the hybrid between them, have been planted right across Scotland (and indeed, Britain) in close proximity, giving the opportunity for new genetic combinations in areas isolated from either parents native range.

Though it cannot be directly inferred from this table, as there may be differences in rotation, European larch far outnumbered Japanese or Hybrid larch until the 30's, but then lost favour- by 1980, it was outnumbered 10:1. However, since the late 70's, it's planting rate has exceeded that of hybrid and Japanese larch (Forestry Commission 1988).

European larch enjoyed a prominent place in Scottish silviculture for 200 years (Taylor 1964). It was used in boat building, as its timber is particularly durable. A 19th century frigate, 'The Atholl', was constructed entirely from Scottish larch and outlasted a similar vessel made of fir by 40 years (Forestry Commission 1988). Even as recently as the second world war, it was used in the construction of motor torpedo boats (Dallimore and Jackson 1948). However, by the 1960s, it was past it's heyday (Taylor 1964), as the use of modern preservatives allowed the timber of the faster rotation spruces and firs to be used just as effectively. It is now valued more for the aesthetic qualities of its woodlands, and can be used as a 'nurse' (Mason and Baldwin 1994) for species such as oak.

Table 2-3: Area of coniferous forest in Scotland (100's of hectares)

(adapted from Forestry Commission woodland census, 1983)

Year	SP	CP	LP	SS	NS	DF	EL	JL/HL	Other	TOTAL
Total	**1444**	**33**	**1039**	**3646**	**547**	**116**	**160**	**521**	**149**	**7656**
pre 1861	49	0	0	0	0	1	2	0	2	**55**
1861-1900	99	0	0	1	1	2	11	0	5	**120**
1901-1910	45	0	0	1	3	1	15	1	2	**68**
1911-1920	39	0	0	2	6	2	8	3	3	**63**
1921-1930	122	5	3	39	34	10	15	15	7	**251**
1931-1940	107	5	6	113	78	5	27	26	6	**374**
1941-1950	143	4	6	212	93	6	14	42	5	**525**
1951-1960	447	10	175	498	156	34	32	216	35	**1604**
1961-1970	290	8	397	1087	124	38	24	107	50	**2124**
1971-1980	103	1	451	1693	53	17	10	111	33	**2472**

'Year' refers to the amount of extant forest that was planted in that year

Key: SP=Scots pine, CP=Corsican pine, LP=Lodgepole pine, SS=Sitka spruce
NS=Norway spruce, DF=Douglas fir, EL=European larch, JL=Japanese larch,HL=Hybrid larch
HL and JL were counted as one class

2.3.1 The Atholl Estate

The estate of the late XIVth Duke of Atholl is the largest in Britain, extending from Dunkeld in Perthsire at its Southern point to beyond Blair Atholl at its Northern boundary. A large part of its income derives from forestry, with much of this consisting of European and hybrid larch (Panting 1997). In fact, the second, third and fourth Dukes planted 10,000 acres of larch, accounting for 2/3rds of their woodlands. This earned them the epithet 'The planting Dukes'.

Figure 2-6: Blair Atholl Castle

2.3.2 Larch on the Estate

At one time, it was thought that the first European larch trees in Britain were planted at Dunkeld in the 1700's. Though, it is now known that at least one tree was flourishing in England by 1662 (Mitchell 1963), the estate was the first to become connected with the silviculture of European larch. The famous Dunkeld larches, 5 trees (of which one survives) brought by Colonel Menzies from London, were planted (by Dunkeld Cathedral) in 1738 by the second Duke of Atholl (Mitchell 1963). They are not the progenitors of as much of the larch on the Atholl estate as once thought, but do represent the start of widespread interest in the forestry value of the tree. Of particular interest are the 15 fine trees on Kennel bank on the Dunkeld House estate, which were planted around 1750 (Edwards 1956) and may have derived from the parents by the Cathedral.

Little pure Japanese larch has been planted on the estate. However, there are 11 ornamental trees on 'The Avenue' just below Kennel bank, and one by Dunkeld House Hotel. They were planted in 1887 from seed imported from Japan by the seventh Duke of Atholl (Edwards 1956).

Hybrid larch represents the bulk of the larch planting on the estate (Panting 1997). Indeed, it was on the Dunkeld House estate that the hybrid larch was first discovered. Seedlings from the aforementioned Japanese trees were planted in a nursery around 1900 (Edwards 1956), and it was observed that some of these differed in colour and growth rate from the usual Japanese seedlings. These seedlings were subsequently planted around the estate (Henry and Flood 1919), since when seedlings taken from either the original Japanese trees or from subsequent hybrid plantations have been planted out and designated 'Hybrid larch' (Panting 1997).

2.4 Outline Of Field Project

As discussed in Chapter 1, a valuable approach to analysing the evolution of hybrids is to track the genetic consequences of bringing two previously isolated species into contact. This is what has occurred on the estate- the two species are distinctly allopatric and the plantations around Dunkeld and Blair Atholl represent the first time these two species have been brought into contact.

The initial phase of the project was to identify species specific markers (Chapter 3). These markers were then applied to a putative F_1 stand (Chapter 4). Originally, a subsequent F_2 stand and its regenerants were to be studied. However, based on the results of the putative F_1 stand, two new study sites were chosen instead. These were chosen because they had been

fenced off to allow regeneration, and genetic analysis of both these putative F_1s (Chapter 5) and their regenerants (Chapter 6) was carried out. The implications of these results, both with respect to the future of plantations of hybrid larch in Scotland and with respect to the current composition of larch on the estate are discussed (Chapter 7).

2.5 Phylogenetic Relationships in Larch

A knowledge of the exact phlyogeny of *Larix* is not crucial for this project. It is known that both *L. decidua* and *L.leptolepis* are exotic to Britain, and do not and have not made contact in their natural ranges. To this extent, it is justified to treat their planting in modern times in Britain as a novel opportunity for evolution, without dwelling on their exact relationship. However, it would support the general discussion of hybridisation could the taxonomic proximity of the two species be determined. A brief review of what is known about the phylogeny of these two species is presented. Furthermore, as part of the survey to find variation between the two species (chapter 3), sequencing of organellar regions was attempted. The results of this brief study are also presented here.

Morphologically, there are too few differences between the species to provide an accepted taxonomy. The most recent non-molecular taxonomy proposed, that shown earlier (LePage and Basinger 1992), is actually a mix of morphological characters and hypothetical evolutionary history, based on a rather poor fossil record. It is notable that the two species studied here are present in the two different main groups, indicating a large genetic divergence. This is to be expected based on the geographical distribution of the two species. However, the main split based on bract exsertion is dubious. In Scotland at least (based on personal observation as well as other phylogenies (Vidakovic 1991), the bracts of the Japanese larch are not exserted such that placement in this group according to this characteristic appears to be erroneous.

Also in support of an ancient split between the two species is a recent and extensive study of isozyme variation in the genus (Semerikov *et al.* 1998). *L.leptolepis* and *L.decidua* are put into highly separated groups on a PCA, and the suggested phylogeny was that the *L.leptolepis* split was the most ancient to occur in this group.

In contrast, the two species are morphologically very similar and hybridise extremely easily. A study of cpDNA restriction sites (Qian, Ennos and Helgason 1995) suggested a rather closer relationship between them than between many other larch species. This was interpreted as possible evidence of a chloroplast transfer event following hybridisation, or simply a close genetic relationship.

2.6 Sequencing Organellar Genes

As part of this project, in a survey looking for inter-specific variation, organellar regions from these two species, as well as chinese larch (*L.potanini)* and tamarack (*L.laricana*) were PCR amplified and sequenced. Though results were not strongly supported and sequences error-prone, there was some evidence of a greater similarity between the two study species than with either of the two other species.

2.6.1 Materials And Methods

Three Japanese trees from the Avenue and three European trees from Kennel bank were used as source material for this study. Samples of *L. laricana* and *L.potanini* came from seed from provenance trials that had been germinated and DNA extracted by X. Hu (Hu 1998).

DNA extraction was performed using the Scotlab Phytopure kit, details of which are given in Appendix A (p.177).

PCR primers used were Taberlet primer pairs (Gielly and Taberlet 1994) ab, cd, and ef. These 'Universal' primers amplify inter-genic spacers between chloroplast tRNA genes, and are commonly used in surveys of organellar variation. Sequences for Chinese larch for these regions were provided by X.Hu (Hu 1998).

For the mitochondria, the COX1 primer was used. This was designed by Sinclair (Sinclair *et al.* 1998) as part of a project to detect variation in Scots Pine, but was successfully applied in larch to amplify a product later used as a radioactive probe (described in Chapter 3).

PCR was carried out as described in Appendix A (p. 182). PCR products were run on an agarose gel and visualised using EtBr. Successfully amplified products were cut out from a 1% agarose gel with a clean scalpel, and the Qiagen Gel-extraction kit (Qiagen) used to remove the agarose and purify the product.

Fluorescent sequencing was performed using the Amersham fluorescent sequencing kit following the manufacturer's protocol. The original PCR primers were used as the sequencing primers in the sequencing reaction. Loading and visualisation of the sequencing product was performed on an ABI automated sequencer by a dedicated laboratory.

Analysis and manipulation of the resulting sequences was carried out using Genejockey, a sequence analysis programme for the Macintosh PC.

2.6.2 Results

Results from sequencing COX1 of the mtDNA were disappointing. Error rates, calculated by the computer programme Genejockey II and a measure of the accuracy of the sequence, were far too high to justify in depth analysis, which relies on a good alignment of the sequences. Two European and two Japanese individuals were each repeated twice, and between run variation appeared as high as between individual and between species variation. Thus, though there was no evidence of variation between these two species, there was also no evidence of variation between the Chinese larch and tamarack, and no conclusions regarding phylogenetic relationships can be drawn.

Sequencing of the products from the Taberlet primer pairs proved slightly more informative. Three individuals each of European larch and three of Japanese larch provided the consensus sequences shown in Appendix B, for the region between primers a and b, and the region between primers e and f. The region between primers c and d was not sequenced. There is no evidence of inter-specific variation between these two species. However, a comparison with the Chinese larch sequence generated by the same primers and also supported by 3 individuals (Hu 1998) revealed 1 well supported difference. This was out of about 700 bp where the error rate was suitably low (other possible differences occur in regions of high error-rate, but these could easily have been artefactual). Thus, it would appear that than Japanese and European larch are more closely related to each other than to Chinese larch.

Also, a good sequence for region ab was achieved for tamarack (*L. laricana*), showing 2 single base-pair differences from Chinese, Japanese and European larch. This is also in contradiction to the phylogeny placed above, further suggesting that European and Japanese larch may be more closely related than generally accepted.

2.6.3 Discussion

The danger of drawing phylogenetic conclusions based on a single region has already been discussed. Thus, however well supported the conserved 1bp difference (relative to chinese larch) in this 700bp region, this cannot be taken as strong evidence that the chloroplast of these two species is closely related. However, when the independent evidence based on restriction analysis (Qian, Ennos and Helgason 1995) is taken into account, there is a plausible argument that the Japanese chloroplast and European chloroplast possess a more recent common ancestor than with Chinese larch (in contrast to the phylogeny presented earlier). Other explanations- aside from small sample or experimental errors- of course exist.

Unequal rates of evolution- due to ecological or selective factors- could also explain the results. For example, *L.decidua* and in particular *L.leptolepis* exist in far smaller population sizes than, for example, *L.potanini*. This affects the rate of selective evolution. If there was a selective process operating on the chloroplast subsequent to their divergence, it might result in more rapid change in the large population. Such an increase in evolutionary rate need not be restricted to selective mutations, as the non-recombinational nature of the chloroplast allows 'hitch-hiking'. This argument is that 'selective-sweep' might be responsible for the average time until fixation of neutral genes being less in populations undergoing selection.

If the close relationship between the chloroplast sequences actually reflects evolutionary history, there are two explanations. The first is that it represents organismal history. This appears unlikely on the basis of geographical distribution and the isozyme study mentioned earlier. However, the close morphological relationships and ease of hybridisation mean that this cannot be formally ruled out.

The second such explanation is that it represents a chloroplast capture event subsequent to a relatively recent hybridisation event, resulting in a taxon with incongruent nuclear and organellar phylogenies. This may be the case, and the divergence at the isozyme loci certainly supports the idea that the two species are not closely related. However, as was clear from the introduction, comparisons between these two classes of markers are error-prone, and there are plenty of ways in which the observed pattern could come from a more recent Japanese-European split than the split from Chinese larch. For example, the Chinese split could have occurred first, allowing time for the two chloroplast lineages to diverge under mutational (and possibly selective) influences. The difference between the Japanese and European isozyme frequencies could be due to population genetic effects (the Japanese population, in particular, is very small and drift more effective) more recently.

2.6.4 Conclusion

In conclusion, the evidence from the chloroplast sequencing suggests that chloroplasts of European and Japanese larch may be more closely related to each other than to those of Chinese larch, though the evidence is not conclusive. It is suggested that though this may reflect a chloroplast capture, it is also possible that it accurately reflects organismal history. Under this interpretation, the isozyme taxonomy must be viewed cautiously, possibly reflecting recent drift which the less variable, slower mutating chloroplast might not demonstrate. In the light of this, it is particularly disappointing that mitochondrial sequencing did not help, as if this interpretation could be supported with this organelle marker as well, the

evidence would be very strong in support of this hypothesis. There is certainly much scope for molecular taxonomy to provide a reinterpretation of the evolutionary history of *Larix*.

2.7 Summary

The widespread planting of two exotic species of larch and their hybrid in recent years in Scotland provides the opportunity for the study of the genetic consequences of hybridisation in larch. This is in contrast to previous studies, where the exact timescale since genetic contact is rarely known, and provides fertile ground for the examination of some of the issues discussed in chapter 1. A particular site, consisting of the Atholl estate and Dunkeld in Scotland was, chosen as a historically interesting and valuable place to start such an investigation.

The phylogenetic relationships between the two parental species should also be resolved, as a theoretically unrelated but experimentally similar and interesting project.

3. Inter-specific markers

3.1 Introduction

In order to determine the genetic composition of larch plantations on the Atholl estate, it was necessary to develop species-specific markers. Morphological and molecular markers were both considered. The greater resolving power of molecular methods with respect to hybridisation has already been discussed. Furthermore, uniparentally inherited markers make it possible to determine the direction of the parental mating. In *Larix,* as with most conifers (Mogensen 1996), the chloroplast genome is paternally inherited (Szmidt *et al.* 1987) and the mitochondrial genome is maternally inherited (Devemo *et al.* 1993). Thus, molecular markers make it possible to study the effect of hybridisation on three differently inherited genomes and it is these that form the bulk of the study.

3.2 Aims

The goals of this initial study were as follows.

1) To find at least one marker capable of distinguishing the mitochondrial genomes of the two species.

2) To find at least one marker capable of distinguishing the chloroplast genomes of the two species.

3) To find at least one, but preferably multiple, markers capable of distinguishing the nuclear genomes of the two species.

4) To perform a feasibility study on morphological studies

The reason for the first two aims is that the two organellar genomes yield qualitatively different information regarding an individual tree's parentage. In most angiosperms, the maternal inheritance of both would make this less informative. In fact, this was one of the motivations behind choosing *Larix* for this study. The third aim is to discriminate as much as possible between F_1s, backcrosses and later generations. The more nuclear markers that could be found, the greater this discrimination. The fourth was to determine whether a morphological approach to the study sites would have been productive.

3.3 Choice Of Molecular Markers

There is a wide range of molecular markers available, with varying eases of use, cost, and patterns of variation. In a study of this sort, choosing markers is one of the most crucial stages of the project, as the wrong choice could result in much lower productivity.

Table 3-1 summarises the properties of the various molecular markers that were considered. These properties are based on considerations for this particular study and the materials,

equipment, and expertise available in the laboratory. Initial method development is an assessment of the time, cost and difficulty of obtaining variable markers. Cost and throughput refer to the cost and rate of processing samples once markers have been obtained.

Table 3-1: Choice of markers for investigating inter-specific variation

Method	Initial method development	Cost/ sample	Throughput	Other
Isozymes	+++	++	+++	pattern of expression may vary non-genetically
RAPDs	+++	+++	++	dominant markers, unreliable
AFLPs	++	++	++	generally dominant markers
microsatellites	+	++	+	technical difficulty
rDNA RFLPs	+++	++	++	complex inheritance
organellar RFLPs	++	++	++	
PCR fragment restriction analysis and SSCP	+(nuclear) ++(organellar)	+++	+	Few sequences known for coniferous nuclear genomes
Sequencing PCR fragments	++	+	+	expensive, slow, but highly informative

+++: Highly suitable ++: Suitable +: Prohibitive

3.4 Materials And Methods

The detailed protocols and recipes for the following techniques can be found in Appendix A. Primer sequences for PCR are shown in Appendix B.

3.4.1 Sample Materials

The majority of the hybrid larch on the Blair Atholl estate derives from the 11 Japanese trees on The Avenue (Figure 3-1 and Figure 2-4). These 100 year old trees were chosen to represent *Larix leptolepis* in the search for species-specific markers. The vast number of potential European parents on the estate made it impossible to screen all of them. However, the most likely parents of the F_1 generation are the 15 European trees above The Avenue (Figure 3-1 and Figure 2-3), and these trees of approximately 250 years old were chosen to represent *L.decidua.*

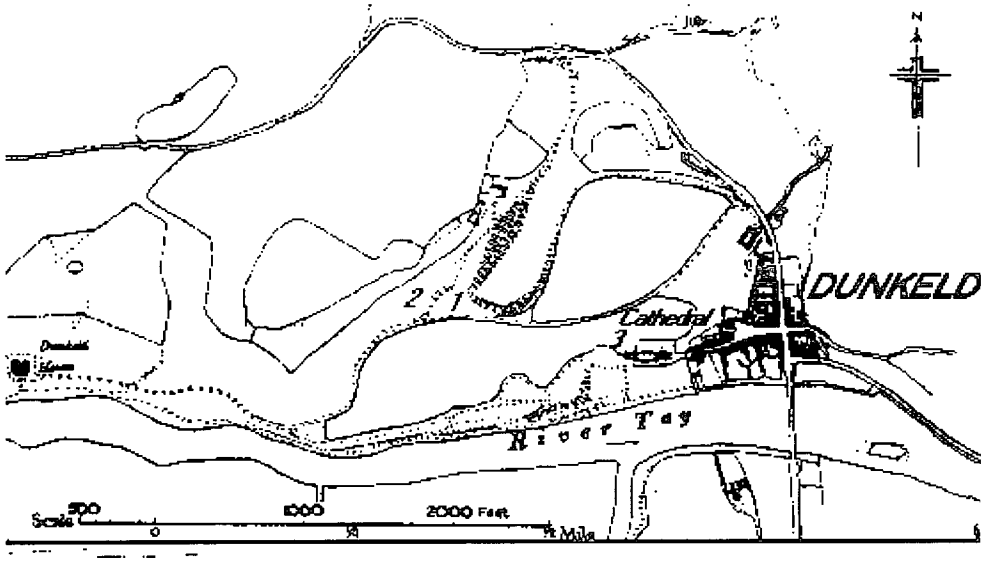
The age and height of the European trees made it problematic to collect directly from them. Fortunately, many of them are represented by clone banks kept by the Forestry Commissions Northern Research Stations at Newton, near Inverness and also Bush, near Edinburgh. This material was generously donated for use in this project.

A series of controlled crosses- with the European larch as maternal parent- had also been set up by the Forestry Commission in the forest of Dean and material had been collected for use prior to this project (Ennos 1995b). This material, though not related to the trees on the estate, was used to check inheritance of the DNA markers.

Several provenances of both European and Japanese larch are maintained at the Northern Research Station near Edinburgh. This material was used for a brief morphological analyses of twig, cone and leaf characters.

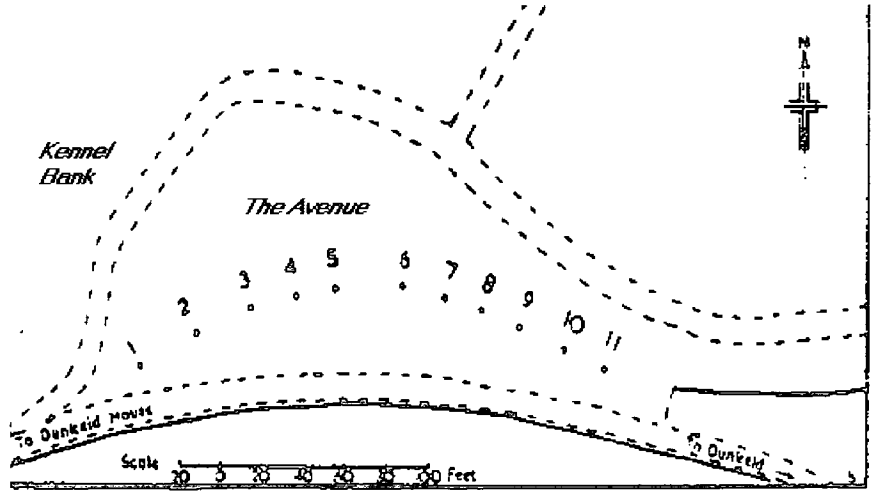
Needles were collected in the summer for DNA extraction, and dormant buds in the winter for isozyme analysis.

Figure 3-1: Locations of the parental trees at Dunkeld (from *Matthews*)



1: The Japanese Trees On The Avenue (Figure 3-2)
2: The European Trees At Kennel Bank
3: The Surviving 'Parent Larch'

Figure 3-2: Map of the Japanese trees on The Avenue(from *Laing*)



3.4.2 Isozyme Analysis

Screening for species-specific isozyme markers was carried out on dormant buds. Shoots were collected between November and March, stored in the fridge, and used as soon as possible (usually within two weeks).

Full details of the process and recipes for buffer solutions are given in Appendix A (p. 181). Removal of the bud cover was necessary to prevent the high levels of resin interfering with the process. Buds were homogenised in a cold extraction buffer to yield a homogenate containing enzymes. The homogenate was then absorbed onto small paper wicks.

Samples were then inserted into gels made up of a polymerised starch solution in a buffer of the required pH and salt concentration (gel buffer). Sponges were used to connect each end of the gel with separate reservoirs of another buffer (electrophoresis buffer). An electric potential was applied across the two electrophoresis buffers, resulting in an electric field across the gel and migration of the components of the sample based on charge, size, and shape.

Slices of the gel were then placed in colorimetric solutions that contains substrates for specific cellular enzymes. Any enzymes capable of utilising these substrates showed up as coloured bands on the gel. In this way, the relative mobility of biochemically identical or similar enzymes was determined.

The procedures and recipes used in this starch gel electrophoresis are based on Cheliak and Patel 1984, with modifications as described in the Appendix (p. 181).

3.4.3 DNA extraction and Quantification

Several DNA extraction methods were used, and are detailed in Appendix A (p.176-179). Extraction of consistent and high quality DNA from needles proved problematic. Initially, the protocol used in the laboratory for extraction of pine DNA was used (Morman 1994). This utilised a buffer containing CTAB, a carbohydrate that under certain salt conditions dissolves DNA. Frozen needles were ground in liquid nitrogen. The resultant powder was incubated in a water bath with the CTAB buffer. This was mixed with chloroform, which dissolved many proteins and carbohydrates while leaving the CTAB soluble DNA in the aqueous phase. Removal of the organic phase, followed by ethanol precipitation of the aqueous phase, yielded a crude 'pellet' of DNA. This was then cleaned further using more chloroform or phenol/chloroform to remove contaminating proteins and carbohydrates, which can inhibit restriction enzyme activity and prevent reproducible PCR amplification.

DNA was quantified by electrophoresis through an agarose gel and ethidium bromide staining (p.179), which causes DNA to fluoresce with an intensity dependent on its concentration. Quantification in a UV spectrophotometer was not consistent with these results, and it has been suggested that CTAB contaminants can interfere and prevent accurate assessment of concentration by spectrophotometry (Doyle and Doyle 1990).

This extraction protocol did not yield any DNA. Replacement of the chloroform as an organic solvent with dichloromethane, which improved yields in some other conifers (Gillies 1995), resulted in substantial amounts of DNA. The reasons for this are unknown, but it suggests that larch needles contain a compound that solubilises DNA in chloroform, but not dichloromethane.

This adjusted protocol was found to be suitable for many applications. However, PCR techniques appeared particularly sensitive to contaminants, and continued attempts to improve the consistency and purity of DNA were made. None of these were used in analysis of the parents.

3.4.4 Southern Blotting And RFLP analysis

Screening for species-specific markers of organellar and nuclear genomes was carried out by radioactive probing of Southern blots. Full experimental details are given in Appendix A (p.187-189). Extracted DNA was digested with a restriction enzyme, resulting in a mix of DNA sequences of varying sizes, bounded by the recognition site of the enzyme. This genomic digest was run on an agarose gel. The gel was soaked in sodium hydroxide, to denature the DNA so as to make it single stranded. After neutralisation of the gel, a nitro-cellulose filter- which binds DNA- was placed on the gel. The filter was covered with layers of tissue in order to draw liquid through the gel by capillary action.The whole 'sandwich' was placed, gel side down, on a wick which was soaking in a reservoir of salt solution. Capillary action draws the solution through the gel and filter, and with it transfers the DNA from the gel to the filter. The filter was then treated by baking or UV irradiation to irreversibly bind the DNA to it.

The blot (or blots) were placed in rotating glass cylinders containing 'hybridisation buffer' in a benchtop hybridisation oven. This equilibrates the blots at the correct temperature for probing. Radioactively labelled probes, homologous to the sequence under analysis, were denatured and mixed with the hybridisation buffer, and left in the rotating cylinders overnight. The single stranded probes bind to their homologues on the blot through DNA base-pairing. The radioactive solution was decanted and stored for disposal, and the blots were rinsed under various degrees of stringency to remove background. This is the most crucial stage of the

operation, as over-stringent washing can remove the signal but under washing can leave unacceptably high background. The resultant radioactive blots were then placed in a specially designed chemi-luminescent cartridge with a film. Radiation striking the luminescent surface causes fluorescence, which leaves a signal on the film. After a suitable length of time (from several hours to two months, depending on concentration and probe), the film was processed in an automatic developer. Dark bands on the film corresponded to fragments in the original digest that had migrated through the agarose gel, with a velocity dependent on fragment length, and showing homology to the probe. Differences between two individuals in banding pattern correspond to differences in sequence- either loss or gain of the restriction site, or reorganisation of the genome.

3.4.5 Obtaining And Preparing Radioactive Probes

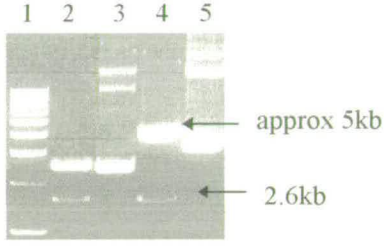
Several different probes were used. For detection of chloroplast variation, clone K140 from a clone library of the chloroplast genome of *Pinus contorta* (Lidholm and Gustafsson 1991) was used. This was chosen because it had already detected variation between the two larch species under investigation (Qian, Ennos and Helgason 1995). The probe was provided as an insertion into pUC plasmid (the vector) that had been transformed into *E.coli* (the host), supplied by the laboratory that constructed the clone library and stored in a 'stab culture'. The probe was isolated by growing the ampicillin resistant transgenic host in bacterial broth overnight. The plasmid was extracted using standard protocols. Digestion of the isolated plasmid with the restriction enzyme Kpn I cut the chloroplast sequence out of the plasmid. The digested plasmid was run through an agarose gel and visualised, and the 9.2kb fragment corresponding to the chloroplast sequence cut out of the gel and extracted.

For detection of mitochondrial variation, it had already been found that the ORF25 genes and COXI genes from wheat distinguished between the two species (Deverno, Charest and Bonen 1993). Rather than use these heterologous probes, a COX I clone prepared from PCR amplified Scots pine DNA(Sinclair, Morman and Ennos 1998) and PCR amplified larch ORF25, using primers designed by Sinclair and Ennos as part of the same study were used.

In order to find variation in the nuclear genome, a 5.2kb rDNA probe from *Taraxacum* was used. This was kindly supplied by Kirsten Wolff, St.Andrews University as, an insert in a 2.6kb pUC plasmid. This probe, designated PTEE5, was transformed into the host by electroporation, and subsequently grown and harvested in the same manner as the chloroplast K140 (Figure 3-3).

Once probes had been isolated, they were denatured and radioactively labelled using a polymerase, short, random primers and radioactive dCTP. This method incorporates radioactive nucleotides into the probe DNA, which is then denatured to make it single-stranded, and placed in the rotating cylinders with the blots and hybridisation buffer as above.

Figure 3-3: Ethidium stained agarose gel showing plasmid PTEE5 before (lane 5) and after (lane 4) digestion with KpnI



Lane 1 is the 1kb Marker, lanes 2 and 3 are from PTEE3 (not used in this study).

Lane 4 is the digested PTEE5 clone (after transformation and harvesting). The 2.6kb vector and the 5.2kb insert can be seen.

Lane 5 is the undigested plasmid containing the insert, with extra bands representing supercoiled plasmid.

3.4.6 PCR Based Methods

Attempts were made to find molecular variation with non-radioactive methods. The PCR reaction utilises a thermostable polymerase, nucleotides and 'primers'- short sequences of DNA (about 20 bp). The target DNA was first denatured, and then cooled in the 'annealing' phase, to allow the primers to base-pair with complementary sequences in the target. The reaction was then heated to 72°C, causing the polymerase to replicate the sequences to which the primers were bound. This was repeated automatically about 40 times. In this way, regions of DNA that have sequences bounded by the two primer sequences were amplified. The protocol for this is described in the Appendix (p.182)

3.4.6.1 Design Of PCR Primers

In this study, 'universal primers', designed from consensus sequences of plant mitochondria and chloroplasts (Demesure *et al.* 1995; Gielly and Taberlet 1994; Sinclair, Morman and Ennos 1998), were used to PCR amplify genomic DNA. These primers were made by comparing sequences of genes that have been characterised in a wide range of species, and designing primers to conserved regions. In this way, primers were produced that could amplify DNA successfully in many species.

Similarly, universal primers to nuclear genes (Strand, Leebens-Mack and Milligan 1997) were used in an attempt to find additional nuclear variation. These primers were designed to

conserved regions of nuclear genes for which sequences existed over a range of species, such as alcohol dehydrogenase. Such primers are quickly becoming recognised as a valuable tool for the analysis of nuclear variation. As discussed in the introduction, there is still a dearth of such primers. In addition to these primers, a primer was designed to the large subunit of Rubisco for which a larch sequence exists (Hutchison 1989).

3.4.6.2 Analysis Of PCR Fragments

Successfully amplified PCR products were digested with a variety of restriction enzymes to screen for sequence variation leading to loss or gain of restriction sites. Digested products were run and visualised on an agarose gel.

Other methods for screening for variation in PCR fragments were also tested, and described in the Appendix (p.185). These included heteroduplex analysis (Cotton 1993), whereby two products from different individuals are mixed together. Denaturing these and reannealing them gives the opportunity for heteroduplexes between different alleles to be formed. Heteroduplexes have a different mobility through a gel than the same size homoduplex, allowing differences between two alleles to be detected. Another method used an enzyme that cuts single stranded regions in order to screen for heteroduplex formation. Finally Single Stranded Conformational Polymorphism analysis (Hongyo *et al.* 1993) was carried out on some fragments. SSCP relies on differences in secondary structure between two single stranded fragments of the same length but slightly different sequences which leads to different mobilities through a gel.

3.4.6.3 RAPD Analysis

Randomly Amplified Polymorphic DNA analysis (Hadrys, Balick and Schierwater 1992) utilises the PCR reaction with short (10mer) primers of random sequence. The PCR reaction was carried out at a very low temperature (about 42°C). Only one primer (as opposed to the usual two) was used. In this way, any sequence that was bounded by the primer sequence and its reverse complement was amplified. The short length of the primer means there are likely to be many such sequences, and a low temperature is needed to allow such short primers to anneal to amplify the template. The product- often containing many bands- was run on an polyacrylamide gel, to provide greater resolution and sensitivity than agarose gel electrophoresis. Originally, silver staining was used, which is a more sensitive detection method than ethidium bromide staining. However, this sensitivity was not required and ethidium bromide staining was used in the remainder of the analysis.

PCR reactions were carried out either in a Perkin-Elmer 480, with 48 samples, or a Perkin-Elmer 9600 with a 96 sample capacity. Use of ammonium buffer rather than the usual potassium buffer (Allnutt *et al.* 1998) was found to be critical. The method is described in the Appendix (p. 184).The polyacrylamide gels had a capacity of 24 samples per gel, excluding markers, and 2 gels could be run simultaneously in the same buffer, theoretically allowing exact comparison across 48 samples.

In order to screen for species-specific markers, 3 individuals from each larch species were amplified using 50 primers. Where a marker was present in all 3 individuals of one species but not another, all 11 Japanese parents and 12 of the European parents were analysed by that primer. Primers that resulted in species-specific markers after this were used in later analyses.

3.4.7 Morphological Analysis

The trees under study varied in age from 250 years to 2 or 3 years old. Thus, characters greatly dependent on age and silvicultural treatment such as crown form, branch number and size characteristics were not considered. The provenance material from NRS was used to examine the feasibility of using characters such as cone form, shoot colour and needle structure as genetic markers.

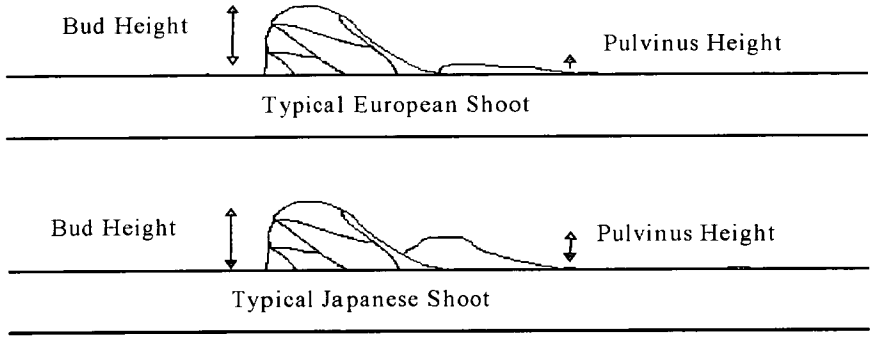
There are microscopic characters of the needles that have been used to differentiate the two species (Laing 1947). However, I was not able to reliably score these. More effort to improve the needle sections used was not deemed worthwhile, as the variability and lack of intermediacy in these characters in the hybrid had already been noted.

From each individual, 5 cones were taken and their lengths and widths (at maximum girth) measured. From each cone, a scale was taken from a point on the cones maximum width and its length and width measured. Curvature of the scale was assessed by measuring the maximum deviation of the scale from its longitudinal axis (using a pair of calipers). A seed was also taken from each cone (where possible) and its length and width measured. Cone shape, seed shape, and scale shape were calculated as length divided by width.

From each individual, 5 shoots were sampled and the length of that year's growth measured. The number of buds on that year's growth was counted, allowing a calculation of the bud density (number of buds/cm) to be made. The colour of each shoot was scored subjectively on a scale of 1 to 5 (5 being deep red, 4 being pink, 3 being pink tinged, 2 being slightly orange, and 1 being straw coloured). The pubescence of each shoot was scored subjectively on a scale of 1 to 3, 1 representing the glabrous nature of European individuals, 3 the 'hairy' Japanese trees and 2 an intermediate state. From each of three shoots, the last bud on the shoot (apart

from the terminal bud) was examined for two characteristics- maximum height of the bud and maximum height of the pulvinus of the bud (Figure 3-4). The ratio of pulvinus height to bud height provided a measurement of the tendency of Japanese pulvinii to be carried further up the bud than in the European tree, and this is denoted by the variable 'pulvinus shape'. Principal Component Analysis (PCA) was used (on Minitab 10) to see how informative these characters might be in further analysis. Principal Component Analysis constructs variables that are linear functions of the measurements taken, such that these variables can be used to present as much variation as possible. These *components* are thus a more useful way of summarising data and detecting patterns than any single variable alone.

Figure 3-4: Diagrammatic representation of measurements taken of European and Japanese buds



3.5 Results

3.5.1 Isozymes

The initial screening was carried out on the Japanese trees and 12 of the European trees. The following 14 enzyme systems were screened, and the results of this are shown in Table 3-2. Those that are marked as 'unscored' meant that activity, and possibly variation, was present, but under the conditions tried no reliable scoring was possible. If there did appear to be variation in these loci, it was not inter-specific and it was not worthwhile optimising conditions.

Table 3-2: Results of initial isozyme assays for detecting variation between parental species

No activity	glucose dehydrogenase (GLUDH), hexokinase (HK), Nicotinamide adenine dinucleotide dehydrogenase (NADH), esterase (EST), Alcohol dehydrogenase (ADH)
Monomorphic	aconitase (ACO) , isocitrate dehyrogenase (IDH),
Unscored	phosphoglucose isomerase (PGI), leucine-amino peptidase (LAP), malic enzyme (ME),
Variation	shikimic acid dehydrogenase (SKDH) , phosphoglucomutase (PGM), glutamic oxaloacetic transaminase (GOT)

From this initial screening, 3 out of 13 systems were taken to be useful in subsequent analyses. The failure of NADH to give activity was disappointing, given that it had previously been found useful in a hybrid larch seed orchard (Hacker and Bergmann 1991).

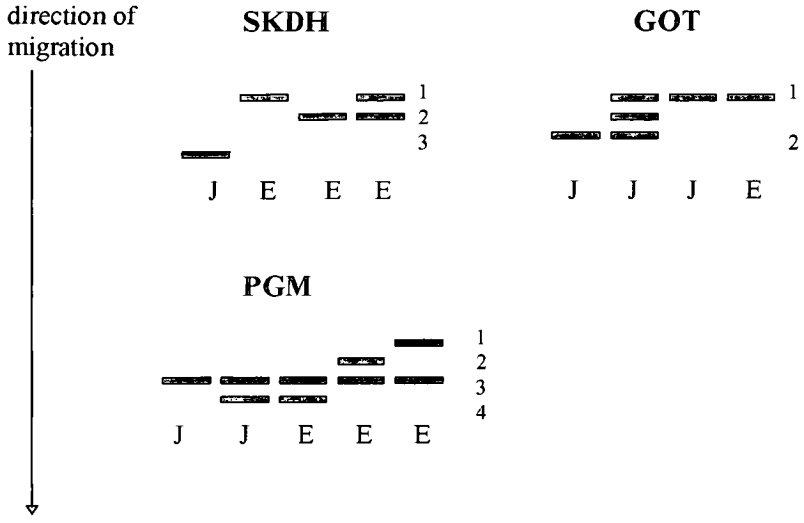
Table 3-3 shows the results for these three loci in the parents, and Figure 3-5 represents the banding patterns seen. Genotypic frequencies at none of the loci showed a deviation from Hardy-Weinberg equilibrium in either the Kennel Bank European parents or the Japanese parents on the Avenue. SKDH-1 is clearly a species-specific marker, with alleles SKDH-1a and SKDH-1b being specific to European parents, and SKDH-1c being specific to Japanese larch. GOT-2a was fixed in the European parent, but was present along with GOT-2b in the Japanese parents. PGM-1 had 4 alleles, the most common PGM-1c being present at a lower frequency in Japanese larch than in European larch. These three loci are therefore potentially useful markers for determining the nuclear composition of hybrids.

Table 3-3: Allele frequencies at three isozyme loci in parent trees

Locus	Allele	Frequency in European parents	Frequency in Japanese parents
SKDH-1	a	10/24	0/22
	b	14/24	0/22
	c	0/24	22/22
GOT-2	a	24/24	10/22
	b	0/24	12/22
PGM-1	a	1/22	0/20
	b	7/22	0/20
	c	13/22	19/20
	d	1/22	1/20

Notes: The allele frequencies are shown with denominators so the number of reliably scored individuals can be seen. 12 out of the 15 European parents were scored, one of which did not work for PGM. All 11 Japanese trees were scored for SKDH and GOT, but only 10 were scored for PGM.

Figure 3-5: Isozyme banding pattern seen in parents



Note: All banding patterns seen in parents for SKDH and GOT are shown, only some of the patterns seen in the European trees are shown for PGM

3.5.2 Mitochondrial RFLPs

The ORF25/Kpn and COX1/Dra probe/enzyme combinations both revealed fixed, inter-specific variation (Figure 3-6). The COX1/Dra probe was fainter and only the ORF25/Kpn results are shown. The Japanese parents all contained a 2.4kb band, whereas the Europeans had a 2.8kb band instead.

Inheritance was checked using the controlled crosses. Five pairs of controlled crosses were used, all of which had a European seed parent and a Japanese pollen parent. One offspring was scored from each of these crosses, except in the fifth cross (designated E361xJ42), where 5 offspring were scored. Offspring from these crosses had the European mitotype, as expected. Figure 3-7 shows the four crosses with their single offspring each and the final cross with 5 offspring.

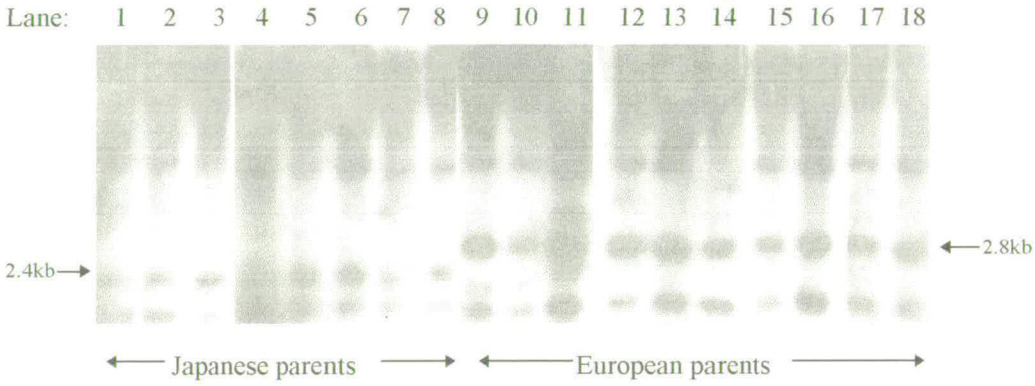
Overall, 10/11 of the Japanese parents from the Avenue and the same 12 European trees used in the isozyme analysis were scored.

3.5.3 Chloroplast RFLPs

The clone K140/Bcl probe/enzyme combination revealed inter-specific variation (Figure 3-8). The Japanese parents had a strong band at 2.9kb, whereas the European parents had a double band, the strongest one being at 3.0kb and the fainter one at 2.85kb. There was one exception to this pattern, with one European parent not having the 3 kb band but having a very strong

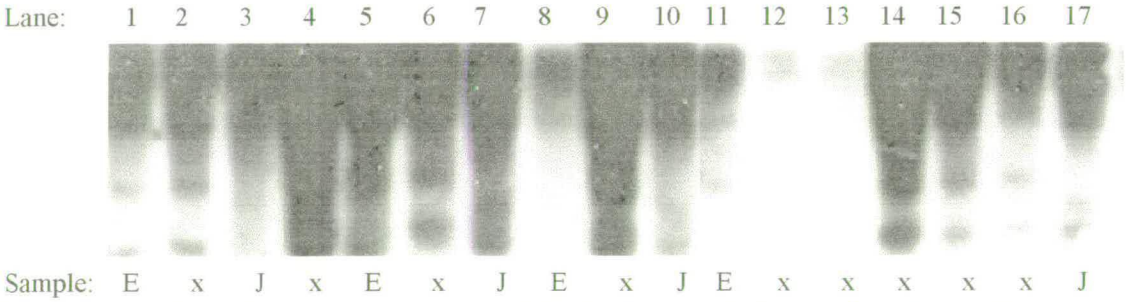
band at 5.0kb instead. This probe/enzyme combination was therefore suited to discriminating the two parents, in agreement with a previous study (Qian, Ennos and Helgason 1995). Paternal inheritance of this marker was verified both in the previous study and by analysis of the controlled crosses, where it was observed that all offspring of controlled crosses had the paternal Japanese chlorotype. Of the Japanese parents, 10/11 trees from the Avenue were scored as were 12/16 of the putative European parents.

Figure 3-6: Autoradiograph showing Southern blot of KPN digested genomic DNA from Japanese and European larch, probed with ORF25 mitochondrial probe



Lanes 1-8 are 8 Japanese parents from the Avenue, and all contain a 2.4kb band.
Lanes 9-18 are 10 European parents from Kennel Bank, and all contain a 2.8kb band.

Figure 3-7: Autoradiograph showing Southern blot of KPN digested genomic DNA From hybrids and parents from controlled crosses between Japanese and European larch, probed with ORF25 mitochondrial probe

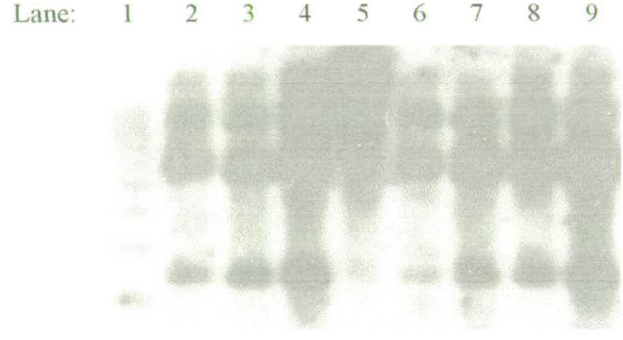


Lanes marked 'E' are the European mothers, 'J' are the Japanese fathers, and 'x' the hybrids.
Lanes 1-3: E361, E361x J8, J8 Lanes 4-5: E351xJ8,E351 Lanes 6-7:E351xJ42,J42
Lanes 8-10: E380, E380 x J42, J42 Lanes 11-17:E361, 5 crosses of E361xJ42, J42
Maternal Inheritance best seen in lanes 1-3 and 15-17, where the hybrids in lanes 2 and 16 can be seen to have the mitotype of the European mothers in lanes 1 and 15.

Figure 3-8: Autoradiograph showing Southern blot of Bcl digested genomic DNA from Japanese and European larch, probed with K140 chloroplast probe

Lanes 1-6: Japanese Parents from the Avenue Lanes 7-18:European Parents from Kennel Bank

Figure 3-9: Autoradiograph showing Southern blot of SAC1 digested genomic DNA from Japanese and European larch, probed with rDNA probe PTEE5



Lane 1:1 kb Marker Lanes 2-5:European Parents(Kennel Bank) Lanes 6-9:Japanese Parents(the Avenue)

3.5.4 rDNA RFLPs

No variation was found in the nuclear genome using probe PTEE5 and 10 restriction enzymes on genomic digests. A representative blot, in this case using the SAC restriction enzyme, is shown in Figure 3-9.

Thus, the rDNA probes were not used in the analysis of further generations.

3.5.5 Analysis Of PCR Fragments

The chloroplast universal Taberlet primer pairs ab, cd, and ef (Gielly and Taberlet 1994) successfully amplified single products in both parental species. None of the mitochondrial universal primers (Demesure, Sodzi and Petit 1995) yielded strong, reproducible single bands. 12 enzymes, including both 4bp and 6bp cutters, were used to digest the cpDNA amplification products from three individuals of each species (from the Avenue and the Kennel Bank parents). No variation, inter- or intra-specific, was found. SSCP analysis of the fragment produced by primer pair ef showed differences from Scots pine (used as a control) but no differences were evident between a representative from each of the two larches (Figure 3-10).

Heteroduplex analysis of the same individuals also failed to detect variation between representatives of the two species.

PCR amplification using nuclear primers was also attempted. Three primers which amplified ADH, calmodulin and Glyceraldehyde 3-phosphate dehyrogenase genes were used with the original authors conditions for starting points (Strand, Leebens-Mack and Milligan 1997), but no successful amplification was achieved.

The RUBISCO primer produced a poor quality double banded product. Experimenting with different PCR conditions, and even attempting to reamplify from one band, failed to produce a product that was suitable for further analysis.

Thus, analysis of organellar PCR products failed to reveal any variation- a result later backed up by the sequencing of some of the same products- and nuclear genes were not amplified successfully. Therefore, analysis of PCR products could not be applied to the detection of hybrids.

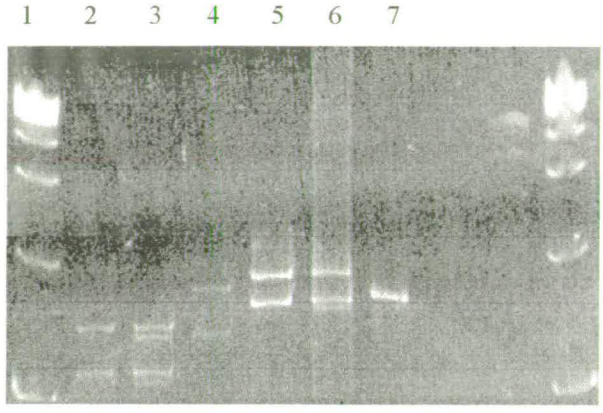
3.5.6 Morphological analyses

Morphological analysis was performed on individuals from the provenance trials at Bush. Table 3-4 shows the characters measured and those which were significantly different between species (from a multivariate analysis of variance, or MANOVA, performed on Minitab 10). Table 3-4 also shows the weight of each character in the first 2 components of the PCA. Only those characters that were significantly different at below the 5% level were used. The subjectively scored, discrete characters (pubescence and colour) were excluded from this analysis to allow an assessment of the use of the more variable quantitative characters. Figure 3-11 shows the first two components plotted against each other. The first two components account for nearly 70% of the variation present in this sample.

The first component provides a degree of separation between the two species. The coefficients are all negative apart from the cone shape which is positive. This reflects the fact that the mean measurement in the European trees is lower than in the Japanese trees for all characters apart from cone shape, such that the first component clearly describes the species that the individual belongs to. The coefficients are all of a similar magnitude (reflecting the fact that only those characters which provided discrimination were used in the PCA in the first place). It is not so easy to provide a biologically meaningful interpretation of the second component.

It can be seen that there are morphological characters that could provide the basis of a separation of the two species. However, it appears unlikely that hybrids would cluster so tightly between the two groups so as to be distinct from the parents.

Figure 3-10: Polyacrylamide gel showing SSCP analysis of Taberlet fragments ab and cd from larch and pine



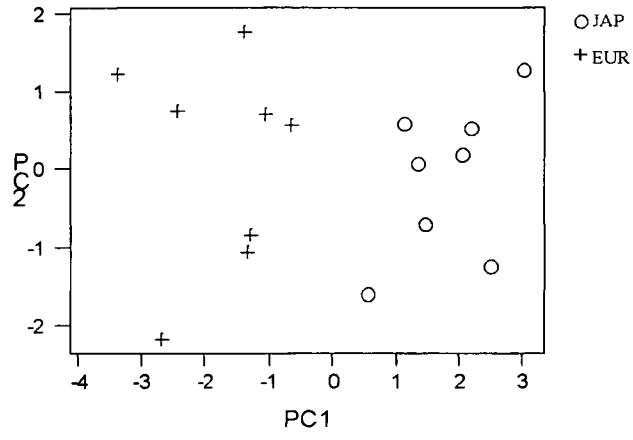
Notes: Lane 1 is a marker. It's weights do not correspond to fragment size, as the fragments are single-stranded in a non-denaturing gel. Lanes 2 and 3 are Taberlet fragments ab from a Japanese and a European parent. Lane 4 is a fragment from a Scots Pine. Lanes 5-7 are Taberlet fragment cd from the same samples. Note the large difference between larch and pine, though no size differences are observed in normal electrophoresis. No difference is seen between the two larch species.

Table 3-4 : Characters measured in morphological analysis, their means in each species, results of ANOVA, and coefficients for those used in PCA

Character	*Mean in Japanese*	*Mean in European*	*Significance of between Species Variation (from MANOVA)*	*PCA 1*	*PCA 2*
Cone length	2.95cm	3.24cm	27.6%	-	-
Cone width	2.24cm	1.99cm	1.4%	-0.319	0.498
Cone shape	1.32	1.64	0.4%%	0.321	-0.262
Scale length	1.18cm	1.19cm	93%	-	-
Scale width	2.59cm	2.78cm	44%	-	-
Scale shape	0.99	0.98	35%	-	-
Scale curve	1.36cm	0.57cm	<0.5%	-0.404	0.391
Seed length	4.61mm	4.50mm	60%	-	-
Seed width	2.59mm	2.78mm	7.4%	-	-
Seed shape	1.79	1.63	0.8%	-0.295	0.280
Bud density	1.25/cm	0.95/cm	<0.5%	-0.325	-0.166
Bud height	13.7mm	10.7mm	<0.5%	-0.371	-0.357
Pulvinus ht.	5.3mm	3.1mm	<0.5%	-0.407	-0.447
Pulv. shape	0.40	0.28	2.8%	-0.368	-0.312
Shoot Colour	1	4.4	<0.5%	-	-
Pubescence	1	2.8	<0.5%	-	-

PCA and ANOVA were performed using Minitab 10. Material was 8 individuals of each species from the NRS provenance trials, 5 samples from each individual were used for cone characters and 3 samples for shoot and bud characters.

Figure 3-11: PCA on morphological characters of individuals from provenance trial



PCA performed using Minitab 10 (see Table 3-4 for details). The first component separates the two species. The two components account for 70% of the variation.

3.5.7 RAPD analysis

Reliability of RAPDs was a major problem. Much of the difficulty was later put down to variations in the quality of de-ionised water. However, the most important factor was likely to be the quality of DNA. Even when the concentration of the samples was accurately equalised, many samples did not amplify. Thus, the concentration of contaminants that could not be removed even after many cleaning up steps was clearly a variable factor affecting PCR based techniques. The result of this was that amplification of all samples in one run was not achieved, and as repeatability between runs was not high, it was impossible to reliably score bands from one run to the next. For this reason, only those bands that were relatively simple to score and consistent from one run to the next were used in the analysis.

Table 3-5 shows the primers that were considered useful, and the frequency of the markers in the two parents. It also shows the results of testing the inheritance of these markers in a controlled cross, 361x42, with a European tree as seed parent. The two parents and 8 offspring were analysed. On the assumption that the parents are homozygous for either the band or absence of the band, one can infer from this whether the band is behaving in a dominant fashion- all offspring should have the band. If there is deviation from this, it is marked with an asterix and an explanation is offered. Photographs are shown in Figure 3-12 through to Figure 3-16. Figure 3-12, the RAPD profile generated using OPA7, demonstrates the existence of a 580bp product found when DNA from either Japanese or hybrid trees is used as a template. Figure 3-13 (using OPAL14) shows a 650bp band absent in European trees (this band is referred to as OPAL14a) and a 520bp band present in most European and hybrid trees, but not in most Japanese trees (referred to as OPAL14b). Figure 3-14, using

OPAL16, also shows two useful markers: the 900bp OPAl16a, absent in DNA from Japanese trees and the 480bp OPAL16b, absent in most European samples. The profile generated by OPK16 is shown in figure 3-15. OPK16a is an 850bp band amplified from all samples other than several European parents, OPK16b is a 700bp band absent in most of the Japanese Larch samples, and OPK16c is a 500bp band (actually, a doublet) present in all European parents. Finally, figure 3-16 shows how the inheritance of each marker was examined using the controlled cross. In this case, the primer used is OPK16.

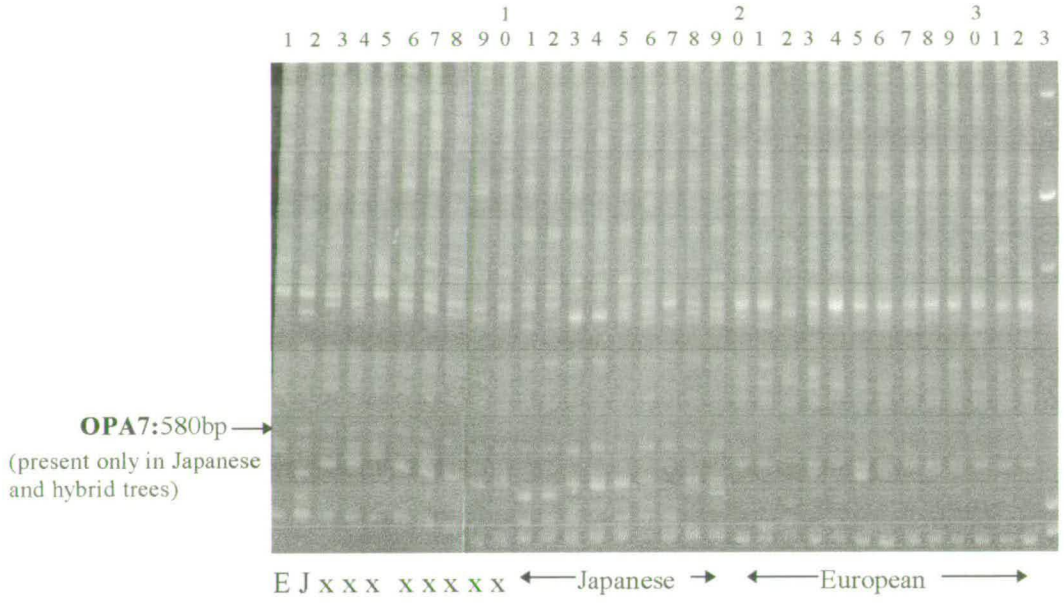
This approach cannot determine whether a RAPD marker is uniparentally inherited. As only one direction of cross is present- Japanese pollen parent and European seed parent- it is not possible to compare reciprocal crosses. Until chapter 5, where further data is presented, the assumption must be made that all markers are biparental.

Table 3-5: Frequencies of potentially useful RAPD markers in Japanese parents, European parents, and hybrids from a controlled cross

PRIMER	*approx. Length (bp).*	*Freq. in Jap. Parents*	*Freq. in Eur. Parents*	*Freq. in controlled cross hybrids*	*Strictly Dominant?*
OPA7	580	10/10	0/13	8/8	Yes
OPAL14a	650	9/9	0/11	8/8	Yes
OPAL14b	520	2/9	11/11	2/7	No*
OPAL16a	900	0/10	13/13	8/8	Yes
OPAL16b	480	8/9	0/13	5/7	No*
OPK16a	850	10/10	6/13	8/8	Yes
OPK16b	700	8/10	0/13	6/8	Yes
OPK16c	500	0/10	13/13	6/8	No*

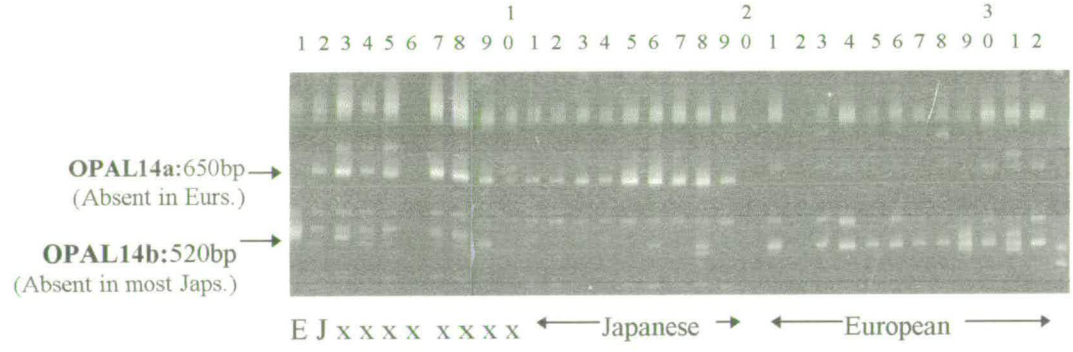
Note: Individuals that did not amplify correctly or reproducibly were not included in the analysis. For example, a frequency of 8/10 (from 11 parents) means 8 individuals had the marker, 2 did not, and one did not amplify or was not scorable.

Figure 3-12: RAPD analysis of parents and cross with OPA7



See legend at foot of page for details of lanes and samples

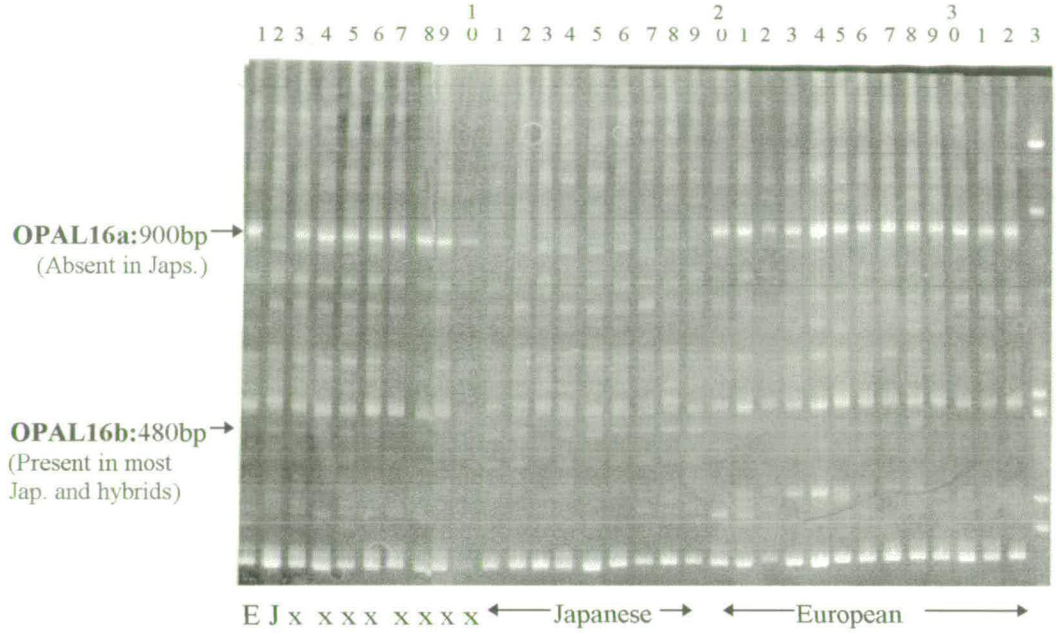
Figure 3-13: RAPD analysis of parents and crosses with OPAL14



For both Figure 3-12 and Figure 3-13, the lanes are labelled thus:

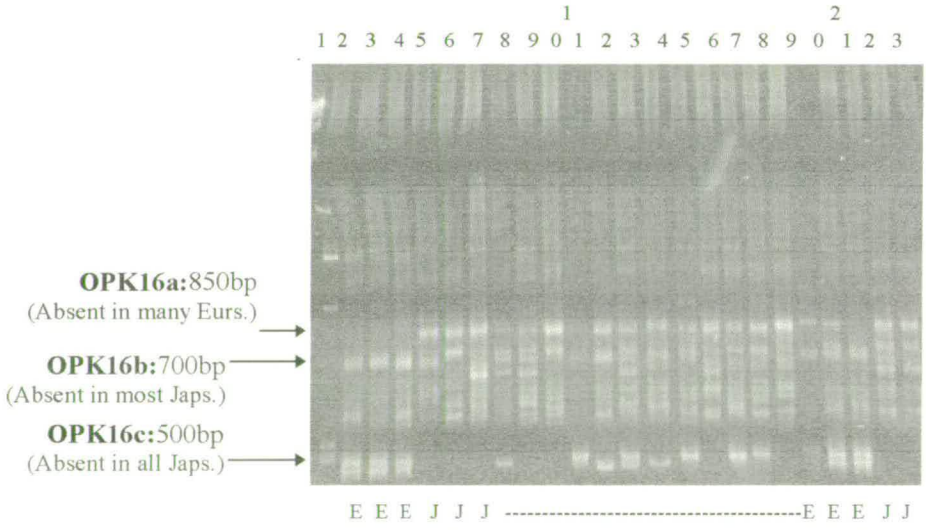
Lane 1: European parent (E361) of controlled cross
Lane 2: Japanese parent (J42) of controlled cross
Lanes 3-10: 8 hybrid offspring of cross E361 x J42
Lanes 11-19: 9 Japanese parents from The Avenue
Lanes 20-32: 13 European parents from Kennel bank
Lane 33: 1kb Marker

Figure 3-14: RAPD analysis of parents and cross with OPAL16



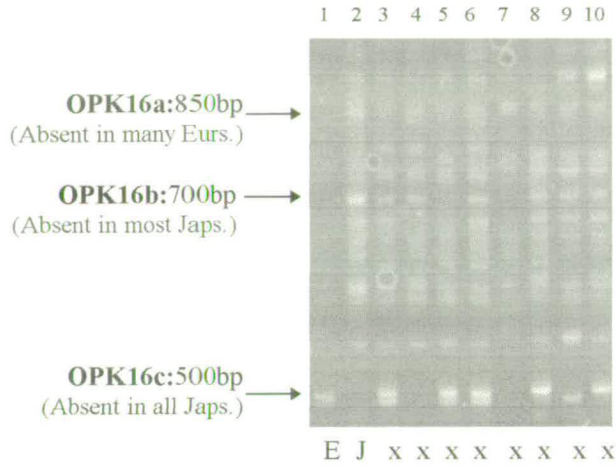
Lane 1: European parent (E361) of controlled cross
Lanes 3-10: 8 hybrid offspring of cross E361 x J42
Lanes 20-32: 13 European parents from Kennel bank
Lane 2: Japanese parent (J42) of controlled cross
Lanes 11-19: 9 Japanese parents from The Avenue
Lane 33: 1kb Marker

Figure 3-15: RAPD analysis of parents with OPK16



Lane 1:Marker Lanes 2-4:European Parents (Kennel Bank) Lanes 5-7: Japanese Parents (the Avenue)
Lanes 8-19: Mixed Parents and Hybrids Lanes 20-22: European Parents Lanes:23-24: Japanese Parents

Figure 3-16: RAPD analysis of controlled cross with OPK16



Lane 1: European Parent (E361) Lane 2: Japanese Parent (J42) Lanes 3-10: Hybrid Offspring

3.6 Discussion

For the RAPD markers, there are two markers that are fixed in the Japanese parents and not present in the European parents, and 'behave' (i.e. are inherited in a strictly dominant fashion) in the cross. OPA7 and OPAL14a are therefore diagnostic of Japanese parentage. Similarly, OPAL16a is diagnostic of European parentage. If these are successfully applied to later generations, and are nuclear markers, all F_1s should inherit all 3 bands. However, later generations will be missing some of these bands. For example, each individual in an F_2 generation would have a $1-0.75^3 = 57\%$ chance of being homozygous for the unamplified allele of at least one of these loci. In theory, therefore, these markers should be useful for determining the nature of a hybrid plantation even given their dominant nature.

There are three RAPD markers that do not 'behave' in the controlled crosses- while one parent of the cross has the band and the other does not, not all the offspring have the band. There are four possible explanations. Firstly, that the controlled crosses have not been successful and some of the offspring are mislabelled as hybrid but are in fact pure. This is not likely, as the individuals without the band are different for each marker. Secondly, that these reflect unreliable amplification products that are either amplified unreliably or are scored unreliably. However, the PCR reaction was repeated at least once for each marker and the results were consistent, so this is ruled out. Thirdly, that these reflect the complex inheritance that can be seen in RAPDs. It has been demonstrated that RAPD bands are often non-additive (Hallden *et al.* 1996), and that interference between various priming sites in the PCR reaction can lead to interactions between loci. In particular, competition for primer may result in one marker causing other markers not to amplify (Hallden *et al.* 1996). This explanation is not ruled out,

and may turn out to be the major problem in using RAPDs in the analysis of later generations. It is also the explanation favoured for OPAL14b, where all European parents (including that of the cross) had the marker, but only 2/7 of the offspring had the marker. Finally, if the parent containing the marker is heterozygous rather than homozygous, half the offspring are expected not to contain the marker. This is the explanation favoured for OPAL16b, where variation is present in the Japanese parents, making it more likely that the parent of the controlled cross is heterozygous. The 6/8 of the offspring that contain the marker is not significantly different from the expected 4/8 if heterozygosity was the case. Whatever explanation is favoured, these markers are considerably less useful for determining the exact nature of an individual tree.

Three isozyme systems and one RFLP for each organelle were found to be of potential use in analysing the genetic composition of hybrids. Alleles at one of the isozyme loci (SKDH-1) are species specific, whereas the other two (GOT-1 and PGM-1) have significantly different allele frequencies which are of potential use. For example, no tree with the GOT-1b allele can be a pure European tree. The organellar markers are species specific. This information alone allows the unambiguous discrimination of F_1 hybrids from parents. It should also allow certain classes of backcrosses and F_2 to be detected, though on their own they do not allow an F_1 to be unambiguously discriminated from an F_2. The diagnostic RAPD markers, even given their limitations, should be able to increase this discrimination, which is important for generations later than the F_1.

It is noted that analysis of the genetic composition of hybrid swarms would be greatly enhanced with the addition of more co-dominant, species specific nuclear markers. However, a combination of these markers will allow the genetic composition of hybrid swarms to be assessed to some extent.

4. Composition Of A Putative F_1 Hybrid Larch Plantation

4.1 Introduction

The genetic markers developed for Japanese and European Larch were used to assess the composition of a stand on the Atholl estate, a few miles from Dunkeld and the sites of the parental trees (Figure 4-1). This stand was planted in 1912 with seed collected from Japanese parents, and is described as an F_1 plantation by the estate foresters (Ennos 1995a). As further generations have been planted from this stand, it was considered an ideal place to start in an analysis of the immediate genetic effects of bringing two allopatric species into contact .

Figure 4-1: Location of the 1912 stand

NOTES: The putative hybrid population is designated compartment 268 by the Atholl estate, and is shown surrounded by a thick black line at the North East of the map, supplied by the estate. The road from Dunkeld up to the Lock of Craiglush was followed, where the car was parked and the track to the site followed on foot.

4.2 Aims

The main aim was to determine the nature of this stand. Described as an F_1, it was expected to consist of a mix of pure Japanese and F_1 hybrid individuals. Characterisation of its nature was essential to proceed with the analysis of later generations. A subsidiary aim was to test the suitability of previously discovered molecular markers to detect hybrids in the field, and to decide upon suitable methods of data analysis and presentation to apply to later generations.

4.3 Materials And Methods

Fifty trees were sampled out of a population of around 250. Sampling was not entirely random, as accessibility of lower branches was a consideration. The trees are over 80 years old and the lowest branches were not always healthy, often appearing dead or nearly so. However, apart from these considerations, sampling was random and covered the whole plantation. The exceptions to this were two trees which were chosen because they were apparently regenerants (about 15 years old). Though regeneration in this stand was uncommon due to browsing by rabbit and deer, these two individuals proved informative in determining the nature of this stand.

Shoots were collected in January so that dormant buds could be subjected to isozyme analysis. The trees were marked and coarsely mapped to facilitate the collection of needle samples in the summer from the same individuals.

All molecular methods are as described in chapter 3 and appendix A. Isozyme analysis was performed on dormant buds using the three systems (GOT, SKDH and PGM). DNA analysis was performed on 1.0g of frozen needle material using the CTAB/Dichloromethane method. Mitochondrial and chloroplast markers were provided by radioactively labelled ORF25 and K140 probes respectively. Parental DNA was analysed simultaneously as a control to confirm that the different haplotypes were being detected. RAPD analysis was not performed at this stage, and neither was detailed morphological analysis carried out. However, the trees were classified as Japanese, European, or Hybrid on the basis of shoot colour and cone shape. This classification was to provide a very rough guide to the consistency of *my* discriminatory ability in the field, and the molecular data.

4.4 Data Analysis

Interpretation of the data is not, in this case, greatly enhanced by use of disequilibria parameters. However, to facilitate comparison with other stands and to demonstrate one way in which such multi-locus data can be analysed, some of these parameters have been

calculated. Full details of the calculations are given in Appendix C. As there are a great many parameters that can be used, and different ways of estimating and interpreting such parameters, a discussion of the approach is required.

4.4.1 Cytoplasmic Disequilibrium

The first, and simplest, of the multi-locus disequilibria is the cytoplasmic disequilibrium between chloroplast and mitochondria. Choosing one chlorotype and one mitotype (arbitrarily) as "+ve", this quantifies the association between these two cytotypes, a positive value indicating they are associated and a negative value indicating a repulsion. As each individual has only one copy of each, these can be calculated as the direct analogue of the gametic linkage disequilibria between a pair of nuclear loci. Thus,

$$D_{CM} = P_{CM} - P_C P_M$$

where D_{CM} is the disequilibrium between chlorotype c and mitotype m, P_{CM} the frequency of individuals with both, and $P_C P_M$ the product of each frequency.

It would be possible to test for a significant disequilibrium based on the sampling properties of this measure. However, for reasons that will become clear when other parameters are discussed, an exact test on the contingency table (similar in concept, but more accurate particularly with small sample sizes, to a standard χ^2 test) has been performed. The significance of this (in terms of deviation from random mating expectations, not a particularly useful hypothesis in this case) is presented alongside the normalised disequilibrium (see below). It should be noted that it is the association (i.e. Ho is that chlorotype and mitotype are independent) and not the parameter itself (i.e. Ho is that D=0) that has been tested. As with the following associations, Genepop (Raymond and Rousset 1997) was used to perform the exact test.

4.4.2 Cytonuclear Disequilibria

The next disequilibria to be calculated were the cytonuclear disequilibria, measuring associations between a nuclear locus and an organellar gene. This is a little more complicated, as associations between nuclear genotypes and cytotypes must be taken into account. Such an association can exist, in principle, even if there is no overall association between allele and cytotype. There are three genotypic disequilibria and one allelic disequilibrium for each nuclear/cytoplasmic combination. However, two of these are fixed by the other two (Asmussen *et al.* 1987), so only two need be calculated. In this case, the allelic disequilibrium

and the genotypic disequilibrium between the "+ve" homozygote and the cytotype have been calculated. Thus,

$D_{CA}=P_{CA}-P_CP_A$

where D_{CA} is the *allelic* disequilibrium between (in this case) the chloroplast and nuclear locus A. P_{CA} is calculated as the frequency of homozygotes with the cytotype plus half the frequency of heterozygotes with the cytotype. This is analogous to the cytoplasmic disequilibrium and linkage disequilibrium for nuclear alleles.

And

$D_{CAA}=P_{CAA}-P_CP_A$

where D_{CAA} is the *genotypic* disequilibrium between the chloroplast and homozygotes at locus A. P_{CAA} is simply the frequency of homozygotes with the chlorotype.

The other two disequilibria, D_{CAa} and D_{Caa} can be calculated from the relationships that must exist between the disequilibria (Asmussen, Arnold and Avise 1987).

$D_{CA} = D_{CAA}+(1/2)\ D_{CAa} = -D_{caa} - (1/2)\ D_{CAa} = (1/2)\ D_{CAA} - (1/2)\ D_{caa}$

These parameters are usually called D, D_1, D_2 and D_3- the notation used here, though unwieldy, makes it easier to determine what each parameter refers to, and allows simple distinction between (for example) chloroplast (D_C) and mitochondrion (D_M) based disequilibria.

Again, the values of D have not been tested directly, rather exact tests either on the genotypic or the allelic contingency tables has been performed. For each set of disequilibria, there are 5 tests that can be performed. One is testing for an allelic association. The next is to test for an overall genotypic association. The final 3 can be regarded as tests of each of the individual genotypic disequilibria, and are performed by converting the 3x2 genotypic contingency table into a 2x2 table (Basten and Asmussen 1997). For example, to test D_{CAA}, the contingency table used has as its nuclear genotype categories 'AA' and 'not AA' (Basten and Asmussen 1997). This tests for an association between a cytotype and the homozygote AA, which may be significant even though there is no significant overall disequilibrium.

4.4.3 Nuclear Disequilibria

The final and most complicated range of disequilibria refer to the disequilibria between nuclear loci. Without gametic information, the gametic, or linkage, disequilibrium cannot be calculated. Thus direct analogues of the cytonuclear and cytoplasmic disequilibria cannot be given. However, it is possible to define a "composite" genotypic disequilibrium that measures the association between genotypes (Weir 1990). In principle digenic, trigenic and quadrigenic

disequilibria can be calculated. In this case, only the digenic disequilibrium which, in many cases, can simply be interpreted as a measure of association between alleles, is calculated. So,

$\Delta_{AB}=n_{AB}/n-2P_AP_B$

where Δ_{AB} is the *composite genotypic* disequilibrium between nuclear loci A and B, n is the sample size and $n_{AB} = 2n_{AABB} + n_{AaBB} + n_{AABb} + (1/2)\, n_{AaBb}$ (i.e. twice the count of the double homozygote plus the count of each homozygote/heterozygote plus half the count of the double heterozygote.). See Appendix C for details.

Testing the significance of this parameter is somewhat unwieldy and inaccurate (Weir 1990). For this reason, the exact test was performed, and for the sake of consistency the same approach was used as with the cytonuclear and cytoplasmic disequilibria.

4.4.4 Normalised Disequilibria

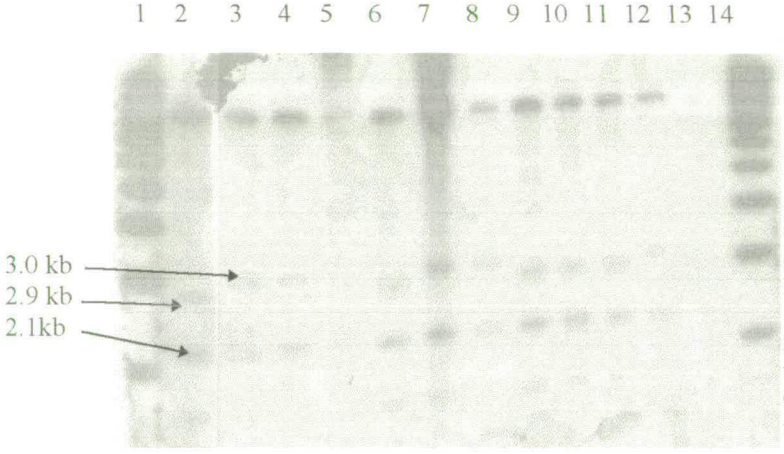
It is not always clear how to interpret a given disequilibrium. The magnitude of a value depends on both the allele frequencies, and the arbitrary naming of alleles. For this reason, normalised disequilibria which range between -1 and +1 have been defined for allelic disequilibria (Lewontin 1988). These have been extended to cytonuclear disequilibria (Asmussen and Basten 1996). It is noted that these normalised values are not independent of allele frequency, and their magnitude does not indicate statistical significance. However, if the significance of associations can be tested in another way (i.e. Fisher's exact test), they provide a useful way of summarising the extent and direction of these associations.

No constraints could be found in the literature for the composite disequilibrium. However, it is relatively easy to find maximum and minimum boundaries for this heuristically, based on the constraints placed on the genotypic contingency table by the marginal frequencies. This method is shown in Appendix C.

4.5 Results

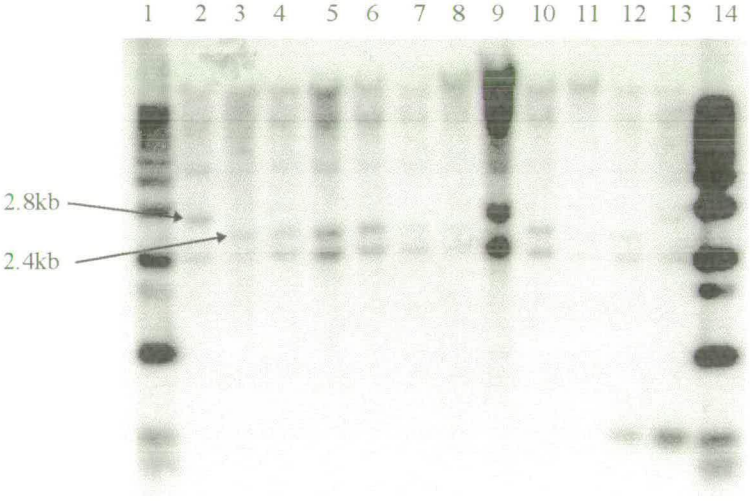
Representative blots of the two organellar genomes are shown in Figure 4-2 and Figure 4-3. Variation can be seen in both genomes.

Figure 4-2: Autoradiograph of Southern blot of Bcl digested genomic DNA from individuals in 1912 stand, probed with K140 chloroplast probe



Lanes 1 and 14 are markers.
Lanes 2 and 3 are Japanese and European parents for control purposes.
The difference between the two genotypes is most easily seen by comparing the distance between the band at 2.1kb and the band at either 3 or 2.9 kb.
Lanes 4-12 are samples from the 1912 'Hybrid' population
Note 4 individuals (4,7,8, and 12) that have the European chlorotype. One of these (lane 7) is a hybrid regenerant.

Figure 4-3: Autoradiograph of Southern blot of KPN digested DNA from individuals in 1912 stand, probed using mitochondrial ORF25 probe



Lanes 1 and 14 are 1kb Markers
Lanes 2 and 3 are European and Japanese parent controls.
Lanes 4-13 are 10 samples from the 1912 'hybrid' population
Of these 10 samples, only that in lane 9 has the European mitotype.

Table 4-1 shows the allele frequencies at each locus, nuclear and organellar, with the frequency of each allele in the putative parents presented alongside for ease of reference. The final column is the significance level of a χ^2 test for Hardy-Weinberg equilibrium. This table shows that as well as having variation at the chloroplast locus, there is also variation at the mitochondrial locus. Approximately 25% of the cp and mt haplotypes are European. The genotype frequencies at two of the isozyme loci (SKDH and GOT) are not in H-W equilibrium, suggesting that this is not the product of a random mating population.

The multi-locus data is summarised in Table 4-2 where various classes have been grouped together on the basis of organellar, SKDH and GOT data. Results for PGM are not presented as part of this table, as this method of presentation is not suitable for many loci or alleles and PGM does not, in any case, provide additional information. Only those individuals for which all loci have been scored are shown (though all individuals scored at the relevant loci are used for frequency and disequilibria calculations). For ease of interpretation, the notation used to present allele frequencies at SKDH-1 is different from that in Table 4-1. SKDH-1a and SKDH-1b, being European specific alleles, will be referred to jointly as SKDH-E. Allele SKDH-1c, which is Japanese specific, will be represented by SKDH-J. This non-standard notation will be used in much of the subsequent discussion. This table presents the important features of the data in an accessible form. Notice the presence of European mitochondria in several individuals, all of which also have the European chloroplast. The full genotype of each of the two regenerants (excluding PGM) is given in Table 4-2 as the two parenthesised classes. The "hypothetical parentage" is based on the hypothesised nature of the stand. It is not actually part of the data-set, rather an inference based on the data and properly belongs in the discussion. However, it is presented alongside the data in Table 4-2 for ease of reference when the discussion is reached.

Table 4-3 shows the various disequilibria described previously. A significant positive association was found to exist between the two cytotypes. The European variants of SKDH are associated with both European cytotypes, and GOT-1a (the allele found in 100% of European parents) is also associated with both European cytotypes. The associations exist at both a genotypic and allelic level, which is expected for hybrid or mixed stands of a young age (Asmussen, Arnold and Avise 1987). PGM shows no significant associations with cytotypes. Finally, there is a strong genotypic association between SKDH-E and GOT-1a, and a negative correlation between SKDH-E and PGM-1b (found at a higher frequency in the Japanese parents than the putative European parents).

The two regenerants are not included in the analysis of disequilibria and allele frequencies, as they are of a different generation and were not part of the random sample.

In terms of the morphological classification, 11 trees were regarded as European, 18 as Japanese and 21 (including the two regenerants) as *possible* hybrids (though all rather more similar to Japanese trees than expected from the literature, with the exception of the two regenerants). The morphologically European trees were found to be pure Europeans by molecular methods (assuming that the hypothesised origin of this stand is correct, see discussion). Similarly, none of the morphologically Japanese trees displayed European molecular characters. However, it was also found that all the hypothetical hybrids with the exception of the regenerants appeared to be pure Japanese trees.

The most important feature to notice is that, on the basis of the organellar data, there is no evidence of any hybrid formation apart from the two regenerants. This is shown, in terms of disequilibria, by the value of 1 for the normalised cytoplasmic disequilibria, indicating maximal association between the cytotypes. That the regenerants are hybrid provides evidence that this methodology was capable of detecting hybrids in the fields. Similarly, there are maximal associations between the most common European nuclear alleles and the European cytotypes in SKDH and GOT.

The next important point to notice is the presence of the 'Japanese specific' SKDH allele (SKDH-1c) in trees with both European organelles. Out of the nine individuals with European mitochondria, two are homozygote for SKDH-1c and two heterozygote.

It is also noteworthy that all the individuals with European mitochondria are homozygote for GOT-1a. All the putative parental European trees were homozygote for this allele.

Table 4-1: Single locus results in the 1912 stand, showing allele frequencies and tests for Hardy-Weinberg disequilibrium

Allele	Allele Freq. in 1912 Stand	Frequency in Putative Parents (Chap. 3)		Hardy-Weinberg Test (p value)
		Japanese	European	
Chloroplast				-
Japanese	32/43=0.74	1.00	0.00	
European	11/43=0.26	0.00	1.00	
Mitochondrion				-
Japanese	32/42=0.74	1.00	0.00	
European	11/42=0.26	0.00	1.00	
SKDH-1a,b	15/78=0.19	0.00	1.00	0.00 *** homozygote excess
SKDH-1c	63/78=0.81	1.00	0.00	
GOT-1a	43/82=0.52	0.45	1.00	0.00*** homozygote excess
GOT-1b	39/82=0.48	0.55	0.00	
PGM-1a	58/72=0.81	0.95	0.32	0.30 n.s
PGM-1b	14/72=0.19	0.05	0.68	
				0.000***

Notes:
1) Allele frequencies are shown with their denominators so that the number of individuals scored for that locus can be seen.
2) The HW test was an exact test performed by Genepop (Raymond 1997). The final cell in the HW column is the global test for all 3 loci.
3) The column containing frequencies in putative parents is presented to aid interpretation.

Table 4-2: Multi-locus results in the 1912 stand, showing genotype frequencies and hypothesised parentage

Chloroplast	Mitochondrion	SKDH-1	GOT-1	Hypothetical Parentage	Frequency
European	European	JJ	aa	European	2
European	European	JE	aa	European	3
European	European	EE	aa	European	6
European	Japanese	-	-	F_1 Hybrid	0
Japanese	European	-	-	F_1 Hybrid	0
Japanese	Japanese	JJ	aa	Japanese	7
Japanese	Japanese	JJ	ab	Japanese	6
Japanese	Japanese	JJ	bb	Japanese	16
(European)	(Japanese)	JJ	ab	(Regenerant)	(1)
(European)	(European)	JE	aa	(Regenerant)	(1)

Notes:
1) The SKDH alleles are denoted by 'J' for SKDH-1c(which was the only allele in the Japanese parents) and 'E' for SKDH-1a and SKDH-1b, which were the only two in the putative European parents.
2)The parenthesised individuals are the two younger trees chosen for their apparent hybrid nature, and are not part of the random sample of the stand.
3)The column labelled "hypothetical parentage" is for examination during the discussion.

Table 4-3: Nuclear, cytoplasmic, and cytonuclear disequilibria in the 1912 stand

Disequilibria	*Type of Association and Significance*	*Absolute D*	*Normalised D*
D_{CM}	Cytoplasmic Allelic***	0.17	1
D_{SG}	Nuclear Genotypic*	0.17	1
D_{GP}	Nuclear Genotypic n.s	0.045	0.24
D_{SP}	Nuclear Genotypic***	0.083	0.75
D_{CS}	Cytonuclear Allelic***	0.094	1
D_{CSS}	Cytonuclear Genotypic ***	0.063	1
D_{CSs}	Cytonuclear Genotypic *	0.063	1
D_{Css}	Cytonuclear Genotypic ***	-0.13	-1
	Overall Cytonuclear Genotypic ***		
D_{MS}	Cytonuclear Allelic***	0.100	1
D_{MSS}	Cytonuclear Genotypic***	0.060	1
D_{MSs}	Cytonuclear Genotypic *	0.080	1
D_{Mss}	Cytonuclear Genotypic ***	-0.14	-1
	Overall Cytonuclear Genotypic ***		
D_{CG}	Cytonuclear Allelic***	0.12	1
D_{CGG}	Cytonuclear Genotypic***	0.14	1
D_{CGg}	Cytonuclear Genotypic n.s	-0.037	-1
D_{Cgg}	Cytonuclear Genotypic***	-0.10	-1
	Overall Cytonuclear Genotypic***		
D_{MG}	Cytonuclear Allelic***	0.13	1
D_{MGG}	Cytonuclear Genotypic***	0.15	1
D_{MGg}	Cytonuclear Genotypic n.s	-0.040	-1
D_{Mgg}	Cytonuclear Genotypic***	-0.11	-1
	Overall Cytonuclear Genotypic***		
D_{CP}	Cytonuclear Allelic n.s	0.037	0.26
D_{CPP}	Cytonuclear Genotypic n.s	0	0
D_{CPp}	Cytonuclear Genotypic *	0.074	0.45
D_{Cpp}	Cytonuclear Genotypic n.s	-0.074	-0.69
	Overall Cytonuclear Genotypic n.s		
D_{MP}	Cytonuclear Allelic n.s	-0.14	-0.77
D_{MPP}	Cytonuclear Genotypic n.s	0	0
D_{MPp}	Cytonuclear Genotypic *	0.063	0.41
D_{Mpp}	Cytonuclear Genotypic n.s	-0.063	-0.41
	Overall Cytonuclear Genotypic n.s		

Notes:

1) Statistical significance is not based on the disequilibria parameters (normalised or absolute), but on the contingency tables for genotypic or allelic associations.

2) The subscripts are as described in Methods, with M and C representing mitochondria and chloroplasts, S representing SKDH, G for GOT and P for PGM.

3) "+ve" alleles are arbitrarily chosen to be the 'most European'. Thus, +ve alleles are the European cytotypes, SKDH-1a and b (treated as one allele, SKDH-E), GOT-1a, and PGM-1a,c and d.

4.6 Discussion(1): The Nature Of The 1912 Stand

Testing for statistical significance of the various associations was based on the null hypothesis of random-mating. This is clearly not the case- significant associations between all classes of markers exist. The expected result, and an alternative hypothesis is that all individuals are the progeny of Japanese seed parents and either European or Japanese pollen parents, with the organelles in particular being species-specific and uniparentally inherited. This hypothesis can be rejected without formal statistical tests, as the presence of European mitochondria in the stand represents an event with zero probability. Approximately a quarter of individuals have the European mitochondria, so this hypothesis must be rejected at any level.

Two explanations for this are presented, one of which is selected as the most plausible.

1) The organelle markers have not been applied or interpreted correctly. Many *ad hoc* explanations could then be put forward. In particular, it could be suggested that both organelle markers are chloroplast markers. In this case, the 11 individuals scored as European parents would be hybrids. This is not totally implausible, particularly given that some of these individuals contained an SKDH-1 allele previously scored as Japanese specific. However, the organellar autoradiographs concurs with the previous study on parents and crosses, which showed the different inheritance patterns of these markers. Furthermore, that the two regenerants analysed at the same time were detected as hybrids demonstrates the success of the methods. Finally, though the presence of 'Japanese specific' SKDH-1c alleles in the hypothetical pure European trees requires further explanation, all these individuals are homozygous for GOT-1a, whereas no GOT-1b alleles (found at 55% in the Japanese parents) were present. This concurs with, though does not statistically imply (due to the low numbers involved) the hypothesis that these trees are not hybrids.

2) This is not a hybrid stand, but a mixture of both parental types. This could have arisen in a number of ways, for example mislabelling of the stand or failure of the Japanese parents to produce any hybrids that year. Formation of hybrids from Japanese seed parents is likely to be particularly variable (due to climactic variations), as the period of overlap between European pollen production and Japanese receptivity is low in any case (Ennos and Qian 1994). European trees would then be explained as either inclusion (purposeful or accidental) in the original planting, or more likely regeneration from previously existing European stands. The major difficulty with this explanation is the presence of the SKDH-1c allele in hypothetical European trees.

This second suggestion appears most likely. Morphologically, the stand appears to consist of a mix of the two parents rather than hybrids, an interpretation that became more certain after comparison with other putative hybrid stands. The molecular data is largely consistent with this hypothesis, and the detection of the two regenerants increases confidence in these methods. The only inconsistency is the presence of SKDH-1c (the 'Japanese' allele) in some of those trees of hypothetical pure European origin. This is quite plausible, given that these European trees are not hypothesised to be of the same origin as the trees on Kennel bank. In fact, this allele has been found at moderate frequencies in some stands of European trees before (Hacker and Bergmann 1991). It is possible that this allele is distinct from but migrates closely with the Japanese allele- some evidence of this was found on careful examination of the gels, though it was never possible to score it objectively as distinct. Thus, if this allowance is made, the results are completely consistent with the hypothesis that this 1912 stand consists of pure parental species (see Table 4-2).

If it is the case that, apart from the two obvious regenerants, all the 1912 individuals are pure parental types, it is possible to subdivide the stand into two subpopulations, Japanese and European, and examine them separately. Table 4-4 shows, for the three isozyme loci (the organellar loci are fixed), the results of a Hardy Weinberg test in each sub-population, and tests for genic and genotypic differentiation from the corresponding parental population (i.e. the European trees are tested against the Kennel Bank trees, and the Japanese against those on the Avenue). In this table, the trees have been classified as European or Japanese if they have been scored for either organelle. Finally, Table 4-5 shows the multi-locus nuclear disequilibria for each of these two populations. Note that only two disequilibria- a different one for each sub-population- have been calculated. All others are undefined, as variation does not exist in that population for one of the loci.

Table 4-4: The 1912 stand split into sub-populations based on cytotype, p-values for exact tests on H-W equilibrium and genic and genotypic differentiation from their putative parents

Locus	European (by cytotype)			Japanese (by cytotype)		
	H-W Test	Genic diff.	Genotypic diff.	H-W Test	Genic diff.	Genotypic diff.
GOT-1	n/a	n/a	n/a	.004*** (Homozygote excess)	0.58 n.s	0.19 n.s
SKDH-1	.23 n.s	.012***	0.23 n.s	n/a	n/a	n/a
PGM-1	.22 n.s	1.0 n.s	0.82 n.s	1 n.s	.68 n.s	0.88 n.s

Notes: Hardy-Weinberg tests performed by the probability-test option of Genepop (Raymond 1997), and exact tests for genic differentiation performed by the population differentiation option.

Table 4-5: Composite genotypic disequilibria for the two sub-populations of the 1912 stand defined by cytotype

Disequilibria	*Sub-Population*	*Absolute Δ*	*Normalised Δ*	*Significance*
Δ_{SP}	European	0.095	0.65	n.s
Δ_{GP}	Japanese	-0.0086	-0.17	n.s

Notes: These are the only two disequilibria defined for these two sub-populations- they have no cytoplasmic variation (by definition). SKDH is fixed in the Japanese sub-population, and GOT is fixed in the European sub-population, so disequilibria involving these cannot be calculated either. Again, significance is based not on D, but an exact test on the genotypic contingency table.

Two interesting results arise from this analysis. First, the European sub-population within the 1912 stand is seen to be significantly different from the Kennel bank population in terms of SKDH-1 (Table 4-4). This supports the hypothesis that the European trees in the 1912 stand might arise from regeneration from another stand. Secondly, the Japanese trees are not at H-W equilibrium with respect to their GOT-1 locus- there is a significant homozygote excess (Table 4-4). This suggests a high degree of inbreeding or assortative mating in the production of this generation. It is formally possible that whatever factor prevented hybrid formation in this year also caused selfing, though no such factor is suggested.

Finally, it can be seen from Table 4-4 that the only associations that are not significant are those involving PGM-1. This suggested a further test: the comparison of the two subpopulations with each other on a *per locus* basis. The aim of this was to ascertain that each locus would be informative in other stands. SKDH-1 and GOT-1 are indeed highly significantly different between the two subpopulations (this is a trivial and obvious result). However, PGM-1 is not significantly different. This suggested that the time and cost involved

in utilising an extra isozyme system would not be justified by the limited extra information, and its use was discontinued.

4.7 Discussion(2): The Use Of Molecular Markers And Cytonuclear Disequilibria

The methods used were adequate in this case to elucidate the nature of the stand. Some caveats must be added to this. Firstly, the organellar blots, particularly those using mitochondrial probes, were time consuming, expensive, and rather variable in intensity and background. PCR based methods, where possible, are recommended for similar work in future. Secondly, the isozyme analysis, while simple and successful in this case, present much additional work in that if DNA work is also to be carried out, separate collections must be carried out, necessitating marking and mapping of the trees. Further, it was somewhat disappointing that even the SKDH-1 system could not be relied upon to provide a species specific marker. With no indication as to which other European trees may be involved in the production of later generations of hybrids, and a clear indication that the European trees are heterogeneous in this respect, it is not possible to determine the frequency of this SKDH-1c allele in the European parents of any given hybrid stand. Insofar as the main aim of this work was to determine the nature of the stand, the organellar results alone provided the most information, the nuclear and cytonuclear data providing just corroborative details. As far as testing the suitability of these markers for further work, the results suggest the need for more nuclear markers (e.g. RAPDs). This was not a serious problem in this early (in fact, parental) 'hybrid' stand, but the lack of species-specific nuclear markers could limit the power of tests in later generations.

As mentioned, the use of disequilibria as applied to this data is limited. As pointed out by Asmussen, Arnold and Avise (1987), any particular set of values could be interpreted in many ways. For example, one cannot infer directly from the normalised cytoplasmic disequilibrium of 1 that no hybrids have been formed. A value of 1 would also occur in a mixed stand with unidirectional mating (for example, the inclusion of the two regenerants, both with the same cytoplasm, does not affect this value). Thus ,the disequilibria parameters, as treated here, have not been instructive in forming the hypothesis- it was simply easier to infer the nature of this stand from the genotypic distribution. Neither have they been used to test the hypotheses. However, it is worth asking if the disequilibria alone *could* allow one to draw the same conclusions. They can serve as a useful shorthand for presenting the associations that exist in this population (i.e. as descriptive statistics). Explicitly, the interpretation of D_{CM} is obvious-

there is as large an association between these cytoplasmic markers as there could be given their frequencies. Likewise, the cytonuclear disequilibria are instructive. European haplotypes are maximally associated with the 'European' nuclear alleles of GOT-1 and SKDH-1. In terms of the genotypic cytonuclear disequilibria, the disequilibria involving the heterozygotes (e.g. D_{CSs}) were negative for GOT-1 and positive for SKDH-1. This represents the fact that heterozygotes of SKDH-1 are found only in trees with European cytoplasm, whereas heterozygotes of GOT-1 are found only in trees of Japanese cytoplasm. A simple explanation for these results is assortative, directional mating in a young stand (Asmussen, Arnold and Avise 1987). This stand represents an extreme example of just such a situation. Finally, the composite nuclear disequilibria are positive and significant in two of the three pairs, indicating association in this stand between alleles of the same species.

Notable by its absence is any treatment of the statistical properties of the parameters , which could be considered a shortcoming. A'correct' way to approach the problem would be to test, individually, each of the disequilibria parameters against increasingly complex hypotheses (Asmussen and Basten 1994). A justification and discussion of why this has been avoided is given in Chapters 5 and 7. Here, it will suffice to say that, unless the population under consideration can be considered to be relatively old and stable, cytonuclear disequilibria appear to be limited as inferential statistics. Furthermore, even the most rigorous treatments place boundaries on the allele frequencies and sample sizes which can be justifiably tested using the parametric approach. For example, computer simulations suggest that even with a sample size of 100, allele frequencies of between 0.3 and 0.8 would be needed before parametric testing would be justifiable (Asmussen and Basten 1994), which is clearly not the case in this larch data set. Given there are currently no statistical treatments of these parameters which hold for this data, it makes sense to use them descriptively but rely on more accurate (though less informative) tests for associations between alleles/genotypes.

4.8 Conclusion

The stand of larch designated as a 1912 plantation of F_1 larch consists of a mixture of the two parents (approximately 3:1 in favour of Japanese), with evidence of hybridisation only found in the regenerants. It is suggested that this situation arose from seed being planted from the Japanese parents on the Avenue in a year where little or no hybrid formation occurred, and that the European trees are regenerants from a previous plantation of different stock than the Kennel bank trees. This unexpected result highlights the uncertainty implicit in analysing plantations when the records are incomplete and often anecdotal. This is not a criticism of

silvicultural practices, as the forester's aim is not to produce stands of trees that are open to easy scientific analysis. However, it does mean that the designation of hybrid status on the estate at this time (unsurprisingly as the hybrid had only just been discovered) is extremely loose, including 'mixed' parental stands. The Japanese trees did not appear to be random-mating, and limitations of the marker systems and use of cytonuclear parameters are suggested.

4.9 Further Work

The results reported were unexpected and thus interesting from an estate forestry viewpoint. However, they necessitated a re-evaluation of the further work required. Firstly, it was not necessary to analyse further, using either molecular or morphological methods, this stand- the conclusions drawn above are seen as robust. Secondly, the 'F_2' planted from this, on which initial collection work had already begun, was no longer of special interest. Though this could (and appears to) contain hybrids, its origin is no more certain than any other stand. This meant that the initial aim of examining the hybrids through three distinct generations was untenable. Thus, new criteria were drawn up for the selection of the next study site. These were that a large proportion of hybrids (on morphological grounds) should be present and that significant regeneration should have occurred. One site was found, where young trees appeared on an area between a plantation of apparent Japanese trees and a small stand of apparent European trees. Work began on this- isozymes confirmed the hybrid nature of some of these- until it was discovered that this was regeneration from a plantation of 'mixed larch' of completely unknown origin. Two more sites were found, planted in the 1930's, which met the criteria and are the subjects of investigations reported in chapters 5 and 6.

5. Composition Of Two Stands Of Hybrid Larch Of Unknown Origin

5.1 Introduction

Stands of hybrid larch have been extensively planted on the Atholl estate since the turn of the century. However, the exact nature of these plantations is unknown. As part of a project to follow the genetic effects of hybridisation in the early generations, it was discovered that a putative stand of F_1 larch, the supposed progenitor of two subsequent generations, was actually a 'mixed parental' stand. As a result, two later plantations, of unknown origin, were analysed. Both of these (adjacent) stands are fenced off to encourage regeneration, allowing the analysis of a generation of known parentage.

5.2 Aims

In order to unravel the planting history of hybrid larch on the estate, molecular markers were used to determine the nature of two 60yr old stands. The markers had already been successfully applied to investigate stands elsewhere on the estate. The determination of the molecular makeup of these individuals was to lead to a hypothesis regarding the parentage of these stands.

5.3 Materials And Methods

The two stands (Figure 5-1), located about 3 kilometres from the castle, were planted in 1933 and 1936. Approximately 50 trees from each stand were sampled, both in the winter (for isozyme analysis) and summer (for DNA analysis) (Figure 5-2).

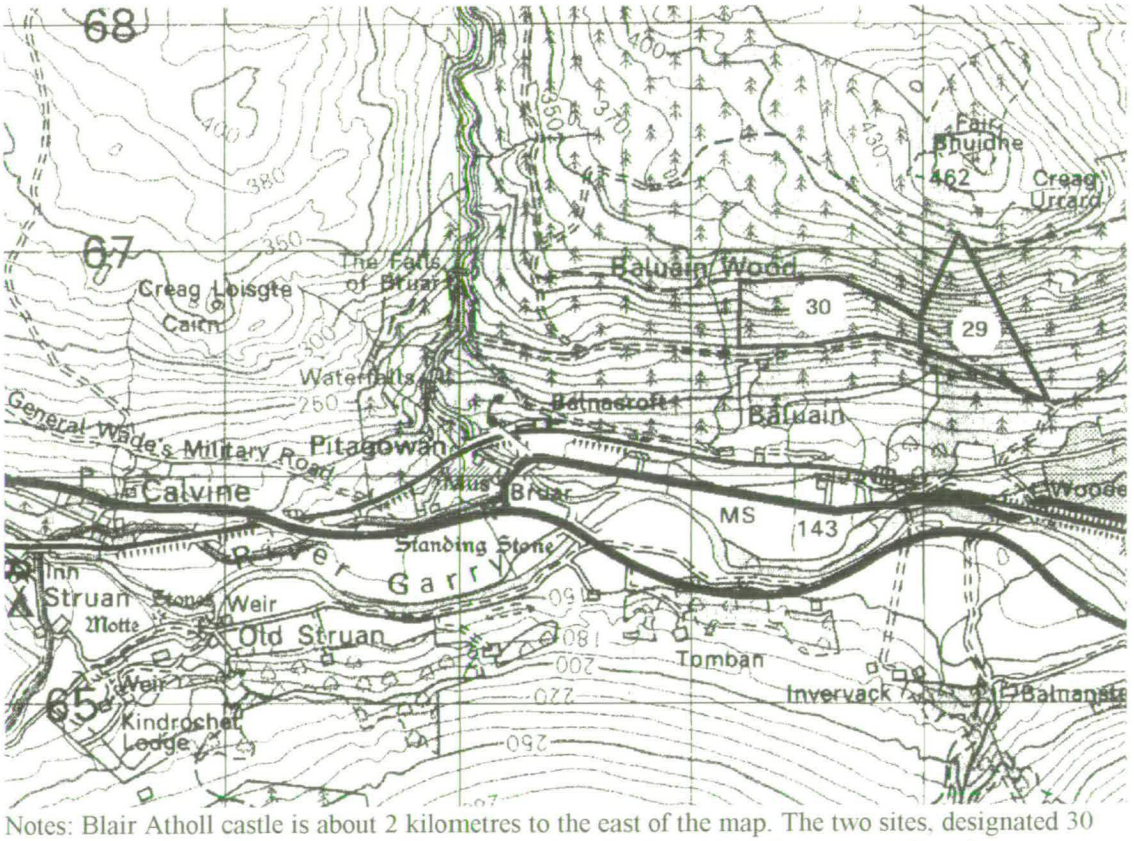
Detailed protocols for the following methods are given in appendix A.

DNA was extracted using the Phytopure kit. 0.1g of frozen tissue was used (p.177). This kit uses an SDS lysis stage followed by incubation with coated silica beads designed to adsorb carbohydrates. This is followed with a dichloromethane extraction, followed by ethanol precipitation. The yield of DNA was found to be lower than a CTAB method. However, it appeared cleaner based on the greater consistency of PCR reactions.

The K140 and ORF25 radioactive probes were used on blots of Bcl and Kpn digested genomic DNA to detect the chloroplast and mitochondrial haplotypes of each individual (p.187). Isozyme analysis by starch gel electrophoresis was carried out for the enzyme systems SKDH and GOT (p.181). RAPD analysis was performed using 10mer primers (p.183), the products of which were run on a poly-acrylamide midi-gel and visualised with ethidium bromide. Only those primers which resulted in consistent, scorable, and informative markers (see chapter 3) are presented.

The isozyme and cytoplasmic data were analysed as described in chapter 4, with disequilibria parameters calculated and presented, but associations tested for by the exact test. There exist added complications due to the inclusion of dominant markers. Associations between loci were tested for in the same way as before, but the use of cytonuclear disequilibria with RAPDs is more complex. First of all, unless assumptions about mating are made, it is possible to measure just one of the four disequilibria parameters, the genotypic disequilibria between the haplotype and the recessive homozygote (i.e. band absent) (Dean and Arnold 1996). The allelic disequilibrium cannot be calculated, as the allele frequencies are not known. Genotypic disequilibria between two RAPD loci or one RAPD and one co-dominant locus have similar limitations. With respect to RAPDs, the method here has been to test for associations between all pairs of loci. However, the only disequilibria parameters involving RAPDs calculated were the genotypic cytonuclear disequilibria (Dean and Arnold 1996).

Figure 5-1: Location of the two stands of hybrid larch (from O.S 1:50 000 map, number 43)



Notes: Blair Atholl castle is about 2 kilometres to the east of the map. The two sites, designated 30 and 29, were planted in 1936 and 1933 respectively. The thick black line surrounding the compartments marks the fence put up to encourage regeneration.

Figure 5-2: Sampling the 1933 hybrid population

5.4 Results

Sample autoradiographs are shown in Figure 5-3 and Figure 5-4. The ORF25/Kpn blot did not reveal any variation in the mitochondrial genome- all individuals scored possessed Japanese mitotype. The K140/Bcl blot demonstrated that both European and Japanese chloroplasts are present in each of the two stands. Examples of the RAPD gels are shown in Figure 5-5 through to Figure 5-8. It can be seen the the RAPD reactions on material from these stands was successful enough to allow calculation of the band frequencies for the markers identified in Chapter 3.

Single-locus results for each population are summarised in Table 5-1, showing allele frequencies for the isozyme loci and haplotype frequencies for the cytoplasmic loci. The two alleles of SKDH-1 found in the Kennel bank trees (i.e. alleles a and b) have been bulked together and termed SKDH-E. SKDH-1c has been termed SKDH-J. Thus, each locus can be treated as having only two alleles. The H-W disequilibrium and the significance of an exact test for H-W disequilibrium are shown for the isozyme loci. The frequency of RAPD bands is also shown in this table but it should be noted that the dominance of these markers means this is not an allele frequency. The final column is the significance level of an exact test for genic differentiation (and genotypic differentiation for the isozyme loci) between the two populations, performed by Genepop. Note that though the RAPD band frequencies do not represent allele frequencies, it is still valid to test the hypothesis that the band frequencies are independent of population. This is the test that the exact test for genic differentiation performs.

Rejection of this hypothesis leads to the conclusion that the two populations are significantly different at a RAPD locus, with regards to the underlying allele frequency or underlying genotypic distribution. This approach is less powerful than tests based on calculating allele frequencies from band frequencies, but allows testing to proceed making no assumptions about H-W equilibrium as is required with other approaches (Lynch and Milligan 1994).

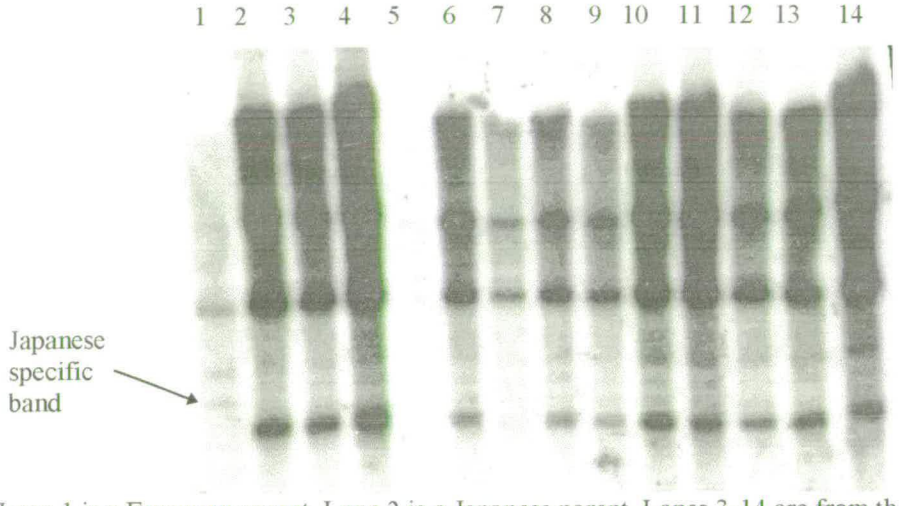
The two populations were found to be significantly different, in terms of gene frequency, at only one locus (OPAL 16b). The 1933 population was at H-W equilibrium for both isozyme loci examined, whereas the 1936 population showed a significant heterozygote excess at both loci. The two populations were shown to be significantly different in terms of SKDH genotype. Both populations contained European chloroplasts at a frequency of approx. 0.8. Finally, a point to note was that no individual contained RAPD marker OPAL16a (Figure 5-7). This marker had been found in all European trees previously examined, and in all the offspring from a controlled cross with the European tree as seed parent (chapter 3). Its absence in this stand requires explanation.

Table 5-2 shows, for each population, the composite nuclear disequilibrium, and the 8 cytonuclear disequilibria associated with the two isozyme loci. As before, associations were tested for with an exact test. A significant association was present between the isozyme loci in the 1936 stand. There were significant allelic associations in both stands between SKDH and the chloroplast loci, and a significant association was found at the genotypic level in the 1936 stand. This could be broken down into a significant positive association between heterozygotes and the European chloroplast, a significant negative association between homozygotes for the Japanese allele and the European chloroplast, but no significant association between European homozygotes and the European chloroplast. No significant cytonuclear associations involving GOT were found.

Table 5-3 presents the results of testing for associations between RAPD loci and chlorotype, and the associated genotypic disequilibria (D_{caa}). Only one significant association is found- between OPK16b and the chloroplast, in the 1933 stand. The negative value indicates an association between the band and the European chloroplast.

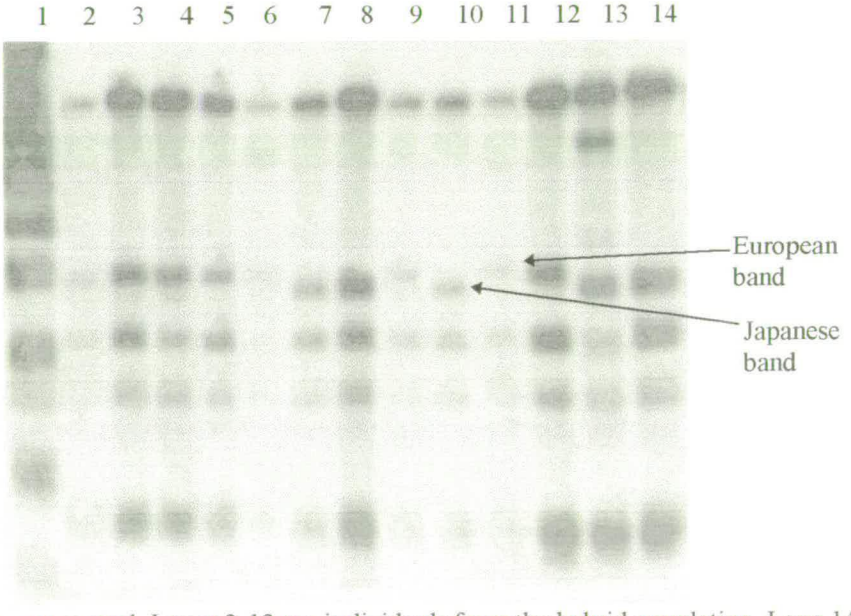
Finally, exact tests were performed for associations between all pairs of RAPD loci. Only one pair of loci in one population showed a significant association (OPAL16 and OPK16b in the 1933 stand), being significant at the 4% level.

Figure 5-3: Autoradiograph of Southern blot of Kpn digested genomic DNA from individuals from 1933 and 1936 stands, probed with ORF25 mitochondrial probe



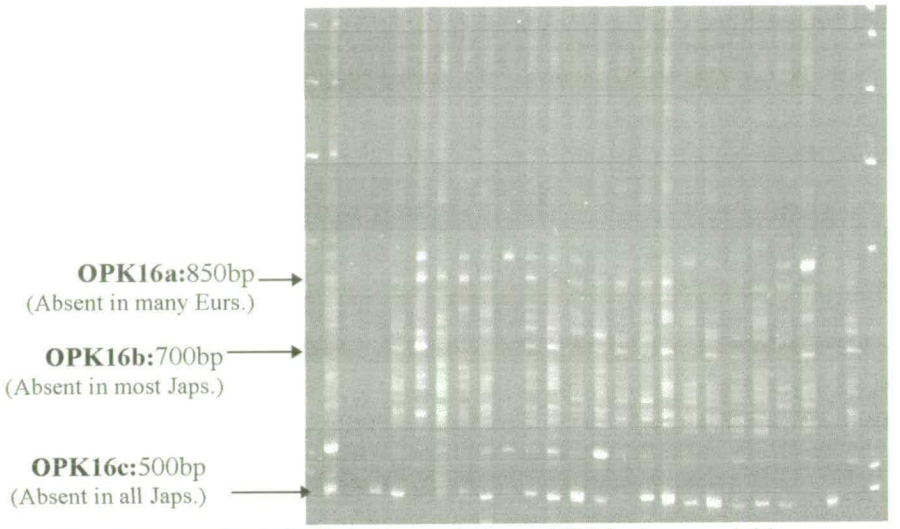
Notes: Lane 1 is a European parent. Lane 2 is a Japanese parent. Lanes 3-14 are from the hybrid populations, and all have the Japanese mitotype.

Figure 5-4: Autoradiograph of Southern blot Of Bcl digested genomic DNA from individuals from 1933 and 1936 stands, probed with K140 chloroplast probe



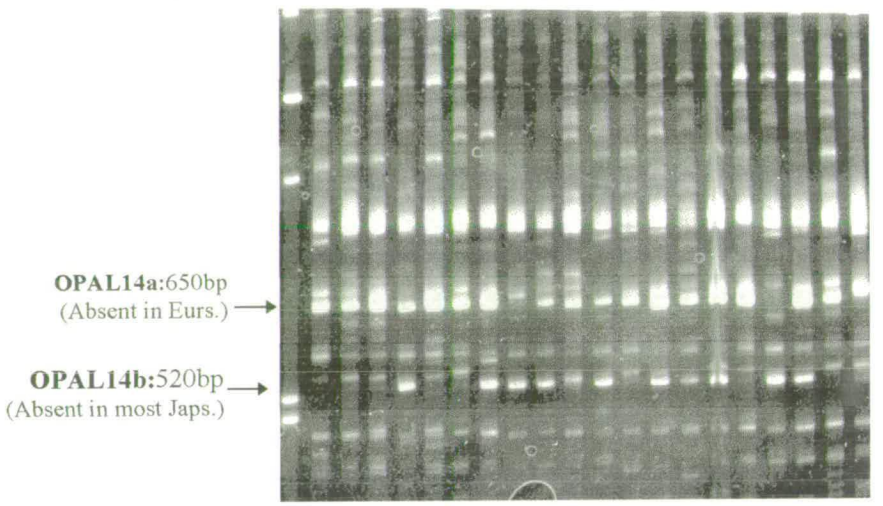
Notes: Lane 1 is a European control. Lanes 2-13 are individuals from the hybrid population. Lane 14 is a Japanese control. Lanes 7,8,10 and 13 can be seen to have the Japanese chlorotype, and the rest have the European chlorotype.

Figure 5-5: RAPD analysis of hybrid populations with OPK16



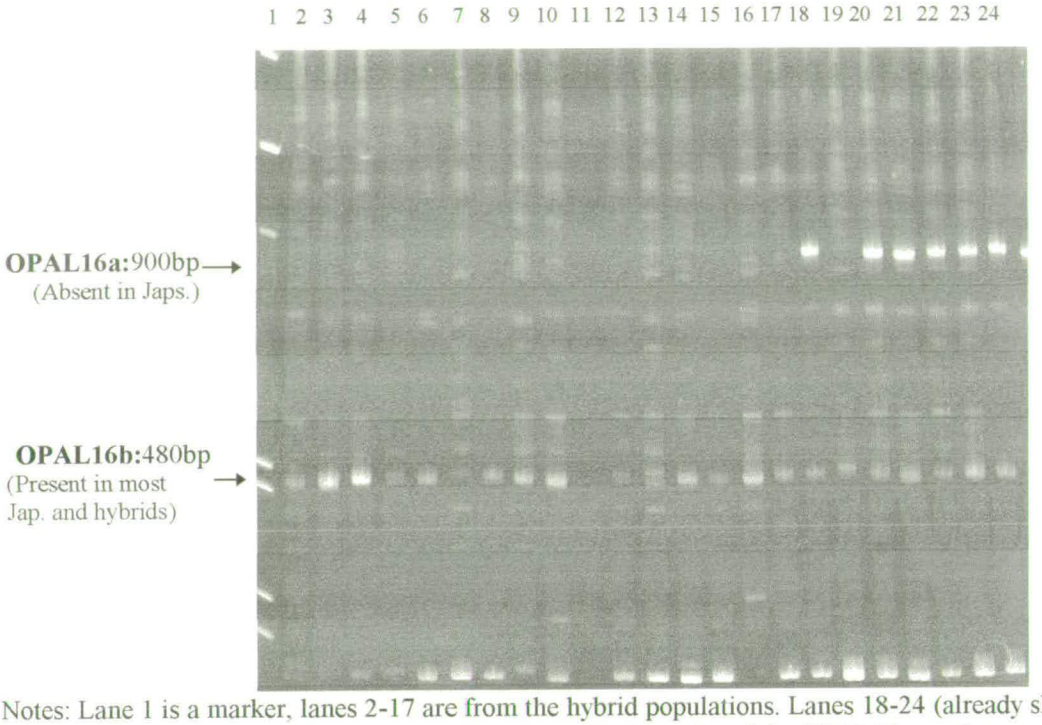
Notes: All samples are from the hybrid populations. All 3 loci are variable.

Figure 5-6: RAPD analysis of hybrid populations with OPAL14



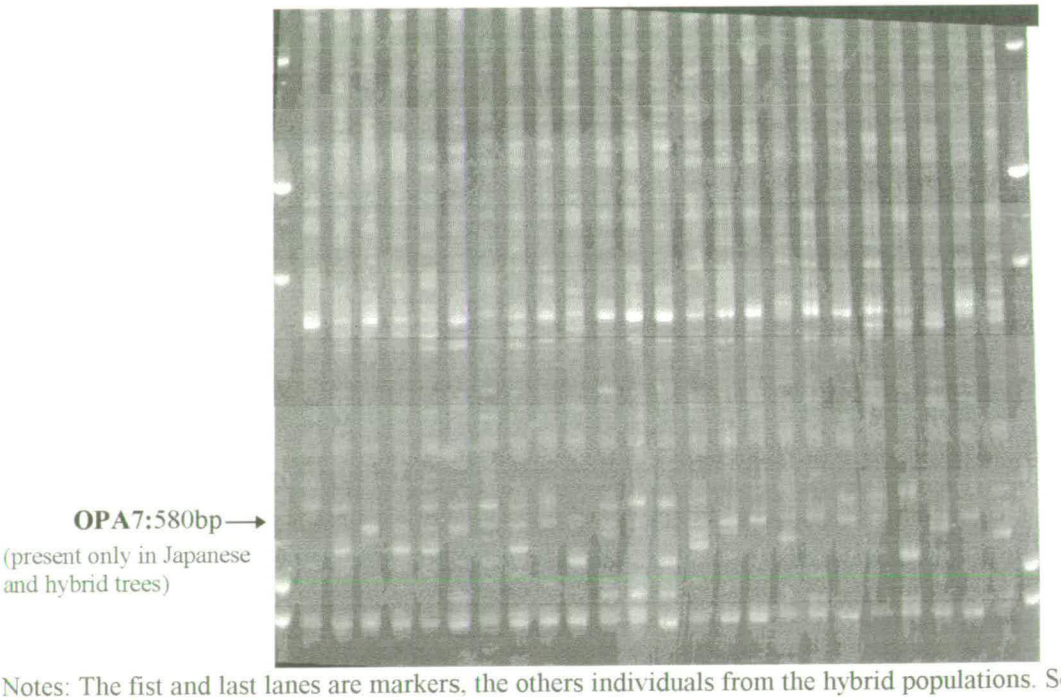
Notes: All lanes apart from the first are samples from the hybrid populations. There is variation at OPAL14b, but only 2 individuals lack the OPAL14a marker.

Figure 5-7: RAPD analysis of hybrid populations with OPAL16



Notes: Lane 1 is a marker, lanes 2-17 are from the hybrid populations. Lanes 18-24 (already shown in Chapter 3) show a controlled cross, and emphasis the intensity of the OPAL16a marker, not present in any of the samples from the hybrid populations. All these samples have the OPAL16b marker.

Figure 5-8: RAPD analysis of hybrid population with OPA7



Notes: The fist and last lanes are markers, the others individuals from the hybrid populations. Some variation can be seen for the 580 bp markers.

Table 5-1: Single locus results in 1933 and 1936 stands, showing allele frequencies and testing for differentiation between the two stands

Locus	Allele/ Band Frequency In 1933 Stand	Allele Frequency In 1936 Stand	p-value for Pop. Differentiation
SKDH-1a,b	*36/94 = 0.38*	*36/76 = 0.47*	Genic:*0.27 n.s*
H.W D	*0.002 n.s*	*-0.112****	Genotypic:*0.00****
GOT-1a	*54/92 = 0.59*	*43/74 = 0.58*	*Genic:1.0 n.s*
H.W D	*0.004 n.s*	*-0.09**	Genotypic:*0.15 n.s*
Chlor. E	*34/43 = 0.79*	*28/36 = 0.78*	*0.85 n.s*
OPAL 14a	*41/45 = 0.91*	*36/41 = 0.88*	*0.61 n.s*
OPAL 14b	*28/44 = 0.64*	*31/41 = 0.76*	*0.09 n.s*
OPAL 16a	*0/44 = 0*	*0/41 = 0*	-
OPAL 16b	*25/37 = 0.68*	*34/36 = 0.94*	*0.00 ****
OPK 16a	*35/41 = 0.85*	*37/41 = 0.90*	*0.47 n.s*
OPK 16b	*30/41 = 0.73*	*26/41 = 0.63*	*0.24 n.s*
OPK 16c	*40/42 = 0.95*	*41/41 = 1.00*	*0.12 n.s*
OPAL 7	*28/42 = 0.66*	*24/31 = 0.77*	*0.18 n.s*

Notes: Frequencies are shown with the denominators so that the sample size for each locus can be read off the table. Differences between populations and HW disequilibrium are tested for with the exact tests of Genepop.

Table 5-2: Tests of isozyme/chlorotype or isozyme/isozyme associations and related disequilibria parameters (D)

D	Association Tested	*1933 Stand*			*1936 Stand*		
		Abs. D.	*Norm D*	*Sig.*	*Abs. D.*	*Norm. D.*	*Sig.*
Δ_{SG}	Nuclear, composite	0.04	0.22	n.s	0.03	0.32	**
D_{CS}	Cytonuclear, allelic	0.05	0.59	**	0.07	0.57	**
D_{CSS}	Genotypic	0.04	1.00	n.s	-0.02	-0.31	n.s
D_{CSs}	Genotypic	0.03	0.28	n.s	0.17	0.83	***
D_{Css}	Genotypic	-0.07	-0.47	n.s	-0.15	-1.00	***
	Overall Gen			n.s			***
D_{CG}	Cytonuclear, allelic	0.03	0.02	n.s	0.01	0.05	n.s
D_{CGG}	Genotypic	0.05	0.68	n.s	0.00	0	n.s
D_{CGg}	Genotypic	-0.05	-0.39	n.s	0.02	0.08	n.s
D_{Cgg}	Genotypic	-0.01	-0.03	n.s	-0.02	-0.30	n.s
	Overall Gen			n.s			n.s

Notes: Associations (allelic or genotypic) were tested for with genepop. Disequilibria parameters (absolute and normalised) were calculated as in chapter 4. The positive alleles were arbitrarily chosen as the 'European' allele as in chapter 4 (i.e. SKDH-1a,b, GOT-1a and the European Chloroplast). Thus, D_{CSs} is the genotypic disequilibrium between the European chloroplast and the heterozygote SKDH (see Chapter 4)

Table 5-3: Cytonuclear disequilibria and exact tests for association between RAPD Loci and chlorotype

D_{Caa} for	*1933 Stand*			*1936 Stand*		
	Abs. D.	*Norm D*	*Sig.*	*Abs. D.*	*Norm. D.*	*Sig.*
OPAL14a	0.02	1.00	n.s	-0.03	-0.27	n.s
OPAL14b	-0.05	-0.32	n.s	-0.01	-0.09	n.s
OPAL16a	n/a	n/a	n/a	n/a	n/a	n/a
OPAL16b	0.04	0.57	n.s	0.01	1.00	n.s
OPK16a	-0.05	-0.48	**	-0.04	-0.39	n.s
OPK16b	0.02	0.23	n.s	0.04	0.54	n.s
OPK16c	0.01	1.00	n.s	n/a	n/a	n/a
OPA7	0.03	0.52	n.s	0.04	0.33	n.s

Notes: For each RAPD locus and each stand, the one cytonuclear disequilibrium has been calculated, as has a normalised value. A positive value represents an association between the band and the Japanese chloroplast. The significance of any association is tested in the third column of each population.

5.5 Discussion

All individuals sampled appeared to descend from Japanese seed parents, on the basis that only Japanese mitochondria were found. The majority (80%) of the chloroplast pool in both populations came from European trees. Thus, there is a clear hybrid nature to these stands, which appear to be different in composition from one another. Discussion of these details will lead to a hypothesis regarding the nature of the stands. It will be noted that the approach used in many cases is non-standard. This is because the conventional genetic parameters and statistics are framed in terms that are unsuited to the situation in these stands. For example, statistics such as Fst are avoided, not because they cannot be used to describe differences between any two populations, but because there is no further application of an Fst in this study. In the terminology of Weir (Weir 1990), the samples here- the stands and the parental species- constitute 'fixed populations' rather than 'random populations'.

5.5.1 Differentiation Between The Two Stands

The aim of this study was to generate a plausible hypothesis regarding the nature of these two stands. The first step towards this was to test the hypothesis that the two stands could be regarded as being samples of the same population. If so, this would increase the statistical significance of any tests regarding the composition of this population. A common approach to similar questions would be the use of Fst statistics (Wright 1951). There are two reasons I have avoided this. Firstly, the use of Fst statistics for organellar loci and especially RAPDs (Lynch and Milligan 1994) requires modification to the usual approach. Secondly, the assumptions behind the use of Fst are based on the populations sharing a common ancestry

(Weir 1990) and that processes such as random drift and migration have led to the distributions seen. While authors often use Fst as a general test for population differentiation, and there is no mathematical reason why this cannot be done, it seems better to restrict this approach to cases where estimates of migration or time of isolation are required.

A more general approach is simply to test the hypothesis that the distribution of markers is independent of population. The hypothesis is, on this data set, most naturally broken down into four hypotheses. First, that the distribution of RAPD markers is independent of population. Second, that the genic distribution of isozyme loci is independent of population. Third, that the genotypic distribution of isozyme loci is independent. Finally, that the genic distribution of chloroplast markers is independent of population.

For the RAPD loci, we see there are 7 loci and therefore 7 tests of the hypothesis of independence. While not strictly independent tests, they can be regarded as approximately so (Weir 1990). For the overall test to be significant at 5%, we require that each locus is tested at 1-(1-0.05)^(1/7). This 'Bonferroni procedure' gives a value of 0.7%. If independence of any RAPD locus can be rejected at this level, independence of RAPD loci can be rejected at 5%. Thus, OPAL 16b, the independence of which was rejected at below the 0.5% level (Table 5-1), leads to rejection of the global hypothesis.

An alternative, and one might expect more powerful, approach to testing whether two populations are different on the basis of their RAPD markers is the use of AMOVA (Excoffier *et al.* 1992). AMOVA (ANalysis of MOlecular VAriance) tests for differences between populations based on their multi-locus genotypes/phenotypes. A distance matrix is constructed from the RAPD scores. In this case, a squared Euclidean distance was used, such that the distance between two individuals is the number of differences over the number of bands scored for both individuals (so that missing data is acceptable). A computer program for running AMOVA then tests whether there are significant differences between populations using a re-sampling approach, and calculates a PhiST genetic distance, analogous to Fst for single-locus data. In this case, the two parental populations and the two hybrid stands were tested. The results are shown in Table 5-4. Numbers above the diagonal are the significance values for the test of the hypothesis that the two populations have a genetic distance of zero, numbers below are the genetic distances themselves. Unsurprisingly, the two parental stands were significantly different from each other - the markers were chosen to be different in each parent, so this is a trivial result. The genetic distance between these two is an irrelevance, as the markers were not randomly selected. Again trivially, the hybrid stands were highly

significantly different from the parents. As far as this section is concerned, the important result is that the two stands come out to be significantly different at their RAPD loci at below the 2% level. Thus, both approaches- testing each RAPD locus individually and AMOVA suggest these two stands do not derive from the same stock.

Table 5-4: Results of AMOVA on the RAPD data- genetic distances and significance

	Japanese Trees	*European Parents*	*1933 Stand*	*1936 Stand*
Japanese Parents	-	0.0000	0.0000	0.0000
European Parents	0.87	-	0.0000	0.0000
1933 Stand	0.37	0.57	-	0.0180
1936 Stand	0.47	0.64	0.03	-

Notes: Figures below the diagonal are the values of PhiST (similar to Fst) between the two populations. Above the diagonal is the significance level for a test that this is significantly greater than 0.

For the isozyme loci, we cannot reject the hypothesis of independence for the genic distribution at either loci, so clearly the global test will not reject it either. However, two tests for genotypic differentiation have been performed, leading to (by the same argument as above) the tests being performed at the 2.5% level for overall rejection of the hypothesis that genotypic distribution of isozyme loci is independent of population. Table 5-1 shows that SKDH leads to rejection of this hypothesis at 2.5%, such that the global test rejects the hypothesis at 5%.

Finally, Table 5-1 shows that the hypothesis that the populations represent samples of the same population with regard to chloroplast frequency can clearly not be rejected.

Thus, the two populations must be treated separately and regarded as having significantly different parentages. It is pointed out that the most important difference is not in terms of allele frequency, but genotypic distribution at one of the isozyme loci, and the multi-locus RAPD profile.

5.5.2 A Maternally Inherited RAPD Marker

It was seen in chapter 3 that the RAPD markers did not all behave in a Mendelian fashion. The offspring of a controlled cross did not all contain a band present in one parent as expected. This is not uncommon. While many authors find 'well-behaved' markers (i.e. those that are inherited in a dominant Mendelian fashion) with seeming ease, there are also many studies showing that inheritance patterns can be unpredictable (Hallden *et al.* 1996). There are two main explanations for this. One is competition and interaction with other loci. This has been experimentally demonstrated (Hallden *et al.* 1996), and amounts to saying that presence/absence of a band may be affected by unscored variable loci. Another often cited problem with RAPDs is experimental irreproducibility, such that a band may spuriously

amplify or, more likely, fail to amplify. This is clearly a problem in this data set, and it is suggested that RAPDs may be more problematic in some species than others depending on the quantity and nature of contaminants in DNA extractions. While every effort was made to minimise this problem -bands were not, for example, scored as absent unless the rest of the RAPD profile was of a similar intensity and pattern to other individuals- extreme caution is advisable when dealing with the RAPD data. In essence, the scored band frequency is only an estimate of the sample band frequency, in contrast to the allele frequency as scored by iszoymes which can be regarded as identical to the sample's allele frequency, except for the presence of null alleles. This is another reason why no attempt to calculate significance using a parametric approach has been made- the sampling properties of RAPD frequencies do not include the additional source of error introduced by this uncertainty.

Given this, and that the nature of the stand is completely unknown, it is futile to attempt reconstruction of the inheritance of a given RAPD marker from the data. However, one interesting result did arise based on the RAPD results alone. Table 5-1 shows that marker OPAL16a was absent from all individuals. This was not failure of the primer to amplify in these individuals- the rest of the profile was consistent (Figure 5-7) and reproducible. This marker was found in all European trees on Kennel bank, and the European parent of the controlled cross, and all its offspring, leading to the reasonable hypothesis that this was a dominant, European specific marker (chapter 3).

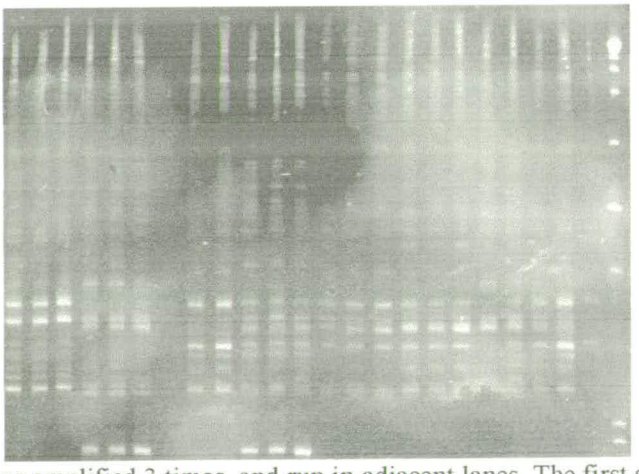
As 80% of the chloroplast pool was European, and other European markers (such as SKDH-1,2) were found at high frequency, the absence of OPAL 16a cannot be explained by sampling effects. The most plausible explanation is that it is a European specific maternal marker, such as a mitochondrial primer. This explains the observation that it is seen in all crosses with a European seed parent, such as the controlled cross, but not in any where maternal ancestry is expected to be entirely Japanese. No European mitochondria were found in either stand, lending weight to this hypothesis.

Studies of RAPDs generally assume nuclear inheritance, though the possibility that some markers may be organellar is often mentioned. In studies where organellar inheritance has been looked for, it is found in varying frequencies- a study in Douglas fir, for example, found that 16% of RAPDs were mitochondrial in origin (Aagaard *et al.* 1998). Assuming that this is the case, another interesting observation can be made. As can be seen in lanes 18-24 of Figure 5-7, the intensity of this marker in European trees and offspring is far higher than any other band. In fact, this caused concern when originally scored as such bands often appear to be

contamination. However, there are good reasons for expecting organellar RAPD markers to be more intense than nuclear markers. There are many copies of each mitochondrial genome for each copy of the nuclear genome. The most basic models of PCR amplification do not expect product intensities to be highly dependent on initial template concentration. Certainly, most anecdotal evidence suggests that the concentration of template DNA is more important with regards to concentration of inhibitors than the DNA itself. Figure 5-9 shows the results of various samples being diluted a hundredfold and amplified (with a different primer). Where amplification is successful, the profile does not appear markedly dependent on concentration. However, if competition plays an important role in the generation of RAPD profiles, *relative* concentrations of DNA template may be important. Specifically, in this case, the greater concentration of organellar template will give the organellar product a 'head-start' over nuclear priming sites for the same primer, and result in a more intense band. A study of RAPD variation in sugar-beet (Lorenz *et al.* 1994) demonstrated that, in this case at least, the average intensity of mitochondrial bands was greater than average band intensity. It could even be argued that this provides evidence for the important role of competition in RAPD amplifications.

In conclusion, it is suggested that this marker represents a segment of the European mitochondrion, and its absence in these stands is consistent with the hypothesis that there is no European maternal ancestry in either population.

Figure 5-9: The limited effect of DNA concentration on RAPD amplification



Notes: Each sample was amplified 3 times, and run in adjacent lanes. The first sample in each repeat used neat DNA template, the second was diluted 1:10 and the 3rd 1:100. Apart from the failure of one of the neat samples to amplify, the constancy across different dilutions of template can be seen.

5.5.3 *Interpretation Of RAPD Data*

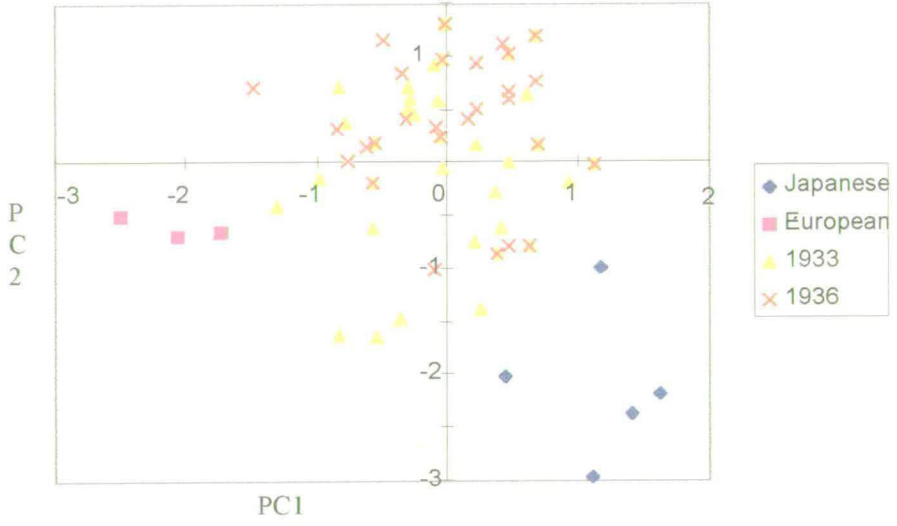
While the many limitations of RAPDs were known before the start of the study, it was hoped that a set of species diagnostic markers could be found that would aid analysis of the stands. As shown in chapter 3, four such markers were found. One of these turned out to be maternally inherited. In a sense, this marker is probably the most important of those used. However, its absence in the plantations means that it provides no further information. Of the other species diagnostic markers (OPAL14a, OPK16c and OPA7a), one (the European specific OPK16c) demonstrated non-Mendelian inheritance, even when only a small sampled of controlled cross offspring were used. Thus, of the 8 markers, we have 2 species diagnostic markers (both specific to the Japanese parent) which appear to be dominant Mendelian markers (OPAL14a and OPA7a). Emphasis is being put on such markers because in almost all studies, it is these markers and these alone which are included in an analysis (Dawson *et al.* 1996; Perron and Bousqet 1997). Thus, the usual analytical methods are not necessarily suited to this study and other approaches will be taken.

The problem of how to analyse dominant RAPD data has lead to a variety of methods becoming commonplace. It has already been shown how AMOVA can be used to demonstrate inter-population differentiation. The next section (5.4.4) shows how they could be used (albeit in a very approximate, error prone fashion) to estimate the allelic composition of a population in terms of parental populations. Following that, 5.4.5 shows how they can be used in the construction of a character or hybrid index. However, one widespread approach, which is included both for the sake of completeness and as an indication of how useful RAPDs might be for such studies, utilises the technique of Principal Component Analysis (PCA) on the binary RAPD data (Manly 1986). This multivariate technique simply reduces the data into a lower number of variables that can describe most of the variation present in a sample. Presenting the results of this graphically can lead to a clearer understanding of the RAPD results. The other advantage of this technique is that no genetic assumptions are made. It can be also used to test for differences between groups, by performing an ANOVA on the grouped principal components. It is useful because, like AMOVA, it is a multi-locus approach and not just based on band frequencies.

Figure 5-10 shows the results of a PCA performed on the two parental groups (using SPSS 6) and the two 1930s stands simultaneously. Marker OPAL16a (maternally inherited) was not included in the analysis, for it would inflate the distance between the two parents but obscure the difference between the hybrids and the Japanese parents. The problem- quite large in this

data set- of missing values has been tackled by replacing each missing value the average value for the whole data set. Note that the parental groups appear particularly small because there are several identical phenotypes within each parent. Three components were extracted, and accounted for 70% of the variation in the sample. Component 1 shows clearly the differences between the parents, with the hybrid stands falling between them. Component 2 also shows some differentiation- this time between the two hybrid stands. A t-test performed on the second component between the two hybrid stands is significant at the 5% level. The third component is not shown- there was no variation between groups for this variable.

Figure 5-10: PCA on RAPD data for the two parental stands and the two hybrid stands



Notes: The Parental groups consist of 11 and 12 samples respectively. However, several in each group have the same phenotype and are therefore displayed only once. A clear difference in the first component (x-axis) can be seen between groups. A small, but significant, difference is found between the two hybrid stands in the second component (y-axis).

This PCA is a good graphical representation of the RAPD data. Note that despite the non-Mendelian inheritance of many of the markers, which makes treatment of genetic parameters awkward, the hybrids clearly cluster between the parents. Furthermore, it has picked out the difference between the two hybrid stands. Note that owing to the dominance of RAPDs, this difference could be allelic or genotypic. This approach could provide a useful way of characterising mixed stands and even individuals.

5.5.4 Using Allele Frequencies To Quantify The Composition Of A Population

Suppose that we want to obtain a very rough guide to the composition of a stand in terms of the two parent species. In other words, we want to estimate the proportion of genes in the

stand that derived from each species. This is clearly a trivial problem for co-dominant, species-specific, diagnostic markers. The overall composition could be described as just the average frequency of alleles specific to one species. Also, this is an error-free estimate of the proportion of alleles from one species in that sample. If , for example, 80% of the alleles derive are diagnostic of one species, then *exactly* 80% of those alleles derived from that species. However, the situation is slightly complicated by the introduction of non-species specific markers. In fact, in a multi-locus model created to estimate the frequency of various classes of individuals in a population, non-species-specific markers were discounted from the analysis due to the added complications (Nason and Ellstrand 1993).

A simple approach to the problem is outlined below for the case of two species (a and b) and two alleles for each locus.

First, estimates of the allele frequencies in each of the two species are made. Call these P_{a1}, P_{a2},...P_{an} for n loci in species a, and P_{b1}, P_{b2},... P_{bn} for n loci in species b. Assume that the study population is composed of a proportion m of genes from species a, and (1-m) from species b. The allele frequency in the study site is denoted by P_{hn} (where the subscript h stands for 'hybrid'). Each locus gives us an estimate of m, given by

$$P_{hn} = m_n P_{an} + (1-m_n) P_{bn}$$

$$\text{so } m_n = (P_{hn} - P_{bn}) / (P_{an} - P_{bn}) \qquad (1)$$

where m_n represents the value of m estimated from that locus.

The least squares estimate is given by the weighted average of these, the weight being the variance of each estimate, if we assume each estimate is independent. This is the same as assuming no linkage disequilibria between the loci, and is the first approximation made.

$$m = \Sigma \sigma^2_n m_n / \Sigma \sigma^2_n \qquad (2)$$

The next approximation is that we take the allele frequencies in the parent species or populations to be known without error. Thus,

$$\text{Var}(m_n) = \text{Var}(P_{hn}) / (P_{an} - P_{bn})^2 \qquad (3)$$

The next approximation is that the variances of the allele frequencies in the study population are the same at each locus. In this case, the weight of each estimate is simply the denominator of (3), or the square of the difference between parental allele frequencies.

$$m = \Sigma w_n m_n / \Sigma w_n$$

$$\text{where } w_n = (P_{an} - P_{bn})^2 \quad (4)$$

By combining this with (1), we get

$$m = \Sigma (P_{hn} - P_{bn})(P_{an} - P_{bn}) / \Sigma (P_{an} - P_{bn})^2 \quad (5)$$

It is readily verified that this is algebraically equivalent to the least-squares approach utilised in the analysis of human hybrid populations (Roberts and Hiorns 1965) in the case of only two populations with diallelic systems, incorporating the restriction that the estimated proportion from each population must sum to 1 (Elston 1971). These authors have a more general approach allowing for more than two parent populations, which has been used in populations of *Aesculus* hybrids (DePamphilis and Wyatt 1990), which reduces to the method outlined above for two populations. These authors also demonstrated that it is not justified to use a minimum chi-squared approach. The approach outlined above is presented because it is more accessible and sufficient for these purposes, and as will be shown below, provides a simple way of incorporating RAPD markers.

It is possible to calculate the standard error of this estimate based on the variation in estimates supplied by different loci. However, this standard error depends on the parental allele frequency being known without error (DePamphilis and Wyatt 1990). This is not the case here. Though it will be shown that this does not bias the estimate of m by much, it will have a greater effect on any estimate of variance and with only two loci providing the individual values, it would not be appropriate to calculate this error.

There are only two such loci here, but applying this method to the data from the 1933 and 1936 stands, using the allele frequencies from the Avenue and Kennel bank as estimates of P_{A1} and P_{A2} gives the results shown in the first row of Table 5-5.

Before progressing onto how the approach is modified to take RAPDs into account, a development of the above approach is suggested. As mentioned, this approach is equivalent, in the two population di-allelic case, to the method of Elston (Elston 1971). One of the assumptions this approach has made is that variances of allele frequencies are the same at each locus. For co-dominant loci, the sampling variance can be given by $(1/2n)(p+D-p^2)$ (Weir 1990), where D is the Hardy-Weinberg disequilibrium at that locus, and this can be included in the weight given in (4) to give the following

$$w_n = (P_{an} - P_{bn})^2(1/2n)(p + D - p^2) \quad (6)$$

This barely affects the proportion, m, calculated above- the negligible effect can be seen by inspecting the second row of Table 5-5. However, it provides a framework for including RAPD loci.

It is possible to construct a similar score based on RAPDs. However, the dominance of these markers yet again causes problems. If any analysis is based on allele frequencies there will always be an assumption - explicit or implicit- about the distribution of the alleles (i.e. that the population is at H-W or a particular disequilibrium). Thus, any approach applied to such markers is bound to be somewhat approximate. The proportion m in the study stand can be estimated by assuming H-W equilibrium, but it is pointed out that this assumption has already been shown to be false for the 1936 stand. RAPD allele frequencies and variances were estimated using asymptotically unbiased estimators of the null allele frequency (Lynch and Milligan 1994). The third and fourth rows of Table 5-5 show the estimated proportion of each stand coming from the Japanese parent based on an assumption of HW equilibrium at all RAPD loci, with and without the assumption of equal variances at each locus. One somewhat arbitrary removal has been made, following suggestions of previous authors (Elston 1971). OPK16c does not appear to be behaving as a species marker. Specifically, it is present in all offspring in the 1936 stand and a surprisingly large proportion of the 1933 stand. For the 1936 stand, this leads to an estimate of '0' for m. Thus, it is hypothesised that some other - possibly non-genetic and certainly non-Mendelian- effect is acting on this marker, and it is removed from this analysis. Whatever the cause, it has been suggested that estimates of m being 0 or less, or 1 or more, should not be included (Elston 1971).

It is not asserted that these are necessarily least squares or unbiased estimates, given the added element of variance in the estimate of sample allele frequencies. However, they are likely to be very close.

The estimates given by the two classes of markers differ, by 4-6% in both cases. Though no confidence interval has been constructed- the assumptions used to construct a standard error are not justified in this case (DePamphilis and Wyatt 1990)- this is not large enough to cause concern. In fact, the differences between individual RAPD loci were far greater than this. Thus, it would appear justified to pool the two classes, as shown in the final two rows of Table 5-5.

Table 5-5: Calculating 'm', the proportion of Japanese contribution to the two hybrid stands using the least squares method (Elston 1971) and variants of this method

Loci used	Assumptions made	m for 1933	m for 1936
Isozyme loci	Equal variances	0.649	0.584
Isozyme loci	Variances based on genotypic distribution and sample size	0.648	0.581
RAPD loci	Variances equal All at H-W Removed OPK16 c	0.607	0.647
RAPD loci	All at H-W Variances based on estimated allele frequencies and sample size Removed OPK 16 c	0.604	0.642
All loci	RAPD loci at H-W Variances all equal	0.617	0.631
All loci	RAPD loci at H-W Variances based on estimated allele frequencies and sample size	0.628	0.602

As can be seen, relaxing the assumption of equal variances does not appear to make much difference when only one class of markers is being examined. However, when the two marker systems are analysed together, the inclusion of the variances does have quite a large effect by skewing the estimate towards that given by the co-dominant systems, which is desirable -the underlying allele frequency of a RAPD marker has a much higher variance than a co-dominant marker.

In conclusion, a restricted case of the methods of Elston (Elston 1971) and Hiorns (Roberts and Hiorns 1965) has been extended to relax the assumption of equal variances at all loci which allows the inclusion of RAPD data using allele frequencies and variances as estimated by the methods of Lynch and Milligan (Lynch and Milligan 1994). The results suggest that despite the demonstrated genotypic differentiation between the two stands, both appear to consist of approximately equal proportions (60%) of nuclear alleles derived from the Japanese parents.

5.5.5 A Character Index

One use of molecular data is the construction of some form of character or hybrid index. This is usually done with species-specific markers (DePamphilis and Wyatt 1990; Perron and Bousqet). A common approach is to score 1 or -1 for every species-specific allele (or band, in the case of RAPDs) depending on which parent it is specific for. Pure parents would have

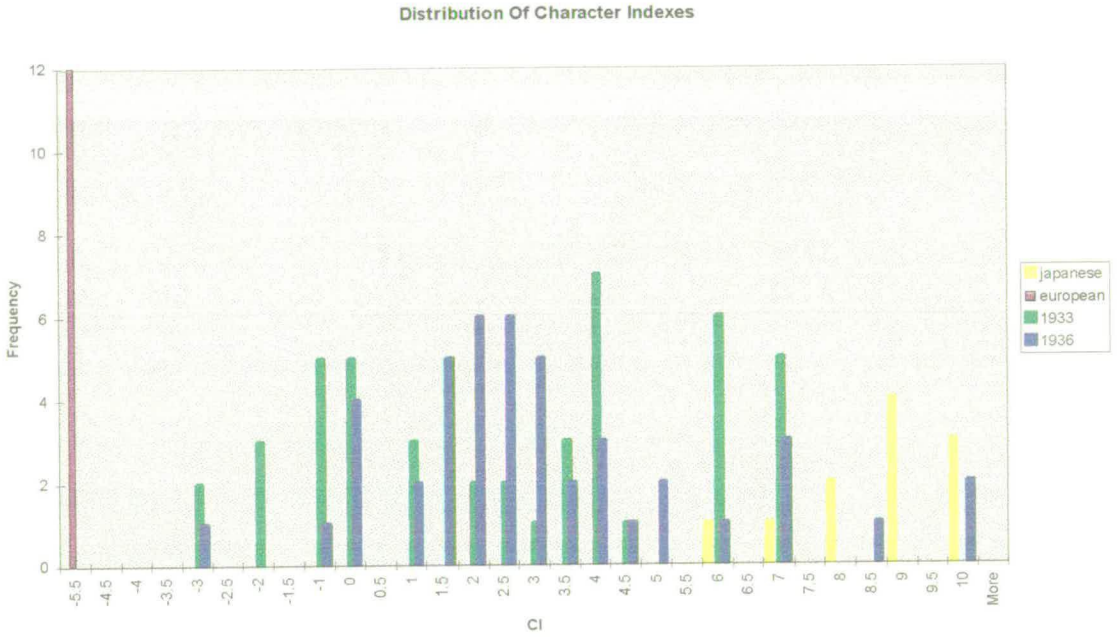
extreme scores, with hybrids scoring intermediate. Markers used can be either species-specific, such as the GOT-1b allele present only in Japanese parents, or diagnostic, such as SKDH-1 which has unique fixed alleles in each parent. Using this approach, SKDH, GOT, and 5 RAPD markers (OK16b, OK16c, OPA7, OPAL14a and OPAL16b), and the chloroplast marker can be used. Missing data, where one cannot afford to eliminate that individual, can be treated by replacing the score for the missing locus by the average score over the other loci. A pure Japanese tree (with all loci scored) would score between 10 (2 SKDH-1c alleles, 2 GOT-1b alleles, 5 RAPD bands (either present or absent) and 6 (the GOT-1b alleles and two of the RAPDs are not fixed). A pure European would score -6. This is clearly a rather arbitrary system and will not have a well-defined distribution. This is partly because of the way missing data has been treated, but more so because of the inclusion of RAPD markers. Only if there were exactly the same number of fixed dominant markers in one species as the other would, one expect, for example F_1 hybrids to score exactly intermediate. On the other extreme, if the only fixed dominant markers were in Japanese trees, the contribution of the RAPD markers to the overall character index would be the same in F_1s as it was in the Japanese parents.

Figure 5-11 shows a multiple histogram showing the distribution of the CI's for each of the parent populations and the two hybrid stands. There is no significant difference in this CI between the two hybrid stands, demonstrating the greater power of AMOVA and PCA. However, as a quick and easy score, it is fairly informative. The two hybrid stands clearly contain a large hybrid character. It appears that there may be more Japanese character than European character to the stands, both with regards to the overall composition and the several individuals with a score at the extreme end of the Japanese range, though this could be deceptive owing to the dominance of the RAPDs discussed above.

A final point is that originally it was intended to construct a character index using all alleles-not just diagnostic ones. However, a problem with this soon becomes apparent, even when using just co-dominant markers. Such a character index can only be constructed if the allele frequencies in the population as a whole are known. This is because the probability that a given (non-specific) allele came from one parent is conditional upon the frequency of that allele in the population. For species-specific or diagnostic markers, this probability is independent of the composition of the population. For example, GOT-1b can be used because a GOT-1b allele has always come from the Japanese parent. The probability that a GOT-1a allele came from the Japanese parent, however, depends on the relative contribution of each

parent to that population. Such a score- based not only on an individual tree but the population from which it comes- could be useful, but is not truly a hybrid index and will not be treated here.

Figure 5-11: Histogram showing distribution of character index scores in the putative parents and two hybrid stands



5.5.6 Disequilibria

The greatest difference between the two stands is not allele frequencies, but disequilibria. One such result has already been mentioned- the 1936 stand has a large deviation from H-W. The same stand also shows a significant association between GOT and SKDH such that the 'European' alleles are associated (Table 5-2). Both stands show cytonuclear disequilibria such that the European SKDH allele is associated with the European chloroplast. In the 1936 stand, there are also significant disequilibria for two of the three genotypic associations (D_{CSs} and D_{Css}) as well as an overall genotypic association. As can be seen from the normalised values, there are significantly fewer (in fact, the minimum possible given the allele frequencies) homozygotes for the Japanese allele with a European chloroplast than chance alone would dictate, and significantly more heterozygotes with the European chloroplast. The interpretation of the 1933 stand, where there is a significant allelic but non-significant genotypic disequilibria is not obvious. This is not likely to be for any genetic or biological reason but due to subtle statistical properties of the exact test.

The 1933 stand exhibited a significant cytonuclear disequilibrium between one RAPD locus and the chloroplast. However, it should be noted that 7 RAPD loci were tested for disequilibrium, all at the 5% level. For the hypothesis that there is significant cytonuclear disequilibrium (between RAPDs and chloroplast) in the stand, the Bonferroni procedure already mentioned necessitates testing each at the 0.7% level for the overall significance to be 5%. This association was *not* significant at this level, so it is not valid to conclude that there is an overall cytonuclear RAPD disequilibrium in either stand.

The same is true of the one pair of RAPD loci to show a significant association when tested individually. In this case, this apparent association is even more insignificant globally- there are more pairs of loci to test and the association was only barely significant at the 5% level in the first place.

In conclusion, there are significant disequilibria involving the isozyme and chloroplast loci, which differ in nature between the two stands. However, there is no case for disequilibrium involving RAPDs. Such disequilibria may well exist at the allelic level, but the power of RAPDs is such that either far greater associations or sample sizes would be needed to detect them.

5.5.7 Nature Of The Two Stands

Having decided that each stand must be treated separately, the parentage of each stand will be analysed in detail. RAPDs are not included in any of the initial hypothesis testing, as this would necessitate making assumptions about their parental distributions. This would mean that every hypothesis would assume a H-W distribution in the parents, and rejection of the hypothesis could always be due to this being incorrect. As there is no importance in whether the parents are in H-W or not, it would not be sensible to include RAPDs in the hypothesis testing.

Once hypotheses are made concerning the nature of the two stands, it is then possible to examine whether the RAPD data is consistent with this. Because of the demonstrated non-Mendelian inheritance of at least some of these markers, and the belief that these RAPD data are not reliable, this will be regarded as a test of the use of RAPD markers in this study, rather than tests of the hypotheses themselves.

Again, the chosen approach is not to use the disequilibria parameters in the usual statistical sense. Disequilibria are not independent of allele frequencies, the allele frequencies of the parents of this stand are not known, and the hypothesis to be tested is not simply random mating. This contrasts with the data-sets to which the statistical properties of disequilibria

have been applied. The examples given in a paper treating these properties (Asmussen, Arnold and Avise 1987) both consisted of data sets involving diagnostic markers where the authors used significant disequilibria to demonstrate non-random mating. Even then, it was noted for one of these examples that a sample size of 800 would have been necessary to detect the resulting disequilibrium with a 90% certainty (Asmussen and Basten 1994). Finally, the approach to treating these disequilibria is generally to test a hierarchical set of hypotheses, culminating in the hypothesis that all disequilibria are non-zero. This would lead to the hypothesis that there was non-random mating, directionality to inter-specific matings, and a limited number of generations since hybridisation. Further qualification of this hypothesis is not suggested by treatments of disequilibria in the literature. As this is an essential part of the study, it was decided to use a fundamentally different approach.

The approach is to adopt a hierarchical sequence of hypotheses, based largely on judgement about which hypotheses are regarded as simplest. The simplest hypothesis is *not* that the stands represent random mating between the two parents, as they are not natural populations but have been planted. By tradition, however, this will be stated as the null hypothesis (and be rejected). The next simplest hypothesis , based on the known history of hybrids on the estate, is that they are populations derived from pollination of Japanese trees by European trees, consisting of a mix of hybrids and Japanese parents. This will be referred to as H_{F1}. The next hypothesis, if this is rejected, is that the stand derives from random mating amongst just such a first generation, and will be termed H_{F2}. It is not considered possible that these two early stands could represent later generation plantations, given that the first hybrid populations were only planted in the 1900's. These hypotheses will be tested assuming diagnostic SKDH, mitochondrial and chloroplast markers, a species specific GOT marker, Mendelian inheritance of nuclear loci, and uniparental inheritance of organellar markers. If these hypotheses are rejected, the next simplest set is that the stands are either "F_1" or "F_2" as before, but that some individuals represent aberrant data. In this case, the hypothesis that can be tested with as little removal of data as possible will be applied, and an explanation of the aberrant data suggested.

5.5.7.1 The 1933 Stand As A First Generation Hybrid Swarm

Considering the 1933 stand, the null hypothesis is that the stand is a product of random-mating between mixed European and Japanese larch. This can be rejected immediately, from the absence of European mitochondria and the presence of genotypes not found in F_1s or parents. The lack of European mitochondria is interpreted in terms of all individuals descending from Japanese mothers, and there being insignificant European seed migration at

the time the stand was planted. The next simplest hypothesis is that these are the immediate progeny of Japanese mothers, with mixed Japanese and European parentage. In this case, the stand should consist of either F_1s with European fathers or pure Japanese trees. The expected proportions of the two genealogical classes, based on a proportion 'h' of hybrids are shown in the first column of Table 5-6.

This can also be ruled out, at least if the assumptions are maintained that the organellar markers are species-specific and the Japanese trees do derive from those on the avenue. There are 7 trees homozygous for the European SKDH-1a and b alleles, with the European chloroplast (but, as all the trees do, containing the Japanese mitochondria). Similarly, there are 18 trees homozygous for the Japanese SKDH-1c allele that contain the European chloroplast. Such individuals could not have arisen from an F_1 mating. In terms of cytonuclear disequilibria, D_{Css} can only be zero under this hypothesis and any deviation from this is significant. No formal testing of the disequilibria parameters is necessary- these 25 individuals (out of nearly 50) all represent events with probability 0 under the F_1 hypothesis.

5.5.7.2 The 1933 Stand As A Second Generation Hybrid Swarm

The next hypothesis to be considered is that the 1933 stand is planted from the progeny of such a stand -referred to from now on as 'the F_1 stand', in shorthand form for 'the putative progenitor F_1 stand'. Note that this F_1 stand is entirely hypothetical. Here, we run into one major difficulty. There is no strong reason for assuming the F_2 stand to have a homogenous parentage- it could easily be a mixed seed-lot from various 'hybrid' plantations around the estate. For the sake of simplicity, I will start off by assuming this is not the case.

The simplest model for this stand would be that the F_1 stand, of mixed hybrids and pure Japanese trees, interbred with no pollen or seed flow from other stands. This is not necessarily the case- in particular, a significant European pollen contribution might be expected. However, it provides a simple working hypothesis.

If the F_1 stand consists of an expected proportion h of hybrids and 1-h of pure Japanese, we can work out the expected 'F_2' composition, which will be a mix of pure Japanese, F_2s and backcrosses (BC_1) to the Japanese parents. The expected genealogical distribution depends not only on h, but on N, the size of the F_1 stand. This is due to genetic sampling in the production of the F_1 stand. The expected distribution in an "F_2" derived from an infinite "F_1", or from an F_1 where the hybrid proportion h is not a random variable but fixed is shown in the final column of Table 5-6 (see Appendix C for derivations).

Table 5-6: Expected genealogical classes in hybrid stands under different hypotheses

Class	Hypothetical Nature Of Stand		
	"F_1"	"F_2" (from " F_1" of size N)	"F_2" (from infinite "F_1")
Japanese	(1-h)	$(1/N)(N+h(1-2N)+h^2(N-1))$	$(1-h)^2$
F_1 (with European father)	h	0	0
BC_1 to pure Japanese father	0	(1/N)h(N-1)(1-h)	h(1-h)
BC_1 to pure Japanese mother	0	(1/N)h(N-1)(1-h)	h(1-h)
F_2	0	(1/N)h(h(N-1)+1)	h^2

Notes: Hypothesis "F_1" refers to the hypothesis that the population was produced by random mating amongst one parent (Japanese), with a proportion h of the pollen contribution coming from another parent (European). Hypothesis "F_2" gives the expected genealogical distribution in the random mating offspring, with no further European pollen contribution, of the "F_1" stand of size 'N'. The final hypothesis gives expected genealogical classes assuming an infinite F_1 stand.

The first step is to estimate the relative contributions of each parent to the F_1 stand. The nuclear contribution in the F_2 has already been estimated as 63%. The F_2 has been found to contain 79% of European chloroplasts. Under the hypothesis of no pollen or seed flow , these provide the best estimates of the nuclear and organellar composition of the F_1. As the chloroplast is paternally inherited, the figure of 79% suggests that 79% of the F_1 stand are F_1s , and 21% are pure Japanese.

A least squares estimate of h, the proportion of the putative F_1 generation that is hybrid, can be found in much the same way as m was calculated. Using similar notation as when m was calculated, h can be estimated from each nuclear locus from:

$$P_{hn} = h_n (P_{an} + P_{bn})/2 + (1-h_n)P_{bn} \qquad (7)$$

where P_{hn} is the estimated allele frequency at locus n in the hypothetical F_1 population, h_n is the estimated frequency of hybrids using locus n, P_{an} is the allele frequency at locus n in the paternal population, and P_{bn} the frequency in the maternal population.

so $$h_n = 2(P_{hn} - P_{an})/ (P_{bn} - P_{an}) \qquad (8)$$

The variance in this estimate, assuming the two parental frequencies are known without error (as was done in calculations of m) is given by:

$$\mathrm{Var}(h_n) = 4\ (\mathrm{Var}\ P_{hn}) / (P_{bn} - P_{an})^2$$

so $$\mathrm{Var}(h_n) = (2/n)(P_{hn} + D_{hn} - P_{hn}^2) / (P_{bn} - P_{an})^2 \qquad (9)$$

where D is the H-W disequilibrium as used in (6).

This approach allows the chloroplast frequency to be taken into account. Chloroplast frequencies given an estimate of h (h_c) from:

$$P_{hc} = (1-h_c)P_{ac} + h_cP_{bc}$$

where the suffix c replaces the suffix n to denote the chloroplast rather than a nuclear locus.

so $h_c = (P_{hc} - P_{ac}) / (P_{bc} - P_{ac})$ (10)

and $Var(h_c) = (Var\ P_{hc}) / (P_{bc} - P_{ac})^2$

so $Var(h_c) = (1/n)(P_{hc} - P_{hc}^2)$ (11)

Finally, the least squares estimate of h is constructed using the individual estimates of h weighted by the reciprocal of their variances:

$$h = \Sigma(h_n/Var\ h_n) \ / \ \Sigma\ (1/Var\ h_n) \quad (12)$$

where n goes through all loci including the chloroplast.

The result of this analysis is an estimate of h of 76.1%.

No standard error of h is given, as the assumption that parental allele frequencies are known without error is not true and such an error would be very inaccurate.

Technically, it is also possible to achieve an estimate of N from the genotypic distribution. The error in such estimates will be high as the dependency of the expected genotypic distribution on N is low. It is therefore sensible to begin by testing the hypothesis with N equal to infinity, essentially assuming that the F_1 population is large enough to ignore genetic sampling (genetic drift).

5.5.7.3 Testing The Hypothesis That The 1933 Stand Is Second Generation

One way of testing this particular hypothesis would be if each tree in the stand could be classified as one of the three possible classes (F_2, BC_1, or Japanese). It would then be a simple matter to test whether these observed frequencies were significantly different from the expected frequencies. If they were not, the simple model that the stand derived from random mating in a population of mixed hybrids and Japanese parents would be plausible. Unfortunately, this cannot be done unambiguously, or even close, with so few markers. The problem is to estimate the numbers of various genealogical classes based on genotypic data, and it is very much more difficult than might be supposed at first. As pointed out by Nason and Ellstrand (1993), attempting to classify each tree based, for example, on a hybrid index (such as that constructed above) is inappropriate and statistically unjustified. The same authors proposed a likelihood method, using co-dominant species specific (not necessarily diagnostic) markers to classify the individual trees into several genotypic classes (Nason and Ellstrand 1993). Even this method does not make full use of the data, as pointed out by the

authors themselves. Furthermore, it becomes prohibitively complex when one attempts to introduce organellar markers (this doubles the number of genotypic classes) and attempts to adopt the model for use with dominant markers proved even more futile. While details will not be given, using this method with just the two isozyme markers yielded an estimate of the proportion of F_2 individuals in the stand of 0. This is owing to the way the genotypic classes are constructed in this method, such that the only genotype 'typical' of an F_2 would be an individual homozygote for the Japanese alleles at both isozyme loci. The expected frequency of this genotypic class under the "F_2" hypothesis can be shown to be less than 1/50, when the frequency of hybrid in the F_1 (h) is taken to be 0.76, and the initial allele frequencies in the parents are those observed. This is a function not only of the small sample size, but more significantly, a function of the initial allele frequency of GOT-1b. This problem derives from the less than complete use of the multilocus genotype in this approach. A different approach must therefore be used.

5.5.7.4 Hypothesis Testing Using Disequilibria: A Simulation Based Approach

A common approach to this form of hypothesis testing is to calculate a set of parameters and test them against the expected parameters based on the distribution under the null hypothesis. However, no method could be found that would provide formulae for expected disequilibria and their variances in this situation. Even parameters based on the null hypothesis of random mating were approximate and unsuitable for all but large samples (Asmussen and Basten 1994). The suggested approach in the past is to test a hierarchical set of hypotheses culminating in a test of the hypothesis that there is non-random mating, directionality to inter-specific matings, and a fairly young stand (Asmussen, Arnold and Avise 1987). However, this approach is simply not suitable in this case- even were this hypothesis found to be the most suitable, it does not provide the detail necessary in this case. Furthermore, the authors ordering of hypotheses in terms of complexity does not make allowances for the specific situation in this population. In fact, it could be argued that the 'simplest' hypothesis of random mating, if accepted, would be extremely difficult to explain.

A general and useful approach to such a problem is to simulate the null hypothesis using a Monte-Carlo method. In this way, the distribution of parameters of interest can be found. Clearly, it is intellectually less satisfying than an exact approach, but is now an accepted way of parametric hypothesis testing.

At the outset, the limitations of the following approach will be made clear. First and foremost, the estimate of h, the proportion of F_1 individuals in the "F_1" stand has been taken to be 0.76,

estimated from RAPD, chloroplast, and isozyme frequencies. No attempt to justify this as a maximum likelihood or unbiased estimator has been made. Instead, this estimate has been made purely for the purposes of hypothesis testing having already rejected the simplest models based on random mating or first generation mixed mating.

Secondly, the expected frequencies of the genealogical classes, and thus the genotypic classes, are not independent of the size of the F_1 (N). This introduces yet another variable. As mentioned, it is justified to treat N as infinity as the simplest hypothesis and use the expected genealogical distributions without adjusting for genetic sampling (i.e. the third column in Table 5-6). As an aside, it was observed that the minimum X^2 estimator for N in this data set did not converge when solved iteratively on a spreadsheet- in other words, this (not necessarily best) estimate of N was infinity. However, the difference between X^2 at N=10 and N=1000 was only 0.5- hence the declaration that an estimate of N from this data is not worth discussing.

Mitochondria have been ignored as no variation is present in the population. However, the explanations given are all based on the observation that the seed contribution of European trees to these stands is negligible.

Finally, though this parametric method can be informative, there is one unavoidable problem- regardless of whether the parameters have a known distribution or just a simulated one. If only a particular set of parameters is of concern, then it is justified to test only these. However, in this case, the model must account not just for cytonuclear disequilibria, but all other associations- there must be 14 exhaustive parameters to completely summarise the data which can belong to any one of 15 possible genotypic classes. Testing just the cytonuclear disequilibria is only a beginning to testing any hypothesis. Thus, though the cytonuclear disequilibria, the allele frequencies, and the composite genetic disequilibria will be tested this way, a better approach will also be used. To avoid the issue of testing the same data set using two different tests, I point out that the parametric approach is included for completeness and to provide an understanding of how it may be achieved, whereas the better goodness-of-fit approach is regarded as the true test of the hypothesis.

The computer simulation (Appendix D) was originally written to look at the long term behaviour of cytonuclear disequilibria, but has been substantially modified to apply it to this particular problem. That it, a test of the following hypothesis:

H_{F2}: The stand in question derives from random mating in a large population itself a mixture of F_1s and pure Japanese parents, with all European genes deriving from pollen. The chance that an individual in the "F_1" is hybrid is 0.76. Species diagnostic markers exist for both SKDH and the chloroplast, and an allele of GOT is specific for the Japanese parent with a frequency of 12/22.

Rejection of this hypothesis would either have to lead to a re-evaluation of h, which could (but not necessarily) result in the same hypothesis with a modified h, or a total re-evaluation of the nature of the stand.

The simulation simply calculates the probability that each of n individuals in the second generation is of a particular genealogical class (as in Table 5-6), and then based on this probability assigns each individual to a class. Based on the individual's class, its genotype is randomly generated from the laws of Mendelian inheritance. This approach is equivalent to assuming an infinite size in the F_1- another model was used to simulate the effects of sampling in this first generation. As only 38 individuals were entirely scored in the sample, n was taken to be 38. The resulting population can be entirely summarised by a 3*3*2 contingency table. The output is then entered into a spreadsheet allowing easy calculation of disequilibria and frequencies. That the programme worked was checked against the results of random mating models (Fu and Arnold 1992) under various conditions.

Table 5-7 shows some of the parameters in the first column (3 allele frequencies, 8 cytonuclear disequilibria, and the digenic composite disequilibria between the two nuclear loci). The second column shows the average, over 1000 simulations, of these parameters. The third column is the 90% probability interval centred on the mean, calculated from the simulations. The final column contains the observed parameters, as presented in more detail in Tables 5-1 and 5-2. As can be seen, all the observed parameters fall within these boundaries, so taking them one at a time, none of these parameters are significantly different (even at the 10% level) from the "expected" parameters from the computer simulation. .

Table 5-7: Genetic parameters in the 1933 stand, simulated and observed

Parameter	Simulated Average	90% Probability Interval	Observed
P_C	0.76	0.64 , 0.86	0.78
P_S	0.38	0.28 , 0.47	0.38
P_G	0.66	0.57 , 0.75	0.59
D_{cS}	0.05	0.01 , 0.08	0.05
D_{cSS}	0.03	0.01 , 0.06	0.04
D_{cSs}	0.02	-0.04 , 0.08	0.03
D_{css}	-0.06	-0.11 , 0.00	-0.07
D_{cG}	0.02	-0.02 , 0.06	0.03
D_{cGG}	0.03	-0.02 , 0.09	0.05
D_{cGg}	-0.02	-0.07 , 0.04	-0.05
D_{cgg}	-0.02	-0.06 , 0.03	-0.01
D_{SG}	0.03	-0.03 , 0.09	0.04

Notes: The first three parameters are the allele frequencies of the three markers, the next 8 are cytonuclear (4 for each locus, starting with the allelic disequilibria), and the final one a nuclear disequilibrium.

5.5.7.5 Testing The Hypothesis Using A Goodness Of Fit Test

Though the fit of the observed disequilibria to the simulated disequilibria is encouraging, a better approach is a goodness-of-fit test such as the χ^2 test, which will test all the data rather than just a subset of its parameters. For this, the genotypic distribution under the null-hypothesis is needed. This is found from the genealogical distribution calculated in Table 5-6, in conjunction with the laws of Mendelian inheritance. Again, RAPDs are ignored- no reasonable way of including them is evident. Partly, it would be problematic to perform a goodness of fit test using any more markers because the expected cell frequencies would become too low.

There are 38 samples with complete genotypic data and 15 possible genotypes in the stand. The expected frequencies in the F_2 derived from an infinite F_1 are simply calculated by constructing a 2 way table of the 15 possible genotypes against the 4 possible genealogical classes. The probability that each genealogical class is a particular genotype (i.e. the probability of having a particular genotype contingent on that genealogical class) is readily worked out from Mendelian inheritance, and the probability that an individual is a particular genealogy has been shown in Table 5-6. The expected proportion in each genotypic class is then simply the sum of the 4 contingent probabilities multiplied by their respective genealogical class probability. Table 5-8 shows the results of the calculations based on h=0.76, using the parental allele frequencies, and a final population size of 38.

From these values, a χ^2 test can be performed. It is sometimes stated that individual frequencies must be at least 1, or even 5 for χ^2 to be an adequate approximation to the X^2 statistic. However, more detailed studies have suggested the following criteria for such a test being suitable.

d.f>1, n>9, and n^2/(no. of cells)>9 , then the χ^2 approximation is valid (D'Agostino and Stephens 1986)

The other issue is the calculation of d.f. There are 15 genotypic classes, giving 14 d.f. The only parameter to be estimated was h. However, this is not a minimum chi-squared estimator (its estimate was based on raw allele data and other RAPD data), and it is therefore not true that the χ^2 distribution is entirely valid. However, critical points *will* lie between d.f=13 and d.f=14 (D'Agostino and Stephens 1986).

From table 5.8, X^2 can be calculated.

$X^2=12.9$

From statistical tables using 13 and 14 d.f,

$22.4< \chi^2_{0.05} < 23.7$

so, $X^2 < \chi^2$ and H_{F2} is accepted.

Actually, because the simulation has already run, it is easy to obtain a "simulated" critical value making no assumptions about the suitability of the χ^2 assumption. If this is done, the critical value is shown to be 24, and the χ^2 approximation and the simulated distribution are in close agreement.

Table 5-8: Genotypic probabilities contingent on genealogical class, and genotypic frequencies

Genotype			Probability Contingent on Genealogical Class				Frequency	
SKDH-1	GOT-1	CP	Jap	F_2	Bcm	Bcf	Exp.	Obs
EE	aa	E	0	0.13	0	0	2.90	4
EE	ab	E	0	0.10	0	0	2.18	3
EE	bb	E	0	0.02	0	0	0.41	0
EJ	aa	E	0	0.26	0	0.17	6.95	5
EJ	ab	E	0	0.20	0	0.26	6.16	7
EJ	bb	E	0	0.04	0	0.07	1.33	3
EJ	aa	J	0	0	0.17	0	1.15	0
EJ	ab	J	0	0	0.26	0	1.80	2
EJ	bb	J	0	0	0.07	0	0.52	1
JJ	aa	E	0	0.13	0	0.17	4.05	4
JJ	ab	E	0	0.10	0	0.26	3.98	1
JJ	bb	E	0	0.02	0	0.07	0.92	3
JJ	aa	J	0.21	0	0.17	0	1.60	1
JJ	ab	J	0.49	0	0.26	0	2.89	3
JJ	bb	J	0.30	0	0.07	0	1.17	1
Genealogical Probs.			0.06	0.58	0.18	0.18		

Notes: For SKDH-1, 'E' refers to both SKDH-1a and SKDH-1b alleles, and 'J' to SKDH-1c alleles (as before). For GOT-1, the alleles are denoted conventionally. For the chloroplast, the European chlorotype is denoted by 'E' and the the Japanese by 'J'.
Appendix C shows how the contingent probabilities were calculated. Genealogical probabilities are from table 5.6, with h set to 0.75, and N to infinity. Expected genotypic frequencies are the sum of the genealogical probabilities x the contingent probabilities, multiplied by 38 to provide expected frequencies. Bcm means Backcross to the Japanese as male, Bcf means the Japanese parent acts as female.

5.5.7.6 The 1933 Stand: Conclusion

Simple considerations showed that the 1933 stand could not be first generation. The next simplest hypothesis (H_{F2}) was taken to be that the stand was planted with seed taken from a large, random mating population with a proportion of hybrids and a proportion of Japanese trees. This proportion was taken to be 0.76, estimated from and consistent with both chloroplast frequencies and 'm', the estimated contribution of each parental gene-pool. The sampling distribution of disequilibria is complex, so a simulation was used to generate the distribution of disequilibria under this hypothesis, and all two-locus disequilibria as well as allele frequencies were found to be within the 90% Probability Interval. The more powerful- because not just a subset of parameters are being examined- goodness of fit test was used to test the hypothesis, which was accepted. This conclusion is consistent with the PCA and AMOVA, which suggested a closer relationship between the stand and the Japanese parents

than the European parents, and the character index, which suggested the possibility of there being several Japanese individuals in the population.

5.5.7.7 The 1936 Stand

Eliminating the possibility that this stand is from a random mating mixed parental population (again, the mitochondria is fixed for the Japanese type, and there are significant disequilibria), the next hypothesis is that it is a mixed F_1/ pure parental type. There are only 3 individuals that cannot be explained under this hypothesis (unlike the 25 in the 1933 stand). This calls for an element of judgement in how best to proceed. Clearly, it is possible that these represent mis-scored (though each was duplicated) individuals or products of Non-Mendelian inheritance. Variation in the European parents does not explain any of these genotypes, and variation in the Japanese parents can be ruled out as all parents have been analysed. However, the procedure adopted above necessitates testing that this hypothesis is the random mating product of an F_1 (i.e. a test of H_{F2}).

This hypothesis can be rejected on two grounds. Firstly, there is still one individual which cannot be explained (the individual with a European nuclear genotype and a Japanese chloroplast). Secondly, it has already been demonstrated that this stand is not the product of random mating, as it has a significant heterozygote excess at both isozyme loci (Table 5.1).

Following the procedure described in 5.5.7, H_{F2} is now tested after removal of the one aberrant genotype. Technically, the H-W disequilibria could be recalculated to confirm that this hypothesis is still to be rejected. However, it is desirable to test this hypothesis using the same methodology as the test for H_{F2} in the 1933 stand, that is a goodness of fit test. Following the same procedure as before, the test statistic X^2 was calculated as 32.5. This is significant at 0.5% if the χ^2 distribution is assumed, and significant at 1.5% if the simulated distribution is used. However, this has used the same estimated value of 'h' as the 1933 stand of 0.76, and a value for N of infinity. To reject H_{F2}, we must use the values of h and N that minimise X^2 (D'Agostino and Stephens 1986). This is easy to perform iteratively on a spreadsheet. The values of h and N that minimise X^2 are 0.76 (very close to the 1933 value) and 2(!) respectively, which reduce X^2 very little (to 30.5). This is still highly significant, and we can reject H_{F2} even after removal of the one "aberrant" individual.

The next hypothesis is that the stand is a mixed F_1, but with three aberrant individuals that must be removed (this was tested after H_{F2}, as more data needed to be removed). Clearly, by definition, every individual in this population can now be classified as an F_1 or a pure Japanese tree, and this hypothesis must be accepted. The ratio of hybrids: Japanese individuals

under this hypothesis, counting those scored at both SKDH-1 and the chloroplast, is 7:22, suggesting a value of 'h' of 0.74.

While not made explicit in the hypothesis, nor an essential part of the model, it was assumed that the mating to produce the F_1 was random. It was suggested by the GOT-1 data that this was not the case. As this suggestion was made after inspection of the data, formal testing of this hypothesis is not possible on this data set. However, the evidence is suggestive. Firstly, in this 1936 stand, all but two (out of 18 scored for GOT-1) hybrid individuals are heterozygous. These individuals are not expected to be in H-W equilibrium. However, based on the frequency of GOT-1a in the Japanese parents of 0.45 (it is fixed in the Europeans), one would expect 10 heterozygotes. It seems plausible that the Japanese parents of the hybrids are not random and that the GOT-1b homozygotes on Kennel bank are enjoying more reproductive success, at least as far as being maternal parents of the hybrids is concerned. This situation is also suggested within the 1933 stand. In both cases, the allele frequency of GOT-1b in the hybrid stand is somewhat higher than the expected frequency based on the frequency in the Japanese parents. However, it is acknowledged that this hypothesis was formed after observing the data and further studies- such as monitoring the reproductive output of the Japanese parents- would be needed to test this.

Some explanation of the three 'aberrant' genotypes is offered. The 1936 stand, planted adjacent to the 1933 stand could easily have received some seed from this stand. In fact, as it is a planted stand, it could easily have received a mixture of seed from all over the estate, effectively making it possible to formulate any hypothesis regarding the nature of any individual that one wishes. Thus, it is possible to explain away any individual that does not fit with the hypothesis. This is not desirable, but unavoidable, given the unnatural nature of the stands. Effectively, it is more of a subjective matter than a statistical one how such individuals are treated. However, I would argue that while these individuals make the hypothesis less satisfactory, they do not invalidate the general conclusion that this stand is essentially F_1 in character.

In conclusion, this stand appears to be first generation rather than second generation like the 1933 stand. Furthermore- and possibly coincidentally- the proportion of hybrids in this stand is almost exactly the same as that hypothesised for the parents of the 1933 stand. It is plausible that this indicates that a fairly consistent 75% hybrid output is achieved by the Japanese parents at Kennel bank, though this may be after selection for hybrids at the seedling status by the foresters of the time. Finally, it is suggested that the GOT-1b homozygotes in the

Japanese parents may be more fecund with regards to hybrid offspring than their neighbours, though this has not been tested.

5.5.8 Consistency Of The RAPD Data

It is now possible to test whether the RAPD data is consistent with these hypotheses. Rather than make complex predictions about their expected distribution in each stand based on their proposed ancestries, simple observation shows they are either uninformative, neither supporting nor rejecting the hypotheses, or inconsistent.

OPK16c is inconsistent with both hypotheses in both stands. This was the marker that was eliminated from the estimate of m discussed earlier, as it was present in such a high proportion in both hybrid stands. As a European specific band, it should have been missing in the 8 putative Japanese trees in the 1936 stand. In fact it was present in all of those scored. This suggests that either the stand is not a mixed parental-F_1 as suggested, it is not inherited in a predictable fashion, or it is not being scored reliably. This is purely a matter of judgement. This RAPD was duplicated, and in both cases gave the same scoring pattern apart from individuals that failed to amplify at all in one or other run, so the scoring is believed to be safe. That it is not inherited in a predictable fashion has already been demonstrated from the controlled crosses. It is not easy to explain how the marker was absent in all 10 of the 11 Japanese parents scored, but present in all of the putative Japanese offspring, without invoking extremely non-Mendelian inheritance. The situation is similar in the 1933 stand, where it is present at far higher frequency than predicted. One possibility is that the band should be amplified in all or most individuals, but that the DNA extractions of the Japanese trees- collected much earlier on in the project from older trees- contained contaminants that inhibited the amplification of this band. In this case, the variation seen between the two parents may be due to non-genetic effects. Another possibility is that it was a bacterial or fungal marker that had not infected the Japanese parents but was present everywhere else. Both of these are speculative, but certainly too much faith should not be put in one RAPD marker when its inheritance has not been adequately tested.

The two Japanese specific markers, OPAL14a and OPA7, present in all Japanese parents but not the Europeans also present a problem. In the 1936 stand, the 8 individuals of hypothetical Japanese nature should contain both bands. In fact, 4 individuals are missing one or the other (but not both) bands. Though this could be due to non-Mendelian inheritance- for example, inhibition by a competing and variable European locus, it is also possible that these are consistent but uninformative. As an example, opal14a was scored for 6 of the 8 supposed

Japanese trees in the 1936 stand, and missing in two of these. We have no way of knowing whether this is due to heterozygotes in the 11 parents (an excess which could easily exist by chance alone), or an inconsistency in our hypothesis. Either way, the RAPD distribution in the population cannot be used to disprove the hypothesis.

Statistical treatment of the RAPD data is not tackled. The complexity was sufficient in co-dominant data to suggest a rigorous treatment of this RAPD data, even were the sample size greater, to be unprofitable. This is not simply due to the small sample size in the populations, but the small parental Japanese genepool. Even under the hypothesis that these Japanese parents derived from random mating, large heterozygote excesses or deficiency would be common through chance alone. Ultimately, the only solution to this would have been to obtain allelic distributions for the parents (by analysing pure offspring of these Japanese trees in a series of controlled crosses, for example). This would have enabled the RAPD data to be used in the hybrids to test hypotheses. If the allelic frequency in the parents was known with any great confidence, then the expected pattern of dominant markers in each stand could be constructed. Without this information, there is simply too much uncertainty in estimates of the allelic distribution to be able to do this.

This criticism of the RAPD data set begs the question of whether the estimates of 'h' and character indexes should be re-evaluated. This would not be justified statistically, as it is simply removing data that does not agree with the hypothesis. However, it is reassuring to note that the goodness of fit tests, not based on the RAPD data, provided their own (not necessarily least squares or unbiased) estimates of 'h' that were in close agreement with the values used.

In conclusion, it is unfortunate that the use of RAPDs was not very profitable. Partly, this was due to the inability to find as many well-behaved species specific markers as expected. More importantly, it became evident that the Japanese parents at least would have had to be more thoroughly analysed before expecting RAPDs to assist the analysis of hybrids. However, the maternal marker discussed earlier was interesting and useful, and these RAPD markers could, with more extensive inheritance checking, be informative in future studies.

5.6 Conclusion

Two adjacent hybrid stands of unknown origin were analysed with two isozyme markers, organellar markers, and RAPD markers. There were significant differences in these two stands in terms of genotypic distributions. Both consisted of large numbers of individuals with

European chloroplasts, some Japanese chloroplasts, and all Japanese mitochondria. SKDH heterozygotes also demonstrated the existence of many hybrids.

A method for estimating parental contributions of nuclear genes to a hybrid population was modified from previous work, to allow the incorporation of RAPD data. This estimated that similar proportions (between 60 and 65%) of the gene-pool derive from the Japanese parents. These values suggest that the European trees must have provided between 70 and 80% of the pollen, as no European contribution through seed was detected. These values are in close agreement with the observed chloroplast frequencies as independent estimates of the European pollen contribution.

The 1936 stand fitted with the hypothesis that it derived from planting seeds from Japanese trees, 75% of which had been fathered by European trees, with the possibility of some Japanese trees being preferred over others. The 1933 stand appeared to be the offspring of a very similar stand (i.e. second generation rather than first generation).

The fit of various parameters, especially cytonuclear disequilibria parameters, to these hypotheses was checked using a computer simulation, but an exact goodness-of-fit test was found to be more satisfactory. This computer simulation could, however, be used to examine the behaviour of cytonuclear disequilibria under a wide range of conditions to see under what circumstances disequilibria are useful tools.

Cytonuclear disequilibria do not appear to be a particularly useful way of analysing this form of data- where the populations are very young- and it is suggested that too much weight has been given to them in much of the literature. If a pattern of mating is occurring that not only affects cytonuclear disequilibria but other parameters, it is preferable to test the hypothesis directly rather than calculate the expected disequilibria and then test these.

That neither stand contained any European seed suggests either a rather low dispersal for European seed on the estate (which was surprising, given the predominance of this species), or difficulty in establishment. That the European genomic contribution in the F_2 was no greater than in the F_1 suggests also a limited pollen dispersal into these two stands- an even more surprising result. These hypotheses will be tested by looking at regeneration occurring in these two stands (chapter 6).

These results suggest that on this estate at least, hybrid stands consist of backcrosses, parents and F_2s as well as just F_1s, and that planting hybrids round the estate has also led to distribution of pure Japanese trees round the estate. The effect of these stands on European plantations has not been touched upon, but would be an interesting study.

6. Molecular Characterisation Of Regenerants In Two Stands Of Hybrid Larch On The Atholl Estate

6.1 Introduction

Two stands of larch on the Atholl estate, planted in 1933 and 1936, were found to differ significantly in their genetic structure (chapter 5). These two stands have been fenced off for 4 years to encourage natural regeneration by keeping out the deer that normally graze on the seedlings. This has been extremely successful- both stands have areas with a high density of young seedlings. The existence of such regeneration allows an analysis of some aspects of the mating system of these populations, and the possible future for similar hybrid plantations to be assessed.

6.2 Aims

The marker systems developed for parental species and controlled crosses (chapter 3) and applied to hybrid stands (chapters 4 and 5) were used to examine young regenerants in the 1933 and 1936 hybrid stands. This enabled a qualitative assessment of the viability, and mating system, of these hybrid stands in a semi-natural environment. In particular, it was hypothesised that much of the regeneration may be European seed from other stands, or have been fathered by a significant European pollen contribution. The consequences of this would be the gradual dilution of hybrid plantations with European genetic material. The main aim of examining the regenerants was to test this hypothesis.

6.3 Materials And Methods

Detailed protocols for all methods used are in Appendix A.

The location of the two stands from which the regenerants were sampled was described in chapter 5.

As isozyme and needle collections had to be made separately, the ideal solution would have been to label each regenerant when collecting needles and buds. However, this was not practical, for two reasons. Firstly, no simple way of securely labelling such young seedlings was suggested- it seemed likely that were such a task undertaken, a significant amount of seedlings would not be identifiable by the second visit. More importantly, particularly for the younger seedlings, the entire seedling was taken for the DNA extraction. Thus, as part of the initial assessment of the regeneration, cytonuclear data were foregone and different individuals were sampled for isozyme and DNA analysis.

Fifty individuals from each stand were sampled for DNA analysis. This sampling was only semi-random, as it was deliberately decided not to sample regenerants in close proximity. This was to prevent over-sampling the progeny of a tree which happened to be enjoying

reproductive success as a result of the environment in its vicinity (i.e. canopy cover, groundwater e.t.c.). Thus, individuals in large patches of regeneration were less likely to be chosen than individuals in smaller patches of regeneration. Also, there was likely to be a bias to larger, more visible individuals, though efforts were made to avoid this.

DNA extraction was performed using the Scotlab Phytopure kit on 0.1g of needle tissue.

Mitochondrial and chloroplast probes were used as previously (chapters 3-5) on genomic blots.

Originally, 50 individuals from each stand were sampled for isozyme analysis. However, as will be discussed, the collection had to be repeated and ultimately just 30 individuals from each stand were analysed.

RAPD analysis of the regenerants was attempted at length, using the same methods and primers as before.

6.4 Results

Figure 6-1 shows a sample autoradiograph from the Bcl enzyme / K140 chloroplast probe RFLP analysis. It is immediately apparent that these results were less than satisfactory. In contrast with previous autoradiographs, there exists much variation in the migration distances of each band from lane to lane. It is easily seen that this is most likely due to impurities that are tightly associated with the DNA. Every band in each 'bad' lane, including the undigested DNA at the top of each lane, is shifted relative to its neighbour. This makes it more difficult to score, and this is reflected in a large number of missing data (because scoring could not be carried out with a great deal of certainty). A more disruptive effect of this is the possibility that the scoring is biased. In this case, it might be that the frequency of the minority Japanese chloroplast is underestimated- the doublet band of the European chlorotype is fairly obvious, whereas it is not easy to distinguish a Japanese individual from poorly digested European individuals when the migration distances are so variable. A subsequent Phenol/Chloroform extraction was conducted in an attempt to remove contaminants followed by a repeat blot but this was unsuccessful.

All regenerants scored had the Japanese mitochondrial haplotype, implying that little of the regeneration has derived from seed from neighbouring European stands. Secondly, there was variation in the chloroplast genome- most regenerants contained European chloroplasts as did their parents, though there were still some individuals containing Japanese chlorotypes.

Isozyme analysis was unsuccessful in the first year of collection. SKDH results were not comparable with previous work, with apparent expression of another locus. Though collection

was carried out at the same time of year as previous samples, the winter had been somewhat milder than usual. Though not proved, or tested, this might suggest a switchover from expression of a 'winter' form of SKDH to a 'spring' form. This switchover may have occurred earlier than in previous years due to the mild winter and confounded the results. Expression of another locus was suggested also in a study of a larch seed orchard (Wirawan 1997). Successful isozyme analysis was carried out in the following year, when samples were collected earlier in the winter.

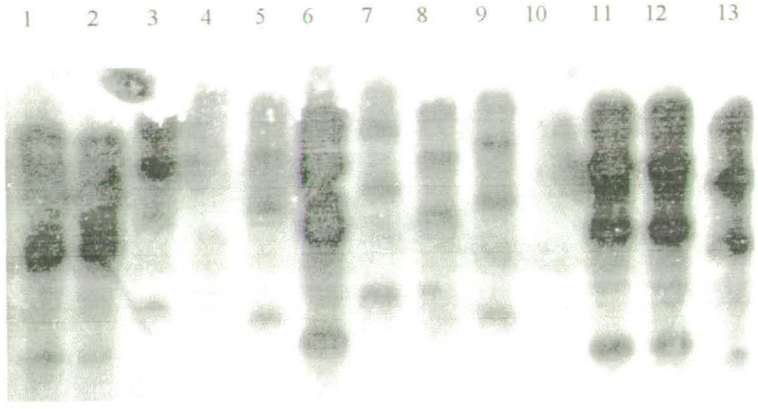
Finally, RAPD analysis, which occupied a great deal of effort, was largely unsuccessful. Gels were less reproducible than before and did not amplify consistently. This may have been due to external effects (water quality, Taq, contaminants etc.) but it is thought more likely that it was an effect of poor DNA quality, already suggested by the variability in the Southern blots that examined cpDNA variation. Unfortunately, no RAPD data could be collected for these regenerants.

Single locus results for the organellar markers and the two isozyme loci are shown in Table 6-1. For the two isozyme loci, the H-W disequilibrium and its significance level in each case is shown. In the regenerants from the 1933 stand, both disequilibria are non-significant. The H-W disequilibrium in the 1936 regenerants is significant at SKDH, and indicates a heterozygote deficit at this locus. However, this cannot be taken to prove that this population does not derive from random mating (see discussion). This table also shows the overall allele frequencies in the regenerant generation.

In Table 6-2, for ease of reference, these allele frequencies are shown alongside the allele frequencies in the planted stand (from Chapter 5). The p-value in this table refers to the significance of the exact test performed by Genepop (Raymond 1997) for genic differentiation at each locus between the regenerant and planted populations. No significant differences are seen. However, this is not the correct approach for testing the hypothesis that there has been no external European contribution to the regenerants. This hypothesis will be discussed in the following section.

Finally, as cytonuclear data were not collected, the usual disequilibria could not be calculated. However, Table 6-3 shows the composite genotypic disequilibria between GOT and SKDH and the significance of the exact test of independence- in neither case was a significant association detected between genotypes.

Figure 6-1: Autoradiograph of Southern blot of Bcl digested genomic DNA from regenerants in 1933 and 1936 stands, probed with Bcl chloroplast probe



Notes: Poor DNA quality meant that band migration was not consistent from one sample to the next.
Lane 1: A European parent Lanes 2-13: Regenerants
Samples 3 and 8 have Japanese chloroplasts, based on the distance between the two fastest running bands (e.g. compare lanes 3 and 5, or lanes 8 and 7). The others have European chloroplasts.

Table 6-1:Single-locus results for regenerants, showing allele frequencies and testing for Hardy-Weinberg disequilibrium

Locus	Allele Frequency In 1933 Regenerants	Allele Frequency In 1936 Regenerant	Overall Allele Frequency, From Pooling Regenerants In Both Stands
SKDH- 1,2	*26/60 = 0.43*	*24/56 = 0.43*	*50/116 = 0.43*
H.W D	*0.048 n.s*	*0.101**	*n/a*
GOT-1	*42/60 = 0.70*	*35/50 = 0.70*	*77/110 = 0.70*
H.W D	*0.01 n.s*	*0.030 n.s*	*n/a*
Chlor. E	*22/26= 0.85*	*22/27 = 0.81*	*44/53 = 0.83*
Mito. E	*0/23 = 0.00*	*0/26 =0 00*	*0/49 = 0.00*

Notes: As usual, the frequencies are shown with their denominators to indicate sample size. For each isozyme locus, the Hardy-Weinberg disequilibrium (D) is calculated, and presented alongside its significance.

Table 6-2: Comparison of allele frequencies between regenerants and planted trees, and significance level of exact test for genic differentiation between them

Locus	1933 Stand			1936 Stand		
	Regens	Parents	p-value	Regens	Parents	p-value
SKDH- 1,2	0.43	0.38	0.61 n.s	0.43	0.47	0.73 n.s
GOT-1	0.70	0.59	0.16 n.s	0.70	0.58	0.20 n.s
Chlor. E	0.85	0.79	0.57 n.s	0.81	0.78	0.71 n.s
Mito. E	1.00	1.00	n/a	1.00	1.00	n/a

Notes: The allele frequencies in the regenerants (from table 6-1) are shown alongside the corresponding frequencies in the planted stands (from chapter 5). The p-value indicates the significance level of the exact test for differences between these two (based on the actual sample numbers).

Table 6-3: Composite genotypic disequilibria (Δ_{SG}) between SKDH-1 and GOT-1 in regenerants from 1933 and 1936 stands

	1933 Stand			*1936 Stand*		
	Abs. D.	*Norm D*	*Sig.*	*Abs. D.*	*Norm. D.*	*Sig.*
Δ_{SG}	0.05	0.27	n.s	-0.08	-0.23	n.s

Notes: See previous chapters for details of how normalised and absolute disequilibria are calculated.

6.5 Discussion

Table 6-2 shows that the regenerants are broadly similar in terms of allele frequencies to their parental stands. At both isozyme loci and also for the chloroplast type there is an increase in the European allele, but these increases do not appear significant when simply testing for frequency differences. However, the simple test for genic differentiation as performed in Table 6-2 is not appropriate for testing whether these populations are derived only from their respective parents.

6.5.1 Composition Of Regenerant Populations

In the previous chapter, a parameter *m* was calculated that estimated the proportion of nuclear alleles that derive from the Japanese parents. It was noted that this is consistent with the observed chloroplast frequencies, but it was *not* based on these frequencies. Indeed, the chloroplast frequency could only be included in the analysis when a particular model for the origin of the stand had been generated.

m based on just the two nuclear loci was calculated for each population as in chapter 5. For both the 1933 and 1936 regenerants, *m* was calculated to be 0.57 (i.e. 57% Japanese alleles). This value is slightly lower than the parental estimates of 63% and 60%- representing a slight increase in the observed frequency of European alleles- but unlikely to be significantly so. However, the assumptions used to generate the variance of m, as mentioned in chapter 5, are felt to invalidate this approach as a method of hypothesis testing.

6.5.2 Testing For Immigration From Neighbouring Stands

As stated at the outset of this chapter, the main point of interest was to determine whether the regenerants had been fathered largely by the individuals in the planted stand, or whether there was a significant pollen contribution from the many European plantations that surrounded these stands. For this reason, it is necessary to test the hypothesis that there is an increase in European alleles from one generation to the next.

Testing *m* alone is not appropriate for detecting European pollen, even if the distribution is known, as the chloroplast frequency also depends on the proportion of immigration. As an

interesting aside, it is not the case that, as might be expected, pollen specific markers provide a more sensitive test of pollen immigration in this situation. In fact, simple calculations or 'experimentation' on a spreadsheet demonstrates that this is only the case when immigration and initial European allele frequencies are low. If all the planted stand had been hybrids, the initial European chloroplast frequency would be one and European pollen could *only* be detected by its impact on the regenerants nuclear composition. Furthermore, the statistical significance of a change in chloroplast frequency is limited relative to a change in isozyme frequency as only one allele rather than two is sampled. This caveat aside, it is still essential to include the chloroplast frequency in a test of pollen immigration.

It is not justified to simply test for genic differentiation at each locus (all non-significant) and conclude that there is no significant contribution from the European trees that form much of the estates forestry, as the three loci are far from independent. This approach would assume that the frequencies at each locus in the planted stand varied independently, whereas it is clear that if the European proportion at one locus had decreased, any increase at another locus would become more significant. A stringent approach, is outlined below, in the absence of a rigorous multi-locus method.

The first point is that the hypothesis that there was no significant seed contribution from European trees on the estate can be accepted. This is because no European mitochondria were detected in the regenerants.

The next step is to test for a significant European pollen contribution. The parental populations are assumed to be exactly as hypothesised in Chapter 5, such that the chloroplast frequency in each is exactly 76% and m (the proportion of Japanese nuclear alleles) is exactly 62%. This gives a chloroplast frequency of 76%, a European SKDH allele frequency of 38%, and a GOT allele frequency of 66%. These are taken to be known without error (the reason this test is so stringent), and used to generate expected allele frequencies in the regenerant population under the hypothesis of no immigration of pollen. Each of the three loci is then tested.

As sample sizes were far lower than the desired 50 per site, the counts were bulked over sites. This is justified, as in terms of all three locus frequencies, the two sites are very similar (coincidentally, the two isozyme loci have identical observed frequencies). These frequencies are shown in Table 6-1.

The null hypothesis is that the regenerant population derives from the parental population described above, with the alternative hypothesis being that there was a European pollen

contribution. Noting that this is a one tail test, we use the Binomial distribution rather than a χ^2 goodness of fit test. The results of this analysis are shown in Table 6-4. The cumulative binomial probabilities were obtained using Excel.

Table 6-4: Distribution of European allele frequency at each locus under hypothesis of no external European pollen contribution, observed allele frequency, and test for significant increase in European allele frequency

Locus	Distribution of European Allele under H_0	Observed Count (Ob)	Pr(X>=Ob)
Chloroplast	X~B(53,0.76)	44	0.08
SKDH	X~B(116,0.38)	50	0.11
GOT	X~B(110,0.66)	77	0.17

Notes: The distribution of each allele count under the null hypothesis is binomial, with number of trials given by the number of scored individuals for the chloroplast or twice this for the nuclear loci. The probability of success is simply the allelic frequency under the hypothesis that the parental population is, or derives from, an infinite population of 76% hybrids (with European fathers) and 24% pure Japanese trees.

The hypothesis has been accepted in all cases, so regardless of the non-independence of these loci, there is no significant evidence for an external European pollen source. It is stressed that the stringency of this test, taking the parental frequencies to be known without error, means that rejection of this hypothesis would not be the same as rejection of the hypothesis that the regenerants derived from the *observed* populations. However, acceptance of the 'stringent' hypothesis can be taken to demonstrate that this, more relevant, hypothesis can also be accepted.

It is, however, pointed out that the power of this test on this data set is not high. This is partly because the sample size was only half that collected. Had the frequencies been the same in the initial, planned samples sizes of 50 per site, both the changes in chlorotype frequency and SKDH-1 allele frequency would have been significant. The other reason for the low power of this test is also due to the unexpectedly high hybrid composition of the parental stands making it harder to detect a further increase in the next generation. Routine calculations using the same one tailed test as above demonstrate that the minimum detectable immigrant pollen level, given the initial frequency, would be about 25% to show a significant change in the SKDH frequency and as high as 40% to show a change in the chloroplast frequency, with these sample sizes. If the sample sizes were doubled, these drop to 18% and 27% respectively.

6.5.3 Least Squares Estimate Of The Proportion Of European Pollen

Having failed to reject the hypothesis that no pollen came from European plantations, the following procedure should be regarded as a way of formulating hypotheses for future studies

rather than a further analysis of these small samples. Thus, though the least squares estimate will not be significantly different from zero, it is still instructive to see how it might be calculated.

A least squares-estimate of the proportion of immigrant pollen is easy to construct, following methodology from chapter 5. Equations 7-11 from chapter 5 can be used, resulting in an estimate *h* of the proportion of regenerants that were fathered by European pollen. Table 6-5 shows how this is achieved. Again, the regenerants are assumed to derive from an infinite population where the allele frequencies are known without error and calculated by assuming the populations derive from a population of 76% hybrids and 24% pure Japanese trees.

Table 6-5: Allele frequencies in regenerants and parents (bulked over both sites), and estimates of the proportion of immigrant pollen (h) calculated for each locus

Locus	Frequency in regenerants (P_h)	Estimated Frequency in 1933/1936 Parents (P_b)	Frequency in European trees (P_a)	h	Var(h)
SKDH	0.43	0.38	1.00	0.165	0.011
GOT	0.70	0.66	1.00	0.235	0.034
CP	0.83	0.76	1.00	0.292	0.046
			Weighted Average	0.199	

This gives a value of about 20%. No claims are made for it being unbiased.

As with previous estimates of *m* and *h* , no error is calculated as the parental allele frequencies are not known without error. This is particularly true in this case. However, it has already been shown that the European pollen contribution is not significant, so the error is not important in any case.

6.5.4 Genetic Structure Of Regenerants

Unlike in previous populations, where cytonuclear data allowed considerable analysis of the genetic structure of these populations (beyond simple allele counts), there is a limit to the fine detail that can be looked at in these samples.

Table 6-3 shows that there is no significant association between the two isozyme loci in either regenerant population, which provides one line of evidence that the regenerants are the product of random mating.

Finally, Table 6-1 indicates that SKDH is not at H-W equilibrium in the 1936 regenerants. However, this is misleading. The correct approach for purposes of determining whether the regenerants are the random mating product of their parents is a global test. This can be

achieved using a Bonferroni procedure as in chapter 5. In this case, as there are two (approximately) independent tests, the correct significance level for each test is $1-(1-0.05)^{1/2} =$ 2.5%. The p-level for SKDH in the 1936 population was just below 5%. Thus, if correct statistical procedures are adhered to, there is no evidence that these populations are not at H-W equilibrium. Again, there is no evidence that the regenerants are not the product of random mating.

Without the cytonuclear data which, in hindsight, should have been collected despite the difficulties, it is not possible to perform the same in-depth analysis of the multi-locus genotypes as was carried out in chapter 5. Had such an analysis been performed, and still failed to detect non-random associations, more confidence could be put in the conclusion that these regenerants have no genetic substructure and are simply the products of random mating amongst their respective parental populations.

6.6 Conclusion

This survey was somewhat limited in its scope, partly because of the disappointing loss of data leading to low sample sizes, loss of RAPD information, and failure to collect cytonuclear data. However, there are interesting conclusions to be drawn.

The procedure adopted means there is no evidence for a pollen contribution from elsewhere on the estate. However, this may well be due to the larger than expected sample sizes necessary to detect such a contribution, and it would require further testing to eliminate the possibility that anything up to 25% of the regenerants were fathered by pure European trees. A least squares estimate suggests that, if further data are to be collected, large enough sample sizes to detect an immigrant pollen proportion of 20% should be analysed. Furthermore, it was discovered during the analysis that isozymes alone would provide a more sensitive means of detecting pollen that chloroplasts alone.

The hybrid stands are viable and fertile, and can produce plenty of regeneration with little or no contribution from other stands. Seed dispersal from other stands is very low, and pollen dispersal is, if not necessarily low, non-significant in this sample.

Further work on determining the evolution of the hybrids on the estate would require a larger sample of regenerants, preferably with cytonuclear data. Characterisation of flowering times would also be useful in determining whether the non-significant pollen input could be due simply to dispersal ranges or, more interestingly, whether a genetically determined mating barrier exists between hybrids and parents. Such a barrier, though very unlikely to be evolved

as such (given that the two parental species have evolved in allopatry), might well increase the opportunity for the hybrids to evolve as a separate race.

The results highlight that even in situations where hybrids and parents are in close proximity, the fate of a hybrid population can only be determined if its mating system and dispersal mechanisms are well characterised. In this particular instance, the data are not conclusive, but the possibility that the hybrid populations could evolve with little input from either parent population is clear.

7. DISCUSSION

7.1 Introduction

The work presented in the previous chapters can be discussed by considering four questions. Firstly, how did the various methods used perform in this study? Secondly, what has been learnt about the nature of the larch on the Atholl estate? Thirdly, what is the likely future of larch on the estate, and indeed in the whole of Scotland, in the light of this knowledge? Finally, what are the consequences of this study for the general question of plant hybridisation, and in particular how should it be approached?

7.2 Tools For The Study Of Plant Hybridisation

A wide range of molecular methods have been used in this study. Many technical difficulties were found and much work was put into researching or testing various methods for detecting variation, and it is these and their implications that will be discussed first.

7.2.1 Morphological markers

Despite some morphological work performed on the two parental species, it was never deemed appropriate to extend this to the field study. This was partly because the cone and seed characters depended on a good cone crop for collection purposes, microscopic characters proved prohibitively difficult to utilise on a large scale, and other morphological characters were simply not discriminatory enough to justify further work.

There is also a general problem with utilising morphology to study hybrid zones. If the makeup of the zone is unknown, it is, in most cases, impossible to use all but the most powerful traits to classify individuals. A study that can convincingly use morphology to discriminate F_1s, F_2s and backcrosses in the wild is not known to this author, and arguing that because an individual tends to one parent it cannot be, for example, an F_1, is clearly suspect. Morphology is thus of limited use in determining the nature of a hybrid population- not necessarily because variation is not there, but interpretation of these traits is impossible.

On the other hand, I would suggest that if molecular work is carried out first to build up a description of the composition of the hybrid population, it is morphology and ecology that then become the most valuable tools. If molecular tools can be used to identify and describe hybrid populations, the more difficult morphological approach might allow some real insights into the nature of the selective and reproductive forces operating on the species and their hybrids. Though morphological approaches will never identify hybrid individuals reliably, molecular approaches alone may never get beyond identification.

7.2.2 Isozymes

The advantage of isozymes lie in their comparative technical simplicity. There is simply not as much to go wrong as with DNA techniques. Though, in this study, only two useful loci were discovered, these two were utilised from the outset with few problems. Inheritance patterns are usually unambiguous, and though scoring requires experience, the method is reliable in that once a system is working, the results can be reproduced without difficulty.

However, there was one unforeseen problem with isozymes. As was discovered, there appear to be differences in expression throughout the season. This makes it important to ensure collection at a constant stage in the growing season. More significant, for this study and any other on hybrids, is that the stage an individual is at in its growing season depends on its species. Thus, if collection is taken at around the point where expression is switching from one locus to another, one species and/or the hybrid may have switched or lost activity prior to the other. This could, plausibly, introduce a systematic bias into the study whereby one species consistently fails to produce results leading to an under-representation of that species' genotype. This is not speculative, as this occurred at one stage in this project when none of the pure Japanese trees gave an isozyme pattern consistent with expectation. The following year, after resampling earlier on in the season, the results were as expected.

That environmental factors could effect the phenotype scored is, of course, a well known phenomenon as far as morphological traits are concerned. It is known to occur with isozymes, but precautions are rarely taken. In future, not only should the inheritance of isozymes be verified, but their temporal stability as well. Given this precaution, a screen for variation between two species using isozymes can be done in two weeks and is a sensible starting point for any such study.

7.2.3 DNA Isolation

One benefit of DNA techniques is that once DNA has been isolated (generally the most time-consuming part of such a study) it can be used extensively, in a variety of ways, without the need for re-collection or re-extraction. At least, that is the theory. In reality, this does not emphasise the prime importance of the DNA extraction step itself. Students of model organisms forget just how much method development has gone into the most basic steps geared towards that species alone. With hindsight, one of the greatest improvements to this project would have been the early development and optimisation of DNA extraction techniques, rather than just finding one that worked. The great advantage of DNA techniques, that the same source can be used as the start for a variety of studies, presupposes that the

same DNA is suitable for all studies. In this case, it was discovered that CTAB extractions were ideal for restriction analysis, but, with this species and this protocol, unsuitable for RAPD analysis. Thus, some samples needed to be re-extracted several times, and some could not be re-extracted at all.

This is not a problem confined to this project alone. Often, it seems, more time is spent trying to overcome the problems associated with poor quality DNA than in solving the root of the problem itself. Trees, and coniferous trees in particular, are often harder to work with because of high levels of tannins and polyphenols in their leaves and needles. The time of collection also appears to be critical. It is suggested that if one species is going to be studied intensively, it is prudent to spend several weeks, if not months, purely on optimising DNA extractions, to avoid settling on a method that may be adequate at the start but turn out to be inadequate subsequently.

7.2.4 RAPDs

A great deal of time and effort was put into developing RAPDs for use with these samples. RAPDs were not originally chosen for this study largely due to their dominant inheritance, a problem that has already been discussed. However, when it became clear that more nuclear markers would benefit the analysis, RAPDs were chosen as the only feasible method to apply to a species for which no other nuclear markers exist.

The technical difficulties involved with a RAPD analysis are many. First and foremost, they are extremely sensitive to the quality of the DNA template. It is this factor that hindered their successful use the most. As mentioned earlier, the ideal solution to this would have been to spend a considerable amount of time optimising the DNA extraction for use with RAPDs. At this stage, however, the samples had already been extracted for use with other DNA methods and it was decided to try to optimise the RAPD analysis for use with these samples. The consequence of this was that the regenerants simply failed to amplify, presumably because of differences in the leaf cell biology of young trees or the stage in the growing season that they were collected. Secondly, reproducibility of many of the bands was poor. It is acceptable in a survey of this kind to simply choose markers that do appear reproducible from run to run, but that there is such variation at all must be a cause for concern. This is partly responsible for the third major problem, that of scoring. Scoring gels from run to run is often near impossible, as the overall profile may be substantially different even though the markers of choice are reproducible.

These technical difficulties aside, as many highly successful RAPD surveys have been carried out in the past, it remains to question the suitability of the technique from a theoretical viewpoint. The dominance of RAPDs, if nothing else, means that even a survey of genetic variation would require 10 times the sample size of a co-dominant marker study to produce the same accuracy of results (Lynch and Milligan 1994). In a study of the structure of a hybrid population, however, the problem is more severe. If each marker is considered as being a single locus, it is impossible to determine the heterozygosity of any locus. How severe this problem is depends on the nature of the project, and partly on luck. If each species can be found to have several species specific RAPD markers fixed within them, it would be possible to use them to distinguish between F_1s, parents and backcrosses- assuming no competition and reliability of amplification. In this project, this did not occur, and the RAPDs could not be used even to distinguish parents from F_1s.

Another difficulty associated with the use of RAPDs, again owing to their dominance, is that the mathematical tools for analysing data are not well developed. This means that many models which can be applied with co-dominant markers cannot be easily modified to include RAPDs- this was found when it became desirable to develop the best fit model categorising each individual into genealogical classes. At the outset of a project, this may seem like a distant and surmountable difficulty. When faced with actually having to make sense out of the data, it often appears that the analysis would have been best performed on far less, but co-dominant markers.

Indeed, the most common and successful method for analysing RAPD data is the use of PCA or AMOVA, neither of which can be easily fitted into genetic models. Rather, they treat the RAPD profile as a phenotype and analyse it in this sense. This is suitable for most studies, but not where the difference at a single locus can be the difference between two genealogical classes.

The demonstration that if DNA from two hybridising species is mixed, the profile is not additive (Hallden *et al.* 1996) is the most convincing demonstration of the lack of suitability of RAPDs in studies of hybrids. If a band can disappear simply due to competition from an almost identical genome, there is little hope of being confident in interpreting the RAPD profile of a hybrid.

One only needs to read a few papers on RAPDs to realise that a large proportion of them draw conclusions pertaining only to the inheritance and suitability of RAPDs for further studies rather than to actually tackling the problem at hand. Considering the study of hybrid

populations, it may be more profitable to spend time developing a single co-dominant marker, using sequencing where necessary, than collecting and attempting to analyse dozens of RAPD markers. The data set will look a lot smaller, but the conclusions that can be drawn will be far more robust.

7.2.5 Organellar Markers

Of all the methods and markers considered, the organellar markers have been the most informative. This is not surprising, as the processes governing natural hybridisation depend on the mating system, and the mating system is most easily determined (from a genetic point of view) by considering uni-parental markers. Much of the further work that will be suggested could be carried out using only organellar markers, and it is the fact that two such markers exist in Larch that make it such an interesting organism for the study of hybridisation.

However, it would be useful to be able to analyse this variation without the time consuming, expensive and hazardous use of radioactivity. Blots can be very variable, and often hard to interpret. A worthwhile extension of the work on organellar markers would be to locate the sequence responsible for the variation, design primers to this region, and perform PCR-RFLP rather than Southern blotting and RFLP.

Whatever the technical approach used, it is hard to see how a molecular analysis of hybridisation would be complete without making use of organellar markers- mitochondrial in animals, either in most plants, and both of them in species such as larch where the inheritance is through both seed and pollen.

7.2.6 Cytonuclear Disequilibria

I include cytonuclear disequilibria in a section about the tools available to researchers deliberately. It was my belief, deriving from their treatment in the literature, that they describe, in a fundamental way, associations between organellar genes and nuclear genes. Thus, by determining the values of disequilibria, and even testing them, some of the essential points of the genetics of a given system are discovered.

My original, naive approach is not uncommon. However, it is clear from this work that this is far from true. Cytonuclear disequilibria summarise *some* of the associations that exist *in that population*. Nothing else is claimed for them by those who developed them, but there appears a danger that they can be used to confuse rather than enlighten. Disequilibria can be due to a variety of causes, and are affected by population sizes and allele frequencies (even when normalised disequilibria are used), not just those more interesting influences such as selection,

assortative mating and so on. Testing of disequilibria is particularly confusing, as the sample sizes needed to justify using the approximations that are made are often extremely high. This study demonstrated that it is possible to use organellar and nuclear data in conjunction without attempting to parameterise them into cytonuclear disequilibria (chapter 5). That work which requires both nuclear and organellar data does so not because cytonuclear disequilibria are of importance, but because associations between nuclear and organellar genes may elucidate the mating system of the population. Cytonuclear disequilibria have not been shown to be the best device for *presenting* such associations, and they are definitely not the only or even the best way of *testing* for them.

7.3 The Status Of Larch Plantations On The Atholl Estate

The Atholl estate is extensively planted with European larch as well as plantations designated as hybrid larch, with no deliberate planting of Japanese larch apart from ornamental trees such as those on The Avenue, which have previously been studied in depth morphologically (Laing 1947). These Japanese trees were examined using molecular tools and shown to share mitochondrial, chloroplast and nuclear markers which distinguished them from some old but healthy European trees planted in proximity.

Given that, until recently, Japanese larch was not planted on the estate apart from these ornamental trees, it could be assumed that all hybrids on the estate descended from these individuals on the maternal side. The paternal parentage of the hybrids could not be so definitively assigned, as pollen could have arrived from anywhere in the vicinity, though it was considered likely that the nearby trees on Kennel Bank would supply the majority of the hetero-specific pollen reaching these trees.

Bearing in mind these considerations, a supposed F_1 stand (which formed the stock for two further generations) was examined. However, it was shown in chapter 4 that this stand contained no hybrids apart from obvious regenerants. This leads to the conclusion that previous descriptions of individual stands as hybrid could not be relied upon to give a true indication of their nature. Thus, even before any attempt to unravel the process of hybridisation was embarked upon, it was realised that the (necessary) use of semi-natural material (as opposed to, for example, provenance trials) introduced an added layer of complexity. This made it necessary to change the focus of the project somewhat. Rather than going straight to an examination of the process of hybridisation on this estate, it became necessary to analyse and qualify at least some of the hybrid plantations. Owing to this, the project could not focus on process as much as it was hoped. Rather, it should be seen as the

necessary backdrop to further work to determine in detail the processes governing the consequences of forced genetic contact.

Subsequent study sites were chosen on the basis of age, health, suspected planting history, and quantity of regeneration. These criteria led to two adjacent stands planted in the thirties being chosen for analysis. Both these stands were found to be more in line with expectation, having a large hybrid nature, though they differed in the genetic structure present. Several assumptions were made in order to form hypotheses about the nature of these sites. One was that they were each formed from seed collected from a single site. This is the most suspect of the assumptions used, as it may well not be the case. This is a particular issue with the 1933 site, which was shown to contain many individuals that could not be F_1 hybrids. The analysis demonstrating the fit of this stand to one derived by random mating from an F_1 generation holds only if the parents were from a single population. If this assumption is violated, it remains the case that the two stands differ in composition and that the 1933 stand cannot be F_1, but further conclusions about its nature may be invalid.

The study of regeneration in these two stands demonstrates clearly that the hybrid populations are fertile and viable, up to the point where there was no detectable genetic input from other populations. If this regeneration is indicative of the hybrid seed still being collected and planted on the estate, it is clear that the hybrid potentially has a future and will not, as might be supposed, simply be eliminated by the more numerous European trees on the estate.

The planting of these hybrids also has implications for the effective fecundity of Japanese trees. Though there has been no deliberate large-scale planting of Japanese larch on the estate (in contrast to Scotland as a whole), many of the trees in hybrid stands were shown to be pure Japanese trees. Thus, the planting of hybrids has not only increased the range of Japanese genetic material but of pure Japanese trees as well.

In conclusion, hybrids of more than the first generation are being produced and planted on the estate leading to the real possibility of an evolutionary future for them. A more detailed picture of the nature of more stands, using the tools developed in this project, would clarify the background essential to the analysis of the process of hybridisation.

7.4 The Future Of Hybrid Larch On The Estate

A major point from chapter 1 was that a knowledge of the early processes of hybridisation should allow a more accurate analysis of the pattern of hybridisation pointed at by many phylogenetic studies. One purpose of this study was to determine the consequences of hybridisation immediately after first contact (in a specific case) in an attempt to provide at

least one example of hybridisation that may, ultimately, lead to such a pattern. It is therefore useful to attempt to extrapolate the results of this study to produce a plausible prediction for the evolutionary future of hybrid larch at Atholl. This extrapolation does not take into account the most important factor, that is the desired forestry outcome (which is reducing the emphasis on all larch in any case). Rather, it is an attempt to put the results of the study into an evolutionary context.

7.4.1 An Evolutionary Dead-End?

In fact, there are several plausible outcomes. The first is that the numerical superiority of European larch, combined with some supposed hybrid inferiority in later generations will result in the hybrid being simply an ephemeral phenomenon, existing for only as long as it is deliberately produced and planted. This situation would be similar to that found when two allopatric species meet and produce less fertile or viable progeny in the contact zone. The evolutionary significance (or lack) of such zones has been discussed in chapter 1. In fact, it was evident from the examples of hybrid zones discussed that whatever the pattern of molecular markers, it was extremely difficult to find convincing evidence that the evolutionary effect of hybridisation would be significant (Goulson and Jerrim 1997; Paige, Capman and Jennetten 1991; Watano, Imazu and Shimizu 1996). Only in one case, the irises (Arnold 1993b), was there any really strong evidence that hybridisation was having an effect on the distribution and nature of the flora in the vicinity. Similarly, this situation may be analogous to that found in many animals where the existence of hybrids is immaterial to the evolutionary path of either parent and unlikely to lead to a novel hybrid taxon.

7.4.2 Creation Of A Novel Hybrid Species

While this study has not looked at viability and fitness of hybrids, an extremely demanding undertaking in any case particularly given the prohibitive life-history of trees, there is evidence that the above 'dead-end' scenario is, though plausible, not likely. This situation differs in two fundamental ways from the usual natural situation. Firstly, both European larch and Japanese larch are exotic, and therefore not especially well adapted to conditions in Scotland. Indeed, the forestry interest in the hybrid is due to the fact that both pure species have traits that are unsuited to Scotland. It is exactly this sort of situation that should allow hybridisation to have a 'creative' role in evolution. It is analogous to the cases where hybrids appear to proliferate in disturbed environments to which neither parent is adapted (Anderson 1949). The hybrids would possess more genetic variation and, if isolated from both parents, may well adapt to

Scottish conditions far quicker. Without studies on selective factors, which are unfortunately largely impractical in this case, this could not be proved. However, it is a major difference from the normal situation where both parental species can be presupposed to be relatively well adapted to the conditions.

Secondly, seed from hybrid populations has been preferentially planted. This is analogous, in the natural situation, to ensuring that first generation hybrids have a large selective advantage. This ensures a greater number of hybrids for natural selection to operate on, increasing the chance that some hybrids that outperform either parent are established. That this could occur is shown by the fact that the regeneration from a second generation hybrid swarm appeared to have involved little or no input from the nearby European trees (chapter 6). This suggests that there would be ample opportunity, in the absence of other influences, for the hybrids to evolve without much competition or input from the parents.

It could therefore be argued that hybrid larch on the Atholl estate may rapidly evolve into a recognisable taxon, distinct from its parents, in much the same way as *I.nelsonii* has in Louisiana (Arnold 1993b) where it exists in close proximity to its parents. The Atholl hybrid has been artificially established in large numbers and at least two populations have shown to be relatively resistant to seed or pollen flow from the European parent, overcoming one difficulty to the formation of a long-lived taxon. The novel conditions relative to either parent certainly provide an opportunity for selection to be effective. The interesting point is that this is plausible even when one parent- the European- is planted in large numbers and close proximity. This shows how important one aspect of further work is- to examine in more detail the seed and pollen dispersal of the trees, for without this, the essential ecological data emphasised in chapter 1 is missing.

Finally, one tantalising possibility which should be addressed is that though the low genetic contact between hybrid and parent may simply be due to low seed and pollen dispersal, it is possible that there exists another mating barrier between hybrid and parent. This is by no means implausible. Even a small difference in, for example, flowering time (a difference which *is* present between the two parents) could reduce the ability of the trees to form backcrosses with the European parents. This is entirely speculative and would need careful testing, but if true, could strengthen the case for the Dunkeld hybrid as an incipient hybrid species, and one that could be described as a native British species.

7.4.3 Creation Of A New Variety Of European Larch

The scenario outlined above is made more likely if the genetic contact between the artificially established hybrid populations and the European populations is low. It should be remembered that this is not a prerequisite- that genetic contact does not prevent separate evolutionary trajectories has been emphasised before (e.g. Goulson and Jerrim 1997). However, there is no doubt that isolating hybrids from their parents enhances the chances that they diverge from them.

That this contact is surprisingly low has already been suggested, though not proven beyond doubt, by the results of Chapter 6. However, it may be that a more likely scenario is that pollen and seed flow, though *statistically* insignificant, will be enough to homogenise the hybrid and European populations. As collection of hybrid seed consists of taking seeds from any population designated as hybrid and replanting them, it would be expected to become progressively 'more European'. Eventually, the larch at Atholl would be a more or less homogenous group, with a largely European genetic makeup. It would undoubtedly contain neutral genes, and most likely have some selective Japanese traits as well. In essence, it would be a variety of European larch that has adapted to the conditions at this locality. An important point is that it will be isolated from the vast majority of the natural range of either parent and is very much a localised phenomenon. The rest of this section will assume that, many generations down the line, a future biologist is attempting to classify this putative 'Atholl larch'.

How this taxon, if it arose, should be viewed depends on ones approach to classification, and continuing the approach from Chapter 1, I would suggest is an insurmountable problem using traditional classification schemes. One approach would be to classify it as a new hybrid species, based purely on the fact that it would contain genes (selective or neutral) from two species. This appears to be the approach taken by the majority of authors (e.g. Rieseberg and Ellstrand 1993). This is justifiable only if the utility of species classification is simply to describe phylogenetic history.

A second approach would be to describe this as a new race, variety, population or other taxon of European larch that has arisen by introgression of Japanese genes into European larch. For many authors, this is logically equivalent to the approach taken above, introgression and inter-specific hybridisation being extremes of the same continuum (Abbott 1992). Alternatively, if it is desired to maintain a special place in classification for hybridisation, it may be better to simply view hybridisation as a mechanism other than mutation for introducing new genetic

variation into a taxon. The Atholl larch remains European larch for as long as the criteria for that species, which must be independent of the fact that some individuals contain genes derived from another species, are satisfied. However, it would be a race that has evolved with the additional influence of introgression. In this case, hybridisation has clearly been influential, but it is not helpful to describe European larch or (the imaginary) Atholl larch as a hybrid species.

One possible problem pertains to the cytoplasmic genes. Even were selective Japanese genes largely eliminated from the gene pool, it would be likely that the population would maintain a large number of Japanese neutral genes and mitochondria, which have been found in all individuals of hybrid populations examined. Were the future biologist to examine the mitochondria of the European larch at Atholl, she might rightly conclude that the European larch had, at this location, 'captured' Japanese mitochondria (the exact analogue of chloroplast capture in most plants, where the chloroplast is maternally inherited). However, if the reasoning of many authors (e.g. D'Orado 1992) was followed, explanations for how such a mitochondria had established itself and spread throughout a population would be sought. The taxonomic 'purity' of European larch might also be questioned, the Atholl larch demonstrating the importance of hybridisation in the evolution of the genus *Larix*.

In reality, the Japanese mitochondria has not spread throughout a European *population*, rather it has become associated with European *genes* - the 'nuclear capture' process described in chapter 1. Secondly, without examining variation at the nuclear locus, the special status of the mitochondrial capture as a process that requires explanation would be void. Finally, any temptation to review the *classification* of this Atholl larch based on the discovery of these Japanese organelles should be treated with extreme caution. If the criteria for species do not specify (as they do not and should not) a particular organellar haplotype, then this should not be used as a basis for a classification. The biologist could infer that hybridisation with Japanese larch may have happened, but not that it had been important.

In summary, I believe that the (unintentional) production of a distinct variety of European larch at Atholl is probable within a few generations. Furthermore, this variety would better be considered to be a European larch that has evolved adaptively and neutrally through introgression from Japanese genes, rather than to be a hybrid species.

7.5 Future Of Hybrid Larch In Scotland

Most of the arguments applied above apply to the consideration of hybrid larch throughout Scotland (and, indeed, Britain as a whole). However, a major difference between the

distribution on the Atholl estate and the majority of other plantations in Scotland is obvious from Table 2-3. Whereas on the estate the majority of non-hybrid larch is European, the general pattern elsewhere is that it is the Japanese larch that is planted in bulk. In fact, much of the larch around the country is designated simply as 'hybrid/ Japanese larch'. This means that not only is there potential non-hybrid genetic material from Japanese rather than European larch, it is likely to be planted in even greater proximity to the hybrid itself.

The consequence of the greater intermingling is that the formation of a distinct hybrid species is made less likely. Furthermore, it has not been verified that a hybrid population will not backcross extensively to neighbouring Japanese populations as they do not with European populations. However, it remains possible that in Scotland as a whole, a new hybrid species could evolve, either through natural or artificial selection providing selective advantages for individuals with combinations of traits from both exotic species.

More likely is the widespread creation of a distinct Scottish race of *Japanese* larch. Unlike at Atholl, where the predominant genetic background is European, the rest of Scotland would involve the backcrossing of the widely spread hybrid with largely Japanese parents. The viability of hybrid populations has been experimentally verified, as has the ability of hybrid individuals to backcross to Japanese parents (Chapter 5). It seems inevitable that the end result in Scotland will be a variety of Japanese larch with a number of European nuclear genes and a proportion of individuals with European organellar genes. Whether this will allow adaptation to Scottish climes is uncertain, but it does seem likely that such widespread genetic contact will result in the Japanese larch absorbing, for example, the winter-frost tolerance of the European species.

The most likely outcome of larch planting in this country since the turn of the century is, I believe, a rapid adaptation to the Scottish environment of both European and Japanese larch. On the Atholl estate, as well as any other estate that plants both hybrids and European larch extensively, the European larch may incorporate the tolerance of Japanese larch to late frost and larch canker, both of which cause silvicultural problems for the planting of European larch. Likewise, and perhaps more convincingly, the widely planted hybrid should allow the Japanese larch to adapt to the cold winter frosts that so limit it in Scotland at the moment. Hybridisation, through introgression, may therefore be altering the evolutionary path of larch in Scotland to an appreciable and important extent.

7.6 Further Work

There remains much work of value to be carried out both on the estate, and possibly throughout Scotland, in order to extract as much information as possible about the process of hybridisation between Japanese and European larch. Here, I will outline suggestions for such work that could be carried out without further method development.

One important extension is to extend the work on regeneration throughout the estate. The goal of this would be to determine the pollen and seed input from European larch elsewhere on the estate. To this end, only organellar genes need be surveyed. If the need to collect buds for isozymes is removed, it would be feasible to sample 100 regenerant individuals/ population (and, if desired, extend the parental sample) and test the hypothesis that the frequency of European mitochondria is zero and the hypothesis that the European chloroplast frequency in the regenerants is larger than that in the parental population (which, assuming no contribution, we would expect to be rejected). If we still failed to uncover evidence of a substantial European pollen/seed pool, the case for possible divergence of a hybrid taxon would be strengthened.

In a similar vein, it would be extremely valuable to survey regenerating European populations. This way, it could be determined if the hybrid was fathering seed in European stands. If so, we might conclude that the idea that hybrids are less viable than their parents is false.

The process of hybridisation could be further dissected through analysis of the seed of the regenerating populations. A particularly interesting question would be whether the seed from the pure Japanese trees in the populations had the same paternity as the seed from hybrid individuals. A greater proportion of Japanese chloroplasts, for example, in the seed from pure trees would suggest some degree of pre-zygotic mating barrier, such as a flowering-time effect. Though a very large sample and a very large effect might lead this to producing detectable and significant cytonuclear disequilibria that could be observed using the approaches already described (i.e. examining the pattern of variation present in regenerants), this more direct analysis of seed would be more sensitive. Furthermore, only the chloroplast of the fertilised seed need be analysed if this approach is used, rather than using nuclear markers to detect cytonuclear disequilibria.

Moving beyond the estate, it would be valuable to perform a wide-scale survey of the 'Japanese/Hybrid larch' in Scotland, and determine unambiguously how much of this is hybrid and how much Japanese. This in turn could identify areas where it would be valuable to examine the process of hybridisation in the detail described above.

Very importantly, but omitted up till now because of the practical difficulties and long term nature of such an experiment, fitness analyses of the parents and various classes of hybrids should be carried out as a starting point for ecologically relevant hypotheses. A restricted version of this approach would be using controlled crosses to produce seed and perform an assay for germination and viability after 1 year of the two parents and their F_1s, using the same individuals as a starting point. This would allow assessment of three important factors. Firstly, the fertility of each of the parents and the F_1 with respect to pollen from either of the parents or the F_1 (for example, is the F_1 more receptive to other F_1 pollen or to a particular parent). Secondly, the germination rate of pure, F_1, backcross and F_2 seeds. Finally, the early viability of each of these classes. While by no means an exhaustive or ecologically detailed fitness analysis, this approach -feasible in a short-term timescale- allows at least some informed discussion about possible hybrid viability/ fertility effects. That the controlled cross is carried out correctly could be assayed using SKDH and the K140 chloroplast marker. A version of this approach is the reason that the study of hybridisation in Louisiana irises was so successful, and as has been argued, is an important element of hybridisation studies.

Finally, the approach of computer modelling has demonstrated its utility at analysing the behaviour of genetic parameters. It is possible that this simple model could be used to provide means and variances of disequilibria under many conditions, including selection and migration, and all forms of 'unusual' situations can be simply incorporated as the starting conditions. Furthermore, it may also be possible to determine a range of selective values, migration rates (from parents) and population sizes under which a recently formed hybrid population might diverge from its parents. Time constraints have meant this has not been possible, but in principle this might be achieved using this or a similar approach.

7.7 Approaches To The Study Of Plant Hybridisation

Specific suggestions as to how to extend this project have been made above. There are also some general points arising that are pertinent to the study of plant hybridisation in general.

The first point is that no study designed to examine the dynamics of hybridisation is complete without an analysis of multiple generations. Indeed, the main weak spot of this project was that the need to identify new study sites after the first was found to be unsuitable severely limited the time available to spend on the regenerant population (in particular, it was not possible to go on to develope the techniques to look at seed genomes). Even so, it was this (limited) analysis of regenerants that promises the most in the way of fruitful study. Without this, the results of the survey, though interesting, are confined entirely to the Atholl estate with

no broader application, as they relate only to the *pattern* present at that particular site. With it, the ongoing process of hybridisation can begin to be dissected.

It might seem obvious that analysis of more than one generation is important. However, as discussed in chapter 1, not all studies even acknowledge that such work could add qualitatively to the value of the results. Probably because of an emphasis on the modelling of hybrid zones, researchers have often been preoccupied with the difficulties in inferring process from pattern, without attempting to examine the process itself.

The situation at the Atholl estate does, of course, lend itself readily to the analysis of multiple generations as there is such a distinct age structure (in my study sites, all trees were either 60 years old or under 5). This is clearly rare in nature, but the importance of this approach is that ways round this should be found. One would be to estimate the age of individuals and see what patterns exist in the temporal dimension. Also, and the only route available if the plant under consideration is an annual, examination of seed genotypes and germination rates could simultaneously look at the genetic changes occurring in the population (or indeed, determine if it is reasonably stable) and provide some valuable fitness information.

A second point is that, where possible, ecological data must be collected. None has been attempted within the confines of this project. Indeed, it might be thought that there is no useful data to collect, as the trees have been planted and managed. However, it would be helpful to know, in detail, the flowering times of each individual and relate that to its genotype, as the two parents do differ. With insect dispersed plants, the behaviour of the pollinators could be very valuable. Without in depth fitness analysis, mating-behaviour is about the only life-history trait that can be analysed in depth.

In general, it appears that the tools used to study plant hybridisation are far more sophisticated than the approaches used. The same tools that are used to study hybrids themselves (be they genetic, ecological, or morphological) should be used to analyse hybridisation. While this project itself does not go much further than some, it is considered to be simply the beginning of a study of hybridisation. Previous work often sees the description of morphological or genetic variation in hybrid zones as the end, whereas it should be just the start.

Appendix A: Materials And Methods

.i DNA Extraction Techniqes

.a CTAB Based Extraction

The most common DNA extraction methods, at least in conifers, are based on the use of Hexadecyltrimethyl ammonium bromide (CTAB). CTAB solubilises DNA in high salt concentrations, but makes it insoluble when salt concentrations are low. This property is the basis of the original CTAB extraction method (Doyle and Doyle 1987), which used a large drop in salt concentration to desolubilise and spin out the DNA. A simplified method was developed (Whittemore and Schaal 1991), which simply uses the CTAB to solubilise the DNA and extractions are performed with the DNA in solution. PVP (Polyvinylpropylene) beads are added, which bind to and help remove many carbohydrates.

A version of this method was being regularly and successfully used to extract DNA from Scots Pines needles (Morman 1994) but did not work in larch. DNA yield was extremely low (sometimes not visible on gels). However, the replacement of the organic solvent chloroform with the related dichloromethane, as suggested from work with tropical trees (Gillies *pers comm* 1995), resulted in high yields. The reason for this is unknown, but it may suggest a carbohydrate or secondary compound of some form that solubilises DNA in chloroform but not DCM. To formulate an improvement on the extraction method would, presumably, require research into the nature of this hypothetical contaminant.

Following the initial extraction, a phenol based clean-up step was undertaken (Sambrook *et al.* 1989). This step removes proteinous contaminants (visible as a white layer) but its contribution to the suitability of DNA for amplification or digestion was uncertain.

This method, replacing the chloroform in the lab protocol with DCM, is described below.

Materials :

2 x CTAB buffer
DNA wash buffer
1 x TE
3M Sodium acetate
RNaseA (5mg/ml)
DCM (dichloromethane)
Phenol:Chloroform 1:1
IPA (Isopropanol), stored at -20°C
Absolute ethanol
Liquid Nitrogen

Method:

The water bath and 2xCTAB were pre-warmed to 65°C. Meanwhile, 1-2 g of frozen needle from each sample was ground to a fine powder in a pre-cooled mortar and pestle using liquid nitrogen. Each sample was transferred to a 15ml polypropylene tube with a chilled spatula and stored at -80°C until all the samples were prepared and the water bath was ready. Then, 5ml

of warm 2X CTAB buffer was added to each sample and mixed well before incubating for 45 mins.

After incubation, the tubes were removed and cooled to room temperature (this step was necessary to avoid a violent reaction on addition of DCM). Extraction of proteins and carbohydrates was carried out by adding 5ml of DCM, mixing well, and centrifuging at 3,200 rpm on a bench-top centrifuge for 10 mins. The top aqueous phase, containing the DNA, was decanted (or removed using a pipette, depending on the nature of the interface) to a new tube, and the DCM extraction repeated a second time to further clean up the preparation. The supernatant was again removed to a new tube, and the DNA precipitated by addition of an equal volume of ice-cold IPA. This was stored at -20°C for at least 1hr, or overnight if required.

If the DNA was clearly visible as a fibrous precipitate it could be 'spooled out' with a shaped glass pipette. Otherwise, and more usually, it was spun for 10 minutes to pellet the DNA and the liquid poured off. The DNA was transferred to a 1.5ml eppendorf, and suspended in 1 ml of wash buffer for 10 minutes, before centrifuging for 10 minutes in a micro-centrifuge. This step removes IPA and excess salt. The wash was decanted and the pellet air-dried for approx. 30 minutes, and then suspended in 480µl of 1xTE and 20µl of 25xRNAase. This was incubated at 37°C, for 1-2 hours or until the pellet was fully dissolved.

This crude extraction could be followed by repeated Phenol-Chloroform extractions (Sambrook, Fritsch and Maniatis 1989) to remove protein, ensuring a final extraction with just chloroform to remove residual phenol.

.b Scotlab Extraction

The parents, crosses and the 1912 'hybrid' stand (chapters 3 and 4) were extracted as above, and these extractions were used successfully in RFLPs and some PCR work. However, CTAB extractions on a trial batch of the individuals from the 1933 and 1936 stands (chapter 5) resulted in DNA that was very variable as regards to its ease of digestion with restriction enzymes. Why this occurred is unknown, though it now seems likely it was due to the stage in the season that the collections were made. The collections from the 1933 and 1936 stands were made earlier in the season, and though the needles appeared fully expanded there may have been some biochemical difference between these younger needles and the slightly older ones used before.

It was decided to try a kit designed for the extraction of DNA from plants. This kit, the Scotlab phytopure kit, used SDS in the lysis phase and modified silica beads to bind to plant

carbohydrates. Though the yield from this method was much lower than that from the CTAB method, the DNA was consistently easier to digest. The method shown below is taken from the manufacturers instructions (Scotlab 1996), with one modification. DCM has replaced chloroform in the initial extraction steps, as this seemed to improve the yield as it did with the CTAB method.

It was realised that much lower quantities of DNA were needed than had been achieved from the CTAB extractions on 2g of needle. Thus, only 0.2g of needle tissue was used. However, the kit was successfully used on larger scales similar to those used in the CTAB extractions.

The method has been slightly modified from the manufacturers instructions, and is therefore described below.

Scotlab DNA extraction of Larch DNA

Materials:

Nucleon Phytopure kit (Scotlab)
- -Reagent A (SDS based lysis buffer)
- -Reagent B (salt solution for lysis)
- -Silica Suspension (binds impurities)

1 x TE

DCM (at 4°C)
IPA (at -20°C)
Ethanol
Liquid nitrogen
DNA Wash buffer

Method:

The water bath was pre -warmed to 65°C. Meanwhile, 0.2g of frozen needle tissue from each sample was ground to a fine powder, in liquid nitrogen and transferred to a 1.5ml eppendorf with a chilled spatula. 500µl of Reagent A was added and mixed well, followed by addition of 160µl of Reagent B. These combined to form the active SDS lysis buffer. This was inverted several times until homogenised and incubated, with frequent shaking, at 65°C for 20 mins. After lysis, the samples were incubated on ice for 20 mins, which increased the relative solubility of carbohydrates and proteins in the organic phase. An initial extraction with an equal quantity of DCM (approx. 750 µl) was performed, spinning at 13,000 rpm for 10 mins to separate the cell debris and organic solvent from the aqueous layer. The aqueous supernatant was removed to a fresh tube, and 40µl of silica suspension and another equal volume of DCM added. This was shaken gently for 10 minutes to allow the modified silica to bind carbohydrates, and the organic soluble compounds to dissolve. Gentle spinning at 2,000 rpm was carried out, and the aqueous phase removed with a pipette, avoiding the silica rich interface.

As with the CTAB extraction, the DNA was precipitated with an equal volume of IPA, washed, and dissolved in TE (50µl). RNAase was no longer used, as tests indicated it did not affect either PCR or restriction digestion.

.c Plasmid DNA Extraction

The use of the K140 chloroplast probe, as well as the PTEE5 ribosomal probe, required the purification of plasmid DNA from *E.Coli* hosts. The basis of this method is that after lysis with a suitable alkaline solution, the bacterial DNA will remain associated with bacterial proteins and the cell membrane and precipitate whereas the lighter plasmid DNA will remain soluble. The method used was the alkaline lysis method of Maniatis (Sambrook, Fritsch and Maniatis 1989)

.ii Agarose Gel Electrophoresis And DNA Visualisation

After DNA of any sort was extracted, or manipulated (e.g. digested, amplified etc.) agarose gels were used to visualise DNA (to estimate concentration or determine the success of a reaction). Agarose gels were also used for Southern blots.

The basis of this method is that heavier (i.e. longer) DNA fragments take longer to migrate through an agarose under the influence of an electric field. By using ethidium bromide, a compound which 'intercalates' DNA and fluoresces under UV light, it is possible to see how far each size class of DNA molecules has migrated. By comparison with a reference sample containing fragments of known sizes (a 'ladder') the length or range of lengths of DNA molecules in the sample can be determined.

Ethidium bromide is highly carcinogenic and care must be taken in its handling. In particular, large quantities should not be disposed of down the sink until deactivated by UV light. For this reason, EtBr was used as a stain rather than a component of gel and running buffer, making it manageable.

Unless specified, gels were 1% agarose. Mini gels were 8cm x 6cm, with a volume of 25ml gel and 8 wells. Midi gels were 15cm x 11cm, with a volume of 100ml and 14 wells. Gel moulds and running chambers were Scotlab's horizontal submarine electrophoresis components.

Materials:

50 x TAE
EtBr (10 mg/ml)
Agarose
10 x Loading Dye
Marker (1kb or 100bp, 1μg/μl)

Method:

Horizontal agarose gels were prepared (generally 1% w/v gels) in 1 x TAE using a microwave to heat and dissolve agarose. 20μl of sample (10μl for mini-gels), was loaded, after dilution with loading buffer and 1xTAE. They were run at a constant voltage of 90v (60v for mini-gels), usually for 3-3 1/2 hrs (1 hr for mini-gels). Staining was achieved by placing gels in a dedicated tray containing 0.5μg/ml EtBr in TAE for 40 mins (20 mins for mini-gel). The EtBr was decanted (for re-use and eventual disposal), and the gel destained (if desired) in water for 10 mins, before visualisation by UV transillumination.

.iii Polyacrylamide Gel Electrophoresis

Occasionally, greater resolution than that provided for by agarose gels was required. In such cases, polyacrylamide gels were used. For SSCP and Heteroduplex analysis, mini-gels (8cm x 6cm) were used. For larger gels (e.g. for RAPD analysis) the Hoefer SE 600 midi-gel rig was used, capable of loading 28 samples and running up to four gels simultaneously under the same conditions. Gels were set-up as according to the manufacturers instructions, with the added refinement of shining a bright lamp during setting. It was found that on dark days, the gel did not set correctly and wells were often mis-formed- the lamp corrected this problem (by speeding up the production of free-radicals in the polymerisation reaction). Gels were 8% Acrylamide, with a 39:1 bis-acylamide:acyrlamide ratio. Gels were run in 0.5xTBE buffer, at constant 20mA/gel for 3-4 hrs depending on the separation required. Gels were removed, stained in EtBr for 20 mins., rinsed, and photographed.

.iv Assessment Of DNA Concentration and Quality

Initially, the concentration of DNA was assessed using a UV spectrophotometer to determine the absorbances at 260 and 280nm (Sambrook, Fritsch and Maniatis 1989). For DNA extractions, this was achieved by diluting approximately 1μl of DNA solution in 1ml of 5mm Tris-HCl, and using a scanning spectrophotometer to measure absorbances. The concentration of the sample can be easily related to its absorbency at 260nm (A_{260}) and its purity to the A_{260}:A_{280} ratio.

However, it was discovered that samples diluted to equal concentrations on the basis of this method were often recorded as very different in concentration. It has been noted that plant

material extracted by CTAB often fails to lead to reliable UV determination and this appeared to be the case here.

Instead, concentrations were equalised by comparing intensities of fluorescence on a gel. These concentrations were compared against a series dilution of marker of known concentration to enable an estimate of the samples concentration to be made. Occasionally, the 'spot' method was used (Sambrook, Fritsch and Maniatis 1989). This involved placing, on parafilm, a series of drops of standards of known concentration with 0.5 ng/µl of EtBR added. Viewing under a transilluminator enabled comparison of the sample with the series dilution to determine concentration.

In order to provide a very rough guide to the quality of extracted DNA, approx. 0.5µg of each sample was digested with 1/2 unit of EcoR1 restriction enzyme (following manufacturers instructions), and run on an agarose gel. This enzyme, fairly sensitive to impurities, enabled very poor quality, indigestible DNA to be discovered early on and re-extracted.

.v Isozyme Gels

Starch gel electrophoresis of isozymes has been commonplace since the 1960s. It is based on the differential migration of proteins, due to differences in shape and charge/weight ratio, through a starch gel under an electric field. All of the methods for running and staining isozyme gel systems were derived from Cheliak (Cheliak and Patel 1984). However, there were variations in the procedure used to pour and set gels, which will be described below.

Gel moulds and running trays were made especially by the workshop, to the specifications of Cheliak. Loading Wicks were made by cutting Whatman 3mm filter paper into small strips, approx. 3mm x 5mm. Gels were prepared in the evening before the day of use.

Materials:

Potato starch, electrophoresis grade
Gel Buffers: 'B' and 'H'
Isozyme extraction buffer
Electrode Buffers: 'B' and 'H'
Stains: SKDH, GOT, and PGM
Bromophenol blue and Xylene Cyanol marker

Method: 27.5g of starch and 255ml of the desired gel buffer (B buffer for GOT, H for SKDH and PGM) was mixed (by shaking) in a 500ml side-arm flask. This was a slight excess of the 250ml for which the gel moulds were designed, but the excess ensured enough solution when small amounts were lost. The buffer was heated slowly over a bunsen, shaking constantly, until the starch had all dissolved and the solution was changing in consistency (becoming clear and more viscous). The solution was allowed to reach boiling point, before removal from heat. A tap vacuum-line was applied to the side-arm and a vacuum applied for approx. 30 sec. (or

as needed) to degas the solution, before it was poured smoothly but quickly into the mould. The solution was allowed to cool for 30 sec., before a plastic plate was lowered over the gel. The plate was place by firmly holding at one corner and lowering slowly to ensure an even distribution of gel throughout the mould. The gel was allowed to cool for 1/2 hr, before being wrapped in cellophane to prevent drying out and left to set overnight.

The next day, large trays of ice with small glass pestles were prepared. Approximately two buds for each sample were placed in each pestle. It was found useful- sometimes essential- to prepare the buds by removal of their scales, which appeared to contain enough sticky residue to impede solubility and mobility of the samples. This was best achieved by peeling off the bud-covering from a shoot with a pair of tweezers, and then picking or scooping out the naked bud. Subsequent procedures are exactly as described by Cheliak (Cheliak and Patel 1984). Samples were ground in 2 drops of extraction buffer, and the cellulose wicks used to soak up the sample. The wicks were wiped on tissue to remove debris, and inserted into the gel (which was sliced across, approximately 5cm from one end). Two or three samples of marker dye were also loaded to visualise progress of the electrophoresis. Gels were run in the fridge, with a safety cut-out, at 40mA constant current. SKDH gels were run for approximately 7 hours, until the slow marker had migrated approximately half-way up the gel. GOT and PGM gels were run for 4 hours, until the fast marker had reached the end of the gel.

After electrophoresis gels were sliced using a fixed cheese wire into two layers (necessary for examining two stains at once, but useful even for singly stained gels to provide a back-up). The gel, trimmed to size, was placed in the smallest tray that would hold the gel and the relevant stain poured over. The gels were incubated at 37°C until bands appeared, approximately an hour for GOT, up to 3 hours for SKDH and PGM. Scoring was achieved over a light box if necessary.

.vi PCR Amplification Of Single-copy Genes

The Polymerase Chain Reaction (PCR) (Mullis and Faloona 1987) was used to amplify single-copy genes for two purposes. First, it was used to produce ORF25 probe for use in the detection of mitochondrial variation. Secondly, it was used to amplify products which were then screened for variation directly. Here, the protocol used for amplifying such products will be described.

Materials:

Promega Taq enzyme (5 units/μl)
Promega x10 KCl buffer
Promega 25mM $MgCl_2$
10 x dNTPs (2.5mM in each nucleotide)
Doubly distilled water
Mineral Oil
1 x Primer (concentration dependent on primer)

Methods:

PCR reactions were carried out in a Perkin-Elmer 480 thermal cycler, in thin-walled 0.6ml reaction tubes. Most reactions were carried out in a 25μl volume, though 50μl reactions, where used, were equally successful.

Generally, DNA was diluted to an approximate concentration of 10ng/μl and 1 μl was used for the PCR reaction. However, it was often necessary to increase or decrease this. Depending on the purity of the template, the nature of the primer, and the accuracy of the assessment on concentration, anything between 1/10th and 10 times this concentration was used.

A 'pre-mix' consisting of the required amount of enzyme, buffer, dNTP and $MgCl_2$ (for a final concentration of 1.5mM) was made, and added to each tube pre-loaded with template DNA. Primer was the last component to be added, followed by mixing, centrifugation and an overlay of 25μl of mineral oil to prevent evaporation.

Amplification was then carried out, the programme being used depending on the primer (see Appendix Table A-1).

.vii RAPD Analysis

The PCR reaction (described above) was used at low temperatures with single, short (10bp) primers to amplify up uncharacterised regions of the genome, and is based on the method of Allnutt *et al.* (1998).

Materials:

Promega Taq enzyme (5 units/µl)
Promega x10 NH_4Cl buffer
Promega 25mM $MgCl_2$
10 x dNTPs (2.5mM in each nucleotide)
Doubly distilled water
1 x Primer (concentration dependent on primer)

Methods:

Reactions were carried out in a Perkin-Elmer 9600 thermal cycler, in volumes of 10µl. 10ng of DNA was used, to which a premix containing 1.5mM $MgCl_2$, 250 µM dNTP, 1x NH_4Cl buffer, and 0.5 units of Taq was added.

The PCR cycle used in this case was 30 secs each at 94°C, 72°C and 42°C. The cycle was repeated 40 times.

Appendix Table A-1: Programmes used for PCR amplification in a PE480 thermal cycler

Primer (approx. 100nM)	Melting (at 94°C)	Annealing Phase	Extension (at 72°C)	Cycles	Reference
ORF25	1 min	90 sec. at 60 °C	60 sec.	35	(Sinclair,1998)
COX1	1 min	90 sec. at 60 °C	60 sec.	35	(Sinclair,1998)
RUBISCO	1 min	90 sec. at 50-65 °C	90 sec.	35	-
Taberlet ab	1 min	90 sec. at 55 °C	60 sec.	35	(Gielly and Taberlet)
Taberlet cd	1 min	90 sec. at 55 °C	60 sec.	35	(Gielly and Taberlet)
Taberlet ef	1 min	90 sec. at 55 °C	60 sec.	35	(Gielly and Taberlet)
ADH	1 min	90 sec. at 45°C	90 sec.	30	(Strand, Leebens-Mack and Milligan 1997)
Calmodulin	1 min	90 sec. at 45°C	90 sec.	30	(Strand, Leebens-Mack and Milligan 1997)
Glyceraldehyde 3-pdh	1 min	90 sec. at 45°C	90 sec.	30	(Strand, Leebens-Mack and Milligan 1997)
Mitochondrial primers	1 min	90 sec. at 45-60°C	90 sec.	35	(Demesure, Sodzi and Petit 1995)

.viii Screening For Variation in Single-copy Genes

After amplification of a PCR product was confirmed by agarose gel electrophoresis, screening for variation was carried out. There were a number of ways of doing this, which are detailed below.

.a Restriction Analysis

Approximately 8µl of PCR reaction was sufficient. It was not found necessary to remove the PCR buffer before digestion, though it is possible that with some enzymes, restriction may have been inhibited to some extent. 1 unit of enzyme, sufficient buffer, and ddH_2O were added to make the PCR product up to 25µl and incubated overnight at the required temperature (according to manufacturers instructions). The digestion DNA was ran and visualised on a 2% agarose gel, aiding separation of smaller fragments.

.b SSCP

Single Stranded Conformational Polymorphism is widely used in medical research to screen for disease causing mutations (Cotton 1993), though has found its way into ecological and evolutionary genetics (Lessa and Applebaum 1993; Watano *et al.* 1995). It is based on the fact that in single stranded form, DNA is capable of taking up a secondary structure through intra-molecular base pairing. Thus, sequences of the same length and weight but slightly different sequence composition may take up different structure and have differing mobilities in subsequent electrophoresis.

Though generally a radioactive method, it was possible to modify a cold SSCP method (Yap and McGee 1992) for use as a preliminary screening method utilising available lab-equipment, as described below.

Materials:

0.5 M NaOH, 10 mM EDTA
Loading Dye
100ng PCR product
8% Polyacrylamide Mini-Gel

Methods:

Approximately 100ng (usually about 8 µl) of PCR product was mixed with 1µl of 0.5M NaOH, 10 mM EDTA and heated to 42°C for 5mins. This denatures the DNA. 1µl of loading dye was added, and the sample immediately loaded into the mini-gel. The gel was run for approx. 1 1/2 hr at 50V, until the slower dye ran off. Time between denaturing and running of the gel was minimised as much as possible. The gel was stained in EtBr for 10 minutes and visualised under UV.

.c Heteroduplex Analysis

Heteroduplex analysis (Gross and Nilsson 1995; Lessa and Applebaum 1993) is a relatively simple way of screening for sequence variation in PCR fragments. The putative variants are mixed together in equal quantities (approx. 8µl of each PCR reaction) and melted by heating for 3 mins to 95°C. On reannealing, a fraction of the resultant double stranded DNA will be 'heteroduplex; due to complementary strands from the different PCR products annealing. If there is any mismatch, a 'bubble' of single stranded DNA is formed, which retards the motion of the duplex through a polyacrylamide gel. 8% Mini polyacrylamide gels, as described for SSCP, were used. Double stranded samples run quicker than single stranded samples, so they were run for only 1/2 hr at 50V, until the marker had reached the bottom of the gel. The first two samples were controls, a PCR product from each parent being melted and reannealed. The third sample was a mix of these two. If this sample showed a slower band not present in either of the other two samples, the presence of a mismatch could be inferred.

.d Sequencing

A 50µl PCR reaction with the required primers was carried out. The product was cut out from the gel under UV with a clean scalpel blade. The slice was weighed, and the Qiagen Gel extraction kit was used, according to the manufacturers instructions (Qiagen 1997), to purify clean reaction product. This was diluted into 30µl, and 2µl was run on a 1% agarose gel to estimate concentration. For the sequencing, for a 'half-reaction', approx. 50ng of DNA was required which was usually approximately 5µl of purified product. The Amersham fluorescent sequencing kit was used according to manufacturers instructions (Amersham Life Sciences), using a PCR PE thermal cycle to perform the thermal cycling. The product was then sent to a University based sequencing laboratory where the sample was run, visualised, and the resulting chromatogram digitised and returned for analysis.

.ix Transformation Of E.coli

In order to produce probes for use in surveying for variation at the rDNA locus, plasmids containing the required probe sequence were obtained. These were transformed into *E.coli* host, which were then grown and harvested as described in Maniatis, using ampicillin and chloramphenicol media for selective purposes (Sambrook, Fritsch and Maniatis 1989).

Transformation was carried out using an electroporation method. The protocol followed was exactly as the manufacturer's instructions, using a dedicated machine. Pre-prepared competent *E.Coli* cells were mixed with the plasmid, and shocked in a 50µl capacity cuvette. These cells

were then serially diluted (neat, 1:10 and 1:100) and spread on chloramphenicol/ampicillin plates. The plasmid carried resistance to both antibiotics, and colonies were seen on all plates representing transformed cells. These cells could then be used to grow and produce large quantities of the plasmid.

.x Preparation And Labelling Of Probe Templates

.a From Plasmid DNA

Chloroplast probe K140 and rDNA probe PTEE5 were prepared by extracting plasmid DNA (pg. 179) from the host *E.coli*. The plasmid DNA was then incubated overnight with the required restriction enzyme (*Kpn*I). The product of this digestion was run on a mini-gel, and a gel-extraction kit (Qiagen 1997) used to extract the probe DNA (e.g. the 9.8 kb fragment for K140). The extracted DNA was then quantified by agarose gel electrophoresis and diluted to 100ng/µl.

.b From PCR Reactions

Mitochondrial probe ORF25 was prepared by amplifying a Japanese larch parent with ORF25 primers to produce 200µl of reaction. This was then purified using a PCR purification kit (Qiagen), and quantified by agarose gel electrophoresis and diluted to 100ng/µl

.c Labelling

Labelling was based on the method of random priming (Feinberg and Vogelstein 1983). This method uses a mixture of short oligonucleotides which randomly bind to denatured template DNA. Incubation with a polymerase and radioactive CTP (labelled with ^{32}P) incorporates radioactivity into the double stranded product. The procedure was greatly simplified by use of the 'Ready-to-go' labelling kit (Pharmacia Biotech), which uses dry beads containing all the necessary components apart from the template and the radioactive CTP. The procedure simply involves heat denaturing the required amount of probe (about 500 ng) at 94°C for 3 minutes, dilution with water, and addition of CTP and the labelling bead followed by incubation for 1hr. This labelled probe is then ready for use. Full details are given in the manufacturers instructions (Pharmacia Biotech).

.xi Southern Blotting and RFLP Analysis

Southern blotting is the process of transferring single stranded DNA from an agarose gel onto a nitro-cellulose filter to which it can be irreversibly bound. Following this, radioactively labelled probe is incubated with the filter, and binds to complementary regions of the DNA

bound to the filter. Washing removes unbound probe, and overlaying with a suitable film allows the pattern of radioactivity on the filter to be visualised.

If the agarose gel is of a restriction digest of the genomic DNA, and the probe to a particular gene, only those fragments containg a region of the gene will show up. Thus, if the gene straddles or is in a region where the samples differ in their restriction pattern, there will be a difference in the banding pattern seen on the autoradiograph. This makes Southern blotting a suitable method for Restriction Fragment Length Polymorphism analysis. Strictly, RFLP analysis does not require Southern blotting- any method that allows differences in restriction sites to be assessed is RFLP (indeed, PCR-RFLP, as was attempted above, is becoming more common).

.a Southern Blotting

The basic protocol for Southern blotting is out of Maniatis (Sambrook, Fritsch and Maniatis 1989), with some modifications.

Materials:

Denaturing solution
Neutralising solution
20x SSC
Hybridisation solution
Hybond N Nitrocellulose filter
Whatman 3M Filter paper
Paper towels

Methods:

Genomic DNA was digested overnight with the required enzyme. For mitochondrial blots, approx. 15µg of DNA were used, and the enzyme used was *Kpn*I (5 units). For the chloroplast blot, only 5µg of DNA was required, and was digested with *Bcl* (5 units). The reaction was performed in 20µl, with water and buffer as required.

Digested DNA was run on a 0.8% agarose midi-gel (15cm x 11cm) for 3 1/2 hours at 90V. After visualisation to check the digestion had occurred correctly, the gel was placed in a tray and covered with denaturing solution for 45 minutes. The denaturant was then decanted, the gel rinsed in water, and the tray filled with neutralising solution. Neutralisation was allowed to proceed for 45 minutes.

Meanwhile, the blotting apparatus was set up. Perspex plates were placed in a tray as the support (about 1.5cm high). The support was entirely wrapped in filter paper, wetted with 20xSSC. Five filter papers cut to size (15cm x 11cm) were wetted in 20xSSC and placed on the support. A 15cm x 11cm Hybond filter was cut, labelled with a pencil, and wetted in water. A notch cut in the corner of the filter allowed orientation.

The neutralised gel was cut at the corner to match that in the filter paper, and placed face down on the filter papers. Air bubbles were smoothed out, and the Hybond filter carefully lowered over the gel. This was then topped with 3 wetted filter papers, and a stack of paper towels cut to size was placed above these. This was weighted by a 500g weight. The tray was filled up to the top of the support with 20xSSC, and left for 14hrs (overnight).

The following morning, the Hybond filter was rinsed in 10xSSC, and while wet (not dripping) was placed in a UV cross linker following the manufacturers instructions to bind the DNA to the filter. The filter was then ready for use.

.b Probing

The blot was soaked in 6xSSC and sandwiched with nylon membranes. In this way, multiple blots (up to 3) could be layered and probed together. These were rolled up to fit into glass tubes. 20-40 ml (depending on the number of blots) of hybridisation buffer, warmed to 65°C, was poured in. The tubes were sealed and place in a Hybaid hybridisation oven, rotating slowly and warmed to 65°C. Pre-hybridisation proceeded for 1-2 hrs.

The radioactive probe was heated to 94°C for 3 minutes to denature it, and pipetted into the tube. 20 µl of probe was sufficent/ 50 ml of hybridisation buffer. Hybridisation proceeded overnight.

.c Washing and Visualising

Washing the blot to remove spurious signal was the most important part of the process. The chloroplast blots required more washing than the mitochondrial blots, and mitochondrial blots were often underwashed to ensure a readable signal.

Stringency of washes was varied by increasing SDS and reducing SSC concentration, both of which made the wash more stringent. All washes were at 65°C.

The radioactive hybridisation buffer was decanted and stored for disposal. The first 2 washes, of 5 minute duration were in 2xSSC and no SDS. These low-stringency washes were followed by two 20 minute washes in 2xSSC and 0.5% SDS. For the chloroplast blot, 2 further 20 minute washes in 1xSSC, 0.5% SDS were also carried out. Radioactivity of the filter was monitored at this stage, and if very high (suggesting a high intensiity of background signal), the high stringency washes were repeated.

Finally, the blot was washed in 2xSSC at room temperature for 10 minutes. The moist blot was then sealed in a plastic bag (with no air bubbles) and laid down in Hybaid chemi-lumescent cartridges with x-ray film overlaid. Two films were used for the mitochondrial blots

to allow further exposure if the first was under exposed. For chloroplast blots, 4 hrs exposure was often sufficent, whereas mitochondrial blots were examined after 2 weeks.

Films were developed in an automatic developer, and scored.

.xii Recipes

.a DNA Extraction, Storage and Visualisation

50 x TE: 0.5M Tris-Hcl 0.05mM EDTA pH 7.4

2 x CTAB buffer (not autoclavable): 2% (w/v) CTAB 1.4M NaCl 20mM EDTA,Na2 100mM Tris-HCl pH 8.0 5% (w/v) PVP

DNA Wash Buffer: 30% (v/v) 0.3M Sodium Acetate 70% Ethanol

5xTBE: 24g Tris base 27.5g boric acid 10mM EDTA in 1l water

50xTAE: 242g Tris base 57.1ml glacial acetic acid 50mM EDTA in 1lwater

.b RFLP And Southern Blotting

20xSSC: 175.3g NaCl 88.2g Citric acid pH 7.0 in 1l water

Gel Denaturing Solution: 1.5M NaCl 0.5N NaOH

Gel Neutralising Solution: 1M Tris-Cl pH 7.4 1.5M NaCl

Hybridisation Buffer: 4xSSC 2xDenhardt's reagent 1% SDS 10%Dextran Sulphate 20mM Tris(pH 7.6)

.c Isozymes

Vegatative Extraction Buffer: 8% (w/v) PVP 0.3M sucrose 0.5mM EDTA 1mM dithiothreitol 1mM Ascorbic acid 0.1% (w/v) BSA 0.2mM Pyridoxal 5-phosphate 0.4mM NAD 0.3mM NADP

H gel buffer: 4.19g/l L-Histidine pH 7.0 with NaOH

H running buffer: 117.64g/l Citric acid pH 7.0 with NaOH

B gel buffer: 12.1g/l Tris base pH 8.5 with citric acid

B running buffer: 2g/l NaOH 18.5g/l Boric acid

SKDH Stain: 50ml 0.2M Tris-Cl pH 8.0 25mg SKDH 0.02% $MgCl_2$ 0.5 ml of10mg/ml NADP 1 ml of 10mg/ml NBT 0.5ml of 5mg/ml PMS

GOT Stain: 50ml 0.2M Tris-Cl pH 8.0 300mg Aspartic acid 150mg ketoglutaric acid 7.5mg Pyridoxal 5-phosphate 225mg Fast blue BB salt

PGM Stain: 25ml Tris-HCl, pH 8.0 150mg glucose-1-phosphate 25 units glucose-6-phosphate dehyydrogenase 0.02% $MgCl_2$ 0.5ml NADP 0.5ml NBT 0.5ml PMS

Appendix B: Sequences And Primers

.i Consensus sequence of Taberlet Fragment ab in Japanese and European larch, Chinese larch (from Hu 1998), and Tamarack

```
                                  10         20         30         40         50
                                   |          |          |          |          |
Japanese and European                                     TGGAATANAACTCCTCATTTCATTGGCGTGA
Chinese                 CTAACCTCTGAGCTAAGCAGGCTCAATGGAAATATAACTCCTCATTTCATTGGCGTGA
Laricana                                  CAGGCTCAATGGAAATATAACTCCTCATTTCATTGGTGTGA
                         60         70         80         90         100        110
                          |          |          |          |          |          |
Japanese and European   GATCTATAGATTCTTTTGGAATCCTAACAATTATAGCGCGAATCAGATTCAATGACG
Chinese                 GATCTATAGATTCTTTTGGAATCCTAACAATTATAGCGCGAATCAGATTCAATGACG
Laricana                GATCTATAGATTCTTTTGGAATCCTAACAATTATAGCGCGAATCAGATTCAATGACG
                            120        130        140        150        160        170
                             |          |          |          |          |          |
Japanese and European   CAATCCAGATTAAAATTACAACGGAACATCGTCTGAATGTAGATTCCTTCAAGGGAA
Chinese                 CAATCCAGATTAAAATTACAACGGAACATCGTCTGAATGTAGATTCCTTCAAGGGAA
Laricana                CAATCCAGATTAAAATTACAACGGAACATCGTTTGAATGTAGATTCCTTCAAGGGAA
                                180        190        200        210        220
                                 |          |          |          |          |
Japanese and European   AGAAAAGGGAAGTAAGGGTCAATTGATATCGATCGTTTTACTCACTATTCCAAATCG
Chinese                 AGAAAAGGGAAGTAAGGGTCAATTGATATCGATCGTTTTACTCACTATTCCAAATCG
Laricana                AGAAAAGGGAAGTAAGGGTCAATTGATATCGATCGTTTTACTCACTATTCCAAATCG
                        230        240        250        260        270        280
                         |          |          |          |          |          |
Japanese and European   ACTAGGGGAGGATAATAACATTGCATTGAAAATGCAGAAATAATATAATGATTCCCA
Chinese                 ACTAGGGGAGGATAATAACATTGCATTGAAAATGCAGAAATAATATAATGATTCCCA
Laricana                ACTAGGGGAGGATAATAACATTGCATTGAAAATGCAGAAATAATATAATGATTCCCA
                            290        300        310        320        330        340
                             |          |          |          |          |          |
Japanese and European   GTCATATTCGATTGGGGTAGAGATAGAGATGGCGAGAGAAGAGGAGTAGGGGATGAT
Chinese                 GTCATATTCGATTGGGGTAGAGATAGAGATGGCGAGAGAAGAGGAGTAGGGGATGAT
Laricana                GTCATATTCGATTGGGGTANAGATAGAGATGGCGAGAGAAGAGGAGTANGGGATGAT
                                350        360        370        380        390
                                 |          |          |          |          |
Japanese and European   ATTTCTCTCATCCAATATTGAGCAAATATCCAATGAATAACGCTGATGAAGATTACA
Chinese                 ATTTCTCTCATCCAATATTGAGCAAATATCCAATGAATAACGCTGATGAAGATTACA
Laricana                ATTTCTCTCATTCAATATTGAGCAAATATCCAATGAATAACGCTGATGAAGATTACA
                        400        410        420        430        440        450
                         |          |          |          |          |          |
Japanese and European   TAATAGGATTAGATCAAGGTATGATCTCGTTCTCTGATAAGGGGATATGGCGAAATC
Chinese                 TAATAGGATTAGATCAAGGTATGATCTCGTTCTCTGATAAGGGGATATGGCGAAATC
Laricana                TAATANGATTAGATCAAGGTATGATCTCGTTCTCTGAGAAGGGGATATGGCGAAATC
                         460
                          |
Japanese and European   GGTA
Chinese                 GGTA
Laricana                GGTA
```

Notes: The Japanese and European sequence is a consensus from two individuals of each species, both forward and reverse.
The Chinese sequence is a consensus from three individuals, forward and reverse.
The Laricana sequence is a consensus from two individuals, both forward and reverse.
Sequences in italics represent primer sites.
Laricana differs from the other species at 2bp, at positions 156 and 354.

.ii Consensus sequence of Taberlet Fragment ef in Japanese and European larch, and in Chinese larch (from Hu 1998)

```
Japanese and European                                                    TTTCTC
Chinese                 GTTCAAGTCCCTCTATCCCCACCCAGGTTCGTTCCCGAACGACTGATCTATTTTCTC
                           10        20        30        40        50        60
                            |         |         |         |         |         |
Japanese and European   CAATTCCGTTAGTTCGAATCCATTCTCACTTCTCGATTCTTTTACCTCACTATTTTA
Chinese                 CAATTCCGTTAGTTCGAATCCATTCTCACTTCTCGATTCTTTTACCTCACTATTTTA
                              70        80        90        100       110       120
                               |         |         |         |         |         |
Japanese and European   TTTATTCATGAAGAGAAGAAATTAGAACATGAATCTTTCCATCCATCTTATGAAAAG
Chinese                 TTTATTCATGAAGAGAAGAAATTAGAACATGAATCTTTCCATCCATCTTATGAAAAG
                                 130       140       150       160       170
                                  |         |         |         |         |
Japanese and European   TTGGGTTGATCAGTTGATCATATGATAAAATCATTTTGTGATATATGATCCACATAG
Chinese                 TTGGGTTGATCAGTTGATCATATGATAAAATCATTTTGTGATATATGATCCACATAG
                          180       190       200       210       220       230
                           |         |         |         |         |         |
Japanese and European   ATGATATCATTTGGAAATTATTCGATCGCAGTCAATTTTTTATCATATTAGTGGCTT
Chinese                 ATGATATCATTTGGAAATTATTCGATCGCAGTCAATTTTTTATCATATTAGTGGCTT
                             240       250       260       270       280       290
                              |         |         |         |         |         |
Japanese and European   CCAGATCGAAAATAATAAAGATTATTCTAAAAACTAGGAAAAATCCTTTTTTTCCTT
Chinese                 CCAGATCGAAAATAATAAAGATCATTCTAAAAACTAGGAAAAATCCTTTTTTTCCTT
                                300       310       320       330       340
                                 |         |         |         |         |
Japanese and European   ATTTTTAGTTGACACAAGTTCAAACCTTGTACCAGGATGATNCACAGGGAAGAGCCG
Chinese                 ATTTTTAGTTGACACAAGTTCAAACCTTGTACCAGGATGATCCACAGGGAAGAGCCG
                          350       360       370       380       390       400
                           |         |         |         |         |         |
Japanese and European   GGATAGCTCNGTTNGTAGAGCAGAGGACTGAAMATCCTCGTGTCACCAGTTCAAATC
Chinese                 GGATAGCTCAGTTGGTAGAGCAGAGGACTGAAAATCCTCGTGTCACCAGTTCAAATCCTC
```

Notes: The Japanese and European sequence is a consensus from two individuals from each species, using both forward and reverse primers.
The Chinese sequence is a consensus from three individuals, using both forward and reverse primers. Sequences in italics represent primer sites.
Notice the 1bp difference between Chinese larch and both European and Japanese larch at position 256.

.iii Primer Sequences for amplifying targeted regions of the genome

Target	Direction	Sequence
Taberlet ab	Forward Reverse	CATTACAAATGCGATGCTCT TCTACCGATTTCGCCATATC
Taberlet cd	F R	CGAAATCGGTAGACGCTACG GGGGATAGAGGGACTTGAAC
Taberlet ef	F R	GGTTCAAGTCCCTCTATCCC ATTTGAACTGGTGACACGAG
ORF25	F R	CAAGWRACTCCTCGACGGA GTCTTTTCGCACTTAGGCGC
COXI	F R	ATTATCACTTCCGGTACTGG AGCATCTGGATAATCTGG
Calmodulin	F R	AGCCTNTTCGACAAGGATGG AGTGANCGCATAACAGTT
Glyceraldehyde 3-pdh	F R	GATAGATTTGGAATTGTTGAGG AAGCAATTCCAGCCTTGG
Alcohol dehydrogenase	F R	TACTTITGGGAACGIAAGGTA TCICCIACACTCTCIACAAT
Rubisco	F R	ATGCATGCAGGTGTGGC GTTGTCGAATCCGATGAT

.iv RAPD primer sequences

Primer	*Sequence*	*Primer*	*Sequence*
OPA7	GAAACGGGTG	OPAL16	CTTTCGAGGG
OPAL14	TCGCTCCGTT	OPK16	GAGCGTCGAA

Appendix C: Disequilibria and Calculations Of Expected Genealogical Class Frequency

.i Disequilibria

All the definitions and derivations of cytonuclear disequilibria are from Asmussen (Asmussen, Arnold and Avise 1987; Asmussen and Basten 1996), except those for RAPDs which have been modified from these definitions (Dean and Arnold 1996). However, some of the notation has been modified. The composite genotypic disequilibria between two nuclear genes is taken from Weir (Weir 1990), though the use of maximum and minimum bounds is my own. Cytoplasmic disequilibria (between mitochondria and chloroplast) follow naturally from other disequilibria.

.a Notation

Due to the large numbers of disequilibria possible in this project, a consistent notation is necessary, which is explained here. All notations are for bi-allelic systems.

D refers to a disequilibrium.

P refers to a frequency (either allelic or genotypic, depending on subscript).

n refers to count (either allelic, or genotypic, depending on subscript).

The subscript M refers to the mitochondria, and C to the chloroplast. The subscripts G, S, and P are used to denote the 3 nuclear markers GOT, SKDH and PGM. More generally, A and B are used to denote nuclear markers. Nuclear and cytoplasmic allele frequencies are denoted by a single subscript. Nuclear genotype frequencies are denoted using two subscripts (e.g. P_{AA} means frequency of the A homozygote). Cytonuclear allelic disequilibria or frequencies are denoted by two subscripts, one for each system. Cytonuclear genotypic disequilibria or frequencies are denoted by three subscripts, one for the organelle and two to denote the genotype of the nuclear marker. Capitals are used to denote the (arbitrary) positive allele, and lower case the negative allele. In the context of this project, capitals are used to denote the allele that is most common in the European parents, usually species-specific or diagnostic.

Reference to haploid cytonuclear frequencies is made (e.g. P_{CA}). This is interpreted as the frequency of the A allele in the presence of a C cytoplasm, and is therefore equal to the genotypic frequency of the A homozygote in the presence of a C cytoplasm plus half the genotypic frequency of the heterozygote.

i.e. $P_{CA} = P_{CAA} + 0.5P_{CAa}$

The only nuclear disequilibria are composite- not genotypic disequilibria- and can simply be denoted by the two nuclear systems involved.

.b Cytoplasmic Disequilibrium

The disequilibria between two uniparentally inherited markers is most naturally defined as the departure from the expected frequency of a chosen genotypic class. In the simple case of a bi-allelic system:

$$D_{CM} = P_{CM} - P_C P_M$$

This is the only disequilibrium for these two markers. Using the convention of choosing the positive alleles to be representative of the same species, a positive value indicates that con-specific markers are associated.

Appendix Table C-1 shows the relationship between cytoplasmic genotypic frequency, the expected frequencies, and the disequilibrium.

Appendix Table C-1:Cytoplasmic frequencies in terms of disequilibrium (D_{CM})

	Mitotype		
Chlorotype	*M*	*m*	*Marginal Total*
C	$P_C P_M + D_{CM}$	$P_C(1-P_M) - D_{CM}$	P_C
c	$(1-P_C)P_M - D_{CM}$	$(1-P_C)(1-P_M) + D_{CM}$	$(1-P_C)$
Marginal Total	P_M	$(1-P_M)$	1

.c Normalised Cytoplasmic Disequilibrium

The normalised cytoplasmic disequilibrium is found by determining the maximum and minimum boundaries for D. For a positive absolute disequilibrium, the normalised value is found by dividing by the maximum D. For a negative absolute disequilibrium, the normalised value is found by dividing through by the magnitude of the minimum D.

Maximum values are found by maximising P_{CM} within the constraints given by the marginal totals P_M and P_C. Appendix Table C-1 shows that if the frequency of the positive chloroplast allele is greater than that of the mitochondrion, then the maximal D_{CM} is given by $\max P_{CM} = P_M$, otherwise $\max P_{CM} = P_C$. The minimum disequilibrium arises from minimising P_{CM}, within the constraints given by the marginal totals $(1-P_C)$ and $(1-P_M)$. If $P_c+P_m < 1$, then the minimum frequency is 0. If $P_C + P_M > 1$, then the minimum frequency is $P_C + P_M - 1$.

Appendix Table C-2 shows the results of calculating D from these values.

Appendix Table C-2: Conditions for maximising or minimising D_{CM}

Condition	Minimum D	Maximum D
$P_C + P_M > 1$ $P_C + P_M < 1$	$=-(1-P_C)(1-P_M)$ $=-P_CP_M$	
$P_C > P_M$ $P_C < P_M$		$P_M(1-P_C)$ $P_C(1-P_M)$

.d Cytonuclear Disequilibria

Asmussen, Arnol and Avise (1987) break the association between nuclear and cytoplasmic genes into four components. The three genotypic disequilibria, measuring associations between each of the three possible nuclear genotypes and the cytotypes, are defined in the following way. These definitions are for the association between chloroplast and nuclear locus A.

$$D_{CAA} = P_{CAA} - P_{AA}P_C$$

$$D_{CAa} = P_{CAa} - P_{Aa}P_C$$

$$D_{Caa} = P_{Caa} - P_{aa}P_C$$

The relationship between these disequilibria, the expected genotypic frequencies and the actual frequencies is shown in Appendix Table C-3. It can be seen from this table that

$$D_{CAA} + D_{CAa} + D_{Caa} = 0$$

Appendix Table C-3: Cytonuclear genotypic frequencies in terms of disequilibria (Asmussen, Arnold and Avise 1987)

	Nuclear Genotype			
Cytotype	*AA*	*Aa*	*aa*	*Marginal Total*
C	$P_{AA}P_C + D_{CAA}$	$P_{Aa}P_C + D_{CAa}$	$P_{aa}P_C + D_{Caa}$	P_C
c	$P_{AA}P_c - D_{CAA}$	$P_{Aa}P_c - D_{CAa}$	$P_{aa}P_c - D_{Caa}$	P_c
Marginal Total	P_{AA}	P_{Aa}	P_{aa}	1

The fourth cytonuclear disequilibria is the allelic disequilibria.

$$D_{CA} = P_{CA} - P_CP_A$$

This measures the association between the cytotype and each nuclear allele. It is possible to use this even though gametic information is not present, as the haploid nature of the organelle genome makes 'phase' an irrelevance. From this, and the definitions of the genotypic disequilibria, it is easily shown that

$$D_{CA} = D_{CAA} + \frac{1}{2}D_{CAa}$$

Along with the relationship between the three genotypic disequilibria, it means that these four disequilibria actually break down into just two independent measures, the other two being calculable from these via the following equations

$$D_{CA} = D_{CAA} + \frac{1}{2}D_{CAa} = \frac{1}{2}D_{CAA} - \frac{1}{2}D_{Caa} = -\frac{1}{2}D_{CAa} - D_{Caa}$$

The convention of calculating and testing all four must be viewed cautiously, as it may suggest a greater degree of information from these measurements than actually exists. However, this parameterisation can help to understand the nature of the genetic structure.

.e Normalised Cytonuclear Disequilibria.

Using the same approach as above for the cytoplasmic disequilibrium, normalised cytonuclear disequilibria can be found by calculating the maximum and minimum disequilibrium for a given data set (constrained by the genotypic marginal totals). If positive, the absolute disequilibrium is divided by the maximum disequilibrium to obtain the normalised value. If negative, it is divided by the magnitude of the minimum D. The following results in Appendix Table C-4 are from Amussen (Asmussen and Basten 1996).

Appendix Table C-4: Maximising and minimising cytonuclear disequilibria (Asmussen and Basten 1996)

D	*minimum D*	*Condition*	*Maximum D*	*Condition*
D_{CAA}	$-P_{AA}P_C$ $-(1-P_{AA})(1-P_C)$	$P_C \leq 1-P_{AA}$ $P_C \geq 1-P_{AA}$	$(1-P_{AA})P_C$ $P_{AA}(1-P_C)$	$P_C \leq P_{AA}$ $P_C \geq P_{AA}$
D_{CAa}	$-P_{Aa}P_C$ $-(1-P_{Aa})(1-P_C)$	$P_C \leq 1-P_{Aa}$ $P_C \geq 1-P_{Aa}$	$(1-P_{Aa})P_C$ $P_{Aa}(1-P_C)$	$P_C \leq P_{Aa}$ $P_C \geq P_{Aa}$
D_{Caa}	$-P_AP_C$ $-\{P_{AA}P_C+P_{aa}(1-P_C)\}/2$ $-(1-P_A)(1-P_C)$	$P_C \leq P_{aa}$ $P_{aa} \leq P_C \leq 1-P_{AA}$ $P_C \geq 1-P_{AA}$	$(1-P_A)P_C$ $\{P_{AA}(1-P_C)+P_{aa}P_C\}/2$ $P_A(1-P_C)$	$P_C \leq P_{AA}$ $P_{AA} \leq P_C \leq 1-P_{aa}$ $P_C \geq 1-P_{aa}$

.f Composite Nuclear Disequilibria

Without gametic information, it is impossible to calculate the linkage disequilibrium normally used to measure associations between two nuclear alleles. However, it is possible to calculate 'composite' genotypic disequilibria which uses the fact that though the digenic frequencies representing alleles in repulsion and alleles in coupling cannot be observed, their sum can be. These definitions are from Weir (Weir 1990).

$$\Delta_{AB} = D_{AB} + D_{A/B}$$

This defines the composite nuclear disequilibria between loci A and B to be the sum of the gametic and nongametic linkage disequilibria, which cannot actually be observed in this

situation. However, the maximum likelihood estimator for this (the parameter used throughout the main text as Δ_{AB}) is given by

$\Delta_{AB} = P_{AB} - 2P_AP_B$

where P_{AB} is interpreted as the frequency of AB alleles in repulsion + the frequency of AB alleles in coupling, so

$P_{AB} = 2p_1 + p_2 + p_4 + p_5/2$ (refer to Appendix Table C-5)

Appendix Table C-5: Notation used for genotypic frequencies

	BB	*Bb*	*bb*	*Marginal total*
AA	p_1	p_2	p_3	t_1
Aa	p_4	p_5	p_6	t_2
aa	p_7	p_8	p_9	t_3
Marginal total	t_4	t_5	t_6	1

This digenic disequilibrium is not the only one that can be calculated. There are also trigenic and quadrigenic disequilibria. However, these have not been included in this analysis, as they are smaller, have larger variances, and are harder to interpret.

.g Normalised Composite Disequilibria

Normalised composite disequilibria are derived here. Again, the minimal/ maximal values of Δ are needed. However, the approach is somewhat different as the situation is more complicated. Consider Appendix Table C-5 which shows how the various genotypic classes are numbered. The following approach shows how to work out the values that maximise Δ. The genotypic frequencies for the 'maximum Δ contingency table' must be calculated in the order given.

To maximise Δ, we must first maximise p_1 (as it contributes the most to Δ), then p_2 and p_4 (whose coefficients are both one, but are independent and can be maximised independently), and finally p_5. This maximisation must, of course, be carried out constrained by the marginal totals denoted by 't'. Appendix Table C-6 shows how this is done.

Appendix Table C-6: Calculating the maximum Δ from a nuclear contingency table

Genotypic Frequency	*Frequency to maximise Δ*
Mp_1	min (t_1,t_4)
Mp_2	min (t_5,t_1-Mp_1)
Mp_4	min (t_2,t_4-Mp_1)
Mp_5	min (t_5-Mp_2,t_2-Mp_4)

Notes: Mp refers to the genotypic frequency that will maximise Δ; refer to Appendix Table C-5 to see which genotypic class each p and t refer to.

Minimising Δ is more complicated. I show a heuristic method that I have used, but it is likely that rather more efficient heuristic or algebraic methods exist. The basis of this method is to repeatedly try and 'remove' as much of the counts from each of the genotypic frequencies in turn (starting at the 'least important' p_5 and finishing with p_1), followed by a redistribution to allow yet more to be removed. The steps are shown below, and it is assumed that a contingency table of the form in Appendix Table C-5 has been set up. It has not been formally proved, but agrees with solutions found using the iterative solver of Excel 5.0.

Algorithm for Finding the Genotypic Frequencies that Minimise Δ

1. Reduce p_5 and p_9 by min(p_5, p_9). Increase p_6 and p_8 by this amount.
2. Reduce p_2 and p_6 by min (p_2, p_6). Increase p_3 and p_5 by this amount.
3. Reduce p_4 and p_8 by min (p_4, p_8). Increase p_7 and p_5 by this amount.
4. Reduce p_1 and p_5 by min (p_1, p_5). Increase p_2 and p_4 by this amount.
5. Reduce p_6 and p_8 by min (p_6, p_8). Increase p_5 and p_9 by this amount.
6. Repeat steps 1 to 5 until there is no change.

The contingency table thus arrived at can be used to calculate the minimum Δ. In all real cases, and many hypothetical ones, the results of applying this algorithm was checked against the solution found by Excel and agreed in every case.

.ii Expected Frequency of Genealogical Classes Under 'F_2' hypothesis

In chapter 5, it was proposed that a second generation hybrid population (the 1933 stand) was derived from a first generation hybrid population which was produced by mixed pollination of one species by another, with seed from one species only. The proportion of hybrids produced by the maternal species under this system was denoted h, and the size of a resulting F_1 population N. The expected genealogical composition of the F_2 produced by random mating within this population is derived below. Note that the 'F_1' is purely hypothetical.

Let h be the chance that an individual in the F_1 generation is hybrid

H be the number of hybrids in the putative F_1 population.

N be the size of the F_1 population

Then,

$H \sim B(N,h)$ (i.e. H is distributed Binomially, N trials and probability h)

$\therefore E(H) = hN$

$Var(H) = h(1-h)N$

$\therefore E(H^2) = hN(hN+1-h)$

This result is used in the following derivations.

The probability that a tree in the F_2 generation is an F_2 individual, E(F2), is the probability that both its parents are F_1. Each of these probabilities is (H/N), so

$$\begin{aligned} E(F2) &= E(\frac{H}{N} \times \frac{H}{N}) \\ &= \frac{1}{N^2} E(H^2) \\ &= \frac{h}{N}(h(N-1)+1) \qquad (2) \\ &= h^2 \qquad (\text{as } N \to \infty) \end{aligned}$$

The probability that a tree in the F_2 generation is a backcross, E(BC), is the probability that one of its parents is pure maternal ((1-H)/N) and the other is hybrid(H/N), so

$$\begin{aligned} E(BC) &= E(2 \times \frac{H}{N} \times (1 - \frac{H}{N})) \\ &= \frac{2}{N} E(H - \frac{H^2}{N}) \\ &= \frac{2}{N}(hN - h(hN+1-h)) \\ &= \frac{2h}{N}(N-1)(1-h) \qquad (3) \\ &= 2h(1-h) \qquad (\text{as } N \to \infty) \end{aligned}$$

Half of these are backcrosses with the F_1 as seed parent, the other half with the F_1 as pollen parent.

Finally, there can be no F_1s in this generation, as the putative F_1 population contained parents of only one species. Thus, the expected proportion of pure parents is found by subtraction:

$$\begin{aligned} E(P) &= 1 - E(BC) - E(F2) \\ &= 1 - (\frac{2h}{N}(N-1)(1-h)) - \frac{h}{N}(h(N-1)+1) \\ &= \frac{1}{N}(N + h(1-2N) + h^2(N-1)) \qquad (4) \\ &= (1-h)^2 \qquad (\text{as } N \to \infty) \end{aligned}$$

The results 2-4 are the results displayed in Table 5-6 and subsequently used to calculate the expectations of each genotypic class frequency, in Table 5-8.

.iii Genotypic Probabilities Contingent on Genealogical Class

Table 5-8 uses these expected genealogical class frequencies in conjunction with the contingent probabilities of genotype given genealogical class. These contingent probabilities are calculated using Mendelian laws in a simple fashion, based on the 'F_2' hypothesis.

Using the same notation as previously to represent allele and genotype frequencies, but with the added notation of a superscript J to represent frequency in the seed parent and E to represent the frequency in the pollen parent, the following section shows how these are calculated.

There are four possible genealogical classes (Parent J, F_2, BC_1 to parent J as pollen parent, and BC_1 to parent J as seed parent). There are 3 possible genotypes at each of the two nuclear loci, and 2 for the paternally inherited marker. There is no variation at the maternal marker, as all trees descend from the same species on the maternal side. Appendix Table C-7 below shows the probability of each single locus genotype given genealogical class.

Appendix Table C-7:Single locus genotypic probabilities contingent on genealogical class

GENEALOGICAL CLASS	AA	NUCLEAR GENOTYPE Aa	aa
PureA	P_A^{J2}	$2P_A^J(1-P_A^J)$	$(1-P_A^J)^2$
F2	$(\frac{1}{2}P_A^J+\frac{1}{2}P_A^E)^2$	$2(\frac{1}{2}P_A^J+\frac{1}{2}P_A^E)(1-\frac{1}{2}P_A^J-\frac{1}{2}P_A^E)$	$(1-\frac{1}{2}P_A^J-\frac{1}{2}P_A^E)^2$
BC1 to A Male	$P_A^J(\frac{1}{2}P_A^J+\frac{1}{2}P_A^E)$	$P_A^J(1-\frac{1}{2}P_A^J-\frac{1}{2}P_A^E)+(1-P_A^J)(\frac{1}{2}P_A^J+\frac{1}{2}P_A^E)$	$(1-P_A^J)(1-\frac{1}{2}P_A^J-\frac{1}{2}P_A^E)$
BC1 to A Female	$P_A^J(\frac{1}{2}P_A^J+\frac{1}{2}P_A^E)$	$P_A^J(1-\frac{1}{2}P_A^J-\frac{1}{2}P_A^E)+(1-P_A^J)(\frac{1}{2}P_A^J+\frac{1}{2}P_A^E)$	$(1-P_A^J)(1-\frac{1}{2}P_A^J-\frac{1}{2}P_A^E)$
	C	CHLOROPLAST GENOTYPE c	
Pure A	P_C^J	$1-P_C^J$	
F2	P_C^E	$1-P_C^E$	
BC1 to A Male	P_C^J	$1-P_C^J$	
BC1 to A Female	P_C^E	$1-P_C^E$	

It is a simple matter to calculate these probabilities by substituting in the relevant observed frequencies (e.g. $P_C^J=0$, $P_C^E=1$). The expected multi-locus genotypic frequency is calculated by multiplying these probabilities together for each locus (in this case, two nuclear loci and the chloroplast locus). The calculations giving rise to the results used in Table 5-8 were performed using a spreadsheet version of Appendix Table C-7.

Appendix D: Computer Model

This program was written in C minus minus, a shareware programme available on the internet. It is similar to C, but without pointers, making it easier and more flexible to programme. Thus, it will not run in C without modification.

```
/* HYBRID SWARM SIMULATION */
/* initialise the global variables now */
/*supply adultnum,seednum,locino*/
/*supply toten, and totsim*/
/*Uses seedgenome(individual,locus,chromosome) and adultgenome(.,.,.,) to describe
each individual*/
/*fitstrings name the selective effects included*/
/*fiteffect name what the effect is on*/
/*selection(fitstring,fiteffect)*/
/*adultfitness(individual,seed fitneess(0)pollen fitness)*/
fiteffect[0]="Seed fitness"
fiteffect[1]="Pollen fitness"
fiteffect[2]="seedling viability"
fitstring[0]="Additive nuclear"
seedno=48
locino=2
adultno=48
totgen=2
totsim=100
totmetasim=11
gen=1
sim=1
metasim=1
RANDM=RAND_MAX+1
#define MTLOC locino][0
#define CPLOC locino+1][0
#define QUANT locino+2][0
main()
{
 printf("Seed?\n");
 scanf("%d",seed);
 srand(seed)
 ofp=fopen("sim3.dat","w")
 print_header(ofp)

 /*START OF METSIMLOOP (used to vary starting conditions)*/
 for (metasim=1;metasim<=totmetasim;++metasim){

  /*START OF SIMLOOP (repeats each initial condition 'totsim' times)*/
  for (sim=1;sim<=totsim;++sim){
   printf("metasim:%d sim:%d     ",metasim,sim)
     /*initialise pop. assuming 11 native(initial conditions are set here)*/
   for (i=1;i<12;++i){
    for (j=0;j<2;++j){
     adultgenome[i][j][0]=1
     adultgenome[i][j][1]=1}
    adultgenome[i][CPLOC]=1
    adultgenome[i][MTLOC]=1
    adultgenome[i][QUANT]=100.00}
   adultno=11

   /*initialise selective coefficients (0, in this case)*/
   for (i=0;i<3;++i){
    selection[0][i]=0}

   /*start of generation(in this case, the effect of varying migration rates is being
tested)*/
   polmig1=0.75-(metasim-1)*1.0/20
   polmig2=(0.75-polmig1)/(1.0-polmig1/2.0)
   simend=0
   gen=0
   calc_stats(adultno,adultgenome,stats)
   print_stats(stats,ofp)
   while (simend==0){
    ++gen
    simend=1
```

```
    if (gen==1) polmig=100*polmig1
    if (gen==2) polmig =100*polmig2

    printf("gen %d\n",gen)
    printf(".")

    /*Mate Adults*/
    for (i=1;i<=seedno;++i){
     mother=rand()*adultno/RANDM+1.0
     father=rand()*adultno/RANDM+1.0
     /*allow pollen migration of migrant with frequence pollenmig*/
     If (rand()*100/RANDM+1>polmig) {
      seedgenome[i][CPLOC]=adultgenome[father][CPLOC]
      seedgenome[i][MTLOC]=adultgenome[mother][MTLOC]

seedgenome[i][QUANT]=(adultgenome[father][QUANT]+adultgenome[mother][QUANT])/2.0
      for (j=0;j<locino;++j){
       pallele=rand()*2/RANDM
       mallele=rand()*2/RANDM
       seedgenome[i][j][0]=adultgenome[father][j][pallele]
       seedgenome[i][j][1]=adultgenome[mother][j][mallele]
      }
     }
     else {
      seedgenome[i][CPLOC]=0
      seedgenome[i][MTLOC]=adultgenome[mother][MTLOC]
      seedgenome[i][QUANT]=(adultgenome[mother][QUANT])/2.0
      for (j=0;j<locino;++j){
       mallele=rand()*2/RANDM
       seedgenome[i][j][0]=0
       seedgenome[i][j][1]=adultgenome[mother][j][mallele]
      }
     }
    }

/*grow seedlings unselectively(replace by module at end of file to include
selection)/
    if (gen==1) adultno=48
    for (i=1;i<=adultno;++i){
     adultgenome[i][MTLOC]=seedgenome[i][MTLOC]
     adultgenome[i][CPLOC]=seedgenome[i][CPLOC]
     adultgenome[i][QUANT]=seedgenome[i][QUANT]
     for (j=0;j<locino;++j){
      adultgenome[i][j][0]=seedgenome[i][j][0]
      adultgenome[i][j][1]=seedgenome[i][j][1]
     }
    }
   /*end of generation*/
    calc_stats(adultno,adultgenome,stats)
    print_stats(stats,ofp)
    if (gen<totgen) simend=0
   }
/*end of sim*/
  }
/*end of metasim*/
 }
 fclose(ofp)
getch();
}

cumulative_prop (number,probability)
{
cumulativeprob[1]=probability[1]
for (i=2;i<=number;++i)
cumulativeprob[i]=cumulativeprob[i-1]+probability[i]
for (i=1;i<=number;++i)
cumulativeprob[i]=cumulativeprob[i]/cumulativeprob[number]
return cumulativeprob
}

/*this calculates disequilibria and allele frequencies*/
calc_stats(number,genome,stats)
{
stats.cpfre=0.0
stats.mtfre=0.0
stats.quan=0.0
for (j=0;j<locino;++j){
```

```
stats.nucfre[j]=0.0
stats.hetero[j]=0.0
stats.homo[j]=0.0}
for (i=0;i<2;++i){
  for (j=0;j<4;++j){
    for (k=0;k<3;++k) {
    stats.contab[i][j][k]=0}}}
pab=0
pam=0
pac=0
for (i=1;i<=number;++i){
for (j=locino-1;j>=0;--j){
alleles=genome[i][j][0]+genome[i][j][1]
stats.nucfre[j]=stats.nucfre[j]+(alleles)/(2.0*number)
stats.hetero[j]=stats.hetero[j]+(genome[i][j][0] != genome[i][j][1])/(1.0*number)}
stats.contab[0][alleles][genome[i][CPLOC]] =
stats.contab[0][alleles][genome[i][CPLOC]]+1
stats.contab[1][alleles][genome[i][MTLOC]] =
stats.contab[1][alleles][genome[i][MTLOC]]+1
pab=pab+genome[i][0][0]*genome[i][1][0]+genome[i][0][1]*genome[i][1][1]
stats.quan=stats.quan+1.0*genome[i][QUANT]/number
}
for (j=0;j<=1;++j) {
stats.contab[j][3][0]=stats.contab[j][0][0]+stats.contab[j][1][0]+stats.contab[j][2][
0]
stats.contab[j][3][1]=stats.contab[j][0][1]+stats.contab[j][1][1]+stats.contab[j][2][
1]
stats.contab[j][0][2]=stats.contab[j][0][0]+stats.contab[j][0][1]
stats.contab[j][1][2]=stats.contab[j][1][0]+stats.contab[j][1][1]
stats.contab[j][2][2]=stats.contab[j][2][0]+stats.contab[j][2][1]
stats.contab[j][3][2]=stats.contab[j][0][2]+stats.contab[j][1][2]+stats.contab[j][2][
2]
}
stats.cpfre=stats.contab[0][3][0]*1.0/number
stats.mtfre=stats.contab[1][3][0]*1.0/number
stats.cpdiv=stats.cpfre*(1-stats.cpfre)
stats.mtdiv=stats.mtfre*(1-stats.mtfre)
stats.D=pab/(2.0*number)-stats.nucfre[0]*stats.nucfre[1]
stats.Dc[1]=(stats.contab[0][0][0]*1.0 +stats.contab[0][1][0]*0.5)/number-
stats.cpfre*stats.nucfre[0]
for (j=2;j<=4;++j)  {
stats.Dc[j]=(stats.contab[0][j-2][0]*1.0/number)-stats.contab[0][j-
2][2]*stats.contab[0][3][0]*1.0/(number*number)
}
stats.Dm[1]=(stats.contab[1][0][0]*1.0 +stats.contab[1][1][0]*0.5)/number-
stats.mtfre*stats.nucfre[0]
for (j=2;j<=4;++j)  {
stats.Dm[j]=(stats.contab[1][j-2][0]*1.0/number)-stats.contab[1][j-
2][2]*stats.contab[1][3][0]*1.0/(number*number)
}
for (j=0;j<locino;++j){
stats.nucdiv[j]=stats.nucfre[j]*(1-stats.nucfre[j])}
stats.popsize=number
}


print_header(ofp)
{
fprintf(ofp,"l1 l2 cp mt l1o l2o CpO mto D Dc DcAA DcAa Dcaa Dm DmAA DmAa Dmaa Pop Qu
gen sim metasim\n")
}



print_stats(stats,ofp)
{
fprintf(ofp,"%f %f %f %f %f %f %f %f %f %f %f %f %f %f %f %f %f %f  %d %d %d %d
\n",stats.nucfre[0],stats.nucfre[1],stats.cpfre,stats.mtfre,stats.nucdiv[0],stats.nuc
div[1],stats.cpdiv,stats.mtdiv,stats.D,stats.Dc[1],stats.Dc[2],stats.Dc[3],stats.Dc[4
],stats.Dm[1],stats.Dm[2],stats.Dm[3],Stats.Dm[4],stats.popsize,stats.quan,gen,sim,me
tasim)
}

/*SELECTIVE BIT this is not used in this study, but can be included. It assumes
different fitness traits are multiplicative*/
/*   calculate fitnesses*/
/* for (i=1;i<=adultno;++i){*/
```

```
/*trait*/
/*    for (trait=0;trait<3;++trait){
adultfitness[trait][i]=1.0*/
/*additive nuclear*/
/*  for (j=0;j<2;++j){
adultfitness[trait][i]=adultfitness[trait][i]+selection[0][trait]*(adultgenome[i][j][
0]+adultgenome[i][j][1])/(locino*2)
}
}
}*/
    /*make seedlings incorpoating pollen output (adultfitness[0]) and seed output
(adultfitness[1])
    pollenprob=cumulative_prop(adultno,adultfitness[0])
    seedprob=cumulative_prop(adultno,adultfitness[1])
    for (i=1;i<=seedno;++i){
     r=rand()*1.0/RAND_MAX
     mother=1
     while (seedprob[mother]<r){
      ++mother}
     r=rand()*1.0/RAND_MAX
     father=1
     while (pollenprob[father]<r){
      ++father}
 ENDOF SELECTIVEBIT*/
```

REFERENCES

Aagaard, J. E., K. V. Krutovskii and S. H. Strauss. 1998. RAPD markers of mitochondrial origin exhibit lower population diversity and higher differentiation than RAPDs of nuclear origin in douglas fir. *Molecular Ecology* **7:** 801-812.

Aas, G. 1993. Taxonomical impact of morphological variation in *Quercus robur* and *Q.petraea:* A contribution to the hybrid controversy. *Ann.Sci.For.Suppl.1* **50:** 107s-113s.

Abbott, R. J. 1992. Plant invasions, interspecific hybridization and the evolution of new plant taxa. *Trends In Ecology & Evolution* **7:** 401-405.

Abbot, R. J. 1998. *Pers. comm.*

Abbott, R. J., P. A. Ashton and D. G. Forbes. 1992. Introgressive origin of the radiate groundsel, *Senecio-vulgaris* var *hibernicus*: Aat-3 evidence. *Heredity* **68:** 425-435.

Abbott, R. J., and A. J. Lowe.1996. A review of hybridization and evolution in British *Senecio.*, in *Hind, D. J. N. and H. J. Beentje (Eds.). Compositae: Systematics; International Compositae Conference, Vol. 1, Kew, England, UK, July 25-August 5, 1994 Pages: 679-689.*

Adam, K. D., and W. H. Koehler. 1995. Evolutionary genetic considerations on the goals and risks in releasing transgenic crops., pp. 59-79 in *Transgenic organisms: Biological and social implications.*, edited by J. Tomiuk, K. Woehrmann and A. Sentker. Birkhaeuser Verlag, New York.

Alavin, M., S. Mathews and C. Hughes. 1991. Chloroplast DNA variation in *Gliricidia sepium* Leguminosae : Intraspecific phylogeny and tokogeny. *American Journal of Botany.* **78:** 1576-1585.

Allnutt, T.R., P. Thomas, A.C.Newton, M.Gardner. 1998. Genetic Variation in *Fitzroya Cupressoides* cultivated in the British Isles, assessed using RAPDs. *Edinburgh Journal of Botany* **55:**329-334.

Allan, G. J., C. Clark and L. H. Rieseberg. 1997. Distribution of parental DNA markers in *Encelia virginensis* (Asteraceae: Heliantheae), a diploid species of putative hybrid origin. *Plant Systematics and Evolution* **205:** 205-221.

Amersham Life Sciences. 1997. ThermoSequenase cycle sequencing pre-mix kit

Anderson, E. 1948. Hybridization of the habitat. *Evolution* **2:** 1-9.

Anderson, E., 1949 *Introgressive hybridization.* John Wiley and Sons, New York.

Arnold, 1997 *Natural Hybridisation.* Oxford University Press, Oxford (U.K.)

Arnold, J. 1993a. Cytonuclear disequilibria in hybrid zones. *Annual Review of Ecology and Systematics* **24:** *521-554.*

Arnold, M. L. 1993b. *Iris nelsonii*: origin and genetic composition of a homoploid iris species. *American Journal of Botany* **80:** 577-583.

Arnold, M. L. 1993c. Rarity of hybrid formation and introgression in Louisiana irises. *Plant Genetics Newsletter* **9:** 14-17.

Arnold, M. L., B. D. Bennett and E. A. Zimmer. 1990a. Natural hybridization between *Iris fulva* and *I.hexagona*: pattern of ribosomal DNA variation. *Evolution* **44:** 1512-1521.

Arnold, M. L., C. M. Buckner and J. J. Robinson. 1991. Pollen-mediated introgression and hybrid speciation in Louisiana USA irises. *Proc. Natl. Acad. Sci. U S A* **88:** 1398-1402.

Arnold, M. L., J. L. Hamrick and B. D. Bennet. 1990b. Allozyme variation in Louisiana irises: a test for introgression and hybrid speciation. *Heredity* **65:** 297-306.

Arnold, M. L., J. J. Robinson, C. M. Buckner and B. D. Bennet. 1992. Pollen dispersal and interspecific gene flow in louisiana irises. *Heredity* **68:** 399-404.

Asmussen, M. A., J. Arnold and J. C. Avise. 1987. Definition and properties of disequilibrium statistics for associations between nuclear and cytoplasmic genotypes. *Genetics* **115:** 755-768.

Asmussen, M. A., and C. J. Basten. 1994. Sampling theory for cytonuclear disequilibria. *Genetics* **138:** 1351-1363.

Asmussen, M. A., and C. J. Basten. 1996. Constraints and normalized measures for cytonuclear disequilibria. *Heredity* **76:** 207-214.

Aubert, J., and M. Solignac. 1990. Experimental evidence for mitochondrial DNA introgression between *Drosophila* species. *Evolution* **44:** 1272-1282.

Bachmann, K. 1994. Tansley review no. 63, Molecular markers in plant ecology. *New Phytologist* **126:** 403-418.

Bacilieri, R., A. Ducousso and A. Kremer. 1996a. Comparison of morphological characters and molecular markers for the analysis of hybridization in sessile and pedunculate oak. *Annales Des Sciences Forestieres (Paris)* **53:** 79-91.

Bacilieri, R., A. Ducousso, R. J. Petit and A. Kremer. 1996b. Mating system and asymmetric hybridization in a mixed stand of European oaks. *Evolution* **50:** 900-908.

Bain, J. F., and R. K. Jansen. 1996. Numerous chloroplast DNA polymorphisms are shared among different populations and species in the aureoid *Senecio* (Packera) complex. *Canadian Journal Of Botany* **74:** 1719-1728.

Baker, H. G. 1947. The ecotypes of *Melandrum dioicum*. *New Phytologist* **47:** 131-145.

Baldwin, B. G., M. J. Sanderson, J. M. Porter, M. F. Wojciechowski, C. S. Campbell *et al.*. 1995. The ITS region of nuclear ribosomal DNA: A valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Gardens* **82:** 247-277.

Baranger, A., A. M. Chevre, F. Eber and M. Renard. 1995. Effect of oilseed rape genotype on the spontaneous hybridization rate with a weedy species: An assessment of transgene dispersal. *Theoretical and Applied Genetics* **91:** 956-963.

Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones., pp. 13-45 in Harrison, R.G (Ed.) *Hybrid zones and the evolutionary process*, Oxford University Press, Oxford.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16:** 113-148.

Barton, N. H., and G. M. Hewitt. 1989. Adaptation, speciation and hybrid zones. *Nature* **341:** 497-503.

Basten, C. J., and M. A. Asmussen. 1997. The exact test for cytonuclear disequilibria. *Genetics* **146:** 1165-1171.

Benkrima, L., and P. Aderkas. 1992. In vitro embryogenesis in larch., pp. 412-416 in Schmidt W.C, and K. J. McDonald (Eds.) *Ecology and management of Larix forests:A look ahead*, Intermountain Research Station, Whitefish,Montana.

Birky, C. W., T. Marayuma and P. Fuerst. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* **103:** 513-527.

Blocker, L. 1992. Aesthetics of larch forests, pp. 151-152 in Schmidt W.C and K. J. McDonald (Eds.) *Ecology and management of Larix forests:A look ahead*, Intermountain Research Station, Whitefish,Montana.

Bobola, M. S., R. T. Eckert, A. S. Klein, K. Stapelfeldt, K. A. Hillenberg *et al.* 1996. Hybridization between *Picea rubens* and *Picea mariana*: Differences observed between montane and coastal island populations. *Canadian Journal of Forest Research* **26:** 444-452.

Brochmann, C. 1984. Hybridization and the distribution of *Argyranthemum coronopifolium* (Asteraceae-Anthemideae) in the Canary Islands. *Nordic Journal Of Botany* **4:** 729-736.

Brubaker, C. L., J. A. Koontz and J. F. Wendel. 1993. Bidirectional cytoplasmic and nuclear introgression in the new world cottons *Gossypium babadense* and *G.hirsutum. American Journal of Botany* **80:** 1203-1208.

Cellino, M. L., and J. Arnold. 1993. The effects of cytoplasmic male sterility on cytonuclear disequilibria in hybrid zones. *Genetica (Dordr)* **88:** 37-50.

Cheliak, W.M., and S.A.Pitel. 1984. *Techniques for starch gel electrophoresis of enzymes from forest tree species.* Information report PI-X-42, Petawawa National Forest Institute.

Clapham, A. R., T. G. Tutin and D. M. Moore. 1987. *Flora of the British Isle.* Cambridge University Press, Cambidge.

Cotton, R. G. H. 1993. Current methods of mutation detection. *Mutation Research* **285:** 125-144.

Cracraft, J. 1989. Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation., pp. 28-59 in Otte D, and J. A. Endler (Eds.), *Speciation and its consequences,* Sinauer, Sunderland.

Cronquist, A. 1987. A botanical critique of cladism. *Systematic Botany* **53:** 1.

Cruzan, M. B., and M. L. Arnold. 1993. Ecological and genetic associations in an iris hybrid zone. *Evolution* **47:** 1432-1445.

Cruzan, M. B., and M. L. Arnold. 1994. Assortative mating and natural selection in an iris hybrid zone. *Evolution* **48:** 1946-1958.

D'Agostino, R., and M. A. Stephens (Editors). 1986. *Goodness of fit techniques.* Marcel Dekker inc., New York.

Dallimore, and Jackson. 1948. *Handbook of Coniferae.* Arnold, London.

Dawson, I. K., A. J. Simons, R. Waugh and W. Powell. 1996. Detection and pattern of interspecific hybridization between *Gliricidia sepium* and *G. maculata* in Meso-America revealed by PCR-based assays. *Molecular Ecology* **5:** 89-98.

Dean, R., and J. Arnold. 1996. Cytonuclear disequilibria in hybrid zones using RAPD markers. *Evolution* **50:** 1702-1705.

Demesure, B., N. Sodzi and R. J. Petit. 1995. A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* **4:** 129-131.

DePamphilis, C., and R. Wyatt. 1990. Electrophoretic confirmation of interspecific hybridization in *Aesculus* (Hippocastanacae) and the genetic structure of a broad hybrid zone. *Evolution* **44:** 1295-1317.

Deverno, L. L., P. J. Charest and L. Bonen. 1993. Inheritance of mitochondrial DNA in the conifer *Larix. Theoretical and Applied Genetics* **86:** 383-388.

Dong, J., and D. B. Wagner. 1993. Taxonomic and population differentiation of mitochondrial diversity in *Pinus banksiana* and *Pinus contorta. Theoretical and Applied Genetics* **86:** 573-578.

Dorado, O., L. H. Rieseberg and D. Arias. 1991a. Chloroplast DNA introgression and the invasion of southern California by sunflowers. *American Journal of Botany* **78:** 181.

Dorado, O., L. H. Rieseberg and D. M. Arias. 1992. Chloroplast DNA introgression in southern California sunflowers. *Evolution* **46:** 566-572.

Doyle, J. J. 1992. Gene trees and species trees: Molecular systematics as one character taxonomy. *Systematic Botany.* **17:** 144-163.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure from small quantities of fresh tissue. *Phytochemical Bulletin* **19:** 11-15.

Doyle, J. J., and J. L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus* **12**.

Dumolin-Lapegue, S., B. Demesure, S. Fineschi, C. V. Le and R. J. Petit. 1997. Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146:** 1475-1487.

Easton, E. 1997. Genetic variation and conservation of the native aspen (*Populus tremula)* resource in Scotland, *PhD Thesis* p.265, University of Edinburgh, Edinburgh.

Echelle, A. F., and P. J. Conner. 1989. Rapid geographically extensive introgression after secondary contact between two pupfish species (*Cyprinidon, Cyprinidontidae*). *Evolution* **43:** 717-727.

Edwards, M. V. 1956. The Hybrid larch, *Larix x Eurolepis* Henry. *Forestry* **29:** 29-43.

Edwards-Burke, M. A., J. L. Hamrick and R. A. Price. 1997. Frequency and direction of hybridization in sympatric populations of *Pinus taeda* and *P. echinata* (Pinaceae). *American Journal of Botany* **84:** 879-886.

Elston, R. C. 1971. The estimation of admixture in racial hybrids. *Annals of Human Genetics* **35:** 9-17.

Emms, S. K., and M. L. Arnold. 1997. The effect of habitat on parental and hybrid fitness: Transplant experiments with Louisiana irises. *Evolution* **51:** 1112-1119.

Ennos, R. 1995a. *Personal communication*

Ennos, R. A. 1994. Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* **72:** 250-259.

Ennos, R. A. 1995b. *Personal communication*

Ennos, R. A. 1998. *Personal communication*

Ennos, R. A., and T. Qian. 1994. Monitoring the output of a hybrid larch seed orchard using isozyme markers. *Forestry* **67:** 63-74.

Ennos, R. A., W. T. Sinclair, X. Hu and A. M. Langdon. 1999. Using organelle markers to elucidate the history, ecology and evolution of plant populations. pp. 1-19 in Hollingsworth, P.M., Bateman, R.H. and Gornall, R.J. (Eds.), *Molecular Systematics and Plant Evolution,* Taylor and Francis,London.

Excoffier, L., P. E. Smouse and J. M. Quattro. 1992. Analysis of Molecular Variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131:** 479-491.

Feinberg, A. P., and B. Vogelstein. 1983. A Technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* **132:** 6-13.

Flake, R. H., L. Urbatsch and B. L. Turner. 1978. Chemical documentation of allopatric introgression in *Juniperus*. *Systematic Botany* **3:** 129-144.

Forestry Commission. 1983. Census of Woodlands and Trees 1979-82 (Scotland)

Forestry Commision. 1988. Larch: A tree for all seasons.

Fowler, D. P. 1978. Population improvement and hybridization. *Unasylva* **30:** 21-26.

Frank, S. A. 1989. The evolutionary dynamics of cytoplasmic male sterility. *American Naturalist* **133:** 345-376.

Fu, Y.-X., and J. Arnold. 1992. Dynamics of cytonuclear disequilibria in finite populations and comparison with a two-locus nuclear system. *Theoretical Population Biology* **41:** 1-25.

Funk, V. 1985. Phylogenetic patterns and hybridization. *Annals of the Missouri Botanical Gardens* **72:** 681.

Gallez, G. P., and L. D. Gottlieb. 1982. Genetic evidence for the hybrid origin of the diploid plant *Stephanomeria diegensis*. *Evolution* **36:** 1158-1167.

Gielly, L., and P. Taberlet. 1994. Chloroplast DNA polymorphism at the intrageneric level and plant phylogenies. *Comptes Rendus De l'Academie Des Sciences Serie III Sciences De La Vie* **317:** 685-692.

Gillespie, J. H. 1987. Molecular evolution and the neutral allele theory. *Oxford Surveys in Evolutionary Biology* **4:** 10-37.

Gillespie, L. J., L. L. Consaul and S. G. Aiken. 1997. Hybridization and the origin of the arctic grass *Poa hatzii* (Poacea): evidence from morphology and chloroplast DNA restriction site data. *Canadian Journal of Botany* **75:** 1978-1997.

Gillies, A. 1995. *Personal communication*

Gottlieb, L. D. 1984. Genetics and morphological evolution in plants. *American Naturalist* **123:** 681-684.

Goulson, D., and K. Jerrim. 1997. Maintenance of the species boundary between *Silene dioica* and *S.latifolia* (red and white campion). *Oikos* **79:** 115-126.

Gower, S. T., and J. H. Richards. 1990. Larches deciduous conifers in an evergreen world. *Bioscience* **40:** 818-826.

Grant, P. R., and B. R. Grant. 1992. Hybridization of bird species. *Science* **256:** 193-197.

Grant, V. 1981. *Plant Speciation*. Columbia University Press, New York.

Gross, R., and J. Nilsson. 1995. Application of heteroduplex analysis for detecting variation within the growth-hormone-2 gene in *Salmo trutta* (brown trout). *Heredity* **74:** 286-295.

Gyllensten, U., and A. C. Wilson. 1987. Interspecific mitochondrial DNA transfer and the colonisation of Scandinavia by mice. *Genet.Res.Camb* **49:** 25-29.

Hacker, M., and F. Bergmann. 1991. The proportion of hybrids in seed from a seed orchard composed of 2 larch species (*L.europaea* and *L.leptolepis*). *Annales Des Sciences Forestieres* **48:** 631-640.

Hadrys, H., M. Balick and B. Schierwater. 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology* **1:** 55-63.

Hallden, C., M. Hansen, N. O. Nilsson and A. Hjerdin. 1996. Competition as a source of errors in RAPD analysis. *Theoretical and Applied Genetics* **93:** 1185-1192.

Hanson, M. R. 1991. Plant mitochondrial mutations and male sterility. *Annual Review of Genetics* **25:** 461-486.

Harlan, J. R., and J. M. J. De Wet. 1963. The compilospecies concept. *Evolution* **17:** 497-501.

Harrison, R. G. 1986. Pattern and process in a narrow hybrid zone. *Heredity* **56:** 337-349.

Harrison, R. G. 1990. Hybrid Zones:Windows on an evolutionary process., pp. 70-128 in Futuyma D, and J. Antonovics (Eds.) *Oxford Surveys in Evolutionary Biology,* Oxford University Press, Oxford.

Hauber, D. P., and W. L. Bloom. 1983. Stability of a chromosomal hybrid zone in the *Clarkia nitens* and *C.speciosa* ssp. *polyantha* complex. *American Journal of Botany* **70:** 1454-1459.

Henry, A., and M. G. Flood. 1919. The history of the Dunkeld hybrid larch, *Larix eurolepis,* with notes on other hybrid conifers. *Proceedings of the Royal Irish Academy* **35:** 55-66.

Hewitt, G. M. 1988. Hybrid zones- natural laboratories for evolutionary studies. *Trends In Ecology And Evolution.* **3:** 158-167.

Hillis, D. M., and C. Moritz. 1996. *Molecular Systematics*. Sinauer, Sunderland.

Hodges, S. A., J. Burke and M. L. Arnold. 1995. Formation of a hybrid population: Production of F1 progeny constrain the frequency and genotypes of future hybrid generations. *American Journal Of Botany* **82:** 72.

Holtmeier, F. 1992. European larch in middle Europe with special references to the central Alps, pp. 41-49 in Schmidt W.C, and K. J. McDonald (Eds.) *Ecology and management of Larix forests:A look ahead*, Intermountain Research Station, Whitefish,Montana.

Hongyo, T., G. S. Buzard, R. J. Calvert and C. M. Weghorst. 1993. Cold SSCP - a simple, rapid and nonradioactive method for optimized single-strand conformation polymorphism analyses. *Nucleic Acids Research* **21:** 3637-3642.

Howard, D. J. 1993. Reinforcement: Origin, dynamics and fate of an evolutionary hypothesis, pp. 46-69 in Harrison R.G (Ed.) *Hybrid Zones and the Evolutionary Process*, OUP, Oxford.

Howard, D. J., R. W. Preszler, J. Williams, S. Fenchel and W. J. Boecklen. 1997. How discrete are oak species? Insights from a hybrid zone between *Quercus grisea* and *Quercus gambelii*. *Evolution* **51:** 747-755.

Hu, X. 1998. Genetic marker studies of the *Larix gmelinii* complex and the development of genetic marker theory for plant populations, PhD Thesis (Edinburgh).

Hudson, R. R. 1990. Gene genealogies and the coalescent process. *Oxford Surveys In Evolutionary Biology* **7:** 1-44.

Hughes, M. A. 1996. *Plant Molecular Genetics*. Longman, Harlow.

Hutchison, K.W. 1989. Nucleotide sequence of the small subunit of ribulose-1, 5-bisphosphate carboxylase from the conifer *Larix laricana*.Genbank Database (accession X16039)

Janich, J., R. Scherg, F. W. Woods and V. W. Rurron. 1981. *Plant Science: an introduction to world crops.* Freeman and co., San Francisco.

Jiggins, C. D., W. O. Mcmillan, W. Neukirchen and J. Mallet. 1996. What can hybrid zones tell us about speciation? The case of *Heliconius erato* and *H. himera* (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society* **59:** 221-242.

Jinks, J. L., and R. M. Jones. 1958. Estimation of the components of heterosis. *Genetics* **43:** 223-234.

Jonseil, B. 1984. The BSC concept reexamined, pp. 159-168 in Grant. W (Ed.) *Plant Biosystematics*, Academic Press, London.

Kadereit, J. W. 1994. Molecules and morphology, phylogeny and genetics. *Botanica Acta* **107:** 369-373.

Keim, P., K. N. Paige, T. G. Whitham and K. G. Lark. 1989. Genetic-analysis of an interspecific hybrid swarm of *Populus* -occurrence of unidirectional introgression. *Genetics* **123:** 557-565.

Kimura, M. 1983. *The neutral theory of molecular evolution.* Cambridge University Press.

Klier, K., M. J. Leoschke and J. F. Wendel. 1991. Hybridization and introgression in white and yellow ladyslipper orchids. *Journal of Heredity* **82:** 305-319.

Knobloch, I. W. 1971. Intergeneric hybridisation in flowering plants. *Taxon* **21:** 97-103.

Laing, E. V. 1947. Studies on the genus *Larix*. *Scottish Forestry* **58:** 6-32.

LePage, B., and J. F. Basinger. 1992. The evolutionary history of the genus *Larix*, pp. 19-29 in Schmidt W.C and K. J. McDonald (Eds.) *Ecology and management of Larix forests:A look ahead*, Intermountain Research Station, Whitefish, Montana.

Lessa, E. P., and G. Applebaum. 1993. Screening techniques for detecting allelic variation in DNA-sequences. *Molecular Ecology* **2:** 119-129.

Levin, D. A., J. Francisco-Ortega and R. K. Jansen. 1990. Hybridization and the extinction of rare plant species. *Conservation Biology* **10:** 10-16.

Lewontin, R. C.1988. On measures of gametic disequilibrium. *Genetics* **120:** 849-852.

Lewontin, R. C., and L. C. Birch. 1966. Hybridization as a source of variation for adaptation to new environments. *Evolution* **20:** 315-336.

Li, B. L., and G. W. Wyckoff. 1994. Breeding strategies for *Larix decidua, L.leptolepis* and their hybrids in the U.S. *Forest Genetics* **1:** 65-72.

Lidholm, J., and P. Gustafsson. 1991. The chloroplast genome of the gymnosperm *Pinus-contorta* - a physical map and a complete collection of overlapping clones. *Current Genetics* **20:** 161-166.

Lorenz, M., A. Weihe and T. Boerner. 1994. DNA fragments of organellar origin in RAPD patterns of sugar beet (*Beta vulgaris* L.). *Theoretical and Applied Genetics* **88:** 775-779.

Lynch, M., and B. G. Milligan. 1994. Analysis of population genetic-structure with RAPD markers. *Molecular Ecology* **3:** 91-99.

Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* **46:** 523-536.

Manly, B. F. J. 1986. *Multivariate statistical methods.* Chapman and Hall, London.

Marayuma, T., and C. W. Birky. 1991. Effects of periodic selection on gene diversity in organelle genomes and other systems without recombination. *Genetics* **127:** 449-451.

Marchant, A. D. 1988. Apparent introgression of mitochondrial DNA across a narrow hybrid zone in the *Caledia captiva* species complex. *Heredity* **60:** 39-46.

Mason, W. L., and E. Baldwin. 1994. Performance of pedunculate oak after 40 years in mixture with European larch and Norway spruce in Scotland. *Scottish Forestry* **49:** 5-14.

Forestry Commission. 1954. Japanese larches at Dunkeld, Perthshire

Mayr, E. 1942. *Systematics and the Origin of Species.* Columbia University Press, New York.

Mccauley, D. E. 1994. Contrasting the distribution of chloroplast dna and allozyme polymorphism among local-populations of *Silene alba* -implications for studies of gene flow in plants. *Proc. Natl. Acad. Sci. USA* **91:** 8127-8131.

Mcdade, L. 1990. Hybrids and phylogenetic systematics i. Patterns of character expression in hybrids and their implications for cladistic analysis. *Evolution* **44:** 1685-1700.

Mcdade, L. 1993. Species concepts insight from groups that are difficult at the species level. *American Journal of Botany* **80:** 117.

Mitchell, A. F. 1963. The history of the introduction of the hybrid larch to Britain. *Scottish Forestry* **17**.

Mogensen, H. L. 1996. The Hows and Whys of Cytoplasmic Inheritance in Seed Plants. *American Journal of Botany* **83:** 383-404.

Mooney, H. A., and J. A. Drake. 1986. *Ecology of biological invasions of North America and Hawaii.*

Moore, W. S. 1977. An evaluation of narrow hybrid zones in vertebrates. *Quarterly Review of Biology* **52:** 263-277.

Morman, J. 1994. *Personal communication*

Mullis, K., and F. Faloona. 1987. The Polymerase Chain Reaction, pp. 335-350 in *Methods in enzymology*, edited by R. Wu. Academic Press, London.

Nason, J. D., and N. C. Ellstrand. 1993. Estimating the frequencies of genetically distinct classes of individuals in hybridized populations. *Journal of Heredity* **84:** 1-12.

Neale, D. B., K. A. Marshall and R. R. Sederoff. 1988. Inheritance of Chloroplast and Mitochondrial DNA in Conifers, pp.89-100 in Hallgren J.E (ed.), Proc. Frans Kempe Symp., *Molecular Genetics Of Forest Trees*, Umea.

Nei, M. 1987. *Molecular evolutionary genetics.* Columbia University Press, New York.

Nelson, E. C. 1980. What is the correct name for the Dunkeld hybrid larch? *Irish Forestry* **37:** 112-118.

Nixon, K. C., and Q. D. Wheeler. 1990. An amplification of the phylogenetic species concept. *Cladistics* **6:** 211-223.

Olmstead, R. G., and J. A. Sweere. 1994. Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the Solanaceae. *Systematic Biology* **43:** 467-481.

Ostenfeld, C. H., and C. S. Larsen. 1930. *The Species of the Genus Larix and their Geographical Distribution*. Det. Kgl. Danske Videnskabernes Selskab. Biologiske Meddelelsker.

Paige, K. N., W. C. Capman and P. Jennetten. 1991. Mitochondrial inheritance patterns across a cottonwood hybrid zone - cytonuclear disequilibria and hybrid zone dynamics. *Evolution* **45:** 1360-1369.

Palmer, J. D. 1990. Contrasting modes and tempos of genome evolution in land plant organelles. *Trends In Genetics* **6:** 115-120.

Palmer, J. D. 1992. Mitochondrial DNA in plant systenatics:applications and limitations, pp. 36-49 in *Molecular Systematics of Plants*, edited by P. E. Soltis, D. E. Soltis and J. J. Doyle. Chapman and Hall, New York.

Panting, R. 1997. *Personal communication*

Paques, L. E. 1989. A critical-review of larch hybridization and its incidence on breeding strategies. *Annales Des Sciences Forestieres* **46:** 141-153.

Perron, M., and J. Bousqet. 1997. Natural hybridization between black spruce and red spruce. *Molecular Ecology* **6:** 725-734.

Pharmacia Biotech. 1996. Ready to Go DNA Labelling Beads.

Potts, B. M., and J. B. Reid. 1988. Hybridization as a dispersal mechanism. *Evolution* **42:** 1245-1255.

Powell, J. R. 1983. Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from *Drosophila*. *Proc. Natl. Acad. Sci. USA* **80:** 492-495.

Pradhan, A. K., S. Prakash, A. Mukhopadhyay and D. Pental. 1992. Phylogeny of *brassica* and allied genera based on variation in chloroplast and mitochondrial DNA patterns molecular and taxonomic classifications are incongruous. *Theoretical and Applied Genetics* **85:** 331-340.

Pyle, R. L., and J. E. Randel. 1994. A review of hybridization in marine angelfishes. *Environmental Biology of Fishes* **41:** 127-145.

Python, M. 1969. How to recognise trees from quite a long way away. *Monty Python's flying circus* **3**

Qiagen. 1997. QIAquick PCR and Gel Extraction Kits.

Qian, T., R. A. Ennos and T. Helgason. 1995. Genetic relationships among larch species based on analysis of restriction fragment variation for chloroplast DNA. *Canadian Journal of Forest Research* **25:** 1197-1202.

Randolph, L. F., I. S. Nelson and R. L. Plaisted. 1967. Negative evidence of introgression affecting the stability of Louisiana Iris species. *Cornell University Agricultural Experimental Station Memoir* **398:** 1-56.

Raybould, A., A. Gray, M. Lawrence and D. Marshall. 1991. The evolution of *Spartina anglica* C.E Hubbard: the origin and genetic variability. *Biological Journal of the Linnean Society* **43:** 111-126.

Raymond, M., and F. Rousset. 1997. Genepop 3.1b (shareware for PC compatible computers), Montpellier,France .

Rieseberg, L. H. 1991. Homoploid reticulate evolution in *Helianthus* (asteraceae) -evidence from ribosomal genes. *American Journal of Botany.* **78:** 1218-1237.

Rieseberg, L. H. 1996. Homology among RAPD fragments in interspecific comparisons. *Molecular Ecology* **5:** 99-105.

Rieseberg, L. H. 1997. Hybrid origins of plant species. *Annual Review of Ecology and Systematics* **28:** 359-389.

Rieseberg, L. H., S. Beckstrom-Sternberg and K. Doan. 1990a. *Helianthus annuus* ssp *texanus* has chloroplast DNA and nuclear ribosomal rna genes of *Helianthus-debilis* ssp *cucumerifolius*. *Proc. Natl. Acad. Sci. USA* **87:** 593-597.

Rieseberg, L. H., S. M. Beckstrom-Sternberg, A. Liston and D. M. Arias. 1991a. Phylogenetic and systematic inferences from chloroplast DNA and isozyme variation in *Helianthus asteraceae*. *Systematic Botany.* **16:** 50-76.

Rieseberg, L. H., R. Carter and S. Zona. 1990b. Molecular tests of the hypothesised hybrid origin of two diploid *Helianthus* species. *Evolution* **44:** 1498-.

Rieseberg, L. H., H. C. Choi and D. Ham. 1991b. Differential cytoplasmic versus nuclear introgression in Helianthus. *Journal of Heredity* **82:** 489-493.

Rieseberg, L. H., and N. C. Ellstrand. 1993. What can molecular and morphological markers tell us about plant hybridization? *Critical Reviews in Plant Sciences* **12:** 313-241.

Rieseberg, L. H., B. Sinervo, C. R. Linder, M. C. Ungerer and D. M. Arias. 1996. Role of gene interactions in hybrid speciation: Evidence from ancient and experimental hybrids. *Science* **272:** 741-745.

Rieseberg, L. H., and D. E. Soltis. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends In Plants* **5:** 65-84.

Rieseberg, L. H., F. C. Van and A. M. Desrochers. 1995. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* **375:** 313-316.

Rieseberg, L. H., and J. F. Wendel. 1993. Introgression and its consequences in plants., pp. 70-109 in Harrison R.G (Ed.) *Hybrid zones and the evolutionary process*, Oxford University Press, Oxford.

Roberts, D. F., and R. W. Hiorns. 1965. Methods of analysis of the genetic composition of a hybrid population. *Human Biology* **37:** 38-43.

Rodriguez, A., and D. M. Spooner. 1997. Chloroplast DNA analysis of *Solanum bulbocastanum* and *S. cardiophyllum*, and evidence for the distinctiveness of *S. cariophyllum* subsp. *ehrenbergii* (sect. Petota). *Systematic Botany* **22:** 31-43.

Saitou, N., and M. Nei. 1986. The number of nucleotides required to determine the branching order of three species, with special reference to the human-chimpanzee-gorilla divergence. *Journal of Molecular Evolution* **24:** 189-204.

Sambrook, J., E. F. Fritsch and T. Maniatis. 1989. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, New York.

Sang, T., D. J. Crawford and T. F. Stuess. 1995. Documentation of reticulate evolution in peonies using ITS sequences of nrDNA: implications for biogeography and concerted evolution. *Proc. Natl. Acad. Sci. USA* **92:** 6813-6817.

Schabel, H. G., C. H. Lee and G. W. Wyckoff. 1992. Exotic larches: experiences in Wisconsin, pp. 19-29 in Schmidt W.C and K.J. McDonald (Eds.) *Ecology and management of Larix forests:A look ahead*, Intermountain Research Station, Whitefish, Montana.

Schmidt, W. C., 1992 Around the world with *Larix:* an introduction, pp. 6-18 in Schmidt W.C and K.J. McDonald (Eds.) *Ecology and management of Larix forests:A look ahead*, Intermountain Research Station, Whitefish, Montana.

Schnabel, A., and M. A. Asmussen. 1992. Comparative effects of pollen and seed migration on the cytonuclear structure of plant populations ii. Paternal cytoplasmic inheritance. *Genetics* **132:** 253-267.

Scotlab, 1996 Nucleon Phytopure Plant DNA Extraction Kit

Semerikov, V. L., L. F. Semerikov and M. Lascoux. 1998. Intra- and interspecific allozyme variability in Eurasian *Larix* Mill.species. *In press* .

Sinclair, W. T., J. D. Morman and R. A. Ennos. 1998. Multiple origins for Scots pine in Scotland: evidence from mitochondrial DNA variation. *Heredity* **80:** 233-240.

Smith, R. L., and K. J. Sytsma. 1990. Evolution of *Populus-nigra* section aigeiros: introgressive hybridization and the chloroplast contribution of populus-alba section populus. *American Journal of Botany* **77:** 1176-1187.

Soltis, D. E., and R. K. Kuzoff. 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* **49:** 727-742.

Soltis, D. E., D. R. Morgan, A. Grable, P. S. Soltis and R. Kuzoff. 1993. Molecular systematics of saxifragaceae sensu stricto. *American Journal of Botany* **80:** 1056-1081.

Soltis, D. E., and P. S. Soltis. 1993. Molecular data and the dynamic nature of polyploidy. *Critical Reviews in Plant Science* **12:** 243-275

Soltis, D. E., P. S. Soltis and J. J. Doyle (Editors). 1990. *Molecular Systematics of Plants.* Chapman and Hall, New York.

Soltis, D. E., P. S. Soltis, D. L. Nickrent, L. A. Johnson, W. J. Hahn *et al.* 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Annals of the Missouri Botanical Gardens* **84:** 1-49.

Stace, C. A. 1991. *The New Flora of the British Isles.* Cambridge University Press, New York.

Stebbins, G. L. 1950. *Variation and Evolution in Plants.* Columbia University Press, New York.

Strand, A. E., J. Leebens-Mack and B. G. Milligan. 1997. Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology* **6:** 113-118.

Sutton, B. C. S., S. C. Pritchard, J. R. Gawley, C. H. Newton and G. K. Kiss. 1994. Analysis of sitka spruce - interior spruce introgression in British- Columbia using cytoplasmic and nuclear-DNA probes. *Canadian Journal of Forest Research* **24:** 278-285.

Szmidt, A. E. 1991. Phylogenetic and applied studies on chloroplast genome in forest conifers., pp. 185-196 in *Biochemical Markers in the Population Genetics of Forest Trees*, edited by S. Fineschi, M. E. Malvolti, F. Cannata and H. H. Hattemer. SPB Academic Publishing, The Hague.

Szmidt, A. E., T. Alden and J. Hallgren. 1987. Paternal inheritance of chloroplast DNA in *Larix*. *Plant Molecular Biology* **9:** 59-64.

Takei, F. 1992. Distribution and management of Japanese larch, pp. 58-65 in Schmidt W.C and K.J. McDonald (Eds.) *Ecology and management of Larix forests:A look ahead*, Intermountain Research Station, Whitefish, Montana.

Taylor, C. J.,1964. The place of larch in Scottish silviculture. *Scottish Forestry* **18:** 189-198.

Templeton, A. R. 1981. Mechanisms of speciation- a population genetic approach. *Annual Review of Ecology and Sytematics* **12:** 23-48.

Templeton, A. R. 1989. The meaning of species and speciation: a genetic perspective., pp. 3-27 in *Speciation and its consequences.*, edited by D. Otte, and J. A. Endler. Sinauer, Sunderland.

Templeton, A. R. 1996. Contingency tests of neutrality using intra-interspecific gene trees: the rejection of neutrality for the evolution of mitochondrial cytochrome oxidase II gene in the hominoid primates. *Genetics* **144:** 1263-1270.

Van Raamsdonk, L. W. D. 1995. The effect of domestication on plant evolution. *Acta Botanica Neerlandica* **44:** 421-438.

Venkatesh, C. S., and Vakshasya. 1977. Effect of selfing, crossing and interspecific hybridization in *Eucalyptus camaldulensis*, in *3rd World Consortium On Forest Tree Breeding*, Canberra.

Vidakovic, M., 1991 *Conifers: morphology and variation*. Graficki Zavod Hrvatske.

Wang, X. R., and A. E. Szmidt. 1990. Evolutionary analysis of *Pinus densata* (Masters) a putative Tertiary hybrid.II.(CpDNA variation). *Theoretical and Applied Genetics* **80:** 641-647.

Wang, X. R., A. E. Szmidt, A. Lewandowski and Z. R. Wang. 1990 .Evolutionary analysis of *Pinus densata* (Masters) a putative Tertiary hybrid.I.(Allozyme variation). *Theoretical and Applied Genetics* **80:** 635-640.

Wang, X.-R., and A. E. Szmidt. 1994. Hybridization and chloroplast DNA variation in a *Pinus* species complex from Asia. *Evolution* **48:** 1020-1031.

Warwick, S. I., J. F. Bain, R. Wheatcroft and B. K. Thompson. 1989. Hybridization and introgression in *Carduus nutans* and *Carduus acanthoides* reexamined. *Systematic Botany* **14:** 476-494.

Watano, Y., M. Imazu and T. Shimizu. 1995. Chloroplast DNA typing by PCR-SSCP in the *Pinus pumila-P. parviflora* var. *pentaphylla* complex (Pinaceae). *Journal of Plant Research* **108:** 493-499.

Watano, Y., M. Imazu and T. Shimizu. 1996. Spatial distribution of CpDNA and MtDNA haplotypes in a hybrid zone between *Pinus pumila* and *P. parviflora* var. *pentaphylla* (Pinaceae). *Journal of Plant Research* **109:** 403-408.

Waters, E. R. 1995. An evaluation of the usefulness of the small heat shock genes for phylogenetic analysis in plants. *Annals of the Missouri Botanical Gardens* **82:** 278-295.

Weir, B. S. 1990. *Genetic Data Analysis.* Sinauer Associates, Sunderland, Massachusetts.

Wendel, J. F., and V. A. Albert. 1992. Phylogenetics of the cotton genus (*Gossypium*): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. *Systematic Botany* **17:** 115-143.

Wendel, J. F., A. Schnabel and T. Seelanan. 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc. Natl. Acad. Sci. USA* **92:** 280-284.

Wendel, J. F., J. M. Stewart and J. H. Rettig. 1991. Molecular evidence for homoploid reticulate evolution among australian species of *Gossypium*. *Evolution* **45:** 694-711.

Whittemore, A. T., and B. Schaal. 1991. Interspecific gene flow in sympatric oaks. *Proc. Natl. Acad. Sci. USA* **88:** 2540-2544.

Whitton, J., D. E. Wolf, D. M. Arias, A. A. Snow and L. H. Rieseberg. 1997. The persistence of cultivar alleles in wild populations of sunflowers five generations after hybridization. *Theoretical and Applied Genetics* **95:** 33-40.

Wilson, M. A., B. Gaut and M. T. Clegg. 1990. Chloroplast DNA evolves slowly in the palm family Arecaceae. *Molecular Biology and Evolution* **7:** 303-314.

Wilson, P. 1992. On inferring hybridity from morphological intermediacy. *Taxon* **41:** 11-23.

Wirawan, E. E. 1997. Measuring the genetic efficiency of hybrid larch seed orchards using isozyme markers., pp. 80 in *Institute of Ecology and Resource Management*. Edinburgh, Edinburgh.

Wise, C. A., M. Sraml and S. Easteal. 1998. Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. *Genetics* **148:** 409-421.

Wolfe, A., and W. Elisens. 1994. Nuclear ribosomal DNA restriction-site variation in *Penstemon* section *Peltanthera* (Scrophulariaceae): An evaluation of diploid hybrid speciation and evidence for introgression. *American Journal of Botany* **81:** 1627-1635.

Woodruff, D. S. 1989. Genetic anomalies associated with *Cerion* hybrid zones: the origin and maintenance of the new electromorphic variants called hybrizymes. *Biological Journal of the Linnean Society* **63:** 281-288.

Wright, S. 1951. The genetical structure of populations. *Annals of Eugenics* **15:** 323-354.

Yap, E. P. H., and J. McGee. 1992. Nonisotopic SSCP detection in PCR products by ethidium bromide staining. *Trends in Genetics* **8:** 49.

Yde-Andersen, A. 1979. Host spectrum, host morphology and geographic distibution of larch canker, *Lachnellula willkommii*. *European Journal of Forest Pathology* **9:** 211-219.

Young, N. D. 1996a. An analysis of the causes of genetic isolation in two Pacific Coast iris hybrid zones. *Canadian Journal of Botany* **74:** 2006-2013.

Young, N. D. 1996b. Concordance and discordance: A tale of two hybrid zones in the Pacific Coast irises (Iridaceae). *American Journal of Botany* **83:** 1623-1629.

Zobel, B., and J. Talbert. 1981. Hybrids in tree improvement., pp. 345-373 in *Applied Forest Tree Improvement*. Waveland Press, Illinois.