Tree Genetics & Genomes
DOI 10.1007/s11295-008-0157-x

ORIGINAL PAPER

Natural hybridization within seed sources of shortleaf pine (*Pinus echinata* Mill.) and loblolly pine (*Pinus taeda* L.)

Shiqin Xu · C. G. Tauer · C. Dana Nelson

Received: 20 December 2007 / Accepted: 2 March 2008

Abstract Shortleaf and loblolly pine trees (n=93 and 102, respectively) from 22 seed sources of the Southwide Southern Pine Seed Source Study plantings or equivalent origin were evaluated for amplified fragment length polymorphism (AFLP) variation. These sampled trees represent shortleaf pine and loblolly pine, as they existed across their native geographic ranges before intensive forest management. Using 17 primer pairs, a total of 96 AFLPs between shortleaf pine and loblolly pine were produced and scored on the sample trees and two control-pollinated F1 interspecies hybrids and their parents. In addition, the well known isocitrate dehydrogenase (*IDH*) isozyme marker was scored for all trees. *IDH* detected two putative hybrids among the loblolly pine samples and two among the shortleaf pine samples, while either 13 or 12 putative hybrids were detected using all AFLP markers and *IDH* and either NewHybrids or Structure software, respectively. Results of this study show that later generation hybrids can be reliably identified using AFLP markers and confirmed that *IDH* is not a definitive marker for detecting hybrids; that is, at least in some seed sources, the alternative species' *IDH* allele resides in the source species. Based on all the markers, hybridization frequency varied geographically, ranging from 30% in an Arkansas seed source to 0% in several other seed sources. The hybridization level was higher in populations west of the Mississippi River than in populations east of the river; the shortleaf pine hybridization rates were 16.3% and 2.4% and the loblolly pine rates were 4.5% and 3.3%, west and east of the river, respectively. The results suggest that hybridization between these two species is significant but varies by seed source and species, and the potential for the unintended creation of hybrids should be considered in forest management decisions regarding both natural and artificial regeneration.

Keywords Genetic diversity

Communicated by: R. Sederoff

S. Xu · C. G. Tauer (✉)
Department of Natural Resource Ecology and Management,
Oklahoma State University,
Stillwater, OK 74078, USA
e-mail: chuck.tauer@okstate.edu

C. D. Nelson
Southern Institute of Forest Genetics, Southern Research Station,
USDA Forest Service,
Saucier, MS 39574, USA

Introduction

Shortleaf pine (*Pinus echinata* Mill.) and loblolly pine (*Pinus taeda* L.) are both of considerable economic importance in southeastern USA. Both species can be used for construction lumber, plywood, posts, poles, paper, and other physical and chemical products. They have broad geographic ranges and a large sympatric region (Fig. 1).

Research has shown that shortleaf pine and loblolly pine have similar karyotypes (Saylor 1972, Islam-Faridi et al. 2007), so they are expected to cross with each other. As early as 1933, artificial hybrids between these two species were created by the Institute of Forest Genetics, Placerville, CA and reported by Schreiner (1937). In nature, however, there are other conditions such as flowering time that affect possible hybridization. Loblolly pine has mature male and receptive female strobili from the end of February to the middle of March, while shortleaf pine has mature male and receptive female strobili about 2 to 3 weeks later; however, peak maturity and receptivity timing may vary by as much

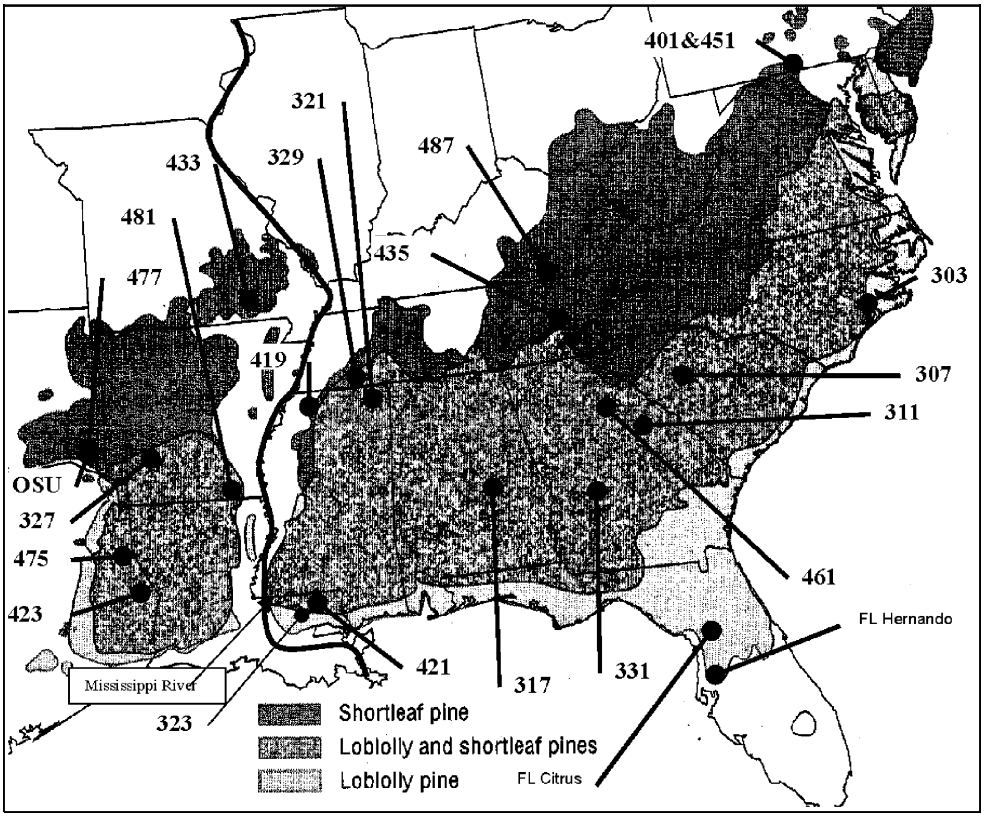

Fig. 1 The natural ranges of shortleaf pine and loblolly pine and approximate sample origins (300's are loblolly pine and 400's are shortleaf pine)

as 3 weeks among trees in the same stand (Dorman and Barber 1956). Male and female strobili maturity is also affected by seasonal climatic fluctuations, which may lead to overlapping timing of receptivity. Thus, hybridization between the two species is possible in sympatric populations in some years.

As early as 1953, researchers reported trees with morphologies intermediate between shortleaf pine and loblolly pine, suggesting that hybrids do occur naturally in the sympatric region (Hare and Switzer 1969; Zobel 1953). Some trees in loblolly pine populations in sympatric regions were found to have resistance to fusiform rust (Henry and Bercaw 1956) similar to that found in shortleaf pine. Likewise, some trees in shortleaf pine stands showed resistance to littleleaf disease, to which shortleaf pine is susceptible and loblolly pine is resistant. Various morphological characters have been used to study hybrids between shortleaf pine and loblolly pine (Mergen et al. 1965; Cotton et al. 1975) with limited success (Hicks 1973). Recent studies using isozyme markers, in particular, the isocitrate dehydrogenase (*IDH*) marker, have revealed a relatively high level of hybridization among trees in shortleaf pine and loblolly pine populations in west-central Arkansas (Raja et al. 1998; Chen et al. 2004) and somewhat lower levels in Georgia and range wide (Edwards et al. 1997).

The isoenzyme *IDH* has been reported to be a good marker (Huneycutt and Askew 1989) for finding first generation hybrids, but due to normal Mendelian segregation, its value in finding later generation hybrids is diminished. Given the recognized problems with morphological and isozyme markers, it is clear that additional markers are needed to reliably detect hybrids. DNA-based markers appear most promising, and a few have been developed and used to identify putative hybrids of shortleaf pine and loblolly pine (Chen et al. 2004; Edwards et al. 1997) with varying success. In the current study, we explore the use of amplified fragment length polymorphisms (AFLPs) combined with *IDH* to identify hybrids, in particular, later generation hybrids, between shortleaf pine and loblolly pine to estimate the hybridization level throughout their natural ranges and to examine possible differences between western and eastern populations. The AFLP and *IDH* marker data were analyzed using the software NewHybrids version 1.1 beta (Anderson and Thompson 2002; Anderson 2003) and Structure version 2.2 (Prichard et al. 2000; Falush et al. 2003, 2007).

Because loblolly pine grows faster than shortleaf pine (except on the driest sites) and is generally preferred to shortleaf pine in artificial regeneration practices (Schultz 1997), more and more shortleaf pine has been replaced with

improved loblolly pine on intensively managed forest lands (South and Buckner 2003). The US Forest Service (USFS) is one of only a few organizations that regenerate shortleaf pine, usually with natural regeneration. As a result, the shortleaf pine stands regenerated by the USFS are becoming surrounded by more and more loblolly pine. Thus, it is reasonable to ask if the hybridization level is increasing in naturally regenerated shortleaf pine in areas surrounded by expanding loblolly pine plantings. The hybridization level may play a very important role in shortleaf or loblolly pine genetic integrity in the future. If we can estimate how intensive forest management affects hybridization levels, we can deduce how intensive loblolly pine management may affect both shortleaf pine and loblolly pine genetic integrity in the long term. Thus, the samples analyzed for this study were collected from Southwide Southern Pine Seed Source Study (SSPSSS) plantings where the trees grown were from seed collected in 1951 and 1952, when man's influence due to forest management was minimal. The hybridization levels of these samples will provide a reference or base level to evaluate the effects of current and future forest management. In addition, this information can serve to develop guidance for shortleaf pine and loblolly pine genetic conservation.

Materials and methods

Needles (leaves) and cones of shortleaf pine and loblolly pine were collected from 11 seed sources of each species (Fig. 1). The seed sources were established by collecting cones from 20 or more trees at each origin, and the resulting seeds were mixed. Trees grown from these seeds were planted into SSPSSS plantings (Wells and Wakeley 1970), which we subsequently sampled for this study. The seed sources and sample sizes are given in Table 1.

Collected needles and cones were placed in plastic bags and kept cool with blue ice in coolers during overnight shipment. Upon arrival in the laboratory, the needles were frozen at −80°C for later use. Cones were placed on laboratory benches to air dry. When the cones opened, the seeds were collected. The seeds were stored frozen at −20°C for later use.

AFLP analysis

Total DNA was extracted from needles of the shortleaf pine samples using a modified cetyltrimethylammonium bromide protocol (Doyle and Doyle 1988) of CGT's laboratory. A DNeasy Plant Mini Kit (Qiagen) was used for isolation of DNA from needle tissue of each loblolly pine sample. The primers and the AFLP marker development protocols described by Remington et al. (1999) and Remington and O'Malley (2000) were utilized in this study, including the use of *Eco*RI and *Mse*I restriction enzymes and the same 48 primer pairs. The primer pairs include all combinations of the selective nucleic acid sequences for *Eco*RI primers (5′-ACA-3′, 5′-ACC-3′, 5′-ACG-3′ and 5′-ACT-3′) with the selective nucleic acid sequences for *Mse*I primers (5′-CCAG-3′, 5′-CCCG-3′, 5′-CCGC-3′, 5′-CCGG-3′,

Table 1 The origin and sample size of the shortleaf pine and loblolly pine sources in this study

Shortleaf pine[a]				Loblolly pine[a]			
Source ID	State	County	No. of tress	Source ID	State	County	No. of tress
401[b]	PA	Franklin	4	303	NC	Onslow	9
419	MS	Lafayette	5	307	SC	Newberry	10
421	LA	St. Helena	5	311	GA	Clarke	10
423	TX	Angelina	7	317	AL	Clay	11
433	MO	Dent	8	321	MS	Prentiss	10
435	TN	Morgan	9	323	LA	Livingston	10
451[b]	PA	Franklin	10	327	AR	Clark	11
461	GA	Clarke	8	329	TN	Hardeman	10
475	TX	Cherokee	10	331	GA	Spalding	10
477	OK	Pushmataha & McCurtain	8	OSU[c]	OK	McCurtain	11
481	AR	Ashley	10	FL[d]	FL	Hernando & Citrus	10
487	TN	Anderson	9				

[a]Ninety-three shortleaf pine and 92 loblolly pine samples were collected by Oklahoma State University Forest Resources Center personnel, Idabel, OK, USA. In addition, ten loblolly pine samples (of comparable age to the SSPSSS samples, i.e., average age=56, range 34–67) from an allopatric region of Florida were provided by Gregory Powell, University of Florida, Gainesville, FL, USA

[b]Four hundred one belongs to the original collection made in 1951 and 451 to the collection made in 1955

[c]Not part of the SSPSSS, rather a local collection of equivalent age

[d]Present day collection from allopatric region

5′-CCTG-3′, 5′-CCAA-3′, 5′-CCAC-3′, 5′-CCCA-3′, 5′-CCGA-3′, 5′-CCTA-3′, 5′-CCTC-3′ and 5′-CCTT-3′).

A LI-COR 4300 DNA Analyzer was used for the AFLP fragment separation and detection. Because a single gel holds 64 samples, we had to load the 205 samples per primer pair onto four gels. Shortleaf pine Z15, loblolly pine SE631 (also designated GFC-631), and two of their controlled-pollinated progeny (i.e., F1 hybrids) were used as controls on each gel to ensure that the same loci were scored for all 205 samples loaded across the multiple gels. Shortleaf pine Z15, originally from North Carolina, was provided by Dr. Bruce Bongarten, formerly Warnell School of Forest Resources, University of Georgia and currently College of Environmental Science and Forestry, State University of New York, Syracuse, NY, USA. Loblolly pine SE631, originally from the west central Piedmont of Georgia, and the control-pollinated progeny (Hyb1 and Hyb2) were provided by CDN and Larry Lott of the Southern Institute of Forest Genetics, Saucier, MS, USA.

IDH analysis

In conifers, *IDH* is a codominant marker; thus, the haploid megagametophyte tissue of the germinating seed is the preferred tissue for analysis. For trees for which seeds were available (110 trees), ten germinated seeds from each tree were used to obtain megagametophyte tissue and the *IDH* genotype as described by Yeh and Layton (1979) and Huneycutt and Askew (1989). For the remaining samples (95 trees), needles were used for the *IDH* analysis. About 0.05 g of needle tissue from each tree was ground in the same buffer used for the megagametophytes tissue. The protocol used for the *IDH* analysis followed that of Raja et al. (1997).

Hybrid analysis

NewHybrids

The software NewHybrids version 1.1 beta (Anderson and Thompson 2002; Anderson 2003) was used to analyze the AFLP and *IDH* data to identify putative hybrids in the populations sampled. Both dominant and codominant markers can be used, although very few studies using NewHybrids with dominant markers have been reported (Anderson 2003). The software assumes that the samples come from a population where hybridization may have occurred and that a random sample of individuals is genotyped at multiple unlinked loci. The software considers six genotype categories: pure species 1, pure species 2, F1 hybrids, F2 hybrids, and the first backcross generation to pure species 1 or pure species 2, with the results giving the estimated probabilities that each individual belongs to each of the different genotype categories.

All 96 AFLP bands that were polymorphic across the two species were scored and used in the analysis. According to the software's author (Anderson, 2006, personal communication), it is not necessary to select species-specific markers and that optimal results are expected by using all markers that are polymorphic across the two species. To see how well or poorly the individuals of known category would be classified with the data collected, shortleaf pine Z15, loblolly pine SE631, and their controlled pollinated progeny (Hyb1 and Hyb2) were used as controls (described above). The z option and s option were used for the control trees, as their genotype categories were known and they were sampled separately from the 205 sample trees. The Markov Chain Monte Carlo (MCMC) analysis ran 42,000 sweeps after burn-in. The random number seeds were two and seven. The Jeffreys Prior was used for Θ (allele frequencies) and π (mixing proportions).

Structure

The program Structure v2.2 (Prichard et al. 2000; Falush et al. 2003, 2007) was also used to investigate the level of admixture or hybridization in the population samples. As both Structure and NewHybrids are relatively new software programs and use different analytical approaches and because confirmation or refutation of the results was essential, both programs were utilized.

All 205 samples trees were included in the Structure analysis. In addition, the four control trees (Z15, SE631, Hyb1, and Hyb2) were added to assist in calibrating the results. We expected that the two parents would fall into the parental species classes in terms of admixture level and that the hybrids would fall near the 50% admixture level, as they are control-pollinated F1 hybrids. The controls were used to calibrate thresholds for placing trees into specific hybrid categories. Thresholds were set by considering the theoretical admixture value of each hybrid category (i.e., F1 is 50%, BC1 is 75%, etc.) and then moving to an intermediate level between the categories. A level of 75% towards the next theoretical value was chosen, as it was the smallest proportion that accommodated the four controls (correctly classifying the two F1s as F1s and the two parents as pure species).

Four data sets were used in the Structure analyses. Two contained all 96 AFLP markers and two contained only those markers that appeared to be most informative for differentiating shortleaf from loblolly pine. These markers were selected based on their band frequencies in trees that were identified as either shortleaf or loblolly pine by their seed source. Markers that had band frequencies >0.7 and

<0.3 or vice versa in the two species or markers that had band frequency differences of >0.4 between the two species were selected. This subset of markers included 51 of the 96 AFLPs. In addition, one each of the datasets either included or excluded the *IDH* marker, resulting in datasets having 97, 96, 52, and 51 markers, each with 209 trees. For each data set, we assumed $K=2$ populations (essentially shortleaf pine and loblolly pine) and the 'Admixture' ancestry model and made four replicate runs using 10,000 burn-in and 40,000 MCMC iterations. Where no criteria were used to select markers, the 'Correlated' allele frequency model was assumed, while the 'Independent' model was used where markers were selected. In addition to these admixture analyses assuming $K=2$ populations, we made a series of runs to determine if we could pick up structure in the data sets beyond $K=2$. The same datasets and assumptions were used except that the probability of K values were computed and evaluated from $K=1$ to 6.

Results

Of the 48 AFLP primer pairs screened, 17 were selected based on their propensity to reveal polymorphic markers between shortleaf pine and loblolly pine (Table 2). The five most informative primer pairs produced nine or more AFLP markers each and together accounted for 65 of the 96 markers scored in this study. Figure 2 shows part of a typical AFLP gel picture produced by primer pair M-CCTCxE-ACG.

IDH results

A typical result for *IDH* using needle tissue is shown in Fig. 3, where the loblolly pine *IDH* allele migrates faster than the shortleaf pine *IDH* allele. Both homozygotes and heterozygotes for these two alleles of *IDH* can be clearly scored.

Assuming the *IDH* heterozygotes can be properly classified as hybrids, then, two loblolly pine hybrids (327-2 and 321-4) and two shortleaf pine hybrids (433-1 and 433-2) can be identified in our sample of 205 trees. We note that trees 433-1, 321-4, and 327-2 were found to be hybrids using seed tissue, and 433-2 was detected using needle tissue.

NewHybrids results

The 96 AFLP and the *IDH* data were analyzed by NewHybrids version 1.1 beta (Anderson and Thompson

Table 2 The 96 AFLPs polymorphic in both shortleaf pine and loblolly pine

Primer pairs	Number of markers	Markers
M-CCAGXE-ACG	22	S1[a] (80[b]), L1(95), A1(100), L2(120), L3(125), S2(145), L4(204), L5(230), L6(270), L7(275), A2(856), A3(99), A4(102), A5(104), A6(105), A7(110), A8(135), A9(140), A10(141), A11(148), A12(160), A13(240)
M-CCTGXE-ACG	12	S4(70), A24(80), A25(155), L9(204), L10(320), A27(78), A28(81), A29(101), A30(102), A31(120), A32(145), A33(254)
M-CCGAXE-ACG	12	L15(76), L16(90), A40(256), A41(300), A42(55), A43(98), A44(100), A45(105), A46(110), A47(120), A48(280), A49(290)
M-CCCGXE-ACG	10	A14(120), A15(215), L8(256), S3(270), A16(271), A17(946), A18(124), A19(200), A20(208), A21(255)
M-CCTCXE-ACC	9	A50(60), A51(75), L17(78), S7(80), A52(101), A53(142), A54(250), A55(68), A56(150)
M-CCGAXE-ACC	6	L13(70), L14(100), A36(80), A37(90), A38(125), A39(150)
M-CCGGXE-ACT	4	A58(145), S9(254), A59(55), A60(145)
M-CCTCXE-ACG	4	A65(111), A66(180), L19(345), L20(160)
M-CCTGXE-ACC	3	S5(105), L11(225), A26(275)
M-CCCAXE-ACG	3	L18(230), A61(80), A62(125)
M-CCTAXE-ACG	3	A63(90), S10(120), A64(142)
M-CCCGXE-ACA	2	A22(120), A23(133)
M-CCCAXE-ACT	2	L12(165), A35(202)
M-CCCAXE-ACT	1	S6(155)
M-CCCAXE-ACC	1	A34(144)
M-CCTTXE-ACC	1	S8(90)
M-CCGCXE-ACT	1	A57(150)

[a]Name of the marker
[b]Estimated size of the marker

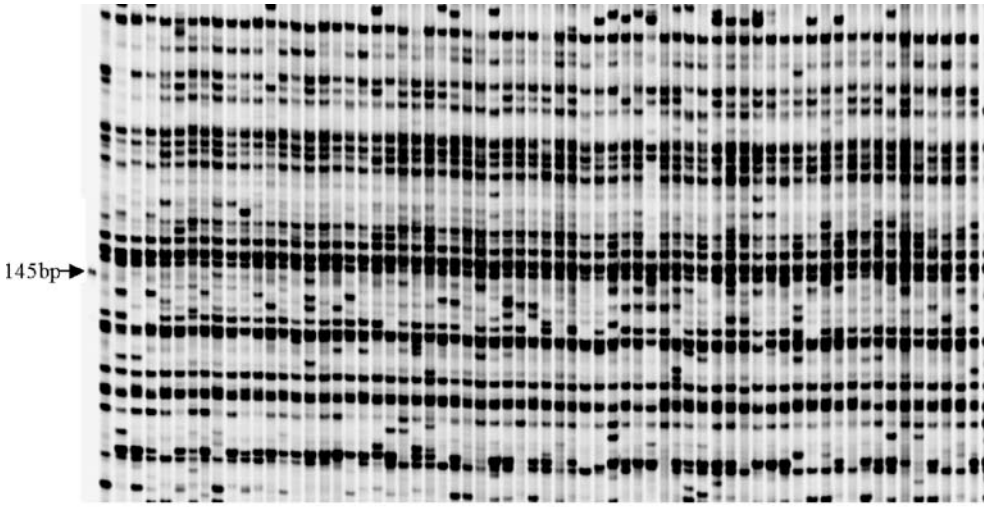

Fig. 2 A part of the AFLP gel picture produced by primer pair M-CCTCxE-ACG

The 1st lane: a molecular standard; the 2nd lane: shortleaf pine Z15; the 3rd lane: loblolly pine 631; the 4th and 5th lanes: hybrids between Z15 and 631; the remaining lanes: samples from the SSPSSS collection.

2002; Anderson 2003). The results clearly identify tree 433-2 as a hybrid, with a 0.96 probability of being a backcross to shortleaf pine (B1S). The tree 433-2 was also identified as a hybrid with *IDH*, and it is the only tree identified as hybrid by both *IDH* and the NewHybrids analysis. The other three *IDH* classified hybrids were categorized as follows: 327-2 and 321-4 both loblolly pine (prob=0.99) and 481-7 shortleaf pine (prob=0.98).

The NewHybrids analysis found 13 hybrids (using the Structure cut-off values, see "Materials and methods" and below), nine in the shortleaf pine samples and four in the loblolly pine samples (Table 3). None of the hybrids identified were F1s; rather, they were of later generation hybrid origin. Of the nine shortleaf pine hybrids, two came from Dent, MO, USA (source 433), three were from Ashley, AR, USA (source 481), two were from Franklin, PA, USA (source 451), and one each was from Angelina, TX, USA (source 423) and Pushmataha, OK, USA (source 477). As seen in Fig. 1, shortleaf pine seed sources 423, 433, 477, and 481 are located west of Mississippi River. Except for the two hybrids from Franklin (source 451), all the hybrids found in the shortleaf pine samples in this study were from west of the Mississippi River. Accordingly, the shortleaf pine hybridization rate is 16.3% (7/43) west of the river and 2.4% (2/50) east of the river.

Loblolly pine hybrids identified included one tree each from Onslow, NC, USA, Newberry, SC, USA, Prentiss, MS, USA, and Clark, AR, USA (sources 303, 307, 321 and 327, respectively). One hybrid is from west of the Mississippi River, and three are from east of the river. Thus, the hybridization rates in the loblolly pine population are 4.5% (1/22) west of the river and 3.3% (3/90) east of the river. For all samples across species, the hybridization rates are 7.7% (5/65) in western populations and 0.7% (1/140) in eastern populations. The hybridization rates for shortleaf pine and loblolly pine range-wide as sampled in this study are 9.7% (9/93) and 3.6% (4/112), respectively. In total, 6.3% (13/205) of the samples were hybrids.

Structure results

The log probability of *K* plots from the Structure v2.2 analysis clearly indicated that two populations are present in the sample. The results over all *K* values were very similar for the four data sets; therefore, only the results of the full data set are reported in this paper. For the most part,

Fig. 3 A needle tissue based *IDH* stained starch gel identifying a hybrid individual

Lane 1: a natural hybrid: lanes 2 to 8 and 12 to 17: loblolly pine samples: line 9. 10 and 11: shortleaf pine samples: arrow: indicts migration direction of *IDH*.

Table 3 Results from the Structure analysis showing the first and last tree in the complete data set (obs 1 and 209), 12 putative hybrid trees, and the two F1 hybrid controls (Hyb1 and Hyb2), in italics, compared to results from the *IDH* and NewHybrid analyses

Samples[a]		Structure[b]		IDH[c]	NewHybrids[d]					
obs	tree	prlob	type	IDH	Prlob	prsl	prF1	prF2	prBs	prBL
1	Z15	0.001	Sht	S	0.000	0.999	0.000	0.000	0.00 0	0.000
//										
73	433-1	0.008	Sht	H	0.000	0.996	0.000	0.000	0.003	0.000
//										
83	477-1	0.018	Sht	S	0.000	0.947	0.000	0.000	0.053	0.000
84	423-3	0.026	Sht	S	0.000	0.976	0.000	0.000	0.024	0.000
85	*433-5*	*0.043*	*B3S*	S	0.000	0.927	0.000	0.000	0.073	0.000
87	*451-6*	*0.050*	*B3S*	S	0.000	0.898	0.000	0.000	0.102	0.000
86	*423-1*	*0.045*	*B3S*	S	0.000	0.816	0.000	0.000	0.184	0.000
88	*451-7*	*0.070*	*B3S*	S	0.000	0.691	0.000	0.000	0.309	0.000
89	*481-5*	*0.089*	*B2S*	S	0.000	0.576	0.000	0.001	0.423	0.000
90	*477-8*	*0.099*	*B2S*	S	0.000	0.575	0.000	0.001	0.424	0.000
92	*481-9*	*0.214*	*B1S*	S	0.000	0.265	0.000	0.007	0.728	0.000
93	*433-2*	*0.265*	*B1S*	H	0.000	0.011	0.000	0.025	0.964	0.000
94	*481-7*	*0.289*	*B1S*	S	0.000	0.027	0.000	0.055	0.918	0.000
95	*Hyb1*	*0.496*	*F1x*	H	0.000	0.000	0.072	0.653	0.365	0.029
96	*Hyb2*	*0.679*	*F1x*	H	0.003	0.000	0.353	0.220	0.011	0.413
97	*307-4*	*0.691*	*B1L*	L	0.442	0.000	0.000	0.347	0.000	0.211
98	*327-10*	*0.802*	*B1L*	L	0.877	0.000	0.000	0.081	0.000	0.042
99	*303-3*	*0.899*	*B2L*	L	0.959	0.000	0.000	0.001	0.000	0.041
100	321-9	0.964	Lob	L	0.950	0.000	0.000	0.000	0.000	0.049
//										
110	327-2	0.989	Lob	H	0.998m	0.000	0.000	0.000	0.000	0.002
//										
121	321-4	0.994	Lob	H	0.999	0.000	0.000	0.000	0.000	0.000
122	SE631	0.994	Lob	L	0.999	0.000	0.000	0.000	0.000	0.001
//										
209	OSU-7	0.999	Lob	L	0.999	0.000	0.000	0.000	0.000	0.000

The tree nearest to being classified as hybrid by Structure on both the shortleaf pine and loblolly pine ends are given (obs=84 and 100, respectively), as are the F1 hybrid parents (obs 1, tree Z15, and obs 122, tree SE631), and the four *IDH* heterozygous individuals (obs 73, 93, 110, and 121)

[a]Trees sampled, where obs=rank (1=most shortleaf-like, 209=most loblolly-like) and tree=tree identification (seed source number-tree number)

[b]Structure results, where prlob=mean (n=4 structure runs, full data) estimated proportion of genome from loblolly and type=tree type as determined by prlob, with Sht=shortleaf pine, B3S, B2S, and B1S=third, second, and first generation backcross to shortleaf pine, respectively, F1x=artificial F1 hybrid, B1L and B2L=first and second generation backcross to loblolly pine, and *Lob*=loblolly pine

[c]IDH results, where S=homozygous shortleaf pine alleles, H=heterozygous, L=homozygous loblolly pine alleles

[d]NewHybrids results, where prlob/prsl/prF1/prF2/prBs/prBl=percent probability that a tree is loblolly/shortleaf/F1/F2/backcross1/backcross2, respectively

the trees were clearly grouped by species as expected by their seed source origin; however, a few trees were intermediate or admixed and thus are apparently hybrids. Table 3 lists the trees identified as hybrid and the proportion of the genome of each tree estimated to be of loblolly pine origin. The cut-offs estimated as the proportion of genome from loblolly pine were F1, 0.6875 upper and 0.3125 lower; BC1, 0.8437 upper and 0.1562 lower; BC2, 0.9219 upper and 0.07812 lower; and BC3, 0.9609 upper and 0.0391 lower.

Of the 205 sample trees, 12 were classified as hybrids, with no trees being first generation hybrids (i.e., F1), five trees being second generation (i.e., BC1), three trees being third generation (i.e., BC2), and four trees being fourth generation (i.e., BC3). Except for loblolly pine 321-9, from the Prentiss source, the trees identified as hybrid by the Structure analysis were exactly the same 12 trees identified as hybrid by the NewHybirds analysis. Nine of the 12 hybrid trees were from shortleaf pine seed sources, while the other three were from loblolly pine sources.

The *IDH* marker results compared to the Structure results were the same as for the NewHybrids results. Tree 433-2 was the only individual identified as hybrid by the *IDH* marker, which was also identified as hybrid by both NewHybrids and Structure.

Discussion

The *IDH* isoenzyme is a codominant marker reported as useful in detecting hybrids between shortleaf pine and loblolly pine (Huneycutt and Askew 1989; Edwards et al. 1997; Chen et al. 2004). To date, it is the only locus thought to be fixed for different alleles between these two species, thus efficiently detecting F1 hybrids. However, Mendelian genetics dictates that 50% of the F2 hybrids will be homozygous at the *IDH* locus as well as in the first backcross generation. Implicit, then, is that more than 50% of the later generation hybrids will be homozygous at the *IDH* locus. Therefore, more markers are needed to reliably identify both early and later generation hybrids. This study demonstrated that AFLPs can identify later generation hybrids that cannot be detected by *IDH* alone. Furthermore, the value of the *IDH* marker for reliably detecting even F1 hybrids is called into question. The genome proportions for the four trees heterozygous for *IDH* are given in Table 3. One of these trees is clearly shortleaf pine, two trees are clearly loblolly pine, and only one is in the intermediate hybrid range. Two each of the *IDH* heterozygous trees were from shortleaf and loblolly pine seed sources, respectively. The two shortleaf pines were from Missouri (source 433), while the loblolly pine sources are from north Mississippi (source 321) and Arkansas (source 327). Some of these seed sources also included true hybrids, so either these alternate species *IDH* alleles are selectively favored and have survived many generations of backcrossing or they naturally were or have become resident alleles at a low frequency in some populations of the other species. Unfortunately, the distinction between these alternate explanations of the *IDH* heterozygotes in the pure species trees is not apparent.

Clearly, there are hybrid individuals in some populations of both shortleaf pine and loblolly pine. The *IDH* locus will identify some of these, but not all, and an *IDH* heterozygote is not necessarily a hybrid. This suggests that earlier estimates of the level of hybridization between these two species by Chen et al. (2004), Raja et al. (1997), and Edwards et al. (1997) are in question. The lack of F1s in the populations in this study, as explained below, might suggest this study's estimate of the level on hybridization is low.

The full data set (96AFLPs and *IDH*) proved to be much more reliable than the *IDH* marker alone. Both NewHybrids and Structure definitively identified the same group of 12 trees as hybrids. The single tree difference between the two analyses (tree 321-9) was in the B3L category, i.e., a fourth generation hybrid, and it is not surprising that the loblolly genome percentage estimate at this stage of backcrossing could vary by the few percentage points between the two analysis. Chen et al. (2004) and Edwards et al. (1997) also presented evidence that some hybrids found in their studies might be later generation hybrids.

It can be argued that F1 individuals would not be expected in these two population samples. The seed collection trees resulting in the seedlings planted in these seed source tests were originally selected to be representative of the species. It is fair to assume that hybrid like individuals would have been avoided both during seed collection and in the nursery grading process, although detection of hybrids in the seedling stage is difficult. It is perhaps more surprising, given this collection process, that the number of hybrids found was relatively high.

We also note the presence of a tendency for the backcrosses to be made to shortleaf pine. In addition to the higher number of B2S and B3S trees (6) compared to B2L and B3L trees (1), there is also a notable break in genome proportion distribution for individuals at the loblolly end of the distribution compared to the shortleaf end (Table 3). This suggests a tendency for preferential backcrossing to shortleaf pine. Nine of the 12 hybrid trees were from shortleaf pine seed sources, while the other three were from loblolly pine sources.

This study found a relatively high level (30%) of hybridization between shortleaf pine and loblolly pine in Arkansas. These results are higher than previous studies (15% by Raja et al. 1997; 14% by Chen et al. 2004), but their results were based on the *IDH* marker, which we have shown to be unreliable and to not identify all hybrid trees in a population. It is worth noting that even though Raja et al. (1997) and this study used samples from SSPSSS plantings (trees originally selected to represent the species), the hybridization level of some seed sources were surprisingly high, 25% in Missouri and 30% in Arkansas in this study and 34% in southeastern Arkansas in the study of Raja et al. (unpublished data). One possible explanation is that even though the trees were originally selected based on their morphological traits, later generation hybrids, in particular backcrosses, often have morphology similar to the backcross parents (Edwards et al. 1997; Chen et al. 2004). It is conceivable that such hybrids were selected as representative of shortleaf or loblolly pine, either as a seed parent or as a seedling from the nursery.

This study agrees with previous studies (Edwards et al. 1997; Raja et al. 1997; Chen et al. 2004) that the hybridization level, particularly in shortleaf pine, is higher in populations west of the Mississippi River than eastern populations. In the current study, the shortleaf pine hybridization rates were 16.3% (7/43) west of the river and 2.4% (2/50) east of the river and, in loblolly pine populations, 4.5% (1/22) and 3.3% (3/90) west and east of the river, respectively. These shortleaf pine estimates are in agreement with estimated levels of 15% by Raja et al.

(1997) and 14% by Chen et al. (2004). Edwards and Hamrick (1995) reported a hybridization level at 4.6% west of the Mississippi River vs. 1.1% east of the river in shortleaf pine. The different percentages of hybridization level reported in different studies may be due to sample size, sample location, time of sampling, and/or the presence or absence of the alternate species allele as resident in the population sampled. The results presented in this study suggest that all previous estimates of hybridization levels between shortleaf pine and loblolly pine based on the *IDH* alleles are not reliable.

The shortleaf pine hybrids found in this study were from Missouri, Texas, Oklahoma, Pennsylvania, and Arkansas. Edwards and Hamrick (1995) also detected shortleaf pine hybrids in Missouri, Oklahoma, and Arkansas, as well as from Mississippi, Alabama, Georgia, South Carolina, and Virginia at a lower frequency. In this study, that of Raja et al. (1997, unpublished data), and that of Edwards and Hamrick (1995) hybrids were found in shortleaf pine populations far north of any natural loblolly pine populations. Schmidtling et al. (2005) pointed to two possible reasons leading to hybrids in allopatraic shortleaf pine populations. The first explanation was that gene flow could be due to long-distance pollen transport from south to north. The second one was that loblolly pine ranged farther north 5,000 to 7,000 years ago because the climate was warmer during the Hypsithermal geological period. Possibly, the apparently later generation hybrids or simply the resident alternate allele found in the allopatric region today result from F1 hybrids formed during the Hypsithermal geological period.

The loblolly pine hybrids found in this study occurred as scattered individuals, one each in South Carolina, North Carolina, Mississippi, and Arkansas. The total number of hybrids was low, but there did not seem to be a relationship between the location of the shortleaf pine hybrids relative to the location of the loblolly pine hybrids. The authors are not aware of other surveys of shortleaf × loblolly pine hybrids in the loblolly pine species.

The hybridization frequency between the two species varies among populations from different places in all the studies. All studies, however, concluded that hybrids do occur at what may be relatively high levels, and this study provides confirmation of those conclusions. This relatively high hybridization level in some locations may have forest management implications when relying on natural regeneration of shortleaf pine if current practices are leading to increased levels of hybridization in remaining native populations.

It is important to note that *IDH* is not definitive in the identification of hybrid individuals between shortleaf pine and loblolly pine. Any individual tree heterozygous for *IDH* may or may not be a hybrid. Further study of *IDH* to determine why such a highly conserved gene differs between these two similar pine species and to elucidate its apparent persistence in the alternate form in both shortleaf and loblolly pine populations would be of interest.

References

Anderson EC (2003) User's guide to the program NewHybrids version 1.1 beta. http://ib.berkeley.edu/laboratorys/slatkin/eriq/software/software.htm. Cited 25 Aug 2003

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. Genetics 160:1217–1229

Chen JW, Tauer CG, Bai G, Huang Y, Payton ME, Holley AG (2004) Bidirectional introgression between *Pinus taeda* and *Pinus echinata*: Evidence from morphological and molecular data. Can J For Res 34:2508–2516

Cotton MH, Hicks RR, Flake RH (1975) Morphological variability among loblolly and shortleaf pines of east Texas with reference to natural hybridization. Castanea 40:309–319

Dorman KW, Barber JC (1956) Time of flowering and seed ripening in southern pines. USDA Forest Service, Southeastern Forest Experiment Station, Station Paper 72.

Doyle JJ, Doyle J (1988) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Edwards MA, Hamrick JL (1995) Genetic variation in shortleaf pine, *Pinus echinata* Mill. (Pinaceae). For Genet 2:21–28

Edwards MA, Hamrick JL, Price RA (1997) Frequency and direction of hybridization in sympatric populations of *Pinus taeda* and *P. echinata* (Pinaceae). Am J Bot 84:879–886

Falush D, Stephen M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Falush D, Stephen M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7:574–578

Hare RC, Switzer GL (1969) Introgression with shortleaf pine may explain rust resistance in western loblolly pine. USDA Forest Service, Research Note SO-88. Southern Forest Experiment Station, New Orleans, LA. pp2

Henry BW, Bercaw TE (1956) Shortleaf-loblolly hybrid pines free of fusiform rust after 5 years' exposure. J For 54:779

Hicks RR Jr. (1973) Evaluation of morphological characters for use in identifying loblolly pine, shortleaf pine and loblolly shortleaf hybrids. Castanea 38:182–189

Huneycutt M, Askew GR (1989) Electrophoretic identification of loblolly pine shortleaf pine hybrids. Silvae Genet 38:95–96

Islam-Faridi, MN, Majid MA, Nelson CD (2007) Chromosomal locations of the ribosomal DNA genes in shortleaf pine. In: Kabrick JM, Dey DC, Gwaze D (eds) Shortleaf pine restoration and ecology in the Ozarks. Proceedings of a symposium. Gen Tech Rep NRS-P-15 USDA Forest Service No Res Sta 215 p.

Mergen F, Stairs GR, Snyder EB (1965) Natural and controlled loblolly shortleaf pine hybids in Mississippi. Forest Sci 11: 306–314

Prichard JK, Stephens M, Donnelly P (2000) Inference of population structure using mulilocus genotype data. Genetics 155:945–959

Raja RG, Tauer CG, Wittwer RF, Huang YH (1997) Isoenzyme variation and genetic structure in natural populations of shortleaf pine (*Pinus echinata* Mill.). Can J For Res 27:740–749

Raja RG, Tauer CG, Wittwer RF, Huang Y (1998) Regeneration methods affect genetic variation and structure in shortleaf pine (*Pinus echinata* Mill.). For Genet 5:171–178

Remington DL, O'Malley DM (2000) Whole-genome characterization of embryonic stage inbreeding depression in a selfed loblolly pine family. Genetics 155:337–348

Remington DL, Whetten RW, Liu BH, O'Malley DM (1999) Construction of an AFLP genetic map with nearly complete genome coverage in *Pinus taeda*. Theor Appl Genet 98: 1279–1292

Saylor LC (1972) Karyotype analysis of the genus *Pinus*-subgenus *Pinus*. Silvae Genet 21:155–163

Schmidtling RC, Myszewski JH, McDaniel CE (2005) Geographic variation in shortleaf pine (Pinus echinata Mill.)—cortical monoterpenes. Proceedings to Southern Forest Tree Improvement Conference vol. 28, Jun. 20–23, Raleigh, North Carolina, pp. 161–167

Schreiner EJ (1937) Improvement of forest trees. 1937 Yearbook of Agriculture, USDA pp 1242–1279

Schultz RP (1997) Loblolly pine: the ecology and culture of loblolly pine (*Pinus taeda* L.). USDA Ag Handb 713, pp. 11–24 and 11–25

South DB, Buckner ER (2003) The decline of southern yellow pine timberland. J For 101:30–35

Wells OO, Wakeley PC (1970) Variation in shortleaf pine from several geographic sources. For Sci 16:415–423

Yeh FC, Layton C (1979) The organization of genetic variability in central and marginal populations of lodgepole pine *Pinus contorta spp. latifolia*. Can J Genet Cytol 21:487–503

Zobel BJ (1953) Are there natural loblolly-shortleaf pine hybrids? J For 51:494–495