

High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.)

Andrew J. Eckert · Barnaly Pande · Elhan S. Ersoz ·
Mark H. Wright · Vanessa K. Rashbrook ·
Charles M. Nicolet · David B. Neale

Received: 20 March 2008 / Revised: 19 June 2008 / Accepted: 3 August 2008 / Published online: 5 November 2008
© Springer-Verlag 2008

Abstract The development and application of genomic tools to loblolly pine (*Pinus taeda* L.) offer promising insights into the organization and structure of conifer genomes. The application of a high-throughput genotyping assay across diverse forest tree species, however, is currently limited taxonomically. This is despite the ongoing development of genome-scale projects aiming at the construction of expressed sequence tag (EST) libraries and

the resequencing of EST-derived unigenes for a diverse array of forest tree species. In this paper, we report on the application of Illumina's high-throughput GoldenGate™ SNP genotyping assay to a loblolly pine mapping population. Single nucleotide polymorphisms (SNPs) were identified through resequencing of previously identified wood quality, drought tolerance, and disease resistance candidate genes prior to genotyping. From that effort, a 384 multiplexed SNP assay was developed for high-throughput genotyping. Approximately 67% of the 384 SNPs queried converted into high-quality genotypes for the 48 progeny samples. Of those 257 successfully genotyped SNPs, 70 were segregating within the mapping population. A total of 27 candidate genes were subsequently mapped onto the existing loblolly pine consensus map, which consists of 12 linkage groups spanning a total map distance of 1,227.6 cM. The ability of SNPs to be mapped to the same position as fragment-based markers previously developed within the same candidate genes, as well as the pivotal role that SNPs currently play in the dissection of complex phenotypic traits, illustrate the usefulness of high-throughput SNP genotyping technologies to the continued development of pine genomics.

Electronic supplementary material The online version of this article (doi:10.1007/s11295-008-0183-8) contains supplementary material, which is available to authorized users.

Communicated by D. Grattapaglia

A. J. Eckert
Section of Evolution and Ecology,
University of California at Davis,
One Shields Avenue,
Davis, CA 95616, USA

B. Pande · D. B. Neale (✉)
Department of Plant Sciences, University of California at Davis,
One Shields Avenue,
Davis, CA 95616, USA
e-mail: dbneale@ucdavis.edu

E. S. Ersoz
Institute for Genomic Diversity, Cornell University,
Ithaca, NY 14850, USA

M. H. Wright
Department of Molecular Biology and Genetics,
Cornell University,
Ithaca, NY 14850, USA

V. K. Rashbrook · C. M. Nicolet
DNA Technologies Core Facility, Genome Center,
University of California at Davis,
One Shields Avenue,
Davis, CA 95616, USA

Keywords Loblolly pine · *Pinus taeda* · Linkage mapping · Single nucleotide polymorphisms · GoldenGate™ · Genotyping · Genomics

Introduction

Genomic approaches to the understanding of genetic variation rely on the ability to contextualize the genomic location of a marker. Most often, that understanding is geared toward the dissection of complex traits into their genetic components. For example, the continued development of genetic marker types

and technologies to rapidly assess those markers has resulted in the construction of genetic linkage maps for a number of conifer species. Those maps have played key roles in the dissection of quantitative traits (Jermstad et al. 2001; Sewell et al. 2000, 2002) and facilitated the development of comparative linkage mapping and genomics (Krutovsky et al. 2004; Pavy et al. 2008). The power of such approaches as applied to conifers where genome sizes range from approximately 10 to 30 Gb (Murray 1998; Ahuja and Neale 2005) is highly dependent on the ability to genotype large numbers of loci. Technological advances in the ability to discover and genotype single nucleotide polymorphisms (SNPs) for conifers offer a promising solution to this deficiency.

The centrality of loblolly pine (*Pinus taeda* L.) to the development of pine genetics, genomics, and the US forest products industry positions it as the prime candidate for the application of high-throughput sequencing and genotyping technologies. The genetic map for loblolly pine is one of the most developed for conifers. Linkage maps with comparable or greater marker densities are available for only a few additional conifers—*Cryptomeria japonica* D. Don (Tani et al. 2003), *Picea abies* (L.) Karst. (Acheré et al. 2004), *Pinus pinaster* Ait. (Ritter et al. 2002), and *Pinus sylvestris* L. (Komulainen et al. 2003). Loblolly pine is a diploid ($2n=24$), long-lived, and largely outcrossing member of the genus *Pinus* with a natural range extending throughout southeastern North America. It is the foremost species harvested for timber and paper production in the United States with a farm gate value second only to maize. Its central role in the US forest products industry has motivated the creation of several breeding programs over the last 50 years, as well as the numerous molecular genetic and genomic projects currently underway (cf. <http://dendrome.ucdavis.edu>). Vital to those efforts, however, is the production of increasingly detailed genetic maps.

Isozymes were used to produce the first genetic maps for loblolly pine (Adams and Joly 1980; Conkle 1981). Linkage groups in those maps were defined using 12 to 20 markers with the first map representing linkage relationships between just three loci all within a single linkage group (Adams and Joly 1980). Additional isozyme markers extended the number of detected linkage groups to five (Conkle 1981). Subsequent developments in DNA-based marker technologies dramatically increased marker availability with concomitant increases in the ability to detect linkage groups. For example, the first DNA-based markers, restriction fragment length polymorphisms (RFLPs), expanded the loblolly pine linkage map to a density of approximately 70 markers with further increases to 191 using randomly amplified DNA (RAPDs) markers and to 508 using amplified fragment length polymorphism (AFLPs) markers (Devey et al. 1994; Remington et al. 1999). Those increases in marker availability were instrumental to the convergence of inferred

maps to approximately the same number of linkage groups as the haploid number of chromosomes for loblolly pine. Increasingly novel marker types also facilitated the construction of integrated consensus maps (Sewell et al. 1999; Brown et al. 2001) and the testing of several evolutionary hypotheses concerning genome organization (Kinlaw and Neale 1997; Zhou et al. 2002).

The primary focus of early linkage mapping efforts was the dissection of complex phenotypic traits. For example, a large number of quantitative trait loci (QTL) has been documented for several physical and chemical wood property traits (Groover et al. 1994; Sewell et al. 1999, 2002; Neale et al. 2002; Brown et al. 2003). Intervals for QTL in conifers, however, are often observed to be on the order of 15 cM (Brown et al. 2003) where a single centimorgan can represent up to 3,000 kb of DNA sequence (González-Martínez et al. 2007). Such poor genomic resolution makes the discovery of causative relationships between genotype and phenotype impractical. Further developments in marker technologies were thus motivated by these large intervals, as well as the need to identify orthologous loci at the DNA sequence level (Brown et al. 2001). This resulted in the production of the first DNA sequence-based markers from expressed sequence tag (EST) libraries (Harry et al. 1998). Such markers, termed EST polymorphisms (ESTPs), were subsequently integrated into existing maps as anchor points (Brown et al. 2001; Temesgen et al. 2001), thus facilitating the development of comparative linkage mapping across the Pinaceae (Krutovsky et al. 2004) and the identification of candidate genes collocating with QTLs (Brown et al. 2003).

The use of ESTPs, however, still did not facilitate the precise mapping of complex phenotypes to genotypes, which has only just begun to be addressed using association genetic methodologies (Neale and Savolainen 2004). Those methodologies rely on the discovery of SNPs within EST-derived unigenes or putative candidate genes through the process of resequencing within a small diversity panel followed by the genotyping of discovered SNPs within a large, unrelated sample of trees. Associations are then inferred between variation at SNPs and the variation in a quantitative phenotype (Long and Langley 1999; Cardon and Bell 2001). Several significant associations of candidate gene SNPs with wood quality (González-Martínez et al. 2007), drought tolerance (González-Martínez et al. 2008), and disease resistance (Ersoz, unpublished dissertation) phenotypes have already been found in loblolly pine. Renewed interest now exists to understand the genomic location and context of those significant associations, as well as the linkage relationships among genes comprising the entire functional gene space.

The ability to discover SNPs, associate them with phenotypes of interest, and integrate them into existing

linkage maps is dependent upon SNP genotyping technology. That technology has progressed from single SNP assays (cf. González-Martínez et al. 2007) to multiplexed, high-throughput assays capable of genotyping hundreds to thousands of SNPs at a time. Applications of the high-throughput technologies to conifer species, however, are currently limited taxonomically (cf. Pavy et al. 2008). In this paper, we report the first application of a highly multiplexed, high-throughput SNP genotyping platform to a species of *Pinus* using 48 progeny from a loblolly pine pedigree in an effort to map wood quality, drought tolerance, and disease resistance candidate genes. The results are subsequently extended to a discussion of the importance of high-throughput SNP genotyping platforms to the development of pine genomics.

Materials and methods

Plant material

Linkage mapping in loblolly pine has primarily been conducted with progeny from two three-generation outbred pedigrees constructed with unrelated first-generation selections obtained from the North Carolina State University—Industry Cooperative Tree Improvement Program. Those pedigrees have previously been used to map isozyme, RAPD, RFLP, AFLP, ESTP, and single sequence repeat (SSR) markers (Devey et al. 1994; Groover et al. 1994; Remington et al. 1999; Sewell et al. 1999; Brown et al. 2001; Temesgen et al. 2001). The first of those pedigrees consists of 95 full-sib progeny and is referred to as the *base* pedigree. The second of those pedigrees consists of 172 full-sib progeny and is referred to as the *qtl* pedigree. A random sample of 48 progeny from the *qtl* pedigree were selected for the SNP genotyping assay.

Detailed descriptions of DNA isolation protocols for needle tissue from all progeny can be found elsewhere (Devey et al. 1994; Sewell et al. 1999). In brief, DNA was isolated using a modified hexadecyltrimethylammonium bromide procedure, quantified using the PicoGreen assay (Invitrogen, Carlsbad, CA, USA), and standardized to Illumina-specified concentrations for SNP genotyping (50–250 ng/ μ L).

Development of SNP markers

Identification and characterization of SNP markers is fully described elsewhere (Brown et al. 2004; González-Martínez et al. 2006; Ersoz, unpublished dissertation). Candidate genes for wood quality, drought tolerance, and disease resistance phenotypes were selected based on the following criteria: colocation with a QTL, plausible physiological function, and previous identification as causative of the

phenotype of interest in another plant species. Haploid megagametophyte DNA samples obtained from 32 trees sampled across the natural range of loblolly pine were subsequently resequenced for SNP discovery. Alignment and SNP identification were conducted manually with established software (e.g., *Phred*, *Phrap*, *Consed*). A total of 1,008 SNPs were discovered across 79 candidate genes. The first set of 384 SNPs which were discovered were subsequently chosen for genotyping. Those 384 SNPs were submitted to Illumina to obtain design scores, which provide a reliability metric for whether or not a particular SNP will convert into genotype data using the GoldenGate™ assay. The 384 selected SNPs are collectively referred to as the loblolly pine 384 OPA.

Marker nomenclature and informatics

Markers were named following the guidelines established for the TreeGenes database. Each marker is designated by a standardized name that conveys unique information linking together the type of marker, the database in which that marker has been deposited (e.g., dbSNP), the laboratory in which the marker was developed, a locus-specific name or number, and/or a species code. Detailed information on these naming conventions can be found at the forest tree Comparative Mapping Database website (CMAP; <http://dendrome.ucdavis.edu/cmap/>).

SNP genotyping

Multiplexed genotyping was carried out using Illumina's GoldenGate™ assay according to the manufacturer's protocol (cf. Landegren et al. 1988; Oliphant et al. 2002; Fan et al. 2003). In brief, the assay involves generating hundreds of templates with specific target and address sequences using allele-specific extension followed by ligation and amplification with universal primers. Fluorescent products are hybridized to precoded beads on an array matrix from which the signal intensities are subsequently determined using Illumina's BeadArray Reader.

Signal intensities were quantified and matched to specific alleles using BeadStudio software v 3.1.14 (Illumina). A minimum threshold of 0.35 for the GenCall₅₀ score (GCS) was chosen for inclusion of SNP genotypes. The GCS for each SNP was based on the median score across all 48 genotyped progeny. The GCS is a quality metric that was used to assess the reliability of progeny samples to be grouped into genotypic clusters with values near 1 indicating high reliability. We correlated the GCS with the Illumina design score using simple linear regression to better understand the relationship between realized data quality and the design score. This analysis was carried out using the *lm* and *anova.lm* functions in R (Ihaka and

Gentleman 1996). Manual reclustering of homozygous and heterozygous clusters was performed when necessary using heritability information and positive controls obtained from four randomly replicated progeny samples. Genotyping was conducted at the DNA Technologies Core Facility located at the University of California at Davis (<http://www.genomecenter.ucdavis.edu>).

Map construction

All SNP loci were assessed for conformity to Mendelian segregation ratios prior to linkage mapping. Goodness-of-fit χ^2 tests were conducted using expected Mendelian ratios of 1:2:1 and 1:1 for SNPs that were heterozygous in both parents or in just a single parent, respectively. Markers of the former type were termed both informative (BI), while those of the latter type were termed maternally informative (MI) or paternally (PI) informative depending on whether it was the mother or father tree that was heterozygous. Multi-allelic markers (i.e., non-SNP markers) can also segregate in a 1:1:1:1 ratio if the parents are heterozygous for different diploid sets of alleles. Markers of that type were termed fully informative (FI). All tests were conducted using a significance threshold of $P=0.05$. Loci showing extreme segregation distortion (i.e., $P<0.001$) were omitted from further analysis.

Segregating SNPs were subsequently combined with 290 other markers (62 ESTP, 11 isozyme, 26 RAPD, 182 RFLP, eight SSR markers) to infer an integrated sex-averaged linkage map for the *qtl* pedigree ($n=172$ progeny) using JOINMAP v 1.4 (Stam 1993; Grattapaglia and Sederoff 1994). Only 48 of the 172 progeny had SNP genotype data. Marker data were obtained from the previously published maps of Devey et al. (1994), Sewell et al. (1999), Brown et al. (2001, 2003), and Krutovsky et al. (2004), as well as from unpublished SSR data kindly provided by C. Dana Nelson and C. Echt.

The map for the *qtl* pedigree was subsequently combined with a sex-averaged, integrated linkage map for the *base* pedigree ($n=95$ progeny) to produce a consensus linkage map as described by Sewell et al. (1999) and Krutovsky et al. (2004). The sex-averaged map for the *base* pedigree was constructed from 95 markers (34 ESTP, five isozyme, 56 RFLP markers). During the construction of all maps, markers were assigned to tentative linkage groups by comparing linkages formed at limit of detection (LOD) thresholds of 3.0 to 8.0 at increments of 0.5. Markers were grouped at a LOD threshold of 4.0 and ordered within linkage groups using a LOD threshold of 0.1. The order of markers was forced to conform to previously published linkage maps when necessary. Recombination frequencies were converted to map distances in centiMorgans (cM) using the Kosambi mapping function (Kosambi 1944). Linkage map results are deposited

and freely available at the forest tree CMAP database website (TreeGenes accession: TG944).

Results

Data summary

Of the 384 SNPs queried with the GoldenGate™ assay, 351 had GCS scores greater than 0.35, and of those, 257 (66.9%) were successfully genotyped. Of those 257 SNPs, 70 were clearly segregating within the *qtl* pedigree. Patterns of segregation for those 70 SNPs were consistent with heritability information (Fig. 1) and the median quality of genotype calls was high ($GCS_{\text{median}}=0.772$; Fig. 2). Eleven of the 70 segregating SNPs had GCS less than 0.70, which required manual reclustering and interpretation. Only a single SNP had a call frequency less than 1.0 with that frequency being 0.98. The call frequency refers to the fraction of samples for which a genotype could be declared for each SNP locus. Of the 314 SNPs not segregating in the *qtl* pedigree, 187 were monomorphic, 109 were difficult to cluster or were at odds with heritability information, and 18 failed to convert in the assay. The relationship between GCS and the design score for SNPs with GCS greater than 0.35 was not significant ($F_{1,349}=1.205$, $P=0.273$). The 70 segregating SNPs represent variation within 31 candidate genes, each of which was represented by one to eight SNPs (Table S1 of the Electronic supplementary material).

Linkage analysis

Minimal segregation distortion was found for the 70 segregating SNPs ($n=4$; $\chi^2_{\text{max}}=17.53$; $P_{\text{min}}=0.0002$). On average, SNPs within the same candidate gene mapped to the same linkage group and general area within that linkage group. The exceptions ($n=10$) to these patterns were typically correlated with segregation distortion, low GCS, or putative paralogs. These exceptions were removed prior to the construction of the final linkage maps. An additional 20 SNPs were mapped to the same position (i.e., <0.5 cM apart) as other SNPs within the same candidate gene. These SNPs were also removed prior to the construction of the final linkage maps due to their redundancy. Therefore, a total of 30 of the 70 segregating SNPs were removed prior to the construction of the final linkage maps.

The resultant sex-averaged linkage map for the *qtl* pedigree contained 15 linkage groups defined by 325 markers. This map had a total length of 1,145.0 cM. Parent-specific maps were generally similar in structure and marker ordering to the sex-averaged map. Both the maternal and paternal maps had 15 linkage groups with the maternal map having a total length of 975.3 cM and the

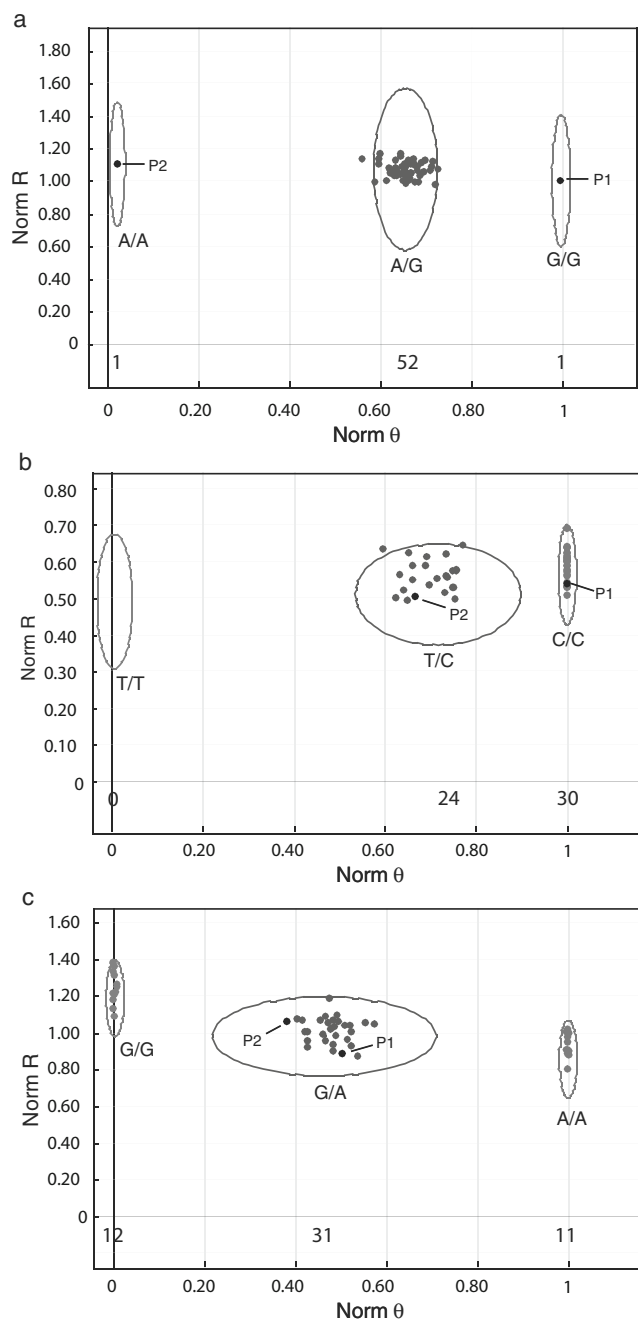


Fig. 1 Exemplar patterns of segregation observed for three SNPs selected from those observed to segregate in the loblolly pine *qtl* pedigree. Segregation patterns varied from uninformative (**a**) to parentally informative (**b**) to both informative (**c**). Each point represents a sample and is positioned along axes corresponding to its observed normalized fluorescent dye intensity for each of two dyes. Parents are labeled as *P1* and *P2*. *Ovals* enclose samples for which genotypes were called with strong confidence. *Numbers* above the *x*-axis are counts of the number of samples within each cluster. The positive controls are included in these plots, so that $n=54$

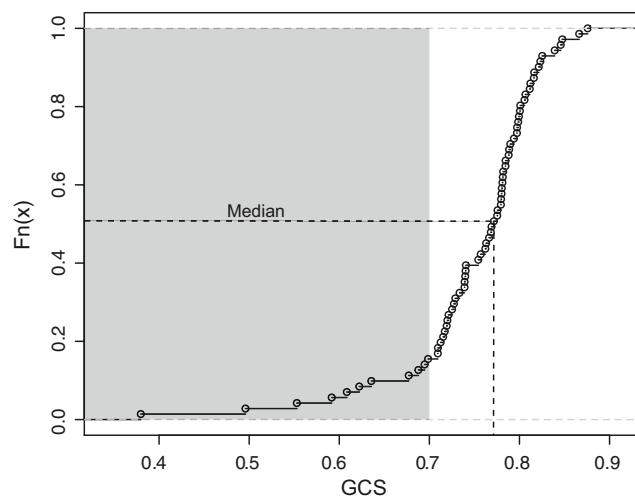


Fig. 2 Cumulative distribution of the GCS quality metric used to assess the reliability of genotypic clustering for the 70 SNPs segregating in the loblolly pine *qtl* pedigree. The *gray region* designates the range of values for which the GCS is not as reliable as an indicator of quality. Manual reclustering and interpretation was performed for SNPs in this region ($n=11$)

paternal map having a total length of 857.0 cM. Four of the 31 candidate genes were not incorporated into the sex-averaged map inferred for the *qtl* pedigree with two further candidate genes forming the basis of small (i.e., four markers in length) linkage groups. Linkage group organization and map length for the sex-averaged and paternal maps inferred for the *base* pedigree were similar to those presented by Sewell et al. (1999).

The inferred loblolly pine consensus map had 12 linkage groups incorporating 373 markers with a total length of 1,227.6 cM (Table S2 of the Electronic supplementary material; Fig. 3). Forty-seven of those markers were shared across pedigrees with a range of zero to eight shared markers per linkage group (mean=4, SD=2). The average number of markers and the spacing between those markers per linkage group were 31 ± 11 and 3.9 ± 1.6 cM, respectively. Of those 372 markers, 35 were SNPs representing 27 candidate genes. The distribution of mapped candidate genes across linkage groups was variable, ranging from a low of zero for linkage groups 2, 4, and 12 to a high of six for linkage group 6. Four of the 31 candidate genes (*nac-1*, *lp3-3*, *ptlim-2*, *myb-2*) represented by five SNPs were not incorporated into one of the 12 linkage groups.

Discussion

Detailed genetic maps are crucial to the empirical understanding of genetic variation, especially for organisms without reference genome sequences. The level of genomic detail and usefulness of those maps, however, is dependent on the ability to genotype a large number of functional and

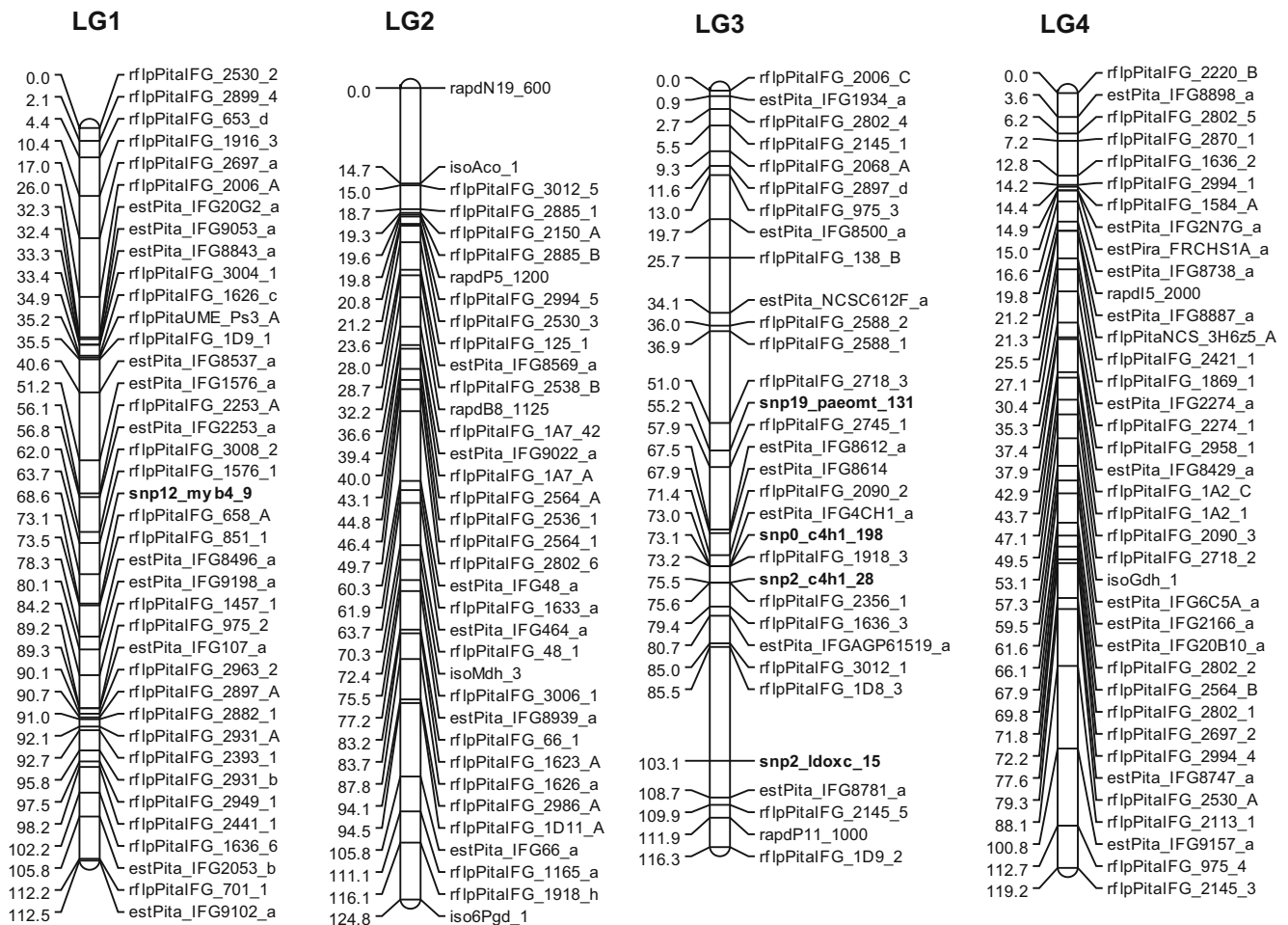


Fig. 3 Consensus linkage map for loblolly pine derived from the *qtl* and *base* pedigrees. This map was inferred from the sex-averaged linkage map from each pedigree (see the “Materials and methods” section). Markers are named according to the guidelines available at

the CMAP database website (<http://dendrome.ucdavis.edu/cmap/>). **Bold type** identifies SNPs. Centimorgan distances (Kosambi) are indicated to the *left* of each linkage group

putatively adaptive genetic markers. In this paper, we have shown that the high-throughput, multiplexed GoldenGate™ genotyping assay recovers approximately 67% of queried SNP loci and that many of those loci can be placed on existing linkage maps.

The GoldenGate™ assay was originally designed for high-throughput genotyping in model organisms or those with reference genome sequences. For example, the first phase of SNP genotyping for the International HapMap project was largely completed using this technology. The benefits of such large datasets are clear with HapMap SNP data being downloaded greater than 500,000 times from the project’s website (Thorisson et al. 2005). Those data have aided in exploring the human genome for signatures of positive selection (Sabeti et al. 2007), recombination hotspots (Myers et al. 2005), and SNPs associated with common disease phenotypes (Clayton et al. 2005). The transfer of this technology outside of model organisms has been slow, but the genetic and genomic implications of

similar data types for conifers, and forest trees in general, are the same as those in model organisms. This is especially true given the large effective population sizes, rapid decay of linkage disequilibrium, relatively little population structure, and ample amounts of genetic diversity often observed in conifer populations, all of which make conifer populations amenable to dissection of complex traits using association genetic methodologies (Neale and Savolainen 2004; Neale 2007).

Data obtained from the GoldenGate™ assay reported in this paper suggest that this assay performs well within the complex and large genome of pines. For example, 66.9% of the queried SNPs translated into easily interpreted genotypic clusters consistent with heritability information and positive controls (Fig. 1). The observation that approximately 27% ($70/257=0.272$) of the successfully genotyped SNPs were segregating within the *qtl* pedigree, moreover, is consistent with expected heterozygosity (H_e) estimates previously reported for diverse conifer species and for

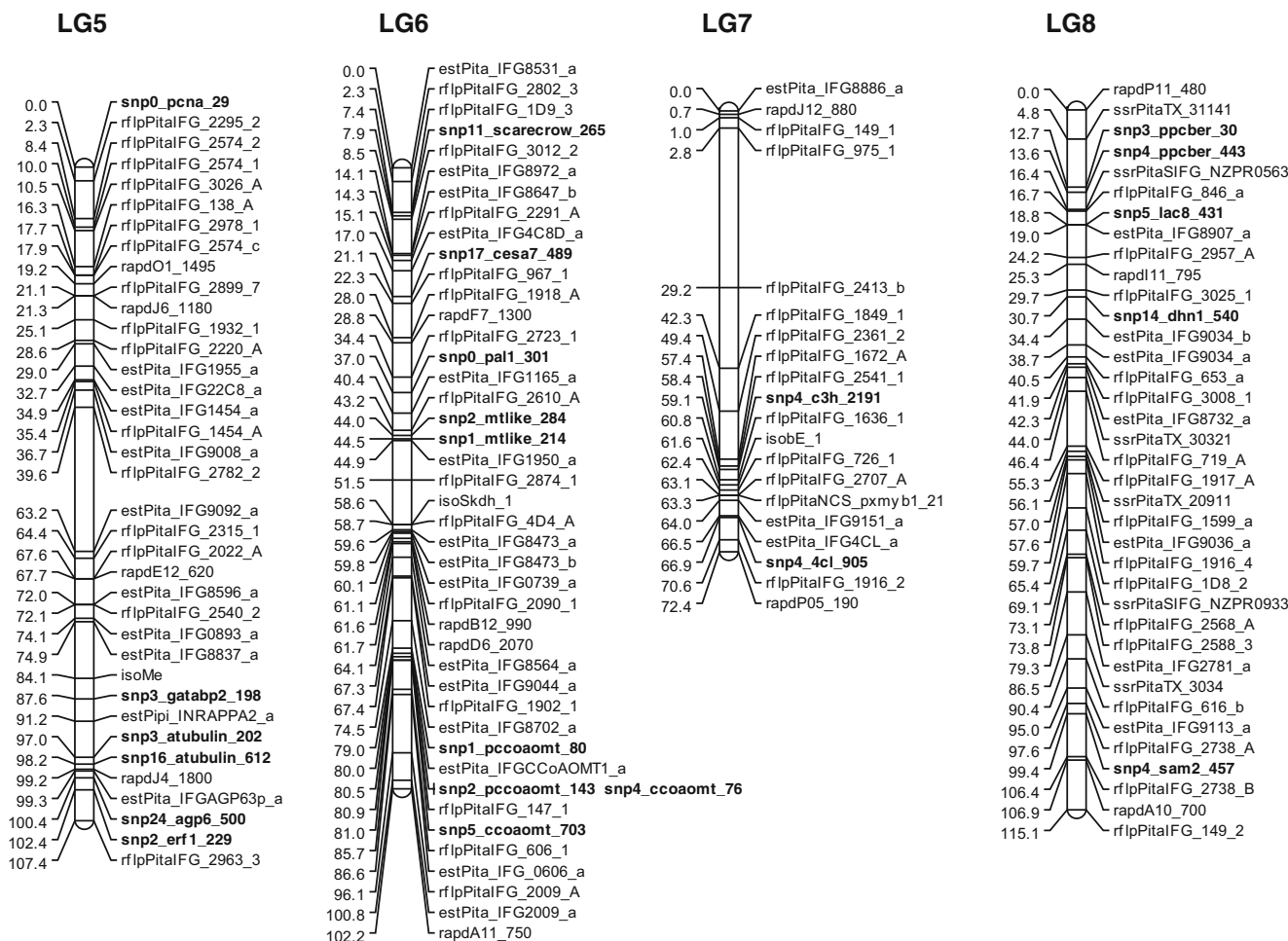


Fig. 3 (continued)

loblolly pine in particular (loblolly pine $H_e=0.282-0.362$; Ledig 1998). The reported conversion rate is also very similar to a recent application of the GoldenGate™ platform in two spruce species where 69.2% and 77.1% of the queried SNPs converted (Pavy et al. 2008). Conversion rates for both of the applications in conifers, however, were lower than applications in barley (Rostoks et al. 2006), bovines (McKay et al. 2007), and soybean (Hyten et al. 2008).

The lower conversion rate for pine, and conifers in general, compared with the other plant and animal species assayed using the GoldenGate™ platform can be explained by their relatively high levels of nucleotide polymorphism (Brown et al. 2004; González-Martínez et al. 2006; Savolainen and Pyhäjärvi 2007). For example, Brown et al. (2004) reported an average per site estimate of nucleotide diversity (π) at silent sites of 0.00658 for loblolly pine, which is several times greater than that of humans and many agriculture plant species to which the GoldenGate™ technology has been applied. Consistent with this argument are the similar figures reported for Norway spruce [*Picea abies* (L.) Karst], which also has moderate levels of nucleotide

diversity (Heuertz et al. 2006). Much of the dropout in applications of this platform to conifers, therefore, is likely to be a function of the common resequencing strategy employed in these species where a small sample of trees are used to discover SNPs and a much larger sample of trees is genotyped for downstream analyses and applications. The dropout then occurs due to the presence of undetected SNPs within the primer regions for the larger sample of trees being genotyped. This is consistent with the nonexistent relationship between GCS and design score. Given the cost per data point of this platform (~0.05 USD), however, a conversion rate around 70% still gives an order of magnitude greater data quantity per unit price than lower throughput and more expensive platforms with higher conversion rates.

Successfully genotyped SNPs were also incorporated into existing linkage maps for loblolly pine. Their integration into a consensus map based largely on anonymous markers did not substantially change the inferred map length, marker ordering, or linkage group organization as that presented previously (Sewell et al. 1999; Brown et al.

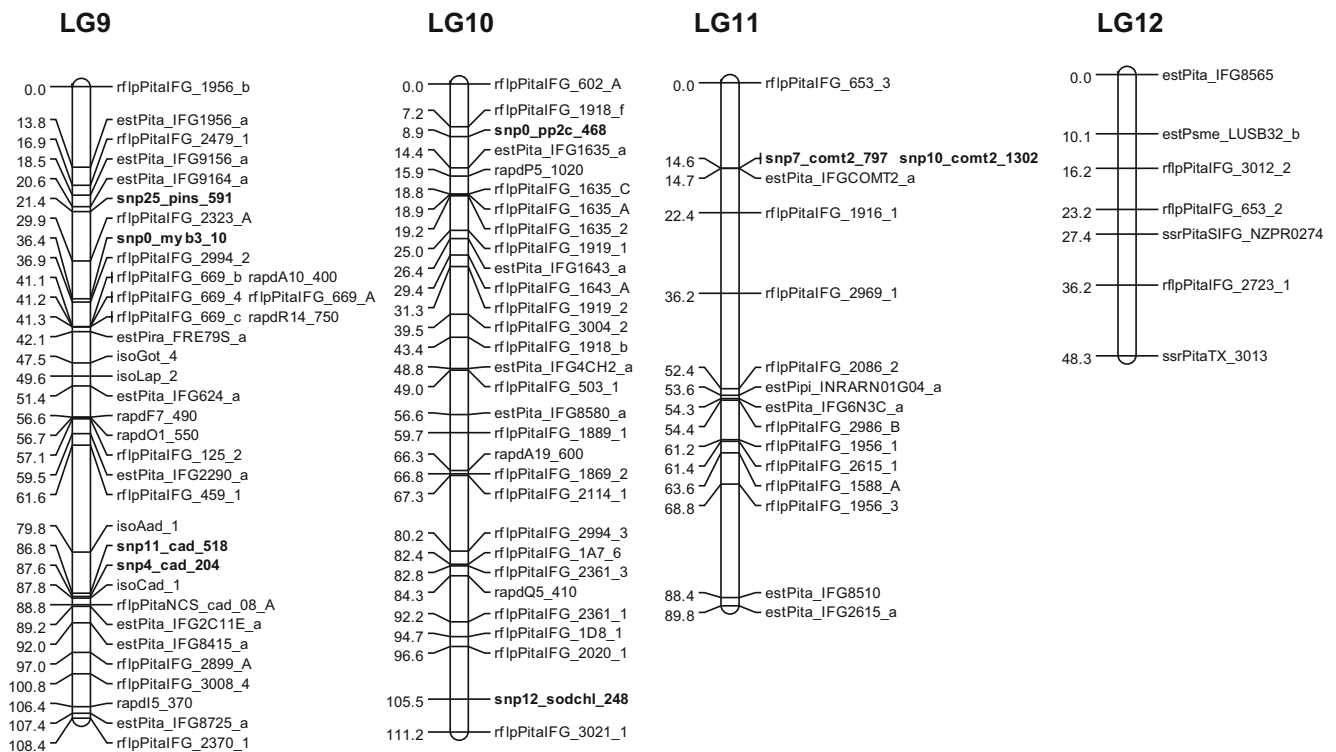


Fig. 3 (continued)

2001, 2003; Krutovsky et al. 2004). The map positions of only four candidate genes differed from those reported previously (Brown et al. 2003, 2004; González-Martínez et al. 2006; Ersoz, unpublished dissertation)—*cesA7*, *lp3-3*, *pAEOMT*, and *sod-chl*. Previous studies were unable to assign three of those four candidate genes (*cesA7*, *pAEOMT*, *sod-chl*) to linkage groups. Those candidate genes clearly mapped to linkage groups 6, 3, and 10 in this study, respectively. Lastly, *lp3-3* was placed in linkage group 2 by González-Martínez et al. (2006) and remains unlinked in the map presented in this paper. Those differences may be a function of the method used to genotype polymorphisms, FP-TDI (Kwok 2001; Gill et al. 2003) in previous studies and GoldenGate™ in this study or by the inclusion of additional dominant markers in the map presented in this study (Hu et al. 2004). In general, results between the FP-TDI and GoldenGate™ platforms are similar with differences largely attributable to the platform-specific ability to design oligonucleotides in the flanking regions of the target SNP (unpublished data).

In total, 27 candidate genes for wood quality, drought tolerance, and disease resistance were placed within 12 linkage groups. Six of those 27 candidate genes are newly incorporated in this version of the loblolly pine consensus map. In several instances, those candidate genes were mapped to the same linkage group using multiple marker types. For example, the cinnamyl alcohol dehydrogenase (*cad*) gene was mapped to linkage group 9 using isozyme

(isoCad_1), RFLP (rflpPitaNCS_CAD_08_A), ESTP (estPitaIFG_2C11E_a), and SNP (snp11_cad_518, snp4_cad_204) markers. The greatest map distance between any two of those markers was only 2.4 cM.

The emergence of high-throughput low-cost sequencing has begun to revolutionize the concept of the genetic map. Recent efforts to characterize the entire functional gene space of loblolly pine (cf. <http://dendrome.ucdavis.edu/adept2>) through resequencing of EST-derived unigenes have discovered a number of SNPs larger than even the highest multiplexity of the GoldenGate™ assay. Those markers will be instrumental in the integration of SNP genotyping with current association (González-Martínez et al. 2007, 2008) and previously completed QTL studies (Sewell et al. 1999, 2000, 2002; Brown et al. 2001, 2003) because now the set of functional genes putatively associated with a QTL or available for association tests is not solely limited by inferred orthology to distantly related angiosperms. The results presented in this study also open the door for applications of even higher throughput technologies such as the Infinium platform from Illumina (Steemers and Gunderson 2007). The result for pine genetic mapping is that consensus maps will ultimately be populated with as many functional markers as can be genotyped and that segregate within established mapping pedigrees. Information such as this will form the basis of a truly genomic approach to the study of genome-wide variation where the object of inquiry has now become the entire functional gene space rather than a small set of candidate genes.

Acknowledgements We would like to thank Jill Wegrzyn and Jennifer Lee for the bioinformatic support, as well as Kathleen Jermsted for the help in the production and interpretation of linkage maps. We would also like to thank C. Dana Nelson and Craig Echt for providing the SSR data. This research was supported by NSF grant DBI-0501763.

References

- Acheré V, Faivre-Rampant P, Jeandroz S, Besnard G, Markussen T, Aragones A, Fladung M, Ritter E, Favre J-M (2004) A full saturated linkage map of *Picea abies* including AFLP, SSR, ESTP, 5S rDNA and morphological markers. *Theor Appl Genet* 108:1602–1613
- Adams WT, Joly RJ (1980) Linkage relationships among twelve allozyme loci in loblolly pine. *J Heredity* 71:199–202
- Ahuja MR, Neale DB (2005) Evolution of genome size in conifers. *Silvae Genet* 54:126–137
- Brown GR, Kadel EE III, Bassoni DA, Kiehne KL, Temesgen B, van Buijtenen JP, Sewell MM, Marshall KA, Neale DB (2001) Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. *Genetics* 159:799–809
- Brown GR, Bassoni DL, Gill GP, Fontana JR, Wheeler NC, Megraw RA, Davis MF, Sewell MM, Tuskan GA, Neale DB (2003) Identification of quantitative trait loci influencing wood property traits in loblolly pine III. QTL verification and candidate gene mapping. *Genetics* 164:1537–1546
- Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc Natl Acad Sci USA* 101:15255–15260
- Cardon LR, Bell JI (2001) Association study designs for complex diseases. *Nat Rev Genet* 2:91–99
- Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, Smink LJ, Lam AC, Ovington NR, Stevens HE, Nutland S, Howson JMM, Faham M, Moorhead M, Jones HB, Falkowski M, Hardenbol P, Willis TD, Todd JA (2005) Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* 37:1243–1246
- Conkle MT (1981) Isozyme variation and linkage in six conifer species. In: Conkle MT (ed) *Proceedings of the Symposium on Isozymes of North American Forest Trees and Forest Insects*. General Technical Report GTR-PSW-048, USDA Forest Service, Berkeley
- Devey ME, Fiddler TA, Liu BH, Knapp SJ, Neale DB (1994) An RFLP linkage map for loblolly pine based on a three-generation outbred pedigree. *Theor Appl Genet* 88:273–278
- Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelplnagel J, Chee MS (2003) Highly parallel SNP genotyping. *Cold Spring Harbor Symp Quant Biol* 68:69–78
- Gill GP, Brown GR, Neale DB (2003) A sequence mutation in the cinnamyl alcohol dehydratase gene associated with altered lignification in loblolly pine. *Plant Biotech J* 1:253–258
- González-Martínez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006) DNA sequence variation and selection of tag SNPs at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* 172:1915–1926
- González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007) Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* 175:399–409
- González-Martínez SC, Huber D, Ersoz E, Davis JM, Neale DB (2008) Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity* 101:19–26
- Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-test cross: mapping strategy and RAPD markers. *Genetics* 137:1121–1137
- Groover AT, Devey ME, Fiddler TA, Lee JM, Megraw RA, Mitchell-Olds T, Sherman BK, Vujcic SL, Williams CG, Neale DB (1994) Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics* 138:1293–1300
- Harry DE, Temesgen B, Neale DB (1998) Codominant PCR-based markers for *Pinus taeda* developed from mapped cDNA clones. *Theor Appl Genet* 97:327–336
- Heurtz M, De Paoli E, Källman T, Larsson H, Jurman I, Morgante M, Lascoux M, Gyllenstrand N (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics* 174:2095–2105
- Hu X-S, Goodwillie C, Ritland KM (2004) Joining genetic linkage maps using a joint likelihood function. *Theor Appl Genet* 109:996–1004
- Hyten DL, Song Q, Choi I-Y, Yoon M-S, Specht JE, Matulmulli LK, Nelso RL, Shoemaker RC, Young ND, Cregan PB (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. *Theor Appl Genet* 116:945–952
- Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *J Comput Graph Stat* 5:299–314
- Jermstad KD, Bassoni DL, Jech KS, Wheeler NC, Neale DB (2001) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir I. Timing of vegetative bud flush. *Theor Appl Genet* 102:1142–1151
- Kinlaw CS, Neale DB (1997) Complex gene families in pine genomes. *Trends Plant Sci* 2:356–359
- Komulainen P, Brown GR, Mikkonen M, Karku A, Garcia-Gil MR, O'Malley D, Lee B, Neale DB, Savolainen O (2003) Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. *Theor Appl Genet* 107:667–678
- Kosambi DD (1944) The estimation of map values from recombination values. *Ann Eugen* 12:172–175
- Krutovsky KV, Troggio M, Brown GR, Jermstad KD, Neale DB (2004) Comparative mapping in the Pinaceae. *Genetics* 168:447–461
- Kwok P-Y (2001) SNP genotyping with fluorescence polarization detection. *Human Mutat* 19:315–323
- Landegren U, Kaiser R, Sanders J, Hood L (1988) A ligase-mediated gene detection technique. *Science* 241:1077–1080
- Ledig FT (1998) Genetic variation in pines. In: Richardson DM (ed) *Ecology and biogeography of Pinus*. Cambridge University Press, Cambridge, pp 251–280
- Long AD, Langley CH (1999) The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Res* 9:720–731
- McKay SD, Schnabel RD, Murdoch BM, Aerts J, Gill CA, Gao C, Li C, Matukumalli LK, Stothard P, Wang Z, Van Tassell CP, Williams JL, Taylor JF, Moore SS (2007) Construction of bovine whole-genome radiation hybrid and linkage maps using high-throughput genotyping. *Anim Genet* 38:120–125
- Murray BG (1998) Nuclear DNA amount in gymnosperms. *Ann Bot* 82(Suppl A):3–15
- Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005) A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310:321–324
- Neale DB (2007) Genomics to tree breeding and forest health. *Curr Opin Genet Dev* 17:539–544
- Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends Plant Sci* 9:325–330

- Neale DB, Sewell MM, Brown GR (2002) Molecular dissection of the quantitative inheritance of wood property traits in loblolly pine. *Ann For Sci* 59:595–605
- Oliphant A, Barker DL, Stuelpnagel JR, Chee MS (2002) BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques (Suppl)*:56–58 (60–61)
- Pavy N, Pelgas B, Beauseigle S, Blais S, Gagnon F, Gosselin I, Lamothe M, Isabel N, Bousquet J (2008) Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC Genomics* 9:21
- Remington DL, Wheten RW, Liu B-H, O'Malley DM (1999) Construction of an AFLP genetic map with nearly complete genome coverage in *Pinus taeda*. *Theor Appl Genet* 98:1279–1292
- Ritter E, Aragonés A, Markussen T, Acheré V, Espinel S, Fladung M, Wrobel S, Faivre-Rampant, Jeandroz S, Favre JM (2002) Towards construction of an ultra high density linkage map for *Pinus pinaster*. *Ann For Sci* 59:637–643
- Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proc Natl Acad Sci USA* 103:18656–18661
- Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES, The International HapMap Consortium (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913–918
- Savolainen O, Pyhäjärvi T (2007) Genomic diversity in forest trees. *Curr Opin Plant Biol* 10:162–167
- Sewell MM, Sherman BK, Neale DB (1999) A consensus map for loblolly pine (*Pinus taeda* L.) I. Construction and integration of individual linkage maps from two outbred three-generation pedigrees. *Genetics* 151:321–330
- Sewell MM, Bassoni DL, Megraw RA, Wheeler NC, Neale DB (2000) Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.) I. Physical wood properties. *Theor Appl Genet* 101:1273–1281
- Sewell MM, Davis MF, Tuskan GA, Wheeler NC, Elam CC, Bassoni DL, Neale DB (2002) Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.) II. Chemical wood properties. *Theor Appl Genet* 104:214–222
- Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J* 3:739–744
- Stemers FJ, Gunderson KL (2007) Whole genome genotyping technologies on the BeadArray platform. *Bitechnol J* 2:41–49
- Tani N, Takahashi T, Iwata H, Mukai Y, Ujino-Ihara T, Matsumoto A, Yoshimura K, Yoshimuru H, Murai M, Nagasaka K, Tsumura Y (2003) A consensus linkage map for Sugi (*Cryptomeria japonica*) from two pedigrees, based on microsatellites and expressed sequence tags. *Genetics* 165:1551–1568
- Temesgen B, Brown GR, Harry DE, Kinlaw CS, Sewell MM, Neale DB (2001) Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). *Theor Appl Genet* 102:664–675
- Thorisson GA, Smith AV, Krishnan L, Stein LD (2005) The International HapMap project website. *Genome Res* 15:1592–1593
- Zhou Y, Bui T, Auckland LD, Williams CG (2002) Undermethylated DNA as a source of microsatellites from a conifer genome. *Genome* 45:91–99