1 Association mapping for wood quality and growth traits in *Eucalyptus globulus* ssp.
2 *globulus* Labill identifies nine stable marker-trait associations for seven traits
3 Saravanan Thavamanikumar^{1,2§}, Luke J. McManus^{1,2}, Peter K. Ades^{1,2}, Gerd
4 Bossinger^{1,2}, Desmond J. Stackpole^{2,3,§}, Richard Kerr⁴, Sara Hadjigol^{2,3}, Jules S.
5 Freeman^{2,3,5}, René E. Vaillancourt^{2,3}, Peng Zhu⁶ and Josquin F.G. Tibbits^{1,2,##}

6 ¹Department of Forest and Ecosystem Science, The University of Melbourne, Water Street,
7 Creswick, Victoria 3363, Australia

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

10 ³School of Plant Science, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001,
11 Australia

12 ⁴Plant Plan Genetics Pty Ltd. PO Box 1811, Mount Gambier, South Australia, 5290,
13 Australia

14 ⁵CRN Research Fellow, Faculty of Science, Health, Education and Engineering, University
15 of the Sunshine Coast, Locked Bag 4, Maroochydore, Queensland 4558, Australia

16 ⁵South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, Guangdong
17 ⁶10650, People Republic of China

18 [§]Present address: CSIRO Agriculture Flagship, GPO Box 1600, Acton, ACT 2601, Australia

19 [#]Present address: Department of Primary Industries, Biosciences Research Division, Victorian
20 AgriBiosciences Centre, 1 Park Drive, La Trobe University research and Development Park,
21 Bundoora, Victoria 3083, Australia

22 [§]Present address: PT Riau Andalan Pulp and Paper, Pangkalan Kerinci, Riau province,
23 Indonesia 28300

24
25 Author for Correspondence:

26 Josquin F. G. Tibbits

27 Tel: +61 3 90327055

28 Email: Josquin.Tibbits@dpi.vic.gov.au
29
30
31

ABSTRACT

The moderate to high levels of nucleotide diversity and low linkage disequilibrium found in many forest tree species make them ideal candidates for association mapping. Here we report candidate gene-based association mapping results for complex wood quality and growth traits in *Eucalyptus globulus* Labill. ssp. *globulus*, the most widely grown eucalypt in temperate regions of the world. Ninety-eight single nucleotide polymorphisms (SNPs) from 20 wood quality candidate genes were assayed in a discovery population consisting of 385 trees sourced from a provenance-progeny trial. Twenty-five selected SNPs with significant associations ($P < 0.05$) in the discovery population were assayed for validation in 296 trees sourced from an independent 2nd generation breeding trial. To account for background genetic structure, mixed models were used in the association analyses. Two associations identified in the discovery population were independently supported in the validation testing. However, combining the discovery and validation results in a combined-analysis we discovered nine stable marker-trait associations for seven traits. These associations link underlying complex wood and growth phenotypes to earlier putative selection signatures opening new avenues to accelerate the dissection of these traits.

Key Words: *Eucalyptus globulus*, association mapping, wood quality, candidate genes, SNP.

INTRODUCTION

Identifying the mutations that underlie variation in complex phenotypic traits is one of the fundamental aims of molecular genetics. Association mapping provides statistical evidence for associating a marker polymorphism with a phenotypic target trait. When other confounding factors are removed association mapping can identify causal mutations where these are genotyped, or use linkage disequilibrium (LD) to associate the phenotypic trait variation with sequence polymorphism (Lander and Schork 1994). Association mapping can be undertaken either at a candidate gene level or at the genome-wide level. The candidate gene approach is attractive for complex trait dissection in non-model species where high density genotyping assays or whole genome level assays are still in the early stages of development. The candidate gene approach has been used in a number of forest tree species with the identification of several quantitative trait nucleotides (QTN) that associate with a wide variety of traits including the wood property traits, early wood specific gravity and percentage latewood (Gonzalez-Martinez et al. 2007), percentage earlywood, tracheid cell wall thickness and average ring width (Beaulieu et al. 2011), microfibril angle (Thumma et al. 2005), cellulose content (Thumma et al. 2009), density (Dillon et al. 2010), growth traits such as height and girth (Lepoittevin et al. 2012) and a range of other traits including carbon isotope discrimination (Gonzalez-Martinez et al. 2008), flowering traits (Ingvarsson et al. 2008), growth cessation (Ma et al. 2010), disease resistance (Quesada et al. 2010) and cold-hardiness related traits (Eckert et al. 2009; Holliday et al. 2010). In most of these studies the species or populations studied were unstructured and only in a few cases have associations been validated in independent genetic material (Dillon et al. 2010; Thumma et al. 2010; Thumma et al. 2009; Thumma et al. 2005).

A main driver in determining mapping resolution (i.e. the ability to identify actual causative sites as opposed to markers in LD with the causative site) is the rate and pattern of decay of

LD. All outcrossing tree species studied so far have shown a rapid decay in LD (Brown et al. 2004; Garcia-Gil et al. 2003; Gonzalez-Martinez et al. 2006; Ingvarsson 2005; Krutovsky and Neale 2005; Thavamanikumar et al. 2011; Thumma et al. 2005) and association mapping in such species will have high resolution making the identification of causative sites possible (Gaut and Long 2003). At the same time, this rapid decay in LD is also a concern because even physically proximal markers can be completely unlinked, and therefore offer no predictive power of a QTN that may in fact reside physically close.

Association mapping is population-based, and as such can suffer from population structure confounding which can significantly increase the rate of false positives (Balding 2006; Lander and Schork 1994). Such spurious associations can be largely avoided by either performing association studies in homogenous populations (Lander and Schork 1994; Pritchard et al. 2000b) or by modelling the structure (Pritchard et al. 2000a). Likewise, kinship structure can also increase false discovery rates and modelling techniques that account for relatedness, including cryptic relatedness, have been developed and their use is generally advocated (Aste and Balding 2009; Yu et al. 2006; Zhao et al. 2007b).

Here we use a candidate-gene based association mapping approach with the aim of identifying markers associated with complex wood quality and growth traits in *Eucalyptus globulus* ssp. *globulus*. *E. globulus* ssp. *globulus* (hereafter referred to as *E. globulus*) is the most widely utilised subspecies of the species complex and in its natural range in south eastern Australia it occurs in coastal western, eastern and south eastern Tasmania, on King and Flinders Islands in Bass Strait, and in Victoria in the Otway Ranges, Wilsons Promontory and areas immediately north towards the Strzelecki Ranges. The international estate is underpinned by a number of breeding programs, some of which utilize advanced quantitative methods to achieve gains (McRae et al. 2004). In Australia, the breeding cycle is now moving only into its third generation from the wild (Jones et al. 2006), highlighting the enormous

gain that is still to be captured from breeding. Despite the demonstrated gains achieved through quantitative breeding, gain is expected to be significantly accelerated with the adoption of molecular marker assisted selection (MAS) (Grattapaglia and Kirst 2008) and/or the more recent development of genomic selection (GS) (Meuwissen et al. 2001; Resende et al. 2012a; Resende et al. 2012b; Resende et al. 2012c). MAS and GS can accelerate gain through a number of mechanisms including: 1) increasing selection intensity, and 2) by overcoming temporal impediments such as age to trait expression, enabling earlier capture and utilisation of elite germplasm. While GS does not require the identification of causative mutations, doing so will provide markers which retain marker/trait correlation making them more likely transferrable across populations and informative of the biology behind trait expression (Thavamanikumar et al. 2013).

In this study, 20 wood quality candidate genes were selected, based on their known functional roles in cellulose and lignin biochemical pathways and/or on mutation and expression studies in other organisms and/or based on putative selection signatures (Coleman et al. 2009; Hu et al. 1999; Kawaoka et al. 2006; Patzlaff et al. 2003; Qiu et al. 2008; Schindelman et al. 2001; Spokevicius et al. 2007; Szyjanowicz et al. 2004; Thavamanikumar et al. 2014; Zuo et al. 2000). Prior work identified a large number of polymorphisms in this set of candidate genes (Thavamanikumar et al. 2011) and a subset of these was selected for inclusion in this association mapping study. Association testing was undertaken with raw data for wood quality and growth traits and estimated breeding values (EBVs) for Volume, Density and Pulp Yield at harvest age. As significant structure has previously been reported between races of *E. globulus* (Steane et al. 2006; Steane et al. 2011) we followed general recommendations and fitted models to account for population structure and familial relatedness (Yu et al. 2006).

MATERIALS AND METHODS

Plant material and phenotypic evaluation

A Gunns Ltd *Eucalyptus globulus* provenance-progeny trial, planted in 1989 near Latrobe in Tasmania, Australia, was used as the association discovery population (hereafter referred to as the ‘discovery population’). Samples from this trial were also used for polymorphism discovery (Thavamanikumar et al. 2011). The trial was established with 570 open-pollinated (i.e. half-sib) *E. globulus* families collected from wild mother trees, with each family represented by a two-tree plot, in each of five replicates. To avoid spurious association arising from familial relatedness we sampled 385 trees from this trial (Figure 1) selecting only one tree per family. Cambial scrapings were collected and DNA extracted using the methods described in Tibbits *et al.*, (2006).

The validation population was a sample of 296 trees from an eight-year-old Southern Tree Breeding Association (STBA) second generation breeding trial growing near Frankland, Western Australia (hereafter referred to as the ‘validation population’). This trial contained 115 *E. globulus* families generated from controlled pollination crosses from 46 paternal and 45 maternal parents. Overlap between the trials was limited with only three fathers (pollen parents) in the Frankland trial originating from the Latrobe trial. None of these three fathers was sampled in the discovery population making the populations sampled from Latrobe and Frankland completely independent. This population was specifically chosen for validation to contrast the expected sources of bias with that of the discovery population. The validation population is expected to show kinship between individuals but little effect of population structure due to the mating of parents from different population backgrounds having been used to create the offspring tested whereas the discovery population is expected to have population structure but little to no kinship. For the validation population we sampled leaves

from the 296 trees and DNA was extracted using a modified CTAB protocol based on Doyle and Doyle (1987).

For most traits both deregressed EBVs (DEBVs) and entry mean adjusted phenotypic estimates were used in association analysis. This approach was taken as each provides a different degree of accuracy for the additive, non-additive and error components of variation. While presented separately we do not aim to treat these estimates as separate traits but rather look for consistency between them. The phenotypic evaluation of trees in the discovery population was undertaken as part of a separate study on the quantitative genetics of *E. globulus* at the University of Tasmania, Australia (Stackpole et al. 2010a,b; 2011). All trees within the validation trial were cored in 2004 (at the age of four) and these cores were used for the estimation of wood basic density, and for NIR-predictions of cellulose content, pulp yield, extractive content and klason lignin content using the same methods as for the discovery population. Consistent with a 2 step approach adjusted entry means were used in place of raw phenotypic data for both the discovery and validation analyses (Lepoittevin et al. 2012; Stich et al. 2008) with field design (environmental) effects removed using SAS version 9.3. The raw data were provided by the Southern Tree Breeding Association (STBA). For both populations DEBVs for harvest age volume, wood basic density and pulp yield were also provided by the STBA. These breeding values were predicted in the multi-generation, multi-trait and multi-site genetic evaluation of all germplasm in the Australian National *E. globulus* Breeding Program using the TREEPLAN® software (McRae et al. 2004) with no base specified, and deregression carried out prior to association model fitting ($DEBV = EBV/r^2$, where r^2 is the reliability of the EBV) following the methods of Garrick *et al.*, (2009). A summary description of the phenotypic trait set used in this study is provided in Table 1.

SNP genotyping

Twenty functional candidate genes, for wood and fibre formation, were selected on the basis of mutation and expression studies (Coleman et al. 2009; Hu et al. 1999; Kawaoka et al. 2006; Patzlaff et al. 2003; Qiu et al. 2008; Schindelman et al. 2001; Spokevicius et al. 2007; Szyjanowicz et al. 2004; Thavamanikumar et al. 2014; Zuo et al. 2000), nucleotide diversity patterns potentially departing from neutral molecular evolution (Eveno et al. 2008; Thavamanikumar et al., unpublished; Gonzalez-Martinez et al. 2006) and association studies (Dillon et al. 2010; Thumma et al. 2009) in other plant and tree species.

Over one thousand SNP within the twenty genes used in this study, were discovered previously by direct sequencing of PCR products from a SNP discovery panel of 11 to 28 trees (Thavamanikumar et al. 2011). From these, 151 were selected based on position, potential function, minor allele frequency and low LD with other SNPs. Several of these SNPs also showed potential selection signatures in genetic differentiation-based outlier tests (S. Thavamanikumar *et al.*, unpublished).

The iPLEX Gold assay (Sequenom Inc.), based on matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS; Bouakaze et al. 2011), was used to genotype SNPs in 385 individuals from the discovery population. Genotyping was performed at the Southern Cross Plant Genomics facility (Lismore, Australia). Details of the genotyping protocol and a complete primer list are given in Appendix A1. A subset of SNPs (25) producing significant associations in the discovery population was genotyped in 296 individuals in the validation population. Hardy-Weinberg equilibrium (HWE) in the discovery population, across four population clusters identified via STRUCTURE (Pritchard et al. 2000a) analysis for individual SNPs, were estimated using FSTAT version 2.9.3.2 (Goudet 2002). Linkage disequilibrium between SNPs was estimated from unphased SNP genotypic data using GEVALT (Davidovich et al. 2007) both (a) across all samples ignoring

population structure and (b) on each of the four populations separately as identified by STRUCTURE.

Association analysis

Quantitative and molecular studies in *E. globulus* have shown moderate genetic structure exists between races and populations (Dutkowski and Potts 1999; Steane et al. 2006). To account for this structure, a mixed linear model (MLM) (Yu et al. 2006) fitting terms for both population structure and familial relatedness was fit using TASSEL version 2.0.1 (Bradbury et al. 2007). Best Linear Unbiased Estimates (BLUE) of SNP effects against all traits were obtained from mixed models in TASSEL. For the discovery population, general co-ancestry (Q) was predicted using the model-based clustering method implemented in STRUCTURE v2.3.1 (Falush et al. 2003; Falush et al. 2007; Pritchard et al. 2000a) with data from 18 microsatellites to estimate Q matrix (see Appendix A2). Assuming no prior population groupings and using the admixture model, the number of genetically homogeneous groups (K) was chosen by comparing the log probability of data at different values of K (from K = 1 to K = 10), using 100,000 MCMC repetitions following a burn-in of 60,000 repetitions. For the association analysis, we used Q data derived from STRUCTURE runs with K=4 as it is biologically sensible, most consistent with other marker based studies of population boundaries in *E. globulus* (Yeoh et al. 2012) and is one of the recommended solutions using the ΔK method of Evanno et al. (2005). For the discovery population, we utilised data from 18 microsatellites and 98 SNPs to derive a matrix of realised pairwise kinship and inbreeding coefficients (K) using the software SpaGeDi (Hardy and Vekemans 2002) as described in Ritland (1996). For the validation population, pedigree information was used to generate a predicted kinship matrix using the R-package kinship (<http://www.r-project.org/>).

To test the effect of population structure and kinship in the association results, association analysis were also conducted without including Q or K. In the discovery population, three models were tested: a) GLM (no Q or K), referred as “GLM”, b) GLM (with Q), referred as “GLM (Q)” and MLM (with Q and K), referred as “MLM (Q+K)”. In the validation population, two models were tested: a) GLM (no Q or K), referred as “GLM” and b) MLM (with K), referred as “MLM (K)”. As the MLM model fits SNP independently the estimates of effects can be biased upward (Kemper et al. 2012). So to improve estimates of SNP effects conditional on the effect of all other SNPs, all the SNPs were fitted simultaneously as random effects using ridge regression as implemented in TASSEL version 3.0. These Best Linear Unbiased Predictors (BLUP) of SNP effects are presented along with the BLUE estimates. Readers should be cautioned that the small sample sizes in the populations used in this study will make these effect size estimates uncertain and a priority of future research will be to re-estimate these in subsequent larger studies.

Combined analysis

A combined analysis of marker-trait associations was performed using Stouffer’s method (Stouffer et al. 1949) implemented in COMPARE2 (Abramson 2004). P-values from association analyses obtained from discovery and validation population were combined using the “Meta-analysis” module of COMPARE2. COMPARE2 assumes that P-values are obtained from independent tests of the same hypothesis. In the combined analysis, we tested 21 SNPs that had data in both the discovery and validation populations and where allelic effects were in the same direction in both populations. The inverse of the squared standard error of the effect size of each population was included in the analysis as a weighting factor. To account for multiple testing, Q-values (False Discovery Rate, FDR) (Storey and Tibshirani 2003) were calculated in R (<http://www.r-project.org/>).

244 **Tests on KOR_2328 SNP**

245 A complicated nonsynonymous substitution pattern is encoded by the adjoining SNP loci
246 KOR_2328 and KOR_2329 at the 263rd codon of the Korrigan gene. To gain insight into a
247 possible function we investigated amino acid substitution tolerances across the Korrigan gene
248 using SIFT ("Sorting Intolerant From Tolerant"; Ng and Henikoff 2003). The analysis was
249 based upon a multiple sequence alignment of 43 endo-1,4- β -D-glucanase (Korrigan)
250 sequences from different plant species, collected through PSI-BLAST. As we could not
251 genotype the second SNP (KOR_2329), due to incompatibility with the genotyping assay, we
252 estimated amino acid frequencies at the 263rd position in the population samples from
253 pairwise frequencies observed in phased Sanger DNA sequencing data across 28 individuals
254 (S. Thavamanikumar *et al.*, unpublished).

255 **RESULTS**

256 **SNP Genotyping in the discovery population**

257 In a trial run to test the SNP assays 151 SNPs were genotyped across 32 individuals using
258 MASS ARRAY. Amongst this SNP set a small number, especially those in close proximity
259 to other SNPs, showed assay problems in multiplexing leaving accurate genotyping assays
260 for 123 of the 151 polymorphisms. These 123 were subsequently genotyped across a further
261 353 individuals from the discovery population. Of the 123 SNPs assayed, 98 SNP from 20
262 genes (on an average 5 SNPs per gene) which had at least 80% data were used in the
263 discovery population association analysis (Table 2). At a Bonferroni-corrected threshold P-
264 value of 0.0005, none of the SNPs deviated from Hardy Weinberg Equilibrium expectations
265 when tested within the four population clusters.

266 **Linkage disequilibrium between SNPs in discovery population is low**

Association tests of SNP loci showing LD are non-independent. To confirm that selected SNP are predominantly independent in the discovery population we estimated LD between SNP in the discovery population and found, as expected, it to be very low when estimated across all samples. Ignoring population clustering, only 1.7% of the pairwise comparisons had r^2 values greater than 0.33 (Supplementary Table 1). More than two thirds of the higher estimates arise from comparisons within 4CL, consistent with the pattern found during SNP discovery (Thavamanikumar 2010). Only one SNP pair from different genes, CSA3_4186 / MYB2_1380, was found to have an $r^2 > 0.5$. Similar results were obtained when LD was computed within population clusters with only 1.17 to 1.85% of pairwise comparisons having r^2 values greater than 0.33. The SNP pair CSA3_4186 / MYB2_1380 exhibited strong LD ($r^2 > 0.5$) in three of the four population clusters with r^2 ranging from 0.58 to 0.81. Overall, despite the small physical distance between genotyped SNPs, the low pairwise r^2 values suggest that the tested SNP can be considered as separate factors (Supplementary Table 1).

Comparison of different models used in association analysis

Accounting for population structure and/or kinship will help reduce false positives in association analysis. To investigate these sources of variation, different models including and excluding the Q and K terms were used in association analysis. Model comparisons are based on number of SNPs associated with different traits at $p < 0.05$ (Table 3) and based on deviation of observed P-values from expectation using qq-plots (Figure 2 and 3). Overall, association analysis performed without Q or K yielded the greatest departures from the expected P-value distribution with greater excess of low P-values and therefore a likely higher rate of false positives in both discovery and validation populations, with the greatest departures in the discovery population. This is expected because there is known geographical structure to the genotypic variance present in the population studied for the phenotypes (Dutkowski and Potts 1999). Inclusion of Q and/or K in the models did not make as large an

impact for the growth traits (Diameter at Breast Height (DBH) and Volume_DEBV) and these have been shown to be less affected by population structure (Dutkowski and Potts 1999). On the other hand, accounting for population structure and/or kinship reduced the false positives rate for physical and chemical wood traits. In the discovery population, the significant associations found were similar between GLM (Q) and MLM (Q+K) reflecting the known lack of kinship amongst the trees (Figure 4). As using MLM (Q+K) in the discovery population and MLM (K) in the validation population produced the least departure from expectation in the qq-plots, and therefore were the models likely to have the lowest false positive rates, we present the data only from these two models in the following section.

Discovery population marker-trait associations

The MLM (Q+K) model, which accounts for both population structure, familial relatedness and individual inbreeding, was used to test genetic associations between 98 polymorphisms from 20 wood quality genes and a range of wood quality and growth traits. Results from association analysis performed in discovery population are presented in Supplementary Table 2. Of the 98 SNPs tested with 12 traits (1176 association tests), 39 showed a significant association ($P < 0.05$) with one or more traits. Because individual SNPs often show association with more than one trait (either due to non-independence between traits or pleiotropy), a total of 66 marker-trait associations (5.6% of total number of tests) were detected.

Validation of marker-trait associations in an independent population

SNPs having marker-trait associations with $P < 0.05$ were carried forward into validation assays in an independent population. This relaxed threshold was used as the low sample numbers in discovery meant that statistical power to discover associations was low and the “cost” of a non-validation was also low. Two criteria excluded associations with $P < 0.05$

from testing in the validation population; those that appeared to result from few individuals of a rare genotype ($MAF < 5\%$) and those where a SNP showed incompatibility in the iPLEX Gold assay. Overall, 25 of the 39 SNPs producing significant associations in the discovery population were genotyped in 296 individuals from the validation population and 21 of these, having at least 80% data, were included in the validation analysis using the MLM (K) model (Supplementary Table 3). At $P < 0.05$, 25 associations for 15 SNPs were identified.

Combined-analysis results

The discovery and validation population results were combined using Stouffer's Z method. This approach enables the P-values from the two studies to be combined while accounting for the differences in the standard error of the effect size in each study. This approach assumes all associations are independent and linear. The combined analysis was carried out for a subset of 21 SNPs that had sufficient data in both the discovery and validation populations. Based on the combined analysis, we identified nine significant associations ($Q < 0.1$) involving seven SNPs from six genes associated with seven traits (Table 4). For these nine SNPs, ridge regression was used to estimate SNP effects by fitting all the SNPs simultaneously as random effects. Invariably, for all the SNPs effect estimated by MLM (Q+K) or MLM (K) was always higher than the effect estimated by ridge regression (Supplementary Table 4) suggesting the possibility for overestimation of SNP effects when SNPs are fitted as fixed effects.

Details of most supported SNP/Trait Associations

SNP CDPK_516, an intronic SNP, showed a significant association with the growth trait - Volume_DEBV in both the discovery and validation populations (Combined-analysis $P = 0.003$; $Q = 0.034$). The minor allele (A) for this SNP was associated with low net present value (NPV) for Volume_DEBV in both populations. CDPK_516 is also associated with

Density_DEBV (Table 4). For the CSA3_1101 association with Density, the minor allele A was associated with low density.

A total of four associations had Q-values < 0.1 in the combined-analysis for the chemical wood property traits Pulp-Yield_DEBV, Extractives and Klason lignin. The strongest association was between KOR_2328 and Pulp-Yield_DEBV where the A and G alleles in both populations associated with higher Pulp-Yield_DEBV while T allele genotypes associated with lower Pulp-Yield_DEBV (Figure 5). KOR_2328 is a triallelic non-synonymous exonic (exon 3) SNP with A, G and T alleles producing 6 different genotypes (AA, AG, AT, GT, GG, and TT). KOR_2328 in combination with KOR_2329 code for a complicated non-synonymous substitution pattern (Thavamanikumar 2010) which results in a combination of six amino acids (Figure 6). As we were only able to genotype KOR_2328, the genotypic groups represent mixtures of amino acids. To estimate the relative proportions of each amino acid class within genotypic classes we estimated the frequencies of the amino acids based on a subset of 28 phased KOR_2328/KOR_2329 genotypes from the discovery population (Thavamanikumar 2010). Overall KOR_2328 TT genotypes represent a mixture of approximately 85% Phe:Phe while GG genotypes represent approximately 56% Val:Val. The genotypic classes AG and AA largely represent a dosage response for Threonine and the genotypic class AT represents a dosage response for Phenylalanine while the GT class is mostly composed of Phe:Val (Figure 6). In both the discovery and validation populations these genotypic classes behave similarly with additive gene action predicted for both substitutions of Thr and Phe. To investigate if amino acid substitutions are tolerated at this codon a SIFT analyses was conducted. SIFT predicts the tolerance levels for amino acid substitution based on the degree of conservation of amino acid residues in alignment from related sequences and is an indirect way of inferring if an amino acid position is likely to confer a phenotypic effect (Kumar et al. 2009). To perform the SIFT analysis, 43 Korrigan

and Korrigan-like sequences were utilised with the results predicting that substitutions of any amino acid would be tolerated at codon 263. Although 10 different amino acids are observed in this position in the 43 homologous sequences searched no sequences coded for Phenylalanine (TTT), while a substantial proportion (12 out of 43 sequences) coded for Threonine.

SuSy3_886 associated with Extractives in both the populations. The minor G allele is associated with higher Extractive content, although the SNP effect appears small. PCBER_1601, an intronic mutation, associates with Klason Lignin variation. This SNP is in strong LD with another SNP within the same gene, PCBER_1714, both in all samples ($r^2 = 0.96$) and within population cluster analysis ($r^2 = 0.91$ to 1) in the discovery population. The minor allele 'C' is associated with lower Klason Lignin.

DISCUSSION

We used an association mapping approach to discover nine statistical associations validated in an independent population, arising from seven SNPs in six candidate genes, with a range of wood quality and growth traits. Consistent with the expectation of quantitative trait architecture, the estimated variance explained individually by our associations was each less than 5%. Overall, more associations were found in genes involved in cellulose biosynthesis than for other traits (Fisher Exact Test: $P = 0.0052$), and these associations were consistently with Density or Pulp Yield traits. As cellulose is a major component of the plant cell wall and the major component of pulp both these traits are correlated with cellulose content at the overall genetic level (Stackpole et al. 2010b). Along with the CDPK_516 association, these “cellulose biosynthesis” gene associations had the most stable results across the discovery and validation populations. The single lignin biosynthesis gene (PCBER) returned a combined-analysis validated association with Klason Lignin content. PCBER plays a role in

phenylpropanoid biosynthesis and localises to secondary xylem parenchyma where it catalyses lignan synthesis (Kwon et al. 2001; Vander Mijnsbrugge et al. 2000). Transgenic down regulation of PCBER has been shown to result in lower lignin content (Boerjan et al. 2006). While based on small numbers, our results indicate that variation in cellulose biosynthesis genes may be a stronger driver of variation in density and pulp yield than variation in lignin biosynthesis genes and therefore, where research resources are limited, future focus in *E. globulus* should preference other catalytic genes from the cellulose biosynthesis pathway over lignin biosynthesis pathway genes.

Among the SNP-trait associations identified, the association between CDPK_516 and Volume_DEBV improved in the validation over the discovery analysis. While a naive analysis, CDPK_516 has an estimated 12% decrease in the AA genotype frequency in the validation (second generation breeding) population compared to the discovery (base) population, with this change consistent in direction with selection for growth through breeding. CDPKs are calcium binding serine/threonine protein kinases (Romeis et al. 2001) with functional roles in plant defence responses (Ivashuta et al. 2005; Romeis et al. 2001) and abiotic stress signalling pathway responses (Li et al. 2008). CDPK gene expression has been shown to directly correlate with growth in *Panax ginseng* (Kiselev et al. 2010) indicating that variation in CDPK may underlie variance in growth traits in many species. Also, CDPK_516 was recently identified as an outlier, based on F_{ST} outlier tests (Hadjigol 2012), indicating a possible adaptive signature. While SNP CDPK_516 is significantly associated with Volume_EBV in both populations it may not be causative as it is in strong LD with SNP CDPK_589.

SNP KOR_2328 was associated with Pulp-Yield_DEBV in both populations (Table 4, Figure 5). Korrigan is a member of the endo-1,4- β -D-glucanase (EGases, EC 3.2.1.4) family and was first identified in an extreme dwarf mutant with distinct architectural alterations in the

primary cell wall (Nicol et al. 1998). It locates to the plasma membrane and most likely acts at the plasma membrane-cell wall interface (Nicol et al. 1998). While it is known that Korrigan is involved in cellulose biosynthesis its exact role is still not clear. In *Pinus pinaster*, the Korrigan gene was shown to co-locate with a wood property QTL (Pot et al. 2006) while high differentiation between Corsican and Aquitaine populations in *P. pinaster* and a significant negative Tajima's D-value in *P. radiata* suggested Korrigan as a potential target of selection in these species (Pot et al. 2005). Our discovery of a validated association of KOR_2328 and Pulp-Yield_DEBV is consistent with the functional role of Korrigan in cellulose biosynthesis and is the first report of an association between allelic variation in Korrigan and a wood property trait in any species.

In *E. globulus*, cellulose content and pulp yield are highly positively correlated with the additive genetic correlation estimated at 0.91 (s.e. 0.02; Stackpole et al. 2010b). KOR_2328 also associated with raw Pulp Yield and raw Cellulose content (Supplementary Figure 1), although the strength of these associations was weaker, possibly due to the increased variance in these raw phenotypes. In a recent phenotypic study in *E. globulus* (Stackpole et al. 2010b; Figure 2), the geographic distribution of Pulp Yield was reported to exhibit broad-scale geographic structure along latitudinal clines (Stackpole et al. 2010b). As this distribution contrasts in some respects to the major population genetic groupings it may be that pulp yield is under selective pressure (Stackpole et al. 2011). Recent work by Hadjigol (2012) identified SNP KOR_2328 as an outlier based on F_{ST} outlier tests. Our association result potentially links this putative selection signature in Korrigan to a trait providing supporting evidence that change in cell wall composition (i.e. cellulose content) is adaptive in *E. globulus*. While many studies have not found any significant trend for outlier loci to show association with adaptive traits (for example, Renaut et al. 2011) our findings suggest that wider application of

genome-scans for selection signatures to identify candidate gene and candidate SNP for QTNs underlying complex wood traits will likely be worthwhile in *E. globulus*.

While selection for cell wall composition remains to be proven, consideration of the functional consequences of the KOR_2328 variants may shed some light on potential drivers. KOR_2328 is a triallelic non-synonymous exonic (exon 3) SNP with A, G and T alleles producing 6 different genotypes. The association results indicate that the Phe encoded allele gives the lowest cellulose content, the Thr allele the highest and other alleles intermediate. Our SIFT analysis indicates that variation at this site in Korrigan is widespread amongst other plant taxa and does not appear strongly deleterious however, Phe was not observed in any other species indicating that reduction in cellulose content may not be widely tolerated. Cellulose is a major component of all higher plant cell walls forming the fibrillar component (Delmer 1999). Variation in cellulose content can confer changes in final cell size and shape (Fagard et al. 2000) as well as in the cell's (and the stem's) ability to withstand compressive or tensile forces (Spokevicius et al. 2007). Cellulose content is known to vary within stems (Yang et al. 2011) and between cells, with an extreme response observed in the development of tension wood cells that are almost entirely cellulose (Qiu et al. 2008). In *E. globulus* tension wood is often characterized by the presence of gelatinous fibres in which the secondary wall is almost entirely cellulose (Washusen 2002). As such, lower cellulose content may confer advantage in plant capacity to tolerate drought through an increased tolerance to cavitation from more rigid cell walls (i.e. higher relative lignin content), or conversely, higher cellulose content may lead to an increased ability to withstand tensile forces from wind and gravity in high growth environments. These scenarios would see Phe alleles favoured in drier/more drought prone environments whereas the Thr alleles would be favoured in wetter, taller forest environments, and this could be directly tested in future work. Other potential drivers are also possible such as in *Arabidopsis*, where the specific activation

463 of novel defence pathways from inhibition of cellulose synthesis and alteration of secondary
464 cell wall integrity could contribute to the generation of an antimicrobial-enriched
465 environment hostile to pathogens (Hernandez-Blanco et al. 2007) and as such the alterations
466 in cellulose content from KOR_2328 may also be defence associated.

467 The SuSy3_886 SNP was significantly associated with Extractive Content. Sucrose
468 Synthase, in the presence of UDP, converts sucrose into UDP-glucose and fructose (Sturm
469 and Tang 1999) supplying carbon to many biosynthetic processes. Strong SuSy activity is
470 observed during wood formation (Schrader and Sauter 2002) and it is thought to be the main
471 enzyme supplying UDP-glucose to cellulose biosynthesis. SuSy activity increases
472 dramatically at the sapwood-heartwood transition zone, where phenolic heartwood
473 extractives are synthesised (Yang et al. 2003). Phenolic heartwood extractives are likely
474 involved in defending the tree stem from wood rotting fungi and bacteria as a significant
475 negative correlation between extractive levels and wood decay was observed in *E. globulus*
476 (Stackpole et al. 2011). The association with extractive content is functionally plausible and
477 potentially highlights an important role for SuSy in controlling carbon supply to extractive
478 biosynthesis.

479 A number of lignin biosynthesis gene and density associations have been reported in the
480 literature (Dillon et al. 2010; Tchou et al. 2011) but, to the best of our knowledge, our SNP
481 CSA3_1101 association is the first report of a cellulose biosynthesis gene association with
482 wood density in any species. Cellulose synthases (CesAs) are localized to plasma membrane
483 and their key function is the production of the β -1,4-glucan biopolymer cellulose (Richmond
484 2000). CesAs have been shown to be expressed in all plant tissue and cell types (Richmond
485 2000), however, different members of the CesA family have been reported to express in a
486 tissue specific manner and there is considerable difference in expression levels of different

CesA family members during primary and secondary cell wall biosynthesis (Kalluri and Joshi 2004; Ranik and Myburg 2006).

If not properly accounted for, background genetic structure can increase the rate of false positive marker-trait associations (Pritchard et al. 2000b). Most forest tree association mapping studies reported to date have been conducted in species that exhibit little population structure (Lepointevin et al. 2012; Sexton et al. 2012; Thumma et al. 2009). Molecular studies in *E. globulus* (Steane et al. 2006; Steane et al. 2011) have consistently found strong latitudinal differentiation between the Victorian and Island (including Tasmania) populations and weaker longitudinal structure between the major population centres. In more recent studies with SNP (Kulheim et al. 2011) and DArT markers (Cappa et al. 2013), the presence and patterns of population structure in *E. globulus* were strongly supported. In this study, we also found these patterns and accounted for these patterns of population differentiation using Q estimates from a STRUCTURE analysis with the prior K set to four ancestral populations. As demonstrated in Cappa et al. (2013) the inclusion of population structure and/or kinship in our association analysis models is highly likely to have reduced the false positive rate. Among the traits studied here, DBH was less influenced by population structure while the wood quality traits Pulp Yield and Cellulose content were more influenced. Similar results were reported by Stackpole et al. (2010b) from a quantitative genetics study in *E. globulus* in which they have found very low levels of subrace-level differentiation for DBH and high differentiation for Pulp Yield, Cellulose and Density. These low levels of differentiation between subraces for DBH observed in both molecular and quantitative studies may partially explain the low heritability of this trait.

While false discovery rate corrections (FDR) can control type I error rates at the experiment level, and are commonly applied (Gonzalez-Martinez et al. 2008; Ingvarsson et al. 2008; Zhao et al. 2007a), they are likely to increase the incidence of false negatives. In non-human

systems the “cost” of false negatives is likely higher than the cost of false positives but in breeding applications, non-linked markers will still carry information that can be utilised in a breeding context. Rather than conservatively correcting for multiple testing, validation through re-testing in an independent population is preferable. Such validation is a standard requirement in both candidate gene based (Bottcher et al. 2009; Traurig et al. 2009) and genome-wide association mapping studies in humans (Jallow et al. 2009; Kathiresan et al. 2008) and permits discrimination of real and false positives without inflating the false negative rate. While preferred, independent replication for validation is often difficult to achieve in practice. In some cases, where allelic effects are very small, as we see in most of the tree association mapping studies, many thousands of individuals may need to be tested (Munafo and Flint 2004). In this study, only two associations identified in the discovery population were independently supported in the validation testing. This low rate of validation should not be interpreted as disproving the associations identified in the discovery population as “winner’s curse” (Zollner and Pritchard 2007) or the “Beavis effect” (Xu 2003) can lead to overestimated effect size in initial discovery (Hirschhorn and Altshuler 2002). Combining results (combined-analysis) from numerous studies provides a means to integrate results across many studies that are unable to be jointly analysed and/or that individually may not show significant association (Munafo and Flint 2004). Combined-analysis provides stronger support for the effect of a SNP on a trait compared to a highly significant result from a single study. Combined-analysis was used in this study to combine results from the discovery and validation populations and nine associations were significant at an FDR of 10% or better. As tree association mapping studies are generally underpowered, combining results from two or more studies appears a sensible approach to increase the power by making use of all available information.

To further investigate our findings, we looked at the co-location between the significant genes and QTL reported in a multi-family study of *E. globulus* (Freeman et al. 2011; Freeman et al. 2013). In summary, each of the genes with validated SNPs was placed in the QTL study and SNPs assayed in parents of each pedigree. In all cases, except KOR_2328 where the variants were not segregating in any pedigree, the genes showing supported associations co-located with QTL for (the same or other) wood property traits. Furthermore, with the exception of PCBER and one of the QTL co-locating with CSA3, each of the best supported associated SNPs were polymorphic (Freeman et al. 2013) in the same family in which the co-located QTL segregated, lending more support that these SNPs capture or underlie QTL. These co-location results strengthen our findings and highlight the continued importance of QTL mapping illustrating how these two approaches are complementary, as the combination of several independent lines of evidence is more confirmative than reliance on a single stringent analysis (Renaut et al. 2011).

Conclusions

This association mapping study on growth and wood quality traits in *E. globulus* has identified nine stable marker-trait associations for seven traits. The linking of SNP from genes showing selection signatures to traits opens new avenues to accelerate the dissection of these traits. The specific localisation of signals to individual SNP in this study, consistent with observations in other tree association studies, supports the assertion that mapping resolution is likely to be extremely high and that informed approaches to candidate SNP selection is advisable to limit the number of tests and increase the potential for discovery. This high mapping resolution is also exciting as it is much more likely that SNP identified are truly causative enabling a direct route from association discovery to functional validation. While alternative approaches such as genomic selection are likely to be more widely deployed to predict breeding values within populations, association study results should find

utility in the development and validation of genomic selection prediction models and, because the associations are more likely to hold across populations, are also likely to be directly useful in breeding high value traits across different breeding programs (Thavamanikumar et al. 2013). As such, association studies will likely remain useful for the further development of MAS for industrial breeding. Also, association study results, such as those reported here, enable the application of new approaches in the study of adaptation in natural populations.

Acknowledgements

The authors wish to thank Gunns Ltd for providing access to the provenance-progeny trial to collect DNA samples, and the Southern Tree Breeding Association for access to the validation population and provision of phenotypic data and breeding values. Dr Chris Harwood is thanked for his constant support throughout the project and for comments on the manuscript. Professor Brad Potts and Dr. Pauline Garnier-Géré are thanked for their comments on the manuscript. Funding support for this project was provided by the CRC for Forestry (www.crcforestry.com.au).

Data Archiving Statement

Genotype (SNP) data and covariates (population structure and Kinship estimates) were submitted to the TreeGenes Database (<http://dendrome.ucdavis.edu/treegenes/>; accession number TGDR033). Phenotype data are commercial data and are archived in the Southern Tree Breeding Association (STBA) DataPlan database. These can be provided on request.

581 **References**

- 582 Abramson JH (2004) WINPEPI (PEPI-for-Windows): computer programs for epidemiologists
583 Epidemiologic Perspectives & Innovations 1:6
- 584 Astle W, Balding DJ (2009) Population Structure and Cryptic Relatedness in Genetic Association
585 Studies Statistical Science 24:451-471
- 586 Balding DJ (2006) A tutorial on statistical methods for population association studies Nature Reviews
587 Genetics 7:781-791
- 588 Beaulieu J et al. (2011) Association Genetics of Wood Physical Traits in the Conifer White Spruce and
589 Relationships With Gene Expression Genetics 188:197-214
- 590 Boerjan W, Polle A, Vander Mijnsbrugge K (2006) A role in lignification and growth for plant
591 phenylcoumaran benzylic ether reductase US patent 2006/0015967
- 592 Bottcher Y et al. (2009) Adipose tissue expression and genetic variants of the bone morphogenetic
593 protein receptor 1A gene (BMPRI1A) are associated with human obesity Diabetes (New York)
594 58:2119-2128
- 595 Bouakaze C et al. (2011) Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass
596 Spectrometry-Based Single Nucleotide Polymorphism Genotyping Assay Using iPLEX Gold
597 Technology for Identification of Mycobacterium tuberculosis Complex Species and Lineages J
598 Clin Microbiol 49:3292-3299 doi:10.1128/jcm.00744-11
- 599 Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ram-doss Y, Buckler ES (2007) TASSEL: Software for
600 Association Mapping of Complex Traits in Diverse Samples Bioinformatics 23:2633-2635
601 doi:10.1093/bioinformatics/btm308
- 602 Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage
603 disequilibrium in loblolly pine Proceedings of the National Academy of Sciences of the
604 United States of America 101:15255-15260
- 605 Cappa EP, El-Kassaby YA, Garcia MN, Acuña C, Borralho NMG, Grattapaglia D, Marcucci Poltri SN
606 (2013) Impacts of Population Structure and Analytical Models in Genome-Wide Association
607 Studies of Complex Traits in Forest Trees: A Case Study in *Eucalyptus globulus*
608 PLoS ONE 8:e81267 doi:10.1371/journal.pone.0081267
- 609 Coleman HD, Yan J, Mansfield SD (2009) Sucrose synthase affects carbon partitioning to increase
610 cellulose production and altered cell wall ultrastructure Proceedings of the National
611 Academy of Sciences 106:13118-13123
- 612 Davidovich O, Kimmel G, Shamir R (2007) GEVALT: An integrated software tool for genotype analysis
613 BMC Bioinformatics 8:36
- 614 Delmer DP (1999) Cellulose biosynthesis: Exciting times for a difficult field of study Annu Rev Plant
615 Physiol Plant Mol Biol 50:245-276 doi:10.1146/annurev.arplant.50.1.245
- 616 Dillon SK, Nolan M, Li W, Bell C, Wu HX, Southerton SG (2010) Allelic Variation in Cell Wall Candidate
617 Genes Affecting Solid Wood Properties in Natural Populations and Land Races of *Pinus*
618 *radiata* Genetics 185:1477-1487
- 619 Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue
620 Phytochemical Bulletin 19:11-15
- 621 Dutkowski GW, Potts BM (1999) Geographic Patterns of Genetic Variation in *Eucalyptus globulus* ssp.
622 *globulus* and a Revised Racial Classification Aust J Bot, 47:237-263
- 623 Eckert AJ et al. (2009) Association Genetics of Coastal Douglas Fir (*Pseudotsuga menziesii* var.
624 *menziesii*, Pinaceae). I. Cold-Hardiness Related Traits Genetics 182:1289-1302
- 625 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the
626 software STRUCTURE: a simulation study Mol Ecol 14:2611-2620 doi:10.1111/j.1365-
627 294X.2005.02553.x
- 628 Eveno E et al. (2008) Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate
629 genes as revealed by genetic differentiation analyses Mol Biol Evol 25:417-437

- Fagard M et al. (2000) PROCUSTE1 encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of arabidopsis Plant Cell 12:2409-2423 doi:10.2307/3871238
- Falush D, Stephens M, Pritchard JK (2003) Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies Genetics 164:1567–1587
- Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles Mol Ecol Notes 7:574-578 doi:10.1111/j.1471-8286.2007.01758.x
- 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:O8
- 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 doi:10.1111/nph.12237
- Garcia-Gil MR, Mikkonen M, Savolainen O (2003) Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris* Mol Ecol 12:1195-1206
- Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium Plant Cell 15:1502-1506
- 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 172:1915-1926
- Gonzalez-Martinez SC, Huber D, Ersoz E, Davis JM, Neale DB (2008) Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination Heredity 101:19-26
- 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
- Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2) Available from <http://www2.unil.ch/popgen/softwares/fstat.htm>
- Grattapaglia D, Kirst M (2008) Eucalyptus applied genomics: from gene sequences to breeding tools New Phytol 179:911-929
- Hadjigol S (2012) Evidence for natural selection acting on genes affecting lignin and cellulose biosynthesis in *Eucalyptus globulus* (MSc thesis). University of Tasmania
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels Mol Ecol Notes 2:618-620
- Hernandez-Blanco C et al. (2007) Impairment of cellulose synthases required for Arabidopsis secondary cell wall formation enhances disease resistance Plant Cell 19:890-903 doi:10.1105/tpc.106.048058
- Hirschhorn JN, Altshuler D (2002) Once and Again--Issues Surrounding Replication in Genetic Association Studies Journal of Clinical Endocrinology Metabolism 87:4438-4441
- 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 Phytol 188:501-514
- Hu WJ et al. (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees Nat Biotechnol 17:808-812
- Ingvarsson PK (2005) Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European Aspen (*Populus tremula* L., Salicaceae) Genetics 169:945-953
- 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 178:2217-2226
- Ivashuta S et al. (2005) RNA interference identifies a calcium-dependent protein kinase involved in Medicago truncatula root development Plant Cell 17:2911-2921 doi:10.1105/tpc.105.035394
- Jallow M et al. (2009) Genome-wide and fine-resolution association analysis of malaria in West Africa Nat Genet 41:657-665 doi:10.1038/ng.388

- Jones TH, Steane DA, Jones RC, Pilbeam D, Vaillancourt RE, Potts BM (2006) Effects of domestication on genetic diversity in *Eucalyptus globulus* For Ecol Manage 234:78-84
- Kalluri UC, Joshi CP (2004) Differential expression patterns of two cellulose synthase genes are associated with primary and secondary cell wall development in aspen trees Planta 220:47-55 doi:10.1007/s00425-004-1329-z
- Kathiresan S et al. (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans Nat Genet 40:189-197 doi:10.1038/ng.75
- Kawaoka A, Nanto K, Ishii K, Ebinuma H (2006) Reduction of lignin content by suppression of expression of the LIM domain transcription factor in *Eucalyptus camaldulensis* Silvae Genet 55:269-277
- Kemper KE, Daetwyler HD, Visscher PM, Goddard ME (2012) Comparing linkage and association analyses in sheep points to a better way of doing GWAS Genetics Research 94:191-203 doi:doi:10.1017/S0016672312000365
- Kiselev KV, Turlenko AV, Zhuravlev YN (2010) Structure and expression profiling of a novel calcium-dependent protein kinase gene PgCDPK1a in roots, leaves, and cell cultures of Panax ginseng Plant Cell Tissue and Organ Culture 103:197-204 doi:10.1007/s11240-010-9767-8
- Krutovsky KV, Neale DB (2005) Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir Genetics 171:2029-2041
- 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 Phytol 191:1041-1053
- Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm Nature Protocols 4:1073-1082 doi:10.1038/nprot.2009.86
- Kwon M, Davin LB, Lewis NG (2001) In situ hybridization and immunolocalization of lignan reductases in woody tissues: implications for heartwood formation and other forms of vascular tissue preservation Phytochemistry 57:899-914 doi:10.1016/s0031-9422(01)00108-x
- Lander ES, Schork NJ (1994) Genetic Dissection of Complex Traits Science 265:2037-2048
- Lepoittevin C, Harvengt L, Plomion C, Garnier-Géré P (2012) Association mapping for growth, straightness and wood chemistry traits in the *Pinus pinaster* Aquitaine breeding population Tree Genet Genom 8:113-126
- Li A, Wang X, Leseberg CH, Jia J, Mao L (2008) Biotic and abiotic stress responses through calcium-dependent protein kinase (CDPK) signaling in wheat (*Triticum aestivum* L.) Plant Signaling & Behavior 3:654-656
- 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 186:1033-1044
- McRae TA, Pilbeam DJ, Powell MD, Joyce GW, Tier KB Genetic evaluation in eucalypt breeding programs. In: Borralho NMG, Pereira JS, Marques C, Coutinho J, Madeira M, Tome' M (eds) Eucalyptus in a ChangingWorld, Aveiro, Portugal, 11-15 October 2004. RAIZ, Instituto Investigação de Floresta e Papel, pp 189-190
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps Genetics 157:1819-1829
- Munafo MR, Flint J (2004) Meta-analysis of genetic association studies Trends Genet 20:439-444 doi:10.1016/j.tig.2004.06.014
- Ng PC, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function Nucleic Acids Res 31:3812-3814 doi:10.1093/nar/gkg509

- Nicol F, His I, Jauneau A, Vernhettes S, Canut H, Höfte H (1998) A plasma membrane-bound putative endo-1,4--D-glucanase is required for normal wall assembly and cell elongation in Arabidopsis The EMBO Journal 17:5563–5576
- Patzlaff A et al. (2003) Characterisation of a pine MYB that regulates lignification Plant J 36:743-754
- 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 Phytol 167:101-112
- Pot D et al. (2006) QTLs and candidate genes for wood properties in maritime pine (*Pinus pinaster* Ait.) Tree Genet Genom 2:10-24
- 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
- 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 Phytol 179:94-103
- 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
- Ranik M, Myburg AA (2006) Six new cellulose synthase genes from Eucalyptus are associated with primary and secondary cell wall biosynthesis Tree Physiology 26:545-556
- Renaut S, Nolte AW, Rogers SM, Derome N, Bernatchez L (2011) SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus* spp.) Mol Ecol 20:545-559 doi:10.1111/j.1365-294X.2010.04952.x
- Resende MDV et al. (2012a) Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees New Phytol 194:116-128
- Resende MFR et al. (2012b) Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments New Phytol 193:617-624
- Resende MFR et al. (2012c) Accuracy of Genomic Selection Methods in a Standard Dataset of Loblolly Pine (*Pinus taeda* L.) Genetics 190:1503-1510
- Richmond T (2000) Higher plant cellulose synthases Genome Biology 1:reviews3001.3001 - reviews3001.3006
- Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients Genet Res 67:175-185
- Romeis T, Ludwig AA, Martin R, Jones JDG (2001) Calcium-dependent protein kinases play an essential role in a plant defence response EMBO J 20:5556-5567 doi:10.1093/emboj/20.20.5556
- Schindelman G et al. (2001) COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in Arabidopsis Genes Dev 15:1115-1127
- Schrader S, Sauter JJ (2002) Seasonal changes of sucrose-phosphate synthase and sucrose synthase activities in poplar wood (*Populus x canadensis* Moench robusta) and their possible role in carbohydrate metabolism J Plant Physiol 159:833-843
- Sexton TR et al. (2012) Pectin Methylesterase Genes Influence Solid Wood Properties of Eucalyptus pilularis Plant Physiol 158:531-541 doi:10.1104/pp.111.181602
- Spokevicius AV et al. (2007) β -tubulin affects cellulose microfibril orientation in plant secondary fibre cell walls The Plant Journal 51:717-726
- Stackpole DJ, Vaillancourt RE, Aguilar Md, Potts BM (2010a) Age trends in genetic parameters for growth and wood density in Eucalyptus globulus Tree Genetics and Genomes 6:179-193
- Stackpole DJ, Vaillancourt RE, Alves A, Rodrigues J, Potts BM (2011) Genetic Variation in the Chemical Components of Eucalyptus globulus Wood G3: Genes, Genomes, Genetics 1:151-159 doi:10.1534/g3.111.000372

781 Stackpole DJ, Vaillancourt RE, Downes G, Harwood CE, Potts BM (2010b) Genetic control of kraft
782 pulp yield in *Eucalyptus globulus* Canadian Journal of Forest Research 40:917-927

783 Steane DA, Conod N, Jones RC, Vaillancourt RE, Potts BM (2006) A comparative analysis of
784 population structure of a forest tree, *Eucalyptus globulus* (Myrtaceae), using microsatellite
785 markers and quantitative traits Tree Genet Genom 2:30–38

786 Steane DA et al. (2011) Population genetic analysis and phylogeny reconstruction in *Eucalyptus*
787 (Myrtaceae) using high-throughput, genome-wide genotyping Mol Phylogen Evol 59:206-224
788 doi:10.1016/j.ympev.2011.02.003

789 Stich B, Möhring J, Piepho H, Heckenberger M, Buckler ES, Melchinger AE (2008) Comparison of
790 Mixed-Model Approaches for Association Mapping Genetics 178:1745-1754

791 Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies Proceedings of the
792 National Academy of Sciences of the United States of America 100:9440-9445
793 doi:10.1073/pnas.1530509100

794 Stouffer SA, Suchman EA, De Vinney LC, Star SA, Williams RJ (1949) The American soldier:
795 Adjustment during army life vol 1. Princeton University Press, New Jersey

796 Sturm A, Tang GQ (1999) The sucrose-cleaving enzymes of plants are crucial for development,
797 growth and carbon partitioning Trends Plant Sci 4:401-407

798 Szyjanowicz PMJ, McKinnon I, Taylor NG, Gardiner J, Jarvis MC, Turner SR (2004) The irregular xylem
799 2 mutant is an allele of korrigan that affects the secondary cell wall of *Arabidopsis thaliana*
800 Plant J 37:730-740

801 Tchin BL, Ho WS, Pang SL, Ismail J (2011) Gene-associated Single Nucleotide Polymorphism (SNP) in
802 *Cinnamate 4-Hydroxylase (C4H)* and *Cinnamyl Alcohol Dehydrogenase (CAD)* Genes from
803 *Acacia mangium* Superbulk Trees Biotechnology 10:303-315

804 Thavamanikumar S (2010) 'Using genetic association studies for the improvement of wood and fibre
805 properties in *Eucalyptus globulus* ssp. *globulus* Labill' (PhD thesis). The University of
806 Melbourne

807 Thavamanikumar S, McManus LJ, Tibbits JFG, Bossinger G (2011) The significance of single
808 nucleotide polymorphisms (SNPs) in *Eucalyptus globulus* breeding programs Australian
809 Forestry 74:23-29

810 Thavamanikumar S, Southerton S, Bossinger G, Thumma B (2013) Dissection of complex traits in
811 forest trees — opportunities for marker-assisted selection Tree Genetics & Genomes:1-13
812 doi:10.1007/s11295-013-0594-z

813 Thavamanikumar S, Southerton S, Thumma B (2014) RNA-Seq Using Two Populations Reveals Genes
814 and Alleles Controlling Wood Traits and Growth in *Eucalyptus nitens* PLoS ONE 9:e101104
815 doi:10.1371/journal.pone.0101104

816 Thumma BR, MacMillan CP, Southerton SG, Williams D, Joyce K, Ravenwood IC (2010) Accelerated
817 breeding for high pulp yield in *E. nitens* using DNA markers identified in 100 cell wall genes:
818 The Hottest 100 (Research Report) Forest and Wood Products Australia Research Reports
819 PNC052-0708

820 Thumma BR, Matheson BA, Zhang D, Meeske C, Meder R, Downes GM, Southerton SG (2009)
821 Identification of a Cis-Acting Regulatory Polymorphism in a Eucalypt COBRA-Like Gene
822 Affecting Cellulose Content Genetics 183:1153-1164

823 Thumma BR, Nolan MR, Evans R, Moran GF (2005) Polymorphisms in cinnamoyl CoA reductase (CCR)
824 are associated with variation in microfibril angle in *Eucalyptus* spp Genetics 171:1257-1265

825 Tibbits JFG, McManus LJ, Spokevicius AV, Bossinger G (2006) A rapid method for tissue collection
826 and high-throughput isolation of genomic DNA from mature trees Plant Molecular Biology
827 Reporter 24:81-91

828 Traurig M et al. (2009) Common variation in SIM1 is reproducibly associated with BMI in Pi ma
829 Indians Diabetes (New York) 58:1682-1689 doi:10.2337/db09-0028

- Vander Mijnsbrugge K, Meyermans H, Van Montagu M, Bauw G, Boerjan W (2000) Wood formation in poplar: identification, characterization, and seasonal variation of xylem proteins *Planta* 210:589-598
- Washusen R (2002) Tension wood occurrence in *Eucalyptus globulus* Labill. II. The spatial distribution of tension wood and its association with stem form *Australian Forestry* 65:127-134
- Xu SZ (2003) Theoretical basis of the Beavis effect *Genetics* 165:2259-2268
- Yang J et al. (2003) Novel gene expression profiles define the metabolic and physiological processes characteristic of wood and its extractive formation in a hardwood tree species, *Robinia pseudoacacia* *Plant Mol Biol* 52
- Yang SS et al. (2011) Using RNA-Seq for gene identification, polymorphism detection and transcript profiling in two alfalfa genotypes with divergent cell wall composition in stems *BMC Genomics* 12:199
- Yeoh SH, Bell JC, Foley WJ, Wallis IR, Moran GF (2012) Estimating population boundaries using regional and local-scale spatial genetic structure: an example in *Eucalyptus globulus* *Tree Genet Genom* 8:695-708 doi:10.1007/s11295-011-0457-4
- Yu JM et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness *Nat Genet* 38:203-208
- Zhao J et al. (2007a) Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa* *Genome* 50:963-973
- Zhao KY et al. (2007b) An *Arabidopsis* example of association mapping in structured samples *PLoS Genet* 3:e4
- Zollner S, Pritchard JK (2007) Overcoming the winner's curse: Estimating penetrance parameters from case-control data *Am J Hum Genet* 80:605-615 doi:10.1086/512821
- Zuo JR, Niu QW, Nishizawa N, Wu Y, Kost B, Chua NH (2000) KORRIGAN, an *Arabidopsis* endo-1,4-beta-glucanase, localizes to the cell plate by polarized targeting and is essential for cytokinesis *Plant Cell* 12:1137-1152

857 **Supporting Information**

858 **Supplementary Figure 1** Genotypic effects of SNP KOR_2328 on Pulp Yield and Cellulose
859 based on association analysis.

860 **Supplementary Table 1** Squared allele frequency correlations (linkage disequilibrium, r^2)
861 between 98 SNPs genotyped in discovery population across all populations where $r^2 > 0.33$.

862 **Supplementary Table 2** SNP-trait association results in the discovery population

863 **Supplementary Table 3** SNP-trait association results in the validation population

864 **Supplementary Table 4** SNP effects estimated by ridge regression (RR) and mixed linear
865 model (MLM) in the discovery population

866 **Appendix A1** SNP genotyping using iPLEXTM Gold reaction

867 **Appendix A2** Genetic structure of *E. globulus* samples

868

869 **Figure Legends**

870 **Figure 1** Geographical distribution of *E. globulus* trees belonging to different races sampled
871 in the association discovery population planted in Latrobe, Tasmania. Races are given in bold
872 letters and localities with plain letters with number of trees sampled from each locality in
873 parentheses. Inset: Location of discovery (Latrobe, Tasmania) and validation (Frankland,
874 Western Australia) populations.

875 **Figure 2** Trait-wise QQ-plots for different association models used in the discovery
876 population. See text for models' abbreviation.

877 **Figure 3** Trait-wise QQ-plots for different association models used in the validation
878 population. See text for models' abbreviation.

879 **Figure 4** Heatmaps of the pair-wise relationship coefficients of samples from discovery
880 population (a) and validation population (b)

881 **Figure 5** Genotypic effects of KOR_2328 in the discovery (top) and validation population
882 (bottom) based on association analysis. Error bars represent standard errors. Number of trees
883 is presented next to each genotype class in the x-axis. This association was validated using a
884 combined-analysis. Combined-analysis (Weighted Stouffer's method) results: $P = 0.0002$; Q
885 $= 0.0043$.

886 **Figure 6** Top: KOR_2328, a triallelic SNP from Korrigan gene. At the 263rd codon, the first
887 site (2328) is tri-allelic (G/A/T) and the second site (2329) is bi-allelic (C/T) which resulted
888 in a coding combination for six amino acids. Bottom: Predicted amino acid frequencies for
889 individuals genotyped with KOR_2328 SNP in the association and validation populations.
890 Frequencies were predicted based on the frequencies observed for KOR_2328 and

891 KOR_2329 SNPs in the Sequencing data consisting of 28 individuals (S. Thavamanikumar *et*
892 *al.*, unpublished).

893

894

895 **Table 1** Description of phenotypic traits used in this study

Traits	Description	Unit
<u>Growth traits</u>		
DBH ^a	Diameter at breast height (1.3 m) over bark	cm
Physical wood properties		
Density	Basic density determined by gravimetric method	Kg m ⁻³
Chemical wood properties		
Cellulose	Predicted cellulose content using near infrared spectra (NIR)	%
Pulp Yield	Predicted pulp yield using NIR	%
Klason Lignin	Predicted klason lignin content using NIR	%
Extractives	Predicted extractive content using NIR	%
S/G ratio ^b	Ratio of syringyl-like lignin structures (S) and guaiacyl-like lignin structures (G)	Ratio
<u>Deregressed Estimated breeding values</u>		
Density_DEBV	Harvest Age predicted breeding value for basic wood density	NPV ^c
Pulp-Yield_DEBV	Harvest Age predicted breeding value for pulp yield	NPV
Volume_DEBV	Harvest Age predicted breeding value for volume	NPV

896

897 ^aDBH is measured at different ages;, DBH was measured at 4 (DBH04), 8 (DBH08) and 16 (DBH16) years of age in the discovery population;
898 DBH was measured at 2 (DBH02), 4 (DBH04) and 7 (DBH07) years of age in the validation population

899 ^bS/G ratio was measured only in the discovery population

900 ^cNet Present Value (\$Aus)/Hectare

901

902 **Table 2** List of SNPs used in the association analysis in discovery population

Gene	<i>Eucalyptus grandis</i> ID ^a	Gene product	Function	SNPs
4CL	Eucgr.C02284	4-coumarate: CoA ligase	Lignin biosynthesis	12
AQP1	Eucgr.D01836	Aquaporin	Water transport	4
CAD	Eucgr.H03208	Cinnamyl alcohol dehydrogenase	Lignin biosynthesis	2
CCoAOMT1	Eucgr.I01134	caffeoyl-CoA O-methyltransferase 1	Lignin biosynthesis	1
CCoAOMT2	Eucgr.G01417	caffeoyl-CoA O-methyltransferase 2	Lignin biosynthesis	4
CesA1	Eucgr.D00476	Cellulose Synthase 1	Cellulose biosynthesis	2
CesA3	Eucgr.C00246	Cellulose Synthase 3	Cellulose biosynthesis	6
COBL4	Eucgr.J01393	Cobra-Like gene	Cellulose biosynthesis	10
Korrigan	Eucgr.G00035	Membrane-bound endo-1, 4-β-D-glucanase	Cellulose biosynthesis	9
LIM1	Eucgr.F02243	LIM transcription factor	Transcription factor involved in lignin biosynthesis	11
LTP	Eucgr.B00824	Lipid transfer protein	Lipid transfer	9
MYB1	Eucgr.G01774	Myb transcription factor 1	Transcription factor involved in lignin biosynthesis	1
MYB2	Eucgr.G03385	Myb transcription factor 2	Transcription factor involved in lignin biosynthesis	3
NAP1	Eucgr.D01822	Nitrilase associated protein	Auxin biosynthesis	4
PCBER	Eucgr.F01584	Phenylcoumaran benzylic ether reductase	Lignin biosynthesis	3
CDPK	Eucgr.I01536	Calcium Dependent Protein Kinsase	Plant defence	2
SAMS	Eucgr.K00588	S-adenosylmethionine synthase	Methionine Metabolism	2
SuSy1	Eucgr.C02844	Sucrose Synthase	Cellulose biosynthesis	4
SuSy3	Eucgr.C00769	Sucrose Synthase	Cellulose biosynthesis	7
TUB1	Eucgr.D01847	Beta tubulin	Guide cellulose microfibril orientation	2
Total				98

903 ^aObtained from *Eucalyptus grandis* genome browser: <http://www.phytozome.net/eucalyptus.php>

904

905 **Table 3 SNP-Trait associations (P<0.05) identified from different models used in association analysis**

Model	DBH02	DBH04	DBH08*	DBH16	Density	Cellulose	Pulp Yield	Klason Lignin	Extractives	S/G ratio	Density_DEBV	Pulp-Yield_DEBV	V
Discovery Population													
GLM	NA	2	3	4	29	41	38	32	33	46	32	53	
GLM (Q)	NA	3	3	3	7	3	3	4	6	8	12	20	
MLM (Q+K)	NA	3	3	3	7	3	2	4	7	10	9	11	
Validation Population													
GLM	3	3	2	NA	4	1	5	2	4	NA	3	12	
MLM (K)	1	2	2	NA	4	1	1	3	2	NA	3	3	

906

907

*DBH07 in the validation population

908 **Table 4** SNP-Trait associations identified in combined-analysis

Locus	Trait	SNP	SNP Type ^a	n ^b	F ^c	P ^d	Discovery Population		n	F	P	Validation Population		Combined-analysis		
							SNP effect (BLUP) ^c	SNP effect (BLUE) ^e				SNP effect (BLUP)	SNP effect (MLM)	Z ^f	P ^g	Q ^h
COBL4_1391	DBH04	A/T	NSS	356	3.76	0.0243	-0.0174	-1.7959	293	1.07	0.3440	-0.3248	-3.9957	2.41	0.016	0.086
COBL4_723	DBH04	C/T	Promoter/5'UTR	368	4.24	0.0150	-0.0524	-6.3653	292	0.94	0.3904	0.0702	-4.8075	2.56	0.011	0.071
COBL4_723	Extractives			366	0.60	0.5467	-0.0002	-0.1194	291	5.30	0.0055	-0.0663	-0.4491	2.8	5.10E-03	0.039
CSA3_1101	Density	A/T	SS	372	2.29	0.1025	-0.5897	-0.9684	290	5.84	0.0033	-1.9955	-26.8019	3.3	9.60E-04	0.021
KOR_2328	Pulp-Yield_DEBV	A/G/T	NSS	349	3.25	0.0070	0.2682	NA	244	2.22	0.0473	-0.4231	NA	2.89	0.004	0.036
PCBR_1601	Klason Lignin	C/G	Intronic	371	9.51	0.0001	-0.0747	-0.5217	292	1.37	0.2555	0.0109	0.0411	3.51	4.50E-04	0.02
CDPK_516	Density_DEBV	A/G	Intronic	364	1.22	0.2975	1.2642	2.7492	286	3.59	0.0244	-0.0757	5.5220	2.36	0.018	0.086
CDPK_516	Volume_DEBV			364	2.01	0.1352	-1.2382	-10.7613	286	5.54	0.0053	-0.4095	-9.2868	2.97	0.003	0.034
Susy3_886	Extractives	A/G	Promoter/5'UTR	368	3.17	0.0432	0.0304	0.0534	292	3.78	0.0238	0.0075	0.1804	2.35	0.019	0.086
Susy3_909	Volume_DEBV	A/G	Promoter/5'UTR	366	0.34	0.7144	0.4626	3.6841	289	5.79	0.0034	1.1368	9.4715	2.95	0.003	0.034

909 ^aNSS – Nonsynonymous substitution, SS - Synonymous substitution, UTR – Untranslated Regions

910 ^bNumber of samples

911 ^cF and P-values obtained from a Mixed Linear Model (MLM) implemented in TASSEL version 2.0.1

912 ^dSNP effect (minor allele) estimated from ridge regression analysis by fitting all SNPs simultaneously as random effects.

913 ^eSNP effect (minor allele) estimated from mixed linear model by fitting SNPs as fixed effects.

914 ^fStouffer's Z Score from the combined-analysis

915 ^gCombined P-values from the combined-analysis using Stouffer's Z method

916 ^hQ values obtained for combined P-values by FDR multiple testing correction

918

919

a.

b.

924

925

Discovery Population

Validation Population

926

927

Individual	Site 2328	Site 2329	Site 2330	Codon	Amino Acid
1	G	T	T	GTT	Valine
2	A	T	T	ATT	Isoleucine
3	G	C	T	GCT	Alanine
4	A	C	T	ACT	Threonine
5	T	C	T	TCT	Serine
6	T	T	T	TTT	Phenylalanine

Thr : Thr = 79.0%	Ala : Thr = 22.2%	Ala : Ala = 6.3%	Ser : Thr = 6.8%	Ser : Ala = 1.9%	Ser : Ser = 0.6%
Ile : Thr = 19.8%	Val : Thr = 66.7%	Val : Ala = 37.5%	Phe : Thr = 82.1%	Phe : Ala = 23.1%	Phe : Ser = 14.2%
Ile : Ile = 1.2%	Ala : Ile = 2.8%	Val : Val = 56.3%	Ser : Ile = 0.9%	Ser : Val = 5.8%	Phe : Phe = 85.2%
	Val : Ile = 8.3%		Phe : Ile = 10.3%	Phe : Val = 69.2%	
AA	AG	GG	AT	GT	TT

928

929

930 **Appendix A1 SNP genotyping using iPLEXTM Gold reaction**

931 iPLEXTM Gold reaction based on MALDI-TOF MS developed by SEQUENOM was used as per the iPLEXTM Gold Application Guide for
932 scoring polymorphisms. SEQUENOM's Assay Design Software was used to design amplification primers for multiplex reactions. This software
933 checks for possible dimer formation. A generic 10-mer tag (ACGTTGGATG) was added to the 5' end of each amplification primer. Primers
934 were designed to produce 80 to 120 bp amplicons (Table 1). Extension primers were designed with 3' ends immediately adjacent to the targeted
935 polymorphic site.

936 Eleven to fourteen plex reactions were designed. Polymerase Chain Reaction (PCR) was performed in 5 µL reactions. Each reaction contained 1
937 µL of appropriate genomic DNA (5- 10 ng/µL), 1.25 x PCR buffer (15mM MgCl₂), 1.625 mM MgCl₂, 500 µM dNTP mix, 100 nM primer mix
938 and 0.5 U Taq. Thermocycling conditions were: an initial denaturation step at 94°C for 15 minutes, 45 amplification cycles at 94° C for 20
939 seconds, 56° C for 30 seconds and 72°C for one minute and a final extension step at 72°C for 3 minutes.

940 Post PCR, reactions were treated with SAP (Shrimp Alkaline Phosphatase) to neutralise unincorporated dNTPs. Two microlitres of SAP mix
941 containing 0.5 units of SAP and 1x SAP buffer was added to the PCR product. This mix was then incubated at 37⁰ C for 40 minutes before
942 denaturation of SAP *via* incubation at 85⁰ C for five minutes.

943 To the SAP treated PCR product, 9 µL of iPLEX Gold reaction master mix was added. This mix contained 0.222x iPLEX buffer plus, 0.5x
944 iPLEX terminator mix, 0.625 µM primer mix and 0.5x iPLEX enzyme. This mix was thermocycled as follows: an initial denaturation step at

945 94°C for 30 sec followed by an annealing step at 52° C for five seconds and extension step at 80° C for five seconds. The annealing and
 946 extension cycle was repeated four more times for a total of five cycles and then looped back to a 94° C denaturing step for five seconds before
 947 entering the five cycle annealing and extension loop again. The five annealing and extension steps with the single denaturing step were repeated
 948 an additional 39 times (a total of 40 times). A final extension followed at 72° C for three minutes.

949 iPLEX Gold reaction products were cleaned using resin. The cleaned iPLEX Gold reaction products were nanodispensed onto a spectroCHIP.
 950 Spectra were then acquired using the MassARRAY Analyzer Compact mass spectrometer.
 951

SNP	Forward primer	Reverse primer	Sequencing primer
4CL_1787	ACGTTGGATGTCTGAAGATTGCTCGAG CAC	ACGTTGGATGACCCGATGCGGAAATTA GAG	GAAATTAGAGGATCGTCTTATT CA
4CL_1920	ACGTTGGATGCAGTGTCAACCCATGAG TAG	ACGTTGGATGAGTTTTTCAGCCGCTACTA TG	tTTTTCAGCCGCTACTATGTGTC CT
4CL_1955	ACGTTGGATGTGACAGGGGAAAACAA AAGG	ACGTTGGATGTAGTAGCGGCTGAAAAC TCC	GGCTGAAAACCTCCGGCTAC
4CL_2121	ACGTTGGATGGCTTGACTTGTAACAAA TGGG	ACGTTGGATGAAGGAAAAGGACTTTTC AGG	GACTTTTCAGGCCCTC
4CL_2219	ACGTTGGATGCCGGCAATTTCTAAATC CGC	ACGTTGGATGCTAGAAGCCAAGCCACA AAG	CAAAGAAAGATGGTCCAAG
4CL_222	ACGTTGGATGCTCACGCGAAATTGAAG AAG	ACGTTGGATGTGGGTATGCGATTGAAA TCC	ATTGAAATCCTTCGAGATCGC
4CL_2232	ACGTTGGATGGCCACAAAGAAAGATGG TCC	ACGTTGGATGCCGGCAATTTCTAAATC CGC	aTGCGATAAAACAAAACATTAAA TAA
4CL_294	ACGTTGGATGTCGCGTGAGGTAGATTT TTC	ACGTTGGATGAACACGGTCAAACACGC TTG	TCCCTCCTGTCACGACATTTCC
4CL_3817	ACGTTGGATGCAACATCTAATGCACCC TCC	ACGTTGGATGTCTTGAATTATTGTGGAC CG	gggTTGTGGACCGTCGAAAC
4CL_4464	ACGTTGGATGGGGCAGGGTTGAAAAAG AAA	ACGTTGGATGTTGAGGAAGGACCTAAG AGC	GTTGGCCTCCGGTGTTT
4CL_485	ACGTTGGATGTTCTCCTTCGTTAGGTT	ACGTTGGATGAGATGGAAAAGAGGCAT	ATGCGGATCAAACATTG

	GG	GCG	
4CL_713	ACGTTGGATGAAGGGACAGAAAAGGC CATC	ACGTTGGATGACCCGAGAGTGATATCA AGG	AGGAGGGGAAAGAAAATTAA gACCTGCTATTTGTTTTACATTT
AQP1_2977	ACGTTGGATGTAGAGAAGCTAACTGTA CCG	ACGTTGGATGCATCCCTTTCATAAACCT GC	AT
AQP1_3084	ACGTTGGATGAGGGCGACAGAAAACA ATAG	ACGTTGGATGAGATCCTCCTGTTCTAAC CC	TAACCCAAATGTTAGCAGA
AQP1_3111	ACGTTGGATGAGGGCGACAGAAAACA ATAG	ACGTTGGATGAGATCCTCCTGTTCTAAC CC	GCCTTCTCTTCCGTGTAATCT
AQP1_3146	ACGTTGGATGCCCAAGTGTAATCTAAC AGC	ACGTTGGATGGCCTTCTCTTCCGTGTAA TC	TTGTTTTCTGTGCGCCCTAA gAGGAAGAATCCAAAAGAGAGA
CADp_183	ACGTTGGATGAGGTATTGCTTGCTGAG GTG	ACGTTGGATGGAAACGAGACGGACCAA TAG	C AGGAGTGACGGGGATACATATG
CADp_282	ACGTTGGATGGGTTGCTGGTAGCGTTTT TG	ACGTTGGATGGTAGGAGCATAGATGAG GAG	GA
CCoAOMT1_83	ACGTTGGATGGACACAACACCTCAAGT TGG	ACGTTGGATGGCTTTTGTGGAATCGAT GG	GGAATCGATGGATGATTCT
CCoAOMT2_16	ACGTTGGATGTGTTGTCGTATCCGATCA 69	ACGTTGGATGCCGACAAGGACAACCTAC ATC	GACAACTACATCAACTACCACA
CCoAOMT2_17	ACGTTGGATGATCGACCTGGTCAAGGT 22	ACGTTGGATGGTAGTACCGGACGTACT TGC	tCCACGGAGCCGTTCCACA
CCoAOMT2_25	ACGTTGGATGCTGTTATTTCCCTGAGCT 8	ACGTTGGATGCTTGGGATTTGCCTGTTT GC	aagGCGTCTACCCAAGAGAGCCA GAG
CCoAOMT2_92	ACGTTGGATGTGGCCATGGACATCAAT 0	ACGTTGGATGAATCGATCTTGTGGGCA AGG	TCTTGTGGGCAAGGCCGGCCTT CT
COBL4_1391	ACGTTGGATGTACTGCAGGTTGTAAGG GAC	ACGTTGGATGACTGCTCCAAGTTCAAA GCC	CAAAGCCGCTATACCTCAT GTGTAACACTGAAACATGTGGT
COBL4_1925	ACGTTGGATGGTCACATTCCATGTCACT GC	ACGTTGGATGCCGAAGAACCATGGTGT AAC	TG
COBL4_1995	ACGTTGGATGCACAGCAGCTGGGATTT TTC	ACGTTGGATGCAGTGACATGGAATGTG ACC	ATGTGACCTGCACTTATTCCCA
COBL4_2214	ACGTTGGATGACTCCTGCATCACCAAT AGC	ACGTTGGATGGAGCGCTTATGTGCTCA ATG	CACCCGAAGCCTGTTTAC
COBL4_2233	ACGTTGGATGGATAAATCACCCGAAGC CTG	ACGTTGGATGTTTGGAGTCACTCCTGCA TC	TGCATCACCAATAGCCG

COBL4_2655	ACGTTGGATGTTCCATACGCAGATGAC ACG	ACGTTGGATGACTGAACATTCCCAAGT GGC	ATTCCCAAGTGGCCCCGCTTCC AT
COBL4_2725	ACGTTGGATGTTGGGAATGTTTCAGTCC GAG	ACGTTGGATGTTCCCTTGGAATGCCCA ACC	ATGCCCAACCTTGCTTAA
COBL4_2828	ACGTTGGATGATGAGAAGGACCCAAGA GAC	ACGTTGGATGGATTCATACCCTTTTCTG CC	CTGCCAAATTCAGCACGTC
COBL4_723	ACGTTGGATGAATCAATCGCGTCTATA TGG	ACGTTGGATGCACCATATATCTTGCGA GTTC	TGCGAGTTCGCTTCAGCAG
COBL4_781	ACGTTGGATGATCTTGCGAGTTCGCTTC AG	ACGTTGGATGCGAGAAAGACAAATCAA TCGC	AAATCAATCGCGTCTATATGG
CSA1_2516	ACGTTGGATGCGACTTCTTCGTGAGTA CAG	ACGTTGGATGAGCAAGGATGGAAAGCA CTG	TGATCAAAGGTGGTTCTTTCA
CSA1_913	ACGTTGGATGTGCCATTGTCGATCGAC TTC	ACGTTGGATGGGTTTCCTTTTTATTAG CG	cCCTTTTATTAGCGAACTCAA TG
CSA3_1101	ACGTTGGATGCGGGACAAAGTTTTGCT CAC	ACGTTGGATGTCCTGAGGATGACGATA ACG	ctcACGCTCAATTTCCATCAGTTA T
CSA3_1767	ACGTTGGATGGAATGCAGGATAGATGA GGC	ACGTTGGATGCCGGTATGGATTGATCTT GC	TTGCTCGATGCAATTGG
CSA3_1872	ACGTTGGATGCGCTACAGGATATTGAA CCC	ACGTTGGATGATCCAGGATCCAGGAGA AAG	AACCAAAGACCAAATGC
CSA3_2724	ACGTTGGATGGAAAATGGGTACCCTTC TGC	ACGTTGGATGGGCTGGACTTTGTCTTTG AG	ggGAAGTACATTTAGGGGCTCG GGG
CSA3_3734	ACGTTGGATGTGATTGCTGCCCATGCTT TG	ACGTTGGATGGATCTGCAGCATCCCCA TTG	GCACTATGCTTAGAATACTT
CSA3_4186	ACGTTGGATGCTATAGAGGCACCCACT TAG	ACGTTGGATGGCAACAAGGTAGCAAAG CTC	TAGCAAAGCTCTGGGTATT
KOR_187	ACGTTGGATGAACGGCTAGTGAAACGT TGC	ACGTTGGATGCCACACCGATTTACCAA GTC	TTACCAAGTCTCCCTCC
KOR_2328	ACGTTGGATGATCTCGGCAGCAAGATC TGA	ACGTTGGATGGATCATTACTGCTGGAT GCG	atACATTGATTATAAACGGCCTG TC
KOR_2493	ACGTTGGATGCATCCCAATAACTAGTG GAG	ACGTTGGATGCATTTTTTCGAGGCAACA GCG	GGAGATACAGTGCTGGT
KOR_2496	ACGTTGGATGCATCCCAATAACTAGTG GAG	ACGTTGGATGCATTTTTTCGAGGCAACA GCG	CGAGGGAGATACAGTGCTGGTT CA
KOR_2863	ACGTTGGATGATTGAAAGTCTCAACGG	ACGTTGGATGAGGTGTATAATTTGCCA	cccGCCACTCTTTCAGATCTAAA

	CCA	CTC	AGTA
KOR_4184	ACGTTGGATGAAACATGGGAGGAACCG CAG	ACGTTGGATGAGGGAATGCAGGACTGG TTG	cccCCTTATCAGGCGACGGGA
KOR_742	ACGTTGGATGTCTCTCTCACCCACTGTC G	ACGTTGGATGGCTTCAGCTCCCAGCAA TAG	gATATCAAGAAAATCAAGAACA GG
KOR_801	ACGTTGGATGGTACATGCTCATCCTCG CTC	ACGTTGGATGTTGCTGGGAGCTGAAGC AGA	CTGAAGCAGACCCAGCAGTT
KOR_813	ACGTTGGATGCTGAAGCAGACCCAGCA GTT	ACGTTGGATGGTACATGCTCATCCTCGC TC	CCACAGTGGATGCAGAA
LIM_1106	ACGTTGGATGTACACCAGGAAATCCTT ACC	ACGTTGGATGACGCGAGACAAATGTGT AGG	ATGTGTAGGCTGTAAGAG
LIM_1157	ACGTTGGATGCTCAATGAAGCAGCCTT TGG	ACGTTGGATGAAGGTAAGGATTTCTTG GTG	TTTCCTGGTGTATTTATAGAAC
LIM_1159	ACGTTGGATGAAGGTAAGGATTTCTTG GTG	ACGTTGGATGCTCAATGAAGCAGCCTT TGG	GTCATGTCCTCGCATCA
LIM_1343	ACGTTGGATGAGAGCTGCTTCAAATGC ACC	ACGTTGGATGCAGTAGAGTTTCCCCTC GTG	ACGTAGTTGGATGGGCTGATCA C
LIM_444	ACGTTGGATGGGGCTGATCTTAAGGTT CTC	ACGTTGGATGCTGAATGATATTCGCTTG GG	cTGTAGAAATGTGTGGGGGC
LIM_537	ACGTTGGATGGCCTTTTGACTGATCATC CG	ACGTTGGATGTAAATCATGCGTGACCC CAG	tATCATGACAATTCATCATAACT
LIM_564	ACGTTGGATGGCCTTTTGACTGATCATC CG	ACGTTGGATGTAAATCATGCGTGACCC CAG	CGTGACCCCAGCTTCTT
LIM_795	ACGTTGGATGCAGCTCTTCAAGAGAAC TGG	ACGTTGGATGAACAGAGTGCATGACTT GCG	TTTTACCTTCAAAGCTTTTTTC
LIM_901	ACGTTGGATGTCTGGCTTTGCAATCTTG GG	ACGTTGGATGATGCACTCTGTTTCTGTC CC	TTCTTCATACAAAATGTCCTCT
LIM_977	ACGTTGGATGGACTGATTTAGCCTAAC TGC	ACGTTGGATGTTGCAAAGCCAGAGAAA CCC	cccCCGCGACTATGGTCTTGCCT
LIM_978	ACGTTGGATGTTGCAAAGCCAGAGAAA CCC	ACGTTGGATGGACTGATTTAGCCTAAC TGC	cTTTAGCCTAACTGCAAACTT
LTP_211	ACGTTGGATGTGTAGATGACCAGGGAA ACG	ACGTTGGATGACACTGGCTTATAGCTA GCG	ACTAAACTCATCTCGATCT
LTP_410	ACGTTGGATGCAGAATCATTACTGCTG TGC	ACGTTGGATGTACAACCCACATCAGC ATC	CACATCAGCATCACCGAAAGCA

LTP_430	ACGTTGGATGCAGAATCATTACTGCTG TGC	ACGTTGGATGCACATCAGCATCACCGA AAG	aAGTGTTCGGGAAAGTTG
LTP_507	ACGTTGGATGCTGCACAGCAGTAATGA TTC	ACGTTGGATGTGAGGATCAGCATGCTC AAG	ATCAGCATGCTCAAGATGCCCT TT
LTP_676	ACGTTGGATGAGTGGCAGATCTCATT CAG	ACGTTGGATGGGGTCAGCGTGATGCAT TTG	tgaagTTATAGGAGCCAAATTGGT GA
LTP_730	ACGTTGGATGAGGAGGATGTGAGGAG AAAG	ACGTTGGATGTGGCTCCTATAAAATTCAT GC	cACCAAAGCCAAATGCATCAC
LTP_754	ACGTTGGATGTGGCTCCTATAAAATTCAT GC	ACGTTGGATGAGGAGGATGTGAGGAGA AAG	TGTGAGGAGAAAGAACTAAGA GTG
LTP_835	ACGTTGGATGCAAAGCAGAAGCAATGA GGG	ACGTTGGATGATGTACAGCGGCCTCA AC	GGAGCAGCACCGCCAGAAGA
LTP_925	ACGTTGGATGTCTGCTCCTTGATCTTGC TG	ACGTTGGATGTGACATGCAGGCCGATA GAG	ATAGAGCTGAGCGCGTGCGTGA
MYB1_474	ACGTTGGATGAGGTAATTGATCCACCG GAG	ACGTTGGATGTTGCCTACATAAGAGCG CAC	cCCGCGGGCCTCCTCCGCTG
MYB2_1380	ACGTTGGATGGCACGTCTTCCTGGAAG AAC	ACGTTGGATGATCGCTCTCGTTTGACGA AG	gagCGTTTGACGAAGATGTAG
MYB2_1436	ACGTTGGATGCCATGCATGTAGCTAAG ACG	ACGTTGGATGTCCAATCCACAAGACAT AGC	CAAGACATAGCTGCAGG
MYB2_1625	ACGTTGGATGGTTTTTACCTTCCCCTC AG	ACGTTGGATGAGTCCCACTTTTGCATGC TC	TAACTTGAATCATCTCCCATACT
NAP1_221	ACGTTGGATGGAATGAAGACAGCCGTT CTC	ACGTTGGATGAGTTTTGGCCTTGAGCTC TG	aTTGAGCTCTGTGGTAATTGTT
NAP1_286	ACGTTGGATGATGAACTTGGGGCAGTA CAC	ACGTTGGATGAGAGCTCAAGGCCAAAA CTC	TCATCACCGTAAGTGATTTTCTC T
NAP1_317	ACGTTGGATGCAACTGTGTACTGCCCC AAG	ACGTTGGATGAACTGATTTGACCTGGG TCG	gCTTCTTAAGAATGAGAAGAAG ATA
NAP1_346	ACGTTGGATGAACTGATTTGACCTGGG TCG	ACGTTGGATGCTGTGTACTGCCCCAAG TTC	cccgTCTTCTCATTCTTAAGAAGT TAC
PCBR_1601	ACGTTGGATGCAGATCACCTGTGCCA ATC	ACGTTGGATGATCATTCGGCCGGTAGA GAG	AGAGAGTCAAGAGGATTAACGT GT
PCBR_1651	ACGTTGGATGTGATTGGCACAGGGTGA TCT	ACGTTGGATGACTGTGATATCACCAC GTC	aATCGCCTTCACCAAGCTCTC
PCBR_1714	ACGTTGGATGGTGGTGATATCGACAGT	ACGTTGGATGACCTTAACGTTGCCAGC	cccTCTTGGTCTGATCCGCCAT

	GGG	TTC	
CDPK_516	ACGTTGGATGGAGATTCTCACAAC TTC	ACGTTGGATGTGGGATGCTGGGATT TA	GCTGGGATTTAGACAGTGGGGC
	CCC	GAC	T
CDPK_589	ACGTTGGATGCCCTTTGTAGTAGTAG TT	ACGTTGGATGAGACAGGATAAGATAG T	tGATAAGATAGTGGCCAGTCA
	TGG	GGC	
SAMS_744	ACGTTGGATGCTTGT CAGGAATCTTTCC	ACGTTGGATGTCTCGTACGCCATTGGA	GAGCCTCTTTCTGTTTTCTGTTGA
	AG	GTG	
SAMS_819	ACGTTGGATGAAGGAGAGCTTTGACTT	ACGTTGGATGTTCAAGAATCTGCCACT	GATCCAGGTTGATGGAAATCAT
	CAG	GCC	
Susy1_1310	ACGTTGGATGTGGAGGGAGAACACATT	ACGTTGGATGGCGATCTGTTTTTTGGCTT	CTGCAGAACATGATGGTGAAT
	CTG	ATG	CAAGAACTTCTCGAGAGTGCAC
Susy1_1510	ACGTTGGATGAGATGATCCAGCTCCTG	ACGTTGGATGTGGGAACCCTATCCAAG	GG
	TTG	AAC	ggTGATTTCTTGATTAGGAGGAT
Susy1_1684	ACGTTGGATGGCAAATGAACAAGTTGA	ACGTTGGATGGGTGGCCAGGTACTTGT	AT
	ATGG	TTT	
Susy1_1733	ACGTTGGATGCCATTCAACTTGTTCA TT	ACGTTGGATGATTTCTTCTCTAGGGCA	CAGGATGTAAACAACCTATAAA
	TGC	CG	CAAGAGGGTCTCATCCAAACGT
Susy3_1008	ACGTTGGATGTGAGAGGATTCAGACAT	ACGTTGGATGATTACGATGAGCCAAGA	TC
	GGC	GGG	cctccAAAAGAAAGACTTACTTTC
Susy3_1641	ACGTTGGATGGCAAGTGACCGAGTTTC	ACGTTGGATGCGTCCATGTGGTGCTAA	CAT
	TTC	AAG	
Susy3_4612	ACGTTGGATGAGGGAAAAAGAGCACA	ACGTTGGATGATTTGCCATGTATACCTG	CACCACATTATACCCACT
	GTTC	CC	
Susy3_593	ACGTTGGATGCCTTAACAGTCACGTTG	ACGTTGGATGGTTTAATACCGTGAGCG	GGTTGCCCTGTTCAGGTCGC
	CAC	AAC	
Susy3_64	ACGTTGGATGCGAACTTTGACCTTGGA	ACGTTGGATGACAGTCAGATCAGCCAG	ACGATTACTCCCACAGAGAT
	GT	AAG	
Susy3_886	ACGTTGGATGCGGTTTTTGGTGGGATTT	ACGTTGGATGCTGCAATACAGAAACGA	GGATCCAAATAACCCAACGCC
	GAC	CAG	AGAAACGACAGTAGTATAAGAA
Susy3_909	ACGTTGGATGCGGTTTTTGGTGGGATTT	ACGTTGGATGCTGCAATACAGAAACGA	AC
	GAC	CAG	
TUB_1475	ACGTTGGATGAAGTTCTGGGAGGTTGT	ACGTTGGATGTAGTACACGTTCACTCG	cGTCACCAGTATAGTTCCCCG
	GTG	CTC	
TUB_1543	ACGTTGGATGTGCGCAGGCTGTCCATA	ACGTTGGATGAGCGAGTGAACGTGTAC	gGTACTACAATGAGGCGAG
	GTC	TAC	

952

953

954 **Appendix A2 Genetic structure of *E. globulus* samples**

955 The genetic structure among the sample of *E. globulus* collected from the discovery population was investigated using the computer program
956 STRUCTURE and a data set derived from 18 microsatellite loci. Of these microsatellites, six (CRC2, CRC5, CRC7, CRC8, CRC10 and CRC
957 11) were designed by Steane et al. (2001), two (EMBRA 10 and EMBRA 11) by Brondani et al. (1998), four (EMBRA23, EMBRA30,
958 EMBRA38 and EMBRA63) by Brondani et al. (2002), two (EMBRA 210 and EMBRA 213) by Brondani et al. (2006), one (Eg131) by
959 Thamarus et al. (2002) and three (EMBRA362, EMBRA712 and EMBRA747) by D. Grattapaglia (pers. comm.). Multiplex PCRs were carried
960 out using a PCR cycle comprising: 15 minutes at 95°C, 30 cycles of 30 seconds at 94°C, 90 seconds at 57°C, 1 minute at 72°C; followed by a
961 final 10 minutes at 60°C. The PCR products were separated on an ABI 377 DNA sequencer at the Australian Genomic Research Facility
962 (AGRF) in Adelaide and analysed using GeneMapper® software (Version 3.7). The SSR loci were highly polymorphic and informative with on
963 average 21.7 alleles per loci.

964

965

966

967

968

969

970

971

972

973 Figure 1. Graph of ΔK calculated from 18 microsatellites for *E. globulus* microsatellite following the method of Evanno et al. (2005). Evanno et
974 al. (2005) provides a method for finding the most likely number of clusters of genetically homogeneous groups of individuals (K). The
975 secondary peak at $K=4$ was deemed the best solution as it made the most biological sense.

976

977

978 Figure 2 Proportion of membership at
 979 $K=4$ for all sampled individuals of *E.*
 980 *globulus* from the discovery population
 981 based on microsatellite data. Each
 982 individual is represented by a thin
 983 vertical line that can be partitioned into
 984 up to four coloured segments
 985 representing the individual's estimated

986 membership (Q) into the K genetic clusters. Individuals are grouped by region (above the figure) and race (below the figure).

987

988

989

990 **References**

991

992 Brondani R, Williams E, Brondani C, Grattapaglia D (2006) A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230
993 microsatellite markers for the genus. *Bmc Plant Biology* 6 (1):20

994 Brondani RPV, Brondani C, Grattapaglia D (2002) Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative
995 microsatellite markers. *Molecular Genetics and Genomics* 267:338-347

996 Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis*
997 and *E. urophylla*. *Theoretical and Applied Genetics* 97 (5-6):816-827

998 Steane DA, Vaillancourt RE, Russell J, Powell W, Marshall D, Potts BM (2001) Development and characterisation of microsatellite loci in *Eucalyptus globulus*
999 (Myrtaceae). *Silvae Genetica* 50 (2):89-91

1000 Thamarus KA, Groom K, Murrell J, Byrne M, Moran GF (2002) A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre, and floral
1001 traits. *Theoretical and Applied Genetics* 104 (2-3):379-387

1002

1003

1004

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Thavamanikumar, S;McManus, LJ;Ades, PK;Bossinger, G;Stackpole, DJ;Kerr, R;Hadjigol, S;Freeman, JS;Vaillancourt, RE;Zhu, P;Tibbits, JFG

Title:

Association mapping for wood quality and growth traits in *Eucalyptus globulus* ssp *globulus* Labill identifies nine stable marker-trait associations for seven traits

Date:

2014-12-01

Citation:

Thavamanikumar, S., McManus, L. J., Ades, P. K., Bossinger, G., Stackpole, D. J., Kerr, R., Hadjigol, S., Freeman, J. S., Vaillancourt, R. E., Zhu, P. & Tibbits, J. F. G. (2014). Association mapping for wood quality and growth traits in *Eucalyptus globulus* ssp *globulus* Labill identifies nine stable marker-trait associations for seven traits. *TREE GENETICS & GENOMES*, 10 (6), pp.1661-1678. <https://doi.org/10.1007/s11295-014-0787-0>.

Persistent Link:

<http://hdl.handle.net/11343/283248>