Anchored Reference Loci in Loblolly Pine (*Pinus taeda* L.) for Integrating Pine Genomics

Garth R. Brown,*,† Edward E. Kadel III,* Daniel L. Bassoni,* Kristine L. Kiehne,* Berhanu Temesgen,*,[1] J. P. van Buijtenen,[2] Mitchell M. Sewell,*,[3] Kimberly A. Marshall* and David B. Neale*,†

*Institute of Forest Genetics, Pacific Southwest Research Station, U.S. Department of Agriculture Forest Service, Davis, California 95616 and †Department of Environmental Horticulture, University of California, Davis, California 95616

Manuscript received October 15, 2000
Accepted for publication July 12, 2001

ABSTRACT

Anchored reference loci provide a framework for comparative mapping. They are landmarks to denote conserved chromosomal segments, allowing the synthesis of genetic maps from multiple sources. We evaluated 90 expressed sequence tag polymorphisms (ESTPs) from loblolly pine (*Pinus taeda* L.) for this function. Primer sets were assayed for amplification and polymorphism in six pedigrees, representing two subgenera of Pinus and a distant member of the Pinaceae, Douglas-fir (*Pseudotsuga menziesii* [Mirb.] Franco). On average, 89% of primer sets amplified in four species of subgenus Pinus, 49% in one species of subgenus Strobus, and 22% in Douglas-fir. Polymorphisms were detected for 37–61% of the ESTPs within each pedigree. Comparative mapping in loblolly and slash pine (*P. elliottii* Englm.) revealed that ESTPs mapped to the same location. Disrupted synteny or significant disruptions in colinearity were not detected. Thirty-five ESTPs met criteria established for anchor loci. The majority of those that did not meet these criteria were excluded when map location was known in only a single species. Anchor loci provide a unifying tool for the community, facilitating the creation of a "generic" pine map and serving as a foundation for studies on genome organization and evolution.

COMPARATIVE genome analysis has revealed a remarkable conservation of gene order in species from diverse mammalian orders and in families of major agronomic plants. For example, the rice genetic map can be divided into a set of linked genes, known as linkage blocks, that can be shuffled to represent the genetic maps of other cereal genomes (Moore *et al.* 1995). Conserved linkage relationships allow the multidirectional transfer of genetic information among species and the prospect of integrating knowledge of DNA sequence and allele variation, biochemistry, metabolism, and physiology with phenotypic information. Also important are the insights into genome organization and evolution provided by comparative genome analysis (Lagercrantz and Lydiate 1996; Kellogg 1998).

Comparative mapping requires a genetic map from different species, each consisting of common markers that are necessary for map alignment. For map comparisons to be meaningful, a common marker must detect the orthologous locus in each species, which can be established by DNA and amino acid sequence homology, conserved map location, and ultimately, functional complementation. Markers based on expressed sequences, such as restriction fragment length polymorphisms (RFLPs) detected by cDNA probes and, more recently, by polymerase chain reaction (PCR)-based markers derived from expressed sequence tags (ESTs), have been widely used in plants and mammals to "anchor" maps of different species. The public availability of sets of anchor loci (O'Brien *et al.* 1993; Lyons *et al.* 1997; Van Deynze *et al.* 1998; Davis *et al.* 1999) has led to the rapid advancement of comparative mapping in mammals and plants. Other marker types, such as random amplified polymorphic DNAs (RAPDs) and amplified fragment length polymorphisms (AFLPs), have limited use in comparative mapping because they detect polymorphisms predominantly in noncoding regions that are poorly conserved among species. Similarly, simple sequence repeats (SSRs) appear to have limited utility beyond a narrow range of related pine species (Echt *et al.* 1999).

The genus Pinus (the pines) is composed of about 100 species representing 20% of all gymnosperms. Pines are the most widespread tree genus in the Northern Hemisphere and are of major ecological and economic

Corresponding author: David B. Neale, Institute of Forest Genetics, Pacific Southwest Research Station, U.S. Department of Agriculture Forest Service, Department of Environmental Horticulture, University of California, 1 Shields Ave., Davis, CA 95616.
E-mail: dneale@dendrome.ucdavis.edu

[1] *Present address:* California Department of Justice, DNA Laboratory, Berkeley, CA 94710.

[2] *Present address:* Department of Forest Science, Texas A&M University, College Station, TX 77843.

[3] *Present address:* Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831.

TABLE 1

Mapping populations of the Conifer Comparative Genomics Project

Genus	Subgenus	Section	Subsection	Species	Source
Pinus	Pinus	Pinus	Australes	*P. taeda*	Weyerhaeuser Co.
				P. elliottii	J. P. van Buijtenen, Texas A&M University
			Sylvestres	*P. pinaster*	C. Plomion, INRA[a]
				P. sylvestris	O. Savolainen, University of Oulu
			Oocarpae	*P. radiata*	C. Echt, FR[b]
	Strobus	Strobus	Strobi	*P. lambertiana*	Institute of Forest Genetics[c]
Pseudotsuga	NA	NA	NA	*P. menziesii*	Weyerhaeuser Co.

[a] Institut National de la Recherche Agronomique, Cestas, France.
[b] Forest Research, Rotorua, New Zealand.
[c] USDA Forest Service, Pacific Southwest Research Station, Davis, California.

significance. Although the two main subgenera, Pinus (hard pines) and Strobus (soft pines), diversified by the end of the Cretaceous (66 mya), cytological studies have shown little karyotype differentation within the genus. All pines are diploid, have a haploid chromosome number of 12, and have similar complements of median to submedian chromosomes (Sax and Sax 1933; Pederick 1970). These characteristics are noted commonly throughout the Pinaceae and suggest that a high degree of colinearity between homologous chromosomes of pines and other conifers might exist. A demonstration of conserved gene orders could have profound implications to the manner and pace in which genome research in conifers is conducted. For example, the genetic map of a "map-rich" species could be used to select evenly spaced markers for systematically creating maps in less-characterized species.

Genetic maps based on RFLPs detected by cDNA probes have been constructed for loblolly pine (*Pinus taeda* L.; Devey *et al.* 1994; Groover *et al.* 1994; Sewell *et al.* 1999) and Monterey pine (*P. radiata* D. Don; Devey *et al.* 1999). Devey *et al.* (1999) conducted a comparative genetic analysis of these two species on the basis of 60 shared RFLP markers and found no evidence of chromosome rearrangement. Previously, Ahuja *et al.* (1994) demonstrated the feasibility of using mapped loblolly pine cDNAs as hybridization probes in a range of conifers. All tested clones hybridized to the five pine species studied, and a large number also hybridized to four other genera of the Pinaceae. Although these studies provided a communal set of anchor loci, RFLP technology is not widely used in conifers. This is in part due to the large genome size of conifers in general [the *C* value of loblolly pine is 21–23 pg (Wakamiya *et al.* 1993)] and the abundance of multigene families (Kinlaw and Neale 1997). In addition, the comparable simplicity of PCR-based markers and the use of half-sib mapping populations derived from limited quantities of haploid endosperm (megagametophyte) tissue has led to the creation of numerous genetic maps of conifer species based on RAPDs, AFLPs, and SSRs (Cervera *et al.* 2000). Therefore, the majority of current conifer genetic maps are essentially species- or even pedigree-specific, with little extractable information for comparative analyses.

The availability of thousands of loblolly pine ESTs presents the opportunity to develop new anchor loci for conifers to support comparative genome analysis. Temesgen *et al.* (2001) developed a strategy for identifying and mapping genetic polymorphisms in PCR products of loblolly pine pedigrees generated by primer sets derived from EST sequence data. It included efforts to reduce the amplification of multiple members of gene families by placing one primer within or near the 3′-untranslated region (UTR) and optimizing the melting profile of the expected product for denaturing gradient gel electrophoresis (DGGE) by adding GC clamps (Meyers *et al.* 1985). This study evaluates 90 polymorphic ESTs (ESTPs) derived from loblolly pine for use as anchored reference loci to facilitate the integration of genetic mapping efforts of the Conifer Comparative Genomics Project (CCGP; http://dendrome.ucdavis.edu/Synteny). Five pine pedigrees, including one of slash pine (*P. elliottii* Engelm.), maritime pine (*P. pinaster* Ait.), Scots pine (*P. sylvestris* L.), Monterey pine, and sugar pine (*P. lambertiana* Dougl.), and Douglas-fir (*Pseudotsuga menziesii* [Mirb.] Franco) were tested for PCR amplification and segregating ESTPs. A genetic map of slash pine was also constructed by using ESTP, RFLP, RAPD, and isozyme markers, and a comparative analysis with loblolly pine was performed. This work demonstrates the feasibility of unifying the isolated mapping programs of pines and other conifers through a common marker system.

MATERIALS AND METHODS

Mapping populations: Most ESTPs in loblolly pine were mapped genetically using two immortalized reference mapping populations, referred to as the *base* and *qtl* pedigrees

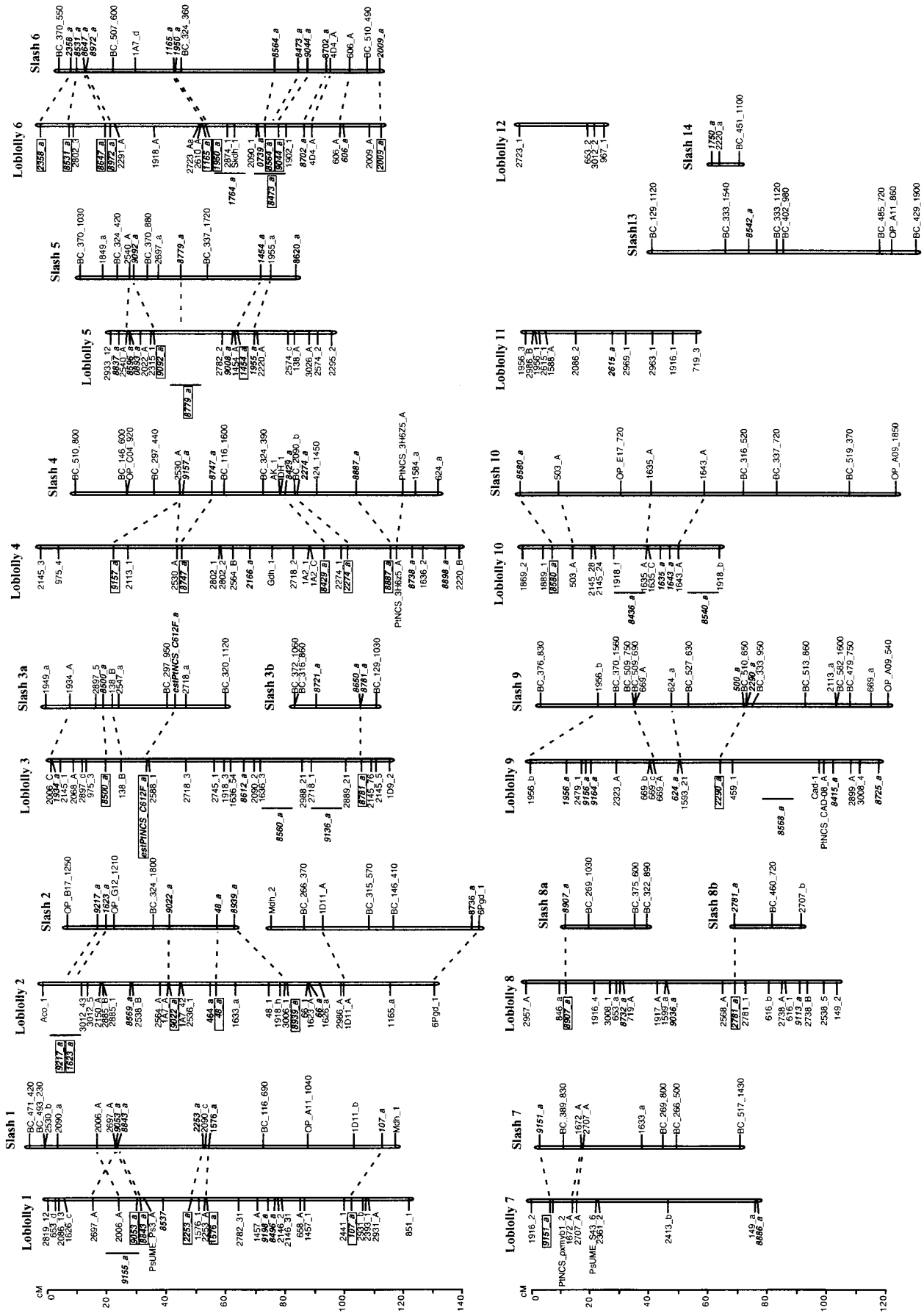
TABLE 2

Clone number, accession number, and primer sequences of 38 ESTPs developed for this study

Clone[a]	Accession no.[b]	Forward primer	Reverse primer
PtIFG_8419	AA739501	GCAACCTCCCATACCAAGAC	cgcgcggACCTTGTTGTGGATGGCG
PtIFG_8429	AA739505	GAGGCATTTATGAGGGAACG	cgcgcggGTTGAAAGCGACTCCAAAGG
PtIFG_8436	AA739508	TGAACTTGTCGGTGCTTAGG	cgcgcggGACAAGGATCCAACTCTCAGC
PtIFG_8471	AA739529	ggcccggcggAGTCAGAGGCCATGTTTTGG	ggcTGAACCACTACATTCCCCTCC
PtIFG_8473	AA739526	ccggccgggCTCTGGAGAACGACTGCAAG	TACAGTAGCCTCTCTGGCCC
PtIFG_8480	AA739530	AGTGGCCAAAGAGATATGGG	gcggccggTATTTCTTGCCCGAAGGC
PtIFG_8496	AA739978	cccggccgcGATCGGCGGTGGCATAAG	AAATGTCATGTACGGACGGG
PtIFG_8500	AA739539	ggcccggcggGCAGATCGGACGATTAAAGG	gccTCTGTACAAAACCGGATGGG
PtIFG_8520	AA739553	CGGAGTCATCTGTGTTTTGG	GTCGGACATAAACTCATTTCTG
PtIFG_8531	AA739558	cccggccgcgTTCAGCTGGAAACACCTCAC	CTCGTGATAGCACAGCAAATACTC
PtIFG_8537	AA739563	cccggccgcgCATTCATTTGAGTGGCTGGG	TAAAGTTCAGAACAATCCAACG
PtIFG_8540	AA739564	ggcccggcggGCACTGCTACTCCATTCTCG	ggccgcAACTGTGAAGCAATGGGGAC
PtIFG_8542	AA739566	ggcccggcggTTGGCATATGGAGGCATG	gcGATGCCCAAAGCATGACATC
PtIFG_8560	AA739577	ggcccggcggccTGTTAGGGATGAGTATGCGC	ggccgCACTCTCTCAGGAAAAGCTCG
PtIFG_8568	AA739584	ggcccggcggTGAGGCCACTAGAGGAGACTG	ggccgTCCAATAGCAATTATAGCCAGC
PtIFG_8580	AA739884	ACTGGATTCCGGAGGATCAC	TGGAAACCGTCTACAGTCGC
PtIFG_8596	AA739594	cccggCTGCTGATGATCGAATCTGG	AGCTAAAGTTGGCTGGCATC
PtIFG_8612	AA739606	GAAGGGCACTATGAAGCTGC	gcggccggAACTAGGAATCCCAAATTCCC
PtIFG_8620	AA739611	gccgTGGATTGATAGGTCGGATGC	ggcccggcggCATTCGCATGGGAACATTG
PtIFG_8647	AA739625	ggccgggTTGGTCCGCTGATTGGAG	ACTTACGGTGGAGACCTTACAC
PtIFG_8650	AA739628	cccggccgccggcccCTCTCAAGGAGCCGAACAAG	ACC CTT GGA TTT TCC CGT C
PtIFG_8656	AA739632	gggcggcggcccAAGGTGTAATGGAAGGGTGC	ccggcgggAAAGATATCATTGCTGGCTATAC
PtIFG_8721	AA739669	cccggccgccggcccTGCCCAAGTGCAATCATG	CGCTTGCTCAGATCATGG
PtIFG_8725	AA739673	AGCGCTGAATGATGTCTTGG	gccgggcccggcCCAAACTTACACCATGCTCG
PtIFG_8732	AA739680	ggcccggcggccggTGAAGTTCTGAGTTTGGCGG	gccATGCACAGGATGAACCCTTC
PtIFG_8736	AA739683	ggcccggcggccggTGGCATCACCAACCCAAC	gcTTCCAAATCGAGCGAAAATAG
PtIFG_8738	AA739685	gTCACAGACCTGAACACTGCG	ggcccggcggCCAAAACCGACTATCTTGGG
PtIFG_8747	AA739692	GCGGTTGGAAAATGTGCTAG	ggcccggcggGCTGCAATGCCAAGTACATC
PtIFG_8779	AA739708	ccggccgggAGGATTCGCGGTGAAAGG	TCAAGTCAGTTGCTTGAACATC
PtIFG_8781	AA739709	ccggccgggGGTTTGCTCACATGTTCGTG	TCCTTCATAAATGCATTCTCGTC
PtIFG_8837	AA739754	ccggccgggTCGGTTACTTGGGACCTACTG	TCCCATGCAATATCATTTGTG
PtIFG_8843	AA739759	ccggccgggTCGCTGATGGCTTGAAGG	CCACCGAAATATATGCCTGTC
PtIFG_9136	AA739940	gccggccgggcCCTTGTCCAAGACTGTAAG	CCGATTTGAAACACACATTC
PtIFG_9155	AA739957	gcgcccgGTTGGTGTTAGGCAGTCATGG	TGAAATTCCCAGCCCATATC
PtIFG_9164	AA739966	CAGCAAAGAGGCTTCAAAGG	gcggccTACTGCGAATGCAACTCCAG
PtIFG_9198	AA739984	CGGCGGTGGCATAAGTTAC	gcggccgggTCCAAGTCTTCTCAATCCGG
PtIFG_9217	AA740001	cccggccTCCAACTGTACAACAGCACCC	TTCAAGCTTGATCCACAACG
PtNCS_6C12F	AA556811	GCCTGCTATCGAATCCAGAA	CCAGACAACCCAAATGAAGG

[a] Clone refers to the loblolly pine xylem cDNA from which the EST sequence was derived. Primer sequences are given 5′ to 3′. Lowercase letters denote nucleotides added as GC clamps and are not homologous to the corresponding EST sequence.

[b] GenBank accession number.


Slash 1
Loblolly 1
Slash 2
Loblolly 2
Slash 3a
Slash 3b
Loblolly 3
Slash 4
Loblolly 4
Slash 5
Loblolly 5
Slash 6
Loblolly 6
Slash 7
Loblolly 7
Slash 8a
Slash 8b
Loblolly 8
Slash 9
Loblolly 9
Slash 10
Loblolly 10
Loblolly 11
Loblolly 12
Slash13
Slash 14
cM

(Temesgen *et al.* 2001). DNA samples and segregation data are publicly available for these populations as a comparative genome analysis resource (see http://dendrome.ucdavis.edu/Synteny/refmap.html). ESTPs that did not segregate in either reference population were also tested in a third pedigree, the *prediction* pedigree belonging to the Weyerhaeuser Company. A two-generation mapping population of slash pine maintained by the Texas Forest Service was derived from the mating of seed parent D4PC40 and pollen parent D4PC13. After germination, the megagametophytes were removed from F_1 seeds for use in RAPD analysis and the seedlings were grown. Ninety-two F_1 seedlings were planted near College Station, Texas. DNA was extracted from megagametophytes and needles using the FastPrep System (Qbiogene, Carlsbad, CA) and a modified hexadecyltrimethylammonium bromide procedure (Devey *et al.* 1991), respectively. Members of the CCGP provided DNA samples of other conifer pedigrees examined (Table 1).

Genetic markers: The majority of EST primer sets assayed were reported in Temesgen *et al.* (2001). These were designed to amplify fragments from (1) single- or low-copy cDNAs from a library of random-primed poly(A) RNA isolated from loblolly pine needle tissue (with accession nos. <PtIFG_3000) or (2) the 3′ end of cDNAs, including a portion of the presumed 3′-UTR, from a directionally cloned library of loblolly pine xylem poly(A) RNA (with accession nos. >PtIFG_8000). Thirty-eight new primer sets (Table 2) targeting the 3′-UTR were also designed using methods described in Temesgen *et al.* (2001). EST sequences are available at http://web.ahc.umn.edu/biodata/nsfpine/ and through GenBank. Primer sequences for PtNCS_6C12F were provided by C. S. Echt (unpublished results).

For simplicity, one standard PCR reaction mix and cycling regime, including a hot start and a touchdown (Harry *et al.* 1998), was used to amplify genomic DNA of all species. PCR was considered to have failed in a species when repeated attempts produced either no product or weak or inconsistent amplification. Amplification products from the parents of each pedigree were screened for polymorphisms by one or more electrophoretic methods, including 2% agarose gels, 4 or 10% nondenaturing polyacrylamide gels (PAGE), and denaturing gradient gels. DGGE was performed according to Temesgen *et al.* (2001). A DCODE Universal Mutation Detection System apparatus (Bio-Rad, Hercules, CA) was used for both PAGE and DGGE. Segregation of putative polymorphisms in all pedigrees was confirmed by subsequent analysis of PCR products from six progeny templates.

Three additional sources of genotypic data were obtained for the slash pine mapping population to provide sufficient markers for linkage analysis. RFLP analysis was performed as described in Devey *et al.* (1991) using cDNA and genomic DNA probes previously mapped in loblolly pine. Clones that revealed simple hybridization patterns in loblolly pine DNA were chosen preferentially. Most clones were also tested in Monterey pine by Devey *et al.* (1999). RAPD analysis was performed according to methods described by Nelson *et al.* (1993) on haploid megagametophyte DNA of the F_1 seeds. A subset of 69 RAPD primers segregating in D4PC40 and spaced at 10- to 20-cM intervals were selected from previous work (J. P. van Buijtenen, unpublished results). Isozyme analysis was performed on vegetative buds of the mapping population. Protein extraction was performed according to Neale *et al.* (1984) and electrophoresis followed the methods of Conkle *et al.* (1982).

DNA sequencing: A comparative sequence analysis was perfomed for 10 ESTP amplification products from loblolly and slash pines to assess the homology between putatively orthologous loci. For each ESTP, a single allele was sequenced from loblolly pine haploid megagametophyte tissue. Diploid templates of a homozygous individual were sequenced preferentially in slash pine. DNA sequences were generated from both strands with the primers used for PCR amplification and the BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA). Fragments were detected on an ABI 377 DNA Sequencer.

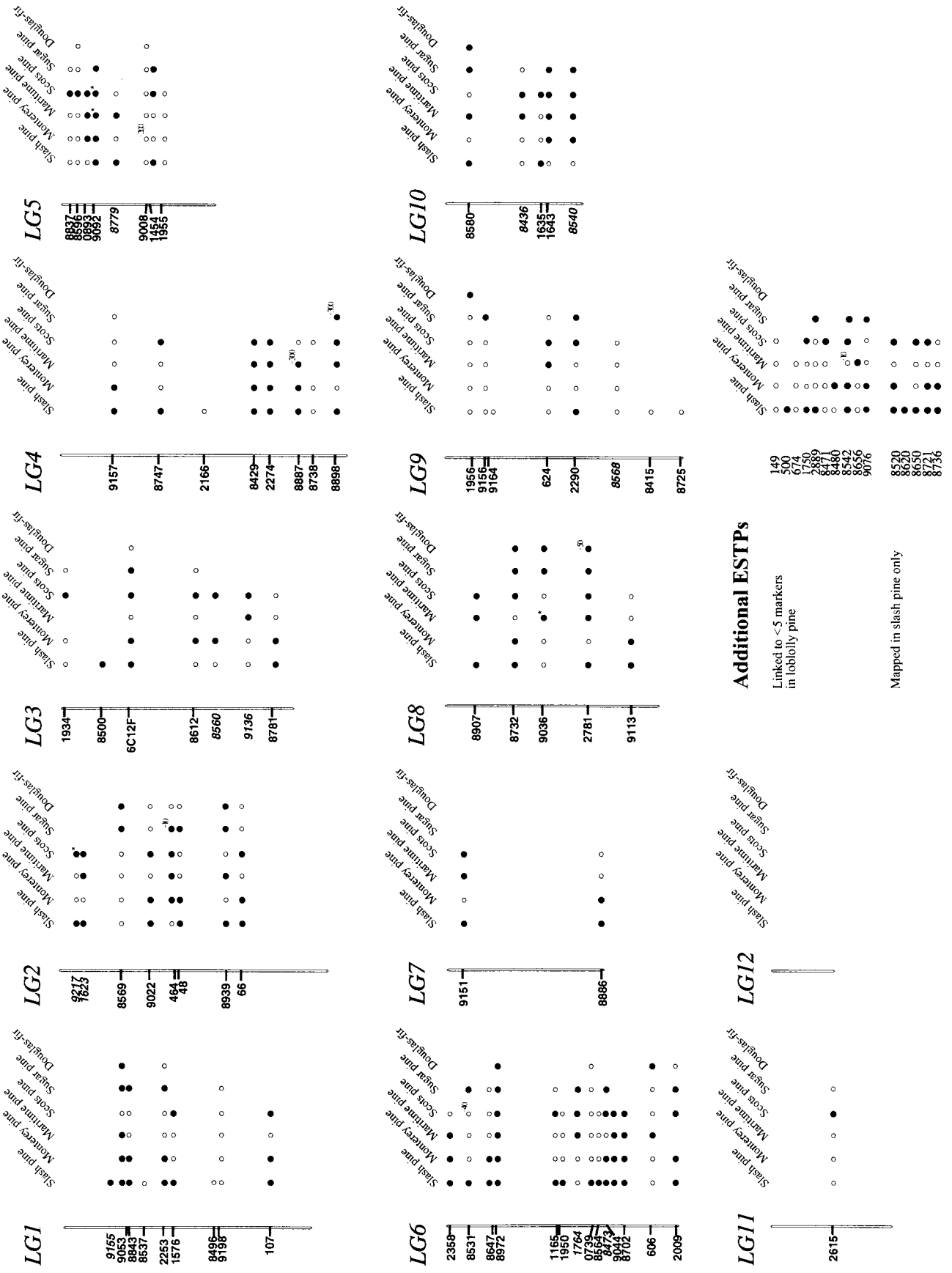
Linkage analysis: Genotypic data were scored visually by two independent readers. Markers that deviated from the expected segregation ratio at the 5% significance level were not eliminated since some deviations from Mendelian ratios are expected in pedigrees of this size (Remington *et al.* 1999).

A reference genetic map of loblolly pine was constructed from all available genotypic data. A consensus map of the *base* and *qtl* pedigrees was generated with Mapmaker (Unix version 3.0; Lander *et al.* 1987; Lincoln and Lander 1992) and JoinMap (Unix version 1.4; Stam 1993). The methods developed by Sewell *et al.* (1999) were employed, although only framework markers from the original data set and ESTPs were retained to avoid minor discrepancies in marker ordering that can arise when analyzing a large number of linked loci. Framework markers within each set of parental meioses were defined as those spaced at ≥10-cM intervals, having <10% missing data, and with interval support ≥3. Preference was also given to markers segregating in both parents and/or both pedigrees to facilitate map integration. Linkage analysis of markers segregating in the *prediction* pedigree was performed separately using JoinMap only. A relative assignment on the consensus map of the *base* and *qtl* pedigrees of ESTPs segregating in the *prediction* pedigree was determined by aligning homologous RFLP markers flanking the ESTPs. Linkage analysis in slash pine was performed essentially as described in Sewell *et al.* (1999). Because of the smaller data set, a LOD ≥3 was used in the JoinMap analysis.

Selection of anchored reference loci: An informal definition of orthology was chosen to be inclusive during the early development of PCR-based anchor loci in pines. An orthologous anchor locus was classified as a primer set that amplified a single locus with known map location in the reference species loblolly pine. Such a primer set must also amplify a single locus of similar size in other species. Amplification products from different species must be highly homologous, as determined directly by DNA sequencing or by inference from similar mobilities on denaturing gradient gels. Finally, the map location of an anchored reference locus must be conserved between loblolly pine and other species.

Several primer sets reported by Temesgen *et al.* (2001) were

Figure 1.—Genetic maps of loblolly and slash pine. Loblolly pine linkage groups are on the left; slash pine on the right. Loci in boldface and italic type are loblolly pine ESTPs. ESTPs to the left of the loblolly pine linkage groups were mapped in the *prediction* pedigree and their positions are estimated only. Loci connected by a dotted line were detected by the same marker and boxed loci denote ESTPs meeting criteria for anchor loci. Experiment fields (IFGREF, loblolly pine reference map; IFGELL, slash pine linkage map) have been omitted. Source fields for ESTP and RFLP markers have also been omitted unless the clone originated from a laboratory other than IFG. Centimorgan distances (Kosambi) are indicated on the scale to the left.


LG1
LG2
LG3
LG4
LG5
LG6
LG7
LG8
LG9
LG10
LG11
LG12
Slash pine
Monterey pine
Martinne pine
Scots pine
Sugar pine
Douglas-fir
Additional ESTPs
Linked to <5 markers in loblolly pine
Mapped in slash pine only

described as producing complex banding patterns after DGGE. Primers yielding a high level of background staining with a superimposed band that was present or absent were excluded from this analysis. However, primer sets that produced multiple fragments without background staining after DGGE were retained at this early stage for evaluation.

Nomenclature and informatics: The locus nomenclature used is according to guidelines for submitting data to TreeGenes, the forest tree genome database (http://dendrome.ucdavis.edu/TreeGenes). A mapped object is defined by its experiment, source, accession number, and locus identifier fields. For example, an ESTP derived from PtIFG_9053 and mapped in slash pine in this study is referenced as IFGELL_ estPtIFG_9053_a. The experiment field for markers on the loblolly pine framework map is IFGREF. For brevity, however, experiment fields have been omitted. Source fields are included only to denote clones from laboratories other than the Institute of Forest Genetics (IFG) and the supplier of RAPD primers. A capitalized locus identifier (*e.g.,* IFGREF_ PtIFG_2006_A) was given to RFLP markers that detected the orthologous locus in both loblolly pine pedigrees or in both loblolly pine and slash pine.

RESULTS

Loblolly pine reference genetic map: A genetic map of framework markers was constructed to serve as the reference for comparative analyses among pine genomes (Figure 1). The 12 linkage groups (LGs) consisted of 155 RFLPs, 75 ESTPs, and 5 isozyme loci spanning 1165 cM(K). The genetic length is similar to the conservative estimate of 1227 cM(K) from Sewell *et al.* (1999), indicating that the overall length and integrity of the map had been maintained. LGs 1–11 corresponded to those reported by Sewell *et al.* (1999) with the incorporation of LG 12/*consensus* onto LG 9 and LG 17/*qtl-pat* onto LG 8. Several slight discrepancies in ordering involved only tightly linked markers. The numbering of LG 12, equivalent to LG 14/*consensus* of Sewell *et al.* (1999), was consistent with the comparative analysis of loblolly and Monterey pines described by Devey *et al.* (1999).

Five to 14 ESTPs were mapped on each of LGs 1–6 and LGs 8–10 (Figure 1). Only two ESTPs were mapped to LG 7, one to LG 11, and none mapped to LG 12. Ten ESTPs were linked to less than five markers at LOD 4.

Amplification and detection of polymorphisms in other conifers: The amplification of 90 EST primer sets and subsequent detection of polymorphisms among the CCGP pedigrees are depicted graphically (Figure 2). All primer sets amplified the DNA of slash pine, the closest relative of loblolly pine studied here. Among three other hard pines, 75–79 primer sets amplified genomic DNA, suggesting that loblolly pine EST primer sets would have broad utility within the subgenus Pinus. Only 44 primer sets amplified sugar pine DNA, a representative of the subgenus Strobus. Douglas-fir templates were amplified by only 20 primer sets, a value similar to that reported by Perry and Bousquet (1998) for amplification across genera using Picea-derived EST primers. For most species/ESTP combinations, amplification products were of the same size, as estimated from agarose gels. In seven instances, amplification in another species produced a fragment from 10 to 300 bp larger, indicating that an insertion had occurred. An additional fragment not present in loblolly pine was observed in four cases, suggesting that a nonorthologous fragment was likely amplified.

Polymorphisms were detected by agarose gel electrophoresis, nondenaturing PAGE, or DGGE for an average of 52% of ESTP primer sets among all species. Maritime pine was the least polymorphic (37%) and slash pine was the most polymorphic (61%; Figure 3). If synteny and colinearity are conserved among pine species, the distribution of ESTPs across the majority of homologous linkage groups in the hard pines is sufficient to allow a low-resolution comparative analysis. Coverage of most linkage groups in sugar pine and Douglas-fir is currently scant.

DNA sequence similarity: It is expected that orthologous loci exhibit high nucleotide similarity between species. Ten ESTPs, ranging from 180 to 457 bp in length, were amplified from one individual of loblolly and slash pines for comparative sequence analysis (Table 3). One to five base substitutions between species were observed and no insertions or deletions were detected. Nucleotide identity averaged 99.4%, strongly supporting that the amplified loci were not paralogs.

Sequencing of all amplification products from different species was beyond the scope of this research. However, since the mobility of a DNA fragment during DGGE is dependent on the sequence and its melting properties, and not length, the mobility of amplification products from different species can be used to infer homology. Similar DGGE mobilities were observed for the majority of ESTP amplification products from all species (Figure 4). The few exceptions derived from the amplification of a significantly different fragment size (*e.g.*, PtIFG_8887 in maritime pine) or of additional fragments not produced in loblolly pine (*e.g.*, PtIFG_9036 and PtIFG_9217 in maritime pine).

Figure 2.—Representation of the potential utility of loblolly pine ESTPs as anchored reference loci. The 12 linkage groups of loblolly pine are shown (LGs 1–12). Successful PCR amplification in a species is shown by a circle; solid circles denote amplification products that are polymorphic in the pedigree examined, and open circles denote monomorphic products. Superscripts indicate amplification products of a different size than observed in loblolly pine. Asterisks indicate the amplification of an additional fragment not observed in loblolly pine. ESTPs mapped in the prediction pedigree are shown in italics and their location on the consensus map is approximated. All loci, with the exception of IFGREF_estPtNCS_C612F_a, are prefixed by IFGREF_estPtIFG.

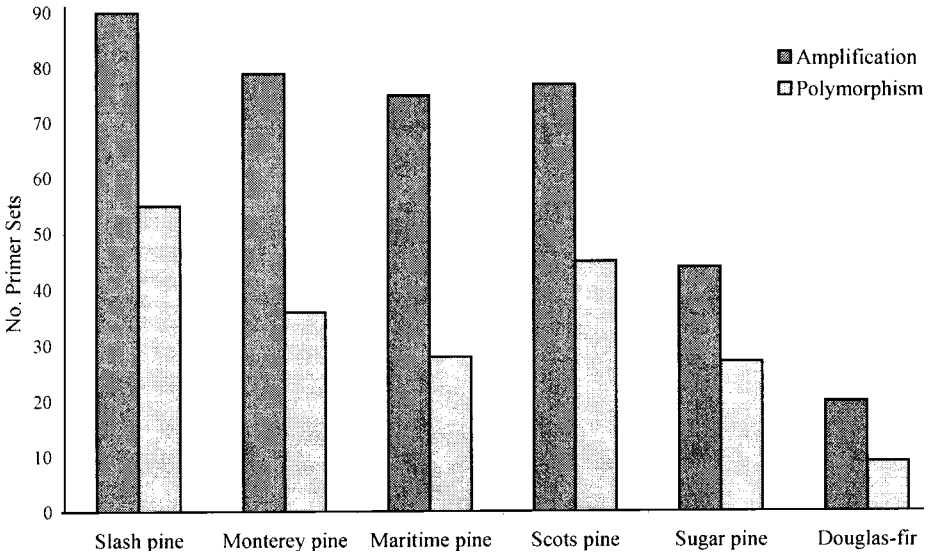


FIGURE 3.—The number of loblolly pine ESTP primer sets producing an amplification product in other species and the number of segregating polymorphisms detected by gel-based methods in the CCGP pedigrees. A total of 90 ESTP primer sets were tested.

Linkage map of slash pine: The genetic segregation data set for slash pine consisted of 170 loci (52 ESTP, 44 RFLP, 69 RAPD, and five isozyme markers) segregating in one full-sib family. Selection of framework markers was limited to the much larger maternal set of meioses since paternal groups consisted of two or three linked markers only. A linkage map consisting of 154 loci (45 ESTP, 41 RFLP, 63 RAPD, and five isozymes) distributed across 15 linkage groups was constructed (Figure 1). Linkage groups ranged from 10 to 124 cM(K) in length with a total map distance of 1115 cM(K). Unlinked markers and marker pairs were excluded.

Comparative mapping in loblolly and slash pine: Based on 60 putatively orthologous ESTP, RFLP, and isozyme markers mapped in both loblolly and slash pine, 9 homologous linkage groups were identified (Figure 1). Slash pine LGs 1–7, 9, and 10 were numbered in accordance with those of the loblolly pine framework map and the Monterey pine genetic map (DEVEY *et al.* 1999). Several large blocks of syntenic and colinear markers, particularly on LGs 2 and 6, were evident. The order of markers on other linkage groups was identical between species except for two discrepancies involving marker pairs on LGs 1 and 4. Both pairs consisted of one paternally informative locus and one locus segregating in both parents in the slash pine population. Therefore, the discrepancies may have arisen from having too few shared markers for parental map integration by JoinMap rather than from genuine disruptions in colinearity.

LGs 3a, 8a, and 8b had only one shared ESTP between slash pine and loblolly pine. The slash pine LGs were tentatively assigned to their homologous linkage group in loblolly pine by the position of estPtIFG_8781_a, estPtIFG_8907_a, and estPtIFG_2781_a, respectively. However, the orientation of markers flanking the ESTP is unknown. Amplification products of estPtIFG_8907 and estPtIFG_2781 were among those sequenced in both species. A single base substitution between species was detected for both ESTPs, supporting these two assignments. The two remaining linkage groups (slash pine LGs 13 and 14) could not be distinguished as homologs of loblolly pine LGs 11 and 12.

Anchored reference loci in pines: Of the 90 ESTPs in loblolly pine, 35 met the criteria established for an anchored reference locus. estPtIFG_48, estPtIFG_8473, and estPtIFG_8647, which mapped to the same location in loblolly and slash pine but revealed multiple amplification products after DGGE, were also provisionally included. The primary reason for excluding an ESTP as an anchored reference locus was that it had been mapped in only a single species. Mapping of ESTPs segregating in other CCGP pedigrees is in progress and should lead to the inclusion of most of the 90 primer sets reported here.

TABLE 3

Nucleotide identity between loblolly and slash pine at 10 ESTP loci

ESTP	Nucleotide identity (%)
estPtIFG_107	422/423 (99.8)
estPtIFG_1950	455/457 (99.6)
estPtIFG_2253	360/365 (98.6)
estPtIFG_2781	447/448 (99.8)
estPtIFG_8500	222/223 (99.6)
estPtIFG_8580	179/180 (99.4)
estPtIFG_8907	287/288 (99.7)
estPtIFG_8972	303/307 (98.7)
estPtIFG_9053	282/283 (99.6)
estPtIFG_9151	290/291 (99.7)
	Mean = 99.4%

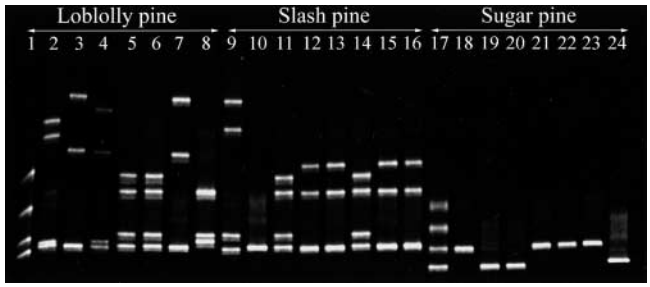


FIGURE 4.—Denaturing gradient gel showing segregation and similar mobility of homoduplexes in different species. DNA fragments amplified by PtIFG_9053 primers in loblolly pine (lanes 1–8), slash pine (lanes 9–16), and sugar pine (lanes 17–24) are shown. Lanes 1 and 9 are the maternal parents, lanes 2 and 10 are the paternal parents, and lanes 3–8 and 11–16 are F_1 progeny. Lane 17 is sugar pine seed tree 5701 (diploid) and lanes 18–24 are products from haploid megagametophytes of tree 5701 seeds. Homoduplexes are the faster one or two migrating bands of the diploid template reactions.

DISCUSSION

The international genome mapping community in forestry is small and many different tree species are involved. Integrating these efforts through a common marker system and shared mapping populations is an important next step in the future progress of forest tree genomics. Elements of the infrastructure needed are now available to begin the systematic comparative analyses of pine genomes. These include the reference anchor loci and other ESTPs described, a consortium through which biological resources are exchanged, and a common nomenclature to facilitate the bioinformatics of comparative mapping.

It is essential that only sets of orthologous loci be used as anchor loci for comparative mapping. Comparative sequence analysis provides *a priori* knowledge of orthology for highly conserved genes. Sequence divergence between loblolly and slash pines was extremely low in a sample of 10 reference anchor loci, supporting that the orthologous locus had been amplified and mapped in both species. Sequence analysis is not without limitations, however, since for genes that are poorly conserved across species, the appropriate level of sequence similarity distinguishing paralogs from orthologs is a subjective decision. Conserved map location as the sole criterion must be evaluated carefully since it may be a circular argument for orthology and lead to a biased view of genome conservation. Although most loblolly pine anchor loci are defined by conserved map position and only inferred sequence homology, the extensive evidence for gene colinearity in vertebrates and plants leads us to believe that most assumptions of orthology will prove to be valid. Nevertheless, the assumptions made must be considered a working hypothesis (ANDERSSON *et al.* 1996) and may be disproved as additional evidence accumulates.

The results of TEMESGEN *et al.* (2001) reflect those of PERRY and BOUSQUET (1998) in that very similar members of a gene family are sufficiently diverged in their 3′-UTRs to allow PCR amplification of individual genes. Our sequence analysis further suggests that the 3′-UTRs of orthologous loci in closely related species, such as loblolly and slash pines, are more similar to one another than to 3′-UTRs of parologs within a species. This has yet to be investigated beyond the subsection Australes, but may be more widely applicable to the subgenus Pinus given that the majority of loblolly pine primer pairs amplify DNA of the hard pines tested. Targeting PCR amplification toward the 3′-UTR promises to provide a suitable number of orthologous loci for more detailed comparative mapping within the subgenus.

The efforts of the CCGP will be broadened to encompass the taxonomic range of the Pinaceae. Improvement in the performance of loblolly pine primers across subgeneric and generic boundaries could be achieved by targeting highly conserved regions of genes. This is a useful strategy for single-copy genes, in which case amplifying the orthologous locus in multiple species is assured. However, the number of single-copy genes in the large, complex genomes of pines is limited, and lower levels of sequence polymorphism in conserved regions could prevent the genetic mapping of many markers. For more distantly related species, comparative mapping with PCR-based markers is inherently limited by the homology requirements between primers and annealing sites. As GALE and DEVOS (1998) have pointed out, had PCR been invented 5 years earlier, and as a result RFLP technology used much less extensively for comparative mapping, the extent of gene order conservation in plants may still be unknown. Therefore, to expand beyond the subgenus Pinus will require optimizing PCR conditions for each primer set and development of several hundred more ESTPs. Other members of the CCGP have developed high throughput methods to map ESTPs (CATO *et al.* 2001), which will considerably augment the number of available anchor loci in pines.

The low-resolution comparisons of loblolly pine with Monterey pine (DEVEY *et al.* 1999) and slash pine have indicated no disrupted synteny or significant disruptions in gene colinearity. At this level of resolution, only the largest rearrangements, if present, would be detected. However, the primary role of the reference anchor loci is in recognizing and identifying homologous linkage groups, thereby establishing a framework by which the location of genes in one species can be predictive of their location in the others. Important traits encoded by single genes, including disease resistance and crown shape, have been identified in tree species with less advanced genome projects than loblolly pine (DEVEY *et al.* 1995; LEHNER *et al.* 1995). Identifying the homologous region of the loblolly pine genome

could lead to sequencing of large insert clones mapped to the region and the identification of candidate genes for those traits. In like manner, markers flanking quantitative trait loci (QTL) affecting wood quality traits in loblolly pine (Groover *et al.* 1994; Sewell *et al.* 2001) could be an entrance for similar studies in other forest trees. Comparative mapping of QTL across several species is one means of QTL validation, which is an important step toward their incorporation into breeding strategies.

Selection of the best anchored reference loci for pines and the Pinaceae is a dynamic process and the list of markers reported here will be updated and continually improved. Ultimately, the criteria will include the following: (1) strong PCR amplification of a single product from both Pinus subgenera and at least one other genus of Pinaceae, (2) known genetic map location in loblolly pine and at least one other Pinaceae species, (3) uniform genome coverage, (4) known DNA sequence in several species, and (5) putative function. Developing anchor loci should rely on community involvement. We encourage the use of the reference mapping populations to place additional gene-based markers on the loblolly pine consensus map and to evaluate them for inclusion as anchor loci. The synthesis of genetic information in pines and other conifers could lead to the creation of a generic map and the practical applications of comparative mapping that follow.

We thank the members of the Conifer Comparative Genomics Project for their participation and contributions, Valerie Hipkins and the U.S. Department of Agriculture (USDA) Forest Service-National Forest Gel Electrophoresis Laboratory for isozyme analysis in slash pine, Nicholas Wheeler, Clem Lambeth, and the Weyerhaeuser Co. for providing the loblolly pine reference mapping populations, and Geoffrey Gill for DNA sequencing and analysis. Funding for this project was provided by the USDA Plant Genome National Research Initiative (USDA NRI grant no. 95-37300-1632). Trade names and commercial products or enterprises are mentioned solely for information and no endorsement by the USDA is implied.

LITERATURE CITED

Ahuja, M. R., M. E. Devey, A. T. Groover, K. D. Jermstad and D. B. Neale, 1994 Mapped DNA probes from loblolly pine can be used for restriction fragment length polymorphism mapping in other conifers. Theor. Appl. Genet. **88:** 279–282.

Andersson, A., M. Archibald, S. Ashburner, S. Audun, J. Barendse *et al.*, 1996 Comparative maps of vertebrates. Mamm. Genome **7:** 717–734.

Cato, S. A., R. C. Gardner, J. Kent and T. E. Richardson, 2001 A rapid PCR-based method for genetically mapping ESTs. Theor. Appl. Genet. **102:** 296–306.

Cervera, M. T., C. Plomion and C. Malpica, 2000 Molecular markers and genome mapping in woody plants, pp. 375–394 in *Molecular Biology of Woody Plants,* edited by S. M. Jain and S. C. Minocha. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Conkle, M. T., P. D. Hodgkiss, L. B. Nunnally and S. C. Hunter, 1982 *Starch Gel Electrophoresis of Conifer Seeds: A Laboratory Manual,* Gen. Tech. Rep. PSW-64. Pacific Southwest Forest and Range Experiment Station, USDA Forest Service, Berkeley, CA.

Davis, G. L., M. D. McMullen, C. Baysdorfer, T. Musket, D. Grant *et al.*, 1999 A maize map standard with sequenced core markers, grass genome reference points, and 932 expressed sequence tagged sites (ESTs) in a 1736 locus map. Genetics **152:** 1137–1172.

Devey, M. E., K. D. Jermstad, C. G. Tauer and D. B. Neale, 1991 Inheritance of RFLP loci in a loblolly pine three-generation pedigree. Theor. Appl. Genet. **83:** 238–242.

Devey, M. E., T. A. Fiddler, B.-H. Liu, S. J. Knapp and D. B. Neale, 1994 An RFLP linkage map for loblolly pine based on a three-generation outbred pedigree. Theor. Appl. Genet. **88:** 273–278.

Devey, M. E., A. Delfino-Mix, B. B. Kinloch, Jr. and D. B. Neale, 1995 Random amplified polymorphic DNA markers tightly linked to a gene for resistance to white pine blister rust in sugar pine. Proc. Natl. Acad. Sci. USA **92:** 2066–2070.

Devey, M. E., M. M. Sewell, T. L. Uren and D. B. Neale, 1999 Comparative mapping in loblolly and radiata pine using RFLP and microsatellite markers. Theor. Appl. Genet. **99:** 656–662.

Echt, C. S., G. G. Vendramin, C. D. Nelson and P. Marquart, 1999 Microsatellite DNA as shared genetic markers among conifer species. Can. J. For. Res. **29:** 365–371.

Gale, M. D., and K. M. Devos, 1998 Comparative genetics in the grasses. Proc. Natl. Acad. Sci. USA **95:** 1971–1974.

Groover, A., M. E. Devey, J. Lee, R. Megraw, T. Mitchell-Olds *et al.*, 1994 Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. Genetics **138:** 1293–1300.

Harry, D. E., B. Temesgen and D. B. Neale, 1998 Codominant PCR-based markers for *Pinus taeda* developed from mapped cDNA clones. Theor. Appl. Genet. **97:** 327–336.

Kellogg, E. A., 1998 Relationships of cereal crops and other grasses. Proc. Natl. Acad. Sci. USA **95:** 2005–2010.

Kinlaw, C. S., and D. B. Neale, 1997 Complex gene families in pine genomes. Trends Plant Sci. **2:** 356–359.

Lagercrantz, U., and D. Lydiate, 1996 Comparative genome mapping in Brassica. Genetics **144:** 1903–1910.

Lander, E., P. Green, J. Abrahamson, A. Barlow, M. Daley *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181.

Lehner, A., M. A. Campbell, N. C. Wheeler, T. Poykko, J. Glossl *et al.*, 1995 Identification of a RAPD marker linked to the pendula gene in Norway spruce (*Picea abies* (L.) Karst. f. pendula). Theor. Appl. Genet. **91:** 1092–1094.

Lincoln, S., and E. Lander, 1992 Systematic detection of errors in genetic linkage data. Genomics **14:** 604–610.

Lyons, L. A., T. F. Laughlin, N. G. Copeland, N. A. Jenkins, J. E. Womack *et al.*, 1997 Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat. Genet. **15:** 47–56.

Meyers, R. M., S. G. Fischer, T. Maniatis and L. S. Lerman, 1985 Modification of the melting properties of duplex DNA by attachment of a GC-rich DNA sequence as determined by denaturing gradient gel electrophoresis. Nucleic Acids Res. **13:** 3131–3145.

Moore, G., K. M. Devos, Z. Wang and M. D. Gale, 1995 Grasses, line up and form a circle. Curr. Biol. **5:** 737–739.

Neale, D. B., J. C. Weber and W. T. Adams, 1984 Inheritance of needle tissue isozymes in Douglas fir. Can. J. Genet. Cytol. **26:** 459–468.

Nelson, C. D., W. L. Nance and R. L. Doudrick, 1993 A partial genetic map of slash pine (*Pinus elliottii* Englm. var. *elliotti*) based on random amplified polymorphic DNAs. Theor. Appl. Genet. **87:** 145–151.

O'Brien, S. J., J. E. Womack, L. A. Lyons, K. J. Moore, N. A. Jenkins *et al.*, 1993 Anchored reference loci for comparative genome mapping in mammals. Nat. Genet. **3:** 103–112.

Pederick, L. A., 1970 Chromosome relationships between *Pinus* species. Silvae Genet. **19:** 171–180.

Perry, D. J., and J. Bousquet, 1998 Sequence-tagged sites (STS) markers of arbitrary genes: the utility of black spruce-derived STS primers in other conifers. Theor. Appl. Genet. **97:** 735–743.

Remington, D. L., R. W. Whetten, B.-H. Liu and D. M. O'Malley, 1999 Construction of an AFLP genetic map with nearly complete genome coverage in *Pinus taeda.* Theor. Appl. Genet. **98:** 1279–1292.

Sax, K., and H. J. Sax, 1933 Chomosome number and morphology in conifers. J. Arnold. Arbor. **14:** 356–375.

Sewell, M. M., B. K. Sherman and D. B. Neale, 1999 A consensus map for loblolly pine (*Pinus taeda* L.). I. Construction and integra-

tion of individual linkage maps from two outbred three-generation pedigree. Genetics **151:** 321–330.

Sewell, M. M., D. L. Bassoni, R. A. Megraw, N. C. Wheeler and D. B. Neale, 2001 Identication of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). I. Physical wood properties. Theor. Appl. Genet. **101:** 1273–1281.

Stam, P., 1993 Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J. **3:** 739–744.

Temesgen, B., G. R. Brown, D. E. Harry, C. S. Kinlaw, M. M. Sewell *et al.*, 2001 Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). Theor. Appl. Genet. **102:** 664–675.

Van Deynze, A. E., M. E. Sorrells, W. D. Park, N. M. Ayres, H. Fu *et al.*, 1998 Anchor probes for comparative mapping of grass genera. Theor. Appl. Genet. **97:** 356–369.

Wakamiya, I. R., R. J. Newton, J. S. Johnston and H. J. Price, 1993 Genome size and environmental factors in the genus *Pinus*. Am. J. Bot. **80:** 1235–1241.

Communicating editor: O. Savolainen