Genetic diversity in longleaf pine (*Pinus palustris*): influence of historical and prehistorical events

R.C. Schmidtling and V. Hipkins

Abstract: Genetic diversity of allozymes at 24 loci was studied in 23 populations of longleaf pine (*Pinus palustris* Mill.), including three seed orchard populations and an old-growth stand. Overall, the mean number of alleles per polymorphic locus was 2.9, the percentage of polymorphic loci was 92%, and the mean expected heterozygosity was 0.105. These values are comparable with diversity measures found in a similar loblolly pine (*Pinus taeda* L.) study. Diversity measures of the seed orchard sources and the old-growth stand were similar to those in the other natural seed sources. *F* statistics indicate very little inbreeding overall (F_{IS} = –0.002) and low differentiation among populations (F_{ST} = 0.041). All measures of genetic diversity were significantly related to longitude; western sources tended to have more allozyme diversity. Since growth or survival are not related to longitude, and no important climatic variables are related to longitude within the natural range of longleaf, it is proposed that the east–west variation in longleaf pine is a result of migration from a single refugium in the west (south Texas or northeastern Mexico) after the Pleistocene.

Résumé : Les auteurs ont étudié la diversité génétique d'alloenzymes observée pour 24 loci chez 23 populations de pin à longues feuilles (*Pinus palustris* Mill.), dont trois populations de verger à graines et une vieille forêt. Pour l'ensemble de l'étude, le nombre moyen d'allèles par locus polymorphe, le pourcentage de loci polymorphes et l'hétérozygotie moyenne espérée affichaient des valeurs respectives de 2,9, 92% et 0,105. Ces valeurs se comparent aux estimés de diversité obtenus lors d'une étude similaire chez le pin à encens (*Pinus taeda* L.). Les estimés de diversité obtenus pour les populations de verger à graines et pour la vieille forêt étaient similaires à ceux des autres populations naturelles. Les statistiques de *F* indiquent un faible niveau d'endogamie (F_{IS} = –0,002) pour l'ensemble de l'étude, et une faible différenciation de populations (F_{ST} = 0,041). Tous les estimés de diversité génétique démontraient une relation significative avec la longitude; les populations de l'ouest avaient une propension à démontrer un diversité d'alloenzymes accrue. Puisque la croissance et la survie ne démontrent pas de relation significative avec la longitude, et qu'aucune variable climatique importante ne démontre de relation significative avec la longitude au sein de l'aire de répartition de l'espèce, les auteurs proposent que la variation est–ouest notée chez le pin à longues feuilles résulte de la migration à partir d'un seul refuge dans l'ouest (au sud du Texas ou au nord-est du Mexique) après le Pléistocène.

[Traduit par la rédaction]

Introduction

During the late Pleistocene, longleaf pine (*Pinus palustris* Mill.) was undoubtedly absent from the lower Coastal Plain of the southeastern United States, and the area was dominated by a type of boreal forest (Watts 1983). The location of the southern pines during the height of the Wisconsin glaciation is a matter of some speculation (Wells et al. 1991). By historical times, however, longleaf pine had become the predominant species on the Coastal Plain.

Starting before the turn of the century, clear-cutting and high grading followed by sporadic natural regeneration left many areas devoid of forest or sparsely populated by a few genotypes that were inferior to the harvested stands. In the reforestation carried out by the Civilian Conservation Corps in the 1930s, and other individuals and organizations, longleaf pine was largely replaced by slash (*Pinus elliottii* Engelm.) and loblolly pine (*Pinus taeda* L.) because longleaf was difficult to plant and slow in early growth (Croker 1990). As a result, the area of longleaf pine in the southern United States has declined from 12.2 × 10^6 to 3.8 × 10^6 acres (1 acre = 0.405 ha) over the past 30 years alone (Kelly and Bechtold 1990). Longleaf is the most valued of the southern pines (Croker 1990), and there is now renewed interest in restoring longleaf to its historical, commercial, and ecological prominence.

Restoration of longleaf pine will necessarily require a great deal of planting (or perhaps direct seeding) and choosing the proper seed sources will be essential to ensure long-term success. Basic information on population genetics and geographic variation is therefore needed for longleaf pine.

There are relatively few studies of geographic variation or population genetics in longleaf pine. Provenance tests have shown that substantial variation in growth, survival, and disease incidence exists in longleaf pine (Wells and Wakeley 1970). Growth is generally related to latitude or temperature

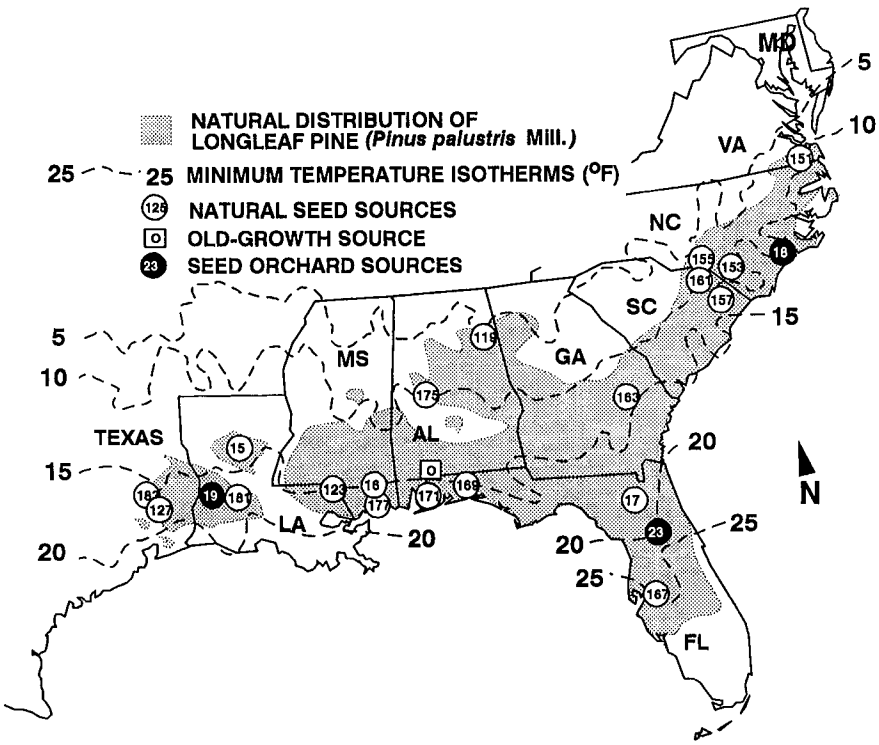
Received February 2, 1998. Accepted June 9, 1998.

R.C. Schmidtling.[1] USDA Forest Service, Southern Institute of Forest Genetics, 23332 Highway 67, Saucier, MS 39574, U.S.A.
V. Hipkins. USDA Forest Service, National Forest Genetics Electrophoresis Laboratory, Camino, CA 95709, U.S.A.

[1] Author to whom all correspondence should be addressed. e-mail schmidtl@datasync.com

© 1998 NRC Canada

Fig. 1. Map of the southeastern United States showing the location of the seed sources and the natural distribution of longleaf pine (adapted from Critchfield and Little 1966). Minimum temperature isotherms are from USDA (1990). Sources with three digits are identical to those from the Southwide Southern Pine Seed Source Study (SSPSSS) (Wells and Wakeley 1970).



at the seed source (Schmidtling and White 1990; Schmidtling and Sluder 1995). Geographic variation in longleaf pine parallels that of other forest tree species; seed sources from warmer climates grow faster than those from colder climates, if these sources are not transferred to very different climates.

Duba (1985) surveyed allozymes in the central part of the longleaf pine distribution. He showed Mendelian inheritance for 19 loci and found a north–south increase in the percentage of loci that were polymorphic.

The fact that there has been little planting of longleaf pine, and therefore very little seed transfer, makes longleaf a much better candidate for studies of geographic variation than loblolly or slash pines. This study explores patterns of genetic variation across the geographic range of longleaf pine, as measured by allozymes. The possible effects of historical and prehistorical events on genetic diversity in the species was also examined.

Materials and methods

Plant materials

Seed were collected from 23 geographic sources of longleaf pine from across the natural range (Fig. 1, Table 1). Sixteen of the sources were of secondary origin, located in provenance test plantings of the Southwide Southern Pine Seed Source Study (SSPSSS), longleaf pine phase (Wells and Wakeley 1970). The longleaf pine phase of the SSPSSS was established using seed collected from natural stands of longleaf pine in the early 1950s. The original specifications for the collections were that seed should be collected from 20 or more trees separated from each other by at least 100 ft (31 m). It is not known how closely the many cooperators who collected the seed adhered to these specifications. Considering the quantity of seed required (500 seedlings per planting for each source, and from 10 to 20 plantings, depending on series), and the usual scarcity of cones on longleaf pine, it is probable that the collections represented more than 20 trees. Many of the collections probably came from harvesting operations, where cones would have been collected from a large number of trees. Since only megagametophytes were analyzed, these samples are representative of the original populations from which they were derived.

Seed from approximately 30 trees from each of the 16 seed sources in the longleaf phase of the study was collected from three different plantings, located in southern Mississippi, southern Alabama, and central Louisiana. Seed was also collected from three natural stands located in areas not represented in the SSPSSS as well as from seed orchards located in central Louisiana, coastal North Carolina, and central Florida. Seed was also collected from a rare, old-growth stand located in southern Alabama (Fig. 1, Table 1).

Enzyme electrophoresis

Isozyme band patterns were investigated using megagametophyte tissue as the enzyme source material from individual tree seed collections. Seeds were sterilized for 5 min in calcium hypochlorite solution and then spread on Petri plates lined with filter paper moistened with 3% hydrogen peroxide. Seeds were placed in a germinator at 20–21°C with a 12 h light : 12 h dark photoperiod, until radicles just emerged from the seed coat, which normally occurred within 3–14 days. Extracts were prepared by crushing an excised megagametophyte in two drops of 0.20 M phosphate buffer (pH 7.5), absorbed onto 2 mm wide paper wicks and frozen at –70°C until electrophoresis. Ten megagametophytes per tree were prepared for isozyme analysis.

Prior to electrophoresis, paper wicks were thawed and inserted into 11% starch gels (Sigma Chemical Co.) that accommodated 48 samples. The preparation and running of the gels are modifications

© 1998 NRC Canada

Table 1. Population origins, number of trees sampled per population, mean number of alleles per polymorphic locus (N_a), percent loci polymorphic (P_l), observed heterozygosity (H_o), and expected heterozygosity (H_e) of 23 populations at 22 loci.

Population			Sample		P_l[a]			
Source	County	State	size	N_a	100%	95%	H_o	H_e
0	Escambia	Alabama	37	2.00 (0.19)[b]	68.2	36.4	0.106 (0.029)	0.103 (0.027)
15	Winn	Louisiana	33	2.27 (0.16)	86.4	40.9	0.134 (0.033)	0.126 (0.029)
16	Harrison	Mississippi	30	2.09 (0.19)	72.7	50.0	0.095 (0.023)	0.106 (0.024)
17	Baker	Florida	15	1.50 (0.14)	40.9	31.8	0.103 (0.034)	0.106 (0.036)
Seed orchard sources[c]								
18	Craven	North Carolina	30	2.05 (0.21)	59.1	45.5	0.096 (0.028)	0.108 (0.029)
19	Rapides	Louisiana	28	2.00 (0.17)	68.2	50.0	0.122 (0.027)	0.129 (0.031)
23	Marion	Florida	16	1.59 (0.17)	45.5	40.9	0.114 (0.038)	0.101 (0.031)
SSPSSS sources								
119	Cleburn	Alabama	31	2.09 (0.19)	81.8	40.9	0.101 (0.021)	0.112 (0.025)
123	Washington	Louisiana	30	1.73 (0.18)	50.0	40.9	0.082 (0.031)	0.081 (0.027)
127	Polk	Texas[d]	30	2.09 (0.15)	77.3	50.0	0.135 (0.031)	0.137 (0.032)
151	Nansemond	Virginia	25	1.59 (0.16)	45.5	22.7	0.069 (0.025)	0.064 (0.023)
153	Bladen	North Carolina	14	1.64 (0.15)	50.0	27.3	0.088 (0.031)	0.089 (0.029)
155	Richmond	North Carolina[e]	17	1.68 (0.18)	50.0	31.8	0.075 (0.030)	0.075 (0.027)
157	Florence	South Carolina	27	1.77 (0.19)	54.5	22.7	0.072 (0.028)	0.081 (0.028)
161	Chesterfield	South Carolina[e]	29	2.05 (0.19)	68.2	40.9	0.093 (0.030)	0.102 (0.030)
163	Treutlen	Georgia	30	1.91 (0.19)	63.6	36.4	0.085 (0.024)	0.092 (0.028)
167	Hillsborough	Florida	24	1.86 (0.18)	59.1	36.4	0.101 (0.032)	0.093 (0.026)
169	Okaloosa	Florida[e]	20	1.82 (0.18)	59.1	36.4	0.105 (0.035)	0.092 (0.028)
171	Baldwin	Alabama	30	2.23 (0.19)	68.2	45.5	0.113 (0.028)	0.109 (0.028)
175	Perry	Alabama	31	2.00 (0.19)	63.6	40.9	0.104 (0.029)	0.113 (0.029)
177	Harrison	Mississippi	30	1.95 (0.19)	63.6	40.9	0.103 (0.030)	0.119 (0.032)
181	Rapides	Louisiana	31	2.27 (0.24)	68.2	54.6	0.138 (0.033)	0.136 (0.032)
183	Polk	Texas	30	2.05 (0.21)	63.6	45.5	0.144 (0.038)	0.147 (0.036)
Mean				1.92	62.1	39.5	0.103	0.105

Note: Enzyme loci AK-1 and AK-2 were not polymorphic and are excluded from the data. SSPSSS source numbers are as in Wells and Wakeley (1970).

[a]At the 100% criterion, a locus is considered polymorphic if more than one allele was detected. At the 95% criterion, the most common allele must have a frequency of 0.95 or less for a locus to be considered polymorphic.

[b]Standard errors of the estimates are given in parentheses.

[c]Located in these counties as well as adjacent counties.

[d]This collection included trees from the two adjoining counties to the east.

[e]These collections were made on deep sand sites.

of Adams et al. (1990) and Conkle et al. (1982). A total of 618 trees were genotyped (using 10 megagametophytes per tree) at 24 isozyme loci using three buffer systems. Buffer system "LB" (gel and tray buffer "A" of Adams et al.(1990)) was used to resolve enzyme systems alcohol dehydrogenase (ADH), phosphoglucose isomerase (PGI), fluorescent esterase (FEST), malic enzyme (ME), aconitase (ACO), leucine aminopeptidase (LAP), and phosphoglucomutase (PGM). Buffer system "SB" (a modification of gel and tray buffer "B" of Adams et al. (1990), where the electrode buffer was pH 8.0), was used to resolve enzyme systems triosephosphate isomerase (TPI), glycerate-2-dehydrogenase (GLYDH), glucose-6-phosphate dehydrogenase (G6PD), 6-phosphogluconate dehydrogenase (6PGD-1), and glutamic oxaloacetate transaminase (GOT). Buffer system "MC8"(a modification of gel and tray buffer "C" of Adams et al. (1990), where the stock solution was adjusted to pH 8.0), was used to resolve adenylate kinase (AK), isocitrate dehydrogenase (IDH), 6-phosphogluconate dehydrogenase (6PGD-2), and malate dehydrogenase (MDH).

Running conditions and stain recipes follow Adams et al. (1990) and Conkle et al. (1982). After the dye marker migrated 8 cm, gels were cut horizontally into four to seven slices, stained, and scored.

Growth data

Variation at allozyme loci is considered nonadaptive, but such variation may reflect variation at other loci. Therefore, it may be useful to compare allozyme variation with the adaptive traits height and plot volume. Since 16 of our sources are from the SSPSSS, 25-year measurements are available (Schmidtling and White 1990) for comparison with diversity indices. A seed-source transfer function was previously developed using data from 17 plantings located across the natural range of longleaf pine (Schmidtling 1997). As is usual for large provenance tests, overall site quality varied widely among plantings, and there was a substantial planting location × seed source location interaction. The simple means calculated for seed sources across all planting sites contain a great deal of extraneous variation. To adjust for differences in site quality (site index), plot means were first standardized by expressing growth as a percent deviation from the local source.

The best independent variables related to climate for predicting growth and plot volume of seed sources was average yearly minimum temperature and its square (Schmidtling 1994). Average yearly minimum temperature has been used by horticulturists to delineate plant hardiness zones (USDA 1990). The data from the 17 plantings was combined by expressing minimum temperatures

© 1998 NRC Canada

Table 2. Summary of F-statistics at 22 loci for 20 populations of longleaf pine.

Locus	F_{IS}	F_{IT}	F_{ST}
PGM-1	0.110	0.133	0.026
PGM-2	0.069	0.102	0.036
LAP-1	0.075	0.095	0.021
LAP-2	0.003	0.018	0.015
ACO-1	0.018	0.053	0.036
ME7	0.087	0.147	0.067
FEST1	0.183	0.205	0.027
PGI-2	–0.054	–0.010	0.042
ADH	–0.015	–0.001	0.013
GOT-2	–0.057	0.021	0.074
GOT-3	–0.067	–0.021	0.044
6PGD1	–0.032	0.014	0.044
G6PD1	0.035	0.062	0.029
G6PD2	–0.017	–0.004	0.013
GLYDH	–0.019	–0.002	0.017
TPI-2	–0.027	–0.003	0.023
MDH-1	–0.053	–0.007	0.043
MDH-2	0.079	0.123	0.048
MDH-3	–0.043	–0.010	0.032
MDH-4	–0.052	0.005	0.054
6PGD2	–0.025	–0.010	0.014
IDH-1	–0.024	–0.003	0.020
Mean	–0.002	0.039	0.041

Note: The three orchard sources are excluded from the calculations.

as a difference between minimum temperatures at the planting site and minimum temperatures at the seed source, producing a seed source transfer function. The data were then fitted to a polynomial model using average yearly minimum temperature and its square as the independent variables (Schmidtling 1997). The model explained 31% of the variation in height growth.

Only the mean deviations of seed sources from the expected values of the previously developed transfer function (Schmidtling 1997) were compared with various measures of diversity from the allozyme data. The model was used here to remove seed source × planting location interactions and give a relatively clear estimate of the growth and survival potential of the seed sources across the natural range of the species, compared with an expected value.

Statistical analysis

Allozyme data were used to provide several estimates of genetic variation using BIOSYS I (Swofford and Selander 1989): mean number of alleles per polymorphic loci (N_a), percent loci polymorphic (P_l, 100% and 95% criteria), observed heterozygosity (H_o), and expected heterozygosity (H_e). An additional measure of diversity computed was (N_r), the number of rare alleles per tree (a rare allele being defined here as one that occurs at a frequency of 0.05 or less in the overall population).

BIOSYS also provided measures of genetic distance (Cavalli-Sforza and Edwards 1967) and F statistics F_{IS}, F_{IT}, and F_{ST} (Wright 1965, Nei 1977). Gene flow was estimated using Wright's (1931) formula:

$$Nm = \frac{(1 - F_{ST})}{4F_{ST}}$$

where N is the effective population size of the recipient population and m is the rate of gene flow. Nm estimates the number of migrants per generation. F_{ST} is considered to be equivalent to G_{ST} (Wright 1978).

Diploid genotypes were also transformed for multivariate analysis using the technique of Smouse and Williams (1982). For each allele at a locus (minus one), the value of 0.5 was assigned when the allele was present and 0 when the allele was absent. The score when the allele at the locus is in the homozygous state would be 0.5 + 0.5 = 1.0, and when it is in the heterozygous state, 0.5 + 0.0 = 0.5. For individuals without the allele the score would be 0. This is equivalent to a measure of the amount of an allele in each individual. Data sets with more than 10 alleles can be assumed to have a normal distribution (Smouse and Williams 1982). Transformed data were analyzed using SAS (SAS Institute Inc. 1990) multivariate analysis of variance and canonical discriminant analysis.

Results and discussion

Population structure

For the entire population, 92% of the loci were polymorphic, and the 24 loci averaged 2.92 alleles per locus. Two of the 24 loci examined, AK-1 and AK-2, were not polymorphic, and were dropped from the analyses. For the 23 populations, N_a averaged 1.92, P_l (95% criterion) was 39.5%, and H_e was 0.105 (Table 1). These values were somewhat lower than those found for the related southern pines (Critchfield and Little 1966), loblolly pine (Williams et al. 1995, Conkle 1981), and shortleaf pine (*Pinus echinata* Mill.) (Raja et al. 1997). In a study that sampled the central part of the longleaf pine distribution, however, Duba (1985) found comparable population means for N_a (1.78) and P_l (39.9%) although his average value for H_o (0.150) was higher. (These values were computed from Duba's Table 6; H_e was not noted.)

Diversity measures of the seed orchard sources were similar to those in the provenance test sources and the old-growth source (Table 1). Neither past logging practices nor tree improvement appears to have affected the genetic resource of longleaf pine significantly.

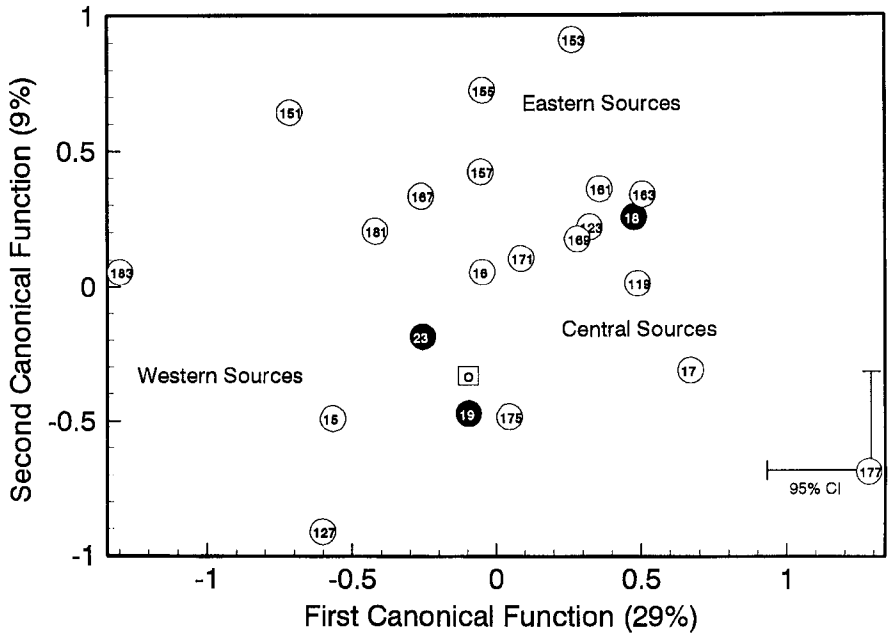
The data indicate that the populations are very close to Hardy–Weinburg equilibrium (Table 1). The exact probability chi-square test for deviation from Hardy–Weinburg equilibrium was significant in only 7 of 313 possible comparisons (at six different loci). This is less than half that expected by chance alone.

Wright's fixation index indicates that only 4.1% of the genetic variation is due to differences among populations (F_{ST} = 0.041; Table 2). This is considerably less than one estimate for loblolly pine (7.8%, Edwards and Hamrick 1995) but is intermediate to two estimates for shortleaf pine (2.6%, Edwards and Hamrick 1995; 8.9%, Raja et al. 1997).

The value for F_{IS} (–0.002) showed a very slight excess of heterozygotes within populations, indicating very little inbreeding. There is a close agreement between H_e and H_o across all populations. The age of the parent trees is certainly a factor. Heterozygosity and genetic diversity increases from embryo to seedling to mature tree, probably because of excess mortality in inbred individuals under the increased stress of competition (Ledig 1986). The SSPSSS trees were just over 40 years of age when seed were collected. The natural stands were probably 30–60 years old, the orchard ortets averaged 80–100 years old, and trees in the old-growth stand ranged from 100 to 350 years old.

© 1998 NRC Canada

Fig. 2. Plot of the first two canonical functions from the multivariate analysis of the transformed allozyme frequency data. Analysis was performed using SAS Institute Inc. (1990).



Most inbred individuals were probably eliminated from all populations when sampled.

Including the three orchard sources in the calculations changed Wright's F statistics very little. With these three sources included, F_{ST} was slightly lower, 0.038; F_{IS} was the same, –0.002; and F_{IT} was also lower, 0.036. This supports the observation that the orchard populations differ only slightly from the natural populations in allozyme diversity.

Gene flow using Wright's formula is Nm = 5.85 migrants per generation in this study, which is greater than the 4 migrants per generation that Wright (1931) considered great enough to prevent differentiation due to drift. This is not surprising considering extensive long-distance pollen flow that has been found in studies of pollen contamination in southern pine seed orchards (Friedman and Adams 1985). Limited drift may be occurring, however, since geographic distance is significantly, although weakly, correlated with genetic distance (r = 0.31, p < 0.01).

In spite of the small F_{ST} value, multivariate analysis of the transformed data showed significant differences among the natural populations in all the standard tests of significance (P < 0.001; SAS Institute Inc. 1990). The populations are also differentiated in a canonical discriminant analysis. The first two canonical functions account for 38% of the variation among populations and appear to be primarily related to source longitude. In a plot of the first and second canonical functions (Fig. 2), western sources are aggregated in the lower left quadrant of the plot, eastern sources in the upper section, and central sources in between.

A separate canonical analysis of only the natural sources verified the expectation that including the seed orchard sources (sources 18, 19, 23) in the analysis did not change the spatial relationships among the natural sources in the scatter plot of the first two vectors. The seed orchard sources as well as the old-growth source (source 0) do not appear to differ substantially from nearby sources in the combined canonical analysis (Fig. 2). They fall within the limits of variation for their respective geographic areas.

Source 177 (southern Mississippi) appears in an anomalous position in Fig. 2, in the lower right-hand corner (although adjacent to the other central sources). The genetic diversity (N_a, P_l, and H_e; Table 1) of this source is very close to the mean; the only qualitative difference between this source and the others is the presence of a private allele at PGM-2 (Table 3). This allele is at a low frequency and would not have much effect on the canonical analysis. The recently collected source from the same county but not the same stand (source 16) plots near the other sources from the central part of the distribution (Fig. 2). Source 177 appeared to be morphologically "typical" longleaf when recently examined in one of the SSPSSS field plantings where seed were collected. Hybridization with loblolly or slash pines cannot be ruled out.

East–west variation

An east–west trend in variation is evident in correlations of allozymic diversity measures with geographic variables (Table 4). All of the diversity parameters are correlated significantly with longitude; diversity decreased from west to east for all parameters studied. P_l was also correlated significantly with latitude, showing a tendency for diversity to decrease from south to north. Duba (1985) also found a decrease in P_l from south to north. An east–west component to variation was apparent in his data but was not statistically significant, probably because his study included only the central part of the range.

East–west variation in longleaf pine is in contrast with those of adaptive traits. The only climatic variable that varies from east to west is rainfall, and rainfall only becomes critical at the far western edge of the natural distribution. The adaptive traits of growth and survival have a strong north–south component of variation (Schmidtling and Sluder

© 1998 NRC Canada

Table 3. Frequency of rare alleles in longleaf seed sources arranged geographically from west to east.

		Seed source										
Locus	Allele	183	127	19	15	181	123	16	177	171	175	0
PGM-1	2			0.036		0.016		0.033	0.036	0.017		
PGM-1	3	0.033	0.017		0.015	0.016	0.067	0.017	0.036	0.017		0.014
PGM-2	2		0.033		0.015	0.032		0.017	0.036			
PGM-2	3		0.017			0.016						
PGM-2	4					0.016						
PGM-2	5											
PGM-2	6								0.054			
LAP-1	2	0.050	0.050	0.036	0.045	0.048	0.017	0.033	0.018	0.050	0.016	0.041
LAP-1	3	0.033	0.033	0.018	0.015	0.016		0.017		0.017	0.081	
LAP-1	4											
LAP-2	2		0.033	0.054	0.030	0.032			0.018	0.017		0.014
LAP-2	3	0.033		0.018	0.030	0.032		0.017	0.018			0.027
LAP-2	4					0.016				0.017		
ME7	4	0.017			0.015					0.017		0.041
FEST-1	3			0.018			0.033	0.017	0.058	0.033	0.033	
PGI-2	4	0.050		0.036	0.061	0.065	0.017	0.017	0.018	0.017		0.027
PGI-2	5	0.050				0.032		0.017	0.054		0.017	0.027
ADH	4				0.015							0.014
GOT-2	2	0.117	0.067	0.036	0.045					0.017	0.017	
GOT-2	3		0.017		0.015					0.033		0.014
GOT-2	4	0.117	0.050	0.018	0.015	0.016		0.017		0.033		
GOT-2	5	0.017								0.017		
GOT-3	3	0.017			0.015		0.017				0.017	0.014
GOT-3	4					0.016						
6PGD-1	3	0.067	0.033		0.045	0.081	0.050	0.017		0.086	0.016	0.041
G6PD-1	2	0.150	0.033	0.036	0.113	0.065		0.050	0.042	0.020	0.033	0.056
G6PD-2	2		0.017		0.015					0.018		
GLYDH	2				0.015							
TPI-2	2							0.017	0.019			
MDH-1	2		0.067		0.015		0.050					
MDH-2	2					0.032		0.033	0.107		0.016	0.027
MDH-4	2		0.050	0.054		0.113		0.050	0.019	0.067	0.032	
MDH-4	4		0.017			0.016						
6PGD-2	2	0.017		0.036	0.030					0.017	0.017	
6PGD-2	3	0.017	0.017									
IDH-1	2			0.018							0.016	
IDH-1	3										0.016	
No. of loci		15	16	13	19	19	7	15	14	18	13	13
Alleles/tree		1.533	1.067	0.821	1.121	1.355	0.500	0.733	1.036	1.000	0.677	0.703

1995), which is most closely related to variation in mean annual minimum temperature. Once differences in minimum temperature have been accounted for, there is no east–west variation in growth and survival in longleaf pine (Schmidtling and White 1990).

A plot of H_e versus longitude of the seed source shows an obvious decrease in variation from west to east (Fig. 3).

The old-growth source and two of the orchard sources fit the east–west model of variation very closely. One of the orchard sources (18) seems to have more variation than expected. The orchard sources consist of individual selections scattered over a wide area, rather than from specific stands. These are old, dominant trees selected for size and form; they would presumably be well adapted to their ecological circumstance. The trees from source 18 are from three adjacent counties in the lower Coastal Plain of North Carolina (Fig. 1). Although the landform appears to be flat and uniform, a difference in elevation of less than 1 m can determine the difference between a wet site and a mesic site. There may be some differential selection for wet-site and mesic-site genotypes that favors greater variability in this population than in the other orchard populations.

At the other extreme, source 123 has considerably less variation than expected (Fig. 3). This is a source identified by Wells and Wakeley (1970) as performing below expectations in growth and survival in the SSPSSS, considering its geographic origin. When variation in this source is viewed in the context of Fig. 3 rather than simply making comparisons in Table 1, it appears likely that this source is genetically depauperate, as Wells and Wakeley proposed. The

© 1998 NRC Canada

		Seed source											
Locus	Allele	169	119	163	167	17	23	161	157	155	153	18	151
PGM-1	2	0.025	0.065									0.033	
PGM-1	3									0.029		0.033	
PGM-2	2	0.025	0.033	0.067				0.017					
PGM-2	3	0.025		0.017							0.071	0.017	
PGM-2	4												
PGM-2	5							0.017	0.037		0.036	0.017	
PGM-2	6												
LAP-1	2		0.016		0.021			0.069	0.056	0.029			0.040
LAP-1	3												
LAP-1	4				0.021								
LAP-2	2	0.025	0.032		0.021				0.019	0.059		0.033	
LAP-2	3			0.017	0.021	0.067		0.017	0.019		0.036	0.050	0.040
LAP-2	4							0.034			0.036		
ME7	4									0.029		0.034	
FEST-1	3		0.052		0.021			0.052				0.034	
PGI-2	4	0.025					0.031	0.069	0.019	0.029		0.033	0.060
PGI-2	5	0.050		0.050	0.021		0.031	0.086	0.037	0.059	0.071	0.017	
ADH	4												
GOT-2	2		0.017	0.017									
GOT-2	3	0.025											
GOT-2	4		0.017						0.019			0.017	0.020
GOT-2	5												
GOT-3	3		0.017										
GOT-3	4												
6PGD-1	3		0.016	0.033	0.042	0.071		0.017	0.130			0.033	0.020
G6PD-1	2		0.037	0.033	0.071		0.100	0.093	0.038	0.029	0.071		0.063
G6PD-2	2			0.017					0.019				
GLYDH	2				0.021								
TPI-2	2							0.034				0.017	
MDH-1	2		0.016										
MDH-2	2		0.032					0.017					
MDH-4	2	0.025	0.016	0.067	0.021	0.033		0.017	0.019	0.059		0.050	0.040
MDH-4	4			0.017									
6PGD-2	2	0.025	0.016			0.033		0.017					
6PGD-2	3			0.017						0.029			
IDH-1	2		0.016				0.033						
IDH-1	3		0.016		0.020								
No. of loci		9	16	11	11	4	4	14	11	9	6	14	7
Alleles/tree		0.500	0.807	0.700	0.560	0.400	0.375	0.700	0.560	0.706	0.643	0.807	0.560

reduced variability is undoubtedly a result of genetic sampling. According to Wells and Wakeley, this collection was made in a stand of 30+ trees that probably seeded in from 4 trees left after a clearcut made around 1905. Although much of the pollen would have come from nearby stands, inbreeding probably occurred. Inbreeding is not evident in the current sample as H_o and H_e are identical (Table 1). Inbred individuals may have been largely eliminated from the stand by the time seed were collected for this study, i.e., at age 40 or more years.

Hybridization could explain the higher genetic variability in the western part of the longleaf pine distribution, if more hybridization occurs in the west than in the east. Longleaf pine is known to hybridize with loblolly pine, and the natural hybrid was named "Sonderregger" when it was first described by Chapman (1922). The hybrid is common, especially in disturbed areas. Because of the distinctive "grass" stage in longleaf pine, F_1 hybrids are easy to identify in nursery beds of longleaf pine. Any putative longleaf pine seedling starting height growth in the first year is undoubtedly a hybrid. Even the backcross of the hybrid to longleaf pine can be distinguished on the basis of height growth in a nursery bed (Lott et al. 1996). Longleaf pine is also characterized by resistance to fusiform rust disease (*Cronartium quercuum* f.sp. *fusiforme*) and susceptibility to brown-spot needle blight (*Mycosphaerella dearnessii* Barr). Loblolly pine is the opposite, i.e., it is susceptible to fusiform rust and resistant to brown spot needle blight. The hybrid is intermediate (Lott et al. 1996). Thus, if a greater incidence of hybridization is the basis for the greater genetic variation in the

© 1998 NRC Canada

Fig. 3. Regression of expected heterozygosity (H_e) with longitude of the seed source for each population.

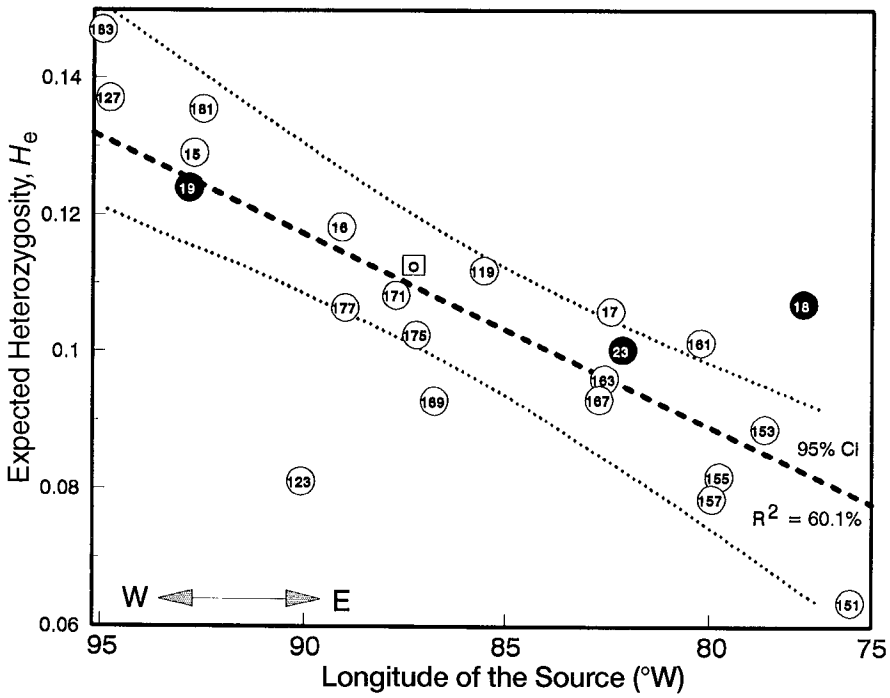


Table 4. Correlations of diversity parameters with latitude, longitude, and average yearly minimum temperature at the seed source.

	Longitude	Latitude	Minimum temperature
N_a[a]	0.604**	–0.125	0.061
H_e[b]	0.775**	–0.393	0.303
P_l[c]	0.718**	–0.544**	0.431*
N_r[d]	0.549**	0.047	0.040

[a]N_a, Mean number of alleles per locus.
[b]H_e, Expected heterozygosity.
[c]P_l, Percent polymorphic loci (95% criterion).
[d]N_r, Number of rare alleles per tree.
*Significant at the 0.05 level.
**Significant at the 0.01 level.

western sources, these sources should be less susceptible to brown-spot needle blight, more susceptible to fusiform rust, and start height growth sooner than the eastern sources.

A complete appraisal of growth and disease incidence for the first 10 years for the SSPSSS sources of this study has been published by Wells and Wakeley (1970). They found an unusually large number of Sondereggers in the Virginia source (151). These hybrids were culled in the nursery from all seed sources, but some remained in the plantings; they were subsequently excluded from the data and removed from the plantings in a thinning operation at age 15. The thinning at age 15 was mainly mechanical, but obvious hybrids and deformed trees were removed. At age 10, sources from Alabama had the greatest number of trees starting height growth. Although fusiform rust disease was not common in the plantings, where it occurred, Wells and Wakeley found that the sources in the central part of the range (Alabama and Georgia) were the most heavily infected. They also found that brown-spot infection was greatest in the western sources. Hybridization and introgression certainly occurs in longleaf pine, but the growth and disease incidence data from the SSPSSS does not support the hypothesis of greater hybridization in the western sources.

Huneycutt and Askew (1989) screened 22 loci in shortleaf and loblolly pine and found only one, the IDH locus, which appeared useful in indicating hybridization with loblolly pine. We have allozyme data in loblolly pine on 14 of the loci screened in the present experiment (R.C. Scmidtling and V.D. Hipkins, in preparation). Unfortunately, the alleles that are common in longleaf are generally common in loblolly. At the IDH locus, the common allele in loblolly migrates at the same rate as the common allele in longleaf and is probably homologous. The two alleles that are rare in longleaf (Table 3) are also rare in loblolly.

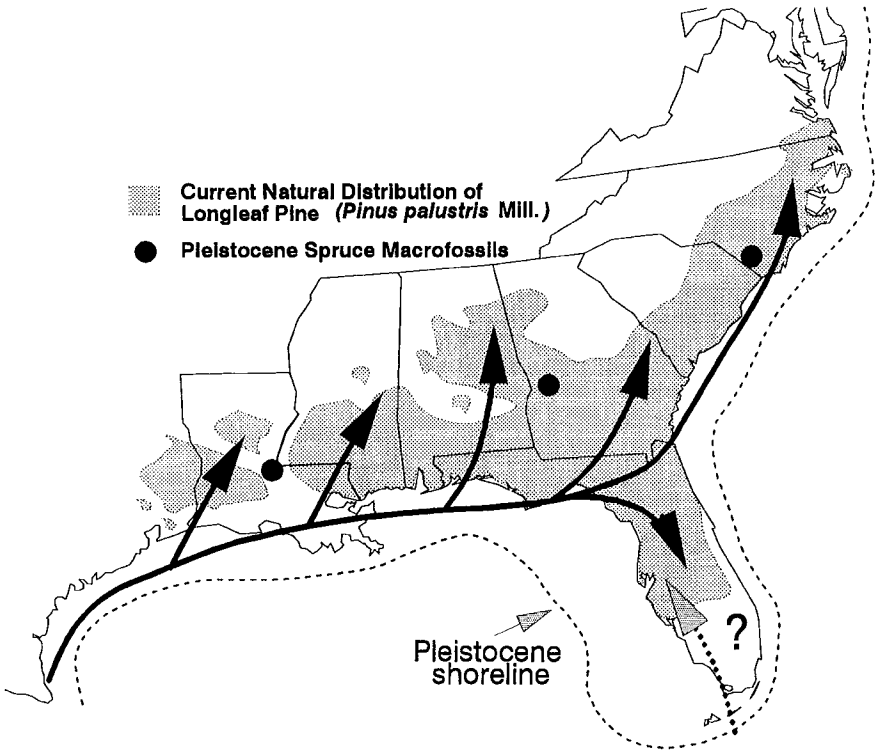
The 6PGD-1 locus might be useful in identifying hybrids. Loblolly has seven alleles at this locus (most of them rare) whereas we found only three in longleaf. The one rare allele in longleaf at this locus (No.3) appears to be homologous to an allele in loblolly that is common, occurring at a frequency of about 0.5. If we assume that the presence of this allele indicates hybridization with loblolly pine, we see no evidence of greater hybridization in the western part of the longleaf distribution; this particular allele is scattered at random from west to east (Table 3). Similarly, allele 2 of LAP-2 (Table 3) and allele 4 of ACO-1 may indicate hybridization but are not concentrated in the western sources.

Post-Pleistocene migration

The east–west trend in variability in longleaf pine becomes less problematic if events during and after the Pleistocene are considered. Very little is known about the location of the southern pines during the Wisconsin glaciation because of the lack of macrofossils. Palynological

© 1998 NRC Canada

Fig. 4. Proposed migration route for longleaf pine from a southern Texas – northeastern Mexico refugium at the close of the Pleistocene. The Pleistocene shoreline assumes a drop in sea level of 100 m. Pleistocene spruce macrofossils are from Watts (1983).



records are not conclusive, because of the difficulty in identifying pine pollen to the species level. Macrofossils of spruce (*Picea* sp.) dating from the Pleistocene have been found within the current range of longleaf pine (Fig. 4) indicating that the climate was boreal (Watts 1983). It is reasonable therefore to assume that longleaf pine was situated south of its present location during the Pleistocene.

Wells et al. (1991) proposed two refugia for the closely related loblolly pine during the Pleistocene: south Florida – Caribbean and southern Texas – northern Mexico. The more austral slash pine may have resided in only one of these refugia: south Florida – Caribbean. Taxonomically, slash pine has obvious affinities with other Caribbean hard pines, and before 1954 (Little and Dorman 1954), slash pine and *Pinus caribaea* (Morelet) from Cuba and the Bahamas were not considered separate species. In addition, slash pine does not occur naturally west of the pineless Mississippi River valley (Critchfield and Little 1966), although it grows very well in western Louisiana and eastern Texas in plantations. Circumstantial evidence favors the hypothesis of a single refugium for slash pine in southern Florida with a migration to the north and west at the close of the Pleistocene.

In contrast to these other species of southern pines, our data favors the hypothesis that longleaf pine was located in a single refugium in southern Texas or northern Mexico and migrated northward and eastward at the close of the Pleistocene (Fig. 4). Our data appears to fit the criteria of Wheeler and Guries (1982) for expansion from a single refugium in the west. Paraphrasing from their paper the populations at the extremity, i.e., in the east, should possess: (*i*) relatively close genetic affinity for one another (Fig. 2); (*ii*) a reduced frequency and distribution of rare alleles resulting from rapidly migrating populations affected by stochastic events (Table 3); and (*iii*) a reduced level of genetic variability, resulting from these same stochastic events (Table 4, Fig. 3).

Table 5. Growth of SSPSSS Seed sources after 25 years in the field compared with H_e.

Seed source	Deviation from model (%)		Expected heterozygosity	
	Height	Volume	H_e	Residuals[a]
183	–0.38	7.26	0.147	0.016
127	0.54	9.67	0.137	0.006
181	0.30	–2.15	0.136	0.012
123	–3.93	–13.38	0.081	–0.037
177	2.56	47.82	0.119	0.004
171	2.07	4.67	0.109	–0.002
175	4.62	11.01	0.113	0.003
169	–2.03	–3.73	0.092	–0.016
119	0.73	–1.74	0.112	0.007
163	3.98	3.29	0.092	–0.004
167	–1.91	–24.77	0.093	–0.003
161	1.61	11.27	0.102	0.013
157	–0.51	–1.32	0.081	–0.007
155	4.97	–0.74	0.075	–0.011
153	0.13	–8.47	0.089	0.004
151	–4.21	–7.57	0.064	–0.016

Note: Height and volume are expressed as percent deviation from a minimum-temperature transfer function (Schmidtling 1994). Seed source numbers are given in Table 1.

[a]Deviations from the east–west regression of Fig. 3.

Our results are remarkably similar to the results in a recent study of genetic variation in *Pinus kesiya* (Royle ex

© 1998 NRC Canada

Gordon) by Myburg and Harris (1997). They found a distinct decrease in allozymic variation from southeast to northwest in Southeast Asia. They also concluded that the pattern in variation was due to a post-Pleistocene migration.

There is some circumstantial evidence for an origin of longleaf pine and western sources of loblolly pine from a common environment some time in the past. Western sources of loblolly pine as well as all sources of longleaf pine have greater resistance to fusiform rust (Wells et al. 1991) and are less susceptible to southern pine beetle (*Dendroctonus frontalis* Zimm.) (Powers et al. 1992) and Nantucket pine tip moth (*Rhyacionia frustrana* Comstock) (Schmidtling and Nelson 1996) than eastern sources of loblolly pine. Thus, it is reasonable to assume that western loblolly populations and all longleaf populations shared an environment at some time in the past where selection for resistance to these pests was important. The proposal that longleaf pine and western sources of loblolly pine both originated in a common refugia in southern Texas – northeastern Mexico fits the circumstantial evidence.

The present climate in southern Texas is too dry for pines but was probably much wetter during the Pleistocene (Watts 1983). Pines do exist just south of the border in Mexico at high elevations (Critchfield and Little 1966). Perry (1991) notes that there are four species of pine in Mexico that possess the peculiar growth habit of the "grass" stage that is characteristic of longleaf pine and uncharacteristic of other southern pines. One of these, *Pinus montezumae* Lamb. occurs in the mountains of northeastern Mexico. The current taxonomic status of these pines does not suggest that they are closely related to longleaf pine, but phylogenetic relationships are always undergoing revision.

Allozyme variation and growth

Variation in allozymes is generally considered to be selectively neutral, which makes them very appropriate for the kind of population analysis described above. On the other hand, variation in allozymes may be a good indicator of overall genetic variability and inbreeding which might be useful in predicting growth and adaptability.

Volume per hectare (in which survival is an important component) and height data, expressed as percent deviation from the expected, are compared with H_e in Table 5. There is a positive but nonsignificant relationship between expected heterozygosity and height ($r = 0.24$, $p = 0.15$) and volume ($r = 0.45$, $p = 0.081$). The correlations still do not reach significance if the residuals from the east–west regression are used (Table 5); the correlations of the residuals with height and with volume are $r = 0.43$ and $r = 0.41$, respectively. None of the correlations among height and volume with the diversity parameters are significant statistically.

The differences among sources for height and volume in Table 5, although statistically significant ($p < 0.01$), are not large, attesting to the applicability of the minimum-temperature seed-transfer model (Schmidtling 1994). The observation by Wells and Wakely (1970) of the poor performance of seed source 123 is verified, as this source also performs below expectations in height and volume when evaluated using the seed-transfer model. The diversity measures for this source are below average overall (Table 1) and are much below average when considering the expected for a source from this longitude (Fig. 3).

Another interesting anomaly in Table 5 is the performance of source 177. Not only is the height of this source greater than the expected, but volume is nearly 50% above the predicted volume in the seed-transfer model. Diversity indices for the source are about average. The source did stand out in the canonical discriminant analysis (Fig. 2). Hybridization with loblolly or slash pine might account for the difference. Hybrids of longleaf pine with loblolly or slash pine grow much faster in early years than longleaf pine (Lott et al. 1996). F_1 hybrids are easy to spot in the nursery and were certainly culled. Perhaps the culling of F_1 backcross hybrids in the nursery and in the field was not quite as successful in this source as in the other sources.

Conclusions

Longleaf pine has somewhat less allozyme variability than the other southern pines but, in general, does not appear to have diminished variation due to past logging practices, except in unusual instances, such as in source 123. Although tree improvement programs seem to have resulted in less genetic variation in loblolly pine (Williams et al. 1995), this does not appear to be a problem in the first-generation longleaf orchard populations included in this study. Allozyme variability does not seem to have much utility in predicting growth, except in cases of greatly diminished genetic variation.

The data in this paper supports the hypothesis that longleaf pine occupied a single, perhaps restricted, refugium in southern Texas or northern Mexico during the Pleistocene. A re-examination of the taxonomic relationship between the pines of northeastern Mexico and the southern pines seems warranted. The data also suggest that population sampling should favor western sources, because of the greater amount of variation in these sources.

Acknowledgements

This study was partially supported by a grant from the Longleaf Restoration Program, USDA Forest Service, Southern Research Station.

References

Adams, W.T., Neale, D.B., Doerksen, A.H., and Smith, D.B. 1990. Inheritance and linkage of isozyme variants from seed and vegetative bud tissue in coastal Douglas-fir [*Pseudotsuga menziesii* var. *menziesii* (Mirb.) Franco]. Silvae Genet. **39**: 153–167.

Cavalli-Sforza, L.L., and Edwards, A.W.F. 1967. Phylogenetic analysis: models and estimation procedures. Evolution, **21**: 550–570.

Chapman, H.H. 1922. A new hybrid pine (*Pinus palustris* × *Pinus taeda*). J. For. **20**: 729–734.

Conkle, M.T. 1981. Isozyme variation and linkage in six conifer species. *In* Proceedings, Symposium on the Isozymes of North American Forest Trees and Forest Insects, *Technical coordinator*: M.T. Conkle. USDA For. Serv. Gen. Tech. Rep. No. PSW-48. pp. 11–17.

Conkle, M.T., Hodgskiss, P.D., Nunnaly, L.B., and Hunter, S.C. 1982. Starch gel electrophoresis of conifer seeds: a laboratory manual. USDA For. Serv. Gen. Tech. Rep. No. PSW-64.

© 1998 NRC Canada

Critchfield, W.B., and Little, E.L., Jr. 1966. Geographic distribution of the pines of the world. USDA For. Serv. Misc. Publ. No. 991.

Croker, T.C., Jr. 1990. Longleaf pine—myths and facts. *In* Proceedings, Symposium on the Management of Longleaf Pine. *Edited by* R.M. Farrar, Jr. USDA For. Serv. Gen. Tech. Rep. No. SO-75. pp. 2–10.

Duba, S.E. 1985. Polymorphic isoenzymes from megagametophytes and pollen of longleaf pine: characterization, inheritance, and use in analyses of genetic variation and genotype verification. Proc. South. For. Tree Improve. Conf. **18**: 88–98.

Edwards, M.A., and Hamrick, J.L. 1995. Genetic variation in shortleaf pine, *Pinus echinata* Mill. (Pinaceae). For. Genet. **2**: 21–28.

Friedman, S.T., and Adams, W.T. 1985. Estimation of gene flow into two seed orchards of loblolly pine (*Pinus taeda* L.). Theor. Appl. Genet. **69**: 609–615.

Huneycutt, J.L., and Askew, G.R. 1989. Electrophoretic identification of loblolly pine – shortleaf pine hybrids. Silvae Genet. **38**: 95–96.

Kelly, J.F., and Bechtold, W.A. 1990. The longleaf pine resource. *In* Proceedings, Symposium on the Management of Longleaf Pine. *Edited by* R.M. Farrar, Jr. USDA For. Serv. Gen. Tech. Rep. No. SO-75. pp. 11–22.

Ledig, F.T. 1986. Heterozygosity, heterosis, and fitness in outbreeding plants. *In* Conservation biology: the science of scarcity and diversity. *Edited by* M.E. Soule. Sinauer Associates, Sunderland, Mass. pp. 77–104.

Little, E.L., Jr., and Dorman, K.W. 1954. Slash pine (*Pinus elliottii*), including south Florida slash pine. Nomenclature and description. USDA For. Serv. Southeast. For. Exp. Stn. Res. Pap. No. 36.

Lott, L.A., Schmidtling, R.C., and Snow, G.A. 1996. Susceptibility to brown-spot needle blight and fusiform rust in selected longleaf pine and hybrids. USDA For. Serv. Tree Plant. Notes No. 47. pp. 11–15.

Myburg, H., and Harris, S.A. 1997. Genetic variation across the natural distribution of the south east Asian pine, *Pinus kesiya* Royle ex Gordon (Pinaceae). Silvae Genet. **46**: 295–301.

Nei, M. 1977. *F*-statistics and analysis of gene diversity in subdivided populations. Ann. Human Genet. **41**: 225–233.

Perry, J.P., Jr. 1991. The pines of Mexico and Central America. Timber Press, Portland Oreg.

Powers, H.R., Jr., Belanger, R.P., Pepper, W.D., and Hastings, F.L. 1992. Loblolly pine seed sources differ in susceptibility to the southern pine beetle in South Carolina. South. J. Appl. For. **16**: 169–174.

Raja, R.G., Tauer, C.G., Wittwer, R.W., and Huang, Y. 1997. Isozyme variation in natural populations of shortleaf pine (*Pinus echinata*). Can. J. For. Res. **27**: 740–749.

SAS Institute Inc. 1990. SAS procedures guide, version 6. 3rd ed. SAS Institute Inc., Cary, N.C.

Schmidtling, R.C. 1994. Using provenance tests to predict response to climatic change: loblolly pine and Norway spruce. Tree Physiol. **14**: 805–817.

Schmidtling, R.C. 1997. Using provenance tests to predict response to climatic change. *In* Ecological issues and environmental impact assessment. *Edited by* P.N. Cheremisinoff. Gulf Publishing, Houston Tex. pp. 633–654.

Schmidtling, R.C., and Nelson, C.D. 1996. Interprovenance crosses in loblolly pine using selected parents. For. Genet. **3**: 53–66.

Schmidtling, R.C., and Sluder, E.R. 1995. Seed transfer and genecology in longleaf pine. Proc. South. For. Tree Improve. Conf. **23**: 78–85.

Schmidtling, R.C., and White, T. 1990. Genetics and tree improvement of longleaf pine. *In* Proceedings, Symposium on the Management of Longleaf Pine. *Edited by* R.M. Farrar, Jr. USDA For. Serv. Gen. Tech. Rep. No. SO-75. pp. 114–127.

Smouse, P.E., and Williams, R.C. 1982. Multivariate analysis of HLA–disease associations. Biometrics, **38**: 757–768.

Swofford, D.L., and Selander, R.B. 1989. BIOSYS-1, a computer program for the analysis of allelic variation in population genetics and biochemical systematics, release 1.7 edition. Illinois Natural History Survey, Champaign, Ill.

U.S. Department of Agrculture (USDA). 1990. USDA plant hardiness zone map. USDA Agric. Res. Serv. Misc. Publ. No. 1475.

Watts, W.A. 1983. A vegetational history of the eastern United States 25,000 to 10,000 years ago. *In* The late Pleistocene. Vol. I. Late-Quaternary environments of the United States. *Edited by* S.C. Porter. University of Minnesota Press, Minneapolis. pp. 294–310.

Wells, O.O., and Wakeley, P.C. 1970. Variation in longleaf pine from several geographic sources. For. Sci. **16**: 28–45.

Wells, O.O., Switzer, G.L., and Schmidtling, R.C. 1991. Geographic variation in Mississippi loblolly pine and sweetgum. Silvae Genet. **40**: 105–118.

Wheeler, N.C., and Guries, R.P. 1982. Biogeography of lodgepole pine. Can. J. Bot. **60**: 1805–1814.

Williams, C.G., Hamrick, J.L., and Lewis, P.O. 1995. Multiple-population versus hierarchical conifer breeding programs: a comparison of genetic diversity levels. Theor. Appl. Genet. **90**: 584–594.

Wright, S. 1931. Evolution in Mendelian populations. Genetics, **16**: 97–159.

Wright, S. 1965. The interpretation of population structure by *F*-statistics with special regard to systems of mating. Evolution, **19**: 395–420.

Wright, S. 1978. Evolution and the genetics of populations. Vol. 4. Variability within and among natural populations. University of Chicago Press, Chicago, Ill.

© 1998 NRC Canada