Theor Appl Genet (2007) 116:87–94
DOI 10.1007/s00122-007-0649-3

ORIGINAL PAPER

# Validating a QTL region characterized by multiple haplotypes

**Claire G. Williams · M. Humberto Reyes-Valdés · Dudley A. Huber**

Received: 16 April 2007 / Accepted: 17 September 2007 / Published online: 16 October 2007


**Abstract** Validation of a quantitative trait locus (QTL) for outcrossing perennial plants is rarely reported due to complexity of plausible genetic models and reliance on field designs already available. Here, a particular marker-QTL haplotype exerted a large, positive effect on height for *Pinus taeda* and its origin could be traced to a founder, $GP_3$, in a three-generation QTL pedigree. To validate this QTL effect, we used an extended $GP_3$-based pedigree. In the validation cross, each of the 46 offspring was clonally propagated from developing seeds using somatic embryogenesis technology. Subsequent analyses were conducted separately for seedlings and for other somatic emblings. For seedlings, the original QTL effect could not be fully validated. For somatic emblings, a strong negative QTL effect was detected in the validation cross; some evidence from another cross-supported the original positive QTL effect. From this part of the analysis, three distinct marker-QTL haplotypes at a single locus could be inferred. Validating QTL haplotypes in readily available field tests was feasible despite the genetic model complexity inherent to outcrossing long-lived perennials.

Communicated by M. Sillanpää.

C. G. Williams
Nicholas Institute for Environmental Policy Solutions,
Duke University, Durham, NC 27708, USA
e-mail: claire-williams@fulbrightweb.org

M. H. Reyes-Valdés
Departamento de Fitomejoramiento,
Universidad Autónoma Agraria Antonio Narro,
Buenavista, Saltillo, COAH 25315, Mexico
e-mail: mhreyes@uaaan.mx

D. A. Huber (✉)
School of Forest Resources and Conservation,
University of Florida, Gainesville, FL, USA
e-mail: dahuber@ufl.edu

## Introduction

For the special case of outcrossing long-lived pedigrees drawn from natural populations, validating a quantitative trait locus (QTL) is often complex and fraught with ambiguity. Validation does require a different pedigree (Lande and Thompson 1990) but these pedigrees should have one or more founders common to the original QTL detection pedigree (Muranty 1996; Verhoven et al. 2006). Otherwise, linkage equilibrium between a marker locus and its QTL (M-QTL) can add new sources of ambiguity (Maliepaard et al. 1997; Williams 1998). Even so, there are many other obstacles which must be considered.

First, one must consider the possibility that the QTL might be polymorphic in one pedigree yet monomorphic in another. Second, multiple QTL alleles per locus are often segregating in highly heterozygous populations and these can confound plausible genetic models especially if non-additive effects are strong (i.e. van Eck et al. 1994; Groover et al. 1994). Third, delayed reproductive onset and longevity places a premium on using readily available tests; otherwise years would be required to breed and test these selections. One alternative to such a lengthy delay is to plant clonally replicated genotypes across several locations (Bradshaw and Foster 1992) then search for suitable pedigrees at a later date. Another method is to conduct a priori founder-origin analyses to pinpoint chromosomal segments of interest (Reyes-Valdés and Williams 2002; Williams and Reyes-Valdés 2007) then search related pedigrees only for these segments. Whatever the method,

optimal QTL validation for outbred, long-lived species is still an open question. Here we present a case study using *Pinus taeda* as an example.

Resolving the complexity of QTL validation for long-lived organisms is illustrated as follows. *P. taeda* (Gwaze et al. 2002, 2003; Reyes-Valdés and Williams 2002) is an outcrossing conifer reported to have a large, positive QTL haplotype. The QTL was detected using two flanking microsatellite markers. The QTL effect explained 12.2% of the phenotypic variance of height at age 4 years (HT4). The QTL haplotype, inferred as $A_{12}$–$Q_1$–$B_5$, was inherited from founder $GP_3$ or 7-56 (Gwaze et al. 2003). Note that this original pedigree tested the QTL effect segregating among offspring from a $GP_3$ descendant, $P_{34}$, rather than from $GP_3$ directly so all QTL haplotypes transmitted by founder $GP_3$ could not be inferred from the original study.

Quantitative trait locus validation began with an initial cross between the original founder $GP_3$ and $GP_5$. Their offspring were clonally propagated using somatic embryogenesis technology (von Arnold 1987) then planted at multiple locations. The next steps were two related studies based on other $GP_3$ pedigrees: one was based on seedlings and the other used additional somatic emblings.

Our study objective was to determine if the putative marker-QTL segment influencing HT4 was consistently expressed across related genetic backgrounds regardless of propagation method.

## Materials and methods

### Marker assays and QTL notation

In the original QTL detection study (Gwaze et al. 2003), PtTX3030 was defined as locus A with alleles 4, 12 and 16 (356, 324 and 300 bp, respectively). This microsatellite has a repeat motif $(TA)_5 \ldots (GGT)_{10}$ which is highly polymorphic; to date, 18 alleles have been found, as follows: 398, 364, 363, 356, 350, 348, 346, 342, 336, 333, 326, 324, 320, 314, 303, 300, 298 and 295 bp within *P. taeda*. Its sequence and its repeat structure is moderately conserved in hard and soft pines as well as sister taxon *Picea rubens* (Kutil and Williams 2001; Liewlaksaneeyanawin et al. 2004). PtTX3030 mapped closely to the putative QTL allele in the original study (Fig. 1). Its flanking marker was microsatellite PtTX3127 for linkage group 13 (Zhou et al. 2003). PtTX3127, shown in Fig. 1 as locus B, was also segregating in the original QTL pedigree for alleles 1, 2 and 5 (185, 181 and 168 bp, respectively). Nuclear microsatellite assays for the validation study were run by an independent lab, VizionSciTec Ltd., in Vancouver, BC, Canada using *P. taeda* microsatellite primer sets for PtTX3030 and PtTX2127. QTL effects are haplotypic and here they refer to effects associated with a respective marker allele inherited from a specific parent; QTL genotypic effect refers to the marker genotype of the individual.

### QTL pedigrees

The outcrossing conifer, *P. taeda*, has historically large, interconnecting populations extending throughout the southeastern US (Al-Rabab'ah and Williams 2002). Populations east and west of the Mississippi River Valley are mildly differentiated at DNA and phenotypic levels. Low linkage disequilibrium characterizes this species (Brown et al. 2004). For the original QTL pedigree, four unrelated founders or grandparents, known as $GP_1$–$GP_4$ were assayed (Gwaze et al. 2003). Parent $P_{34}$ was an offspring of $GP_3$ and $GP_4$. As such, it was thought to have received the $A_{12}$–$B_5$ haplotype from $GP_3$ and a phase-unknown haplotype $A_{16}$–$B_1$ from $GP_4$. The parental cross, $P_{12} \times P_{34}$, produced 90 seedling offspring which were measured for height at age 4 years (Fig. 1; Gwaze et al. 2003).

The first part was based on the validation cross between founders $GP_3$ and $GP_5$ (Fig. 1). These founders originated in wild populations in Williamsburg County, South Carolina ($GP_3$) and Onslow County North Carolina USA ($GP_5$). Their 46 offspring were clonally replicated from developing seeds using somatic embryogenesis then outplanted as emblings. Marker genotypes were assayed from laboratory cultures. For the subsequent step, QTL validation was conducted in two parts using (a) seedling offspring ($N = 97$) from the original QTL cross $GP_3 \times GP_4$ and (b) analysis of nine more marker genotypes from related $GP_3$

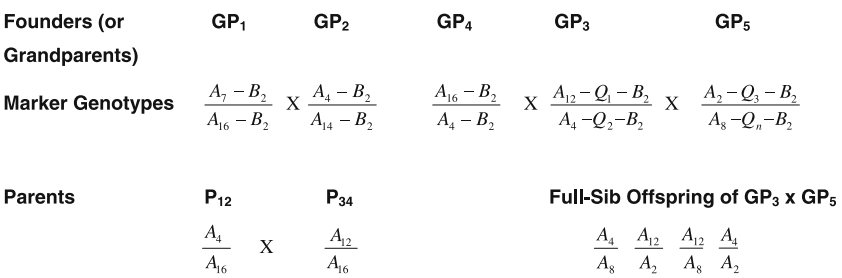


**Fig. 1** Relationship between the original QTL detection pedigree, now corrected, and the initial QTL validation cross $GP_3 \times GP_5$ used in this study. The second marker locus, B, was determined to be monomorphic. This finding collapsed the QTL validation analysis to a single marker

pedigrees in the same somatic embling tests as the validation cross (Fig. 2). The larger field experiment included 17 somatic embling families, seven full-sib families and four open-pollinated seedling families.

## Experimental field test design

For the initial validation cross, emblings were planted in single-tree incomplete block designs planted at each of five random Georgia USA locations: Glynn County; Wayne County; Perry County; Screven County and a nursery in Wayne County. As part of a planned imbalance design, clonal replicates for each of the 46 offspring genotypes ranged from zero to eight per location ($N$ = 1,017). Each location was planted with eight resolvable replications per site. The incomplete block design was generated using CycDesigN Version 2 (Whittaker et al. 2002) with an incomplete block size of 12 trees (six trees each in two adjacent rows).

Quantitative trait locus validation was conducted using data from only one cross within this larger field experiment.

One subsequent analysis, based on seedlings, was conducted using 97 offspring from the cross $GP_3 \times GP_4$ planted at a location separate from five locations planted with somatic emblings (Fig. 2). Another subsequent analysis was based on somatic emblings; these included nine clonal lines from three other crosses ($GP_3 \times OP$, $GP_3 \times GP_6$ and $P_{3-8} \times P_{9-10}$) planted in the same five locations as the validation cross (Fig. 2).

## Statistical analysis

All measurement data (all somatic and seedling families) across five locations were used in an analysis to estimate the fixed effect of location and variances for replication, incomplete block, family, clone within family and their interactions while an indicator variable included marker effects for the appropriate genotypes for the cross $GP_3 \times GP_5$. Thus marker effects were estimated using a single analysis. For the cross $GP_3 \times GP_5$, the QTL effects for height at 4 years (HT4) were estimated using SAS Proc Mixed with a linear model as follows:

$$\begin{aligned} y_{ijklmno} = {} & \mu + T_i + L_j + TL_{ij} + \gamma M_k + \gamma F_l + \gamma MF_{kl} + \gamma LM_{jk} \\ & + \gamma LF_{jl} + \gamma LMF_{jkl} + r_{m(j)} + b_{n(jm)} + \mathrm{fam}_{n(i)} \\ & + \lambda c_{o(n)} + l\mathrm{fam}_{jn(i)} + \lambda lc_{jo(n)} + r\mathrm{fam}_{mn(ij)} + e_{ijklmno.} \end{aligned}$$

Model terms included as fixed effects were as follows: the mean ($\mu$); propagule type as somatic family, full-sib family or open-pollinated family ($T_i$); location ($L_j$); the interaction of propagule type and location ($TL_{ij}$); QTL effects for $GP_3$ ($\gamma M_k$), and for $GP_5$ ($\gamma F_l$); QTL genotype ($\gamma MF_{kl}$); location by QTL effects ($\gamma LM_{jk}$ and $\gamma LF_{jl}$ for the $i$th location); and location by marker genotype effects ($\gamma LMF_{jkl}$) where $\gamma$ is an indicator variable that is 1 when the marker genotype is known and 0 otherwise. Random effects included replication within location ($r_{m(j)}$), incomplete block within location and replication ($b_{n(jm)}$), family within propagule type ($\mathrm{fam}_{n(i)}$), clone within family for propagule type 'somatic' ($\lambda c_{o(n)}$), location by family within propagule type interaction ($l\mathrm{fam}_{jn(i)}$), location by clone within family interaction for propagule type 'somatic' ($\lambda lc_{jo(n)}$), replication by family interaction within location and propagule type ($r\mathrm{fam}_{mn(ij)}$), and residual ($e_{ijklmno}$) for somatic propagules and $e_{ijklmn}$ for zygotic propagules) where $\lambda$ is an indicator variable which is 1 for propagule type 'somatic' and 0 otherwise. Different variances were allowed for propagule type for all random variables containing the subscript '$i$' including the residual. The phenotypic observation for HT4 for somatic propagules was $y_{ijklmno}$ but for other propagule types, the observation for HT4 was $y_{ijklmn}$. An alpha level of 0.05 was used for significance of the fixed effects.

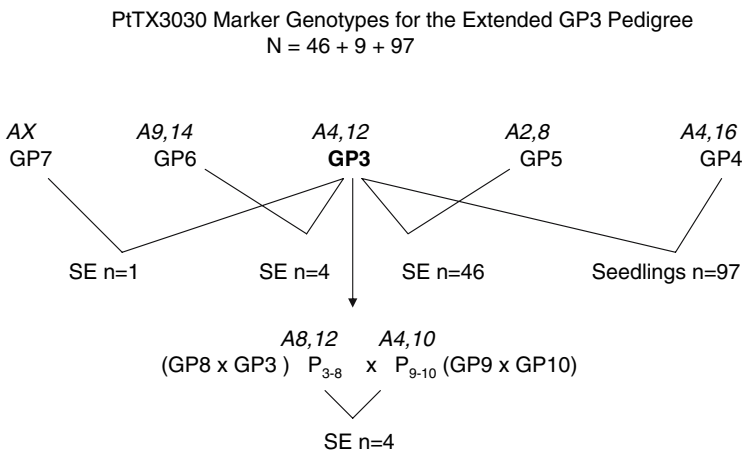


**Fig. 2** The extended pedigree, based on founder $GP_3$, was used for the combined QTL validation analyses conducted in three parts: **a** validation cross which was cloned via somatic embryogenesis or SE ($N$ = 46 genotypes) and **b** subsequent analysis based on seedlings ($N$ = 97) planted at a separate location and **c** subsequent analyses based on other crosses based on somatic emblings planted at the same locations as (**a**)

For the subsequent seedling analyses, the 97 seedlings were planted in a single block. For the subsequent analysis based on somatic emblings, nine marker genotypes were available from three more crosses planted at the same locations as $GP_3 \times GP_5$ (see Eq. 1).

## Results

### Validation: re-genotyping the original QTL pedigree

Re-genotyping $GP_3$ and $GP_4$ showed a new marker inheritance pattern. Locus B or PtTX3127 was monomorphic; its second band was a paralogous locus rather than an allele. This means, contrary to published results (Gwaze et al. 2003), that founders $GP_3$ and $GP_4$ had the same PtTX3127 allele 2 (181 bp). The correct chromosomal location for PtTX3127 has subsequently been reported in a consensus *Pinus* spp. linkage map (C. G. Williams, unpublished). This finding collapsed the QTL validation study to a single marker-QTL segment.

The corrected $GP_3$ haplotypes were now $A_{12}$–$B_2$ and $A_4$–$B_2$ and the corrected $GP_4$ haplotypes were $A_4$–$B_2$ and $A_{16}$–$B_2$ (Fig. 1). The new founder $GP_5$ was segregating for marker PtTX3030 with alleles $A_8$ (342 bp) and $A_2$ (364 bp) (Fig. 1). Note that the M-QTL segment was defined as a haplotype because the chromosomal segment, although of indeterminate genetic distance, can include other polymorphisms or even tightly linked, multiple QTL in addition to marker PtTX3030.

### Validation pedigree using a single marker

The largest QTL genotypic effect detected here was negative, not positive as expected (Tables 1, 2). Large non-additive or M × F interaction effects were detected for the validation cross, $GP_3 \times GP_5$ (Table 3). The combination of $GP_5$ allele $A_4$ and $GP_3$ allele $A_2$ together produced a visibly shorter phenotype (a mean difference of 0.9 m, which was statistically significant at $P < 0.0001$, Fig. 3). The other three marker genotypes ($A_2A_{12}$, $A_8A_{12}$ and $A_4A_8$) had HT4 means which were statistically similar ($P \geq 0.44$). The negative QTL effect was consistently associated with marker genotype $A_2A_4$ across sites (Table 2).

Marker genotype × environment (G × E) effects were also present although most of the effect was due to scaling rather than true rank change; this could be confirmed by re-analysis with and without a log transformation for HT4. The HT4 means at each location for the other three marker genotypes were not statistically different although slight changes in rank did occur across locations (Tables 2, 3). In particular, the $A_2A_4$ marker genotype means showed slight rank change only at site 3 which showed compressed genotypic differences (Table 2). Thus G × E effects here were largely due to scaling rather than rank changes, with the exception of site 3.

The extreme non-additive effect shown in Fig. 2 occurred for marker genotype class $A_{4-8}$ which had the highest HT4 value of +0.24 m and marker genotype class $A_{2-4}$ which had lowest HT4 value of –0.58 m. The $A_2A_4$ genotypic class had the lowest QTL effect at each location; only 18 individuals were present in total but genotypic distortion was not statistically significant (Table 2). As an example, individuals with the $A_2A_4$ marker genotype included three extreme phenotypes which were visibly shorter in the field testing. These had HT4 values of 2.17, –1.98 and –1.22 m relative to their test mean. These three phenotypes exemplified the visible evidence associated with this marker genotype. Further study of this extreme phenotypic effect associated with the $A_2$ allele was not possible here because no other founders in this eastern population had this allele.

The subsequent analysis based on 97 seedlings showed no statistically detectable QTL effect for any marker alleles or genotypes (Tables 4, 5). Either marker allele $A_{12}$, in contrast to Gwaze et al. (2003), did not exert a true positive QTL effect on HT4 measured on seedlings or its statistical power for discerning among marker genotypes in the single block plot was too weak. The QTL effect associated with the $A_{12}$ allele on HT4 could not be statistically detected here.

The subsequent analysis based on somatic emblings also yielded weak QTL validation for a long-lived woody perennial. Here, the $A_{12}$ marker genotype class from $GP_3 \times GP_6$ had a positive influence on HT4 which was statistically higher than the $A_4$ marker genotype (Table 6). All offspring from the other two crosses (Fig. 2) shared a single marker genotype, $A_4$–$A_{12}$ despite the fact that these crosses should have been fully informative (Fig. 2). Also, no allelic contrast was possible for six of the nine marker genotypes.

**Table 1** PtTX3030 allele and genotypic effects (and standard errors) for HT4 for QTL validation cross $GP_3 \times GP_5$ shown here

| Source | PtTX3030 allele or genotype | HT4 effect (m) |
|---|---|---|
| $GP_3$ | $A_4$ | 5.59 ± 0.26 |
| | $A_{12}$ | 5.95 ± 0.25 |
| $GP_5$ | $A_2$ | 5.58 ± 0.26 |
| | $A_8$ | 5.96 ± 0.26 |
| $GP_3 \times GP_5$ | $A_2$–$A_4$ | 5.20 ± 0.29 |
| | $A_4$–$A_8$ | 5.98 ± 0.27 |
| | $A_{12}$–$A_2$ | 5.96 ± 0.27 |
| | $A_{12}$–$A_8$ | 5.94 ± 0.28 |

**Table 2** QTL effects for four marker genotype classes from $GP_3 \times GP_5$ by location

| PtTX3030 microsatellite genotype | Location-1 (count) | Location-2 (count) | Location-3 (count) | Location-4 (count) | Location-5 (count) | Total count |
|---|---|---|---|---|---|---|
| $A_{2–4}$ | –2.83 (3) | –3.37 (5) | –0.74 (5) | –2.40 (2) | –0.30 (3) | 18 |
| $A_{2–12}$ | –0.17 (10) | +0.06 (11) | +0.63 (10) | 0 (11) | +0.20 (7) | 49 |
| $A_{4–8}$ | +0.62 (8) | +1.48 (9) | +0.29 (9) | +0.50 (7) | –0.20 (6) | 39 |
| $A_{8–12}$ | +0.87 (6) | +0.49 (6) | –0.86 (6) | +0.10 (7) | +0.10 (5) | 30 |

Results were based on a total of 46 offspring cloned via somatic embryogenesis then planted across five locations using an alpha-lattice planned imbalance design. Count for each marker genotype per location is shown in parentheses then summed across locations

**Table 3** Results for the validation cross, $GP_3 \times GP_5$

| Source | Type III *F*-value | Pr > *F*-value analysis |
|---|---|---|
| Propagule type | 13.71 | 0.006 |
| Study | 12.53 | <0.001 |
| Propagule × study | 0.99 | 0.481 |
| $GP_3$ alleles | 2.23 | 0.111 |
| $GP_5$ alleles | 3.69 | 0.057 |
| $GP_3 \times GP_5$ | 11.61 | <0.001 |
| Study × $GP_3$ | 0.81 | 0.596 |
| Study × $GP_5$ | 6.88 | <0.001 |
| Study × $GP_3$ × $GP_5$ | 4.49 | 0.002 |

Analysis of variance for height at 4 years (HT4) based on generalized least squares results. Offspring from a single cross ($N = 46$) were tested across five locations using *Pinus taeda* somatic emblings within a larger sample of pedigrees and propagule types

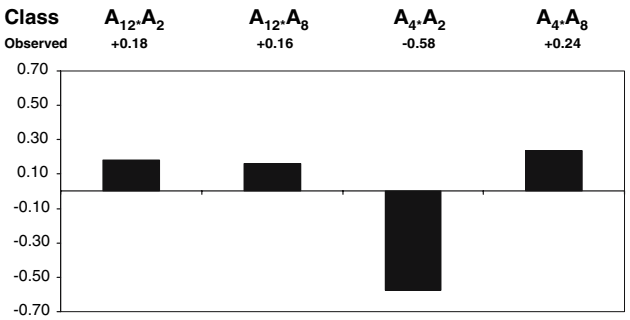


**Fig. 3** Marker genotypic effects estimated for the QTL validation pedigree shown for adjusted height at age 4 years (*HT4*). Only the marker genotype $A_2A_4$ had a phenotype which was statistically different from the other three

## Discussion

The large, positive QTL effect reported in Gwaze et al. (2003) was validated only in the sense that marker allele $A_{12}$ had a consistently positive effect on HT4 when compared to the other alleles; it could not be detected statistically using a different cohort of 97 full-sib seedlings planted in a single block. Part of the explanation is that

**Table 4** Subsequent QTL analysis results for 97 seedlings

| Effect | *F*-value | Pr > *F* |
|---|---|---|
| $GP_3$ | 0.95 | 0.3314 |
| $GP_4$ | 0.22 | 0.6411 |
| $GP_3 \times GP_4$ | 0.68 | 0.4121 |

Founders $GP_3 \times GP_4$ were part of the original QTL pedigree. No marker allele or marker genotype had a statistically significant effect on HT4

**Table 5** PtTX3030 allele and genotypic effects (and standard errors) on HT 4 for 97 seedlings shown here

| Source | PtTX3030 allele or genotype | HT4 effect (m) |
|---|---|---|
| $GP_3$ | $A_4$ | 5.05 ± 0.23 |
| | $A_{12}$ | 5.16 ± 0.23 |
| $GP_4$ | $A_4$ | 5.13 ± 0.23 |
| | $A_{16}$ | 5.08 ± 0.23 |
| $GP_3 \times GP_4$ | $A_4$–$A_4$ | 5.12 ± 0.24 |
| | $A_4$–$A_{16}$ | 4.98 ± 0.24 |
| | $A_{12}$–$A_4$ | 5.14 ± 0.25 |
| | $A_{12}$–$A_{16}$ | 5.18 ± 0.25 |

re-genotyping in this study removed flanking markers and thus reduced statistical power. The magnitude of the QTL effect is now confounded with the recombination distance from the marker locus. Experimental design also contributes in that planting seedlings in a single block was less accurate for QTL estimation than the original QTL study itself which was planted in single-block plots across several locations (Gwaze et al. 2003). Otherwise, there is no doubt that the original analytical methods were correct; QTL analysis based on flanking markers coupled with regression analysis is widely accepted for complex outbred pedigrees (e.g. Haley et al. 1994; Reyes-Valdés and Williams 2002). Statistical efficacy is the more likely explanation here because in all cases, allele $A_{12}$ had a consistently positive effect on HT4 whether measured on seedlings or somatic emblings.

**Table 6** Subsequent QTL analysis results for additional somatic emblings

| Source | Marker genotype | HT4 effect (±SE) in meters |
|---|---|---|
| $GP_3 \times GP_6$ | $A_4$–$A_9$ | 5.44 ± 0.15 |
| | $A_{12}$–$A_9$ | 6.63 ± 0.10 |
| | $A_{12}$–$A_{14}$ | 6.94 ± 0.10 |

Only least squares mean contrasts for only one of the three crosses, $GP_3 \times GP_6$, could be estimated at the genotypic level due to severe marker genotype imbalance. Weak yet independent confirmation is shown for PtTX3030 marker genotypes $A_{12}$–$A_9$ and $A_{12}$–$A_{14}$ relative to $A_4$–$A_9$

This QTL validation experiment also provided a new insight into the nature of interaction: a strong negative QTL genotypic effect was both statistically and visually apparent when $GP_3$'s other marker allele, $A_4$, occurred in combination with marker allele $A_2$ as in the case of the $GP_3 \times GP_5$ cross. This finding is distinct from another deleterious allele transmitted by the same $GP_3$ founder, the *cad*-null mutant (Wu et al. 1999). Although reduced height was originally reported for heterozygote carriers of this lethal allele (Wu et al. 1999), a more recent study by the same group showed no effect on height (Yu et al. 2005) so this QTL locus cannot be linked to the *cad*-null locus. This findings of a negative genotypic effect thus adds a new insight into QTL architectural findings for *Pinus* spp. which currently include (1) large-effect QTL haplotypes are not segregating for the same *P. taeda* mapping pedigree (with a different offspring cohort) also measured for early height (Kaya et al. 1999) and (2) large QTL effects are rarely found as part of the genetic architecture for growth traits (Devey et al. 2004).

Another question which surfaces with the use of a long-lived perennial species is why use HT4 instead of older measurements? While most members of the Pinaceae are harvested at ages 25 years or older, there is benefit in early assessment. Consider the case of somatic embryogenesis technology where the immediate benefit for HT4 data can be reduced field testing burden. Only tissue from developing seeds can be cloned so no phenotypic information is available at the time that somatic embryogenesis takes places. From developing seeds come infinite numbers of somatic embryos which are cultured. Some are frozen and others are outplanted in field tests because the decision to clone must occur before any phenotypic information is directly available for the individual. Somatic emblings in the field tests are then measured early age phenotypic information such as HT4 (see review in Gwaze et al. 2002) and assayed for QTL-linked markers. Using these early selection criteria, one can now eliminate the worst phenotypes from further propagation. In this specific example, one would discard the $A_2A_4$ marker genotypes in the freezer so that these individuals could not planted on a larger commercial scale.

## A proposed model for multiple QTL effects linked to a single marker locus

The right genetic model underlying these QTL validation must account for the following experimental results. First, different marker-QTL haplotypes are segregating in these extended pedigrees. Second, one or more genotypic combinations exert a strong effect on HT4. Founder $GP_3$ can be inferred to have two haplotypes: $A_{12}$–$Q_1$ and $A_4$–$Q_2$. Founder $GP_5$ can be inferred to have at least one (or two) additional QTL haplotypes, designated here as $A_8$–$Q_n$ and $A_2$–$Q_3$. The third QTL can be inferred because only $Q_2$ elicited the non-additive interaction with $Q_3$. The haplotype $A_8$–$Q_n$ transmitted by $GP_5$ cannot be more fully determined here because $Q_n$ exerts a similar effect whether combined with $Q_1$ or $Q_2$. In either case, a simple genetic model predicts that offspring from the $GP_3 \times GP_5$ cross are segregating for three or more different QTL alleles linked to the marker PtTX3030 (although a more complex epistatic model assuming multiple loci rather than multiple alleles cannot be ruled out here).

Our simple M-QTL model for multiple QTL alleles rests on three critical assumptions: (1) that M-QTL linkage disequilibrium is imposed by use of a common founder so that each marker allele is tightly associated with its given Q allele in parts of the extended pedigree, (2) that the marker alleles associated with the large-effect QTL genotype each occur at rare frequency and (3) that somatic emblings have a different epigenetic trajectory relative to seedlings such that the same QTL model is not appropriate for both propagule types. How realistic are these three assumptions?

First, use of the common founder $GP_3$ in all pedigrees should increase linkage disequilibrium but this cannot be estimated without the statistical power inherent to flanking markers. Single-marker analysis, as shown here, is limited by the fact that the magnitude of the QTL effect is confounded with its map distance from its marker. Second, large-effect QTL haplotypes occur at rare frequencies in model organisms (MacKay 2001, p. 407). As a crude test of this possibility, we checked allele frequencies for PtTX3030 in a rangewide *P. taeda* population survey (Al-Rabab'ah and Williams 2002). These data showed that $A_2$ was indeed rare in eastern *P. taeda* populations but occurs at higher frequency in the western part of the species' range (Al-Rabab'ah and Williams 2002; C. G. Williams unpublished data). Testing allele $A_2$ further for its associated QTL effect would thus be likely only if one drew on pedigrees drawn from admixed or western populations. For the third assumption, one would require paired

comparisons of seedlings versus somatic emblings and this has not yet been reported. If developmental trajectory did prove to be different between these propagule types then one would expect QTL validation results to differ between them.

While it is true that combined results for allele $A_4$ did differ between seedlings and somatic emblings here, the two propagule types were not directly compared using the exact same cross or testing design. For example, the 97 seedling offspring, if better replicated beyond a single block plot, could have yielded better phenotypic measurements such that M-QTL results were statistically enhanced. A partial diallel design for QTL validation would be more efficient than repeated use of the same cross at different locations.

The multiple locus model deserves some mention here: the strong negative QTL genotypic effects could be due to multiple loci clustered so that they are linked to the marker PtTX3030 locus on one or both sides of the marker ($Q_{11}$–$A_{12}$–$Q_{22}$ or $Q_{11}$–$Q_{21}$–$A_{12}$). If so, one could hypothesize that different QTL alleles are segregating at each QTL locus ($Q_{11}$–$A_{12}$–$Q_{22}$ versus $Q_{11}$–$A_{12}$–$Q_{21}$) for different $GP_3$-related crosses. Such allelic differences for QTL loci would be confounded with propagule types if the same exact cross was not directly compared. This is a complex yet intriguing QTL model. Verifying such a model requires estimating founder-probability origins, fine-mapping methods and saturated markers closely linked to the QTL region especially if physical map data are not available.

## Summary

From this case study, QTL haplotype validation and inference of multiple QTL alleles segregating at a single marker locus was feasible. The original QTL haplotype was weakly validated yet a negative QTL effect was detected at the same marker locus. Our results show that culling rare, negative-effect QTL haplotypes associated with PtTX3030 alleles $A_2A_4$ offers direct benefit to reducing field testing burden. A multiple-allele QTL haplotype model could be inferred for somatic emblings. Multiple marker alleles, a highly heterozygous species and use of a shared founder between QTL detection and validation pedigrees together yielded plausible QTL models for further hypothesis-testing. This case study for QTL validation in a long-lived perennial species yet to be fully domesticated has relevance to highly heterozygous outcrossing organisms whether forest trees, perennial grasses, invertebrates, insects or marine mollusks.

**Acknowledgements** Research support to C.G. W. was provided by Canada-US Fulbright Program, USDA-Forest Service Southern Experiment Station and USDA Biotechnology Risk Assessment Program. Special appreciation goes to Dr. David Gwaze, Missouri Department of Conservation, Columbia, MO, USA; in addition we received research assistance from Ms. Margarita Gilbert, Cell For, Vancouver, BC, Canada; Dr. David Ellis, USDA; Mr. Paul Belonger, Plum Creek Timber Company, USA and Dr. Craig Newton, Vizion-SciTec Ltd., Vancouver, BC, Canada.

## References

Al-Rabab'ah M, Williams CG (2002) Population dynamics of *Pinus taeda* L. based on nuclear microsatellites. For Ecol Manage 163:263–271

von Arnold S (1987) Improved efficiency of somatic embryogenesis in mature embryos of *Picea abies* (L.) Karst. J Plant Physiol 128:233–244

Bradshaw HD, Foster GS (1992) Marker-aided selection and propagation systems in trees—advantages of cloning for studying quantitative inheritance. Can J For Res 22:1044–1049

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

Devey ME, Carson SD, Nolan MF, Matheson AC, Riini CT, Hohepa J (2004) QTL associations for density and diameter in *Pinus radiata* and the potential for marker-aided selection. Theor Appl Genet 108:516–524

van Eck HJ, Jacobs J, Stam P, Ton J, Stiekema WJ, Jacobsen E (1994) Multiple alleles for tuber shape in diploid potato detected by qualitative and quantitative genetic analysis using RFLPs. Genetics 137:303–309

Groover A, Devey M, Fiddler T, Lee J, Megraw R, Mitchell-Olds T, Sherman B, 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

Gwaze DP, Bridgwater FE, Williams CG (2002) Genetic analysis of growth curves for woody perennials: a case study for *Pinus taeda* L. Theor Appl Genet 105:526–531

Gwaze DP, Zhou Y, Reyes-Valdes MH, Al-Rababah MA, Williams CG (2003) Haplotypic mapping in an outbred pedigree. Genet Res 81:43–50

Haley CS, Knott SA, Elsen J-M (1994) Mapping quantitative trait loci in crosses between outbred lines using least squares. Genetics 136:1195–1207

Kaya Z, Sewell MM, Neale DB (1999) Identification of quantitative trait loci influencing annual height- and diameter-increment growth in loblolly pine (*Pinus taeda* L.). Theor Appl Genet 98:586–592

Kutil BL, Williams CG (2001) Triplet repeat microsatellites shared among hard and soft pines. J Hered 92:327–332

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756

Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA, Ritland K (2004) Single-copy, species transferable microsatellite markers developed from loblolly pine ESTs. Theor Appl Genet 109:361–369

MacKay T (2001) The genetic architecture of quantitative traits, Chap 17, pp 391–411. In: Wagner G (ed) The character concept in evolutionary biology. Academic, New York, 624p

Maliepaard C, Jansen J, van Ooijen JW (1997) Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. Genet Res 70:237–250

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

Reyes-Valdés MH, Williams CG (2002) A haplotypic approach to founder-origin probabilities and outbred QTL analysis. Genet Res 80:231–236

Verhoven KJF, Jannink JL, McIntyre LM (2006) Using mating designs to uncover QTL and the genetic architecture of complex traits. Heredity 96:139–149

Whittaker D, Williams ER, John JA (2002) CycDesigN version 2—a package for the computer generalization of experimental design. CSIRO, Canberra, Australia

Williams CG (1998) QTL mapping in outbred plants. In: Paterson A (ed) Molecular dissection of complex traits. CRC Series, Boca Raton, FL, pp 81–94, Chap 5

Williams CG, Reyes-Valdes MH (2007) Estimating a founder's genomic proportion for each descendant in an outbred pedigree. Genome 50:289–296

Wu RL, Remington DL, MacKay JJ, McKeand SE, O'Malley DM (1999) Average effect of a mutation in lignin biosynthesis in loblolly pine. Theor Appl Genet 99:705–710

Yu Q, McKeand SE, Nelson CD, Li B, Sherrill JR, Mullin TJ (2005) Differences in wood density and growth of fertilized and nonfertilized pine associated with a mutant gene, *cad*-n1. Can J For Res 35:1723–1730

Zhou Y, Gwaze DP, Bui T, Reyes-Valdés MH, Williams CG (2003) No clustering for linkage map based on low-copy and undermethylated microsatellites. Genome 46:809–816