

# Microsatellite analysis of *Pinus taeda* L. in Zimbabwe

CLAIRE G. WILLIAMS\*†‡, CHRISTINE G. ELSIK‡ & RICHARD D. BARNES§

†Faculty of Genetics and ‡Department of Forest Science, Texas A & M University, College Station, Texas 77843–2135, U.S.A. and §Commonwealth Forestry Institute, Department of Plant Sciences, University of Oxford, Oxford, OX1 3RB, U.K.

Deducing the origin of early 20th century introductions of *Pinus taeda* into Zimbabwe is possible given microsatellite markers and clear population differentiation in ancestral U.S. populations. This study was designed to determine whether *P. taeda* introductions into Zimbabwe came from one U.S. region or whether the present-day population is an admixture of introductions from east and west of the Mississippi River Valley. Principal components analysis, Cavalli-Sforza and Edwards' chord distances and presence of diagnostic alleles each indicate that the Zimbabwe population is an admixture. There were five novel alleles in the Zimbabwe population not represented in the indigenous U.S. populations, possibly because of *de novo* mutation, introgression with other introduced North American pines or sampling error.

**Keywords:** conifers, phylogeny, population dynamics.

## Introduction

Microsatellite analysis detects interpopulational admixtures from population differentiation and cladistic relationships among alleles. Pines, indigenous only to the northern hemisphere, have been introduced into southern hemisphere countries to provide forest products. These introduced populations provided the opportunity to trace North American ancestors using microsatellite analysis. Microsatellite markers are ideal for determining the origin of 20th century introductions of *Pinus taeda* in Zimbabwe.

There have been several documented 20th century introductions of *P. taeda* into Zimbabwe prior to 1960 (Mullin *et al.*, 1978; pp. 189–193). The first introduction in Zimbabwe was a collection of two individuals growing near a homestead of the Rhodes Inyanga Estate. The second introduction in 1926 came from seed harvested from South African exotic plantations of unknown provenance. The third and largest introduction arrived in 1927–31 from both the U.S. via South Africa and from collections made in South African plantations. Results from local provenance trials suggest the Georgia Piedmont as the putative source of the third introduction (Mullin *et al.*, 1978; pp. 189–193). From 1957 to 1962, Queensland, Australia supplied three

*P. taeda* selections with a putative origin of south-eastern Georgia or north-eastern Florida.

Today, Zimbabwe is self-sufficient in producing *P. taeda* seed for afforestation, but determining the origin of Zimbabwe *P. taeda* introductions is critical for long-term management of the exotic population. The original provenance(s) of these selections is either the Georgia Piedmont (Mullin *et al.*, 1978) or more likely an admixture of ancestral populations east and west of the Mississippi River Valley. Genetic diversity levels appear comparable to U.S. populations (Williams *et al.*, 1995). The objective of this study is to determine the origin of a highly selected *P. taeda* population in Zimbabwe by comparison with *P. taeda* populations sampled throughout the species' natural range in the southern United States.

## Materials and methods

### *Study organism and sample collections*

*Pinus taeda* L. (Pinaceae) is wind-pollinated, outcrossing and monoecious. The species has a nearly continuous range throughout the southern U.S. Population divergence follows the east–west division of the Mississippi River (Florence & Zink, 1979); in the Atlantic coast populations about 95% of the genetic variation is found within a population rather than between populations (Williams *et al.*, 1995). The eastern range of the species

\*Correspondence. E-mail: claire-williams@tamu.edu

extends from southern Delaware (39°21'N) through Maryland, Virginia, North Carolina, South Carolina, Georgia into central Florida then westwards through Alabama, southern Tennessee and eastern Mississippi. This constitutes the eastern range of *P. taeda*. The Mississippi River Valley forms a natural gap of 50–200 km, delineating the western range of Louisiana, south-western Arkansas, the south-eastern tip of Oklahoma, eastern Texas and the remote Lost Pines population in Bastrop Co., Texas, which is the western-most range of *P. taeda*.

To test for admixtures, data must be available from either two putative sources or from parental populations which represent the original populations that produced the hybrid populations. A set of 41 U.S. samples were collected from grafted archives (Table 1). The archives represent indigenous populations present in the U.S. from 1950 to 1960 prior to domestication. The selected Zimbabwe population was composed of 12 pedigreed Zimbabwe individuals bred from eight

unrelated parents of unknown origin. The Zimbabwe population represents an elite domesticated population which is three generations removed from indigenous U.S. populations.

#### PCR amplification for microsatellite markers

The DNA extraction was based on Doyle & Doyle (1987). The 18 low-copy and genomic microsatellite markers were published in Elsik *et al.* (in press): PtTX2008, PtTX2037, PtTX2123, PtTX2128, PtTX2142, PtTX2146, PtTX2164, PtTX3011, PtTX3013, PtTX3017, PtTX3019, PtTX3020, PtTX3025, PtTX3026, PtTX3030, PtTX3032, PtTX3034, PtTX3037. GenBank accession numbers for sequences were AF143958 to AF143975.

Each 25 µL PCR reaction consisted of 50 ng of template DNA, 50 mM Tris (pH 9.0), 20 mM ammonium sulphate, 100 µg/µL gelatin, 1% DMSO, 100 µM of each dNTP, 200 nM of each forward and reverse primer, 0.625 units Taq polymerase (Life Technologies),

**Table 1** Description of 41 samples of *Pinus taeda* representing U.S. populations from 1950 to 1960

Region	<i>N</i>	County/State	Latitude/Longitude	
Lost Pines	5	Bastrop, Texas (5)	30°	97°
Western Gulf	10	Polk, Texas (1)	30°30'	95°
		Liberty, Texas (1)	30°20'	95°
		McCurtain, Oklahoma (2)	34°30'	95°
		Livingston, Louisiana (2)	30°30'	91°
		Bienville, Louisiana (1)	32°5'	93°10'
		Winn, Louisiana (1)	32°	93°
		Ashley, Arkansas (1)	33°	92°
		Lafayette, Arkansas (1)	33°	93°30'
Mississippi–Alabama	5	Jasper, Mississippi (1)	32°89'	
		Clarke, Mississippi (1)	32°88'	50°
		Winston, Alabama (1)	34°	87°
		Tallapoosa, Alabama (1)	32°40'	85°40'
		Franklin, Alabama (1)	34°	88°
South Carolina–Florida	14	Georgetown, South Carolina (2)	33°50'	79°30'
		Williamsburg, South Carolina (1)	33°	80°
		Francis Marion National Forest, South Carolina (3)	33°	80°
		Morgan, Georgia (1)	33°30'	83°30'
		Cherokee, Georgia (1)	34°50'	84°30'
		Wayne, Georgia (1)	31°50'	82°
		Dixie, Florida (1)	29°30'	83°
		Levy, Florida (2)	29°	82°30'
Marion, Florida (2)	28°	82°15'		
North-east	7	Northampton, North Carolina (1)	37°15'	77°
		Southampton, Virginia (1)	37°20'	77°
		Sussex, Virginia (1)	36°45'	77°
		Accomac, Virginia (1)	37°30'	76°
		Somerset, Maryland (1)	38°15'	75°30'
		Wicomico, Maryland (1)	38°20'	75°30'
		Dorchester, Maryland (1)	38°30'	75°30'

1–2 mM MgCl<sub>2</sub> (optimized for each primer set). PCR products were visualized on silver-stained 6% polyacrylamide denaturing gels using modifications from Echt *et al.* (1996). The molecular weight marker pGem (Promega) was loaded every eight lanes for allele sizing.

### Estimating genetic diversity

Allele frequencies were determined by direct count. Diagnostic alleles for eastern and western parts of the *P. taeda* range were determined by pooling allele frequencies in these two regions exclusive of Mississippi–Alabama. The Mississippi–Alabama population was intermediate between the eastern and western regions and thus omitted in the identification of region-specific diagnostic alleles. Only alleles present in Zimbabwe were treated as diagnostic in either the eastern or western parts of the range. Alleles which were present in one part of the range yet absent in Zimbabwe were excluded because their absence was confounded with population sampling error. Summary analyses were completed using GENEPOP (Raymond & Rousset, 1995) and BIOSYS-2, a version of BIOSYS-1 (Swofford & Selander, 1989) modified for microsatellite analyses by Dr William C. Black IV, Colorado State University (pers. comm.).

Deviations from Hardy–Weinberg equilibrium (HWE) were based on exact tests ( $\chi^2$ ) (Rousset & Raymond, 1995). In addition to the exact test, a Markov-chain algorithm was used for loci with five or more alleles to obtain an unbiased estimate of the exact probability of incorrectly rejecting HWE (Rousset & Raymond, 1995). In all cases, the Markov chain was set to 1000 steps of dememorization and 1000 iterations per batch. Heterozygote deficiency was tested as the alternative hypothesis to HWE. Heterozygote excess, common in the presence of symmetric overdominance or in the case where male and female gametes have different allele frequencies, was separately tested as an alternative hypothesis to HWE. The experimentwise error rates ( $\alpha'$ ) for 108 chi-squared tests (Weir, 1990; p. 109) were 0.996 and 0.102 at the 5% and 0.1% levels, respectively.

The probability of deviation from HWE for pooled loci across populations was determined using Fisher's formula (Raymond & Rousset, 1995). This is a  $\chi^2$  distribution with  $2 \times r$  degrees of freedom.

$$\chi_{\text{pop}}^2 = 2 \left( \sum_{j=1}^r \ln P_j \right),$$

where  $r$  is the number of loci pooled across populations and  $P_j$  is the probability value for each population.

### Principal components analysis

Hybrid populations are not discrete evolutionary lineages and thus phylogenetic trees with admixed populations can violate the fundamental principles of phylogeny reconstruction (Felsenstein, 1982). Principal components analysis should be used in lieu of phylogenetic trees in the case of admixture between two ancestral populations (Cavalli-Sforza *et al.*, 1994; pp. 39–42). Principal components analysis was conducted using the SAS statistical and graphics software package (SAS, 1991).

### Estimating genetic distances and cladistic relationships among populations

Genetic distance estimates for microsatellites are based on underlying mutational processes. At least two known mutational processes are responsible for microsatellite variation: the infinite alleles model (IAM) and the stepwise mutation model (SMM). Unequal crossing-over during meiosis generates new alleles, satisfying the IAM assumptions for novel alleles not previously found in the population (Kimura & Crow, 1964). For SMM, DNA polymerase slippage shifts the number of repeats up or down in a stepwise manner (Levinson & Guttman, 1987). Under SMM, an allele with  $i$  repeats mutates to an allele either in state  $i + 1$  or  $i - 1$  with equal probability (Ohta & Kimura, 1973). The coexistence of two mutational processes operating on different loci or even different alleles within a locus raises the question of an appropriate measure of genetic distance (Estoup *et al.*, 1995). The best choice is a third class of genetic distance estimates that is not based on a mutational model. Cavalli-Sforza and Edwards' chord distance,  $D_c$ , depends on geometric distances rather than on a given mutational process (Cavalli-Sforza & Edwards, 1967). A computer simulation study of various genetic distance measures found  $D_c$  to be the most suitable for microsatellite analysis, even in the presence of a bottleneck effect (Takezaki & Nei, 1996).

If the two populations are represented on the surface of a multidimensional hypersphere using allele frequencies at the  $j$ th locus,  $D_c$  gives the chord distance between two populations. A generalized form of the chord distance (Goldstein & Pollock, 1997) is given as follows:

$$D_c = c \left[ 1 - \sum (x, y)^b \right]^a,$$

where  $x$  and  $y$  are frequencies of alleles of size  $i$  in populations  $x$  and  $y$ , respectively, and  $a$  and  $b$  are constants set at 0.5 where the constant  $c$  equals  $2\sqrt{2/\pi}$ . For multiple loci, the average distance is taken across

loci. The cladistic relationship among U.S. (ancestral) populations was based on the chord distances and the neighbour-joining method (Saitou & Nei, 1987). Neighbour-joining analysis was conducted using MEGA (Kumar *et al.*, 1993).

## Results

Genetic diversity levels were high within all populations except the Lost Pines from Bastrop Co., Texas. The greatest population differentiation for *P. taeda* occurred between east and west of the Mississippi River Valley. High levels of microsatellite polymorphism and population differentiation in the ancestral U.S. populations made it possible to distinguish between an admixture vs. the putative Georgia Piedmont origin for the Zimbabwe population. The Zimbabwe population was an admixture of south-eastern and western *P. taeda* populations. Five alleles in the Zimbabwe population were not detected in the samples of U.S. ancestral populations.

### Genetic diversity estimation

All populations including Zimbabwe had similar levels of genetic diversity if the sample size bias was removed (Table 2). The total allele number was 192 for the full sample (Table 2). The SC-FL region added the most alleles (126) but it was also the population with the largest sample (14/41). The other populations had the following total alleles: Zimbabwe (94), Western Gulf (88), North-east (83), MissAla (67) and Bastrop (63). However, mean number of alleles (MNA) is a good indicator of genetic variation only if the populations are the same size and at mutation-drift equilibrium.

To check for sample size bias, the Pearson product-moment correlation between MNA and the size of each population was estimated. The highly positive correlation ( $r = 0.93$ ) indicated that MNA was higher in the larger populations and that there was a bias in allele

numbers caused by sample size. A uniform size estimate of MNA was added by randomly deleting entries in the larger population so that sample size became uniform in all populations (Table 2) and the total allele number was 172. After adjusting for sample size, the allelic contributions were nearly similar in all populations. The Lost Pines in Bastrop Co., Texas had slightly less genetic diversity.

This is the first estimate of genetic diversity for Bastrop Co., Texas. This population was sampled within a single county, rather than the single sample per county or even regional sampling applied to other populations. The lower genetic diversity level was also expected because Bastrop Co., Texas has partial reproductive isolation. This population is connected to the larger, continuous distribution of *P. taeda* in east Texas by a string of small isolated stands between Bastrop and Houston, Texas.

The HWE test for heterozygosity deficiency was significant at the 0.1% level in eight out of 108 tests (Table 3). Significance at the 5% level was only reported as a suggestive level because of its high experimentwise error rate for 108 tests, raising the probability of incorrectly rejecting HWE. Pooled loci across populations showed that seven loci out of 18 had statistically significant heterozygote deficiency and thus deviated substantially from HWE (Table 3). There was no heterozygote excess detected at any locus.

### Divergence in ancestral populations

The greatest population differentiation in ancestral *P. taeda* populations was east and west of the Mississippi River Valley. The neighbour-joining phylogeny indicated that eastern populations were more similar to one another than to western populations (Fig. 1). The neighbour-joining phylogeny was constructed for the two sets of ancestral U.S. populations to check for population differentiation prior to testing for admixture.

**Table 2** Summary statistics for 18 low-copy microsatellite loci assayed for Zimbabwe samples and five U.S. *Pinus taeda* populations

Source	<i>n</i>	MNA (full)	MNA (uniform)	Mean observed heterozygosity ( $H_o$ )	Expected heterozygosity ( $H_e$ )	% Poly loci (95%)	% Poly loci
Zimbabwe	12	5.2 ± 0.70	3.8 ± 0.50	0.590 ± 0.06	0.623 ± 0.06	88.9	100
Bastrop	5	3.5 ± 0.30	3.5 ± 0.30	0.483 ± 0.06	0.558 ± 0.06	94.4	94.4
Western Gulf	10	5.0 ± 0.50	4.1 ± 0.50	0.492 ± 0.05	0.619 ± 0.05	100	100
Miss/Alabama	5	3.8 ± 0.50	3.8 ± 0.50	0.486 ± 0.08	0.591 ± 0.08	83.3	83.3
SC-FL	14	7.1 ± 1.0	4.6 ± 0.50	0.533 ± 0.06	0.643 ± 0.07	94.4	100
North-east	7	4.7 ± 0.60	4.4 ± 0.60	0.571 ± 0.06	0.628 ± 0.07	94.4	94.4
Total	53		—			100	100

MNA, mean number of alleles per locus.

The eastern and western regions were clearly delineated, as reported by Florence & Zink (1979). The population differentiation between eastern and western *P. taeda* regions and microsatellites made it possible to distinguish between U.S. regional origins and an interpopulational admixture in the selected Zimbabwe population.

**Testing for interpopulational admixture**

The principal components analysis supported the Zimbabwe population as an admixture. This was corroborated by the diagnostic alleles which appear to be specific to eastern or western regions of the *P. taeda* range in the southern U.S. The eastern region included the

Eastern Shore, Virginia, North Carolina (north-east) and South Carolina, Georgia and Florida. The western region was represented by Louisiana, Oklahoma, Arkansas and Texas (Western Gulf) including the westernmost extreme of the *P. taeda* range, Bastrop Co., Texas.

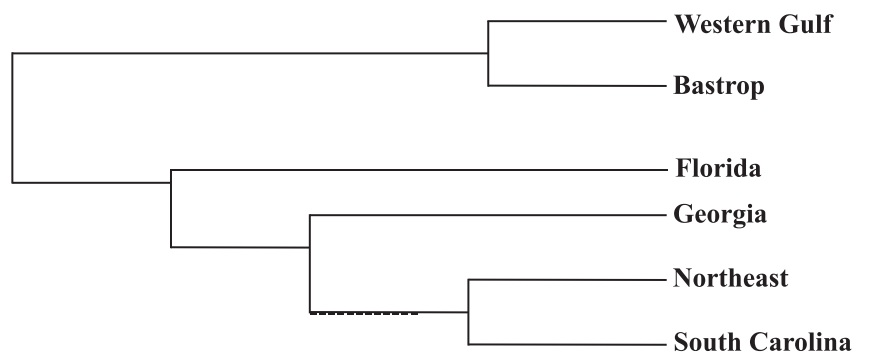
To a lesser extent, chord distances showed the Zimbabwe selected population to be an admixture of populations ancestral to the east and west of the Mississippi River Valley. A finer analysis of sources within the South Carolina–Florida population was conducted to find the central origins. The analysis showed Georgia Piedmont and north-eastern Florida to have the two highest chord distances with Zimbabwe ( $D_c = 0.578, 0.556$ ). In the western region, the chord distance was higher for Bastrop Co., Texas ( $D_c = 0.543$ ). The South Carolina coastal region, the north-eastern range, and Western Gulf regions had slightly lower chord distances ( $D_c = 0.521, 0.475$  and  $0.433$ , respectively).

Principal components analysis showed the Zimbabwe population to be divergent from either U.S. region despite a marked affinity to the eastern region (Fig. 2). The proportion of the total variance accounted for by all three components is 90.6%. The first, second and third principal components accounted for 73.1%, 9.2% and 8.3% of the total variance, respectively. Divergence of the Zimbabwe selected population from other populations was more pronounced with the addition of the third component (Fig. 2).

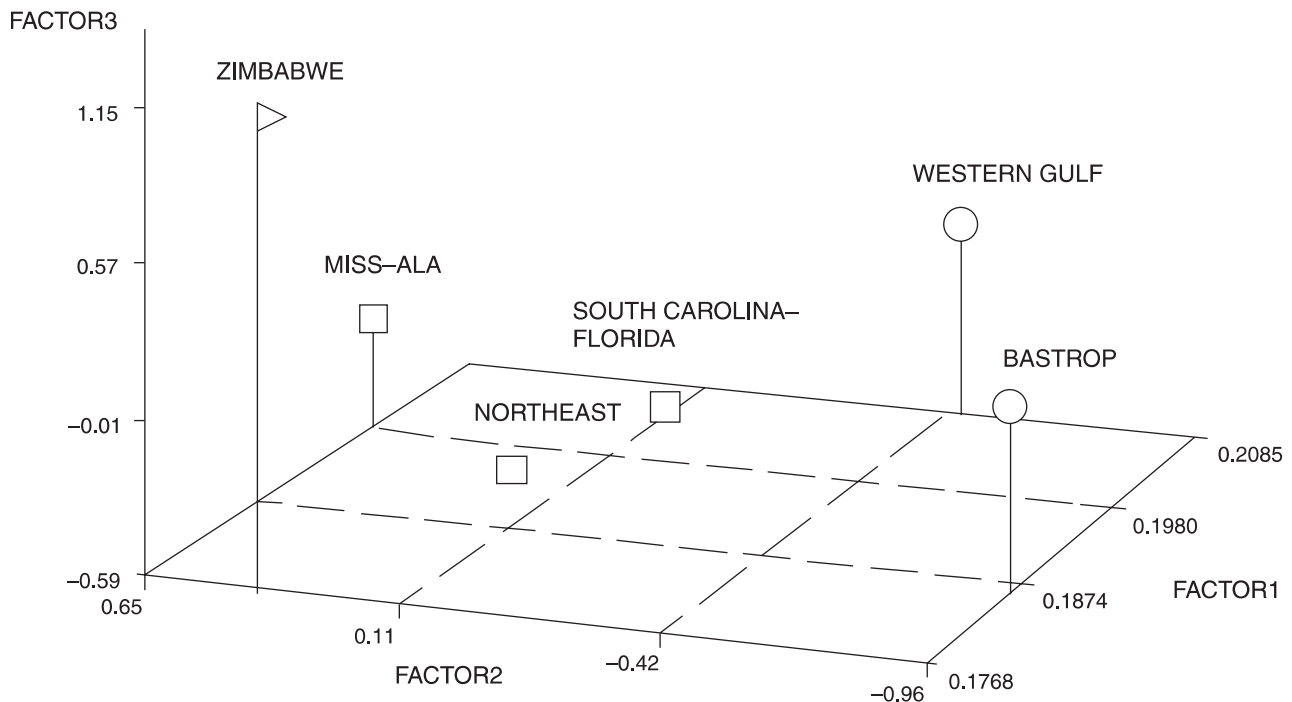
Allelic frequency distribution also confirmed the Zimbabwe population as an admixture. Seven diagnostic alleles which were present in Zimbabwe showed affinity for both western and eastern regions (Table 4). Each had an allele present in the Zimbabwe population which occurred in either the eastern or western regions. Of the seven, five occurred only in the eastern region and two occurred in the western region. The region-specific alleles could not be traced back to one particular parent, suggesting that admixture occurred in preceding generations. For example, the western alleles occurred in cross  $124 \times 13$  at locus *PtTX3037*. The same cross  $124 \times 13$

**Table 3** Probability values (□ for  $P \geq 0.05$ ; ■ for  $P \geq 0.001$ ) for exact test of Hardy–Weinberg equilibrium vs. heterozygote deficiency for *Pinus taeda* microsatellites. Markers developed from a set of total genomic libraries have 1 as second digit

Locus	ZIM	BA	WG	M–A	SC–FL	NE	Pooled loci
PtTX2008							
PtTX2037					□	□	■
PtTX2123				□			
PtTX2128	□						
PtTX2142							■
PtTX2146							
PtTX2164	□		□		□		■
PtTX3011							
PtTX3013							
PtTX3017							
PtTX3019					□		
PtTX3020					■		■
PtTX3025					□		□
PtTX3026							
PtTX3030			■	■	■		■
PtTX3032		□	■	□			■
PtTX3034							
PtTX3037	■	□	■	□	■	□	■



**Fig. 1** Neighbour-joining tree based on Cavalli-Sforza & Edwards (1967) chord distances among three eastern and two western *Pinus taeda* populations indigenous to the southern U.S.



**Fig. 2** Graphical depiction of principal components analysis for Zimbabwe and the U.S. ancestral populations of *Pinus taeda*. Factors 1, 2 and 3 account for 90.6% of the total variance associated with allele distribution and frequencies.

**Table 4** Allele number and frequencies for Zimbabwe which are unique to eastern vs. western parts of the natural range of *Pinus taeda*. Bold letters indicate frequencies of diagnostic alleles for the western part of the *P. taeda* range

Locus	Allele	Zimbabwe	Western		Eastern	
			BA	WG	SC-FL	NE
2037	11	0.125	0	0	0.038	0.071
2128	8	0.042	0	0	0.143	0.286
2142	2	<b>0.042</b>	<b>0.250</b>	<b>0.125</b>	0	0
2164	10	0.136	0	0	0.107	0.167
3011	10	0.167	0	0	0.115	0.143
3030	8	0.056	0	0	0.042	0.083
3037	18	<b>0.208</b>	<b>0.500</b>	<b>0.214</b>	0	0

BA, Bastrop Co., Texas; WG, Western Gulf region of Oklahoma, Arkansas, Louisiana and eastern Texas; SC-FL, South Carolina to Florida; NE, north-eastern *P. taeda* range from Eastern Shore, Maryland to coastal North Carolina.

also had eastern alleles in three cases: *PtTX2128* allele 8, *PtTX2164* allele 10 and *PtTX3011* allele 10.

#### Diagnostic alleles for the Zimbabwe population

The Zimbabwe population had five alleles absent in the U.S. samples. The unique alleles were *PtTX2164* allele 3, *PtTX3030* allele 17, *PtTX3034* alleles 9, 11 and *PtTX3037*

allele 26. The four loci with a total of five alleles either (i) arose from *de novo* mutations (ii) were overlooked in the U.S. sampling or (iii) resulted from introgression with another introduced North American pine.

Three of these four loci (*PtTX3030*, *PtTX3034*, *PtTX3037*) had dinucleotide repeat motifs (Elsik *et al.*, in press). Four of the Zimbabwe alleles were at the smaller end of the size range of the respective locus; all alleles except 3037-26 were separated from the adjacent larger allele by at least two to five nucleotides, suggesting a tendency towards contraction in repeat length or deletions. The distribution of these private alleles also suggested a rare mutation, a low-frequency allele from *P. taeda* or mating with a close relative. Allele 17 in *PtTX3030* occurred twice but the same parent contributed the allele in both cases. Allele 3 in *PtTX2164* occurred four times in the selected Zimbabwe population but one parent contributed the allele in three of the four cases. Both *PtTX3034* alleles 9 and 11 and *PtTX3037* allele 26 each occurred once.

#### Discussion

This is the first microsatellite analysis of population structure for *P. taeda*. Detecting interpopulational admixture is possible given population differentiation between eastern and western regions of the *P. taeda* range (Florence & Zink, 1979) and highly polymorphic

microsatellite markers (Elsik *et al.*, in press). The selected Zimbabwe population is an admixture of eastern and western *P. taeda* populations.

#### High genetic diversity within all populations

All *P. taeda* populations except Bastrop Co., Texas had similar levels of genetic diversity. Zimbabwe's selected population had genetic diversity levels comparable to U.S. populations, as previously reported (Williams *et al.*, 1995). The Lost Pines were heterozygous but had lower genetic diversity levels. The lower genetic diversity resulted from the narrow area for sampling and partial reproductive isolation. The high genetic diversity levels within U.S. populations and relatively low levels of population differentiation fit with *P. taeda*'s life history attributes: historically large populations, an outcrossing species, annual producer of larger quantities of wind-borne pollen and seeds, and an aggressive colonizer. These attributes may explain why Zimbabwe *P. taeda* introductions have retained high genetic diversity levels: the extensive plantings throughout southern Africa, multiple introductions of U.S. germplasm (Mullin *et al.*, 1978; pp. 189–193) and possibly introgression or *de novo* mutations in southern Africa.

Deviation from Hardy–Weinberg equilibrium was caused by heterozygote deficiency. No loci exhibited deviation resulting from heterozygote excess. The most common cause for heterozygote deficiency in microsatellite marker studies is the presence of null alleles. Null alleles occur when there is a mutation within the DNA sequence complementary to one or both primers, preventing PCR amplification of the microsatellite sequence (Callen *et al.*, 1993). Null alleles for some of these loci have been detected in intergenerational *P. taeda* pedigrees (Elsik *et al.* in press) but are undetected in this population survey, leading to downward bias for heterozygosity.

Other causes of heterozygote deficiency include selection or hitchhiking on a linked locus under selection, deviation from panmixis, or population admixture. Of the three, selection seemed to be a likely cause for HWE deviation at locus *PtTX3037*. Unlike the other microsatellite loci, *PtTX3037* exhibited strong heterozygote deficiency at the 5% or 0.1% levels for all populations including Zimbabwe (Table 3). Of the 18 loci, 13 were selected from low-copy gene-rich regions in the pine genome (Elsik *et al.* in press) and thus may be under strong directional selection or closely linked to protein-coding regions which are under selection. Deviation from panmixis or the presence of admixture seemed the least likely cause of heterozygote deficiency. The selected Zimbabwe population had fewer HWE deviations than the U.S. populations and yet it exhibited admixture and a deviation from panmixis (Table 3).

#### *Pinus taeda* in Zimbabwe: a divergent admixture

The strongest evidence for admixture was based on the principal components analysis (PCA). Chord distances differed only slightly and the diagnostic alleles should be subject to closer scrutiny with larger sampling from U.S. populations. PCA results also indicated that Zimbabwe has diverged from its ancestral populations (Fig. 2). Divergence may be accelerated by the same mutational processes which generated the five private alleles found only in the Zimbabwe population.

The unique alleles could have arisen through *de novo* mutations or from introgression with another introduced North American pine species. Both explanations are supported by a novel allele at the *PgI* isozyme locus found in a selected Zimbabwe population; this allele was absent in over 500 U.S. samples throughout the *P. taeda* range (Williams *et al.*, 1995; J. L. Hamrick, pers. comm.). A less likely explanation is that the unique alleles are very low-frequency and thus undetected in these U.S. samples. If so, the frequency of these alleles would have been increased in the Zimbabwe population through drift and selective breeding.

Three of these four loci (*PtTX3030*, *PtTX3034*, *PtTX3037*) have dinucleotide repeat motifs (Elsik *et al.* in press). Dinucleotide repeats tend to have higher mutation rates in model organisms (Chakraborty *et al.*, 1997) and this supports the idea of *de novo* mutations in Zimbabwean *P. taeda* introductions. Additional *de novo* mutations for Zimbabwe may be present yet obscured by allelic size homoplasy where two different repeat changes result in the same allele size (Grimaldi & Crouau-Roy, 1997) or by null alleles (Callen *et al.*, 1993).

The fact that Zimbabwe alleles were at the smaller end of the size range of the respective locus mitigates against *de novo* mutation. Four of the five Zimbabwe alleles were at the smaller end of the size range of the respective locus; all alleles except marker 3037 allele 26 were separated from the adjacent larger allele by at least two to five nucleotides. If so, then these private alleles would have arisen from deletions or from contractions in the repeat motif rather than expansions. This is counter to the observed trend in microsatellite analyses in mammals: random mutations in the repeat motif are expected to be both expansions and contractions and longer alleles expand more than short alleles (e.g. Valdes *et al.*, 1993).

Introgression with *Pinus elliottii*, another introduced species closely related to *P. taeda*, is the more likely source of novel alleles for Zimbabwe populations. Closely related pines hybridize freely and North American pines have an unusual reproductive phenology in southern Africa which favours introgression. The strobilus receptivity between *P. elliottii* and *P. taeda*

species completely overlaps relative to the species range in the southern U.S. (Mullin *et al.*, 1978). Seed collections from plantations in southern Africa would be more likely to contain a higher proportion of interspecific hybrids than U.S. collections. Natural hybrid zones frequently have higher levels of novel alleles (Hoffman & Brown, 1995). The introgression hypothesis is further supported by the high degree of interspecific repeat motif conservation in many low-copy microsatellite markers (C. G. Williams, unpubl. data).

## Acknowledgements

This work was sponsored by the Texas Higher Education Coordinating Board grants 999902-090 and 999902-201, USDA-Forest Service, The Timber Company and NATO Collaborative Research Grant 940699. We thank V. T. Minihan and A. M. Scarpa for technical assistance. Germplasm was supplied by Texas Forest Service, The Timber Company, Westvaco Corporation, Weyerhaeuser Company, Bosch Industries, Chesapeake Corporation, Oklahoma Forest Service and USDA-Forest Service National Forest System.

## References

- CALLEN, D. F., THOMPSON, A. D., SHENN, Y., PHILLIPS, H. A., RICHARDS, R. I., MULLEY, J. C. *ET AL.* 1993. Incidence and origin of 'null' alleles in the (AC)<sub>n</sub> microsatellite markers. *Am. J. Hum. Genet.*, **52**, 922–927.
- CAVALLI-SFORZA, L. L. AND EDWARDS, A. W. F. 1967. Phylogenetic analysis models and estimation procedures. *Am. J. Hum. Genet.*, **19**, 233–257.
- CAVALLI-SFORZA, L. L., MENOZZI, P. AND PIAZZA, A. 1994. *The History and Geography of Human Genes*. Princeton University Press, Princeton, NJ.
- CHAKRABORTY, R., KIMMEL, D. N., STIVERS, D. N., DEKA, R. AND DAVISON, L. J. 1997. Relative mutation rates at di-, tri- and tetranucleotide microsatellite loci. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 1041–1046.
- DOYLE, J. J. AND DOYLE, J. L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.*, **19**, 11–15.
- ECHT, C. S., MAY-MARQUARDT, P., HSEIH, M. AND ZAHORCHAK, R. 1996. Characterization of microsatellite markers in eastern white pine. *Genome*, **39**, 1102–1108.
- ELSIK, C. G., MINIHAN, V. T., HALL, S. E., SCARPA, A. M. AND WILLIAMS, C. G. in press. Low-copy microsatellite markers for *Pinus taeda*, L. *Genome*, in press.
- ESTROUP, A., GARNERY, L., SOLIGNAC, M. AND CORNUET, J. 1995. Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics*, **140**, 679–695.
- FELSENSTEIN, J. 1982. How can we infer geography and history from gene frequencies? *J. Theor. Biol.*, **96**, 9–20.
- FLORENCE, L. Z. AND ZINK, G. 1979. Geographic patterns of allozyme variation in loblolly pine. In: *Proceedings of the Southern Forest Tree Improvement Conference*, Starkville MS, pp. 33–42.
- GOLDSTEIN, D. B. AND POLLOCK, D. D. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *J. Hered.*, **88**, 335–342.
- GRIMALDI, M. AND CROUAEU-ROY, B. 1997. Microsatellite allelic homoplasy due to variable flanking sequences. *J. Mol. Evol.*, **44**, 336–340.
- HOFFMAN, S. M. G. AND BROWN, W. M. 1995. The molecular mechanism underlying the 'rare allele phenomenon' in a subspecific hybrid zone of the California field mouse, *Peromyscus californicus*. *J. Mol. Evol.*, **41**, 1165–1169.
- KIMURA, M. AND CROW, J. F. 1964. The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725–738.
- KUMAR, S., TAMURA, K. AND NEI, M. 1993. MEGA: *Molecular Evolutionary Genetics Analysis*, version 1.0. Pennsylvania State University, University Park, PA 16802.
- LEVINSON, G. AND GUTTMAN, G. A. 1987. Slipped strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.*, **4**, 203–221.
- MULLIN, L. J., BARNES, R. D. AND PREVOST, M. J. 1978. *A Review of the Southern Pines in Rhodesia*. Rhodesian For. Res. Bull. no. 7., Salisbury, Rhodesia.
- OHTA, T. AND KIMURA, M. 1973. The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. *Genet. Res.*, **22**, 201–204.
- RAYMOND, M. AND ROUSSET, F. 1995. GENEPOP (version 1.2): *population genetics software for exact tests and ecumenicism*. *J. Hered.*, **86**, 248–249.
- ROUSSET, R. 1996. Equilibrium values for measures of population subdivision for stepwise mutation processes. *Genetics*, **142**, 1357–1362.
- ROUSSET, F. AND RAYMOND, M. 1995. Testing heterozygote excess and deficiency. *Genetics*, **140**, 1413–1419.
- SAITOU, N. AND NEI, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
- SAS. 1991. *SAS System for Statistical Graphics*. SAS Institute, Cary NC.
- SWOFFORD, D. L. AND SELANDER, R. B. 1989. BIOSYS-1: a Fortran program for comprehensive analysis for electrophoretic data in population genetics and systematics. *J. Hered.*, **72**, 281–283.
- TAKEZAKI, N. AND NEI, M. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389–399.
- VALDES, A. M., SLATKIN, M. AND FREIMER, F. B. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics*, **133**, 737–749.
- WEIR, B. S. 1990. *Genetic Data Analysis*. Sinauer Associates, Sunderland, MA.
- 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.