ORIGINAL ARTICLE

Estimation of number and size of QTL effects in forest tree traits

David Hall[1,2] · Henrik R. Hallingbäck[1,3] · Harry X. Wu[1,4]

Abstract Mapping the genetic architecture of forest tree traits is important in order to understand the evolutionary forces that have shaped these traits and to facilitate the development of genomic-based breeding strategies. We examined the number, size, and distribution of allelic effects influencing eight types of traits using 30 published mapping studies (linkage and association mapping) in forest trees. The sizes of allelic effects, measured as the phenotypic variance explained, generally showed a severely right-skewed distribution. We estimated the numbers of underlying causal effects (n_{qtl}) for different trait categories by improving a method previously developed by Otto and Jones (Genetics 156:2093–2107, 2000). Estimates of n_{qtl} based on association mapping studies were generally higher (median at 643) than those based on linkage mapping (median at 33). Comparisons with simulated linkage and association mapping data suggested that the lower n_{qtl} estimates for the linkage mapping studies could partly be explained by fewer causal loci segregating within the full-sib family populations normally used, but also by the cosegregation of causal loci due to limited recombination. Disease resistance estimates based on linkage mapping studies had the lowest median of four underlying effects, while growth traits based on association mapping had about 580 effects. Theoretically, the capture of 50% of the genetic variation would thus require a population size of about 200 for disease resistance in linkage mapping, while growth traits in association mapping would require about 25,000. The adequacy and reliability of the improved method was successfully verified by applying it to the simulated data.

Keywords Association mapping · Linkage (QTL) mapping · Linkage disequilibrium · QTL number estimate · Size of QTL effect

Communicated by D. Grattapaglia

✉ Harry X. Wu
  harry.wu@slu.se

  David Hall
  david.hall@umu.se

  Henrik R. Hallingbäck
  henrik.hallingback@slu.se

[1] Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, 90183 Umeå, SE, Sweden

[2] Department of Ecology and Environmental Sciences, Umeå University, 90185 Umeå, SE, Sweden

[3] Department of Plant Biology, Swedish University of Agricultural Sciences, 75007 Uppsala, SE, Sweden

[4] Black Mountain Laboratory, CSIRO NRCA, ACT, Canberra 2601, Australia

Introduction

It is fundamentally important to understand the complex architecture of adaptive and commercial traits in order to assess the role of evolutionary forces in shaping the genetic variation and to improve long-term breeding strategy. The introduction of genetic markers held promises of fast and reliable diagnostics of genetic diseases in humans, marker-assisted selection in animal and plant breeding and a better understanding of the genetic basis of adaptation (Neale and Savolainen 2004). In maize and *Arabidopsis*, the genetic basis underlying several traits has been relatively well identified as major portions (>50 and >80%, respectively) of the heritable trait variation was

explained by both large and small effects underlying quantitative trait loci (QTL) (Buckler et al. 2009; Segura et al. 2012). In contrast, human height is considered to be highly heritable but most of the genetic variation was still unexplained by genome-wide association studies (GWAS). This unexplained genetic variation was termed "the missing heritability" (Maher 2008). More recent studies point to an underlying architecture of numerous QTL each having a small effect on human height (Yang et al. 2010). Individually, these QTL thus remain undetected by GWAS methods but, together, they nonetheless explain a majority of the genetic variation. In line with this, theoretical work has shown that alleles of large effect are likely unusual and allele effects have been suggested to follow a negative exponential distribution (Orr 1998). This in turn means that most of the heritability would be explained by many alleles with small effect size (McKown et al. 2014; Porth et al. 2013; Rockman 2012; Yang et al. 2010) and the magnitude of the detected allele effects would form a truncated exponential distribution (Otto and Jones 2000). In studies of flowering time in maize and *Arabidopsis thaliana*, the allele effect distribution of detected QTL has been described as a rapid increase of the number of observed QTL for effect magnitudes just above the detection threshold but thereafter decreasing dramatically at larger effect sizes, thus forming a "long tail" of rare large QTL effects (Buckler et al. 2009; Salomé et al. 2011). This trend appears to be similar in most plants when looking at effect size distribution of detected QTL effects (Ingvarsson and Street 2011).

Historically, QTL were detected through linkage analyses in crossing designs between extreme phenotypes or related species (termed QTL studies). QTL studies identified significant effects associated with genetic segments cosegregating due to physical linkage. For the QTL studies treated in this paper, the extent of such genetic segments ranged from 0.9 to 60 cM in terms of genetic distance (Henery et al. 2007; Jorge et al. 2005; Paterson et al. 1988; Thumma et al. 2010). However, QTL detection at higher resolution could instead be performed through linkage disequilibrium (LD) analyses, also termed association mapping (Lander and Schork 1994; Pritchard et al. 2000). Due to limitations in genotyping technology and to lack of resources, association mapping has so far mostly been performed using the candidate gene approach (Thavamanikumar et al. 2013). Such studies may however fail to find unexpected associations as the candidate genes are preferentially selected in areas indicated to be important by previous QTL studies (Frewen et al. 2000; Ingvarsson et al. 2006) or through other research efforts. Nonetheless, as large-scale genotyping becomes increasingly affordable, the genome-wide approach association mapping will likely be the preferred method as it does not suffer from the candidate gene selection constraints. Notably, both QTL and association studies are often limited by the relatively small mapping population sizes resulting in low statistical power, thus rendering

small or even medium effect QTL statistically nonsignificant and difficult to detect. Such statistically underpowered populations may also suffer from severe inflation of effect size estimates (so-called Beavis effect), in particular if very strict thresholds for significance are applied (Beavis 1994; Beavis 1998; Xu 2003b). The magnitude of the overestimation depends not only on the power of the study but also on the true effect size. If the true effect magnitude is small, a large sample size (e.g., thousand individuals) is required to enable unbiased effect estimation (Beavis 1998).

While the theoretical groundwork in QTL and association studies has been available for some time, accumulated practice and experience now indicate the need for substantial sample sizes given the objective of detecting a large portion of the heritable variation observed (Lynch and Walsh 1998; Sham et al. 2000). If the sampled markers do not include the causal allele or if the LD between the marker and the causal allele is incomplete, power is reduced and the size of the mapping population need to be increased even further (Porth et al. 2013). The number of association studies in forest trees is still limited (Thavamanikumar et al. 2013), but with the promise of current genomics efforts in several species (Birol et al. 2013; Neale and Kremer 2011; Nystedt et al. 2013), the expectations on the dissection of complex traits and efficiency of breeding utilizing the molecular markers for genomic selection (GS) are increasing (Meuwissen et al. 2001). Recent common garden trials in *Populus* along with the completion of the *Populous trichocarpa* genome sequence (Tuskan et al. 2006) have resulted in a few genome-wide association mapping studies on *P. trichocarpa* where the number of markers is much greater than previous candidate gene approaches (Evans et al. 2014; McKown et al. 2014; Porth et al. 2013). However, several mapping studies in forest trees suggest that most traits have a large number of underlying QTL, and to date, there are very few studies that have a mapping population size that provide the power needed to capture a large proportion of the genetic variation (Neale and Savolainen 2004; Thavamanikumar et al. 2013). Investigating the expected number of underlying QTL effects and their distribution can enhance our understanding of the genetic architecture of important quantitative traits, design better experiments to uncover missing factors of heritability, and facilitate implementation of marker-assisted selection (MAS) and genomic selection (Isik 2014; Meuwissen and Goddard 2010; Thavamanikumar et al. 2013). In this paper, we suggest a modified method from Otto and Jones (2000) in order to estimate the number of underlying QTL from the reported size of phenotypic variance explained of significant markers for eight different trait types. We also use simulated data in order to verify our method for association mapping and QTL mapping procedures and to potentially reproduce differences between association mapping studies and QTL mapping studies in terms of number of QTLs and the size distributions of their effects. Furthermore, we provide estimates of the

required population size for QTL mapping and association mapping studies aiming to detect at least 50% of the underlying genetic variation for eight different trait types.

Material and methods

This study contains two main parts where the first *meta-analysis* examines published data from QTL and association mapping studies of forest trees. Through this examination, effect sizes of detected QTL were surveyed and compiled and the underlying number of effective loci, both detected and undetected, were estimated. The second *simulation* part examined several simulated populations based on a common designed genome harboring a specified number of QTL of known effect sizes to allow us to test the validity of the method used to estimate the number of underlying QTL. Furthermore, the simulated data was designed in such a way that differences between the meta-analysis QTL and association studies could be artificially reconstructed and thus verified.

Meta-analysis

We searched the forest tree literature, primarily for association mapping studies reporting several loci significantly influencing a trait and secondly for QTL linkage studies treating similar traits in order to enable comparisons. We selected studies where three or more significant loci had been identified for any single trait in order to compare the effect distributions and to estimate the total number of QTL segregating in the studied QTL and association mapping populations. We examined 15 association studies (of which two conducted genome-wide scans) and 15 QTL studies (see Supplementary Table 1). These studies report several detected effects potentially providing analyses with the greatest statistical power. Because absolute allele effect sizes (α) were generally unavailable, we used the reported correlation coefficients (r^2) or proportion of phenotypic variance explained (PVE) for individual QTL as a measure of the effect size. The thresholds used to declare significance varied among studies, but in order to harmonize the criteria as far as possible, we applied either a (i) a false discovery rate (FDR) threshold of $Q \leq 0.1$, (ii) a chromosome-wide corrected error rate of $p < 0.05$, or (iii) a genome-wide corrected error rate of $p < 0.1$. Sums of PVE estimates over all detected QTL for each trait were also calculated as a measure of the collective PVE of all detected QTL. Even though such sums may be inflated due to LD between QTL, it was still used because proper estimates of the total PVE obtained by multilocus model analyses (Guan and Stephens 2011; Segura et al. 2012; Xu 2003a) were infrequently available. The examined traits were very numerous and diverse, and to simplify comparisons, we categorized them into eight trait types, i.e., wood chemistry (Chem), physical wood properties (Phys),

water use efficiency (Water), phenology (Phen), cold tolerance (Cold), secondary metabolites (Metab), disease resistance (Disease), and growth (Growth).

Number of underlying QTL effects

Several methods have been developed for the estimation of the number of underlying QTL (e.g., Castle 1921; Goddard 2009; Hayes and Goddard 2001; Lande 1981; Otto and Jones 2000; Park et al. 2010). In most cases, these methods are difficult to apply to forest mapping studies due to limited availability of repeated studies in the same species and traits, or due to the lack of good estimates of relevant population parameters. The assumption made about the distribution of QTL effects is that each analysis has a threshold value (θ) at the lowest detectable effect size which is inversely proportional to the power of the analysis.

We made the assumption that the underlying allelic effects follow a negative exponential distribution. This type of distribution has such properties that the number of underlying allelic effects can be estimated from observed characteristics of the detected effects. Otto and Jones (2000) provided a method of estimating the number of QTL underlying a trait from the distribution of allelic effects (α), assuming they were exponentially distributed. However, most association and QTL mapping studies do not provide estimates of allelic effects but rather the PVE. Given that PVE $\propto \alpha^2$, we therefore suggest that its distribution could be represented by the square of an exponentially distributed variable. We adapted the method of Otto and Jones (2000) to estimate the number of underlying QTL given a squared exponential distribution (see Supplementary material 1 for details):

$$f_{pve}(x) = \frac{\lambda}{2\sqrt{x}} e^{-\lambda\sqrt{x}} \qquad (1)$$

where λ is the rate parameter and x is the PVE. The estimation method assumes that all allelic effects are additive and segregate independently (see Supplementary material 1 for further motivation). We thus estimated the number of effective QTL underlying a trait (n_{qtl}) from the heritability (h^2), average PVE of detected QTL (μ_d), and the detection threshold (θ) according to the equation:

$$n_{qtl} = \frac{h^2}{\mu_d - \sqrt{2\mu_d\theta - \theta^2}} \qquad (2)$$

The validity of this method for QTL number estimation purposes was verified using simulated data (see subsection Simulation study). The numbers of underlying QTL were estimated for each trait in each study where three or more effects had been detected. Similarly, we used a modified method also adapted from Otto and Jones (2000) in order to estimate the

lowest detectable PVE (θ) which uses the PVE estimates of the QTL detected in each analysis (Supplementary material 1). As narrow-sense heritabilities are seldom reported in association and QTL mapping studies, we used heritability estimates from the meta-analysis of (Cornelius 1994). For association mapping studies, this implies the use of different heritabilities for the different trait types (for example, 0.24 for growth, 0.40 for physical wood properties, 0.40 for metabolite, 0.70 for phenology, 0.50 for water use efficiency, 0.50 for wood chemistry properties, 0.25 for cold tolerance, and 0.50 for disease resistance traits). It should also be noted that most QTL mapping studies include only single full-sib families in which only half of the additive genetic variance (sometimes called the *segregation deviation*) is available on average (Lynch and Walsh 1998). Assuming the environmental variance to be constant, the *local narrow-sense heritability* within QTL mapping populations (h^2_l) was thus calculated as follows:

$$h^2_l = \frac{h^2}{2-h^2} \tag{3}$$

By combining the n_{qtl} estimate and the trait heritability, we also estimated the mapping population size required to detect QTL accounting for 50% of the genetic variation. We did this by assuming that the QTL having the largest true PVE would be most easy to detect and, given the distribution, subsequent QTL with smaller PVE would be successively detected, with increasing population size, until the desired amount of genetic variance was explained. Within this set of QTL, the QTL with the smallest PVE constitutes the required detection threshold level for explaining the desired amount of genetic variance thereby also determining the needed mapping population size (see Supplementary material 2 for further details) (Lynch and Walsh 1998; Sham et al. 2000). We provide an R-function to estimate number of underlying QTLs and necessary mapping population size from a set of detected QTL in Supplementary material 3.

Simulation study

We simulated QTL and association mapping populations primarily to verify the n_{qtl} estimation method, modified from Otto and Jones (2000), and its application to the distribution of PVE rather than allelic effects. In addition, we wanted to examine the relationship between the phenotypic variation explained and the number of detected QTL, and also the observed discrepancy between association and QTL mapping PVE. This would allow us to compare results from the simulation with the observed data in the meta-analysis and in turn determine whether the observed difference between QTL and association studies could be reproduced from the two types of mapping methods using simulated data.

The simulated genome consisted of a single chromosome comprising 1000 evenly spaced biallelic loci and surrounded by an additional 1000 biallelic dummy loci in order to facilitate crossover events with greater realism. Of the chromosomal loci, 49 were randomly assigned to be causal QTL. The allelic effects (α) were drawn from a negative exponential distribution of with the rate parameter $\lambda = 150$. The distribution of observed PVE depends however on allele frequencies as well as the square of the allelic effects. Consequently, PVE were more variable in the simulated population because allele frequencies were assigned randomly according to a beta distribution $Beta(\alpha, \beta)$, where $\alpha = \beta = 0.7$. This distribution describes a frequency spectrum where rare variants are somewhat more common than intermediate variants as observed in published data (Eckert et al. 2010). We utilized the software "Metagene" described by Hallingbäck et al. (2014) and Sanchez et al. (2008) to simulate the QTL crosses and the association mapping populations. We set the number of crossover events to 3 per generation and offspring individual. The crossovers were randomly distributed across all chromosome and dummy loci (2000 in total). Such a design results in a linkage group with a theoretical genetic length of 150 cM (subsequently confirmed by linkage mapping) and which is fairly representative for larger linkage groups in forest tree species (Acheré et al. 2004). The simulated trait was completely additive with a global narrow-sense heritability (h^2) of 0.5 created by adding environmental random noise to the breeding values. The deviation caused by environmental factors were sampled from a normal distribution $N(0, \sigma^2_E)$ with an environmental variance of σ^2_E. Using the aforementioned design parameters as a framework (loci, allele effects, allele frequencies, crossover distribution, and environmental variation), 1000 unrelated founder individuals were created as a base for the simulated association mapping population. From this founder base, 500 full-sib families were generated by single pairwise mating of the founder individuals with 4 offspring each, thus producing 2000 individuals available for analysis. Using the same framework, simulated QTL mapping populations were produced based on 10 repeated simulations each featuring 6 unique crosses between 4 monoecious parent founders generating 500 offspring for each cross.

For the simulated QTL crosses, the true PVE, of a QTL detected in analysis, was calculated by creating partial breeding values (BV) based on the assigned effects of the loci situated within the significant portion of the QTL peak:

$$BV_{k,i} = \sum_{j\in k} x_{j,i}\alpha_j \tag{4}$$

where α_j is the average genetic effect of allelic substitution at causal locus j and $x_{j,i}$ is the locus genotype (number of alleles 0, 1, or 2). These products ($x_{j,i} \times \alpha_j$) are then summed for all loci detected within the peak of QTL k for individual i. For the simulated association mapping populations, the extent of

physical linkage was assumed to be very short and a detected QTL was thus considered to comprise only the detected locus in isolation (i.e., $k = j$). The partial breeding value for all loci within all detected QTL was calculated simply by extending Eq. 4 to perform summation over all j included in any k. The true PVE for the segregating causal loci within one or all detected QTL (or effective loci in association analyses) then becomes the ratio between the variance of partial breeding values (BV_k) and the variance of phenotypic values (PV):

$$PVE_k = \frac{Var(BV_k)}{Var(PV)} \tag{5}$$

QTL analyses were subsequently performed on all the simulated 60 crosses using the pseudo-testcross approach where loci can be heterozygous in both parents (Grattapaglia and Sederoff 1994). The linkage analysis was performed separately for those loci that segregated in either parent using single marker analysis with the linear regression model:

$$PV_i = \beta_j + g_j x_{ij} + e_{ij} \tag{6}$$

where PV is the phenotypic value being dependent on the fixed intercept β and genetic association effect g plus the random residual e for individual i and marker j and dependent on the numerical marker genotype $x = 0$, 1, or 2. To minimize the discovery of false positives, in a simple but comparable way to the methods used in the meta-analysis, a significance level of 0.05 was applied including a Bonferroni correction based on the number of segregating loci. The loci showing the highest levels of significance (smallest p values) within QTL peaks were assumed to constitute the causal loci whether true or not. Association analyses were performed utilizing the "SNPassoc package" (González et al. 2007) with the single marker linear factor model:

$$PV_i = \beta_j + g_{jk} + e_{ij} \tag{7}$$

where PV, β, and e are the same as described for Eq. 6. However, for association analyses, there were separate fixed genetic effects (g) for each genotype class $k = 0$, 1, or 2 of marker j. Again, a Bonferroni-adjusted significance level of 0.05 was used for detecting QTL. For analysis, we only used loci with a minor allele frequency of at least 0.05. Notably, the approach to perform separate QTL mapping for markers segregating in either parent and the factorial nature of Eq. 7 implies that dominance, as well as additive effects, are included in g. However, as only additive effects were simulated, this should have a negligible impact and it follows trivially that PVE calculations based on estimates from Eqs. 6 and 7 reflect additive genetic variation only. In similarity to the meta-analyses, we estimated the amount of variance explained by all detected QTL by summation of the QTL PVE estimates (sum PVE). We compared sum PVE with the true cumulative

amount of variation (tot. true PVE) attributed to underlying causal loci, colocating with detected QTL for association analyses, or segregating within the detected QTL peaks in QTL mapping analyses (Eqs. 4 and 5). For further comparisons, a total PVE was also estimated based on linear multilocus models. Association analyses were chiefly performed on populations of 500 individuals randomly sampled from the 2000 individuals available from the base population. However, in order to examine the effect of sample size on mapping efficiency and on accuracy of n_{qtl} estimation, we also generated random samples ranging from 100 to 1500 individuals (i.e., 100, 250, 500, 750, 1000, and 1500). We repeated each sampling procedure until we had 100 samples with significantly detected QTL in each sample size class. The detection of false positives was checked, and irrespective of the sampled population size, the FDR was found to be reasonably low (~4%).

To assess the usability and validity of the n_{qtl} estimation method, it was also applied on the simulated datasets in the same fashion as previously described (Eq. 2). However, in order to account for the sampling variation of QTL segregation and allele frequencies, the local narrow-sense heritability within each simulated cross or population was calculated and used in Eq. 2 according to the following:

$$h_l^2 = \frac{Var(BV_{true})}{Var(PV)} \tag{8}$$

where BV_{true} is the known true breeding value of an individual calculated by Metagene (i.e., a further extension of Eq. 4) and PV is the phenotypic value. To evaluate the robustness of the method with respect to varying numbers of true causal loci, three extra simulated datasets were created using the same parameters as previously described but setting the number of causal loci at 25, 50, and 100, respectively. This evaluation was also done to assess the impact of varying degrees of LD between causal loci on n_{qtl} estimation because a higher number of causal loci in a genome of a constant genetic length (same recombination rate) would confer increased LD between the loci. Additionally, we conducted n_{qtl} estimation using QTL detected at a more relaxed significant threshold ($p < 0.2$ after Bonferroni correction) to assess the possible impact of the variable significance thresholds encountered in the meta-analysis.

Results

Meta-analysis

For association studies used in our meta-analyses, the population size ranged from 116 (Ma et al. 2010) to 700 (Eckert et al. 2009) with an average of 446, while the sample size range for QTL studies was 75 (Ukrainetz et al. 2008) to 740 (Pelgas

et al. 2011) with an average of 292. There were 152 and 239 separately defined traits for association studies and QTL studies, respectively, featuring 1520 and 664 detected QTL. Among these studies, 88 and 98 traits, from association and QTL studies, respectively, had three or more detected QTL and could thus be used to estimate the number of underlying effects. The individual effect PVE estimates from all association studies were generally smaller (median = 3.1%) than those of all the QTL studies (median = 6.8%). This resulted in an effect size distribution where association study effects with a PVE of 10% or more were very rare while effects of corresponding size class in QTL studies were substantially more numerous (Fig. 1). In contrast to association mapping studies, the estimated PVE of linkage mapping QTL were very variable and a few QTL identified in linkage mapping studies even explained substantial parts (>20%) of the phenotypic variation (Fig. 1a). The differences in individual PVE estimates were also reflected in the estimates of summed PVE (and multilocus-based total PVE estimates when available) where, in general, larger amounts of variation were explained for QTL studies than for association studies (Fig. 2). For most trait types, individual effect PVE estimates were small (<5%, Table 1) but water use efficiency traits exhibited somewhat larger median PVE both in QTL studies and in association studies (8.4 and 5.0%, respectively). For QTL studies, disease resistance traits showed notably large (median at 62.5%) estimates of total/summed PVE followed by phenology traits at a median of 39.9% (Fig. 2). The large maximum (56.4%) and average (15.2%) PVE for QTL studies of disease resistance (Table 1), moreover, indicated the presence of major QTL for that trait. For association studies, only water use efficiency and cold resistance traits explained substantial parts of the variation (tot/sum PVE in the range 30–40%) and association studies explained more of the total variance than QTL studies

only for cold resistance. This largely stems from the number of detected effects for association studies (ranging from 15 to 19) by far exceeding the number of QTL detected in QTL studies (at most four). The median of individual PVE estimates for cold resistance for association studies was about half that of the corresponding QTL study based estimates (1.8 and 3.8%, respectively) which was similar to other traits. The standardized distributions of PVE effects generally exhibited means that were greater than the corresponding modes (Fig. 3), thus indicating a right-sided skew.

The total variation explained for a trait when summing up individual effects without considering interaction or linkage (sum PVE) varied from 1.2 to 392% in association studies and 1.8 to 97.9% in QTL studies (Fig. 4). The summed amount of variance explained at the same number of detected effects was comparatively greater for the QTL mapping approach than for association mapping. For example, with respect to the sum PVE, five significant QTL effects in QTL studies explained, on average, about 37% of phenotypic variation while five significant effects in association studies explained only 19%.

The median of estimated numbers of underlying QTL effects (n_{qtl}) was 643 and 33 for association and QTL studies respectively while the mode estimate was 399 and 19, respectively (Fig. 5a). When considering the studied trait types separately (Fig. 5b, Table 2), association study n_{qtl} medians ranged from 72 (disease resistance) to 1275 (secondary metabolites) whereas QTL study medians ranged from 4 (disease resistance) to 133 (wood chemistry traits). The disease resistance traits showed the lowest n_{qtl} medians for both QTL and association mapping studies. For traits of traditional interest in genetic tree improvement (growth and physical wood traits), we estimated n_{qtl} at 578 and 801, respectively, for association studies while the corresponding estimates for QTL studies were much lower (17 and 35). In turn, the number of
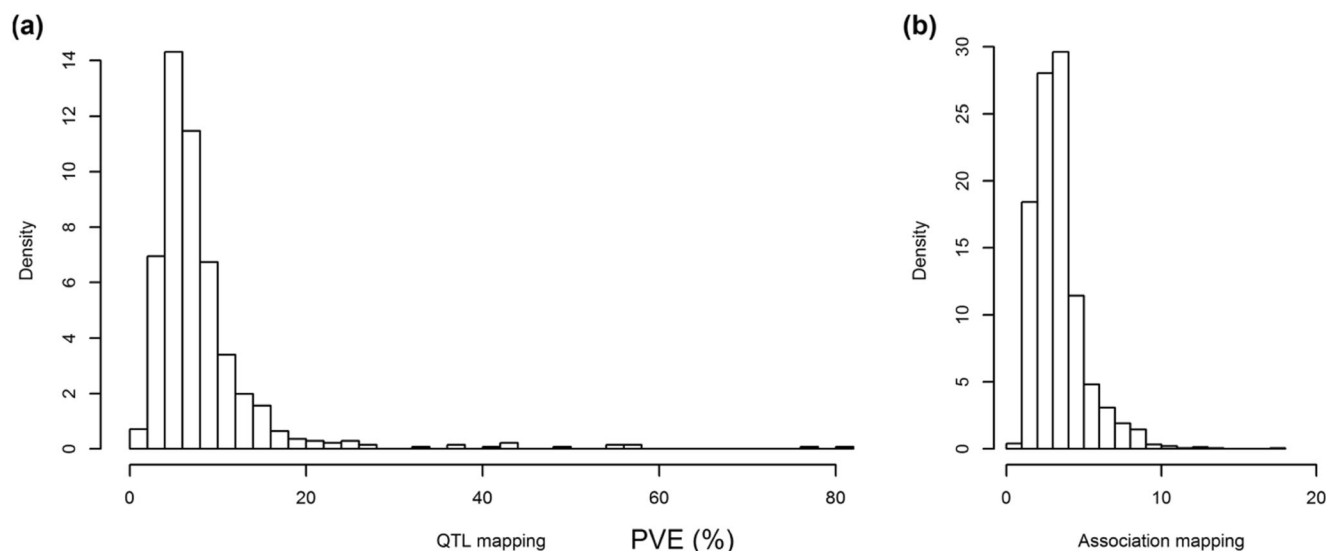
Fig. 1 Histogram of the percentage variance (PVE) explained by individual effects in the examined studies: **a** QTL analyses; **b** association analyses
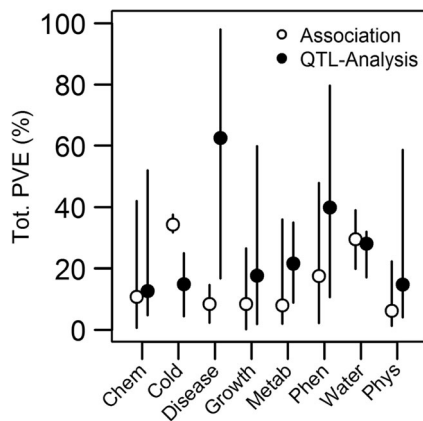
Fig. 2 The distribution of the total phenotypic variance explained (PVE) by detected QTL in published association and QTL studies for each trait type. Multilocus-based total PVE estimates were used when available while summed PVE were included otherwise. *Symbols* show the median of distribution of total/summed PVE reported in the studies (Supplementary Table 1) while *whiskers* represent the total PVE min–max range

individuals required to explain 50% or more of the genetic variation for most traits was found to be substantial, in particular for association mapping (Fig. 5c). Given the n_{qtl} estimates, there would be a need to submit at least 20,000 individual trees to association mapping analysis in order to capture 50% of genetic variation for physical wood quality and growth traits, while less than 2000 trees would be required in a traditional QTL mapping study. For other traits, 2000 trees (disease resistance) to 30,000 (metabolites) were found to be required for association studies while only 200 (disease resistance) to about 4500 trees (wood chemistry traits) were required to detect QTL explaining at least 50% of the genetic variation of QTL mapping populations (Table 2).

Table 1 The size of individual allelic effect (PVE) from the studies used in the meta-analysis

Trait type	Minimum (%)		Median (%)		Average (%)		Maximum (%)	
	A	Q	A	Q	A	Q	A	Q
Chem	1.11	3.50	4.00	6.00	4.09	7.15	9.92	37.00
Cold	0.70	2.80	1.84	3.80	2.03	5.56	5.40	9.80
Disease	0.29	2.20	1.12	5.50	1.46	15.23	3.83	56.40
Growth	1.18	2.30	3.63	4.50	3.87	5.69	13.24	42.20
Metab	1.80	3.40	3.30	5.70	4.93	7.01	18.00	16.90
Phen	1.00	1.00	3.16	6.80	3.35	8.04	12.70	27.60
Water	0.45	4.00	5.00	8.40	4.04	7.99	8.00	15.90
Phys	2.00	2.80	4.10	7.70	4.15	9.18	6.50	25.40
Average[a]	3.27	4.45	3.97	8.37	3.79	7.17	5.41	15.14

Only traits that produce three or more detected QTL are considered
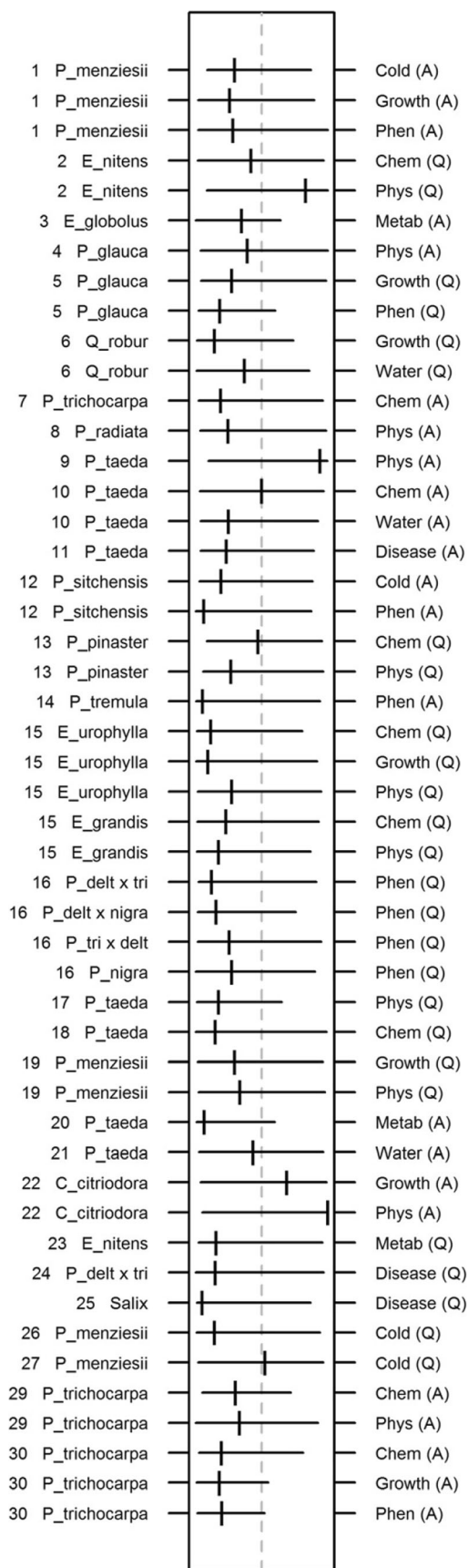
A association study, *Q* QTL studies

[a] The minimum, median, average, and maximum individual PVE estimates averaged over all analyses

Our meta-analyses clearly indicate that both association and QTL mapping populations harbor several to many undetected QTL controlling the traits, where the high PVE for QTL detected in linkage mapping populations result in much lower estimates of total number of underlying effects.

Simulations

QTL mapping analyses

Out of the 49 designed causal loci, only 25.5 on average (standard deviation = 3.35) segregated in the 60 simulated QTL crosses. The local heritability of the simulated trait within each cross varied from 0.16 to 0.59 with an average of 0.40 in comparison to the designated global h^2 of 0.5. Numbers of effects detected from crosses in the simulated population ranged from 1 to 8 (Fig. 6a) with an average of 3.7. Although these QTL were few relative to the loci actually segregating, their PVE estimates were nonetheless large. For example, the largest QTL of each analysis exhibited an estimated PVE of 21.9% on average which is to be compared with the largest causal locus segregating in the cross which, on average, had a true PVE of only 10.1%. Moreover, the estimated average sum PVE of these QTL was as high as 43.5%. Loci situated within the identified QTL peaks in truth accounted for 39.6% of the phenotypic variation, indicating that a considerable portion of influential loci were located within the peak. However, a multilocus-based total PVE estimate suggested that a model including all detected QTL in fact accounted for only 34% of the phenotypic variation. The naively summed PVE estimates were thus only, on average, 3.9% larger than the total assigned PVE and correlated fairly well with this metric (Fig. 6b) but still appeared overestimated in comparison to the proper statistical explanatory power of all QTL jointly. In any case, it is notable that all these average measures of total PVE were close to the average local heritability, suggesting that the detected QTL explained almost all the segregating genetic variation. The relationship between number of effects and summed PVE of the simulated data closely mirrors that of the metadata from the experimental results (Figs. 4 and 6a). Taken together, these observations indicate that the effects of several causal loci were compounded within a single detected QTL peak. It was directly observed that several causal effects were situated within the portion of individual QTL peaks exceeding the significance threshold, thus contributing to the total PVE (Fig. 6a, b). Consequently, if a QTL thus detected would be regarded as a single causal locus per se, its PVE would be greatly overestimated in comparison to the causal locus closest to the peak. However, the inflation is not large if the true total PVE of all adjacent causal loci segregating in the cross, detected or undetected, are accounted for (Lee and Van der Werf 2006) (Fig. 6b).

1	P_menziesii	Cold (A)
1	P_menziesii	Growth (A)
1	P_menziesii	Phen (A)
2	E_nitens	Chem (Q)
2	E_nitens	Phys (Q)
3	E_globolus	Metab (A)
4	P_glauca	Phys (A)
5	P_glauca	Growth (Q)
5	P_glauca	Phen (Q)
6	Q_robur	Growth (Q)
6	Q_robur	Water (Q)
7	P_trichocarpa	Chem (A)
8	P_radiata	Phys (A)
9	P_taeda	Phys (A)
10	P_taeda	Chem (A)
10	P_taeda	Water (A)
11	P_taeda	Disease (A)
12	P_sitchensis	Cold (A)
12	P_sitchensis	Phen (A)
13	P_pinaster	Chem (Q)
13	P_pinaster	Phys (Q)
14	P_tremula	Phen (A)
15	E_urophylla	Chem (Q)
15	E_urophylla	Growth (Q)
15	E_urophylla	Phys (Q)
15	E_grandis	Chem (Q)
15	E_grandis	Phys (Q)
16	P_delt x tri	Phen (Q)
16	P_delt x nigra	Phen (Q)
16	P_tri x delt	Phen (Q)
16	P_nigra	Phen (Q)
17	P_taeda	Phys (Q)
18	P_taeda	Chem (Q)
19	P_menziesii	Growth (Q)
19	P_menziesii	Phys (Q)
20	P_taeda	Metab (A)
21	P_taeda	Water (A)
22	C_citriodora	Growth (A)
22	C_citriodora	Phys (A)
23	E_nitens	Metab (Q)
24	P_delt x tri	Disease (Q)
25	Salix	Disease (Q)
26	P_menziesii	Cold (Q)
27	P_menziesii	Cold (Q)
29	P_trichocarpa	Chem (A)
29	P_trichocarpa	Phys (A)
30	P_trichocarpa	Chem (A)
30	P_trichocarpa	Growth (A)
30	P_trichocarpa	Phen (A)

95% Distribution of allele effects

◀ **Fig. 3** Normalized distribution of effect sizes for each trait type in the examined studies. The *vertical ticks* indicate the median of PVE of the detected effects while *whiskers* show the 95% percentiles

Furthermore, the estimated median number of underlying QTL (n_{qtl}) was only 9.8 in comparison to the 25.5 loci that segregated (Fig. 6c) implying underestimation. The systematic n_{qtl} underestimation of QTL mapping populations was further corroborated by additional simulations using genomes designed with variable numbers of causal loci (Supplementary Fig. 1). Notably, n_{qtl} estimates for genomes comprising 100 causal loci were only somewhat higher (median at 19.8) than for genomes comprising 25 or 50 causal loci (medians at 13.3 and 13.4, respectively). For the genomes comprising 50 and 100 loci, n_{qtl} estimates were substantially and consistently underestimated in comparison to the number of segregating loci. A probable explanation for this result is that the extensive physical linkage between adjacent causal loci within single crosses created the large PVE composite QTL observed. In turn, this resulted in major gene like PVE distributions and thereby giving impression that fewer underlying QTL were present than segregating in reality (see also Eq. 2).

Association mapping analyses

For the simulated association mapping populations with the sample size of 500 individuals, almost all (on average 46.8 out of 49) causal loci segregated. The local heritability estimates, ranging from 0.43 and 0.56 with an average equal to the assigned global heritability of 0.5 were, as expected, more stable than those of the QTL crosses. In similarity to the QTL cross analyses, between two and eight (average at 4.5) QTL were detected. The detected loci showed a sum PVE of 28.3%, and the underlying loci accounted for, on average, 23.2% of the phenotypic variance in reality (Figs. 6a and 7b). These estimates were somewhat lower than the corresponding estimates for QTL mapping crosses and were considerably lower compared to local heritabilities thus indicating an issue of missing heritability. Comparing the summed estimated PVE and the total true PVE, a relationship similar to that of the QTL cross analyses was observed (Figs. 6b and 7b). However, the summed PVE increased with the number of detected effects, at a much lesser rate than for the simulated QTL mapping populations again reflecting the observed metadata results for QTL and association studies (Figs. 4, 6a, and 7a). For example, at five detected effects, the sum PVE estimated for the simulated QTL and association mapping populations were 53.6% (Fig. 6a) and 31.4% (Fig. 7a), respectively. At a sample size of 500, overestimation of summed PVE of QTL in comparison to the true total PVE of the underlying causal loci increased from 0.7 to 9.3% with increasing number of detected effects (Fig. 7a). The estimated number of QTL
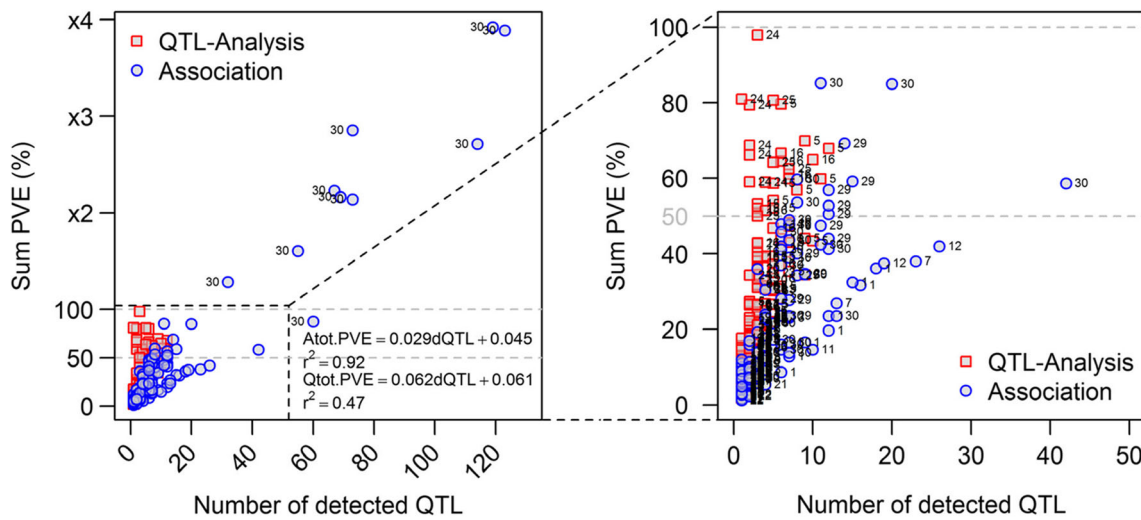
Fig. 4 Scatterplot showing the relationship between the number of detected effects and the sum of phenotypic variance explained estimates derived from single marker analyses (PVE) of experimental data. *Blue circles* are data from association studies and *red squares* from QTL studies. The *numbers at each symbol* signify the study from which the data was obtained (Supplementary Table 1). The *subplot on the right* comprise the same data zoomed in to increase resolution

appeared to be somewhat overestimated for such a population (83, Fig. 7c). The n_{qtl} estimates for this population also appeared to show a rising trend dependent on the number of detected QTL, but this observation was not clearly supported given results from the extra simulation populations with variable numbers of causal loci (Supplementary Fig. 1).

In order to assess the impact of population sample size on association mapping performance and on n_{qtl} estimation, analyses were also performed on populations with sample sizes ranging from 100 to 1500 (Fig. 8). As an example, an association population size of 1500 enabled the detection of a higher number of QTL (on average 11.8, maximum at 14) in comparison to the population size of 500 and the underlying loci accounted for a higher portion of the phenotypic variance (39.6%) on average. There was a rapid decrease in the overestimation of individual effect PVE with the increase of association population size (Fig. 8c) and, given the simulation conditions (49 causal loci and $h^2 = 0.50$), approximately 1000 individuals were shown to be needed in an association population to avoid
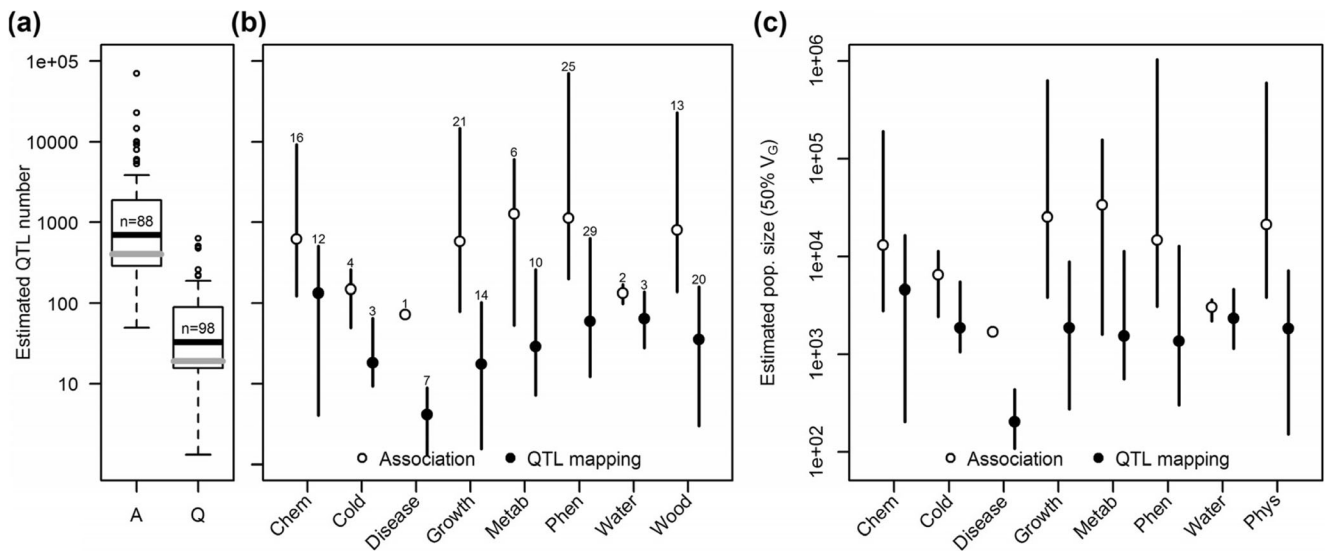
Fig. 5 Estimates of the number of underlying effects (n_{qtl}) for all data shown as a boxplot (**a**) and subdivided between association (*A*) and QTL mapping studies (*Q*), but also shown for data subdivided for each trait type (**b**). Furthermore, the mapping population size required to capture 50% of the genetic variation for each trait type is shown (**c**). For subplot **a**, the *black horizontal indicators* signify the median while the *gray indicators* signify the mode. For subplots **b** and **c**, *points* describe the median of the distribution of estimates while *whiskers* show the 95% percentile of the variation of estimates. The number of analyses underlying the respective n_{qtl} distribution ranges are given as *numbers above the whiskers*

Table 2 Estimated number of underlying QTL and mapping population size needed to capture 50% of the genetic variation when using traditional mapping methods

Study type	Trait type	Number analyses	Detected effects	Estimated effective number of QTL				Required population size			
				Min	Mode	Median	Max	Min	Median	Max	h^{2a}
Association studies	Chem	20	153	123.2	473.2	615.3	9188.4	2783	13,047	190,466	0.5
	Cold	4	68	49.2	144.9	147.9	256.8	2436	6462	11,237	0.25
	Disease	2	12	72.1	72.1	72.1	72.1	1685	1697	1685	0.5
	Growth	31	185	78.4	462.7	578.1	14,507	3844	25,382	628,384	0.24
	Metab	39	65	52.7	414.9	1274.9	6032	1607	33,455	156,547	0.4
	Phen	27	929	197.9	485.3	982.9	69,628	3085	14,707	1,032,734	0.7
	Water	2	11	98.7	133	133	167.2	2196	3043	3641	0.5
	Phys	27	97	138.3	657	801.3	22,859.4	3858	21,191	593,852	0.4
QTL studies	Chem	43	90	4.1	136.4	133.1	505.8	204	4575	16,485	0.33
	Cold	9	21	9.3	13.8	18	64.9	1059	1875	5485	0.14
	Disease	15	45	1.3	4.2	4.1	8.9	108	204	438	0.33
	Growth	31	99	1.5	17.3	17.4	102.2	274	1873	8799	0.14
	Metab	14	45	7.2	24	28.7	257.3	560	1539	11,302	0.25
	Phen	38	185	12.2	38.5	59.3	630.6	302	1366	12,672	0.54
	Water	4	13	27.7	46.8	64.3	136.6	1145	2330	4622	0.33
	Phys	85	166	3	15.9	35.4	158.5	151	1831	7127	0.25

[a] Heritabilities used for the estimation of total QTL numbers are generic assumptions based on quantitative genetic studies on forest trees

the inflation. This overestimation reduction with increasing sample size was also apparent for sum PVE estimates as the previously mentioned positive relationship between sum PVE inflation and number of detected effects was obscured and even reversed by a negative relationship when analyses of several sample sizes were considered jointly (Fig. 8a). Also, the n_{qtl} estimates became less inflated and acquired better stability with increased population

size (Fig. 8d) ranging from substantial overestimation (median at 125) and considerable instability at the population size of 250 to a reasonably accurate estimate (median at 49) and better stability at the population size of 1500. When including the complete base population (2000 individuals) in the analysis, 14 effects were detected whose underlying loci in reality accounted for 41.4% of the phenotypic variance indicating a relatively limited missing
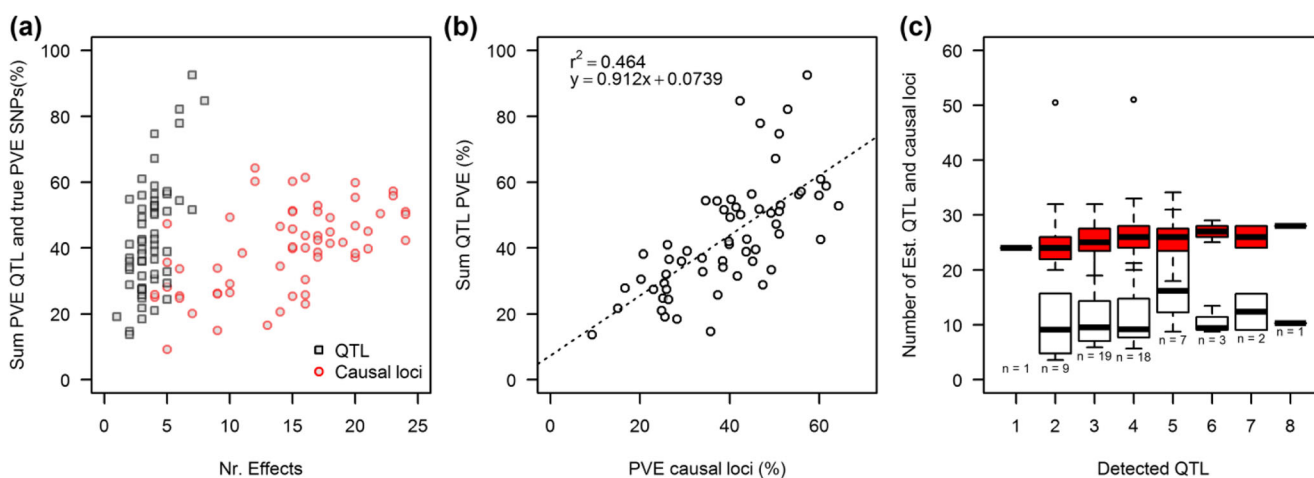
Fig. 6 Results from QTL analyses of 60 simulated crosses with 500 offspring in each. **a** Number of effects detected for QTL studies with the summed percentage of phenotypic variance explained (PVE) and the corresponding total true PVE of causal loci cosegregating within the significant QTL. **b** A linear relationship between the sum of individual PVE for QTLs and the true total PVE of causal loci segregating within the QTL peaks. **c** The estimated number of underlying QTL (*white boxes*) and the segregating number of causal loci (*red boxes*) in relation to the number of QTL actually detected
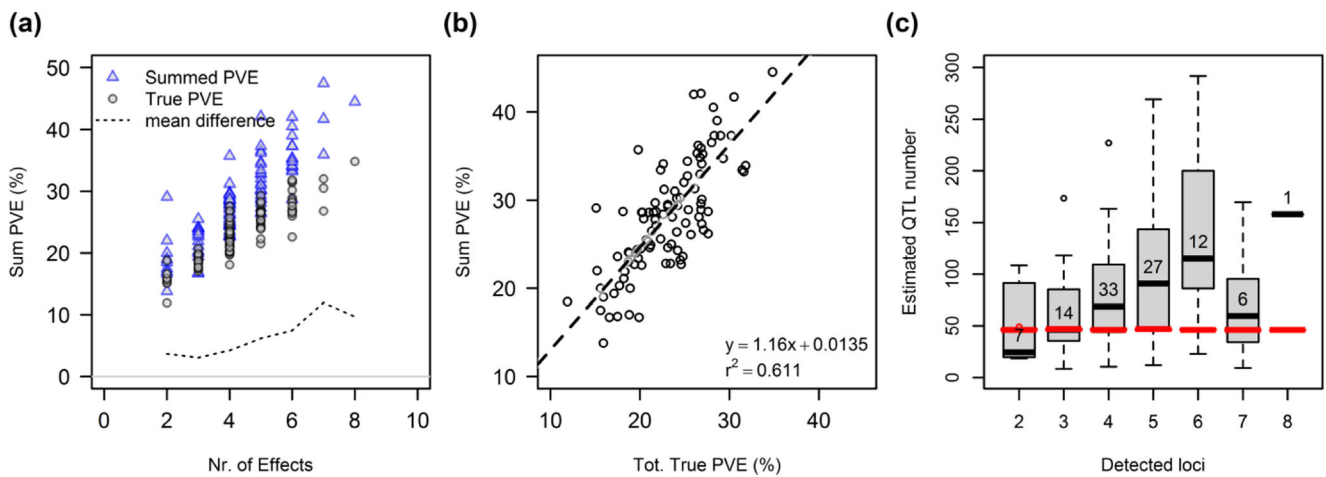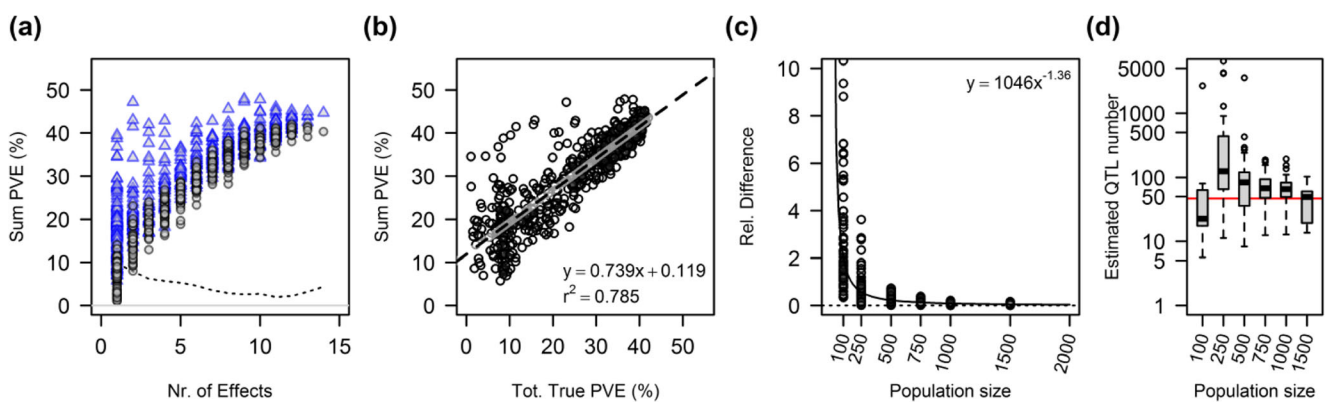
Fig. 7 Results from association mapping analysis of 100 mapping populations with 500 individuals each from simulated data. **a** Summed estimates of percentage of phenotypic variation explained by the detected effects (sum PVE) in relation to the number of detected effects and the total true PVE of the causal loci segregating together with the detected effects with the dashed line showing the mean difference between the two. **b** The summed estimated PVE in relation to the total true PVE of the correspondingly detected effects. **c** The estimated number of underlying effects (n_{qtl}) at different number of detected effects

heritability problem. The estimated number of underlying QTL was 47.8. Corresponding improvements in n_{qtl} estimation with increasing sample size were also observed when simulations of genomes featuring lower (25) or higher (100) numbers of causal loci were considered (Supplementary Fig. 1). In the case of association mapping populations, n_{qtl} estimates also agreed with the number of causal loci in a manner more consistent than for the QTL mapping populations. Additionally, the reliability of the n_{qtl} estimation method was not greatly affected by a relaxation of the significance thresholds used for detection (Bonferroni corrected $p < 0.2$, Supplementary Fig. 2) even

though such a relaxation was associated with a substantially increased FDR (rising from ~4 to at most 15%).

In summary, the results from both types of mapping procedures on simulated data suggest that the method used to estimate the number of underlying effects performs well when the PVE of the effects are estimated using a large-sized association mapping population (Fig. 8d and Supplementary Fig. 1). However, in contrast to association mapping, n_{qtl} estimation based on QTL mapping analyses will consistently underestimate the true number of underlying QTLs, both due to a lower number of loci actually segregating within single crosses and due

Fig. 8 Association analysis results from simulated datasets with a sample size range 100–1500 drawn from a base population. **a** Summed estimates of percentage of phenotypic variation explained by the detected effects (sum PVE) in relation to the number of detected effects (*triangles*) and the total true PVE of the causal loci cosegregating with the detected effects (*circles*), with the dashed line showing the mean difference between the two. **b** The summed estimated PVE in relation to the total true PVE of the correspondingly detected effects. **c** The impact of different population sizes on the PVE inflation of individual detected effects (relative difference expressed as $PVE_{est}/PVE_{true} - 1$). **d** The estimated number of underlying effects (n_{qtl}) at different population sizes. The *red line* describes the true number of causal loci segregating in the base population. Boxplots for estimates of population size 100 and 250 are based on 11 and 60 estimates, respectively, while the boxplots of the remaining population sizes are based on 100 estimates each

to extensive physical linkage resulting in lower mapping resolution and making it difficult to separating adjacent causal loci in analysis (see Fig. 6c and Supplementary Fig. 1).

Discussion

Several QTL mapping analyses have been performed to identify the underlying genetic architecture of traits with economic importance in trees (some of them detailed in Supplementary Table 1). More recently, association mapping methods have been utilized to identify loci that explain the observed phenotypic variation at a higher resolution, mainly through candidate gene approaches. We observed large variation in the results both between and within study types with respect to the number of QTL detected and to the degree which these accounted for the phenotypic variance. Nonetheless, QTL PVE distributions for most individual studies and traits were similarly skewed to the right, indicating that the experimental setup exerted little influence over the allele effect size distribution and the overall L-shaped PVE distribution appeared to agree with our assumption of this parameter adhering to a squared exponential distribution (see also Ingvarsson and Street (2011)). In this respect, the study of simulated data yielded valuable results because we could then ensure that the same base characteristics (genomic setup, crossover rate, global heritability) were used to create both association populations and the QTL crosses. For the simulated data, we therefore expect very little confounding between differences in population parameters and study type, especially in comparison to the studies of the meta-analysis. Furthermore, the usability of the modified n_{qtl} estimation method of Otto and Jones (2000) was validated for simulated association mapping data, at least provided that a large enough population size was used.

QTL vs. association mapping

Individual QTL detected in QTL mapping studies explained on average more of the phenotypic variation than individual QTL detected in association studies, which is consistent with studies comparing effects sizes between classical linkage mapping QTL and influential SNPs identified in association mapping (Kump et al. 2011; Tian et al. 2011). We also found that more of the total variance could, in most cases, be explained in QTL studies than association mapping studies (Fig. 4), and this trend was again observed for the simulated data (Figs. 6a and 7a). The occasionally very high sum of PVE depicted for a few association mapping analyses in Fig. 4a (even biologically unreasonable) should in this context be interpreted cautiously, as they stem from the naive PVE summation of a great number of detected loci that are often in LD

with each other and possibly also inflated due to the Beavis effect (Beavis 1994; Xu 2003a). This is also the main reason behind the discrepancies between Figs. 2 and 4 since the former figure depicts multilocus estimated total PVE when such estimates were available while the latter consistently uses arithmetic sums of PVE estimates (see also Supplementary Table 1 and McKown et al. 2014). Apart from the systematic differences in estimated PVE, estimates of the total number of effects (n_{qtl}) were for most traits considerably lower for QTL studies than for association mapping studies with medians and modes that often differed in orders of magnitude (Fig. 5).

The larger individual and summed PVE estimates and the lower n_{qtl} estimates of QTL studies appear to be a result of several different factors or a combination of them. (i) A major factor is that fewer loci will segregate within the QTL mapping full-sib families. These are generated by a mere two parent sample in comparison to the genetically diverse association mapping populations. This tendency was indeed supported by the simulated data in terms of decreased numbers of segregating loci and by decreased lower local heritabilities. (ii) However, it should also be noted that those few causal loci actually segregating in the QTL mapping crosses would themselves explain larger amounts of the phenotypic variation than in association populations because, within the QTL full-sib family, they would be guaranteed to segregate at intermediate allele frequencies (see Eq. 4.12a in Lynch and Walsh 1998). (iii) Another important factor to consider is that the significant peaks encountered in QTL mapping studies are often regarded as being generated by a causal single locus but may in reality comprise the cosegregation of several linked causal loci. Such cosegregating clusters of effects would consequently increase the estimated effect and would be biased in the case a single locus would be regarded as the sole source of the QTL peak. This phenomenon would then reduce the number of detected QTL and increase the PVE of their effects as we indeed observed in the simulation study (see Fig. 6a, b (Lee and Van der Werf 2006)). Such clusters of cosegregating causal loci could even have contributed to the few QTL with very large effects described for QTL mapping studies (Fig. 1a). An interesting and comparable real data example of these phenomena is the study of Du et al. (2015) where significant markers from an association study on a population of *Populus tomentosa* with 460 individuals were validated in 1200 F1 individuals from a hybrid cross. Those markers that could be validated for the same trait in the F1 population showed on average 58% larger PVE in the F1 cross in comparison to the association mapping population.

Several causal loci segregating within individual QTL peaks is, moreover, the most likely explanation for the systematic n_{qtl} underestimation observed for the simulated QTL mapping populations even taking into account that only a subset of the designed causal loci segregated (see Fig. 6c and Supplementary Fig. 1) and despite the fact that the QTL

actually detected accounted for almost the whole local heritability (0.844 multilocus-based estimate). In contrast, n_{qtl} estimation based on large sample size simulated association mapping data, was reasonably unbiased (Fig. 8d and Supplementary Fig. 1), albeit showing substantial variance. At lower sample sizes, the numbers of underlying QTL were even overestimated rather than the opposite.

Although the observations made here somehow precludes the usability of our n_{qtl} estimation method for QTL cross type data with respect to the true number of causal loci, it may still be informative with respect to how many *effective* QTL peaks, including cosegregating clusters of causal loci, would be available for detection given good enough marker coverage and large enough sample size.

Differences in QTL number estimate among traits

A closer analysis of the traits and studies included in this paper revealed additional differences, not only between study types, but also between trait categories. The most conspicuous feature of the separate trait-type meta-analyses was the high total/summed PVE estimates and low n_{qtl} estimates obtained for disease resistance traits (Table 1, Fig. 5), in particular for the QTL studies. This indicates that disease resistance traits, on average, are influenced by fewer QTL potentially exhibiting larger effects than other traits. However, one should also be aware that QTL mapping studies for disease resistance traits frequently entail interspecies hybrid crosses. This design facilitate the segregation of polymorphic loci that are not necessarily representative of the single species in isolation and whose influence over disease resistance might have been known or suspected a priori (e.g., Hanley et al. 2011). Such interspecies segregating QTL could exert major influence over the trait within a QTL cross population, thus offering another possible explanation for the long tail of a major effect QTL depicted in Fig. 1a. For instance, the disease QTL studies included here focused on biotrophic fungal resistance which appears to have a few large effect resistance loci (Poland et al. 2009). The genetic architecture of biotrophic fungal resistance traits often consists of gene for gene resistance and are likely influenced by dominance and epistasis which, in turn, could inflate effect size estimates if effects are assumed to be purely additive. The included association studies instead focused on resistance to a necrotrophic fungus (Quesada et al. 2010), which kills the host cells to extract nutrients from dead cells. This type of resistance has previously been indicated to have a complex genetic architecture with many underlying small QTL (Poland et al. 2009). The n_{qtl} estimate for association mapping of disease resistance was still low (72), albeit higher than corresponding estimates from QTL studies. Indeed, recent studies evaluating genomic selection supports the large difference in number of underlying QTL for different traits, where rust resistance reached highest predictive ability when

only a few hundred markers were used indicating moderate number of underlying effective loci. In contrast, more than 4500 markers were necessary for the prediction of height and sugar content (Resende et al. 2012).

In general, we expected that QTL detected for high heritability traits (>0.4), such as physical wood traits, water use efficiency, disease resistance, and phenology, would explain more of the phenotypic variation as they should be less influenced by environmental noise and the underlying QTLs should consequently be easier to detect. However, this notion appeared to be true only in the case of phenology (Fig. 2) because the analyses identified a multitude of QTL for this trait category (Table 2). For phenology, it is indeed conceivable that the numerous QTL detected owes to rich and well-characterized candidate gene resources available for this trait category (Andrés and Coupland 2012) rather than high heritability. Even though physical wood traits are considered high heritability traits (Baltunis et al. 2007), only a little more than two effects were found for both association and QTL studies with small average and summed/total PVE (Tables 1 and 2 and Fig. 2). Irrespective of these differences, the median estimates of underlying QTL were found to be close to a thousand for association mapping studies and in the 20–60 range for QTL studies of high heritability phenology and physical wood traits as well as for low heritability traits such as growth (Fig. 5b). This resulted in very high estimates of the number of individuals required for identifying QTL that would explain 50% or more of the genetic variation (Fig. 5c). Theoretically, some of the causes behind the differences observed between QTL and association analyses could contribute to the differences among traits, where highly adaptive traits tend to have their QTL clustered in the genome (Renaut et al. 2013) while growth traits in trees could have similar genetic architecture as height in humans, evenly distributed along the genome (Yang et al. 2011b). Growth traits are however closely correlated to phenology (Rehfeldt 1992) and to dissect its genetic architecture in better detail, further studies which better account for phenology may be needed. The clustering of QTLs among adaptive traits may result in larger effect size estimates than for growth traits due to the cosegregating variation proximal to the detected QTL.

From this study, we can still conclude that problems of missing heritability are to be expected in conifer association mapping analyses and even in the QTL mapping of more complex traits. Missing heritability issues will remain severe as long as study population sizes are kept in the range of hundreds, criteria for QTL significance are kept at a traditionally strict level, and no substantial improvements are made in statistical analysis methods. It should be noted that this likely holds true even if the whole genome would be sufficiently covered with markers for the task (Wurschum and Kraft 2015).

Issues of current and future studies

The n_{qtl} estimation method developed for this study appears to be able to produce reasonable estimates from association mapping data analyses. However, with respect to the meta-analyses of published association and QTL mapping studies, the general lack of heritability estimates resulted in the necessity to assume broad generic heritabilities (Table 2), thus introducing a potential source of noise in n_{qtl} estimation. One should further acknowledge that departures from the underlying assumption, that allelic effects are negative exponentially distributed, could of course affect the reliability of the estimates. There was also considerable variation among the different studies with respect to the nature of the trait, mapping methodology, and available population size (Supplementary Table 1), making detailed comparisons difficult. In hindsight, many of the studies did not have enough individuals in order to detect QTL accounting for the majority of the genetic variation (Fig. 5c). Populations of such limited sample size, also increases the risk of overestimating the QTL effect sizes (Fig. 8c) (Beavis 1998; Xu 2003b). The analyses of simulated association mapping data indicated that a small population size would cause the number of underlying effects to be systematically overestimated due to an upward bias of the detection threshold PVE (θ) (Fig. 8d). Given this, one cannot exclude that the meta-analysis n_{qtl} estimates for the association mapping studies of complex traits, such as growth and physical wood quality, may be biased upward, possibly by as much as three times as indicated in Fig. 8d and Supplementary Fig. 1. Nonetheless, even if such a bias was accounted for, n_{qtl} estimates for most traits of this study would still be very large. In any case, one should treat the n_{qtl} estimates and the number of individuals required for explaining half of the genetic variation as a guide indicator producing the correct n_{qtl} order of magnitude rather than yielding any individual precise value.

Another notable limitation of the n_{qtl} estimation method already admitted by Otto and Jones (2000) concerns its inability to account for LD between loci. According to our simulation results, association mapping analyses did not suffer much from this limitation and the extent of LD for such populations has been observed to be very short, in particular for tree species (Pavy et al. 2012). On the other hand, n_{qtl} estimates were consistently underestimated for QTL mapping populations with respect to numbers of truly segregating QTL and seemed to describe rather the putative effective number of QTL peaks available for detection. A similar limitation encountered in the examined studies, especially association studies, stemmed from the tendency to perform analyses on a single locus-by-locus basis. Like the Beavis effect, this may also inflate individual effects of a significant marker, which may share some of their contribution to the trait with other loci that are in LD (Peres-Neto et al. 2006). Theory predicts that some of the phenotypic variance is explained by cosegregating genetic variation within populations (Latta 1998; Le Corre and Kremer 2003) resulting in an epistatic effect where two (or more) cosegregating variants explain more than their individual effects combined. If this type of genetic architecture is true and handled analytically, much smaller population sizes is required to catch a majority of the PVE that stem from genetic sources, however, estimating the number of effects would be less straightforward. Multimarker association methods (Wurschum and Kraft 2015; Yang et al. 2011a) or genomic prediction analysis could be utilized to better estimate individual effect size, in turn improving our understanding of the evolutionary forces that shaped their genetic architecture. The Bayesian methods suggested by Meuwissen et al. (2001) and other more recent extensions of the methods (Habier et al. 2011) estimate the variance of each marker separately for all markers at once, which could be fruitful for finding candidate loci and more robust estimates of individual effect sizes. These methods still require large training sets of several thousand individuals and are also more computer intensive. However, non-MCMC-based algorithms for BayesB have also been developed and these are much faster, although slightly less accurate (Meuwissen et al. 2009; Shepherd et al. 2010). Applying the latter method to our simulated 49-loci dataset of 2000 individuals resulted in the detection of 20 QTL (Supplementary Fig. 3) at a confidence probability of 95%, to be compared with the 14 QTL detected at a similar confidence level (FDR ≈ 0.04) using traditional single-locus models. The use of improved statistical analysis in conjunction with the multitude of methods available to detect loci under selection (e.g., outlier methods which quantify the differentiation of loci among populations or landscape genetics which link the genetics to the landscape and environment (Beaumont and Balding 2004; Bonhomme et al. 2010; De Mita et al. 2013; Sork et al. 2013) can help unravel the trait architecture further (Evans et al. 2014). Finally, it should be emphasized that a few studies included in this paper did try to reduce or correct for the issues originating from single locus analysis, such as the Beavis effect (Ma et al. 2010) or bias due to LD (McKown et al. 2014; Porth et al. 2013) between significant markers (Peres-Neto et al. 2006). Others used the recommended approach to validate and reestimate the effects in an independent population (Dillon et al. 2012; Dillon et al. 2010; Du et al. 2015; Tian et al. 2014) which also is a considerable improvement in this respect.

In general, it can be said that small effect sizes, limited statistical power and strict significance thresholds constitutes a substantial impediment for the dissection of the genetic architecture of complex traits. This problem is especially serious in the validation of QTL due to the high likelihood that effects considered to be significant in one population are not counted as significant in another even if they were true in both (Dillon et al. 2012; Freeman et al. 2013). For some traits (e.g.,

growth), where most effect sizes appear to be small and heritability is low, a huge effort would be required in order to capture most of the heritability with very large mapping populations replicated in multiple sites (Fig. 5c). Such traits will likely suffer from inescapable missing heritability issues where a large part of the segregating causal variation is undetected due to a combination of low effect size and limited statistical power (Fig. 8a). However, one approach to improve statistical power of low heritability would be the clonal replication of genotypes (possible for several tree species) in order to decrease the environmental noise, increase the genotypic accuracy of individuals, and thus increase the possibility to detect causal segregating alleles. Moreover, with the increased number of available markers, analyses will probably include more causal variants and this will also facilitate a better correction for relatedness and genetic background and better estimates of genetic merit through realized relationships (Hayes et al. 2009). This in turn will improve the detectability of causal markers in a mapping population (Vilhjálmsson and Nordborg 2013) .

Conclusions

1. The method of estimating the number of underlying effects proposed in this study was shown to be approximately accurate when applied to association mapping analyses of high statistical power. The n_{qtl} estimated from QTL mapping populations rather represents underlying and segregating composite QTL clusters, likely due to extensive linkage impairing mapping resolution.

2. The estimated QTL effect sizes in published association studies were generally smaller than the effects from QTL studies. This resulted in distributions implying that allelic effect sizes of 30% (PVE) or more are very rare for QTL studies while effect sizes of 10% or greater are very rare in association studies. These trends were in agreement with the results from simulated data analyses, suggesting that the differences between QTL and association mapping studies largely stems from the segregation of clusters of causal loci at intermediate allele frequencies in QTL mapping populations and the independent segregation of causal loci at lower allele frequencies in association mapping populations. The distributions of estimates of percentage of variance explained were sharply skewed toward the right and appears, in most instances, to be adequately describable by the square of a negative exponential effects.

3. The estimated numbers of underlying QTL were, on average, about 19 times higher for association studies than QTL studies for the same traits. The fewer segregating causal loci in the QTL mapping populations (i.e., limited genetic diversity) in combination with the lack of historical recombination events in these populations ultimately results in fewer

detectable effects. Median numbers of QTL estimated from QTL studies ranged from about 4 for disease resistance to 130 for wood chemistry while the corresponding estimates for association mapping studies ranged from about 70 for disease resistance to 1300 for metabolite traits. For traditional tree breeding traits (growth, physical wood traits, phenology), median n_{qtl} estimates for association mapping studies ranged from around 500 to 1000.

4. Given the assumed parameters and results of this study, the number of individuals needed to explain 50% of the genetic variation for most traits was found to be substantial, particularly for association mapping. For the complex traditional tree breeding traits (growth, physical wood quality, phenology), association mapping analyses would require the sampling of at least 10,000 individuals in order to capture the majority of the genetic variation. Current population sizes (hundreds) appear to be adequate only for the QTL mapping of disease resistance traits, preferably segregating in interspecific hybrid crosses.

Data archiving statement   This manuscript does not use any raw data. There are no data uploaded to any database.

References

Acheré V et al (2004) A full saturated linkage map of *Picea abies* including AFLP, SSR, ESTP, 5S rDNA and morphological markers. Theor Appl Genet 108:1602–1613

Andrés F, Coupland G (2012) The genetic basis of flowering responses to seasonal cues. Nat Rev Genet 13:627–639

Baltunis BS, Wu HX, Powell MB (2007) Inheritance of density, microfibril angle, and modulus of elasticity in juvenile wood of *Pinus radiata* at two locations in Australia. Can J For Res 37:2164–2174

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol 13:969–980

Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: Wilkinson, DB (ed) Proceedings of the 49th Annual Corn and Sorghum Research Conference, Chicago, Illinois, USA., 1994. American seed trade association, Washington, DC, USA., pp 250–266

Beavis WD (1998) QTL analyses: power, precision, and accuracy. In: Paterson AH (ed) Molecular dissection of complex traits, 1 edn. CRC Press, Boca Raton, FL, pp. 145–162

Birol I et al (2013) Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. Bioinformatics 29:1492–1497

Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, Blott S, SanCristobal M (2010) Detecting selection in population trees: the Lewontin and Krakauer test extended. Genetics 186:241–262

Buckler ES et al (2009) The genetic architecture of maize flowering time. Science 325:714–718

Castle WE (1921) An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. Science 54:223

Cornelius J (1994) Heritabilities and additive genetic coefficients of variation in forest trees. Can J For Res 24:372–379

De Mita S, Thuillet A-C, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Mol Ecol 22:1383–1399

Dillon SK, Brawner JT, Meder R, Lee DJ, Southerton SG (2012) Association genetics in *Corymbia citriodora subsp. variegata* identifies single nucleotide polymorphisms affecting wood growth and cellulosic pulp yield. New Phytol 195:596–608

Dillon SK, Nolan M, Li W, Bell C, Wu HX, Southerton SG (2010) Allelic variation in cell wall candidate genes affecting solid wood properties in natural populations and land races of *Pinus radiata*. Genetics 185:1477–1487

Du Q et al (2015) Identification of additive, dominant, and epistatic variation conferred by key genes in cellulose biosynthesis pathway in *Populus tomentosa*. DNA Res 22:53–67

Eckert AJ, Bower AD, Gonzalez-Martinez SC, Wegrzyn JL, Coop G, Neale DB (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). Mol Ecol 19:3789–3805

Eckert AJ et al (2009) Association genetics of coastal Douglas fir (*Pseudotsuga menziesii var. menziesii*, Pinaceae). I. Cold-hardiness related traits. Genetics 182:1289–1302

Evans LM et al (2014) Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. Nat Genet 46:1089–1096

Freeman JS, Potts BM, Downes GM, Pilbeam D, Thavamanikumar S, Vaillancourt RE (2013) Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in *Eucalyptus globulus*. New Phytol 198:1121–1134

Frewen BE, Chen THH, Howe GT, Davis J, Rohde A, Boerjan W, Bradshaw HD (2000) Quantitative trait loci and candidate gene mapping of bud set and bud flush in Populus. Genetics 154:837–845

Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245–257

González JR, Armengol L, Solé X, Guinó E, Mercader JM, Estivill X, Moreno V (2007) SNPassoc: an R package to perform whole genome association studies. Bioinformatics 23:654–655

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and Eucalyptus Urophylla using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121–1137

Guan Y, Stephens M (2011) Bayesian variable selection regression for genome-wide association studies and other large-scale problems. Ann Appl Stat 5:1780–1815

Habier D, Fernando R, Kizilkaya K, Garrick D (2011) Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics 12:186

Hallingbäck H, Sánchez L, Wu H (2014) Single versus subdivided population strategies in breeding against an adverse genetic correlation. Tree Genet Genomes 10:605–617

Hanley SJ, Pei MH, Powers SJ, Ruiz C, Mallott MD, Barker JHA, Karp A (2011) Genetic mapping of rust resistance loci in biomass willow. Tree Genet Genomes 7:597–608

Hayes B, Goddard ME (2001) The distribution of the effects of genes affecting quantitative traits in livestock. Genet Sel Evol 33:209–229

Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. Gen Res 91:47–60

Henery ML, Moran GF, Wallis IR, Foley WJ (2007) Identification of quantitative trait loci influencing foliar concentrations of terpenes and formylated phloroglucinol compounds in *Eucalyptus nitens*. New Phytol 176:82–95

Ingvarsson PK, Garcia MV, Hall D, Luquez V, Jansson S (2006) Clinal variation in phyB2, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). Genetics 172:1845–1853

Ingvarsson PK, Street NR (2011) Association genetics of complex traits in plants. New Phytol 189:909–922

Isik F (2014) Genomic selection in forest tree breeding: the concept and an outlook to the future. New For 45:379–401

Jorge V, Dowkiw A, Faivre-Rampant P, Bastien C (2005) Genetic architecture of qualitative and quantitative Melampsora larici-populina leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. New Phytol 167:113–127

Kump KL et al (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43:163–168

Lande R (1981) The minimum number of genes contributing to quantitative variation between and within populations. Genetics 99:541–553

Lander ES, Schork NJ (1994) Genetic dissection of complex traits science 265:2037–2048

Latta RG (1998) Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. Am Nat 151:283–292

Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. Genetics 164:1205–1219

Lee SH, Van der Werf JHJ (2006) Simultaneous fine mapping of multiple closely linked quantitative trait loci using combined linkage disequilibrium and linkage with a general pedigree. Genetics 173:2329–2337

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Inc., Sunderland, MA

Ma X-F, Hall D, Onge KRS, Jansson S, Ingvarsson PK (2010) Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. Genetics 186:1033–1044

Maher B (2008) Personal genomes: the case of the missing heritability. Nature 456:18–21

McKown AD et al (2014) Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. New Phytol 203:535–553

Meuwissen T, Goddard M (2010) Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185:623–U338

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Meuwissen THE, Solberg TR, Shepherd R, Woolliams JA (2009) A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. Genet Sel Evol 41:2–11

Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. Nat Rev Genet 12:111–122

Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. Trends Plant Sci 9:325–330

Nystedt B et al (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584

Orr HA (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution 52:935–949

Otto SP, Jones CD (2000) Detecting the undetected: estimating the total number of loci underlying a quantitative trait. Genetics 156:2093–2107

Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, Chatterjee N (2010) Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet 42:570–U139

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726

Pavy N, Namroud MC, Gagnon F, Isabel N, Bousquet J (2012) The heterogeneous levels of linkage disequilibrium in white spruce genes and comparative analysis with other conifers. Heredity 108:273–284

Pelgas B, Bousquet J, Meirmans PG, Ritland K, Isabel N (2011) QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments. Bmc Genomics 12:145–167

Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology 87:2614–2625

Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ (2009) Shades of gray: the world of quantitative disease resistance. Trends Plant Sci 14:21–29

Porth I et al (2013) Genome-wide association mapping for wood characteristics in Populus identifies an array of candidate single nucleotide polymorphisms. New Phytol 200:710–726

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67:170–181

Quesada T et al (2010) Association mapping of quantitative disease resistance in a natural population of loblolly pine (Pinus taeda L.). Genetics 186:677–686

Rehfeldt G (1992) Early selection in Pinus ponderosa: compromises between growth potential and growth rhythm in developing breeding strategies. For Sci 38:661–677

Renaut S et al (2013) Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat Commun 4:1827

Resende MFR et al (2012) Accuracy of genomic selection methods in a standard data set of loblolly pine (Pinus taeda L.). Genetics 190:1503–1510

Rockman MV (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution 66:1–17

Salomé PA, Bomblies K, Laitinen RAE, Yant L, Mott R, Weigel D (2011) Genetic architecture of flowering-time variation in Arabidopsis thaliana. Genetics 188:421–433

Sanchez L, Yanchuk AA, King JN (2008) Gametic models for multitrait selection schemes to study variance of response and drift under adverse genetic correlations. Tree Genet Genomes 4:201–212

Segura V, Vilhjalmsson BJ, Platt A, Korte A, Seren U, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet 44:825–U144

Sham PC, Cherny SS, Purcell S, Hewitt JK (2000) Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for Sibship data. Am J Hum Gen 66:1616–1630

Shepherd R, Meuwissen T, Woolliams J (2010) Genomic selection and complex trait prediction using a fast EM algorithm applied to genome-wide markers. BMC Bioinformatics 11:529

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. Tree Genet Genomes 9:901–911

Thavamanikumar S, Southerton S, Bossinger G, Thumma B (2013) Dissection of complex traits in forest trees—opportunities for marker-assisted selection. Tree Genet Genomes 9:627–639

Thumma BR, Southerton SG, Bell JC, Owen JV, Henery ML, Moran GF (2010) Quantitative trait locus (QTL) analysis of wood quality traits in Eucalyptus nitens. Tree Genet Genomes 6:305–317

Tian F et al (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43:159–162

Tian J, Du Q, Li B, Zhang D (2014) Single-nucleotide polymorphisms in the 5′ UTR of UDP-glucose dehydrogenase (PtUGDH) associate with wood properties in Populus tomentosa. Tree Genet Genomes 10:339–354

Tuskan GA et al (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & gray). Science 313:1596–1604

Ukrainetz NK, Ritland K, Mansfield SD (2008) Identification of quantitative trait loci for wood quality and growth across eight full-sib coastal Douglas-fir families. Tree Genet Genomes 4:159–170

Vilhjálmsson BJ, Nordborg M (2013) The nature of confounding in genome-wide association studies. Nat Rev Genet 14:1–2

Wurschum T, Kraft T (2015) Evaluation of multi-locus models for genome-wide association studies: a case study in sugar beet. Heredity 114:281–290

Xu S (2003a) Estimating polygenic effects using markers of the entire genome. Genetics 163:789–801

Xu SZ (2003b) Theoretical basis of the Beavis effect. Genetics 165:2259–2268

Yang J, Lee SH, Goddard ME, Visscher PM (2011a) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Gen 88:76–82

Yang J et al (2011b) Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet 43:519–525

Yang JA et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42:565–U131