Dissection of complex traits in forest trees – opportunities for marker-assisted selection

Saravanan Thavamanikumar^{1,2}, Simon G. Southerton³, Gerd Bossinger^{1,2}, and Bala R. Thumma^{3,*}

¹*Department of Forest and Ecosystem Science, The University of Melbourne, Creswick, Victoria 3363, Australia*

²*Co-operative Research Centre for Forestry, Private Bag 12, Hobart, Tasmania 7001 Australia*

³*CSIRO Plant Industry, GPO Box 1600, Acton, ACT 2601, Australia*

**Corresponding author: Bala R. Thumma, CSIRO Plant Industry, GPO Box 1600, Acton, ACT 2601, Australia*

Email: Reddy.Thumma@csiro.au

Phone: +61-02-62464856

ABSTRACT

Due to their long reproductive cycles and the time to expression of mature traits, marker-assisted selection is particularly attractive for tree breeding. In this review we discuss different approaches used for developing markers and propose a method for application of markers in low linkage disequilibrium (LD) populations. Identification of useful markers for application in tree breeding is mainly based on two approaches, quantitative trait locus mapping (QTL) and association genetic studies. While several studies have identified significant markers, effect of the individual markers is low making it difficult to utilize them in breeding programs. Recently genomic selection (GS) was proposed for overcoming some of these difficulties. In GS, high density markers are used for predicting phenotypes from genotypes. Currently small effective populations with high LD are being tested for GS in tree breeding. For wider application, GS needs to be applied in low LD populations which are found in many tree breeding programs. Here we propose an approach in which the significant markers from association studies may be used for developing prediction models in low LD populations using the same methods as in GS. Preliminary analyses indicate that a modest numbers of markers may be sufficient for developing prediction models in low LD populations. Genomic selection based on large numbers of random markers or small numbers of associated markers is poised to make marker-assisted selection a reality in forest tree breeding.

Key words: *Genomic selection, genomic estimated breeding values, association mapping, QTL mapping, linkage disequilibrium.*

1 **Introduction**

2 Traits that do not follow classical Mendelian monogenic inheritance are termed “complex” or
3 “quantitative” traits (Lander and Schork 1994). Such traits are usually governed by variation
4 in multiple genes and are typically influenced by the environment. One of the major
5 challenges in the field of genetics is to account for all inherited phenotypic variation at the
6 molecular level. In addressing this challenge it is often difficult to identify molecular markers
7 that exhibit perfect co-segregation with complex traits because of their polygenic inheritance
8 and other factors including epistasis, incomplete penetrance, phenocopy and locus
9 heterogeneity (Lander and Schork 1994). However, it is essential to dissect complex traits
10 into their genetic components in order to facilitate breeding through marker-assisted selection
11 (MAS). In tree species, where most phenotypic traits are complex, the use of MAS for tree
12 improvement is particularly attractive because conventional selection is impeded by long
13 generation times and the long delay in the full expression of mature traits (Grattapaglia et al.
14 1996).

15 Attempts at identifying suitable markers for breeding have until recently focused on the
16 mapping of linkage-based quantitative trait loci (QTL). Using this approach significant
17 progress has been made in identifying genomic regions that underlie complex traits in
18 numerous species (Goddard and Hayes 2009; Grattapaglia et al. 2009; Mauricio 2001).
19 However, due to several limitations of the QTL approach (see below), the focus of marker
20 discovery research in trees has increasingly shifted to association genetic mapping. In this
21 paper, we (1) review the methods for dissection of complex traits such as QTL and
22 association studies, (2) discuss the advances made using these studies and limitations of these
23 approaches in applied breeding programs, (3) discuss the current method of genomic
24 selection to overcome some of these limitations and (4) discuss how genomic selection
25 models can be applied in low LD populations using markers identified from association
26 studies.

27 **Quantitative trait locus (QTL) mapping**

28 QTL mapping is generally performed in pedigrees from controlled crosses and involves
29 associating variation in regions of a genome with variation in measured phenotypic traits.
30 Generally, only two alleles segregate at a locus in backcross and F₂ pedigrees constructed in

1 inbred crop species. This is because segregation in the mapping generation results from
2 meiosis in a single parent. However, in outcrossing and highly heterozygous forest trees, up
3 to four alleles can segregate at a single locus (Sewell and Neale 2000). This is because
4 segregation results from separate meioses and crossovers in the two parents (Groover et al.
5 1994). Where a QTL segregates for more than two alleles, there is an increased likelihood
6 that allelic interactions such as simple dominance, over-dominance or under-dominance will
7 occur (Groover et al. 1994).

8 **QTL mapping in forest trees**

9 Until recently, QTL mapping was the principal means of dissecting the genetics of complex
10 traits in forest tree species. QTL have been identified for growth and wood property traits
11 (Brown et al. 2003; Devey et al. 2004; Markussen et al. 2003; Thumma et al. 2010b),
12 vegetative propagation (Marques et al. 2002; Scotti-Saintagne et al. 2005; Thumma et al.
13 2010a), and disease resistance (Newcombe et al. 1996) in a range of forest trees. While
14 several QTL have been identified, individual QTL however typically explain only a small
15 proportion of phenotypic variation. For example, using an outcrossed F₂ family of *E.*
16 *globulus*, a recent study identified a number of QTL influencing growth (diameter at breast
17 height) and wood properties (wood density, cellulose content, pulp yield and lignin content)
18 (Freeman et al. 2009). One to four QTL were identified for each of these traits, which jointly
19 explained between 3.8% and 34.9% of the phenotypic variation in the same population.
20 Because of the small number of individuals used most of these marker effects may have been
21 over estimated (Beavis 1998).

22 Though QTL mapping helps determine the complexity of the genetic architecture underlying
23 phenotypic traits, the mode of action of specific genes within QTL remains unclear as the
24 genomic regions identified are usually large (Doerge 2002). Most QTL span hundreds, and
25 occasionally thousands, of genes (Nadeau and Frankel 2000). For example, 95% confidence
26 intervals for three separate QTL observed for wood basic density in a recent QTL study in *E.*
27 *globulus* (Freeman et al. 2009) covered 39, 40 and 28 centimorgans (cM) respectively.
28 Similarly, in *E. nitens*, five separate QTL identified for wood basic density had 95%
29 confidence intervals ranging from 38 to 50 cM (Thumma et al. 2010b). In eucalypts, a single
30 cM, which represents the distance across which one recombination will occur in 1% of trees
31 in a single generation, was estimated to be approximately 200 kb (Grattapaglia and Bradshaw

1994). Assuming 5 to 10 expressed sequence tags (ESTs) per 200 kb¹, this equates to between 135 and 400 ESTs per QTL. QTL mapping resolution cannot be increased by increasing the number of markers alone as the major limitation to map resolution is the low number of recombination events observed. Although the resolution could be increased by increasing the number of progeny along with more markers, the increased resolution would still likely to be insufficient for identification of candidate genes.

While the marker effects observed within species are small to moderate, marker effects observed in crosses between species are however larger. Shepherd et al (2006, 2008) identified large effect QTL explaining as much as 66% of variation in rooting ability by analysing the genetic architecture of adventitious root system in *Pinus* and *Corymbia*. In *Pinus*, by comparing QTL detected in F₁ and F₂ populations they observed that the size of the QTL effects observed within species was lower compared to that of between species (Shepherd et al. 2006). Similarly using an interpecific cross in *Populus*, (Bradshaw and Stettler 1995) observed growth QTL explaining as much as 45% of genetic variation in stem volume. Some of the large effect QTL observed in these inter-specific crosses may be real. Transgressive segregation may explain the large effect QTL observed in the inter-specific crosses. Complimentary gene action of the recombined alleles in hybrids that were fixed with opposite sign in parents results in transgressive segregation (Stelkens and Seehausen 2009).

Because QTL studies are conducted in pedigrees from a small number of parents, only a fraction of the potential variation underlying a quantitative trait can be investigated. As the interval between the marker and the QTL is generally large, recombination between the marker and the QTL may make it difficult to observe the QTL in a different pedigree (Verhaegen et al. 1997). Since a QTL detected in one pedigree may not segregate in another pedigree, validating QTL results in different pedigrees (Lander and Schork 1994; Rocha et al. 2007) is challenging. For example, in a QTL study in *E. globulus* conducted across multiple pedigrees and sites, only 19 of the 98 QTL, which affected wood properties and growth, were significant in more than one family (Freeman et al. 2011). This can limit the ability of QTL

¹estimated using the *Eucalyptus* Genome Database browser (<http://eucalyptusdb.bi.up.ac.za/>)

studies to guide broader inferences on the genetic architecture of any given trait. In contrast to QTL mapping, association studies can be used to explore population wide variation.

Association genetic mapping

Detection of significant associations between allelic variation and phenotypic variation at a population level is variously termed as association genetics, association mapping or linkage disequilibrium (LD) mapping. Association studies are generally carried out in populations containing several hundreds of unrelated individuals. The resolution of marker-trait associations identified is much higher than in QTL studies as association studies exploit the low LD generally found in natural populations. However to exploit the advantage of high resolution due to low LD, high density markers are needed. Accumulation of numerous recombination events over many generations in natural populations breaks all long-range linkages. This results in short stretches of high LD between the loci and permits fine-scale mapping of marker-trait associations (Nordborg and Tavaré 2002). Moreover, association genetics exploits trait variation existing in the whole population while QTL studies are restricted to variation present between the parents of the pedigree used in the study (Neale and Savolainen 2004). Therefore, as markers are identified in analyses of diverse populations, these markers may be used in most early generation tree breeding populations which are based on a large number of families.

Understanding of the nucleotide diversity (π) patterns in a population is important for association studies. Several studies have examined nucleotide diversity in tree species (Table 1). Nucleotide diversity is measured as average number of base pair (nucleotide) differences per base pair between orthologous (allelic) sequences of two trees drawn at random from a population. In forest trees the estimates of nucleotide diversities have ranged from 0.00156 (*Cryptomeria japonica*; Kado et al. 2008) to 0.01110 (*Populus tremula*; Ingvarsson 2005). While most of the studies have analysed only a small number of genes it appears that the available nucleotide diversity is large enough to be useful for association studies. Compared to studies in humans, association studies are new to forest trees, however, early results are promising and a number of loci associated with a range of traits have been identified (Table 2).

Linkage disequilibrium and its impact on the resolution of association mapping

In addition to nucleotide diversity, linkage disequilibrium (LD) has important implications in association mapping. Linkage disequilibrium is the non-random association of alleles at two or more loci, not necessarily on the same chromosome, in which some combinations of alleles occur more or less frequently in a population than would be expected if alleles were inherited independently (Neale et al. 2002; Nordborg and Tavare 2002). A number of evolutionary phenomena including selection, genetic drift, gene flow, mutation, and factors including an active mating system, population size and structure generate LD (Rafalski and Morgante 2004; Terwilliger and Weiss 1998). Recombination between alleles leads to break down in LD (Terwilliger and Weiss 1998).

In general, LD decays more quickly in outcrossing species than in inbreeding species. For instance, LD decays rapidly in outcrossing species like maize (decays in just 100 bp; Tenaillon et al. 2001), potato (decays within 1 kb; Simko et al. 2006), loblolly pine (decays within 800 bp; Gonzalez-Martinez et al. 2006), Scots pine (decays within 750 bp; Garcia-Gil et al. 2003), Douglas fir (decays within 2 kb; Krutovsky and Neale 2005), white spruce (decays within 65 bp; Pavy et al. 2012), Shining gum (decays within 850 bp; Thumma et al. 2005), and European aspen (decays within 500 bp; Ingvarsson 2005). In contrast, LD extends over longer distances in selfing species like soybean (extends over 50 kb; Zhu et al. 2003), *Arabidopsis* (extends to 250 kb; Hagenblad and Nordborg 2002) and rice (extends to 100 kb; Garriss et al. 2003). LD patterns may vary between different populations of a species that show significant structure (Goddard et al. 2000; Kidd et al. 2004) and between different parts of the genome (Daly et al. 2001; Dawson et al. 2002). In structured populations, LD should be computed within populations as pooling population samples with different allele frequencies will produce LD between unlinked loci (Mangin et al. 2011). Although LD patterns in different populations were studied (Heuertz et al. 2006; Ingvarsson 2005; Pyhajarvi et al. 2007), LD in different regions of the genome is not yet characterized well in forest trees.

The extent of LD has two main implications for association studies. Rapid decay of LD enables high resolution mapping of associations between the marker and the trait (Grattapaglia and Kirst 2008; Thumma et al. 2005) and can even lead to the identification of causative polymorphisms (Gaut and Long 2003; Thumma et al. 2009). However, as high

1 densities of markers are needed to attain an appropriate level of genomic coverage, low LD
2 makes genome-wide association studies (GWAS), such as those carried out in humans
3 (Petukhova et al. 2010) and Arabidopsis (Todesco et al. 2010), unfeasible in many forest tree
4 species. Consequently, candidate gene-based association studies are currently the preferred
5 approach in forest trees (Gonzalez-Martinez et al. 2007; Thumma et al. 2005). In candidate
6 gene-based association studies markers from candidate genes controlling the relevant traits
7 are tested for association while in GWAS random markers spread across the genome are
8 tested. However, as the cost of next generation sequencing continues to fall, GWAS based on
9 sequencing individual genomes may soon become possible at least for species with reference
10 sequences such as *Eucalyptus* and *Populus*.

11 **Experimental design of association studies**

12 Two main designs based on whole populations and individual families are used for
13 conducting association studies in forest trees. In the population-based approach, unrelated
14 individuals or a few individuals from many different families are used while in the family-
15 based approach multiple individuals from a smaller number of different families are used.
16 However, the main drawback of population-based approaches is that they can suffer from
17 demographic problems, such as population structure, which can lead to false positives
18 (Lander and Schork 1994). These spurious associations may be avoided by performing
19 association studies in homogenous populations rather than in structured populations (Lander
20 and Schork 1994; Pritchard et al. 2000b). Alternatively, statistical models can be used to infer
21 population structure (Pritchard et al. 2000a) or to account for multiple levels of relatedness
22 (Yu et al. 2006). Accounting for population structure may sometimes result in false negatives
23 when the polymorphism associated with the trait coincides with the population structure (Yu
24 and Buckler 2006; Zhao et al. 2007).

25 Family-based designs, on the other hand, are robust against false positives associated with
26 population structure. Family-based approaches such as quantitative transmission
27 disequilibrium tests (QTDT) exploit information from linkage and linkage disequilibrium to
28 find associations (Benyamin et al. 2009). Using this approach Gonzalez-Martinez et al.
29 (2008) identified two SNPs from a *dehydrin* and a cell wall gene that were significantly
30 associated with water use efficiency in *Pinus taeda*. While family-based association studies
31 are robust against population structure, in general they have low power and resolution

1 compared to population-based association studies. Another drawback of this method is the
2 requirement for heterozygous parents in the families.

3 **Selection of candidate genes and polymorphisms for association studies**

4 Candidate genes for association studies are selected from expressed sequence tag (EST)
5 libraries derived from tissues likely to influence the trait in question. For wood quality studies
6 in eucalypts, a number of EST libraries have been generated from developing xylem tissue
7 (Paux et al. 2004; Rengel et al. 2009). A large number of candidate genes have been
8 identified based on their 'differential expression' in a range of forest tree species including
9 eucalypts (Foucart et al. 2006; Paux et al. 2004; Qiu et al. 2008), loblolly pine (Allona et al.
10 1998; Whetten et al. 2001) and spruce (Friedmann et al. 2007). With the steady decrease in
11 the cost of next generation sequencing, even larger numbers of candidate genes are expected
12 to become available in the near future, even for species with currently limited genomic
13 resources (Neale and Kremer 2011). Recent developments in genomics technology are also
14 making it possible to identify functional SNPs controlling traits. For example, using next
15 generation methods for transcriptome sequencing it is possible to detect not only candidate
16 genes but also functional markers controlling the expression of these candidate genes. RNA
17 sequencing of *Eucalyptus camaldulensis* seedlings subjected to water stress was recently used
18 to identify several candidate genes along with several potential functional SNPs based on
19 differential allelic expression (Thumma et al. 2012).

20 **SNP genotyping and association analysis**

21 Common SNPs are identified by sequencing candidate genes in several individuals (Thumma
22 et al. 2009; Gonzalez-Martinez et al. 2006). To capture species-wide diversity, trees
23 representing all the major sub-populations or races of a species are typically included
24 (Thavamanikumar et al. 2011). Subsequently, common and haplotype tagging SNPs (Johnson
25 et al. 2001) are selected for genotyping in the larger population. The frequency and position
26 of polymorphisms, the presence of haplotypes and LD between the SNPs are some of the
27 factors influencing the selection of SNPs for genotyping (Tabor et al. 2002). SNPs from
28 coding regions, especially nonsynonymous substitutions (substitutions that will change the
29 amino acid sequence), are generally targeted as these substitutions are more likely to impact
30 phenotype than substitutions in coding regions that do not alter the amino acid sequence

(Terwilliger and Weiss 1998). However, there is increasing evidence that other silent site mutations are similarly likely to confer phenotypic effects by a variety of mechanisms. For example, two intronic mutations in the *cinnamoyl CoA reductase* gene (*CCR*) were reported to associate with microfibril angle (Thumma et al. 2005) and a synonymous SNP in *EniCOBL4A* gene was reported to associate with cellulose content by affecting allelic expression in *E. nitens* (Thumma et al. 2009). Similarly, an intronic SNP in the α -tubulin gene of *Pinus taeda* was strongly associated with microfibril angle (Gonzalez-Martinez et al. 2007). In white spruce, one intronic SNP each from β -expansin and β -TIP (β -Tonoplast Intrinsic Protein) was associated with percentage of early wood and tracheid diameter in the latewood, respectively (Beaulieu et al. 2011). In *Drosophila melanogaster*, a number of intronic mutations from several loci have been found to associate with bristle number (Long et al. 1998; Lyman et al. 1999). Likewise, a number of promoter variants were found to associate with a variety of diseases in human association studies (Clement et al. 2009; Goyenechea et al. 2009; Lan et al. 2009; Marquet et al. 2008). These findings suggest that important markers may be missed if only coding non-synonymous SNPs are selected for genotyping.

Validation of marker-trait associations

False positives in association mapping studies can result from population stratification when sub-populations differ in their SNP allele frequency and mean trait values (Koller et al. 2004) or due to multiple testing errors. A range of methods are now available to account for population structure and to correct for multiple testing errors. Several statistical methods have been proposed to account for this structure when testing for association (Pritchard et al. 2000b; Yu et al. 2006). Validation of marker-trait associations in one or more independent populations is another way of identifying robust markers and reducing false positives. Associations have been successfully validated in independent populations in some studies (Dillon et al. 2012; Dillon et al. 2010; Thumma et al. 2009; Thumma et al. 2005). Functional studies may also help in validating associations. For example, in *E. nitens*, Thumma et al. (2009) have identified a synonymous SNP in the *EniCOBL4A* gene associated with cellulose content. Using allelic expression imbalance studies and nuclear protein binding studies, these authors showed that this synonymous SNP is a *cis*-acting regulatory polymorphism affecting allelic expression. Recently, Beaulieu et al. (2011) have identified 13 SNPs associated with

1 wood traits in white spruce. Expression of three genes was significantly correlated with SNP
2 genotypes indicating the functional significance of the associated SNPs.

3 Low frequency polymorphisms have low power in detecting associations as large populations
4 are required to test the effect of rare variants (Long and Langley 1999). For this reason, and
5 because populations of a few hundred individuals are generally used in association studies in
6 forest trees, common SNPs (> 0.10 MAF, minor allele frequency) are generally used. Recent
7 genome-wide association studies in humans and a number of candidate-gene based
8 association studies in forest trees have shown that the effect of common SNPs is typically
9 modest (Table 2). In view of these results it has been suggested that rare SNPs with large
10 effect on the traits may exist which are not detected by current methods (Goldstein 2009). In
11 order to identify such rare SNPs large populations with thousands of individuals are required
12 (Grattapaglia and Kirst 2008); inevitably there is a trade-off between the power to detect an
13 association and the size of the effect a polymorphism has on phenotype.

14 While several studies have identified markers linked to QTL and SNP markers within the
15 candidate genes affecting a trait, these markers are not currently being used in breeding
16 programs. In addition to the difficulties of using markers from QTL studies as described
17 earlier, the effect of markers identified from both QTL and association studies are small.
18 Often individual markers explain less than 5% of total variation. Two SNPs that were
19 independently associated with microfibril angle (MFA) in *E. nitens* explained between 3.4
20 and 5.9% of total variation in MFA (Thumma et al. 2005). In *Populus tremula*, two
21 nonsynonymous SNPs were independently associated with bud set (Ingvarsson et al. 2008).
22 After correcting for the possible upward bias in the effect size, these two SNPs explained 1.4
23 and 5.9% of the variation in bud set. In a large multilocus association mapping study
24 involving 117 genes in *Pseudotsuga menziesii*, all SNPs independently explained less than
25 5% of the phenotypic variation in traits (Eckert et al. 2009a). In white spruce, 13 SNPs
26 associated with wood quality explained between 3 to 5% of trait variation individually
27 (Beaulieu et al. 2011). A large proportion of the trait variation may be captured by combining
28 the effects of several markers (Yang et al. 2010). The recently developed genomic selection
29 (GS) approach exploits the effects of several markers for predicting phenotypes for selection
30 in breeding programs.

Genomic selection – a new approach for marker-assisted selection in tree breeding

Selecting individuals based on genomic estimated breeding values (GEBVs) is termed genomic selection (Hayes et al. 2009). Genomic selection was pioneered by Meuwissen et al. (2001) and has been quickly adopted by animal and plant breeders (Hayes et al. 2009; Heffner et al. 2009; Lorenzana and Bernardo 2009; Luan et al. 2009). Recently this approach has been applied to tree breeding (Resende et al. 2012a; Resende et al. 2012b; Resende et al. 2012c; Grattapaglia and Resende 2011). In GS, large numbers of progeny from a highly structured population with high LD are genotyped with thousands of markers. The effects of all the markers are used in developing prediction models using ‘training’ or ‘discovery’ populations. These models are then used to predict genomic breeding values of the individuals of the validation population using only the genotype data. The accuracies of the GEBVs are obtained by correlating GEBVs against traditional pedigree-based breeding values. Accuracies of GS models are also obtained by testing the ability of markers to predict phenotypes in one generation by using the model developed in an another generation (Habier et al. 2007). Simulation analyses have indicated that accuracies as high as 0.85 are possible with markers alone (Meuwissen et al. 2001). With high accuracies of GEBVs it is possible to double the genetic gain by shortening the breeding cycle (Hayes et al. 2009). It has been estimated that the application of GS in dairy cattle breeding could lead to cost savings of as much as 92% (Schaeffer 2006).

In GS, as currently applied, a high density of markers is necessary to ensure that at least some of the markers are in LD with the genomic regions affecting the traits. As the number of predictors are generally higher than the number of samples ($p \gg n$), marker effects cannot be estimated using multiple regression by ordinary least squares (Jannink et al. 2010). To overcome multicollinearity and the reduced predictive ability associated with using large number of predictors compared to the samples, statistical models treating markers as random effects are used to develop the prediction models (Jannink et al. 2010). Ridge regression best linear unbiased prediction (RR-BLUP) and Bayesian regression (Meuwissen et al. 2001) are some of the statistical models used for developing prediction models. All of these methods are aimed at reducing the marker variance by shrinking the marker effects toward zero. RR-

1 BLUP assumes equal marker variances while Bayesian models allow each marker to have its
2 own variance (Meuwissen et al. 2001).

3 Traditional best linear unbiased predictions (BLUP) using mixed model equations are now
4 widely used in animal and tree breeding for predicting breeding values (White et al. 2007). In
5 traditional BLUP, phenotypic information from individuals as well as their relatives is used to
6 predict breeding values. Information from relatives in current and past generations is
7 incorporated via the additive relationship matrix, which contains information on the
8 proportion of the genome identical by descent (IBD) between pairs of individuals calculated
9 from their known pedigree relationships (Lynch and Walsh 1998). It has been shown that the
10 accuracy of predictions can be higher when the expected relationship matrix from the
11 pedigree is replaced by the realized relationship matrix generated from markers (Goddard
12 2009; Hayes et al. 2009). As the markers are capable of capturing the random Mendelian
13 sampling variations i.e., variations between the siblings, the accuracy of the predictions
14 would be higher than using the pedigree-based expected relationship matrix (Daetwyler et al.
15 2007). The prediction of GEBVs using the realized relationship matrix is termed genomic
16 best linear unbiased prediction (GBLUP). GBLUP assumes that a large number of genes with
17 equal effects control the traits similar to RR-BLUP (Zhang et al. 2010). GBLUP is equivalent
18 to RR-BLUP and the accuracies of GEBVs from the two methods are similar (Hayes et al.
19 2009). Accuracies of GEBVs from GBLUP or RR-BLUP are higher if a trait is controlled by
20 a large number of QTL while the accuracies from Bayesian methods are higher if a trait is
21 influenced by a small number of QTL (Daetwyler et al. 2010).

22 When the density of markers is high, most of the genetic effects will be captured by the
23 markers. However, when the marker density is low inclusion of a polygenic term in the
24 model may help in capturing genetic variance not captured by the markers (Haley and
25 Visscher 1998). Simulation studies have shown that when marker density is high the accuracy
26 of GEBVs did not improve with the inclusion of polygenic effect in the model. At lower
27 marker densities however the accuracy of the GEBVs improved with the inclusion of the
28 polygenic effect. However, the inclusion of polygenic effect reduced the bias between the
29 estimated variance components and the simulated values even for high density markers
30 suggesting increased bias in the estimated values without the polygenic effect at higher
31 marker densities (Calus and Veerkamp 2007; Solberg et al. 2009).

1 Simulation studies have also shown that inclusion of polygenic effect will improve the
2 persistence of marker effects over several generations (Solberg et al. 2009). At higher
3 densities markers capture family relationships leading to spurious associations, i.e.,
4 association due to family relationships rather than markers linked to the causative QTL. Such
5 spurious associations due to family relationships decay faster over time compared to
6 associations due to markers in LD with the QTL. By including the polygenic effect in the
7 model, associations due family relationships are captured by polygenic term leaving the QTL
8 variation to be captured by the markers. This will lead to persistence of the accuracies as the
9 predictions are based on the markers in LD with the QTL which decay more slowly over time
10 (Solberg et al. 2009). In association studies similar approaches (marker based kinship
11 estimates) are used to control for relatedness among the individuals (Yu et al. 2006).

12 GS models utilize both LD between markers and the QTL as well as the family relationships
13 in estimating GEBVs. RR-BLUP mainly exploits the family relationships while Bayesian
14 methods use both family relationships as well as LD between the markers and the QTL in
15 estimating the GEBVs (Habier et al. 2007). As the contribution of genetic relationships
16 towards the accuracy of GEBVs is halved in each descendant generation following the
17 generation in which the model was developed, the ranking of the individuals may change in
18 different generations for the methods such as RR-BLUP which mainly exploit the genetic
19 relationships in estimating GEBVs. Bayesian methods on the other hand exploit LD between
20 markers and the QTL better than RR-BLUP and therefore the accuracies of Bayesian
21 methods will persist over several generations (Habier et al. 2007).

22 **Genomic selection in forest trees**

23 Simulation studies in forest trees have indicated that in populations with small effective size
24 and high LD, GS can substantially improve the efficiency of conventional tree breeding
25 (Grattapaglia and Resende 2011; Iwata et al. 2011). Only a few markers per genetic distance
26 (two to 20 markers/cM depending on the effective population size) are needed to reach the
27 accuracies of selection based on traditional BLUP estimates. Simulation studies have shown
28 that selection efficiencies (gain with markers in comparison to phenotypic selection) of more
29 than 100% could be obtained with GS (Grattapaglia and Resende 2011). In a proof of concept
30 study, Resende et al. (2012c) tested the utility of GS for predicting growth traits in *Pinus*
31 *taeda*. By using genotype information from 4825 markers in 800 individuals from a

1 structured population they developed prediction models for height and DBH. Accuracies of
2 GS for these traits ranged from 0.64 to 0.74. Compared to phenotypic selection, a selection
3 efficiency of between 1.53 and 2.00 was obtained with GS. Similarly, Resende et al. (2012a)
4 have observed accuracies of between 0.55 and 0.88 for pulp yield and growth traits in two
5 *Eucalyptus* breeding populations using more than 3000 markers. While the accuracies of the
6 models were found to be high, the accuracy for predicting across different populations and
7 sites were low in the two studies suggesting significant genotype by environment (G x E)
8 interactions. Significant G x E interactions between populations may arise due to differences
9 in allele frequencies and LD patterns between the populations. Results from these studies
10 however clearly show that marker-assisted selection with GS has the potential to significantly
11 improve the efficiency of traditional tree breeding. In practical tree breeding programs, GS
12 effectively replaces the prolonged testing phase and thus drastically reducing the breeding
13 cycles (Figure 1). In *Eucalyptus* breeding, GS would result in 50% reduction in breeding
14 cycle resulting in economic return of 20 times on the investment (Resende et al. 2012a).

15 The main limitation of applying GS in forest tree breeding is the need to develop breeding
16 populations with high LD using small effective populations. Most of the early generation tree
17 breeding populations generally contain large numbers of open pollinated families (Eldridge et
18 al. 1994). Linkage disequilibrium will typically be low in such populations. Other difficulties
19 associated with GS in high LD populations are the potential breakdown in the accuracy of
20 prediction models in advanced generations and the inability of the models to predict traits
21 across sites and/or populations due to genotype X environment interactions and differences in
22 LD between the populations (Hayes et al. 2009). As the populations used for GS are derived
23 from a few parents, loss of genetic diversity and increase in the rates of inbreeding are of
24 concern. However, as markers can capture random Mendelian segregation variations,
25 application of GS within a given population results in reduced inbreeding compared to
26 traditional BLUP-based selections (Hayes et al. 2009; Heffner et al. 2009; Jannink et al.
27 2010). Using traditional BLUP two siblings from a full-sib family may receive similar
28 breeding values if their phenotypes are similar. Therefore both of them will be selected using
29 traditional BLUP methods leading to increased selection of individuals with similar genetic
30 makeup. However, with markers, the siblings can be separated by capturing random
31 Mendelian variations among the siblings leading to lower inbreeding (Hayes et al. 2009).

1 Some of the limitations of using GS in high LD populations may be overcome by using
2 markers identified from association studies. Once markers from several candidate genes
3 explaining a significant proportion of a trait are identified, allele effects of all the markers can
4 be used to generate marker-based breeding values as in GS. The pre-selected markers may be
5 used in random effects models as in GS or in fixed effects models (fixed regression-least
6 squares) to estimate the marker effects and to predict GEBVs. As the markers identified from
7 association studies are often in high LD with traits the predictions should be valid over
8 several generations. However, the identification of flip-flop effects (Lin et al. 2007; same
9 marker showing significant association in different populations at different sites but with
10 allelic effects swapped) in recent association studies (Southerton et al. 2010; Dillon et al.
11 2012) indicate that results of GS using markers from association studies may also be
12 population specific unless stable markers with allelic effects in the same direction from
13 across sites are used.

14 Lack of large number of markers is the main impediment for developing prediction models in
15 low LD populations. Recent developments in sequencing and genotyping technologies should
16 accelerate identification of markers associated with the traits. Transcriptome sequencing can
17 be used to identify candidate genes and SNPs at genome wide scale (Thumma et al. 2012).
18 Potential genes and markers showing differential expression can be used in association
19 studies to test the effect of the SNPs in different populations.

20 In a preliminary study to test the efficacy of using markers from association studies for
21 application of GS in low LD populations, we used a modest number of markers (90 SNPs)
22 associated with several wood quality traits at a marginal significance level in *E. nitens*
23 (Thumma et al. unpublished data) to develop prediction models. Prediction models were
24 developed using RR-BLUP, Bayesian LASSO and fixed regression models. Using these
25 prediction models we were able to predict several wood traits with accuracies ranging from
26 0.20 to 0.47 in three different populations. These predictive abilities are similar to those
27 reported for several wood quality traits in a highly structured and high LD population of
28 loblolly pine (Resende et al. 2012c) and the predictive ability of pulp yield (0.38) in
29 Eucalyptus (Resende et al. 2012a). Results from this preliminary study indicate that GS
30 models can be applied in low LD populations using a modest number of markers. Increasing
31 the number of markers should potentially increase the accuracies of predictions.

1 **Conclusions**

2 QTL mapping has been used extensively in forest trees to identify genomic regions that
3 contribute to variation in complex traits. However, the low resolution of QTL mapping limits
4 its application to tree breeding. Association mapping has recently been used successfully to
5 identify several SNP markers in forest tree species which could be used in breeding
6 programs. Due to the moderate to high nucleotide diversity and low LD in most forest tree
7 species candidate gene-based association studies are most likely to yield success. However,
8 the low proportion of phenotypic variation explained by individual SNPs is consistent with
9 earlier results from QTL studies. The small effect of individual QTL/SNPs makes it difficult
10 to apply these markers in breeding programs. Genomic selection has been proposed to
11 overcome the limitations of using markers in breeding programs. By using large numbers of
12 markers in prediction models GS can significantly improve the selection efficiency in
13 breeding programs. So far the application of GS has been restricted to highly structured
14 populations with high LD. For wider application in forest tree breeding programs, GS needs
15 to be tested in low LD populations. For this, markers previously identified from association
16 studies can be used to develop prediction models. As the markers from association studies are
17 in general in high LD with the causal variants, prediction models from such markers should
18 persist over several generations. Preliminary results indicate a modest number of markers
19 associated with several traits at nominal significance level may be enough for developing
20 prediction models in low LD populations. However, identification of sufficient markers is
21 important for developing prediction models with high accuracy in low LD populations.
22 Application of GS in high LD advanced generation populations or candidate gene-based
23 genomic selection in low LD early generation populations promises to have a major impact
24 on forest tree breeding.

Figure Legend

Figure 1: Improving the efficiency of tree breeding with GS. Genomic selection effectively replaces expensive and time consuming progeny trials which may take between 8-20 years depending on the species and the breeding system thus shortening the breeding cycle. With the accuracies of genomic selection similar to traditional phenotype based selection, marker genotypes of seedlings can be used for selecting superior individuals for the next breeding cycle. The selected individuals can be used for field testing and clonal trials. The genotype and phenotype information from some of these individuals can be used to re-train and refine the model.

Table 1. Nucleotide diversity and linkage disequilibrium studies in forest tree species

Common name	Scientific name	Function of genes studied	Number of genes studied	Mean total nucleotide diversity (θ_n)	LD decay	References
Loblolly pine	<i>Pinus taeda</i>	Wood formation	19	0.00398	Within 2000 bp	Brown et al. (2004)
		Drought response	18	0.00507	Within 800 bp	Gonzalez-Martinez et al. (2006)
Maritime pine	<i>Pinus pinaster</i>	Wood formation	8	0.00241	-	Pot et al. (2005)
		Drought response	11	0.00548	-	Eveno et al. (2008)
Radiata pine	<i>Pinus radiata</i>	Wood formation	8	0.00186	-	Pot et al. (2005)
Scots pine	<i>Pinus sylvestris</i>	Cold-related	14	0.00600	Within 200 bp	Wachowiak et al. (2009)
		Allozyme coding	6	0.01000 ^a	No appreciable decay	Pyhajarvi et al. (2011)
Douglas Fir	<i>Psuedotsuga menziesii</i>	Cold Hardiness and wood formation	18	0.00655	Within 1500 bp	Krutovsky and Neale (2005)
		Cold Hardiness	121	0.00435	Within 1000 bp	Eckert et al. (2009b)
Norway spruce	<i>Picea abies</i>	Growth cessation	22	0.00208	Within 100 bp	Heuertz et al. (2006)
White spruce	<i>Picea glauca</i>	Multiple functions	105	0.00430	Within 65 bp	Pavy et al. (2012)
European aspen	<i>Populus tremula</i>	Multiple functions	5	0.01110	Within 500 bp	Ingvarsson (2005)
		Multiple functions	77	0.00420	Within 200 bp	Ingvarsson (2008)
Balsam poplar	<i>Populus balsamifera</i>	Multiple functions	590	0.00280	No appreciable decay	Olson et al. (2010)
Japanese cedar	<i>Cryptomeria japonica</i>	Multiple functions	7	0.00251	-	Kado et al. (2003)
		Multiple functions	5	0.00213	-	Kado et al. (2006)
		Multiple functions	10	0.00156	-	Kado et al. (2008)

^a Watterson's theta: θ_w

Table 2. Association mapping studies in forest tree species

Common name	Scientific name	Markers used		Traits studied	Marker-trait associations found	Annotation of the associated SNPs ^b	Independent validation attempted	Percentage of phenotypic variation explained	References
		Association testing	Background markers ^a						
Shining gum	<i>Eucalyptus nitens</i>	SNP (25)	SSR (14)	Wood quality	2	I	Yes	4.6	Thumma et al. (2005)
		SNP (11)	-	Wood quality	1	S	Yes	-	Thumma et al. (2009)
Tasmanian blue gum	<i>Eucalyptus globulus</i>	SNP (195)	SSR (16)	Foliar secondary metabolites	37	S, NS, I	No	1.8 to 6.1	Kulheim et al. (2011)
Spotted gum	<i>Corymbia citriodora</i> subsp. <i>variegata</i>	SNP (74)	SNP (74)	Wood quality, growth, disease resistance	9	S, I	Yes	3.4 to 8	Dillon et al. (2012)
European aspen	<i>Populus tremula</i>	SNP (41)	SSR (26), SNP (39)	Phenology	2	NS	No	1.5 to 5	Ingvarsson (2008)
Blackbutt	<i>Eucalyptus pilularis</i>	SNP (52)	SSR (12)	Wood quality	2	NS	No	4.1 to 4.2	Sexton et al. (2010)
Loblolly pine	<i>Pinus taeda</i>	SNP (58)	SSR (22)	Wood quality	4	I, NS	No	2.2 to 3.6	Gonzalez-Martinez et al. (2007)
		SNP (46)	-	Carbonisotope discrimination	4	I, NS	No	0.5 to 3.4	Gonzalez-Martinez et al. (2008)
		SNP (3938)	SSR (23)	Disease resistance	10	S, NS, UTR	No	4.8 to 7.2	Quesada et al. (2010)
Radiata pine	<i>Pinus radiata</i>	SNP (149)	SNP (149)	Wood quality	10	S, I	Yes	2.0 to 6.5	Dillon et al. (2010)
Maritime pine	<i>Pinus pinaster</i>	SNP (384)	SNP (98-109)	Growth and wood quality	2	S	Yes	10.0 to 11.4	Lepointevin et al. (2011)
Douglas Fir	<i>Pseudotsuga menziesii</i>	SNP (228)	Isozymes (25), SSR (6)	Cold-Hardiness	30	I, S, NS	No	1.9 to 4.3	Eckert et al. (2009a)
		SNP (113)	SSR (26)	Growth Cessation	6	S, NS	No	6.2 to 12.7	Ma et al. (2010)
White Spruce	<i>Picea glauca</i>	SNP (944)	SNP (944)	Wood quality	25	S, NS, UTR	No	2.6 to 5.4	Beaulieu et al. (2011)
Sitka Spruce	<i>Picea sitchensis</i>	SNP (339)	SNP (98)	Bud set timing and cold-hardiness	45	S, NS	No	0.7 to 5.4	Holliday et al. (2010)

^a Used for identification/accounting of population structure^b I – Intron, S – Synonymous, NS – Nonsynonymous, UTR – Untranslated Regions

References

- Allona I, Quinn M, Shoop E, Swope K, St Cyr S, Carlis J, Riedl J, Retzel E, Campbell MM, Sederoff R, Whetten RW (1998) Analysis of xylem formation in pine by cDNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 95 (16):9693-9698
- Beaulieu J, Doerksen T, Boyle B, Clement S, Deslauriers M, Beauseigle S, Blais S, Poulin PL, Lenz P, Caron S, Rigault P, Bicho P, Bousquet J, MacKay J (2011) Association Genetics of Wood Physical Traits in the Conifer White Spruce and Relationships With Gene Expression. *Genetics* 188 (1):197-214
- Beavis WD (1998) QTL analyses: power, precision, and accuracy. In: Paterson AH (ed) *Molecular Dissection of Complex Traits*. CRC Press, New York, pp 145-162
- Benyamin B, Visscher PM, McRae AF (2009) Family-based genome-wide association studies. *Pharmacogenomics* 10 (2):181-190
- Bradshaw HD, Stettler RF (1995) Molecular-Genetics of Growth and Development in *Populus*. 4. Mapping Qtls with Large Effects on Growth, Form, and Phenology Traits in a Forest Tree. *Genetics* 139 (2):963-973
- Brown GR, Bassoni DL, Gill GP, Fontana JR, Wheeler NC, Megraw RA, Davis MF, Sewell MM, Tuskan GA, Neale DB (2003) Identification of quantitative trait loci influencing wood property traits in loblolly pine (*Pinus taeda* L.). III. QTL verification and candidate gene mapping. *Genetics* 164 (4):1537-1546
- Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences of the United States of America* 101 (42):15255-15260
- Calus MPL, Veerkamp RF (2007) Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. *Journal of Animal Breeding and Genetics* 124 (6):362-368
- Clement K, Stunff CI, Meirhaeghe A, Dechartres A, Ferrieres J, Basdevant A, Boitard C, Amouyel P, Bougneres P (2009) In obese and non-obese adults, the cis-regulatory rs361072 promoter variant of PIK3CB is associated with insulin resistance not with type 2 diabetes. *Molecular Genetics and Metabolism* 96 (3):129-132
- Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* 185 (3):1021-1031
- Daetwyler HD, Villanueva B, Bijma P, Woolliams JA (2007) Inbreeding in genome-wide selection. *Journal of Animal Breeding and Genetics* 124 (6):369-376
- Daly MJ, Rioux JD, Schaffner SE, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. *Nature Genetics* 29 (2):229-232

- Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, Dibling T, Tinsley E, Kirby S, Carter D, Papaspyridonos M, Livingstone S, Ganske R, Lohmussaar E, Zernant J, Tonisson N, Remm M, Magi R, Puurand T, Vilo J, Kurg A, Rice K, Deloukas P, Mott R, Metspalu A, Bentley DR, Cardon LR, Dunham I (2002) A first-generation linkage disequilibrium map of human chromosome 22. *Nature* 418 (6897):544-548
- 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. *Theoretical and Applied Genetics* 108 (3):516-524
- 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 Phytologist* 195 (3):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 (4):1477-1487
- Doerge RW (2002) Multifactorial genetics: mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics* 3 (1):43-52
- Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV, St. Clair JB, Neale DB (2009a) Association Genetics of Coastal Douglas Fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-Hardiness Related Traits. *Genetics* 182 (4):1289-1302
- Eckert AJ, Wegrzyn JL, Pande B, Jermstad KD, Lee JM, Liechty JD, Tarse BR, Krutovsky KV, Neale DB (2009b) Multilocus Patterns of Nucleotide Diversity and Divergence Reveal Positive Selection at Candidate Genes Related to Cold Hardiness in Coastal Douglas Fir (*Pseudotsuga menziesii* var. *menziesii*). *Genetics* 183 (1):289-298
- Eldridge K, Davidson J, Harwood C, van Wyk G (1994) *Eucalypt domestication and breeding*. Clarendon press, Oxford,
- Eveno E, Collada C, Guevara MA, Leger V, Soto A, Diaz L, Leger P, Gonzalez-Martinez SC, Cervera MT, Plomion C, Garnier-Gere PH (2008) Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution* 25 (2):417-437
- Foucart C, Paux E, Ladouce N, San-Clemente H, Grima-Pettenati J, Sivadon P (2006) Transcript profiling of a xylem vs phloem cDNA subtractive library identifies new genes expressed during xylogenesis in *Eucalyptus*. *New Phytologist* 170:739-752
- Freeman J, Potts B, Downes G, Thavamanikumar S, Pilbeam D, Hudson C, Vaillancourt R (2011) QTL analysis for growth and wood properties across multiple pedigrees and sites in *Eucalyptus globulus*. *BMC Proceedings* 5 (Suppl 7):O8
- Freeman JS, Whittock SP, Potts BM, Vaillancourt RE (2009) QTL influencing growth and wood properties in *Eucalyptus globulus*. *Tree Genetics and Genomes* 5 (4):713-722

- Friedmann M, Ralph SG, Aeschliman D, Zhuang J, Ritland K, Ellis BE, Bohlmann J, Douglas CJ (2007) Microarray gene expression profiling of developmental transitions in Sitka spruce (*Picea sitchensis*) apical shoots. *Journal of Experimental Botany* 58 (3):593-614
- Garcia-Gil MR, Mikkonen M, Savolainen O (2003) Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. *Molecular Ecology* 12 (5):1195-1206
- Garris AJ, McCouch SR, Kresovich S (2003) Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.). *Genetics* 165 (2):759-769
- Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. *Plant Cell* 15 (7):1502-1506
- Goddard KAB, Hopkins PJ, Hall JM, Witte JS (2000) Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. *American Journal of Human Genetics* 66 (1):216-234
- Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136 (2):245-257. doi:10.1007/s10709-008-9308-0
- Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews - Genetics* 10 (6):381-391
- Goldstein DB (2009) Common Genetic Variation and Human Traits. *N Engl J Med* 360 (17):1696-1698
- Gonzalez-Martinez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006) DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* 172 (3):1915-1926
- Gonzalez-Martinez SC, Huber D, Ersoz E, Davis JM, Neale DB (2008) Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity* 101 (1):19-26
- Gonzalez-Martinez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007) Association Genetics in *Pinus taeda* L. I. Wood Property Traits. *Genetics* 175:399-409
- Goyenechea E, Collins LJ, Parra D, Abete I, Crujeiras AB, O'Dell SD, Martinez JA (2009) The -11391 GA/ polymorphism of the adiponectin gene promoter is associated with metabolic syndrome traits and the outcome of an energy-restricted diet in obese subjects. *Hormone and Metabolic Research* 41 (1):55-61
- Grattapaglia D, Bertolucci FLG, Penchel R, Sederoff RR (1996) Genetic Mapping of Quantitative Trait Loci Controlling Growth and Wood Quality Traits in *Eucalyptus grandis* Using a Maternal Half-Sib Family and RAPD Markers. *Genetics* 144 (3):1205-1214
- Grattapaglia D, Bradshaw HD, Jr. (1994) Nuclear DNA content of commercially important Eucalyptus species and hybrids. *Canadian Journal of Forest Research* 24 (5):1074-1078
- Grattapaglia D, Kirst M (2008) Eucalyptus applied genomics: from gene sequences to breeding tools. *New Phytologist* 179 (4):911-929

- Grattapaglia D, Plomion C, Kirst M, Sederoff RR (2009) Genomics of growth traits in forest trees. *Current Opinion in Plant Biology* 12 (2):148-156
- Grattapaglia D, Resende MDV (2011) Genomic selection in forest tree breeding. *Tree Genetics & Genomes* 7 (2):241-255
- Groover A, Devey M, Fiddler T, Lee J, Megraw R, Mitchel-Olds T, Sherman B, Vujcic S, Williams C, Neale D (1994) Identification of Quantitative Trait Loci Influencing Wood Specific Gravity in an Outbred Pedigree of Loblolly Pine. *Genetics* 138 (4):1293-1300
- Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177 (4):2389-2397
- Hagenblad J, Nordborg M (2002) Sequence variation and haplotype structure surrounding the flowering time locus FRI in *Arabidopsis thaliana*. *Genetics* 161 (1):289-298
- Haley CS, Visscher PM (1998) Strategies to utilize marker-quantitative trait loci associations. *Journal of Dairy Science* 81:85-97
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* 92 (2):433-443
- Heffner E, Sorrells M, Jannink J (2009) Genomic Selection for Crop Improvement. *Crop Science* 49 (1):1
- Heuertz M, De Paoli E, Kallman T, Larsson H, Jurman I, Morgante M, Lascoux M, Gyllenstrand N (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics* 174 (4):2095-2105
- Holliday JA, Ritland K, Aitken SN (2010) Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist* 188 (2):501-514
- Ingvarsson PK (2005) Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European Aspen (*Populus tremula* L., Salicaceae). *Genetics* 169 (2):945-953
- Ingvarsson PK (2008) Multilocus Patterns of Nucleotide Polymorphism and the Demographic History of *Populus tremula*. *Genetics* 180 (1):329-340
- Ingvarsson PK, Garcia MV, Luquez V, Hall D, Jansson S (2008) Nucleotide polymorphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (*Populus tremula*, Salicaceae). *Genetics* 178 (4):2217-2226
- Iwata H, Hayashi T, Tsumura Y (2011) Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*. *Tree Genetics & Genomes* 7 (4):747-758
- Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics* 9 (2):166-177
- Johnson GCL, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RCJ, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SCL, Clayton DG, Todd JA (2001) Haplotype

- tagging for the identification of common disease genes. *Nat Genet* 29 (2):233-237
- Kado T, Matsumoto A, Ujino-Ihara T, Tsumura Y (2008) Amounts and patterns of nucleotide variation within and between two Japanese conifers, sugi (*Cryptomeria japonica*) and hinoki (*Chamaecyparis obtusa*) (*Cupressaceae sensu lato*). *Tree Genetics & Genomes* 4 (1):133-141
- Kado T, Ushio Y, Yoshimaru H, Tsumura Y, Tachida H (2006) Contrasting patterns of DNA variation in natural populations of two related conifers, *Cryptomeria japonica* and *Taxodium distichum* (*Cupressaceae sensu lato*). *Genes & Genetic Systems* 81 (2):103-113
- Kado T, Yoshimaru H, Tsumura Y, Tachida H (2003) DNA Variation in a Conifer, *Cryptomeria japonica* (*Cupressaceae sensu lato*). *Genetics* 164 (4):1547-1559
- Kidd KK, Pakstis AJ, Speed WC, Kidd JR (2004) Understanding human DNA sequence variation. *Journal of Heredity* 95 (5):406-420
- Koller DL, Peacock M, Lai D, Foroud T, Econs MJ (2004) False Positive Rates in Association Studies as a Function of Degree of Stratification. *Journal of bone and mineral research* 19 (8):1291-1295
- Krutovsky KV, Neale DB (2005) Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir. *Genetics* 171 (4):2029-2041
- Kulheim C, Yeoh SH, Wallis IR, Laffan S, Moran GF, Foley WJ (2011) The molecular basis of quantitative variation in foliar secondary metabolites in *Eucalyptus globulus*. *New Phytologist* 191 (4):1041-1053
- Lan MY, Chang YY, Chen WH, Kao YF, Lin HS, Liu JS (2009) Serotonin transporter gene promoter polymorphism is associated with body mass index and obesity in non-elderly stroke patients. *Journal of Endocrinological Investigation* 32 (2):119-122
- Lander ES, Schork NJ (1994) Genetic Dissection of Complex Traits. *Science* 265:2037-2048
- Lepoittevin C, Harvengt L, Plomion C, Garnier-Géré P (2011) Association mapping for growth, straightness and wood chemistry traits in the *Pinus pinaster* Aquitaine breeding population. *Tree Genetics & Genomes* 8 (1):113-126
- Lin PI, Vance JM, Pericak-Vance MA, Martin ER (2007) No gene is an island: The flip-flop phenomenon. *American Journal of Human Genetics* 80 (3):531-538
- Long AD, Langley CH (1999) The Power of Association Studies to Detect the Contribution of Candidate Genetic Loci to Variation in Complex Traits. *Genome Research* 9:720-731
- Long AD, Lyman RF, Langley CH, Mackay TFC (1998) Two Sites in the Delta Gene Region Contribute to Naturally Occurring Variation in Bristle Number in *Drosophila melanogaster*. *Genetics* 149 (2):999-1017

- Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoretical and Applied Genetics* 120 (1):151-161
- Luan T, Woolliams JA, Lien S, Kent M, Svendsen M, Meuwissen THE (2009) The Accuracy of Genomic Selection in Norwegian Red Cattle Assessed by Cross-Validation. *Genetics* 183 (3):1119-1126
- Lyman RF, Lai C, Mackay TFC (1999) Linkage disequilibrium mapping of molecular polymorphisms at the scabrous locus associated with naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics Research* 74 (03):303-311
- Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA
- Ma X-F, Hall D, St. Onge KR, Jansson S, Ingvarsson PK (2010) Genetic Differentiation, Clinal Variation and Phenotypic Associations With Growth Cessation Across the *Populus tremula* Photoperiodic Pathway. *Genetics* 186 (3):1033-1044
- Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C (2011) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity*:1-7
- Markussen T, Fladung M, Achere V, Favre JM, Faivre-Rampant P, Aragonés A, Perez DD, Harvengt L, Espinel S, Ritter E (2003) Identification of QTLs controlling growth, chemical and physical wood property traits in *Pinus pinaster* (Ait.). *Silvae Genetica* 52 (1):8-15
- Marques CM, Brondani RPV, Grattapaglia D, Sederoff R (2002) Conservation and synteny of SSR loci and QTLs for vegetative propagation in four Eucalyptus species. *Theor Appl Genet* 105:474-478
- Marquet S, Doumbo O, Cabantous S, Poudiougou B, Argiro L, Safeukui I, Konate S, Sissoko S, Chevereau E, Traore A, Keita MM, Chevallard C, Abel L, Dessein AJ (2008) A functional promoter variant in IL12B predisposes to cerebral malaria. *Human Molecular Genetics* 17 (14):2190-2195
- Mauricio R (2001) Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nat Rev Genet* 2 (5):370-381
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157 (4):1819-1829
- Nadeau JH, Frankel WN (2000) The roads from phenotypic variation to gene discovery: mutagenesis versus QTLs. *Nature Genetics* 25 (4):381-384
- Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. *Nat Rev Genet* 12 (2):111-122
- Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends in Plant Science* 9 (7):325-330
- Neale DB, Sewell MM, Brown GR (2002) Molecular dissection of the quantitative inheritance of wood property traits in loblolly pine. *Annals of Forest Science* 59 (5-6):595-605

- Newcombe G, Bradshaw HD, Chastagner GA, Stettler RF (1996) A major gene for resistance to *Melampsora medusae* f sp *deltoidae* in a hybrid poplar pedigree. *Phytopathology* 86 (1):87-94
- Nordborg M, Tavaré S (2002) Linkage disequilibrium: what history has to tell us. *Trends in Genetics* 18 (2):83-90
- Olson MS, Robertson AL, Takebayashi N, Silim S, Schroeder WR, Tiffin P (2010) Nucleotide diversity and linkage disequilibrium in balsam poplar (*Populus balsamifera*). *New Phytologist* 186 (2):526-536
- Paux E, Tamasloukht M, Ladouce N, Sivadon P, Grima-Pettenati J (2004) Identification of genes preferentially expressed during wood formation in Eucalyptus. *Plant Molecular Biology* 55 (2):263-280
- 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 (3):273-284
- Petukhova L, Duvic M, Hordinsky M, Norris D, Price V, Shimomura Y, Kim H, Singh P, Lee A, Chen WV, Meyer KC, Paus R, Jahoda CAB, Amos CI, Gregersen PK, Christiano AM (2010) Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. *Nature* 466 (7302):113-117
- Pot D, McMillan L, Echt C, Le Provost G, Garnier-Gere P, Cato S, Plomion C (2005) Nucleotide variation in genes involved in wood formation in two pine species. *New Phytologist* 167 (1):101-112
- Pritchard JK, Stephens M, Donnelly P (2000a) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155:945–959
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association Mapping in Structured Populations. *Am J Hum Genet* 67:170-181
- Pyhajarvi T, Garcia-Gil MR, Knurr T, Mikkonen M, Wachowiak W, Savolainen O (2007) Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics* 177 (3):1713-1724
- Pyhajarvi T, Kujala ST, Savolainen O (2011) Revisiting protein heterozygosity in plants-nucleotide diversity in allozyme coding genes of conifer *Pinus sylvestris*. *Tree Genetics & Genomes* 7 (2):385-397
- Qiu D, Wilson IW, Gan S, Washusen R, Moran GF, Southerton SG (2008) Gene expression in Eucalyptus branch wood with marked variation in cellulose microfibril orientation and lacking G-layers. *New Phytologist* 179 (1):94-103
- Quesada T, Gopal V, Cumbie WP, Eckert AJ, Wegrzyn JL, Neale DB, Goldfarb B, Huber DA, Casella G, Davis JM (2010) Association Mapping of Quantitative Disease Resistance in a Natural Population of Loblolly Pine (*Pinus taeda* L.). *Genetics* 186 (2):677-686
- Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics* 20 (2):103-111
- Rengel D, San Clemente H, Servant F, Ladouce N, Paux E, Wincker P, Couloux A, Sivadon P, Grima-Pettenati J (2009) A new genomic

- resource dedicated to wood formation in *Eucalyptus*. *Bmc Plant Biology* 9:36
- Resende MDV, Resende MFR, Sansaloni CP, Petroli CD, Missiaggia AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria DA, Pappas GJ, Kilian A, Grattapaglia D (2012a) Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytologist* 194 (1):116-128
- Resende MFR, Munoz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, Resende MDV, Kirst M (2012b) Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist* 193 (3):617-624
- Resende MFR, Muñoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF, Kirst M (2012c) Accuracy of Genomic Selection Methods in a Standard Dataset of Loblolly Pine (*Pinus taeda* L.). *Genetics* 190 (4):1503-1510
- Rocha RB, Barros EG, Cruz CD, Rosado AM, Araujo EFd (2007) Mapping of QTLs related with wood quality and developmental characteristics in hybrids (*Eucalyptus grandis* X *Eucalyptus urophylla*). *Revista Arvore* 31 (1):13-24
- Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics* 123 (4):218-223
- Scotti-Saintagne C, Bertocchi E, Barreneche T, Kremer A, Plomion C (2005) Quantitative trait loci mapping for vegetative propagation in pedunculate oak. *Annals of Forest Science* 62 (4):369-374
- Sewell MM, Neale DB (2000) Mapping quantitative traits in forest trees. In: Jain SM, Minocha SC (eds) *Molecular biology of woody plants*, vol 1. Kluwer Academic Publishers, The Netherlands, pp 407-423
- Sexton TR, Henry RJ, McManus LJ, Henson M, Thomas DS, Shepherd M (2010) Genetic association studies in *Eucalyptus pilularis* Smith (blackbutt). *Australian Forestry* 73 (4):254-258
- Shepherd M, Huang SW, Eggler P, Cross M, Dale G, Dieters M, Henry R (2006) Congruence in QTL for adventitious rooting in *Pinus elliottii* X *Pinus caribaea* hybrids resolves between and within-species effects. *Molecular Breeding* 18 (1):11-28
- Shepherd M, Kasem S, Lee DJ, Henry R (2008) Mapping species differences for adventitious rooting in a *Corymbia torelliana* x *Corymbia citriodora* subspecies variegata hybrid. *Tree Genetics & Genomes* 4 (4):715-725
- Simko I, Haynes KG, Jones RW (2006) Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. *Genetics* 173 (4):2237-2245
- Solberg T, Sonesson A, Woolliams J, Meuwissen T (2009) Reducing dimensionality for prediction of genome-wide breeding values. *Genetics Selection Evolution* 41 (1):29
- Southerton SG, MacMillan CP, Bell JC, Bhuiyan N, Downes G, Ravenwood IC, Joyce KR, Williams D, Thumma BR (2010) Association of allelic variation in xylem genes with wood properties in *Eucalyptus nitens*. *Australian Forestry* 73 (4):259-264

- Stelkens R, Seehausen O (2009) Genetic distance between species predicts novel trait expression in their hybrids. *Evolution* 63 (4):884-897
- Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics* 3 (5):391-396
- Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp *mays* L.). *Proceedings of the National Academy of Sciences of the United States of America* 98 (16):9161-9166
- Terwilliger JD, Weiss KM (1998) Linkage disequilibrium mapping of complex disease: fantasy or reality? *Current Opinion in Biotechnology* 9 (6):578-594
- Thavamanikumar S, McManus LJ, Tibbits JFG, Bossinger G (2011) The significance of single nucleotide polymorphisms (SNPs) in *Eucalyptus globulus* breeding programs. *Australian Forestry* 74 (1):23-29
- Thumma B, Sharma N, Southerton S (2012) Transcriptome sequencing of *Eucalyptus camaldulensis* seedlings subjected to water stress reveals functional single nucleotide polymorphisms and genes under selection. *Bmc Genomics* 13 (1):364
- Thumma BR, Baltunis BS, Bell JC, Emebiri LC, Moran GF, Southerton SG (2010a) Quantitative trait locus (QTL) analysis of growth and vegetative propagation traits in *Eucalyptus nitens* full-sib families. *Tree Genetics & Genomes* 6 (6):877-889
- Thumma BR, Matheson BA, Zhang D, Meeske C, Meder R, Downes GM, Southerton SG (2009) Identification of a Cis-Acting Regulatory Polymorphism in a Eucalypt COBRA-Like Gene Affecting Cellulose Content. *Genetics* 183 (3):1153-1164
- Thumma BR, Nolan MR, Evans R, Moran GF (2005) Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* 171 (3):1257-1265
- Thumma BR, Southerton SG, Bell JC, Owen JV, Henery ML, Moran GF (2010b) Quantitative trait locus (QTL) analysis of wood quality traits in *Eucalyptus nitens*. *Tree Genetics & Genomes* 6 (2):305-317
- Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M, Epple P, Kuhns C, Sureshkumar S, Schwartz C, Lanz C, Laitinen RAE, Huang Y, Chory J, Lipka V, Borevitz JO, Dangl JL, Bergelson J, Nordborg M, Weigel D (2010) Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* 465 (7298):632-636
- Verhaegen D, Plomion C, Gion JM, Poitel M, Costa P, Kremer A (1997) Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers .1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theoretical and Applied Genetics* 95 (4):597-608
- Wachowiak W, Balk PA, Savolainen O (2009) Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-

- related candidate genes in Scots pine (*Pinus sylvestris* L.). *Tree Genetics & Genomes* 5 (1):117-132
- Whetten R, Sun Y-H, Zhang Y, Sederoff R (2001) Functional genomics and cell wall biosynthesis in loblolly pine. *Plant Molecular Biology* 47 (1):275-291
- White TL, Adams WT, Neale DB (2007) *Forest Genetics*. CABI Publishing, Wallingford, England
- Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42 (7):565-569
- Yu JM, Buckler ES (2006) Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology* 17 (2):155-160
- Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38 (2):203-208
- Zhang Z, Liu J, Ding X, Bijma P, de Koning D-J, Zhang Q (2010) Best Linear Unbiased Prediction of Genomic Breeding Values Using a Trait-Specific Marker-Derived Relationship Matrix. *PLoS ONE* 5 (9):e12648
- Zhao KY, Aranzana MJ, Kim S, Lister C, Shindo C, Tang CL, Toomajian C, Zheng HG, Dean C, Marjoram P, Nordborg M (2007) An Arabidopsis example of association mapping in structured samples. *Plos Genetics* 3 (1):e4
- Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. *Genetics* 163 (3):1123-1134

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Thavamanikumar, S;Southerton, SG;Bossinger, G;Thumma, BR

Title:

Dissection of complex traits in forest trees - opportunities for marker-assisted selection

Date:

2013-06-01

Citation:

Thavamanikumar, S., Southerton, S. G., Bossinger, G. & Thumma, B. R. (2013). Dissection of complex traits in forest trees - opportunities for marker-assisted selection. TREE GENETICS & GENOMES, 9 (3), pp.627-639. <https://doi.org/10.1007/s11295-013-0594-z>.

Persistent Link:

<http://hdl.handle.net/11343/283141>