Fall 1990

Genetic variability and population structure of eastern white pine in a two hectare stand, Durham, New Hampshire

Petr R. Brym
University of New Hampshire, Durham

Order Number 9108311

Genetic variability and population structure of eastern white pine in a two hectare stand, Durham, New Hampshire

Bryn, Petr R., Ph.D.

University of New Hampshire, 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

GENETIC VARIABILITY AND POPULATION STRUCTURE OF EASTERN WHITE PINE IN A TWO HECTARE STAND, DURHAM, NEW HAMPSHIRE

By

Petr Brym
BS Stanford University, 1981

DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy

in

Genetics

September, 1990

This dissertation has been examined and approved.

Robert T. Eckert

Dissertation director, Robert T. Eckert,
Associate Professor of Forest Genetics

Yun-Tzu Kiang

Yun-Tzu Kiang, Professor of Plant Biology
and Genetics

Thomas D. Lee

Thomas D. Lee, Associate Professor of
Plant Biology

Albert E. Luloff

Albert E. Luloff, Associate Professor of
Leisure Management and Tourism and
Resource Administration and Management

C. Tattersall Smith, Jr.

C. Tattersall Smith, Associate Professor of
Forest Ecology and Soils

July 18, 1990

Date

ACKNOWLEDGEMENTS

I would like to thank Dr. Robert T. Eckert for his patience, guidance, and for the opportunity to learn; my dissertation committee members Dr. Yun-Tzu Kiang, Dr. Thomas D. Lee, Dr. Albert E. Luloff, and Dr. C. Tattersall Smith; Dr. James P. Barrett and Dr. Robert Westfall for their advice on statistical data analysis; Dr. Donald M. Green for his encouragement; Joseph Danahy, Linda Isaacson, Mary Skip and Karen Stapelfeldt for their technical assistance; John Carlson and Robert Cooke for always being there when I needed their help; my parents Franta and Eva Kalousek for teaching me about the importance of education; and my wife Donna and daughter Katie for their endless patience. Katie, I hope this work saves at least one tree for you and your children.

PRB

This study was carried out as part of the McIntire-Stennis Project 11 at the New Hampshire Agricultural Experiment Station, University of New Hampshire, and Regional Project NE-27.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

GENETIC VARIABILITY OF EASTERN WHITE PINE IN A TWO HECTARE STAND

by

Petr Brym
University of New Hampshire, September, 1990

The genetic structure of a two hectare stand of eastern white pine (Pinus strobus L.) was studied by relating foliar allozyme composition of individual trees to tree location and age. Using starch gel electrophoresis, the trees were found to have variants for six of the seven enzymes analyzed. Trees of similar age and with common allozymes were found in clusters, suggesting that genetic structure existed in the stand. The genetic structure may have originated during establishment of several seedlings from common parents when openings were created in the canopy. Knowing that a genetic structure, such as that observed in this study, may exist in "natural" forests is important to foresters who are concerned with maintaining current levels of genetic variability in those forests.

Evaluation of sampling strategies revealed that at least 50 randomly sampled trees would be required to correctly estimate allele frequencies, the number of alleles per locus, and heterozygosity in this stand. Smaller samples were likely to yield incorrect estimates by chance. To estimate polymorphism, between 100 and 250 trees would be needed. Reliable detection of deviations from allelic frequencies expected under the Hardy-Weinberg equilibrium required 25 to 425 randomly chosen trees, depending on allozyme frequencies for the enzyme in question. Differences in estimates for larger samples compared to smaller ones were due mainly to detection of additional low frequency alleles in the larger samples. These results indicate that cost and labor can be saved by selecting approximately ten percent of trees in a stand such as the one used in this study to obtain estimates for the first three measures, but more extensive and costly samples must be taken when estimating polymorphism and testing for deviation from the Hardy-Weinberg Equilibrium.

INTRODUCTION

Eastern white pine has been harvested, often indiscriminately, since the colonization of the United States. It has been subjected to clear cutting and "high-grading", or removal of the most desirable phenotypes for timber production, as well as more reasonable silvicultural practices. Although much is known about geographic variation in growth and survival of the species (Kuser and Hobbs 1985, Genys 1977, Jeffers 1977, Garrett et al. 1973), and recently, information regarding variation at the allelic level has become available (Eckert and Ryu 1983, Ryu and Eckert 1983, Eckert et al. 1981), the distribution of genetic variation at the local stand level is not well known. Understanding genetic variability of white pine within small geographic areas is important because, as our use of the land intensifies we more directly affect which trees remain in individual stands, and which are allowed to provide seeds for future generations of this very important resource. Forested land in many sections of the Northeast is continuously divided into smaller tracts of forests due to human population growth, as small sections are separated

from other forested land with highways, housing developments and commercial areas. These smaller tracts of land are in danger of unpredictable changes in their genetic resources due to random gene frequency fluctuations commonly observed in small populations. Chance variations in allelic frequencies similar to those due to the founder effect (new population is established by only very few individuals) occur when populations go through such bottlenecks (Ayala 1982). In situations where only a few trees establish a new population or where only a few trees contribute to the next generation in an existing small size stand, a limited gene pool is available. In such situations the next generation of trees also has a limited gene pool to contribute, and that gene pool can have very different gene frequencies compared to the previous generation of trees. Unfortunately, in such situations the change in gene frequencies can be more influenced by chance than by selection, and undesirable genes are just as likely to increase in frequency as are desirable genes. We must learn about local genetic variability before it is irreversibly changed by our activities. Having such information is a prerequisite to understanding what level of genetic variability is needed in white pine stands to allow for successful long-term reproduction and growth of healthy trees.

Without this knowledge protecting our future forests from disease and potential climate changes will be more difficult.

Previous genetic studies of white pine populations, using variation in foliar isozyme to represent genetic variability, have focused on an understanding of genotype distribution over both large and small geographic areas of the species' range. None of these studies included all the trees in the stand (Eckert and Ryu 1983, Ryu and Eckert 1983, Eckert et al. 1981). The traditional approach required a sampling protocol which selected a few individuals of each stand per study. The number of trees used to sample natural populations varies widely around 50 (Ryu 1982; Grant and Mitton 1977; Copes 1975; Rudin et al. 1974), but reported sample sizes have more meaning when the sizes of stands represented by the samples also are reported as by Roberds and Conkle (1984), Feret (1974), and Tigerstedt (1973). Reliability of the genetic measures used to estimate genetic variability of stands based on isozyme analysis cannot be verified if the true genetic variability of the stands is not known.

The need for an evaluation of sampling intensities used in isozyme studies of forest tree populations is clear. In its absence reliable sampling methods and intensities cannot be ensured. Validation of sampling methods at a minimum must include intensive data collection in a stand, followed by various strategies for sampling individual trees in order to determine which approach yields the most reliable and accurate results at a reasonable cost. Minimizing the cost of sampling forest stands in genetic studies while ensuring accurate estimates of genetic variability is important because funds provided to researchers can be used more effectively and requests for funding are more likely to be approved for efficiently planned studies. Without funding, or without efficient use of existing funding, our efforts to better understand the genetic variability of white pine, so we can better manage it for future generations, will suffer.

Demographic studies of isozyme patterns have potential for use in assessment of silvicultural impacts on genetics and regeneration in residual stands, but, other than a report by Ledig and Smith (1981), the influence of silvicultural practice on the

genetics of white pine has not been evaluated. In their paper, Ledig and Smith primarily attributed improvement in height of white pine progeny to a reduction in inbreeding. If reduction in inbreeding can help produce better quality white pines, then a serious effort should be made to minimize inbreeding in the remaining "natural" white pine stands. Further, if desirable phenotypic traits of white pines can be correlated with measurable genetic variability, then desirable parents can be identified in the field and silvicultural practices can be employed to help ensure their contribution to future generations. The first step toward this goal would include the development of reliable methods of measuring genetic variability of white pines, and an evaluation of genetic variability, especially as it relates to tree age, location, and stand structure. One procedure for measuring genetic variability in trees is through evaluation of the trees' isozymes. Although isozyme analysis provides information for only about 0.1% of the tree's DNA (Feret 1979), it is one of the very few both relatively affordable and quick procedures available to forest geneticists for evaluating the genetic variability of individual trees and tree populations. Evidence from the literature suggests that heterosis and isozyme heterogeneity are correlated in plants

(Gupta and Singh 1977, Schwartz and Laughner 1969) and that "hybrid molecules" may be somewhat "better" than "non-hybrid" molecules (Scandalios 1974, Scandalios, Liu and Campeau 1972). If this is also true for trees, isozyme analysis would be useful in : (1) measuring homozygosity and heterozygosity, (2) detecting heteromultimers, and (3) tracking genetic markers to evaluate parent contribution to future generations both in unmanaged forests and under controlled conditions. Foliar isozyme analysis is convenient for studying white pine genetic population structure, because foliar samples can be collected in the field quickly, whether trees are located in remote areas, or adjacent to populated regions (Appendix A). Laboratory methods for analyzing many white pine foliar isozymes are available (Ryu 1982, Eckert et al. 1981, Adams and Joly 1980, Gorman and Kiang 1977) and methods for additional enzyme systems are being developed. Methods also exist for preliminary confirmation that the isozymes used for such studies correspond to alleles and therefore are under genetic control (Eckert et al. 1981). These methods test whether isozymes in haploid megagametophytes from heterozygous white pines segregate as they would if they were under genetic control.

The next three chapters address the following objectives: (1) determine electrophoretic variability of 7 foliar enzymes in 511 eastern white pines from a local stand; (2) demonstrate evidence for genetic control of those variants; (3) test for the relationship between tree age and location to the tree's genotype; (4) identify the stand structure of the white pines; and (5) evaluate the reliability of the genetic measures used.

Chapter 1

STUDY OF VARIABILITY OF SEVEN ENZYMES OF EASTERN WHITE PINE USING STARCH GEL ELECTROPHORESIS

INTRODUCTION

Understanding the distributions of trees with similar genotypes can help in evaluating how past, current, and future forest management activities affect the genetic composition of forests. A detailed record of genetic variability in forests can help in searching for possible family structure in forest stands, and in identifying any possible correlations between genetic composition and the quality of the trees in the stands.

Brown and Moran (1981) indicate that isozymes are ideal for monitoring base-line genetic variability, increased vulnerability in plantations, and loss of variation resulting from ecosystem

destruction. Isozyme studies can yield information about the genetic structure of forests more quickly than can traditional field and nursery tests because isozyme methods do not depend on the lengthy reproductive and generation cycle of the white pine. Although DNA (deoxyribonucleic acid) studies can yield more precise information, they are more expensive than are isozyme studies, and it is difficult to analyze large numbers of samples necessary to assess within-species variation. RFLP (restriction fragment length polymorphism) analysis is analysis of DNA fragments, and fragments are not likely to represent functional genes. Random fragments of DNA therefore do not represent genetic information necessary to synthesize functional enzymes. The genetic significance of RFLP variation is therefore not known. Isozymes (enzymes) are products of functional genes and their structure reflects the structure of those genes. Differences in protein structure can affect the protein's functionality (ability to catalyze biochemical reactions) and may alter its behavior under electrophoretic conditions. Some biologically significant differences in genes may therefore be detected by studying differences in isozymes through electrophoretic methods. With the constant removal of natural and semi-natural forest stands, now is

the time to record and evaluate the genetic composition of local forests. Past use and management of forests in the northeastern United States may have irreversibly changed the tree populations and current genetic composition of the forests should be recorded before further change occurs. This record can be used in conjunction with other current studies of the stands, and it may prove to be an invaluable reference for forest management in future years.

Analysis of isozyme gene expression allows researchers to examine the structure of genetic variability in a population with minimum interference from environmental factors. In the case of eastern white pine, isozyme analysis can be accomplished without long-term experiments and without causing harm to the trees. Foliar or megagametophyte isozymes can be used to infer the tree's genotype, and megagametophyte isozymes can be used to obtain preliminary evidence that the isozymes used to infer the tree's genotype are under genetic control. Foliar isozymes are often more convenient for inferring the tree's genotype, because not all trees produce cones, and cones are not produced each year. With some restrictions to ensure that seasonal effects do not influence

results, foliar isozymes can be used for all trees, each year. When available, megagametophyte isozymes are easier to analyze than foliar isozymes, because the enzymes of interest are not exposed to as many destructive chemical compounds in seeds, as they are in needles during extraction procedures in the laboratory.

Although foliar and seed isozymes of trees are used mostly as markers and to study genetic variability without knowing the biological significance of the isozymes, some studies already have demonstrated that isozyme variability can be linked to variation in traits more commonly considered in practice. Ryu and Eckert (1983) found a statistically significant correlation between height growth of eastern white pine in provenance test plantations and malic enzyme, LAP (leucine aminopeptidase), and PGM (phosphoglucomutase) allele frequencies. The authors also analyzed allelic frequencies at ten loci of eastern white pine in relation to geographic origin of clones (Ryu and Eckert 1983). The authors identified four clusters of white pine sources suggesting that broad ecotypes of white pine exist. Ryu and Eckert (1983) did not intensively sample many trees at each location to determine whether local groups of related trees exist within individual stands. Understanding local variability and adaptation to local

conditions is important for foresters because success of forest management efforts depends on the trees' response to both broad and local ecological conditions. In the following chapter, the study of the genetic variability of 511 eastern white pines from one forest stand in New Hampshire is reported.

MATERIALS AND METHODS

The study site

The study site is an "unmanaged" two hectare forest stand in the town of Durham, New Hampshire. Stand boundaries were based on tree species composition, physical barriers and former silvicultural treatments of adjacent stands. The stand contains eastern white pine and a variety of other evergreen and deciduous trees such as eastern hemlock (Tsuga canadensis (L.) Carr), white oak (Quercus alba L.), black birch (Betula lenta L.), and sugar maple (Acer saccharum Marsh). Management records do not show any silvicultural activity in the stand since 1962. The stand was selected for its interesting microsite variability and includes dry and wet areas, low and elevated sites, deep glacial till soils and massive rock outcrops. Eastern white pine is found in all these varied locations. Very few one and two year old seedlings are present and approximately eighty percent of the pines had a diameter greater than 14 cm. Diameter measurements were taken 30 cm from the forest floor. Average age, based on a random sample of one third of the trees, was sixty years.

Isozyme analysis

Foliar tissue samples were collected from all 511 eastern white pines. A slingshot was used to remove one year old foliage from sun-exposed portions of tree crowns (see Appendix A). The foliar samples were too high for pole pruners, and the stand was too close to a populated region to allow the use of a rifle. Samples were stored on ice while in the field and frozen (- 20°C) in the laboratory until analyzed. Horizontal starch gel electrophoresis was used to determine allelic composition of foliar samples for the following seven enzymes: leucine aminopeptidase (LAP E.C. 3.4.11.1), phospho-glucose isomerase (PGI E.C. 5.3.1.9.), glutamate-oxaloacetate transaminase (GOT E.C. 2.6.1.1), shikimate dehydrogenase (SDH E.C. 1.1.1.25), fluorescent esterase (F-EST E.C. 3.1.1.2) phosphoglucomutase (PGM E.C. 2.7.5.1), and malate dehydrogenase (MDH E.C. 1.1.1.37). Each sample was frozen in liquid nitrogen and ground to powder with a mortar and pestle. An extraction buffer (Table 1.1) was added, the mixture was stirred until a paste formed, and then the extracts were absorbed into paper wicks through lens paper covering the paste (Gorman and Kiang 1977). Electrophoresis was carried out at 3° - 5° C. Migration of most enzymes required approximately four hours

under conditions described by Adams and Joly (1980), however malate dehydrogenase gels required twenty hours at 20 milli-amperes. Gel and electrode buffer recipes, and electrophoretic conditions are summarized in Table 1.2. Stain recipes are described in Table 1.3. The Agar overlay method was used for SDH and MDH. F-EST gels were scored immediately, all other stains were developed at least one hour in darkness at 37^{o}C.

Evidence for genetic control of enzyme variants

Preliminary evidence for genetic control of enzyme variants was obtained from segregation ratios in haploid tissue of heterozygous trees for those enzyme systems for which recently published evidence was not available. White pine megagametophytes are haploid, therefore allozymes are expected to segregate at a 1:1 ratio if they are under genetic control. Since cones were not available in the stand, cones were collected from the University of New Hampshire white pine clone bank where the genotypic identity of open pollinated parent trees is known from previous isozyme analyses. The clones are not genetically related to the white pines in the study stand, but they are

representative of trees growing in New Hampshire. Cones were allowed to dry in sunlight, seeds were extracted, and stored at -15°C until needed. Two days preceding analysis, seeds were removed from storage, soaked in 10% H_2O_2 for twelve hours, and retained in distilled water for the remaining 36 hours. The two day seed treatment was carried out in plastic petri dishes, illuminated by a 60 watt tungsten lamp. Megagametophyte tissue of each seed was dissected from the seeds up to 12 hours before analysis, and was stored individually in one drop of water at -15°C. Immediately preceding electrophoretic analysis, the tissue was crushed with a glass rod, and an extract of the paste was absorbed into paper wicks. Electrophoresis and staining of the megagametophyte enzymes was identical to that described for foliar tissue enzymes. Chi-square tests were performed to test for statistically significant deviations of allozyme segregation ratios from the expected ratios of 1:1. Deviation probability of 0.01 was used as a cut-off for selecting statistically significant deviations from expected ratios.

RESULTS AND DISCUSSION

Enzyme variants observed

The electrophoretic analysis showed the presence of enzyme variants for six of the seven enzymes surveyed. Foliar isozyme band patterns will be discussed separately for each enzyme system. Tests of significance for deviation from expected 1:1 ratios of enzyme variants in haploid megagametophyte tissue of seeds from heterozygous mother trees are summarized in Table 1.4. A graphic summary of patterns for each system at corresponding Rf values is shown in Figure 1.1.

Each zone of staining activity on a particular gel is encoded by a gene, with one or more allozymes encoded by allelic forms of the gene. The fastest migrating zone of each enzyme is designated locus one, the following locus two, etc. Alleles at each locus are identified as 1 for the most frequent allele, 2 for the next to most frequent allele, etc. For example, the fast allele of the first MDH locus was also the most frequent allele at that locus and is

represented as MDH-1(1). Heterozygote genotypes are represented as "1/2", and homozygotes as "1/1", "2/2", etc.

Leucine aminopeptidase (LAP) E.C. 3.4.11.1. - A single enzyme band appeared consistently on gels with foliar samples stained for LAP. Two other isozymes, one slower and one faster, were occasionally observed, but not included in this study. The three zones of activity correspond to those seen by Ryu (1982). Eckert et al. (1981) observed similar bands in megagametophyte tissue of heterozygous white pines, but also detected additional loci. 1:1 segregation of allozymes in heterozygous white pine megagametophytes indicated the allozymes were under genetic control in megagametophyte tissue (Eckert et al. 1981).

Phosphoglucose Isomerase (PGI) E.C. 5.3.1.9 - Two zones of activity were observed on gels with foliar samples stained for PGI. One band was observed for PGI-1 and two variants were observed for PGI-2. PGI-2 bands of homozygotes appeared darker than those of heterozygotes. The heterozygotes displayed a dimeric pattern, with a darker band between the two

lighter fast and slow bands. This type of pattern was reported for other conifers by Adams and Joly (1980), and by Guries and Ledig (1978). Ryu (1982) detected a third band, but that band occurred at low frequencies in his study, and the allele for that band probably does not exist in this stand. Both allozymes of PGI-2 used in this study segregated in a 1:1 ratio in heterozygous megagametophyte tissue.

Phosphoglucomutase (PGM) E.C. 2.7.5.1 - Two zones of activity were seen on gels with foliar samples stained for PGM. The fast migrating zone contained one dark band similar to that observed by Eckert et al. (1981) in megagametophyte tissue. Three different lighter bands were seen in the slow migrating zone, both in homozygous and heterozygous monomeric form. Heterozygote bands were lighter than homozygote bands. All six possible band combinations were observed at the slow locus. Ryu (1982) observed nearly identical numbers of enzyme bands at compatible Rf values in foliar tissue and in the mega-gametophyte. Eckert et al. (1981) observed 1:1 variant segregation ratios in megagametophyte tissue of heterozygous

eastern white pines, suggesting the allozymes are under genetic control.

Glutamate-oxaloacetate Transaminase (GOT) E.C. 2.6.1.1. - Two zones of activity were observed on gels stained for GOT, the fast zone had one isozyme band and the slow zone had two variants. The fast allele of GOT-2 was present in only two trees. No heterozygotes were detected. A third, even slower zone was occasionally visible, but was not intense enough for consistent scoring. Three GOT zones with compatible Rf values and nearly identical numbers of enzyme bands have been reported for both megagametophyte and foliar tissue, with two monomeric bands at the second locus (Ryu 1982). Eckert et al. (1981) used segregation of allozymes in heterozygous white pine megagametophytes to demonstrate genetic control of the GOT variants.

Fluorescent Esterase (F-EST) E.C. 3.1.1.2 - Two loci containing two isozyme variants each were scored for foliage. F-EST was more difficult to score than other enzymes, because

stains appeared for a limited amount of time, and the stains often were faint. Repeated runs and adjacent replicates were necessary to verify scoring of approximately half of the trees. Heterozygotes exhibited a dimeric pattern as reported for eastern white pine (Ryu 1982) and for ponderosa pine foliage (Mitton et al. 1979). The allozymes for the first locus segregated in megagametophyte tissue as expected (Table 1.4). No seeds were available to test segregation of the second locus. Segregation tests of the variants at the second locus of F-EST should be conducted when seeds from a heterozygous tree become available.

Shikimate Dehydrogenase (SDH) E.C. 1.1.1.25 - One zone of three isozyme variants was observed for this enzyme in foliar samples. All six possible monomeric patterns were seen in the trees sampled for this study. Homozygote bands stained more intensely than did heterozygote bands. The first and second allozymes segregated as expected in megagametophyte tissue. No seeds were available to test segregation of the third allozyme because no heterozygotes were found with the third allozyme.

Malate Dehydrogenase (MDH) E.C. 1.1.1.37 - Two zones of activity were seen for MDH in foliar samples, a fast zone with three enzyme variants and a slow zone fixed for one variant. All three homozygote and three dimeric heterozygote patterns of bands were observed at MDH-1. Megagametophyte enzyme band pairs from heterozygous mother trees segregated in a 1:1 ratio as expected for a single locus under genetic control (Table 1.4). Homozygote stains of MDH-1 appeared darker than heterozygote stains, but comparison of the foliar band patterns to those of megagametophytes was hampered by differential migration of isozymes from the tissues. Chemicals present in the foliar extracts, not found in the megagametophyte extracts may be responsible for this problem (Pitel and Cheliak 1983).

The frequencies of the isozymes described above, and their applicability to searching for genetic population structure in a white pine forest are reported in Chapter 2. The preliminary verification of genetic control of the expression of the isozymes through the analysis of their segregation in haploid megagametophyte tissue shows that these isozymes can be used to

represent genetic variability of the white pines. Segregation of variants for the second locus of F-EST and the third variant of SDH should be verified when seeds from a heterozygous tree become available. In general, a sufficient number of isozymes were observed in this study to allow consideration of population structure, but some previously used enzyme systems were not included because the methods, or their implementation during this study, did not yield sufficiently reliable and repeatable results. Only enzyme systems for which scoring could be confirmed in repeated runs were included in this study. Future studies of population structure of white pine should include additional enzyme systems to allow for greater flexibility in selecting desirable loci. With improved electrophoretic methods, and with experience gained from other electrophoretic studies, it should be possible to increase the number of enzyme systems used in eastern white pine studies.

Chapter 2

WITHIN-STAND CLUSTERING OF EASTERN WHITE PINE GENOTYPES

INTRODUCTION

A large proportion of genetic variability in eastern white pine appears to be within stands, rather than among stands (Ryu and Eckert 1983), but the spatial distribution of the genetic variability within stands is not well understood. Traditional methods used for genotyping trees are too time consuming and costly for studying a large number of trees of a single population, in order to facilitate detection of groups of genotypically similar trees. Electrophoretic analysis of tree enzymes can help in understanding genetic variability and structure in forest stands because the variability of tree enzymes reflects variability in the genes that code for those enzymes, and many trees can be sampled relatively quickly and inexpensively when foliar samples are

analyzed with electrophoretic methods. Gaining a better understanding of local genetic variability in eastern white pine is important because it could be applied to decisions affecting stand management, including thinning and seed collection for breeding purposes. Studies of stand population structure should proceed without delay, because the most useful knowledge will be gained in natural and semi-natural (unmanaged) forest stands, which are rare in most regions of the white pine species range. These stands are decreasing in number with the establishment of new communities and intensified use of land and natural resources.

Researchers face two general difficulties when using isozyme analysis for this purpose. (1) Many forest tree species, including some conifers, appear to have weakly differentiated populations as measured by the variability of allozymes, and show small differences in allozyme frequencies between populations (Yeh et al. 1985). This makes it very difficult to identify small sub-populations that may reflect local conditions or reproductive history. (2) The relationship between allozyme frequencies and more commonly observed traits such as tree quality, height, age, and location is difficult to evaluate because many factors are

involved in determining whether a tree is successfully established and survives to maturity in a particular location within a stand. For this reason, multivariate data analysis methods are necessary for evaluating the relationship between allozyme frequencies and tree location. These methods allow researchers to consider the relationship between many allozymes, and many physical traits of the tree at the same time (Gittins 1979). Multivariate tests of differentiation of allozyme loci can in some cases yield higher levels of statistical significance than the average level from single loci (Smouse et al. 1982, Mitton 1977, Smouse and Spielman 1977). When considered jointly, a large number of small differences may be equivalent to a small number of large differences (Yeh et al. 1985). By evaluating several loci and several physical traits of trees in a population at the same time, patterns and relationships may be discovered that cannot be detected with univariate methods. Multivariate methods can help detect subtle but important interactions and may therefore provide a better understanding of a population structure and the underlying factors affecting it (Yeh et al. 1985).

This chapter describes the use of canonical variate scores to study the relationship of tree location to the allelic composition of eastern white pines growing in a two hectare unmanaged forest stand. Evidence suggesting that tree establishment may be in part responsible for clustering of trees with similar genotypes is presented.

MATERIALS AND METHODS

Data analysis

The raw isozyme scores reported in Chapter 1 were converted into continuous genetic scores (Smouse et al. 1982) so the data could be evaluated with canonical correlation analysis (Gittins 1979). These "Smouse Scores" represent the traditionally categorical isozyme (allelic) information through continuous variables, which measure the contribution of an allele to a genetic locus. By convention, for example, the first of two alleles of a diploid locus contributes a value of 0.5 to the total "Smouse Score" of that locus. An individual homozygous for that allele would then have a score of 1.0, a heterozygous individual would have a score of 0.5, and an individual homozygous for the other allele would have a score of 0.0. A locus with three alleles would be described by two "Smouse Scores", etc. Conversion of all the loci used in this study is described in detail in Appendix B.

Canonical correlation analysis was used to evaluate the strength of relationship between the set of genotypic variables represented by isozymes, and the within-stand location of the trees represented by X-Y coordinate points. To obtain the X-Y coordinate points, a map of the stand was made using aerial photographs and ground measurements with a compass and measuring tape. Using the stand's outer boundary and internal reference lines, the location of each tree was recorded on the map. The map was then digitized using a Summagraphics Digitizing board attached to a VAX/VMS computer system to represent tree locations with X and Y coordinate points. The Y coordinate points represent the south-east to north-west direction, and the X coordinate points represent the south-west to north-east direction. The relation of age and genotype to the spatial distribution of the trees was explored. Canonical correlation analysis, a form of multivariate correlation analysis, was selected because it allows evaluation of the relationship between two sets of variables. Let $X_1 \ldots X_i$ represent the set of i isozyme variables, and $Y_1 \ldots Y_j$ represent the set of j tree location and phenotypic variables. Canonical correlation analysis generates pairs of linear canonical variates of the form:

$$V_k = a_{k1}X_1 + a_{k2}X_2 + \ldots + a_{ki}X_i$$

$$W_k = b_{k1}Y_1 + b_{k2}Y_2 + \ldots + a_{ki}Y_j$$

where $a_{k1} \ldots a_{ki}$ are the i canonical coefficients for associated isozyme variables, $b_{k1} \ldots b_{kj}$ are the j canonical coefficients for associated location and phenotypic variables, and k is the number of canonical correlation variate pairs. These coefficients reflect the importance of the raw variables in the canonical variate. The correlation between the canonical variates is maximized in the analysis (Marriott 1974). Any relationship discovered between the two sets of data of this process can be traced back to the individual original variables by viewing the correlations between the raw variables and the canonical variate scores. In the analysis it was assumed that the sets of isozyme and location and phenotypic variables were jointly normally distributed and Wilks' lambda was used to test the null hypothesis that there was not any residual linear association between the two variable sets, for which another pair of canonical variate scores could account.

Plots of predicted canonical variate scores for the enzymes were used to identify regions of the stand which contain trees

with similar genotypes. Predicted scores, rather than raw scores, were used because predicted scores show any existing patterns more clearly. Predicted scores tend to reduce the effect of outliers. Maps of trees showing both the location and predicted score of trees allow researchers to visually identify general patterns in the geographic distribution of trees with similar scores. Plots showing the distribution of trees with selected enzyme variants important in the formation of the canonical variate scores were used to explain the basis for some of the observed patterns. The relationship of tree age to allelic composition of individual trees, and the distribution of trees with similar age were compared to explore the patterns that were observed. Location of trees with similar allelic composition, as measured by the canonical variate scores, was compared to the location of different soil conditions determined through soil analysis. The analysis was done at the Agriculture Experiment Station and Cooperative Extension Service, University of New Hampshire, Durham, N.H. 03824. Soil samples used in this analysis were taken at 13 locations in the stand, as shown in Appendix C3. The soil samples were evaluated for salt, nitrogen, P, K, Ca, and Mg content. Soil pH was measured, and texture was determined.

RESULTS AND DISCUSSION

Relationship of tree location to genotype

Canonical correlation analysis of enzyme scores and tree location variables showed two significant ($p<0.05$) canonical correlations (Table 2.1) suggesting that a non-random relationship exists between a tree's genotype, represented by isozyme scores, and where it is located in the stand, measured by X and Y coordinate points. The canonical correlation coefficient (r) is a measure of the relationship between each pair of canonical variates (Gittins 1979). The presence of a large canonical correlation coefficient would suggest a strong relationship between the location and genotype of the individual trees. Only the first two canonical correlation coefficients in table 2.1 are interesting, because they are statistically significant. The first correlation accounted for 11.6 percent of variation in the data, the second correlation accounted for 6.8 percent of variation in the data.

Although these correlation coefficients are low compared to correlations commonly observed in biological studies, they are important because they help to demonstrate a relationship between genotypes and tree location that is otherwise difficult to detect. The two correlations between the first two pairs of canonical variates were statistically significant ($p < 0.01$ and 0.05 respectively) and sufficiently strong for identifying spacial distribution patterns. Based on this analysis, it appears that trees are not randomly scattered in the stand, and that some trees with similar genotypes exist together in identifiable areas in the stand.

Two independent relationships were observed between the isozyme scores and the locations of the trees. The first relationship is represented by the first pair of canonical variates V_1 and W_1 (Table 2.2) and the second is represented by V_2 and W_2. Using the canonical variates to study the spacial distribution of the trees allowed for considering several loci at the same time, and relating the presence or absence of their alleles to the location of the trees in the stand. Being able to consider several loci at once is important, because the contribution to the canonical model of any one locus is influenced

by the effect of other loci and many ecological factors. For example, our ability to detect the contribution of a locus coding for an efficient protein, which in turn may allow for rapid growth rate, depends on the overall health of the tree, available nutrients, climate, location, and age of the tree.

Let V represent a linear combination of the enzyme variables and let W represent a linear combination of the X and Y coordinate point (location) variables. In canonical correlation analysis the linear combination of each set of variables is calculated so that the correlation between V_1 and W_1 is maximized (Gittins 1979). The second set of canonical variables (V_2 and W_2) is then extracted, using a portion of the remaining variance not accounted for by the first set of canonical variables, etc. The relative importance of individual raw variables that contributed to the formation of the canonical variate scores may be determined from correlations between the canonical variates and raw variables (Table 2.2). Larger correlation coefficients indicate a stronger relationship between either the location or enzyme variables and their respective canonical variates. A larger correlation coefficient therefore indicates that the canonical

variate more strongly represents the overall variability of the raw variables.

Scores for the F-EST-2 and SDH loci were most highly correlated with V_1, the first canonical variate score, which is a linear combination of the raw genetic scores. Scores for the PGI-2 locus were most highly correlated with V_2, the second canonical variate score representing genetic variability. The Y values and squared values of Y, the coordinate points for tree location in the SE to NW direction in the stand were most highly correlated with both W_1 and W_2, the first and second canonical variate scores for tree location. (For presentation purposes the Y axis is nearly horizontal.) This indicates that the canonical variates represent the distribution of trees by F-EST and SDH (first canonical pair) and PGI (second canonical pair) isozyme scores in predominantly the Y or the SE to NW direction (see Figure 2.1). The XY term is also important, and represents the bi-directional north to south pattern seen in Figure 2.1b. This diagram shows that predicted canonical variate scores change in two directions (X and Y). A change in predicted scores represents a change in allelic composition or genotype of trees.

The change in tree genotype in relation to tree location is directional in the stand. The predicted canonical variate scores help demonstrate this pattern (Figure 2.1), because they smooth the effect of minor deviations in the original data. The change in genotype is related to the discrete variation in single enzymes in individual trees. Figure 2.2 shows the distribution of trees homozygous for the slow allele of PGI, one of the enzymes important in forming the second canonical variate, and helps interpret the predicted canonical variate score (PV_2) patterns. The trees with this genotype are found in the north-western quadrant of the stand. Trees in this section of the stand also have lower predicted PV_2 scores (Figure 2.1b).

Additional canonical correlation analysis (Table 2.3), using a reduced set of 158 trees for which age is known, and including location variables, produced predicted genotype distribution patterns similar to those discussed above. The contribution of raw variables to the formation of the canonical variates also remained similar. The first correlation coefficient was 0.49 ($p < 0.01$). The second canonical correlation was not statistically significant. Age was correlated similarly (Table 2.4) with its

canonical variate (W_1) as were both the X and Y coordinates (r=-0.39 for X and r=0.45 for Y). Canonical correlation analysis for the same 158 trees excluding the variable age yielded almost identical results as did the analysis with age. This suggests that although age is related to enzyme composition of eastern white pines in this stand, the spatial distribution of trees explains most of the variation in the enzyme composition.

Canonical correlation analysis without F-EST-2, the locus for which no seeds were available to test segregation of variants, also yielded similar results (Appendix D). The canonical correlation coefficient was .27 ($p<.01$). The second canonical correlation coefficient was not significant ($p>.05$). Scores for SDH and PGI were still most highly correlated with the first canonical variable (V1), and Y and squared value of Y were also most highly correlated with the first canonical variable. The patterns of genotype distribution throughout the stand were also almost identical. F-EST-2 therefore contributed to the patterns seen in this study, but was not the only locus responsible for the patterns. SDH, the locus which consistently contributed to all the patterns observed in the stand, was the most clearly and consistently staining enzyme.

Plots of age categories (Figure 2.3) show that trees of similar age (up to 40 years old) occur in distinct clusters along the outer perimeter of the study area. One reason for why clusters of younger white pines were found in this stand is explained by Hibbs (1982). The eastern white pines share this forest stand with hemlocks and deciduous hardwoods such as white oak, black birch and sugar maple. The canopy is closed throughout most of the stand and younger white pines are found only where the upper canopy has gaps. As Hibbs explained, white pine can regenerate in established pine-hardwood stands under one of several conditions: (1) when hardwood competition is high, pine may regenerate in small groups and the group of pines buffers the most successful of the group to grow to maturity, (2) understory pine from a previous stand may survive and grow after canopy removal, and (3) pine may regenerate in large canopy openings.

The clusters of younger pines found in this stand could have formed following the death or fall of mature trees. This explanation is supported by the presence of fallen trees and canopy openings where younger pines are growing. This stand therefore

appears to be normal in its development. If only one, or a small number of trees, contributed seeds for the establishment of seedlings in these openings, clustering of trees by similar allelic component would be expected. No cones were seen on the white pines in the study area during the past four years and fewer than ten seedlings were found. It is not known how many mature trees contributed to the establishment of the white pines which play a role in the patterns seen in this study, but cone production, limited to a small number of mature trees, could have resulted in the establishment of adjacent trees with similar genotypes.

Several possible explanations exist for why most of the young trees were found along the outer portions of the stand. The stand is bordered on the southern and western side by a path, which provides sunlight. Some younger trees were seen on either side of this path. The northern and western side of the stand is bordered by a white pine stand that was recently thinned, and provides openings in the canopy along the border. The northern and north eastern side of the stand is bordered by a wet area, where fewer trees grow, and several canopy openings exist. These three

situations provide more sunlight than is available inside of the stand, where mature white pines, tall hemlocks and various deciduous trees form a solid canopy. White pine seedlings therefore have a better chance of establishment along the outside of the stand.

Significance of the observed patterns

The presence of clusters of trees with similar ages and genotypes suggests that either reproduction, establishment or survival of trees in the stand was not random. It also means that if all trees are removed from given areas in the stand without knowing the genotypes of the trees, the frequency of some alleles could be affected to a greater degree than if genotypes were scattered more evenly throughout the stand. Though it is not practical for foresters to consider genotypes of individual trees in a stand during timber stand improvement, or commercial harvest operations, results of this study suggest that foresters should retain representatives of trees found in clusters of similarly aged trees to enhance the chance of retaining trees with genotypes that may be unique to the clusters. In the absence of detailed

information about population structure of a stand, a uniformly spaced shelterwood system that maximizes the number of stems per acre, without causing excessive tree competition for vital resources, should in most cases provide the best chances of retaining the genetic diversity of a stand.

Interpretation of the relationship between the enzyme canonical variate scores and the location canonical variate scores is difficult, because the biological significance of X and Y is not clear for a geographic area of 2 hectares. The coordinates would have more meaning over an extended latitudinal and longitudinal range. In this study, the location information describes the position of a tree in relation to other trees, and possibly differences in ecological conditions such as competition with neighboring trees, nutrition, exposure to the sun, and soil type.

Analysis of soil samples taken throughout the stand showed differences in soil type and quality, but no obvious patterns could be associated with the location of trees with similar genotypes (Appendix C). The following items were evaluated during the soil

analysis: pH, texture, Mg, Ca, K, P, salt content, and total nitrogen. Overall geographic changes in soil texture, pH, and concentration of the measured elements did not correspond with the changes in the values of the canonical variate scores (Appendix C2) which represent the allelic composition of the trees in the stand. The fact that differences in soil type and quality do exist in the stand suggests that a more intense sampling of the soil in the stand is justified, and may lead to the identification of soil type and quality patterns that could be associated with the distribution of trees with similar allelic composition. Ideally, soil should be tested within several feet of each tree, but this type of intense sampling would have been too costly. The value of the canonical correlation analysis therefore is its ability to summarize groups of biological factors of unknown importance, to point out patterns in the data, and to show which raw variables are important in forming those patterns. Those raw variables must then be used to help explain the patterns and to point out practical implications. The clustering of trees by age (Figure 2.3) and by genotype (Figure 2.2) support results from the canonical correlation analysis. Both canonical variate plots and raw variable plots suggest the distribution of trees with similar isozyme makeup is

not random and it is related to age of the trees. The age and genotype clusters are probably closely related to tree establishment, and the clusters may disappear with time as their members die due to age, disease and other natural processes. The more random distribution of older trees supports this idea.

The detection of clusters of trees homozygous for the slow allele of PGI corresponds to results obtained in a study of Scots pine (Tigerstedt et al. 1982). Analysis of mature Scots pine and progeny generation in the understory showed that the progeny may be on the average slightly related, with high proportions of homozygotes. The adult trees conformed to Hardy-Weinberg equilibrium. Tigerstedt et al. (1982) suggested that selection against high proportions of inbreds in young tree populations is important to avoid inbreeding depression in natural populations. Such selection would lead to the situation observed in both their study and in the current study, where older trees with similar genotypes are less frequent and more randomly distributed.

Determining what caused the clustering of trees with similar genotypes would require additional studies, including

experimentation. Information gathered in this and other studies presents some interesting possible explanations, however. The relationship found in this study between age and tree location, and between tree location and genotype suggests that establishment of trees affects the distribution of genotypes. Trees with similar genotypes were not found to have identical ages, but rather similar ages. These trees may have received some of their genes from the same parent, but the contribution of that parent tree spanned several years of reproduction. Hibbs' (1982) explanation of establishment of white pine suggests why we may be finding trees of similar genotype and similar, but not identical age. An opening that provides sunlight for several years after the fall of an older tree would provide the opportunity for establishment of many seedlings from the same parents during those years. As the seedlings grew in size, however, adjacent seedlings would compete for space, nutrients and sunlight, and only a few trees in that former opening would eventually reach the ages observed during this study. The remaining trees would become suppressed and die. Age alone does not determine which trees become dominant and survive. The exact location of the seedling establishment in relation to local moisture, nutrients, other seedlings, and exposure

to sunlight can cause some seedlings of the same age group to survive, and others to die. During the beginning of this study, fifteen young (6-10 years) trees were seen growing adjacent to each other within a ten foot radius in the study area. Many of these trees were tall, had thin crowns, and were bent from the weight of snow and ice. After three years only two trees remained alive in this area. These trees had the same genotype as measured by the methods used in this study. Their age was similar, but the exact age was not determined to avoid damage to the trees. If both of these remaining trees reach maturity, they will continue to contribute to the genotype and age spacial distribution patterns observed in this stand, but the original group of genetically similar trees will not exist.

Another possible explanation includes the possible contribution from trees of neighboring stands. Clustering of trees homozygous for the slow allele of Phosphoglucose Isomerase in the north-west portion of the stand indicates that the parent trees may be located in the northern or western portion of the stand, or that the parents are located north and west of the stand but not in the studied area. This explanation would be consistent with the

contribution of older trees in one stand to the establishment of younger trees in a neighboring stand as described by Roberds and Conkle (1984).

Some of the isozymes related to the genotypic clustering observed in this study occurred in low frequencies. The significance of the patterns therefore depends in part on whether rare alleles are considered biologically important. Rare alleles may not be important for the current population of trees, but they may help white pine survive through environmental changes in the future, especially global climate change. On the other hand, these alleles may be detrimental to long term survival of the young trees, and most of the trees homozygous for these alleles may be lost.

Additional studies are necessary to search for patterns of genotype distribution in other white pine forest stands. Until such studies are completed, this study will serve to remind foresters that local eastern white pine stands can be genetically structured. This information is important to foresters whose objective is to conserve both the current genetic resources and

the distribution of those resources, and who wish to understand better how selective thinning of trees may affect the genotypic composition of future generations.

Chapter 3

THE EFFECT OF SAMPLE SIZE ON ESTIMATES OF GENETIC VARIABILITY IN AN ELECTROPHORETIC STUDY OF EASTERN WHITE PINE

INTRODUCTION

When investigations of genetic variation in local stands and impacts of silvicultural practices are planned, decisions must be made about what are sufficiently large sample sizes to represent tree populations in natural and unmanaged forest stands. Appropriate sample size can be estimated from reviews of theory (Brown and Moran 1981, Ewens 1972, Crow and Kimura 1970) but few studies actually test theoretical considerations. The number of trees necessary to represent natural populations in isozyme studies is often determined by convention, and is affected by seed or foliar tissue availability. Sample sizes are chosen to represent local tree populations, but whether they actually do so is difficult

to determine from existing literature.

Nei (1978) evaluated magnitudes of systematic biases of sample heterozygosity. He demonstrated that small sample sizes may be used if a large number of loci are analyzed and average heterozygosity is low. Analysis of 20 genetic loci is generally considered sufficient in electrophoretic population studies (Ayala 1982). When a small number of loci are used, a large number of individuals per locus should be observed to reduce standard error of average heterozygosity. While comparing performance of biased and unbiased estimators of average heterozygosity, Nei (1978) showed there was little difference between the approaches if at least 50 individuals were used per sample. The number of trees used to sample natural populations varies widely around 50 (Ryu 1982, Grant and Mitton 1977, Copes 1975, Rudin et al. 1974). Reported sample sizes have more meaning when the sizes of stands represented by the samples also are reported as by Roberds and Conkle (1984), Feret (1974), and Tigerstedt (1973). Knowing the stand size, both the geographic area of the stand and the approximate number of trees in the stand, helps in understanding the intensity with which the stand was sampled. A sample of 50

trees randomly selected from a total of 500 trees in two hectares probably provides a different level of sampling intensity than does a 50-tree sample of a 1500-tree stand covering 9 hectares. A 50-tree sample may not be sufficient in the latter case where the chance of encountering "local" populations, or related trees that are different from other groups of trees in the stand is greater. In large stands, effective seed dispersal and pollen flow limits may lead to a genetic population structure that can be detected through more intensive sampling, which calls for sampling more trees per stand.

Little is known about the validity of tests for goodness-of-fit to the Hardy-Weinberg equilibrium genotype distribution in large populations, when small samples are used (Elsten and Forthofer 1977). Relatively large samples are required to detect small deviations commonly expected in natural populations, but large samples are often difficult to obtain. Sample trees must be selected, identified, samples must be collected from the tree, and the samples must be transported to the laboratory for analysis. Trees are often difficult to reach, either because they are far from traveled roads, or because they are tall and foliar samples are

difficult to obtain. The time required to reach each tree and to analyze the sample is costly. Doubling the sample size can in many cases double the time commitment and cost, because processing each additional sample is labor intensive. For these reasons, researchers often choose to study larger areas by sampling fewer trees, rather than small areas with many trees.

The following chapter reports the effect of subsampling the allozyme database of 511 eastern white pines from a 2 hectare forest stand, discussed in Chapter 1. This subsampling shows that if random samples of tree populations are too small, inaccurate estimates of genetic variability can be obtained. Emphasis was placed on accurate estimation of allele frequencies, mean number of alleles per locus, heterozygosity, and polymorphism. The effect of sample size on tests of genotype frequency deviation from the Hardy-Weinberg equilibrium was evaluated. The results of this study answer some questions regarding sampling strategies for eastern white pine, but additional studies are necessary to evaluate the general applicability of these results in other eastern white pine stands, or stands of other tree species.

MATERIALS AND METHODS

Data Analysis

The computer program BIOSYS-1 (Swofford and Selander 1981) was used to calculate the mean number of alleles per locus, average heterozygosity, and the mean proportion of polymorphic loci from isozyme scores of the 7 genetic loci for the 511 eastern white pines in the two hectare forest stand. Isozyme frequencies were tested for deviation from the Hardy-Weinberg equilibrium. Average heterozygosity was calculated in three different ways: (1) "direct-count", or the proportion of individuals sampled that were heterozygous; (2) an estimate based on Hardy-Weinberg expectations; and (3) the unbiased estimate, based on conditional expectations (Swofford and Selander 1981, Levene 1949, Nei 1978). Estimates of polymorphism were evaluated under three different criteria for polymorphism: (1) "no criterion", or the presence of more than one allele per genetic locus; (2) a maximum frequency limit for the most common allele of 0.99; and (3) a maximum frequency limit for the most common allele of 0.95. The

latter two criteria were set to focus on loci having alleles with frequencies too high to be explained merely by recurrent mutation (Hartl 1980), and to reduce bias due to the presence of very rare alleles. The hypothesis that the tree population is in Hardy-Weinberg equilibrium at each variable locus was tested using three approaches: (1) The usual chi-square goodness-of-fit test was performed using observed genotype frequencies and those expected under Hardy-Weinberg equilibrium. This approach is suspect when frequencies of some classes are low and it was repeated for small sample sizes using Levene's (1949) correction for small sample size (Swofford and Selander 1981). (2) When more than two alleles were observed at a locus, genotypes were pooled into three classes: (a) homozygotes for the most common allele, (b) heterozygotes for the most common allele and one of the other alleles, and (c) all other genotypes. Chi-square values were recalculated and used with one degree of freedom. (3) The exact significance probabilities analogous to Fisher's exact test for 2x2 contingency tables were calculated (Swofford and Selander 1981). The pooling procedure described above was used for the exact test, when more than two alleles were present per locus.

The statistical analyses described above were repeated for five different randomly chosen groups of trees in each of the following sample sizes; 5, 25, 50, 75, ..., 475 trees. This approach allowed comparison of estimate fluctuation within various sample sizes. The exact chi-square test was repeated for samples of up to 125 trees. For large samples it becomes computationally burdensome and the usual chi-square test is preferable (Swofford and Selander 1981). Results from analyses including all 511 white pines in the stand were used for reference as true stand values. Random samples from the data set were selected using a BASIC random numbers generating program, the program MINITAB, and the program SimThin (Appendix E). The usefulness of particular sample sizes was evaluated based on: (1) the variability of results of five independent replicates within each sample category, and (2) comparison of the sample results with the total stand results.

RESULTS AND DISCUSSION

Allele Frequencies

A variety of allele frequencies, represented by the frequencies of enzyme variants, were found for twelve genetic loci in the white pine stand (Table 3.1). Five loci, GOT-1, PGI-1, PGM-1, MDH-2 and LAP were monomorphic. The remaining loci had at least two alleles. Three loci, GOT-2, SDH and F-EST-2 contained a high-frequency allele (frequency > 0.92). SDH, PGM-2 and MDH-1 had three alleles.

Low Frequency Alleles

Estimates of low frequency alleles were affected by sample size (Tables 3.2, 3.3, 3.4 and 3.5). Higher coefficients of variation for the mean frequency estimates of low frequency alleles, obtained using small samples, show that repeated small samples yielded less consistent results than were obtained with repeated larger samples (Table 3.4). PGM-2(2) frequency estimates (Table 3.5) illustrate the effect of sample size on the reliability of

estimate. Table 3.5 shows an inverse relationship between sample size (number of trees) and standard deviation of mean frequency. The coefficients of variation confirm these patterns. The true PGM-2(2) frequency in the stand is 0.178, but the estimates based on five 25 tree samples of the population ranged from 0.120 to 0.400. The mean value of the five random samples (0.212) is near the true frequency, however a single 25 tree sample of the population could have resulted in a misleading frequency estimate. Five tree samples yielded even less reliable results, but the variability between estimates decreased for larger sample sizes, as expected. Fifty tree samples provided consistently better estimates of the correct value.

The standard deviations, which reflect the variability of estimates within each sample size class (Bailey 1981), are higher for sample sizes of 5 to 50 trees (0.195 - 0.041), and they fluctuate between 0.031 and 0.003 for the 75 to 475 tree samples. Although the larger samples tended to yield more similar results than did smaller samples because sampling was done with replacement and the larger samples were more likely to contain some common trees, the fact that individual large samples tended

to produce estimates that are closer to the true population measures indicates that larger samples produced more reliable estimates. The standard deviations discussed here are of value in determining what would have been an appropriate number of trees to sample to obtain an accurate estimate of allele frequencies for this stand of white pine.

To apply these results to developing general sampling schemes for research in other stands and other species, the coefficient of variation would be more appropriate. The coefficient of variation can be used to compare results from different experiments involving the same character, possibly conducted by different persons (Steel and Torrie 1980). It is the standard deviation expressed as a percentage of the sample mean. It is independent of the unit of measurement used and could therefore be used to compare the effect of different levels of sampling for different tree species, even if different measures of variability are used. Note that the coefficient of variation for PGM2(2) frequency estimates decreases with the standard deviation as the sample size increases (Table 3.5). The standard deviation was reduced to below 10% of the sample mean for

samples containing 125 trees but exceeded 10% of the sample mean for 225 trees per sample. Because it did not exceed 10% for 250 or more trees per sample, to accurately estimate frequency of PGM2(2) in this study would require 250 randomly selected trees. At that level of sampling, variation is consistently low. Two hundred and fifty randomly selected trees represent almost half of the stand's tree population. Such sampling would be expensive and time consuming during routine studies, and some precision may have to be given up for alleles with frequencies similar to that of PGM-2(2) (0.178) by using smaller samples of 50 or more trees.

Patterns of standard deviations for other alleles are similar to those of PGM-2(2). Standard deviation of the frequencies for most of the alleles decrease sharply in the 5 to 50 tree samples (Table 3.3). The largest reductions in the coefficients of variation also occur for sample sizes between 5 and 50 trees (Table 3.4). These coefficients are useful in that they can be used to compare variation in sampling for alleles of different frequencies. These results are based on sampling with replacement to most closely simulate randomly selecting any trees in the stand. Sampling without replacement was not done because such sampling would

artificially exclude trees from individual samples. Sampling with replacement better simulates routine studies, because during such studies all trees in the stand are available to be randomly selected into the sample. The correlation coefficients (Table 3.4) show a strong and consistent relationship between the coefficients of variation and the number of trees sampled, ranging in values between -0.774 and -0.525 and with a mean of -0.642. The coefficients of variation decrease with increasing sample size, as expected according to statistical theory (Steel and Torrie 1980).

Although the average allele frequencies of the five independent replicates in most sample categories tend to closely approximate the true allele frequencies (Table 3.2), the standard deviations are high for samples under 50 trees (Table 3.3). A notable decrease in standard deviations occurs both between 5- and 50-tree samples and at 100-tree samples. This suggests that samples of fewer than 50 trees would unreliably represent population allele frequencies, and that up to 100-tree samples may actually be necessary to significantly and more consistently reduce random fluctuations in allele frequency estimates for low-frequency alleles. Variability of the mean frequency estimates

based on small sample sizes was observed for both low frequency alleles (SDH(2), frequency=0.054) and more common alleles (PGM-2(2), frequency=0.178). For most alleles random samples of fifty trees were superior to 5 or 25 tree random samples, because chance variation of frequency estimates was reduced. Although increasing sample size to above 50 trees per stand would improve reliability even further, the cost of doing so may be too high in return for the progressively smaller gains in precision as seen in this study. If the objective is to estimate frequencies for rare alleles (frequency < 0.01), larger samples would be necessary despite the increased cost.

Mean Number Of Alleles Per Locus

Estimates of the mean number of alleles per locus, based on allozyme frequencies, increased with sample size. This trend is clearly seen in Table 3.6, and is related to detection of low frequency alleles in larger samples. The chances of detecting a low frequency allele when only a few trees are sampled are small. As the number of trees per sample is increased, the chance of detecting one or more low frequency allele increases. As long as

there are low frequency alleles to be detected in the stand, the estimate of mean number of alleles per locus increases with the intensity of sampling. A minimum of 50 to 75 trees were needed to noticeably reduce the standard deviation and coefficients of variation, 200 trees were needed to reach 0, but even above 200 trees per sample standard deviation fluctuated (Tables 3.7 and 3.8).

Estimates of the mean number of alleles per locus using 5-tree samples ranged from 1.25 to 1.50, while estimates from 25-tree samples ranged from 1.67 to 1.83. Larger samples approximated the true value better, but more than half of the trees were necessary to obtain a more consistently correct estimate. The required sample size for correct estimation of the mean number of alleles per locus in other studies depends on the genetic variability of populations analyzed, especially allele frequencies. Researchers comparing tree populations may need to use samples of at least 50 to 75 trees, when using isozymes as the basis for comparison, when the genetic variability of the populations is similar to that of the stand in this study, and when using approximately the same number of variable loci as were used in this study. Based on the standard deviation in Table 3.7, a 50-tree

sample appears to offer no advantage over 25-tree samples, but 25-tree samples are only one step away from 5-tree samples which yielded undesirable results. By using at least 50 trees, the differences in allele frequency estimates will be less due to random effects of the sampling strategy, and more due to real differences between the populations.

Heterozygosity

The average heterozygosity of the total population is 0.242 (direct-count) and 0.182 (biased and unbiased estimates). The biased and unbiased values of most random samples were similar (Tables 3.6, 3.7 and 3.8). These results support Nei's (1978) conclusions, that the two approaches give similar results if at least 50 individuals are used per sample. The standard deviation of the means decreased from 0.032 to 0.001 between the 5- and 125-tree samples, respectively, and remained between 0.000 and 0.006 for 125 to 475-tree samples. See table 3.8 for corresponding patterns of coefficients of variation. Samples above 125 trees therefore yielded little or no improvement in the reliability of heterozygosity estimates. The direct count approach gave higher heterozygosity estimates, and little precision improvement was

gained using samples of more than 50 trees. Samples of 125 trees were necessary to keep the average standard deviation of mean heterozygosity under 0.008. The average standard deviation ranged between 0.019 and 0.001 for 5-tree to 475-tree samples respectively. Most sample sizes above the 5-tree category gave estimates near (+/- 0.006) the true heterozygosity value, showing the consistency of heterozygosity over different sample sizes as a measure of genetic diversity.

Heterozygosity is generally considered a more reliable measure of genetic variability than is polymorphism, because it tends to be less affected by sample size and by the presence of rare alleles. On the average, a 25- to 50-tree random sample would have yielded a heterozygosity estimate within approximately 2.5% of the actual value.

Polymorphism

The true population polymorphism is 58.33 percent, but using the 0.99 and 0.95 percent criteria, it is only 50.00 percent. To obtain consistent estimates equivalent to the true polymorphism measure, 250 randomly chosen trees were necessary (0.95

criterion). At least 225 trees were needed with the 0.99 criterion, and 425 trees were necessary when no criterion was applied.

Polymorphism is a less desirable measure of genetic variability than is heterozygosity because the measure is not a precise representation of genetic variation, since a slightly polymorphic locus counts as much as a very polymorphic one (Ayala 1982). The decision regarding what criterion of polymorphism to use is usually arbitrary. Polymorphism estimates based on small samples can be misleading (Tables 3.6, 3.7 and 3.8), but some of the problems would have been avoided in this study by using a 0.99 criterion (most common allele having frequency of less than 0.99) and at least 100 randomly chosen trees. This approach should help reduce the occurrence of polymorphism estimates artificially inflated by the presence of rare alleles from random mutations.

Hardy-Weinberg Equilibrium

A population of trees is considered to be in Hardy-Weinberg equilibrium if both gene frequencies and genotype frequencies are constant from generation to generation in the absence of migration, mutation, random genetic drift, and selection (Falconer

1960). Determining whether a population of trees is in Hardy-Weinberg equilibrium is of interest, because finding that a population is not in Hardy-Weinberg equilibrium can be the first step in identifying disruptions in random mating among trees. Identifying such disruptions in the course of forest management is an important part of helping to maintain healthy tree populations, by preserving genetic variability. The extent to which genetic variability helps tree populations adapt to climatic changes, pollution, and diseases is yet to be determined, but experience with some other tree species suggests that genetic variability may be better than homogeneity (Bergmann and Scholz 1987, Geburek et al. 1987). The long life cycles of trees preclude direct evaluation of random mating in forest stands.

A comparison of expected and observed genotype frequencies, using the chi-square test, and the probability (P) values associated with the chi-square values provide a categorical means to detect possible deviations from random mating. These values were calculated with the program BIOSYS-1 (Swofford and Selander 1981). A value of P less than 0.05 is commonly considered "statistically significant", and indicates evidence for an

unacceptable fit of expected to observed frequencies (Hartl 1980).

The "usual test", a chi-square goodness-of-fit test using observed genotype frequencies and those expected under Hardy-Weinberg equilibrium (Swofford and Selander 1981) was not appropriate for GOT-2, PGI-2, and F-EST, because more than two alleles were detected at these loci. The test also is not appropriate where expected frequencies of some classes are low, such as when sample size is less than 50, and the expected frequency is less than 5 (Sokal and Rohlf 1969).

When all trees were used, tests of all loci (except the chi-square test with pooling for PGM-2) were significant ($p<0.01$), suggesting that allele frequencies in the stand are not as would be expected under Hardy-Weinberg Equilibrium. Smaller random samples of data were progressively less dependable as one or more of the five chi-square tests for each sample size did not detect significant deviations (Tables 3.9 - 3.11). The smaller samples were less likely to detect the presence of low frequency alleles, suggesting the alleles were not present in the stand. GOT-2 was an exception, in that only two trees contained the rare allele, and

all chi-square tests for GOT-2 were significant. The following sample sizes were necessary to give statistically significant results for all five chi-square tests performed for each of the five random samples in each category: 475 trees for SDH, 25 trees for PGI-2, 25 trees for F-EST-1 and 100 trees for F-EST-2. The frequencies of the low frequency alleles for these loci were as follows: 0.054 and 0.016 for SDH, 0.477 for PGI, 0.498 for F-EST-1 and 0.066 for F-EST-2. For MDH-1, even some 475 trees samples failed to show significant deviation from expected frequencies. Low frequency alleles for MDH-1 had frequencies of 0.285 and 0.122. The decision to repeat the tests five times was arbitrary to demonstrate the patterns observed in this study, and the reader may choose to set a different criterion for reliability.

The exact chi-square tests of PGM-2 isozyme variant frequencies indicated mating was significantly different from random mating for samples with at least 450 trees (not shown), but the test with pooling failed to detect this significance. This test has limited use because rare alleles were pooled, thus were undetected. Regardless, these results are surprising for PGM, because PGM-2 variant frequencies seem disproportionate. The

most frequent PGM-2 allele had a frequency of 0.773, while the least common allele had a frequency of 0.049. It is possible that trees with the most common allele were more successfully reproduced, or more likely to survive. Depending on the purpose of any given study, the selection of the type of test procedure will be critical in testing for random mating which include, or exclude low frequency alleles. Selection of the probability level for indicating significance also is critical. Using 0.05 as the cut-off level in some cases leads to rejections of the null hypothesis that the observed and expected frequencies are equal, where using 0.01 as the cut-off level would not lead to that conclusion.

Using all 511 trees, most isozymes were found not to follow Hardy-Weinberg equilibrium frequencies in this eastern white pine stand. The results of the chi-square test using all the trees are not surprising, because a predominating variant was found in most enzyme systems. If a mating bias or a selective force is responsible for the presence of very common and relatively rare variants, the selection may not be for the enzyme variants analyzed. It may be for genes linked to those coding for the enzymes studied.

Sample Size Requirements

Sample size requirements observed in this study should be applied with caution to other forest stands of eastern white pine, or other tree species with comparable population variables, such as: genetic variability (see Table 3.1), population size, tree density, tree age, mating system, microsite variability, and former silvicultural treatments. The genetic measures used in this study indicate that this stand is genetically less variable than the average eastern white pine stand, but it falls within the range of values observed in the range wide study (Ryu 1982). Direct comparisons of the values are not appropriate, because Ryu (1982) used additional genetic loci not included in this study. Note the small differences in the measures of heterozygosity, polymorphism, and the number of alleles per locus when the second locus of F-EST, for which seed allozyme segregation data is not available, is excluded from the data (Appendix E). To compare results from different studies of different stands or species, researchers should determine an acceptable level of variation in the sample means, and then use the coefficient of variation to compare results. Based on this study, enough trees should be

sampled to reduce the coefficient of variation to at least 10%. Further reduction would be too costly, because additional reduction of the coefficient would require large increases in sample size.

A 50-tree (approximately 10 percent) sample of the eastern white pine stand would have been an appropriate choice for most purposes in this study for analyzing eleven genetic loci, although chi-square tests were misleading for some loci when even much larger samples were used. Note the apparent step reduction in standard deviations between 5 and 50 trees and between 50 and 100 trees for several of the measures discussed. This suggests that for stands with unknown genetic variability, 100-tree samples may generate more reliable estimates. The 50 to 100 tree sampling level suggested in this case corresponds to results found by El-Kassaby et al. (1983). In their study of Douglas-fir, using random samples, allelic frequency estimates for 27 allozyme loci, and confidence intervals, the authors found that a sample size between 40 and 60 trees is adequate to obtain reliable estimates of population's allelic frequencies. However, if large reduction in the coefficient of variation is desired, up to 250 trees, or approximately 50% of the population, would be needed.

Further work is necessary to determine whether the basic requirement for sampling a 2 hectare stand of approximately 500 eastern white pines is 10 percent of the population, or whether 50 randomly selected trees are sufficient, regardless of the population size. Common sense would suggest that 50-tree samples would not be sufficient for large populations in large geographic regions where seed and pollen dispersal is limited by distance, or where survival of certain genotypes may be affected by local growing conditions and climate. When selecting sample size, researchers will have to clearly define the effective population size, both in the number of trees, and geographic region.

With the wide-spread use of electrophoretic techniques in forest tree studies, increasingly larger numbers of enzymes are analyzed per tree sample. Results of this study will not apply directly to those where many enzymes are used. They may, however, help in pioneering work dealing with new species for which electrophoretic techniques are not fully worked out, in small-scale research projects, or in studies of only a few physiologically important enzymes.

Chapter 4

SUMMARY AND IMPLICATIONS

Electrophoretic Variability of Seven Enzymes

Twelve genetic loci were identified for the seven enzymes surveyed. LAP, PGI-1, PGM-1, GOT-1 and MDH-2 loci had only one variant. PGI-2 and F-EST loci had two variants each, displaying dimeric behavior. GOT-2 had two variants, but no dimers were observed. PGM-2 and SDH had three variants, and MDH-1 had three variants with dimeric patterns. Genetic control has been inferred for LAP, PGI, PGM, GOT, F-EST-1, MDH and the first and second allele of SDH in this, or other studies of white pine. Enzyme variants were found for six of the seven enzymes surveyed, and a total of seven genetic loci coding for these proteins are variable in this stand, with one or more alleles at each variable locus. Although other loci were detected for these enzyme systems, and

staining for several other enzyme systems was successful, those loci were not included in this study because the staining was not sufficiently repeatable. Only those enzymes, for which scoring was consistent in three independent repeated electrophoretic runs using 160 randomly selected trees from the stand, were included in this study. The genetic data collected during this study therefore was useful for the evaluation of local genetic variability, and how it relates to other factors.

Within-Stand Clustering of Eastern White Pine Genotypes

Maps of the stand showing the location of trees with similar predicted canonical variate scores, which represented linear combinations of the isozyme scores, showed directional patterns of increasing variate scores. These patterns suggest that trees in different parts of the stand vary genotypically, and that adjacent trees tend to have similar genotypes. The relationship between genotype and tree location was weak, and location only accounted for 11.6 percent of variation in the genotype variation (first canonical variate pair). Low frequency alleles were responsible for some of the patterns observed. The actual number of isozymes

contributing to this relationship and to these patterns was low, and it is possible that if more enzyme systems were surveyed, that stronger relationships between genotype and location would be found. Age of trees was related to their genotype, and the correlation between age and the canonical variate score for genotype was slightly higher than for either of the two coordinate variables (or their interaction) representing location of trees. Clusters of similarly aged trees were found in the stand. Trees of similar age can have similar genotypes, if their genes were contributed by the same parents and the trees were established in canopy openings at approximately the same time. Older trees would not show such a strong relationship between age and genotype, because most of the members of this reproduction cohort would become suppressed and die, and only one or a few trees in an opening would reach maturity. This scenario is consistent with the findings in this study and with published literature (Hibbs 1982, Tigerstedt et al. 1982).

None of the genotype patterns observed in the stand appeared to be related to soil type and quality, but it is possible that an

insufficient number of soil samples were used to test for this relationship. Ideally, the soil next to each tree should be sampled and the results of each type of soil analysis correlated with the tree's genotype. If this analysis demonstrated that tree genotype varies with soil type or quality, the information would have significant impact on forest management and reforestation. To maximize the use of forested land, foresters would need to consider placement of genotypes not only in terms of location within the geographic range of the species, but also in terms of microsite variability.

The Effect of Sample Size on Estimates of Genetic Variability

Results from this study provide a basis for selecting an appropriate sample size for estimating allele frequencies, the mean number of alleles per locus, polymorphism, heterozygosity and for testing for deviations from the Hardy-Weinberg equilibrium in a white pine stand of similar size, composition and genetic variability as the stand used in this study. The results indicate that a 50-tree sample of the stand would have provided reasonably

accurate estimates for some of the measures, but the following sample sizes would have provided more stable estimates: 50 to 100 trees for estimating allele frequencies, 50 to 75 trees for the mean number of alleles per locus, 25 to 50 trees for heterozygosity, 100 to 250 trees for polymorphism, and between 25 and 475 trees would have been needed to correctly conduct the chi-square test to check for deviation from genotype frequencies expected under the Hardy-Weinberg Equilibrium. The number of trees needed in each case depended on the frequencies of the alleles in question, on the precision needed, and on the type of measure desired. The level of precision is arbitrary and is determined by the researcher. In most cases, in this study, 50 or more trees were needed to obtain estimates that were comparable to the true population measures, but many more trees were needed to improve precision further. Fifty trees represent approximately 10% of the stand. These results agree with the sampling strategies of other researchers, and they indicate that estimates obtained using fewer than 50 trees can be unpredictable and incorrect.

Implications for Future Studies and Forest Management

The results of this study show that a genetic population structure does exist in this white pine forest stand, and that it is related to age and tree location. The observed relationships are weak, but they are recognizable. This means that cutting groups of adjacent trees can lead to elimination of genotypes that are not present in other parts of the stand. Loss of genes in such situations is irreversible, unless seedlings from those trees already are established before the trees are cut, and the seedlings are allowed to reach seed bearing age, or those genes are available from an adjacent stand through gene flow. For example, clearing of one acre lots for house construction could lead to the elimination of unique alleles that are present in a few adjacent trees.

This study also demonstrates that studies which use random samples of less than 50 trees per stand are probably not yielding accurate estimates of genetic measures, unless a larger number of loci are analyzed, than were analyzed in this study. Information from such studies probably cannot be used to test for local genetic

population structure, and the adaptation of specific genotypes to microsite conditions. These facts, combined with the fact that most studies in the past have concentrated on range-wide patterns, mean that unless a serious effort is made to evaluate local genetic variability and genetic adaptation to microsite conditions in the near future, we may never have a good understanding of local genetic structure of white pine forests. Genetic structure of the forests probably changes with time in response to both random events and activities of people. As the average stand size is decreased with greater human population pressures, random events, such as openings in canopy due to lightning strikes, result in greater and more unexpected changes in the tree population structure. Having an understanding of local genetic variability, adaptation and genetic structure is important, because it will allow us to make more efficient use of the ever decreasing land available for growing the eastern white pine. The following steps should be taken to help foresters plant and manage white pine wisely in small stands: (1) repeat this study in two other stands of similar size, species composition, density, and with similar microsite variability, using at least 50 percent of the trees in the

stand; (2) include additional enzyme systems and sample soil adjacent to each tree used; (3) develop guidelines to help foresters identify potential groups of related trees, so that whenever possible, at least one representative of such group is retained during thinning operations; (4) maintain detailed records of these studies for use by future generations of foresters as planting of trees becomes a more common method of reforestation on limited and valuable land; (5) repeat these studies for other valuable tree species, and include any new methods that improve the accuracy and rate of determining tree genotypes. When selecting trees for these studies, select trees systematically to include trees from all representative microsites, geographic locations, and age classes in the stand. If these, or similar steps are not taken before the remaining unmanaged forest stands are cut or extensively thinned, we will never know what is the genetic structure of a healthy forest. Future foresters will not know how genotypes should be distributed during plantings to help create new healthy forests. If these steps are followed, foresters in the near and more distant future will be more likely to plant genotypes that are appropriate for local conditions. We will be able to manage

forest stands to increase the chances of survival of healthy trees, as both local and global climate changes occur.

LIST OF REFERENCES

Adams, W.T. and R.J. Joly. 1980. Genetics of allozyme variants in loblolly pine. J. Heredity 71:33-40.

Ayala, F. J. 1982. Population And Evolutionary Genetics: A Primer. The Benjamin/Cummings Publishing Company, Inc. Menlo Park, California. pp. 1-50.

Bailey, N. T. J. 1981. Statistical Methods In Biology. 2nd Edition, John Wiley and sons, New York.

Bergmann, F. and F. Scholz. 1987. The impact of air pollution on the genetic structure of Norway spruce. Silvae Genetica 36(2):80-93.

Brown, A. H. D. and G. F. Moran. 1981. Isozymes and the genetic resources of forest trees. Symposium On Isozymes Of North American Forest Trees And Forest Insects, July 27, 1979, Berkeley, California.

Copes, D. L. 1975. Isoenzyme study of dwarf and normal Douglas-fir trees. Bot. Gaz. 136(4):347-352.

Crow, J. F. and M. Kimura. 1970. An Introduction To Population Genetics Theory. Harper & Row Publishers, New York.

Eckert, R. T., R. J. Joly and D. B. Neale. 1981. Genetics of isozyme variants and linkage relationships among allozyme loci in 35 eastern white pine clones. Can. J. For. Res. 11:573-579.

Eckert, R. T. and J. B. Ryu. 1983. Foliar isozyme variation in twenty seven provenances of Pinus strobus L.: Relationship of three foliar isozyme alleles with height

growth and multivariate patterns of range-wide variation. Proc. Seventh Norh American Forest Biology Workshop, University of Kentucky, Lexington, KY, USA. (August 1982).

EL-Kassaby, Y. A. and O. Sziklai. 1983. Effect of sample size on the precision of the estimate of allozyme frequencies in a natural stand of Douglas-fir. Egypt. J. Genet. Cytol. 12:345-360.

Elsten, R. C. and R. Forthofer. 1977. Testing for Hardy-Weinberg Equilibrium in small samples. Biometrics 33:536-542.

Ewens, W. J. 1972. The sampling theory of selectively neutral alleles. Theor. Pop. Biol. 3:87:112.

Falconer, D. S. 1960. Quantitative Genetics. The Ronald Press Company, New York.

Feret, P. P. 1979. Isozymes for tree improvement. In: Proceedings of the 26th N.E. For. Tree. Improvement Conference, Penn. State Univ., USDA Forest Service, NE For. Exp. Sta. pp. 102-120.

Feret, P. P. 1974. Genetic differences among three small stands of Pinus pungens. Theoretical and Applied Genetics 44:173-177.

Garrett, P. W., E. J. Shreiner, and H. Kettlewood. 1973. Geographic variation of eastern white pine in the Northeast. USDA Forest Service Research Paper NE-274. 14 pages.

Geburek, T., F. Scholz, W. Knabe, and A. Vornweg. 1987. Genetic studies by isozyme gene loci and tolerance and sensitivity in an air polluted Pinus sylvestris field trial. Silvae Genetica 36(2):49-53.

Genys, J. B. 1977. Results of studies of 16 geographic strains of P. strobus and 4 other white pine species in Maryland. 24th Northeastern Forest Tree Improvement Conference, Proceedings 1977, Pages 66-72.

Gittins, R. 1979. Ecological Applications of Canonical Analysis. In: L. Orloci, C. R. Rao, and W. M. Stiteler (eds.), Multivariate Methods in Ecological Work, pp. 309-535. Int. Co-operative Publishing House, Fairland Maryland.

Gorman, M. B. and Y. T. Kiang. 1977. Variety-specific electrophoretic variants of four soybean enzymes. Crop. Sci. 17:963-965.

Grant, M. C. and J. B. Mitton. 1977. Genetic differentiation among growth forms of Engelman spruce and subalpine fir at tree line. Arctic and Alpine Res. 9:259-263.

Gupta, V. K. and S. P. Singh. 1977. Biochemical basis of hybrid vigor. The genetics of grain weight of Oryza sativa. Theor. Appl. Genet. 49:259-264.

Guries, R. P. and F. T. Ledig. 1978. Inheritance of some polymorphic isoenzymes in pitch pine (Pinus rigida mill.) Heredity 40 (1):27-32.

Hartl, D. L. 1980. Principles Of Population Genetics. Sinauer Associates, Inc., Publishers. Sunderland, Massachusetts.

Hibbs, D. E. 1982. White pine in the transition hardwood forest. Can. J. Bot. 60:2046-2053.

Jeffers, R. M. 1977. Eastern white pine seed source variation in Northern Minnesota planting. Proc. 24th Northeastern Forest Tree Conference.

Kuser, J. E. and B. K. Hobbs. 1985. Comparative height growth of white pine (Pinus strobus) seedlings from 63 seed sources. Bull. N. J. Acad. Sci. 30(1):23-28.

Ledig, F. T. and D. M. Smith, 1981. The influence of silvicultural practices on genetic improvement: Height growth and weevil resistance in eastern white pine. Silvae Genetica 30(1):30-36.

Levene, H. 1949. On a matching problem arising in genetics. Ann. Math. Statistics 20:91-94.

Marriott, F. H. C. 1974. The Interpretation of Multiple Observations. Academic Press, London, 117 pp.

Mitton, J. B., P. Knowles, K. B. Sturgeon, Y. B. Linhart, and M. L. Davis. 1979. Association between heterozygosity and growth rate variables in three western forest trees. In: Proceedings of Symposium on Isozymes of North American Forest Trees and Insects, July 27, 1979, Berkeley, Calif. Gen. Tech. Rep. PSW-48. Pacific Southwest Forest and Range Exp. Sta., Forest Serv., U.S. Dep. Agric., Berkeley, Calif., pp. 27-34.

Mitton, J. B., Y. B. Linhart, J. L. Hamrick and J. S. Beckman. 1977. Observations on the genetic structure and mating system of ponderosa pine in the Colorado front range. Theor. Appl. Genet. 51:5-13.

Mitton, J. B. 1977. Genetic differentiation of races of man as judged by single-locus and multi-locus analyses. Am. Nat. 111:203-212.

Mitton, J. B., Y. B. Linhart, K. B. Sturgeon and J. L. Hamrick. 1979. Allozyme polymorphism detected in mature needle tissue of ponderosa pine. J. Heredity 70:86-89.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583 - 590.

Pitel, J. A., and W. M. Cheliak. 1983. Extraction and characterization of isoenzymes from vegetative tissues of five conifers. Proceedings of the seventh North American Forest Biology Workshop. "Physiology and Genetics of Intensive Culture". July 26-28, 1982. University of Kentucky, Lexington, Kentucky. pp. 321-326.

Roberds, J. H. and M. T. Conkle. 1984. Genetic structure in loblolly pine stands: allozyme variation in parents and progeny. For. Sci. 30:319-329.

Rudin, D., G. Eriksson, I. Ekberg, and M. Rasmuson. 1974. Studies of allele frequencies and inbreeding in Scots pine populations by the aid of isozyme technique. Silvae Genetica 23:1-3.

Ryu, J. B. 1982. Genetic structure of Pinus strobus L. based on foliar isozymes from 27 provenances. Ph.D. Dissertation, The University of New Hampshire, Durham, N.H.

Ryu, J. B. and R. T. Eckert. 1983. Foliar isozyme variation in twenty seven provenances of Pinus strobus L.: Relationship of three foliar isozyme alleles with height growth and multivariate patterns of range-wide variation. Proceedings of the Seventh North American Forest Biology Workshop. "Physiology and Genetics of Intensive Culture." July 26-28, 1982. University of Kentucky, Lexington, Kentucky. pp. 179-187.

Scandalios, J. G. 1969. Genetic control of multiple molecular forms of enzymes in plants: a review. Biochem. Genet. 3:37-79.

Scandalios, J. G. 1974. Isozymes in development and differentiation. Ann. Rev. Plant Physiol. 25:225-258.

Scandalios. J. G., E. Liu and M. A. Campeau. 1972. The effect of intragenic and intergenic complementation on catalase structure and function in maize. A molecular approach to heterosis. Arch. biochem. Biophys. 153:695-705.

Schwartz, D. and W. J. Laughner. 1969. A molecular basis for heterosis. Science 166:626-627.

Smouse, P. E., and R. S. Spielman. 1977. How allocation of individuals depends on genetic differences among populations. Excerpta Med. Int. Congr. Ser. 411:255-260.

Smouse, P. E., R. S. Spielman, and M. H. Park. 1982. Multiple-locus allocation of individuals to groups as a function of the genetic variation within and differences among human populations. The American Naturalist 119:445-463.

Sokal, R. R. and F. J. Rohlf. 1969. Statistical Tables. W. H. Freeman, San Francisco. (Prepared in conjunction with Sokal, R. R. and F. J. Rohlf. BIOMETRY; The Principles And Practices Of Statistics In Biological Research, 2nd edition, c. 1981, W. H. Freeman, San Francisco, 219 p.)

Steel, R. G. D. and J. H. Torrie. 1980. Principles And Procedures Of Statistics, A Biometrical Approach. Second Edition. McGraw-Hill Book Company, New York. pp. 26-27.

Swofford, D. L. and R. B. Selander. 1981. BIOSYS-1: A computer program for the analysis of allelic variation in genetics, User's Manual. Dept. of Genetics and Development, University Of Illinois At Urbana-Champaign. Urbana, Ill. 60801. U.S.A.

Tigerstedt, P. M. A. 1973. Studies on isozyme variation in marginal and central populations of Picea abies. Hereditas 75:47-60.

Tigerstedt, P.M.A., D. Rudin, T. Niemela, and J. Tammisola. 1982. *Competition and neighboring effect in a naturally* regenerating population of Scots pine. Silva Fenn., 16:122-129.

Ward, R. H., and C. F. Sing. 1970. A consideration of the power of the x^2 test to detect inbreeding effects in natural populations. Am. Nat. 104:355-365.

Yeh, F. C., W. M. Cheliak, B. P. Dancik. 1985. Population differentiation in lodgepole pine, Pinus contorta spp. latifolia: a discriminant analysis of allozyme variation. Can. J. Genet. Cytol. 27:210-218.

TABLE 1.1. Extraction buffer for white pine foliage.[1]

COMPONENT	AMOUNT
Germanium dioxide[2]	0.150 g
Sodium borate (decahydrated)	2.200 g
Sodium bisulfite	0.600 g
Soluble polyvinylpyrolidone (PVP-4OT)	8.000 g
L-ascorbic acid	8.000 g
Diethyldithiocarbanic acid	0.550 g
Potassium phosphate buffer (0.16M KH_2PO_4, pH. 7.0)[3]	16.000 ml
2-phenoxyethanol	1.000 ml

[1]Based on recipe of Mitton et al. 1977, adjusted for 50 samples.

[2]Dissolve in 150 ml of boiling water. Cool solution to 2°-3°C.

[3]Mix with dry chemicals to make a paste, then add the germanium dioxide solution, and then add 2-phenoxyethanol.

TABLE 1.2. Enzymes, buffers and electrophoretic conditions used for analysis of eastern white pine foliage.

ENZYMES	References for buffers and conditions for electrophoresis	Gel and Electrode buffers[a]
Leucine aminopeptidase (LAP E.C. 3.4.11.1)	Scandalios 1969	1
Phospho-glucose isomerase (PGI E.C. 5.3.1.9)	Scandalios 1969	1
Glutamate-oxaloacetate transaminase (GOT E.C. 2.6.1.1)	Hamrick 1980[b]	2
Shikimate dehydrogenase (SDH E.C. 1.1.1.25)	Hamrick 1980[b]	2
Fluorescent esterase (F-EST E.C. 3.1.1.2)	Mitton et al. 1977	3
Phosphoglucomutase (PGM E.C. 2.7.5.1)	Mitton et al. 1977	3
Malate dehydrogenase (MDH E.C. 1.1.1.37)	Neale 1980[c]	4

[a]Buffers and conditions: (1) stock a, 12.0 g. LiOH (monohydrate) - 118.9 g. H_3BO_3; pH 8.1; stock b, 62.0 g. Tris 16.0 g. citric acid (monohydrate); pH 8.3; gel, 1 part a and 9 part b; electrode, a only; amperage, 15 - 40 mA; duration, about 4.5 hours; (2) gel, 121.1 g. Tris- 25.2 g. citric acid (monohydrate); pH 8.45; adjust pH with 0.2 M citric acid; electrode, 166.0 g. boric acid - 20.0 g. NaOH; pH 8.0; amperage, 45 - 55 mA; duration, about 4 hours; (3) gel, 14.5 g. Tris - 5.8 g. citric acid (monohydrate); pH 7.7; electrode, 115.875 g. boric

acid - 4.65 g. NaOH; pH 7.5; amperage, 5 - 25 mA; duration, about 4 hours; (4) gel, 1 : 20 dilution of electrode buffer; electrode, 76.838 g. citric acid (anhydrous); pH 6.1; adjust pH with morpholine; amperage, 20 - 25 mA; duration, 20 hours.

NOTE: For each enzyme system, combine distilled water with dry chemicals to total 10 liters of buffer. Use first amperage for 15 minutes, remove wicks, then use second amperage.

[b]Hamrick, 1980; personal communication by Ryu, 1982.

[c]Neale, 1980; personal communication by Ryu, 1982.

TABLE 1.3. Stain recipes for foliar enzymes in eastern white pine.[a]

ENZYME	COMPONENTS	AMOUNT
LAP	L-leucyl-napthamide	35.0 mg
	.2M Tris maleate buffer pH 3.8	89.0 ml
	.2M NaOH	45.0 ml
	Black K salt	90.0 mg
	H_2O	53.0 ml
PGI	1M Tris HC1 pH 8.0	4.6 ml
	MTT	30.0 mg
	PMS	7.5 mg
	G-6-PDH (10 u/ml)	6.5 ml
	Fructose-6-P	10.5 ml
	NADP	10.0 mg
	H_2O	170.0 ml
	$MgCl_2$ (10%)	8.8 ml
PGM	1M Tris HC1 pH 8.0	3.8 ml
	MTT	19.0 mg
	PMS	7.5 mg
	G-1, 6-DP (.05%)	2.5 ml
	G-6-PDH (10 u/ml)	13.0 ml
	NADP	8.0 mg
	H_2O	147.0 ml
	$MgC1_2$ (10%)	3.8 ml
	Na_2 G-1-P (1%)	12.5 ml
GOT	Pyridoxal-5'-phosphate	8.0 mg
	L-aspartic acid	320.0 mg
	ketoglutaric acid	160.0 mg
	1M Tris HC1 pH 8.8	14.0 ml
	Fast blue BB salt	240.0 mg
	H_2O	186.0 ml
F-EST	4-methylumbelliferyl acetate	35.0 mg
	Acetone	20.0 ml
	1M Na acetate pH 5.0	15.0 ml
	H_2O	57.0 ml

	Apply with paper towel, score immediately in UV light	
SDH	05M Tris HC1 pH 8.0	13.5 ml
	MTT	7.5 mg
	PMS	9.0 mg
	(-) shikimic acid	12.0 mg
	NADP	12.0 mg
	Agar solution (1%)	18.0 ml
MDH	.25M Tris HC1 pH 8.6	12.0 ml
	L-malic acid	47.0 mg
	MTT	22.5 mg
	PMS	4.5 mg
	NAD	23.0 mg
	Agar solution (1%)	18.0 ml

[a]Recipes are for three gel slices. Pour stain solutions over sliced gels and incubate in dark at 37°C at least one hour, unless otherwise indicated.

TABLE 1.4. Observed ratios of allozymes from megagametophytes of heterozygous parent trees and chi-square analysis of ratios to expected 1:1 allozyme segregation.

Enzyme	Alleles a[1]	Alleles b[1]	Observed Number a[2]	Observed Number b[2]	x^2(1:1)	Deviation Prob.[3]
SDH	1	2	150	145	0.08	0.78
MDH-1	1	2	131	140	0.29	0.61
	1	3	67	97	5.48	0.02
	2	3	53	44	0.84	0.39
F-EST-1	1	2	52	49	0.09	0.77
PGI-2	1	2	64	81	1.99	0.17

[1] Numbers in the "a" and "b" columns indicate which pair of alleles was considered. For example, SDH a=1 and b=2 indicates that the segregation of the first and second alleles of SDH was tested.

[2] Numbers in the "a" and "b" columns show the count of how many times a particular allele was detected. For example, SDH a=150 and b=145 shows that the first allele was detected in 150 megagametophytes and the second allele was detected in 145 megagametophytes. The ratio is 150:145 with an expected ratio of 147.5:147.5 (or 1:1). Ratios for each enzyme system are based on a different number of observations, because a different number of seeds were available from each of the appropriate heterozygous parent trees.

[3] Deviation probability of 0.01 was used as a cut off for selecting statistically significant deviations from expected ratios.

Table 2.1. Canonical correlation analysis of isozyme variables and location of eastern white pines in a two hectare unmanaged forest stand. [1]

K	r_k	r_k^2	F	NDF	DDF	PR>F
1	0.341	0.116	2.342	50	2251.79	0.0001
2	0.261	0.068	1.496	36	1852.98	0.0297
3	0.145	0.021	0.762	24	1436.25	0.7883

K = number of dimensions
r_k = canonical correlation coefficient
r_k^2 = squared canonical correlation coefficient

F = approximate F based on likelihood ratios testing the null hypothesis (Ho) that the canonical correlation in the current row and all that follow are zero.

NDF=numerator degrees of freedom
DDF=denominator degrees of freedom

N = 508 observations

Isozyme loci recoded to continuous scores (Smouse et al.1982): SDH, PGM-2, GOT-2, PGI-2, MDH-1, F-EST-1, and FEST-2

Location variables: used as X (south-west to north-east) and Y (south-east to north-west) coordinates, X-squared, Y-squared and product of X and Y.

[1] The canonical correlation coefficient (r_k) is a measure of the relationship between the location and isozyme canonical variate scores.

Table 2.2. Canonical correlation analysis of isozyme variables and tree location of eastern white pines in a two hectare unmanaged forest stand. [1]

Isozymes[1]	V1	V2	W1	W2
SDH(1)	<u>-0.5045</u>	0.0992	-0.1721	0.0259
SDH(2)	0.4394	-0.0874	0.1499	-0.0228
PGM-2(1)	0.2551	0.0292	0.0870	0.0076
PGM-2(2)	-0.3226	-0.0820	-0.1101	-0.0214
GOT-2(1)	0.3020	0.1259	0.1031	0.0329
PGI-2(1)	-0.1573	<u>0.8590</u>	-0.0537	0.2245
MDH-1(1)	0.1380	0.0374	0.0471	0.0098
MDH-1(2)	-0.1612	0.0862	-0.0550	0.0225
FEST-1(1)	0.3268	0.2062	0.1115	0.0539
FEST-2(1)	<u>0.7558</u>	0.2490	0.2579	0.0651
Tree Location Variables				
X	0.0022	-0.0701	0.0066	-0.2681
X2	-0.0010	-0.0469	-0.0030	-0.1795
Y	0.2634	-0.1510	<u>0.7721</u>	<u>-0.5777</u>
Y2	0.2750	-0.1463	<u>0.8059</u>	<u>-0.5599</u>
XY	0.1226	-0.1413	0.3593	<u>-0.5406</u>

[1] Isozyme variables were represented using Smouse Scores (Smouse et al. 1982). The correlation coefficients show the relationship between the raw isozyme and location variables and the canonical variate scores. High correlations (>.50) are underlined.

Table 2.3. Canonical correlation analysis of isozyme variables, location, and age of eastern white pines in a two hectare unmanaged forest stand.[1]

K	r_k	r_k^2	F	Num.Df.	Den.DF	PR>F
1	0.498	0.248	1.616	54	733.75	0.0042
2	0.373	0.139	1.074	40	630.48	0.3527

K = number of dimensions
r_k = canonical correlation coefficient
r_k^2 = squared canonical correlation coefficient

F = approximate F based on likelihood ratios testing the null hypothesis (Ho) that the canonical correlation in the current row and all that follow are zero.

NDF=numerator degrees of freedom
DDF=denominator degrees of freedom

N = 158 observations

Isozyme loci recoded to continuous scores (Smouse et al.1982): SDH, PGM-2, GOT-2, PGI-2, MDH-1, F-EST-1, and FEST-2

Location variables: used as X (south-west to north-east) and Y (south-east to north-west) coordinates, X-squared, Y-squared and product of X and Y.

Age: measured in years

[1] The canonical correlation coefficient (r_k) is a meàsure of the relationship between the location and isozyme canonical variate scores.

Table 2.4. Canonical correlation analysis of isozyme variables, tree location, and age of eastern white pines in a two hectare unmanaged forest stand.[1]

Isozymes[1]	V1	V2	W1	W2
SDH(1)	0.1906	0.3794	0.0950	0.1413
SDH(2)	-0.3465	-0.4465	-0.1727	-0.1664
PGM-2(1)	-0.3318	-0.0841	-0.1654	-0.0313
PGM-2(2)	0.2710	0.2706	0.1351	0.1008
GOT-2(1)	0.0000	0.0000	0.0000	0.0000
PGI-2(1)	0.0379	<u>-0.8024</u>	0.0189	-0.2989
MDH-1(1)	0.0285	-0.1421	0.0142	-0.0529
MDH-1(2)	-0.0387	-0.1409	-0.0193	-0.0525
FEST-1(1)	0.3635	0.0944	0.1812	0.0352
FEST-2(1)	<u>0.8138</u>	-0.1887	0.4056	-0.0703
Tree Location Variables				
X	-0.1967	0.2264	-0.3947	0.6076
X2	-0.1933	0.2027	-0.3879	0.5441
Y	0.2252	0.2913	0.4518	0.7819
Y2	0.2401	0.2877	0.4816	0.7723
XY	-0.0545	0.3226	-0.1094	0.8658
AGE	-0.2289	-0.1067	-0.4591	-0.2865

[1] Isozyme variables were represented using Smouse Scores (Smouse et al. 1982). The correlation coefficients show the relationship between the raw isozyme and location variables and the canonical variate scores. High correlations (>0.50) are underlined.

TABLE 3.1. Enzyme variant (allele)[1] frequencies of 511 eastern white pines in a two hectare stand, Durham, NH.

Enzyme Variant	Frequency	Enzyme Variant	Frequency
GOT-1	1.000	PGM-2(2)	0.178
GOT-2(1)	0.996	PGM-2(3)	0.049
GOT-2(2)	0.004	MDH-1(1)	0.593
PGI-1	1.000	MDH-1(2)	0.285
PGI-2(1)	0.523	MDH-1(3)	0.122
PGI-2(2)	0.477	MDH-2	1.000
SDH(1)	0.931	LAP	1.000
SDH(2)	0.054	F-EST-1(1)	0.502
SDH(3)	0.016	F-EST-1(2)	0.498
PGM-1	1.000	F-EST-2(1)	0.934
PGM-2(1)	0.773	F-EST-2(2)	0.066

[1] The fastest migrating zone of each enzyme is designated locus one, the following locus two, etc. Alleles at each locus are identified as 1 for the most frequent allele, 2 for the next to most frequent allele, etc.; the fast allele of the first MDH locus was also the most frequent allele at that locus and is represented as MDH-1(1), for example.

TABLE 3.2. Mean frequency estimates for low frequency enzyme variants obtained using five replicates of each sample size category.

Number of Trees	Enzyme Variant (Allele) SDH(2)	SDH(3)	PGM-2(3)	MDH-1(3)	F-EST-2(2)
5	0.020	0.000	0.020	0.060	0.080
25	0.052	0.028	0.040	0.112	0.064
50	0.054	0.018	0.044	0.119	0.060
75	0.040	0.012	0.038	0.108	0.069
100	0.046	0.014	0.052	0.123	0.065
125	0.062	0.010	0.045	0.107	0.054
150	0.058	0.015	0.046	0.114	0.073
175	0.054	0.014	0.052	0.123	0.062
200	0.053	0.015	0.048	0.117	0.066
225	0.047	0.015	0.047	0.117	0.067
250	0.053	0.015	0.046	0.129	0.063
275	0.047	0.017	0.051	0.118	0.070
300	0.054	0.014	0.049	0.124	0.065
325	0.053	0.014	0.050	0.119	0.073
350	0.054	0.014	0.048	0.117	0.065
375	0.054	0.015	0.047	0.127	0.062
400	0.052	0.016	0.053	0.125	0.068
425	0.055	0.014	0.048	0.122	0.066
450	0.052	0.015	0.050	0.125	0.067
475	0.052	0.016	0.049	0.119	0.065
511*	0.054	0.016	0.049	0.122	0.066

* One sample with all the trees

TABLE 3.3. Standard deviations for mean frequency estimates for low frequency enzyme variants obtained using five replicates of each sample size category.

Number of Trees	Enzyme Variant (Allele)				
	SDH(2)	SDH(3)	PGM-2(3)	MDH-1(3)	F-EST-2(2)
5	0.044	0.000	0.045	0.089	0.084
25	0.018	0.041	0.020	0.013	0.036
50	0.013	0.017	0.009	0.035	0.023
75	0.012	0.009	0.020	0.038	0.024
100	0.010	0.009	0.017	0.025	0.014
125	0.017	0.003	0.012	0.019	0.004
150	0.009	0.008	0.010	0.011	0.014
175	0.011	0.002	0.013	0.021	0.009
200	0.006	0.007	0.008	0.011	0.018
225	0.010	0.008	0.008	0.018	0.010
250	0.007	0.003	0.011	0.009	0.005
275	0.004	0.004	0.004	0.006	0.007
300	0.003	0.002	0.004	0.002	0.005
325	0.004	0.006	0.002	0.009	0.009
350	0.005	0.002	0.006	0.013	0.006
375	0.007	0.003	0.006	0.005	0.008
400	0.003	0.004	0.002	0.003	0.005
425	0.004	0.004	0.004	0.004	0.005
450	0.003	0.002	0.002	0.005	0.003
475	0.002	0.001	0.003	0.007	0.002

TABLE 3.4. Coefficients of variation for mean frequency estimates of low frequency enzyme variants obtained using five replicates of each sample size category.

Number of Trees	Enzyme Variant				
	SDH(2)	SDH(3)	PGM-2(3)	MDH-1(3)	F-EST-2(2)
5	2.236	-	2.236	1.491	1.046
25	0.344	1.481	0.500	0.655	0.559
50	0.248	0.994	0.203	0.296	0.388
75	0.306	0.829	0.526	0.347	0.346
100	0.223	0.687	0.323	0.207	0.211
125	0.278	0.344	0.271	0.179	0.066
150	0.166	0.504	0.222	0.115	0.198
175	0.201	0.151	0.247	0.167	0.142
200	0.121	0.445	0.160	0.091	0.272
225	0.221	0.561	0.162	0.154	0.143
250	0.132	0.205	0.249	0.068	0.086
275	0.088	0.215	0.070	0.533	0.099
300	0.060	0.168	0.093	0.013	0.069
325	0.073	0.427	0.032	0.077	0.120
350	0.087	0.168	0.129	0.113	0.093
375	0.121	0.221	0.137	0.042	0.135
400	0.061	0.256	0.044	0.023	0.078
425	0.081	0.270	0.085	0.031	0.073
450	0.059	0.133	0.036	0.037	0.047
475	0.039	0.080	0.057	0.060	0.033
Correlation Coefficient[1]	-0.525	-0.774	-0.579	-0.621	-0.642

[1] Correlation Coefficient between sample size and coefficient of variation for the mean of that sample size category.

TABLE 3.5. PGM-2(2) frequency estimates observed in five random samples of various sizes in the same forest stand.[1]

Num. of Trees	Replicate 1	2	3	4	5	Mean Freq.	Stan. Dev.	Coeff. Var.
5	0.000	0.400	0.000	0.300	0.000	0.140	0.195	1.392
25	0.220	0.400	0.140	0.120	0.180	0.212	0.112	0.528
50	0.120	0.210	0.190	0.160	0.220	0.180	0.041	0.226
75	0.213	0.173	0.167	0.227	0.180	0.192	0.026	0.138
100	0.205	0.200	0.130	0.175	0.195	0.181	0.031	0.170
125	0.188	0.188	0.188	0.188	0.176	0.186	0.005	0.029
150	0.177	0.197	0.163	0.173	0.163	0.175	0.014	0.080
175	0.189	0.177	0.180	0.180	0.174	0.180	0.006	0.031
200	0.180	0.168	0.188	0.180	0.177	0.179	0.007	0.040
225	0.144	0.173	0.198	0.200	0.198	0.183	0.024	0.133
250	0.194	0.192	0.196	0.170	0.174	0.185	0.012	0.066
275	0.162	0.169	0.198	0.164	0.178	0.174	0.015	0.085
300	0.173	0.183	0.178	0.182	0.188	0.181	0.006	0.031
325	0.194	0.180	0.183	0.178	0.178	0.183	0.007	0.037
350	0.169	0.171	0.184	0.167	0.176	0.173	0.007	0.040
375	0.180	0.187	0.175	0.165	0.173	0.176	0.008	0.047
400	0.160	0.171	0.189	0.181	0.176	0.175	0.011	0.062
425	0.185	0.175	0.181	0.184	0.184	0.182	0.004	0.022
450	0.172	0.177	0.181	0.172	0.178	0.176	0.004	0.022
475	0.177	0.180	0.177	0.173	0.180	0.177	0.003	0.016

[1] The true PGM-2(2) frequency based on all 511 trees is 0.178.

TABLE 3.6. Mean values of heterozygosity (H), polymorphism (P), and number of alleles per locus (A), obtained from five random samples in each sample size category of the same forest stand.[1]

Number of Trees	H-1	H-2	H-3	P-1	P-2	P-3	A
5	0.153	0.170	0.223	35.00	35.00	35.00	1.400
25	0.180	0.184	0.248	45.00	51.67	51.67	1.734
50	0.180	0.181	0.245	45.00	51.67	51.67	1.734
75	0.178	0.179	0.239	45.00	53.33	53.33	1.767
100	0.181	0.182	0.243	45.00	51.67	51.67	1.750
125	0.178	0.179	0.246	50.00	50.00	51.67	1.766
150	0.183	0.183	0.244	50.00	50.00	53.33	1.782
175	0.181	0.182	0.245	50.00	50.00	53.33	1.782
200	0.181	0.181	0.240	48.33	53.33	58.33	1.830
225	0.181	0.181	0.243	48.33	50.00	53.33	1.782
250	0.182	0.182	0.246	50.00	50.00	56.66	1.814
275	0.181	0.181	0.241	50.00	50.00	58.33	1.830
300	0.182	0.183	0.244	50.00	50.00	56.66	1.814
325	0.183	0.183	0.242	50.00	50.00	58.33	1.830
350	0.180	0.181	0.242	50.00	50.00	56.66	1.814
375	0.182	0.182	0.243	50.00	50.00	58.33	1.830
400	0.182	0.183	0.242	50.00	50.00	56.66	1.814
425	0.182	0.182	0.242	50.00	50.00	58.33	1.830
450	0.182	0.182	0.242	50.00	50.00	58.33	1.830
475	0.181	0.182	0.242	50.00	50.00	58.33	1.830
511*	0.182	0.182	0.242	50.00	50.00	58.33	1.830

[1]The following loci were included: SDH, PGM-2, GOT-2, PGI-2, MDH-1, F-EST-1and F-EST-2

H-1 = H (Biased)
H-2 = H (Unbiased)
H-3 = H (Direct)

P-1 = P (0.95 Criterion)
P-2 = P (0.99 Criterion)
P-3 = P (No Criterion)

* One sample with all the trees

TABLE 3.7. Standard deviation of mean heterozygosity (H), polymorphism (P), and the number of alleles per locus (A), obtained from five random samples in each sample size category of the same forest stand.

Number of Trees	H-1	H-2	H-3	P-1	P-2	P-3	A
5	0.029	0.032	0.019	6.97	6.97	6.97	0.109
25	0.009	0.009	0.009	7.45	3.73	3.73	0.067
50	0.007	0.007	0.012	4.56	3.73	3.73	0.067
75	0.007	0.007	0.006	4.56	4.56	4.56	0.036
100	0.006	0.006	0.008	4.56	3.73	3.73	0.057
125	0.001	0.001	0.003	0.00	0.00	3.73	0.036
150	0.004	0.003	0.003	0.00	0.00	4.56	0.044
175	0.003	0.002	0.002	0.00	0.00	4.56	0.044
200	0.003	0.003	0.003	3.73	4.56	0.00	0.000
225	0.006	0.005	0.005	3.73	0.00	4.56	0.044
250	0.003	0.003	0.003	0.00	0.00	3.73	0.036
275	0.003	0.003	0.004	0.00	0.00	0.00	0.000
300	0.001	0.001	0.002	0.00	0.00	3.73	0.036
325	0.001	0.002	0.003	0.00	0.00	0.00	0.000
350	0.003	0.003	0.002	0.00	0.00	3.73	0.036
375	0.002	0.002	0.002	0.00	0.00	0.00	0.000
400	0.001	0.001	0.002	0.00	0.00	3.73	0.036
425	0.002	0.002	0.001	0.00	0.00	0.00	0.000
450	0.000	0.000	0.001	0.00	0.00	0.00	0.000
475	0.000	0.001	0.001	0.00	0.00	0.00	0.000

[1]The following loci were included: SDH, PGM-2, GOT-2, PGI-2, MDH-1, F-EST-1 and F-EST-2

H-1 = S.D. of H (Biased)
H-2 = S.D. of H (Unbiased)
H-3 = S.D. of H (Direct)

P-1 = S.D. of P (.95 Criterion)
P-2 = S.D. of P (.99 Criterion)
P-3 = S.D. of P (No Criterion)

TABLE 3.8. Coefficients of variation for mean heterozygosity (H), polymorphism (P), and the number of alleles per locus (A), obtained from five random samples in each sample size category of the same forest stand.[1]

Number of Trees	H-1	H-2	H-3	P-1	P-2	P-3	A
5	0.190	0.190	0.085	0.20	0.20	0.20	0.078
25	0.050	0.051	0.037	0.17	0.07	0.07	0.039
50	0.037	0.038	0.049	0.10	0.07	0.07	0.039
75	0.040	0.041	0.023	0.10	0.09	0.09	0.020
100	0.031	0.031	0.033	0.10	0.07	0.07	0.032
125	0.008	0.008	0.012	0.00	0.00	0.07	0.020
150	0.020	0.019	0.010	0.00	0.00	0.09	0.025
175	0.015	0.013	0.008	0.00	0.00	0.09	0.025
200	0.018	0.018	0.012	0.08	0.09	0.00	0.000
225	0.031	0.030	0.022	0.08	0.00	0.09	0.025
250	0.014	0.016	0.011	0.00	0.00	0.07	0.020
275	0.015	0.015	0.015	0.00	0.00	0.00	0.000
300	0.007	0.007	0.009	0.00	0.00	0.07	0.020
325	0.007	0.009	0.012	0.00	0.00	0.00	0.000
350	0.017	0.015	0.010	0.00	0.00	0.07	0.020
375	0.008	0.008	0.008	0.00	0.00	0.00	0.000
400	0.003	0.003	0.007	0.00	0.00	0.07	0.020
425	0.008	0.008	0.005	0.00	0.00	0.00	0.000
450	0.002	0.000	0.005	0.00	0.00	0.00	0.000
475	0.002	0.005	0.005	0.00	0.00	0.00	0.000

[1]The following loci were included: SDH, PGM-2, GOT-2, PGI-2, MDH-1, F-EST-1 and F-EST-2

H-1 = S.D. of H (Biased)
H-2 = S.D. of H (Unbiased)
H-3 = S.D. of H (Direct)

P-1 = S.D. of P (.95 Criterion)
P-2 = S.D. of P (.99 Criterion)
P-3 = S.D. of P (No Criterion)

TABLE 3.9. Mean probability (P) values obtained from five significance tests per sample size category, used to detect isozyme variant frequency deviation from frequencies expected under the Hardy-Weinberg equilibrium.

Number of Trees	SDH	PGM-2	GOT-2	PGI-2	MDH-1	FEST-1	FEST-2
5	0.804	0.422	.	0.047	0.229	0.025	0.544
25	0.350	0.675	0.000	0.000	0.570	0.000	0.389
50	0.645	0.468	0.000	0.000	0.291	0.000	0.154
75	0.522	0.582	0.000	0.000	0.266	0.000	0.153
100	0.454	0.260	0.000	0.000	0.215	0.000	0.001
125	0.225	0.261	0.000	0.000	0.391	0.000	0.000
150	0.040	0.252	0.000	0.000	0.200	0.000	0.000
175	0.259	0.329	0.000	0.000	0.030	0.000	0.000
200	0.186	0.336	0.000	0.000	0.022	0.000	0.000
225	0.192	0.326	0.000	0.000	0.050	0.000	0.000
250	0.097	0.069	0.000	0.000	0.023	0.000	0.000
275	0.000	0.012	0.000	0.000	0.021	0.000	0.000
300	0.000	0.221	0.000	0.000	0.006	0.000	0.000
325	0.000	0.019	0.000	0.000	0.001	0.000	0.000
350	0.000	0.058	0.000	0.000	0.001	0.000	0.000
375	0.016	0.041	0.000	0.000	0.001	0.000	0.000
400	0.000	0.006	0.000	0.000	0.000	0.000	0.000
425	0.021	0.070	0.000	0.000	0.000	0.000	0.000
450	0.000	0.007	0.000	0.000	0.000	0.000	0.000
475	0.000	0.006	0.000	0.000	0.000	0.000	0.000
511*	0.000	0.001**	0.000	0.000	0.000	0.000	0.000

* One sample with all the trees

** See Table 3.10 for results from chi-square test with pooling.

TABLE 3.10. Mean probability (P) values obtained from five significance tests per sample size category, using the chi-square test with pooling to detect isozyme variant frequency deviation from frequencies expected under the Hardy-Weinberg equilibrium.

Number of Trees	SDH	SDH*	PGM-2	PGM-2*	MDH-1	MDH-1*
5	-	-	0.025	0.011	0.211	0.121
25	0.286	0.297	0.640	0.645	0.524	0.508
50	0.341	0.348	0.476	0.437	0.361	0.380
75	0.359	0.363	0.553	0.524	0.517	0.505
100	0.392	0.277	0.350	0.351	0.340	0.319
125	0.385	0.353	0.522	0.522	0.487	0.473
150	0.290	0.272	0.243	0.227	0.392	0.372
175	0.161	0.154	0.552	0.553	0.468	0.458
200	0.262	0.249	0.371	0.353	0.141	0.133
225	0.170	0.166	0.251	0.249	0.295	0.280
250	0.076	0.069	0.286	0.272	0.363	0.347
275	0.045	0.040	0.467	0.450	0.151	0.122
300	0.155	0.147	0.234	0.223	0.240	0.231
325	0.119	0.112	0.242	0.232	0.131	0.125
350	0.078	0.061	0.273	0.233	0.083	0.068
375	0.033	0.030	0.148	0.142	0.048	0.045
400	0.068	0.064	0.132	0.126	0.023	0.023
425	0.061	0.058	0.234	0.225	0.029	0.027
450	0.018	0.017	0.097	0.093	0.025	0.023
475	0.008	0.007	0.165	0.158	0.028	0.026
511**	0.002	0.002	0.093	0.088	0.014	0.013

* Levene (1949) correction for small sample size was employed

** One sample with all the trees

TABLE 3.11. Mean probability values obtained from five significance tests per sample size category using observed and expected genotype frequencies, and exact probabilities to detect isozyme variant frequency deviation from frequencies expected under the Hardy-Weinberg equilibrium.[1]

Number of Trees	SDH	PGM-2	GOT-2	PGI-2	MDH-1	FEST-1	FEST-2
			Usual Test[2]				
5	1.000	0.311	.	0.078	0.213	0.046	0.667
25	0.359	0.691	0.000	0.000	0.563	0.000	0.399
50	0.663	0.467	0.000	0.000	0.278	0.000	0.160
75	0.529	0.564	0.000	0.000	0.256	0.000	0.158
100	0.448	0.259	0.000	0.000	0.208	0.000	0.000
125	0.220	0.249	0.000	0.000	0.371	0.000	0.000
			Exact Test[3]				
5	1.000	0.500	.	0.187	0.432	0.127	0.704
25	0.437	0.917	0.020	0.000	0.665	0.000	0.487
50	0.840	0.530	0.010	0.000	0.451	0.000	0.217
75	0.647	0.609	0.007	0.000	0.590	0.000	0.203
100	0.576	0.365	0.005	0.000	0.364	0.000	0.007
125	0.333	0.582	0.004	0.000	0.514	0.000	0.000

[1]Levene's (1949) correction for small sample size was employed.

[2]The usual chi-square goodness-of-fit test is performed using observed genotype frequencies and those expected under Hardy-Weinberg equilibrium (Swofford and Selander 1981).

[3]The exact test for calculating exact significance probabilities is analogous to the Fisher's exact test for 2x2 contingency tables. For large samples (>125 trees) the exact test becomes computationally burdensome (Swofford and Selander 1981).

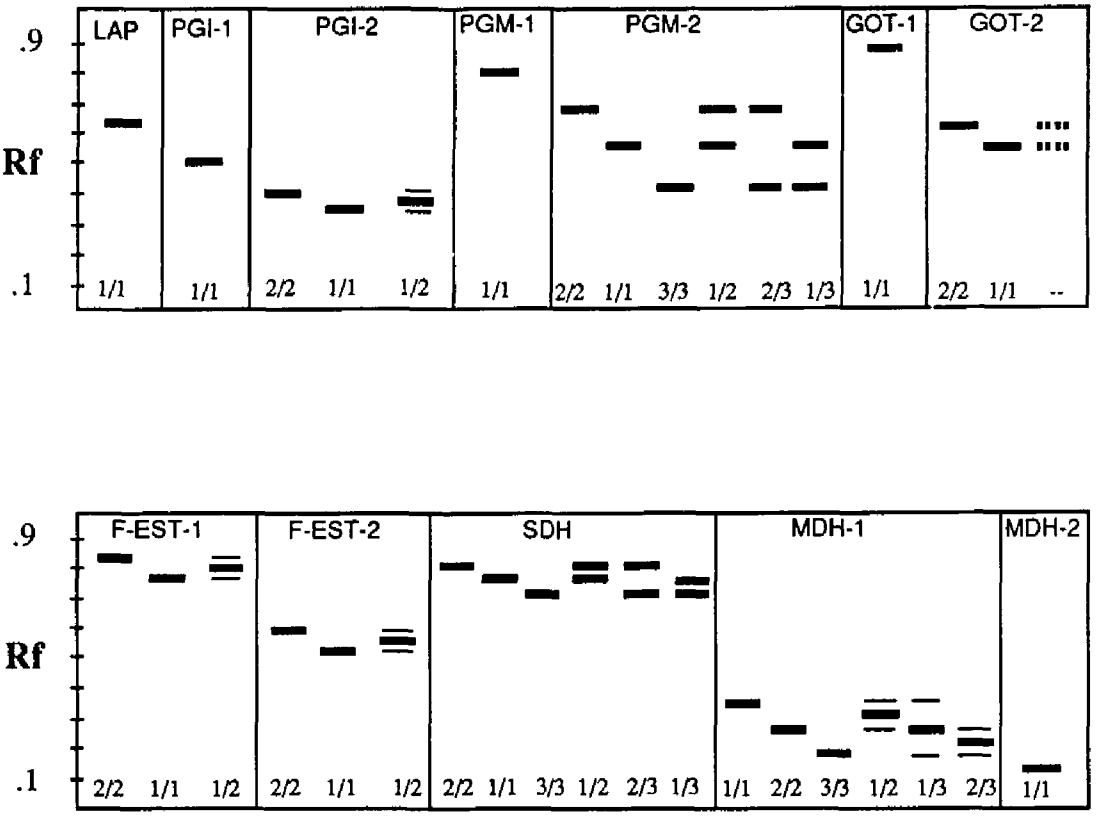


Figure 1.1 -- Band pattern and genotype designation for LAP, PGI, PGM, GOT, F-EST, and MDH in eastern white pine foliage. Rf is the migrational distance of enzymes relative to that of the buffer front (depression in gel), from the origin. Solid lines represent observed genotypes and dotted lines represent expected but not observed genotypes. Alleles at each locus are identified as 1 for the most frequent allele, 2 for the next to most frequent allele, etc.

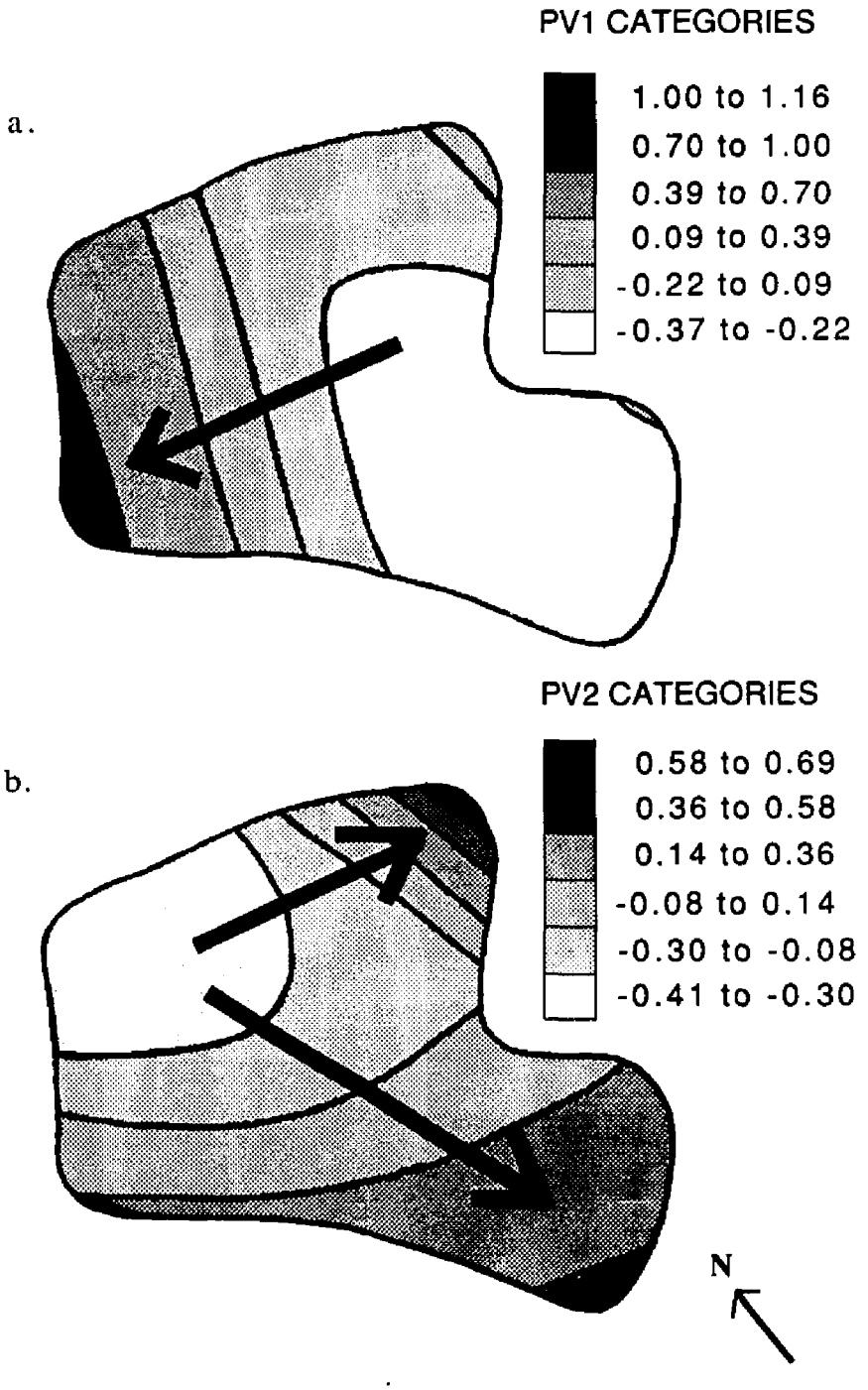


Figure 2.1-- Regions of a two hectare forest stand containing eastern white pines with similar predicted canonical variate scores representing allelic composition for SDH, PGM-2, GOT-2, PGI-2, MDH-1, F-EST-1, and F-EST-2. The scores are predicted from tree location. The arrows show the direction in which predicted scores increase in the stand. Differences in scores reflect allele frequency changes throughout the stand.

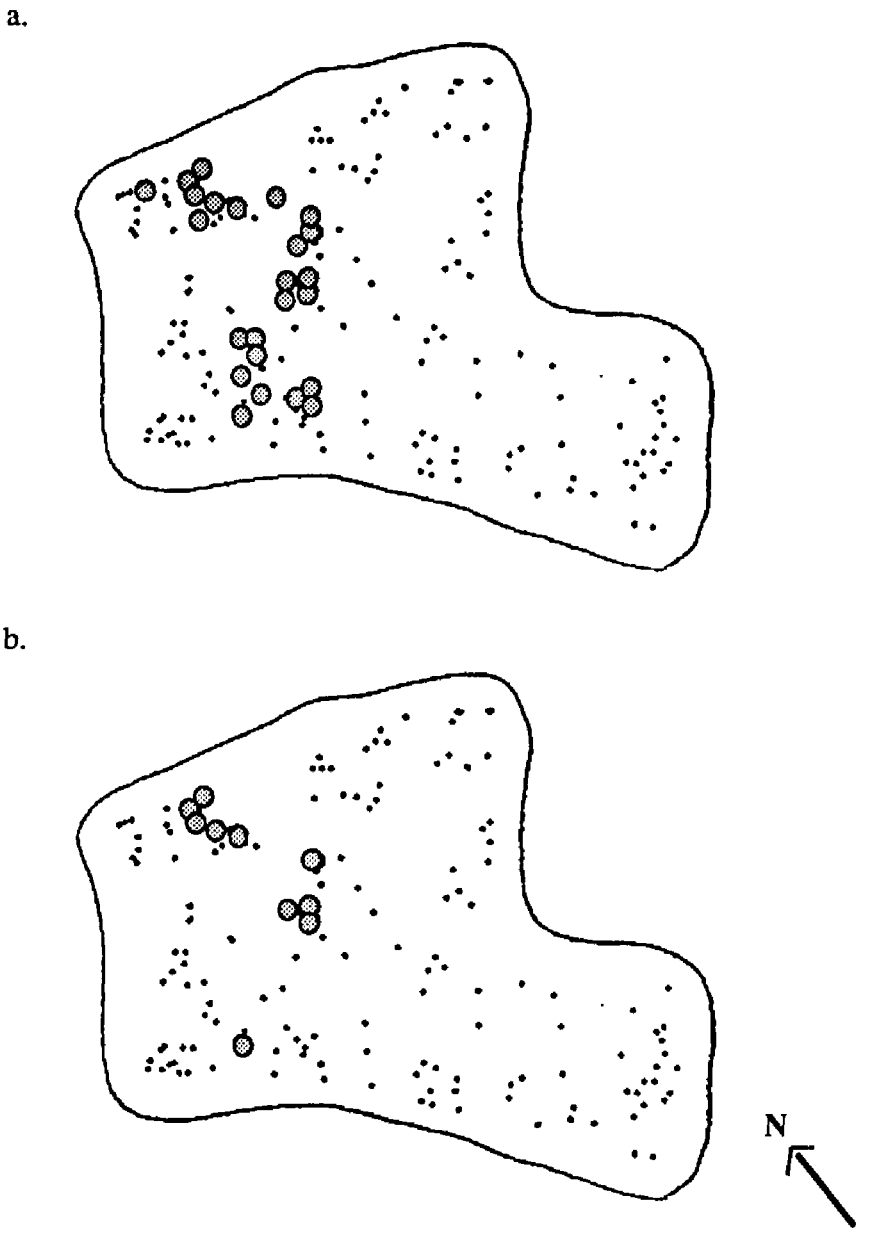


Figure 2.2 -- Spatial distribution of eastern white pines (a.) homozygous for the slow allele of Phosphoglucose Isomerase and (b.) white pines homozygous for Pgi-2(2) for which age is known. The trees, represented by circles, appear to be located in small clusters in the stand. The dots represent other trees in the stand for which age is known, but that are not homozygous for the slow allele.

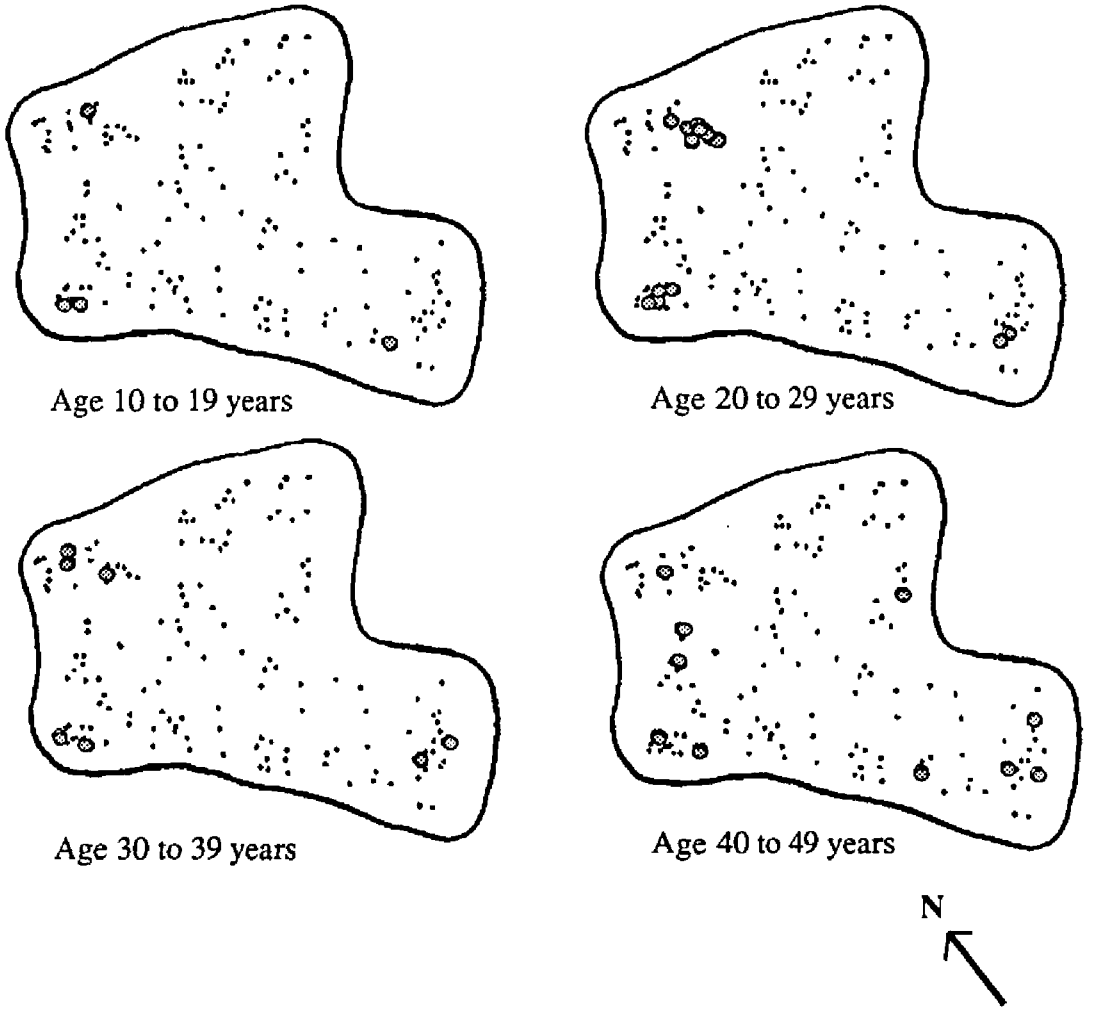


Figure 2.3 -- Spatial distribution of eastern white pines in a two hectare forest stand, showing the location of trees with similar age. The trees in each age category are represented by circles. Other trees are represented by dots.

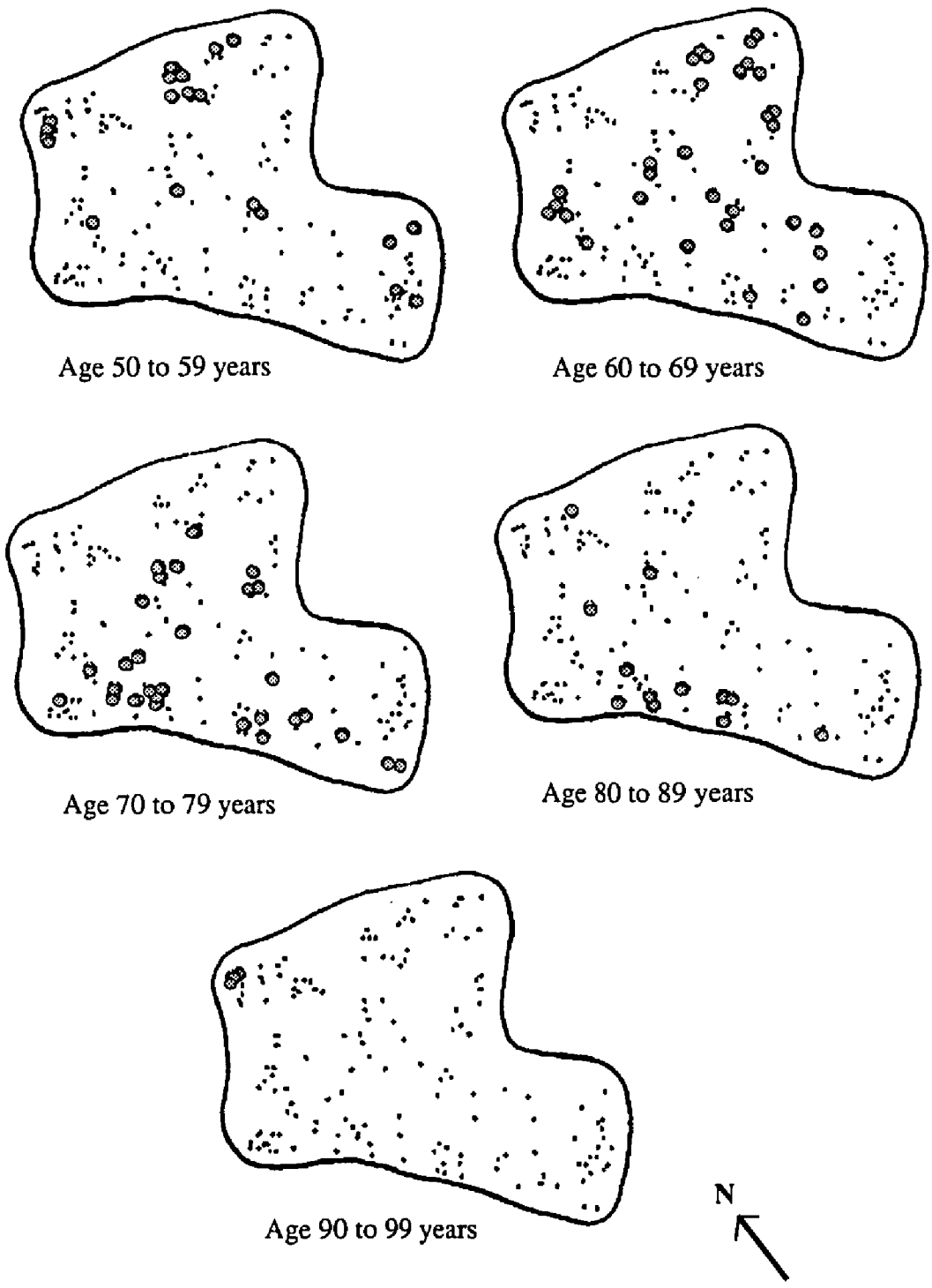


Figure 2.3 (continued) -- Spatial distribution of eastern white pines in a two hectare forest stand, showing the location of trees with similar age.

Appendix A

Slingshots For Collection Of Foliar Tissue From Tall White Pines

Introduction

The foliar tissue samples on small pine branches from the tree crown exposed to sunlight needed for this study were difficult to obtain. Most of the trees were not suitable for safe climbing, they were too tall for cutting instruments and ladders, and the stand was too close to populated areas for safe use of firearms. A customized slingshot was used to remove the desired branches from the white pines. The slingshot was easy to make, inexpensive, provided more support and lesser shocks to the wrist, arm and elbow than the commercial models, and provided sufficient power to break the needed branches.

The Slingshot Design

The basic design and components of the slingshot are described below. The pouch was made from flexible cowhide less than .2 inches thick, 3.5 inches long, and 1.2 inches wide. A hole was punched at each end of the pouch for the elastic bands. The elastic band used was Rubber Amber Latex tubing, approximately (.25x.06) inches. This type performed well both in summer and winter conditions. The tubing was attached to the pouch by forming a loop through pouch holes, and tying the tubing together with carpenter's chalk line. A similar loop was formed on the other end of the tubing, which was slipped over the top end of the slingshot fork. The loop rested in a groove near each tip of the fork. Several pouches of tubing were prepared for easy replacement in the field.

The slingshot fork consisted of a hardwood branch, and was attached to a steel pipe. The steel pipe, at least 10 inches long, served as an arm brace to distribute pressure from the slingshot beyond the wrist, and it added weight to the slingshot to help absorb shock from repeated use during collection trips. Standard

shelf braces were used to join the hardwood branch and steel pipe. The grip was made by wrapping string around the base of the fork and over the braces, screws and clamps. The slingshot was attached to the user's arm with ski straps inserted into sections of bicycle tubes, and a pair of adjustable velcro straps.

Projectiles

Rounded pebbles, approximately 0.75 inches in diameter, were effective for breaking small branches within a thirty yard range. This range was sufficient for all of the trees. Rocks exceeding one inch in diameter were not desirable, because they were too heavy and unsafe. Ocean beach and river beds were a good source of the pebbles.

Usage

When using the slingshot, the user aimed at the silhouette of the branch from which material was needed. Releasing five consecutive pebbles into the silhouette usually yielded at least one satisfactory branch. Aiming at individual twigs twenty-five yards away was not practical for most users. This method was effective

during any season and under almost any weather conditions. Sample collection during cold (below 32^{o}F) winter days was preferred for three reasons: (1) visibility was enhanced in the mixed hardwood and conifer forest; (2) branches were brittle; and (3) cold storage requirements of samples were more easily met. The major obstacles in obtaining desired samples were other branches which prevented free fall of the material. This problem was reduced during windy days, because tree movement helped detached branches fall to the ground.

Using the slingshot described above, samples were collected by one person from at least 20 trees, scattered over one acre area, in less than ninety minutes. This rate of collection included travel between trees, tree identification, and sample packaging. By using the slingshot, the need for firearm permit was avoided, as was the need for securing the firearm and ammunition while not in use.

The slingshot method was effective in all weather conditions. During the study, duplicate samples were collected

during two separate occasions from each of 160 randomly selected white pines, and once from all 511 white pines in the stand. This method was tested for removal of white pine cones from tall trees, and proved useful for situations where only one or two cones were needed from randomly scattered trees.

Precautions

Head and eye protection was used. The rubber tubing, fork, and pouch were frequently inspected for signs of damage due to use and old age. Samples were collected when no other persons were present in the stand.

Appendix B. -- "Smouse Scores" (Smouse et al. 1982) for the genetic loci used to study a two-hectare eastern white pine stand.

LOCUS	NUMBER OF ALLELES	NUMBER OF SCORES	GENOTYPE	SMOUSE SCORE 1ST	SMOUSE SCORE 2ND
LAP	1	0	-	-	-
PGI-1	1	0	-	-	-
PGI-2	2	1	1 1	1.0	-
			1 2	0.5	-
			2 2	0.0	-
PGM-1	1	0	-	-	-
PGM-2	3	2	1 1	1.0	0.0
			1 2	0.5	0.5
			2 2	0.0	1.0
			1 3	0.5	0.0
			2 3	0.0	0.5
			3 3	0.0	0.0
GOT-1	1	0	-	-	-
GOT-2	2	1	1 1	1.0	-
			1 2	0.5	-
			2 2	0.0	-
F-EST-1	2	1	1 1	1.0	-
			1 2	0.5	-
			2 2	0.0	-
F-EST-2	2	1	1 1	1.0	-
			1 2	0.5	-
			2 2	0.0	-
SDH	3	2	1 1	1.0	0.0
			1 2	0.5	0.5
			2 2	0.0	1.0
			1 3	0.5	0.0
			2 3	0.0	0.5
			3 3	0.0	0.0
MDH-1	3	2	1 1	1.0	0.0
			1 2	0.5	0.5
			2 2	0.0	1.0
			1 3	0.5	0.0
			2 3	0.0	0.5
			3 3	0.0	0.0
MDH-2	1	0	-	-	-

Appendix C1. -- Results of soil analysis in two hectare eastern white pine stand.[1]

Sample	pH	Texture	Mg	Ca	K	P	Salts	Nitrogen
A	4.3	SIL	25	100	70	1	<1	0.08
B	4.4	Cl	23	90	90	2	<1	0.11
C	4.4	Cl	18	90	80	2	<1	0.10
D	4.2	L	25	60	60	2	<1	0.16
E	4.5	SIL	10	30	30	2	<1	0.09
F	4.5	L	19	80	50	2	10	0.15
G	4.5	L	21	80	100	4	8	0.18
H	4.7	L	15	60	60	2	<1	0.12
I	4.6	L	11	40	40	3	<1	0.13
J	4.7	L	14	50	40	3	<1	0.17
K	4.8	L	12	40	30	2	<1	0.15
L	4.3	L	20	70	70	2	10	0.17
M	4.3	L	11	60	40	2	<1	0.13

[1] Samples were taken at 13 locations marked in the stand map in Appendix C3.

pH:	5 grams of soil, 5 ml water, electrode method
Salts:	mhos x 10^{-5}
Nitrogen:	Total Nitrogen, percent weight of sample
Mg, Ca, K, P:	Parts Per Million
TEX:	Texture
SIL:	Silt Loam
FSL:	Fine Sandy Loam
CL:	Clay
L:	Loam

Appendix C2. -- Regression of soil quality measurements and the predicted canonical variate scores showing the strength (R) of relationship between individual soil factors and genotype of the trees.

Soil Measurement	R	DF Reg.	DF Res.	F	P
First Canonical Variate Score (PV1)					
pH	0.346	1	11	1.492	0.248
Texture	0.153	2	10	0.119	0.889
Mg	0.065	1	11	0.047	0.832
Ca	0.040	1	11	0.017	0.897
K	0.084	1	11	0.078	0.785
P	0.077	1	11	0.065	0.803
Salts	0.203	1	11	0.471	0.507
Nitrogen	0.042	1	11	0.020	0.891
Second Canonical Variate Score (PV2)					
pH	0.475	1	11	3.209	0.101
Texture	0.302	2	10	0.502	0.619
Mg	0.023	1	11	0.006	0.939
Ca	0.098	1	11	0.106	0.751
K	0.001	1	11	0.000	0.998
P	0.183	1	11	0.381	0.549
Salts	0.032	1	11	0.011	0.918
Nitrogen	0.041	1	11	0.019	0.894

Samples were taken at 13 locations (see Appendix C3).

Salts: mhos x 10^{-5}

Nitrogen: Total Nitrogen, percent weight of sample

Mg, Ca, K, P: Parts Per Million

TEX: Texture

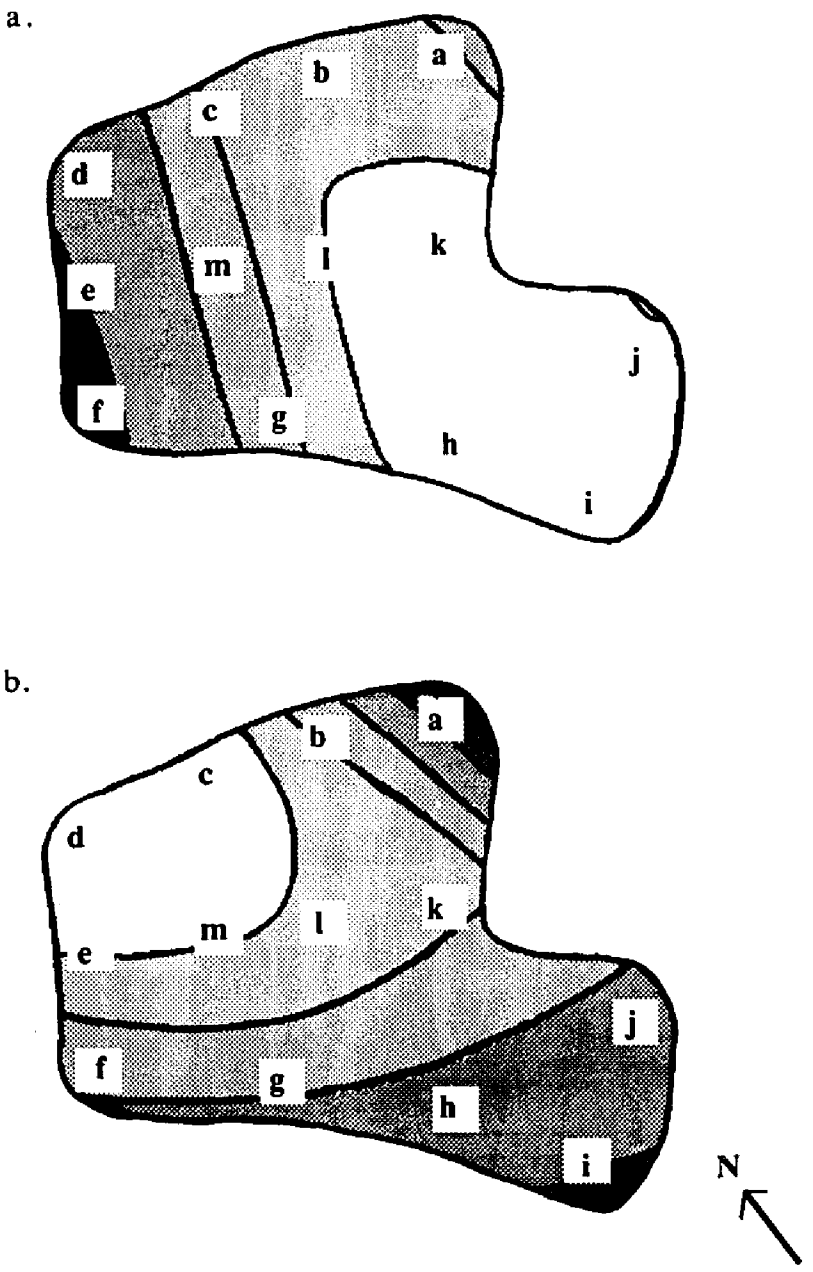
SIL: Silt Loam

FSL: Fine Sandy Loam

CL: Clay

L: Loam

The canonical variate scores represent genotypes of trees in the two hectare eastern white pine stand where the soil samples were taken.



Appendix C 3-- Regions of a two hectare forest stand where soil samples were collected for soil analysis. The letters correspond to sample identification in Appendix C1.

Appendix D1. Canonical correlation analysis of isozyme variables and location of eastern white pines in a two hectare unmanaged forest stand. [1]

K	r_k	r_k^2	F	NDF	DDF	PR>F
1	0.275	0.076	1.701	45	2226.3	0.0027
2	0.201	0.041	1.143	32	1838.1	0.2675

K = number of dimensions
r_k = canonical correlation coefficient (Measures the strength of the relationship between the location and isozyme canonical variate scores.)
r_k^2 = squared canonical correlation coefficient
F = approximate F based on likelihood ratios testing the null hypothesis (Ho) that the canonical correlation in the current row and all that follow are zero.
NDF=numerator degrees of freedom
DDF=denominator degrees of freedom

N = 511 observations

Isozyme loci recoded to continuous scores (Smouse et al.1982): SDH, PGM-2, GOT-2, PGI-2, MDH-1, and FEST-1

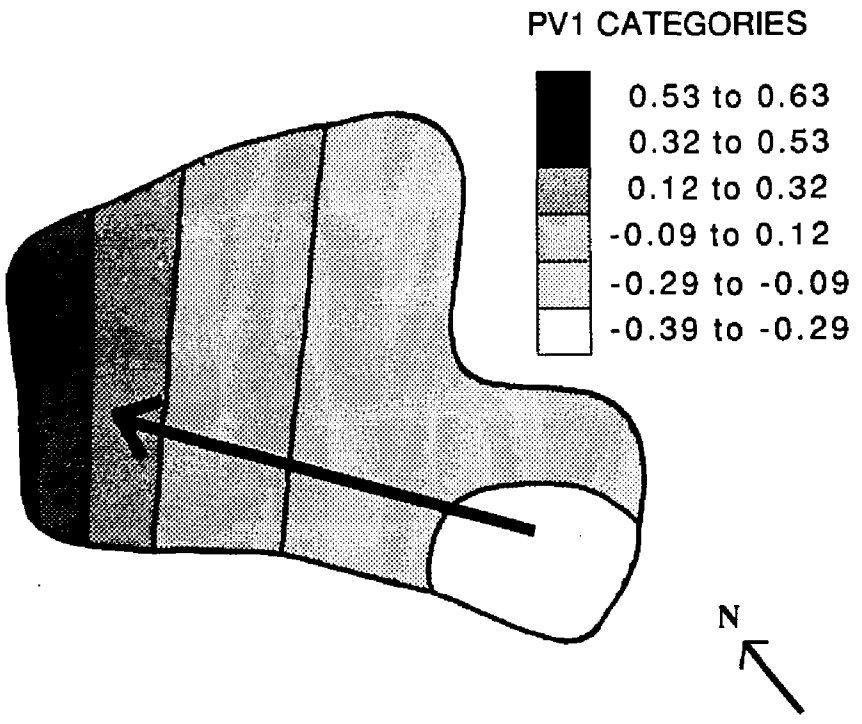
Location variables: used as X (south-west to north-east) and Y (south-east to north-west) coordinates, X-squared, Y-squared and product of X and Y.

[1] The canonical correlation coefficient (r_k) is a measure of the relationship between the location and isozyme canonical variate scores.

Appendix D2. Canonical correlation analysis of isozyme variables and tree location of eastern white pines in a two hectare unmanaged forest stand. [1]

Isozymes[1]	V1	V2	W1	W2
SDH(1)	0.5141	0.1664	0.1413	0.0335
SDH(2)	-0.5916	-0.1867	-0.1626	-0.0376
PGM-2(1)	-0.3065	-0.3225	-0.0842	-0.0649
PGM-2(2)	0.2513	0.2755	0.0691	0.0555
GOT-2(1)	0.4123	0.1622	0.1133	0.0327
PGI-2(1)	-0.4494	0.7925	-0.1235	0.1595
MDH-1(1)	0.1583	0.2811	0.0435	0.0566
MDH-1(2)	-0.2213	-0.2717	-0.0608	-0.0547
FEST-1(1)	0.2754	0.4899	0.0757	0.0986
Tree Location Variables				
X	0.0488	0.0322	0.1775	0.1600
Y	0.2605	-0.0348	0.9479	-0.1730
X2	0.0359	0.0455	0.1305	0.2260
Y2	0.2684	-0.0252	0.9767	-0.1254
XY	0.1685	-0.0019	0.6130	-0.0093

[1] Isozyme variables were represented using Smouse Scores (Smouse et al. 1982). The correlation coefficients show the relationship between the raw isozyme and location variables and the canonical variate scores. High correlations (>.50) are underlined.



Appendix D3.-- Regions of a two hectare forest stand containing eastern white pines with similar predicted canonical variate scores representing allelic composition for SDH, PGM-2, GOT-2, PGI-2, MDH-1, and F-EST-1. The scores are predicted from tree location. The arrows show the direction in which predicted scores increase in the stand. Differences in scores reflect allele frequency changes throughout the stand.

Appendix E1. Mean values of heterozygosity (H), polymorphism (P), and number of alleles per locus (A), obtained from five random samples in each sample size category of the same forest stand.[1]

Number of Trees	H-1	H-2	H-3	P-1	P-2	P-3	A
5	0.154	0.172	0.236	32.72	32.72	32.72	1.378
25	0.186	0.190	0.263	43.63	47.27	47.27	1.712
50	0.186	0.188	0.262	43.63	47.27	47.27	1.712
75	0.183	0.184	0.256	41.81	49.09	49.09	1.748
100	0.186	0.187	0.258	41.81	47.27	47.27	1.730
125	0.185	0.186	0.264	45.45	45.45	47.27	1.748
150	0.187	0.187	0.261	45.45	45.45	49.09	1.766
175	0.188	0.188	0.262	45.45	45.45	49.09	1.766
200	0.186	0.187	0.257	45.45	49.09	54.55	1.820
225	0.186	0.186	0.259	43.63	45.45	49.09	1.766
250	0.188	0.188	0.263	45.45	45.45	52.73	1.802
275	0.186	0.186	0.257	45.45	45.45	54.55	1.820
300	0.188	0.188	0.260	45.45	45.45	52.73	1.802
325	0.187	0.187	0.258	45.45	45.45	54.55	1.820
350	0.186	0.186	0.259	45.45	45.45	52.73	1.802
375	0.187	0.187	0.259	45.45	45.45	54.55	1.820
400	0.188	0.188	0.259	45.45	45.45	52.73	1.802
425	0.187	0.188	0.259	45.45	45.45	54.55	1.820
450	0.187	0.187	0.259	45.45	45.45	54.55	1.820
475	0.187	0.187	0.259	45.45	45.45	54.55	1.820
511*	0.187	0.187	0.259	45.45	45.45	54.55	1.820

[1]The following loci were included: SDH, PGM-2, GOT-2, PGI-2, MDH-1, and FEST-1

H-1 = H (Biased)
H-2 = H (Unbiased)
H-3 = H (Direct)

P-1 = P (0.95 Criterion)
P-2 = P (0.99 Criterion)
P-3 = P (No Criterion)

* One sample with all the trees

Appendix E2. Standard deviation of mean heterozygosity (H), polymorphism (P), and the number of alleles per locus (A), obtained from five random samples in each sample size category of the same forest stand.[1]

Number of Trees	H-1	H-2	H-3	P-1	P-2	P-3	A
5	0.025	0.028	0.029	4.97	4.98	4.98	0.099
25	0.008	0.008	0.010	4.06	4.07	4.07	0.075
50	0.007	0.007	0.012	4.06	4.07	4.07	0.075
75	0.004	0.005	0.007	4.98	4.98	4.98	0.040
100	0.007	0.007	0.008	4.98	4.07	4.07	0.064
125	0.001	0.001	0.002	0.00	0.00	4.07	0.040
150	0.003	0.003	0.003	0.00	0.00	4.98	0.049
175	0.002	0.001	0.003	0.00	0.00	4.98	0.049
200	0.001	0.001	0.002	0.00	4.98	0.00	0.000
225	0.005	0.005	0.005	4.06	0.00	4.98	0.049
250	0.003	0.003	0.003	0.00	0.00	4.07	0.040
275	0.002	0.002	0.005	0.00	0.00	0.00	0.000
300	0.001	0.001	0.002	0.00	0.00	4.07	0.040
325	0.001	0.001	0.003	0.00	0.00	0.00	0.000
350	0.003	0.002	0.003	0.00	0.00	4.07	0.040
375	0.002	0.003	0.002	0.00	0.00	0.00	0.000
400	0.001	0.001	0.002	0.00	0.00	4.07	0.040
425	0.001	0.001	0.001	0.00	0.00	0.00	0.000
450	0.000	0.000	0.001	0.00	0.00	0.00	0.000
475	0.001	0.001	0.002	0.00	0.00	0.00	0.000

[1]The following loci were included: SDH, PGM-2, GOT-2, PGI-2, MDH-1, and FEST-1

H-1 = S.D. of H (Biased)
H-2 = S.D. of H (Unbiased)
H-3 = S.D. of H (Direct)

P-1 = S.D. of P (.95 Criterion)
P-2 = S.D. of P (.99 Criterion)
P-3 = S.D. of P (No Criterion)

Appendix E3. Coefficients of variation for mean heterozygosity (H), polymorphism (P), and the number of alleles per locus (A), obtained from five random samples in each sample size category of the same forest stand.[1]

Number of Trees	H-1	H-2	H-3	P-1	P-2	P-3	A
5	0.163	0.164	0.122	0.15	0.15	0.15	0.072
25	0.045	0.045	0.041	0.09	0.09	0.09	0.044
50	0.037	0.037	0.047	0.09	0.09	0.09	0.044
75	0.025	0.025	0.027	0.12	0.10	0.10	0.023
100	0.039	0.039	0.033	0.12	0.09	0.09	0.037
125	0.007	0.006	0.011	0.00	0.00	0.09	0.023
150	0.017	0.017	0.011	0.00	0.00	0.10	0.028
175	0.009	0.007	0.010	0.00	0.00	0.10	0.028
200	0.007	0.006	0.007	0.00	0.10	0.00	0.000
225	0.026	0.024	0.021	0.09	0.00	0.10	0.028
250	0.014	0.013	0.010	0.00	0.00	0.08	0.022
275	0.010	0.013	0.018	0.00	0.00	0.00	0.000
300	0.004	0.004	0.009	0.00	0.00	0.08	0.022
325	0.008	0.008	0.011	0.00	0.00	0.00	0.000
350	0.014	0.013	0.010	0.00	0.00	0.08	0.022
375	0.011	0.013	0.008	0.00	0.00	0.00	0.000
400	0.007	0.007	0.006	0.00	0.00	0.08	0.022
425	0.006	0.004	0.004	0.00	0.00	0.00	0.000
450	0.000	0.002	0.003	0.00	0.00	0.00	0.000
475	0.004	0.004	0.007	0.00	0.00	0.00	0.000

[1]The following loci were included: SDH, PGM-2, GOT-2, PGI-2, MDH-1, and FEST-1

H-1 = S.D. of H (Biased)
H-2 = S.D. of H (Unbiased)
H-3 = S.D. of H (Direct)

P-1 = S.D. of P (.95 Criterion)
P-2 = S.D. of P (.99 Criterion)
P-3 = S.D. of P (No Criterion)

Appendix F

SimThin

SimThin is a simulation program which combines the friendly interface of the Macintosh computer, graphical data and sound management tools of HyperCard, and the results of the forest genetics research reported in this Dissertation. SimThin has two faces. It is a research tool, allowing researchers to remove individual trees from a stand by clicking on the trees and automatically writing the data of remaining trees to a new data file for statistical data analysis. By comparing the results of analyses for the original data set, and various reduced data sets, researchers can gain a better understanding of whether and how we may be changing the genetic composition of forests with various logging operations. SimThin also is an instructional tool, designed to guide forestry students through a set of preselected thinning choices, and displaying results of each selection. The possible selections include different levels of intensity of tree removal,

house-lot clearing, and clear cutting. Comments and questions built into SimThin help students think about the effects of tree removal on genetic variability of forests. SimThin includes "aerial views" of the forest, maps of trees categorized by age and genotype, and sounds of chainsaws, loggers and falling trees to help bring the forest to the student.

Stand information in SimThin includes tree location and isozyme data for 511 eastern white pines in a 2 hectare forest stand, in Durham, New Hampshire. Also included are height, diameter, age, and crown category information for 158 randomly selected trees in the stand.

SimThin was used several times to help teach forest genetics concepts in a University of New Hampshire genetics course. SimThin was used in this study to generate the age and genotype distribution maps. SimThin generates data sub-samples and sends the data directly to the program BIOSYS for analysis. The BIOSYS FORTRAN subroutines needed to use SimThin this way were converted from the PC-DOS version to run on the Macintosh

computer using the MacFortran compiler.

Future plans for SimThin include adding data sets of other tree species, providing complete access to external data analysis computer programs, and incorporating laser disc technology to include color photographs of the trees and stands. SimThin should help students, researchers and community planners gain a better appreciation of the genetic resources of our forests, and how we may be unknowingly affecting those resources. More importantly, SimThin is designed to do this in just a few minutes, while traditional forest genetics studies can last many years. The methods used in SimThin can be applied in other fields of study, such as wildlife and forest economics.