*Genet. Res., Camb.* (2003), **81**, *pp.* 43–50. *With 4 figures.* © 2003 Cambridge University Press
DOI: 10.1017/S0016672302005992 Printed in the United Kingdom


# Haplotypic QTL mapping in an outbred pedigree

DAVID P. GWAZE[1], YI ZHOU[1]†, M. HUMBERTO REYES-VALDÉS[2], MOHAMMAD A. AL-RABABAH[1] AND CLAIRE G. WILLIAMS[1]*
[1] *Texas A&M University, College Station, TX 77843-2135, USA*
[2] *Universidad Autónoma Agraria Antonio Narro, Departamento de Fitomejoramiento, Buenavista, Saltillo, Coah. C.P. 25315, Mexico*

(*Received 24 June 2002 and in revised form 10 September and 26 September 2002*)

**Summary**

An offspring genome can be viewed as a mosaic of chromosomal segments or haplotypes contributed by multiple founders in any quantitative trait locus (QTL) detection study but tracing these is especially complex to achieve for outbred pedigrees. QTL haplotypes can be traced from offspring back to individual founders in outbred pedigrees by combining founder-origin probabilities with fully informative flanking markers. This haplotypic method was illustrated for QTL detection using a three-generation pedigree for a woody perennial plant, *Pinus taeda* L. Growth rate was estimated using height measurements from ages 2 to 10 years. Using simulated and actual datasets, power of the experimental design was shown to be efficient for detecting QTLs of large effect. Using interval mapping and fully informative markers, a large QTL accounting for 11·3% of the phenotypic variance in the growth rate was detected. This same QTL was expressed at all ages for height, accounting for 7·9–12·2% of the phenotypic variance. A mixed-model inheritance was more appropriate for describing genetic architecture of growth curves in *P. taeda* than a strictly polygenic model. The positive QTL haplotype was traced from the offspring to its contributing founder, $GP_3$, then the haplotypic phase for $GP_3$ was determined by assaying haploid megagametophytes. The positive QTL haplotype was a recombinant haplotype contributed by $GP_3$. This study illustrates the combined power of fully informative flanking markers and founder origin probabilities for (1) estimating QTL haplotype magnitude, (2) tracing founder origin and (3) determining haplotypic transmission frequency.

## 1. Introduction

Quantitative trait locus or loci (QTL) haplotypes can be traced back from offspring to founder by combining founder-origin probabilities with fully informative flanking markers. This can be a complex effort for outbred pedigrees because there are twice as many alleles as founders, markers are heterogeneous for parental mating configuration and marker informativeness is uneven within a linkage group.

The haplotypic method is based on viewing each offspring haplotype as a mosaic of chromosomal segments derived from different founders. Specific haplotypes can be identified then traced back to one or more founders (Reyes-Valdés, 2000; Reyes-Valdés & Williams, 2002). Generalized functions for calculating the probability of the reference founder genome at any point are given in Reyes-Valdés (2000). These functions are based on $x$, which represents a position on a chromosome in centimorgans, considering an arbitrary chromosome end as the zero position; on $r$, which represents the position of a recurrent marker; on $d$, which is the position of a donor marker; on $b$, which is the backcross generation; and on $k$, which is a constant dependent upon the mapping function.

The haplotypic method has been extended to outbred pedigrees by treating one founder as the donor and the other founders as collectively recurrent (Reyes-Valdés & Williams, 2002). Founders were defined as $A_1$, $A_2$, …, $A_n$ in the generalized method but in the

* Corresponding author. Faculty of Genetics, Texas A&M University, TAMU 2135, College Station, TX 77843-2135, USA. Tel: +1 (979) 8623745. Fax: +1 (979) 8456049. e-mail: claire-williams@tamu.edu
† Present address: Cereal Research Centre, Agriculture and Agri-Food Canada, 195 Dafoe Road, Winnipeg, Manitoba, Canada R3T 2M9.

specific case of a three-generation outbred pedigree, the four founders are grandparents and thus defined here as $GP_1$, $GP_2$, $GP_3$ and $GP_4$. In the initial analysis, $GP_3$ would be the donor or reference founder so that $GP_1$, $GP_2$ and $GP_4$ are collectively viewed as recurrent. Note that these other founders are not assumed to be genetically identical but that only one founder can be the donor founder in each subsequent analysis.

The parents are defined as $P_{12}$ and $P_{34}$ with the subscripts denoting their respective founders. The individual offspring are $O_1$, $O_2$, …, $O_n$, where $n$ is their sibship size. Assume that for a given individual $O_i$, the marker composition of the gamete contributed by $P_{34}$ is known. For each marker locus in $O_i$ it is also known whether the allele came from the donor founder $GP_3$ or not. If $GP_3$ contributed the allele then the marker locus is classified as a donor marker. If $GP_3$ did not contribute the allele then the marker locus is classified as a recurrent marker. The marker configuration of individual offspring $O_i$ is determined by applying the six formulae from the backcross model (Reyes-Valdés, 2000) to calculate founder-origin probabilities for all marked chromosomal segments in the individual $O_i$ genome, so that the offspring genome is scored as a mosaic of founder segments.

For example, the chromosome for the reference individual may have the array [**E**, **D**, **R**, **E**], which is defined as **E** = chromosome end, **D** = donor marker and **R** = recurrent marker. Using a backcross model, the probability of any given $x$ contributed by founder $GP_3$ can be calculated for this individual $O_i$. The same process can be performed with reference to each of the other three founders, obtaining an estimate of the proportional contribution of each founder to the individual $O_i$ genome. Fully informative markers for each individual offspring are identified and then defined either as donor or as recurrent markers (Reyes-Valdés & Williams, 2002).

Combining founder-origin probabilities with QTL mapping in outbred pedigrees was first generalized using a simultaneous analysis of all available marker information (Haley *et al*., 1994; Knott *et al*., 1997) and this has been extended to mixed-model inheritance (Nagamine & Haley, 2001). Compared with the simultaneous method, the haplotypic method is less computationally demanding and gives close approximation only if founders are truly outbred. The haplotypic method is not applicable to F2 pedigrees based on inbred lines because it ignores some information in two-allele intercross marker-type configurations (Reyes-Valdés & Williams, 2002).

Other methods have also been proposed for tracing haplotypes in outbred pedigrees, albeit specific to small sibships. Using pedigree analysis theory, Elston & Stewart (1971) and Cannings *et al*. (1978) developed probability functions in complex pedigrees to infer pedigree structure, penetrance rules and mode of transmission of traits. Calculating location scores for human disease mapping via a descent graph approach has been proposed (Sobel & Lange, 1996). A multipoint method for general pedigrees has been developed for linkage analysis of continuous characters, using identity by descent (IBD) correlations at arbitrary points along a chromosome for each relative-pair (Almasy & Blangero, 1998). The haplotypic method requires larger sibships than these methods but is more flexible with respect to extended outbred pedigrees with multiple generations and many founders.

Our experimental approach was to test the haplotypic method with QTL detection using growth rate measured in a woody perennial plant, *Pinus taeda* L. Conifers are suited as models for tracing QTL haplotypes from offspring to contributing founders in outbred pedigrees because they are largely undomesticated and highly heterozygous (Muranty, 1996; Williams, 1998). The major drawback to QTL analysis in any outbred pedigree is that the power of detecting QTLs tends to be low even with large sibships. The power of QTL detection in outbred pedigrees is increased with the use of fully informative markers (Beckmann & Soller, 1988). Given that a QTL is detected, its chromosomal segment or haplotype can be traced to a reference founder. Transmission of a QTL haplotype can then be tracked through the outbred pedigree.

Growth rate in conifers is a composite trait of annual height increments which accrues from indeterminate apical meristem elongation (Greenwood, 1980). In *P. pinaster* heritability values for growth curve parameters are moderate ($h^2 = 0{\cdot}09$–$0{\cdot}14$), and lower compared with height at a single age (Magnussen & Kremer, 1993). Inheritance of growth curves in conifers tends to be lower than those for animals (e.g. Gwaze *et al*., 2002; Rekaya *et al*., 2000). Genetic architecture of growth curves is assumed to be strictly polygenic (Kirkpatrick & Heckman, 1989) but there are no reports of QTL detection studies to test this assumption. Mixed model inheritance with multiple QTL alleles segregating for large and small effects may be a more appropriate model.

Our study objective was to detect QTLs for the growth rate in an outbred *P. taeda* pedigree then determine founder origin, magnitude and transmission frequency for QTL haplotypes. The study reports on the following: (1) finding QTLs influencing the growth rate, (2) testing stability of QTL expression through plant development, (3) estimating phenotypic effects for each QTL haplotype and (4) identifying contributing founder(s) for each QTL haplotype.

## 2. Materials and methods

The experimental approach required selection of an appropriate pedigree, phenotypic measurements,

https://doi.org/10.1017/S0016672302005992 Published online by Cambridge University Press

high-throughput marker data collection, linkage map construction, power calculations, QTL detection and founder analysis.

(i) *Founder-origin probabilities*

In this study, founder-origin probabilities were calculated at 1 centimorgan (cM) intervals by using information from selected fully informative haplotypic markers. QTL haplotypic effects were estimated relative to a reference founder, $GP_3$, which is the fast-growing founder 7-56 widely represented in many breeding programs. QTL detection was conducted on a single outbred three-generation pedigree, with four unrelated grandparents or founders ($GP_1$, $GP_2$, $GP_3$, $GP_4$), the two parents ($P_{12}$, $P_{34}$) and 91 full-sib offspring being used for QTL mapping analysis.

(ii) *Height measurements and growth curve parameters*

Phenotypic measurements were made on the single cross within six genetic tests designed with single-tree plots in 36 replicates per test site. The tests were assessed for height annually at ages 2–5 and at 10 years, defined as HT2, HT3, HT4, HT5 and HT10 (Gwaze *et al*., 2002). The Richards function (Richards, 1959) fits the curvilinear growth of harvest-age conifers. For younger trees, the growth curve analysis adheres to a simple linear relationship between height and age. Thus a linear model was fitted for estimating the growth rate using PROC REG procedure (SAS Institute, Cary, NC):

$$Y = \beta_0 + \beta_1 \times \text{age} + e,$$

where $Y$ is height at a given age, $\beta_0$ is the intercept, $\beta_1$ is the growth rate and $e$ is the random error. Phenotypic data for QTL analyses were collected on 73 to 91 offspring at the different ages. The phenotypic values were adjusted for site and replicate effects. Inheritance patterns and age–age correlations for growth rate of *P. taeda* have been reported previously (Gwaze *et al*., 2002). Growth rate is an appropriate trait for studying growth curves in general because it has a higher heritability value than the other two parameters in the Richards' function, the shape parameter and asymptote (Gwaze *et al*., 2002).

(iii) *Microsatellite data and linkage map construction*

Linked, fully informative microsatellites in a highly heterozygous outbred pedigree were the critical element. A set of 95 microsatellites were mapped to the pedigree using high-throughput marker data collection (Zhou *et al*., 2002). Detailed protocols and primer sets have been published previously (Auckland *et al*., 2002; Zhou *et al*., 2002). For each microsatellite, the forward primer was synthesized and then conjugated at its 5′ end to one of two infrared fluorescent dyes, IRD700 or IRD800 (MWG-BIOTECH, Charlotte, NC). PCR amplification was based on a labelled forward primer and an unlabelled reverse primer yielding fluorescence-labelled PCR products. Gel electrophoresis and visualization of alleles were performed using LI-COR Model IR² 4200 automated DNA sequencer (LI-COR, Lincoln, NE).

A sex-averaged genetic map was constructed with the 95 microsatellites; 55 had an intercross (IC) or multiple intercross (MIC) mating type configuration. Linkage relationships were constructed using the BUILD, ALL and FIXED functions of CRI-MAP (Green *et al*., 1990) with map distances based on Kosambi's mapping function and a LOD 4 threshold. Of the 95 markers, 51 markers were linked in 15 linkage groups. Of the 51 linked markers, 32 had the more informative intercross configurations (IC and MIC). The low-density map covered a total genetic distance of 794·7 cM with an average distance between markers of 21·8 cM.

(iv) *Power calculations and QTL detection*

The power of QTL detection was determined by simulated offspring marker and phenotype data combined with actual marker data for founders and parents. Linkage group 5 was representative of all linkage groups with four markers evenly spaced by about 20 cM. Three of the four markers were fully informative (multiple-allele intercross). The QTL was placed at 38 cM and assumed to have an additive effect of 1 standard deviation with a backcross configuration. Power, defined as percentage of successful QTL detection events, was determined with simulated heritability values of 0·05, 0·10, 0·15 and 0·20. This within-family heritability was estimated using the backcross QTL configuration and assumed the dominance effect was zero (Kearsey & Pooni, 1996). Each QTL analysis was performed with 30 simulations with a threshold value of $P \leqslant 0{\cdot}0034$, equivalent to 0·05 at the genome-wide significance level (Knott *et al*., 1997). The significance thresholds were determined individually using a permutation test (Churchill & Doerge, 1994) with 1000 permutations, performed for each simulation–heritability combination.

QTL mapping was based on linear regression (Martínez & Curnow, 1992; Haley & Knott, 1992; Haley *et al*., 1994). Regression of phenotypic values on those haplotypic probabilities was then used to calculate the $F$ statistics for each linkage group. The $F$ statistic values were plotted in a graphical analysis to locate maximum QTL peaks. A full model comprised of female, male and male × female interaction was tested to detect dominance effects (Knott *et al*.,

https://doi.org/10.1017/S0016672302005992 Published online by Cambridge University Press

1997). Statistical significance was based on permutation tests. Each QTL haplotype could be traced to its respective founder(s) using founder-origin probabilities and fully informative flanking markers.

Analyses were programmed in Mathematica (Wolfram Research, Champaign, ILL). Estimates from the haplotype-based method were also compared with the simultaneous marker method (Haley *et al*., 1994; Reyes-Valdés & Williams, 2002). Significance and magnitude of effects between the one-QTL and two-QTL models were compared to determine the most appropriate model. The two-QTL model was fitted to eliminate the chances of mapping one QTL when there are in fact two QTLs. The search for a two-QTL model was performed (Haley & Knott, 1992; Martínez & Curnow, 1994) by fitting phenotypic values by multiple regression to founder-origin probabilities at two points of the linkage group. The two-QTL model was considered meaningful for locations at two different marker intervals. The analysis for each linkage group and locations was considered in 1 cM increments.

QTL analyses were also performed for informative unlinked markers. For each founder, a binary indicator variable was defined, with an assumed a value of 1 if the marker allele originated in that founder and 0 otherwise, for the corresponding haplotype of each progeny member. For each founder, a regression of the phenotype on the indicator variable was performed and tested for a significant *F* statistic. Multiplicative combinations of indicator variables for different founders were also tested.

Suggestive and significance thresholds (Lander & Kruglyak, 1995) were determined by linkage group. Using Bonferroni's correction and the fact that the number of linkage groups in the map was 15, the groupwise suggestive threshold was $P \leqslant 0{\cdot}0670$ and the groupwise significance threshold was $P \leqslant 0{\cdot}0034$ (Knott *et al*., 1997). The significance thresholds were determined by linkage group using a permutation test (Churchill & Doerge, 1994), with 1000 permutations being performed for each linkage group by trait combination for the one-QTL model; for the two-QTL model, 100 permutations were performed. The proportion of phenotypic variation explained (PVE) by each QTL was the coefficient of determination in the analysis of variance. To test the significance of haplotype effects, growth rate was analysed using the PROC GLM procedure (SAS Institute, Cary, NC).

### (v) *Determining linkage phase for a QTL haplotype*

Linkage phase for the $GP_3$ haplotypes (see Fig. 2) was determined by assaying 19 megagametophytes extracted from $GP_3$'s mature seeds. The conifer megagametophyte is haploid multicellular tissue that serves as the functional megaspore to the egg cell, later providing nutritive tissue around the developing embryo. As the functional megaspore, the megagametophyte is genetically identical to the egg cell. This is equivalent to assaying a female gamete's haplotypic contribution and thus phase or skewed transmission frequency can be determined directly for any founder or parent (see Williams *et al*., 2001). In addition to the megagametophytes, nine adult descendants of $GP_3$ including $P_{34}$ were assayed for the same two microsatellites as an independent test.

## 3. Results

A QTL with a large effect for growth rate was detected. Expression of this QTL effect was stable for height from age 2 to 10 years. Six QTL haplotypes could be traced from offspring to the four founders. The QTL haplotype $A_{12}$-$B_5$ had the largest positive effect; it was a recombinant haplotype inherited from founder $GP_3$ (7-56). The female × male interaction as well as $GP_1$ and $GP_2$ effects were not significant for all traits so inferences were made with respect to the $GP_3$ haplotype.

### (i) *Stability of QTLs for height over the course of plant development*

Power of the experimental design was determined to be efficient for detecting QTLs of large effect. The power of the experiment was higher than expected for 91 offspring due to a preponderance of linked, fully informative markers. Power was 23%, 53%, 87% or 93% for detecting significant QTLs given heritabilities of 0·05, 0·10, 0·15 and 0·20, respectively.

A large QTL was detected on linkage group 13, explaining 11·3% of the phenotypic variance of growth rate. This QTL had an effect of 0·7 standard deviations for growth rate. The QTL was flanked by PtTX3030 and PtTX3127; marker PtTX3030 was most proximal to the QTL. The QTL was significant at the genome-wide level using the one-QTL model (Table 1). The QTL for the growth rate and age-specific heights mapped on the same location on linkage group 13 (Table 1).

The large QTL also explained 7·9–12·2% of the phenotypic variance of height across all ages (Fig. 1). The positive QTL haplotype affecting growth rate and height at all ages was contributed by $GP_3$. The QTL was statistically significant at the genome-wide level for all traits except HT2, the measurement age with the lowest heritability ($h^2 = 0{\cdot}08$). Its QTL expression was significant only at the suggestive level (Table 1). The QTL expression varied at other ages; maximum expression occurred at HT4 and HT5, accounting for 12·2% of the total phenotypic variance.

Several minor QTLs were also detected at the suggestive level but none were stable across all ages.

https://doi.org/10.1017/S0016672302005992 Published online by Cambridge University Press

Table 1. *QTLs which affect the growth rate (RATE) and age-specific height (HT) using haplotype-based mapping*

| Linkage group | Interval | Trait | QTL location (cM-K) | $GP_3$ haplotypic effect (SE) | PVE (%) | *P* value |
|---|---|---|---|---|---|---|
| 13 | PtTX3030–PtTX3127 | HT2 | 3 | +0·169 (0·068) | 7·9 | 0·023 |
| | | HT3 | 0 | +0·252 (0·079) | 10·3 | 0·003 |
| | | HT4 | 1 | +0·350 (0·100) | 12·2 | 0·001 |
| | | HT5 | 1 | +0·430 (0·134) | 12·2 | 0·002 |
| | | HT10 | 2 | +0·542 (0·167) | 10·5 | 0·001 |
| | | RATE | 1 | +0·060 (0·018) | 11·3 | 0·000 |
| 8 | PtTX3013–PtTX2034 | HT2 | 0 | +0·185 (0·063) | 10·8 | 0·012 |

QTL location is shown as Kosambi cM. Positive values for haplotypic effect are relative to the $GP_3$ haplotype. PVE is the percentage of phenotypic variation explained by the markers. *P* values are from a permutation test. The suggestive level is $\alpha \leqslant 0{\cdot}067$ and the genome-wide significance is $\alpha \leqslant 0{\cdot}0034$.

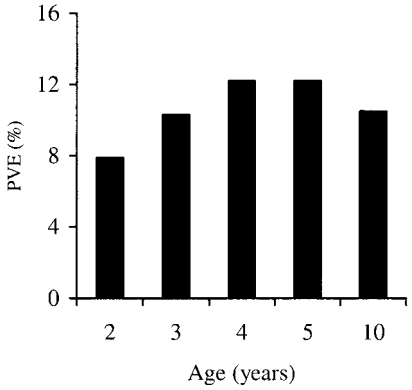


Fig. 1. Pattern of expression of the large, stable QTL on linkage group 13 is shown across *P. taeda* development using the proportion of phenotypic variance (PVE) explained by the QTL. Height measurements were taken at ages 2 to 10 years.

One QTL was detected with the one-QTL model on linkage group 8 for HT2. This QTL on linkage group 8 explained 10·8% of the phenotypic variance of HT2 with the positive QTL haplotype contributed by $GP_3$. Results of single-marker analysis were similar to the interval analysis. Seven additional QTLs were detected from single-marker analyses; all seven were age-specific and did not show stability across ages (Table 2). The test statistics were significant only at the suggestive level and the markers explained 4·4–8·3% of the phenotypic variance (Table 2). The same QTL estimates were obtained using the approach of Haley *et al.* (1994). For linkage groups 8 and 13, the two-QTL results were not a significant improvement on the one-QTL model. None of the test statistics for fitting the two-QTL model was significant at the suggestive level in the other linkage groups, indicating an absence of ghost-QTL effects.

(ii) *Tracing founder origin for each QTL haplotype*

A mixed-model inheritance was more appropriate for describing genetic architecture of growth curves in *P. taeda* than a strictly polygenic model. Using fully informative markers and founder-origin probabilities, a high positive QTL haplotype was identified and then traced from the offspring to its contributing founder, $GP_3$. This large QTL effect could be traced because fully informative markers in this highly heterozygous pedigree flanked the QTL. Marker PtTX3030 was defined as A with alleles 4, 12 and 16 representing band sizes 356, 324 and 300 bp. Marker PtTX3127 was defined as B with alleles 1, 2 and 5 representing band sizes 185, 181 and 168 bp respectively (Fig. 2).

The linked markers flanking the 17·1 cM interval around the large QTL effect located on linkage group 13 were in multiple-allele intercross configuration.

Table 2. *Results of single-marker analysis for loci found to be significantly associated with one or more traits*

| Marker locus | Trait | PVE (%) | *P* value | Reference founder |
|---|---|---|---|---|
| **PtTX2093** | HT10 | 7·5 | 0·009 | $GP_3$ |
| | RATE | 5·4 | 0·027 | $GP_3$ |
| PtTX2122 | HT10 | 6·3 | 0·016 | $GP_2$ |
| PtTX2183 | HT4 | 4·7 | 0·039 | $GP_2$ |
| PtTX3055 | HT4 | 4·6 | 0·043 | $GP_4$ |
| **PtTX3090** | HT3 | 5·2 | 0·029 | $GP_4$ |
| | HT5 | 8·3 | 0·012 | $GP_1$ |
| | HT10 | 4·5 | 0·044 | $GP_4$ |
| | RATE | 7·4 | 0·009 | $GP_1$ |
| **PtTX3110** | HT2 | 7·3 | 0·021 | $GP_2$ |
| | HT3 | 5·2 | 0·030 | $GP_2$ |
| | RATE | 4·4 | 0·047 | $GP_2$ |
| PtTX4140 | HT3 | 4·5 | 0·043 | $GP_1$ |
| | HT5 | 7·1 | 0·020 | $GP_1$ |
| | HT4 | 5·4 | 0·027 | $GP_1$ |

Percentage of phenotypic variation explained (PVE) by the marker locus, *P* value by trait and statistical significance at the suggestive level ($\alpha \leqslant 0{\cdot}05$) are shown. Markers shown in bold are also linked to QTL influencing the growth rate. Reference founder is the founder contributing the positive haplotype for each respective trait.

https://doi.org/10.1017/S0016672302005992 Published online by Cambridge University Press

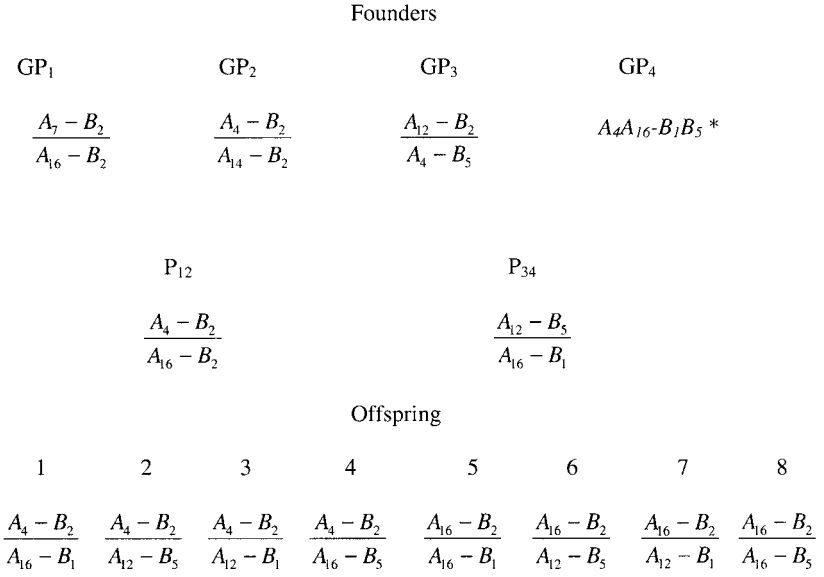


Fig. 2. Segregating haplotypes from four unrelated founders or grandparents to offspring. Markers A and B are PtTX3030 and PtTX3127 in linkage group 13. Asterisk indicates phase unknown. Phase for $GP_3$ was determined by haplotyping 19 megagametophytes and from genotyping eight sibs of $P_{34}$.

There were eight possible founder haplotypes transmitted to the next generation from $GP_3$ and $GP_4$ to $P_{34}$ (Fig. 2). Possible $GP_3$ haplotypes were $A_{12}$-$B_5$, $A_4$-$B_5$, $A_{12}$-$B_2$ and $A_4$-$B_2$, and possible $GP_4$ haplotypes were $A_4$-$B_5$, $A_{16}$-$B_5$, $A_4$-$B_1$ and $A_{16}$-$B_1$. There were four possible founder haplotypes transmitted to the next generation from $GP_1$ and $GP_2$ to $P_{12}$. Possible $GP_1$ haplotypes were $A_7$-$B_2$, and $A_{16}$-$B_2$, and possible $GP_2$ haplotypes were $A_4$-$B_2$ and $A_{14}$-$B_2$ (Fig. 2). Parent $P_{34}$ received a parental $A_{12}$-$B_5$ haplotype from $GP_3$ and phase-unknown haplotype $A_{16}$-$B_1$ from $GP_4$. The adult $P_{34}$ contributed its own parental haplotypes $A_{12}$-$B_5$ and $A_{16}$-$B_1$ as well as its recombinant haplotypes $A_{12}$-$B_1$ and $A_{16}$-$B_5$ to the offspring (Fig. 2). It is important to note here that a recombinant haplotype for the founder (GP) can be transmitted as a parental haplotype by the following generation (P).

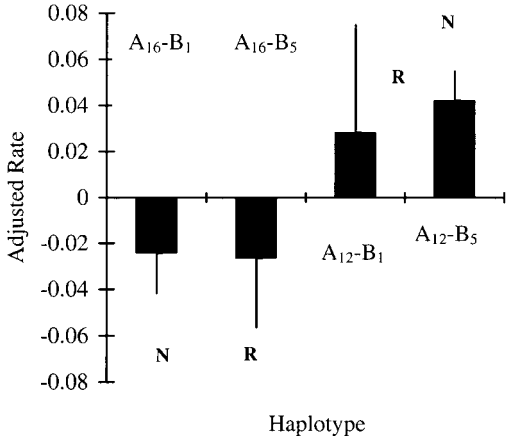


Fig. 3. Phenotypic value of QTL segment. Trait is the growth rate adjusted for replication and site effects. Haplotype $A_{12}$-$B_5$ was inherited from $GP_3$. Note that N codes for the parental (non-recombinant) haplotypes and R codes for the recombinant haplotypes with respect to $P_{34}$.

### (iii) *Estimating of phenotypic effects of a QTL haplotype*

Phenotypic effects varied widely for the four haplotypes contributed by $P_{34}$ (Fig. 3). $P_{34}$ contributed its $A_{12}$-$B_5$ haplotype to some offspring and this haplotype had a large, positive phenotypic effect on the growth rate. This $GP_3$ haplotype had a positive effect on the growth rate which was significantly different ($P \leqslant 0{\cdot}0265$) from the effects from other three QTL haplotypes contributed by $P_{34}$. Haplotype $A_{12}$-$B_5$ had an adjusted growth rate of 0·042 while those haplotypes without allele $A_{12}$ had negative adjusted growth rate values. Those offspring with allele $A_{12}$ had a higher growth rate regardless of whether $B_5$ or $B_1$ was also present (Figs. 3, 4). Offspring that inherited allele $A_{12}$ had an adjusted growth rate of 0·05, while those that inherited the alternative allele $A_{16}$ had an adjusted growth rate of less than 0·01.

### (iv) *Transmission of a QTL haplotype with a large positive effect*

$A_{12}$-$B_5$ was a recombinant haplotype contributed by $GP_3$ to $P_{34}$. Of the 19 megagametophytes assayed from $GP_3$, parental haplotypes $A_4$-$B_5$ and $A_{12}$-$B_2$ were more prevalent with 8 and 10 counts, respectively. None of the 19 megagametophytes had haplotype $A_{12}$-$B_5$ and only one had haplotype $A_4$-$B_2$. Of the nine adult offspring of $GP_3$, only $P_{34}$ had the $A_{12}$-$B_5$ haplotype, an expected result for a $GP_3$ recombinant haplotype. Frequencies of the $A_{12}$-$B_5$ haplotype among $P_{12} \times P_{34}$ offspring had normal transmission ratios.

https://doi.org/10.1017/S0016672302005992 Published online by Cambridge University Press

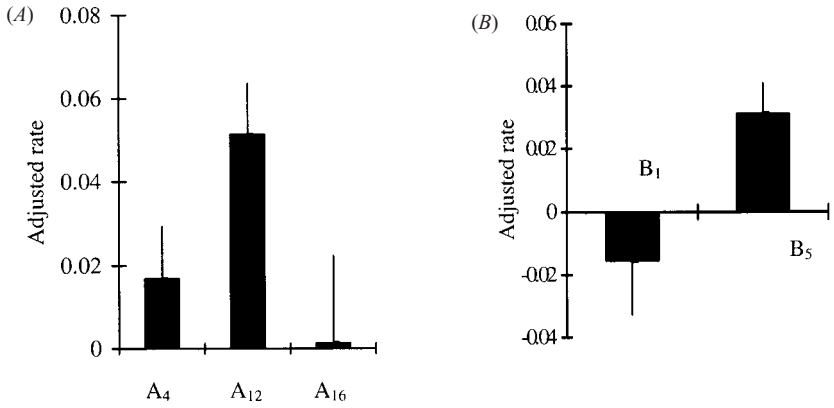


Fig. 4. (*A*) The QTL effect for growth rate for marker A alleles with the standard error for the growth rate. Allele $A_{12}$ was inherited from founder $GP_3$. (*B*) The QTL effect for growth rate for marker B alleles with the standard error for the growth rate.

## 4. Discussion

This study illustrates the combined power of fully informative markers and founder-origin probabilities for estimating QTL haplotype magnitude and tracing haplotypes back to founder. A large QTL effect for growth rate was detected then traced back to $GP_3$, a fast-growing founder.

### (i) *QTL stability across ages*

The QTL with large effect for growth rate was based on an estimated trajectory rather than a trait value, adding imprecision to QTL detection. However, the linear growth rate parameter was estimated from total height across ages and the same QTL was present at each age, adding support to the QTL detected for growth. This research area needs closer investigation in the case of older data requiring curvilinear models such as the Richards' function where several parameters are estimated to describe the growth trajectory of a single tree.

The magnitude of this QTL for height at different ages fluctuated; the low level of QTL expression at HT2 was due to lower heritability ($h^2 = 0{\cdot}08$) at younger ages (Gwaze *et al*., 2002). The stability of this QTL across ages contrasts with a previous study on conifer growth traits that suggests smaller QTL effects for age-specific growth traits are quite unstable across different ages (Kaya *et al*., 1999).

The power simulation indicated that sibship size was adequate at detecting QTLs of large effects but too small for detecting smaller QTL effects. The high LOD threshold (4·0) used in linkage mapping may also have led to a more conservative estimate of number of QTLs because there were fewer markers in each linkage group. Both these factors contributed to the absence of QTLs of small effect. For flanking markers, QTL results were similar whether the analysis was based on a single marker or on a set of flanking markers.

### (ii) *Mixed-model inheritance*

Mixed-model inheritance was more appropriate for describing the genetic architecture of the growth rate in *P. taeda* than a strictly polygenic model. The large QTL effect detected in this study has an effect of 0·7 standard deviations for the growth rate. Thus, under the canonical definition proposed by Smith (1985), this QTL was defined as a major locus. Classification of mixed-model inheritance for *P. taeda* growth rate was appropriate.

In conclusion, QTL mapping in any outbred pedigree is complex due to highly heterozygous founders, multiple alleles per locus and multiple marker configurations. The haplotypic method simplifies this complex analysis by reducing haplotypes within a donor-recurrent framework and viewing each offspring genome as a mosaic of chromosome segments contributed by different founders. This can be done for outbred pedigrees by combining any type of fully informative linked markers with founder-origin probabilities. The haplotypic method has broad application to undomesticated organisms or recalcitrant model systems whether these are outcrossing plants, mammals, fish, insects or marine organisms. The method has substantial value in the study of adaptive variation or interspecific hybridization. The concept was illustrated for a QTL haplotype with a large positive effect on growth rate in *P. taeda* which was traced from offspring to a specific founder. Transmission of QTL-bearing chromosomal segments could be followed across generations and across different stages in the diplohaplontic life cycle.

This work was supported by Agenda 2020 – Department of Energy Grant DE-FC07-001D13877 (to C.G.W.), USDA-Forest Service Southern Experiment Station (to C.G.W.), CONACyT-TAMU grants program (to H.R.V. and C.G.W.) and National Science Foundation Bioinstrumentation Grant NSF DB1-9871390 to the Institute of Plant

Genomics, Texas A&M University. We thank Westvaco, International Paper and Weyerhaeuser Company for access to genetic test data and DNA samples.

## References

Almasy, L. & Blangero, J. (1998). Multipoint quantitative-trait analysis in general pedigrees. *American Journal of Human Genetics* **62**, 1198–1211.

Auckland, L. D., Bui, T., Zhou, Y., Shepherd, M. & Williams, C. G. (2002). *Conifer Microsatellite Handbook*. Raleigh, NC: Corporate Press.

Beckmann, J. S. & Soller, M. (1988). Detection of linkage between marker loci and loci affecting quantitative traits in crosses between segregating populations. *Theoretical and Applied Genetics* **76**, 228–236.

Cannings, C., Thompson, E. A. & Skolnick, M. H. (1978). Probability functions on complex pedigrees. *Advances in Applied Probability* **10**, 26–61.

Churchill, G. A. & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971.

Elston, R. C. & Stewart, J. (1971). A general model for the genetic analysis of pedigree data. *Human Heredity* **21**, 523–542.

Green, P., Falls, K. & Crooks, S. (1990). *Documentation for CRI-MAP*, version 2.4. St Louis: Washington University School of Medicine.

Greenwood, M. S. (1980). Reproductive development in loblolly pine. I. The early development of male and female strobili in relation to the long shoot growth behavior. *American Journal of Botany* **67**, 1414–1422.

Gwaze, D. P., Bridgwater, F. E. & Williams, C. G. (2002). Genetic analysis of growth curves for a woody perennial species, *Pinus taeda* L. *Theoretical and Applied Genetics* **105**, 526–531.

Haley, C. S. & Knott, S. A. (1992). A simple regression model for interval mapping in line crosses. *Heredity* **69**, 315–324.

Haley, C. S., Knott, S. A. & Elsen, J.-M. (1994). Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**, 1195–1207.

Kaya, Z., Sewell, M. M. & Neale, D. B. (1999). Identification of quantitative trait loci influencing annual height- and diameter-increment growth in loblolly pine (*Pinus taeda* L.). *Theoretical and Applied Genetics* **98**, 586–592.

Kearsey, M. J. & Pooni, H. S. (1996). *The Genetic Analysis of Quantitative Traits*. London: Chapman & Hall.

Kirkpatrick, M. & Heckman, N. (1989). A quantitative genetic model for growth, shape and other infinite-dimensional characters. *Journal of Mathematical Biology* **27**, 429–450.

Knott, S. A., Neale, D. B., Sewell, M. M. & Haley, C. S. (1997). Multiple marker mapping of quantitative trait loci in an outbred pedigree of loblolly pine. *Theoretical and Applied Genetics* **94**, 810–820.

Lander, E. & Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**, 241–247.

Magnussen, S. & Kremer, A. (1993). Selection for the optimum tree growth curve. *Silvae Genetica* **42**, 322–335.

Martínez, O. & Curnow, R. N. (1992). Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theoretical and Applied Genetics* **85**, 480–488.

Muranty, H. (1996). Power of tests for quantitative trait loci detection using families in different schemes. *Heredity* **76**, 156–165.

Nagamine, Y. & Haley, C. S. (2001). Using the mixed model for interval mapping of quantitative trait loci in outbred line crosses. *Genetical Research* **77**, 199–207.

Rekaya, R., Carabano, M. J. & Toro, M. A. (2000). Bayesian analysis of lactation curves of Holstein-Friesian cattle using a nonlinear model. *Journal of Dairy Science* **83**, 2691–2701.

Reyes-Valdés, M. H. (2000). A model for marker-based selection in gene introgression breeding programs. *Crop Science* **40**, 91–98.

Reyes-Valdés, M. H. & Williams, C. G. (2002). A haplotypic approach to founder-origin probabilities and outbred QTL analysis. *Genetical Research* **80**, 231–236.

Richards, F. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany* **10**, 290–300.

Smith, C. (1985). Utilization of major genes. In *Genetics of Reproduction in Sheep* (ed. R. B. Land & D. W. Robinson), pp. 151–158. London: Butterworth.

Sobel, E. & Lange, K. (1996). Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *American Journal of Human Genetics* **58**, 1323–1337.

Williams, C. G. (1998). QTL mapping in outbred plants. In *Molecular Dissection of Complex Traits* (ed. A. Paterson), pp. 81–94. Boca Raton, FL: CRC Press.

Williams, C. G., Zhou, Y. & Hall, S. E. (2001). A chromosomal region promoting outcrossing in a conifer. *Genetics* **159**, 1283–1289.

Zhou, Y., Bui, T., Auckland, L. D. & Williams, C. G. (2002). Direct fluorescent primers are superior to M13-tailed primers for automated genotyping of *Pinus taeda* microsatellites. *Biotechniques* **32**, 46–52.