New Phytologist

Research

The molecular basis of quantitative variation in foliar secondary metabolites in *Eucalyptus globulus*

Carsten Külheim[1], Suat Hui Yeoh[1,2], Ian R. Wallis[1], Shawn Laffan[3], Gavin F. Moran[1] and William J. Foley[1]

[1]Research School of Biology, Australian National University, Canberra 0200 ACT, Australia; [2]Institute of Biological Sciences, Faculty of Science, University of Malaya, Lembah Pantai, 50603 Kuala Lumpur, Malaysia; [3]School of Biological, Earth and Environmental Science, University of New South Wales, Randwick 2052 NSW, Australia

Author for correspondence:
Carsten Külheim
Tel: +61 2 61257190
Email: carsten.kulheim@anu.edu.au

Received: *7 February 2011*
Accepted: *13 April 2011*

New Phytologist (2011) **191**: 1041–1053
doi: 10.1111/j.1469-8137.2011.03769.x

Key words: *Eucalyptus*, flavonoids, genetic association, plant secondary metabolites, plant–herbivore interaction, terpenes.

Summary

• *Eucalyptus* is characterized by high foliar concentrations of plant secondary metabolites with marked qualitative and quantitative variation within a single species. Secondary metabolites in eucalypts are important mediators of a diverse community of herbivores.

• We used a candidate gene approach to investigate genetic associations between 195 single nucleotide polymorphisms (SNPs) from 24 candidate genes and 33 traits related to secondary metabolites in the Tasmanian Blue Gum (*Eucalyptus globulus*).

• We discovered 37 significant associations (false discovery rate (FDR) $Q < 0.05$) across 11 candidate genes and 19 traits. The effects of SNPs on phenotypic variation were within the expected range ($0.018 < r^2 < 0.061$) for forest trees. Whereas most marker effects were nonadditive, two alleles from two consecutive genes in the methylerythritol phosphate pathway (MEP) showed additive effects.

• This study successfully links allelic variants to ecologically important phenotypes which can have a large impact on the entire community. It is one of very few studies to identify the genetic variants of a foundation tree that influences ecosystem function.

Introduction

Variations in the concentration of plant secondary metabolites (PSMs) both within and between species have long been of interest for ecologists because they mediate interactions with a diverse suite of other organisms. Both genetic and environmental effects contribute to variation in the phenotypes of PSMs, but for the past 25 yr environment-based explanations have been most common. This is best summarized by > 2600 citations of two key papers (Bryant *et al.*, 1983; Coley *et al.*, 1985). However, the importance of environmental influences on the concentrations of PSMs has been overemphasized (Hamilton *et al.*, 2001), because a large body of evidence shows that variation in PSMs has a high heritability (Hamilton *et al.*, 2001; Andrew *et al.*, 2007). This implies that there must be specific gene variants that affect the concentration of these PSMs. This does not preclude an environmental contribution through genotype × environment (G × E) interactions. Nonetheless, although few G × E studies are able to identify the specific environmental factors underlying phenotypic differences (Tobler & Carson, 2010), they can indicate the extent to which traits are plastic (Andrew *et al.*, 2010).

Recent syntheses (e.g. LeRoy *et al.*, 2006; Whitham *et al.*, 2006; Bailey *et al.*, 2009) have emphasized the importance of genetic variations in PSMs of foundation tree species in influencing interactions at the population, community and ecosystem levels. These syntheses, which are part of a broader 'genes to ecosystems' framework, are extremely useful in thinking about community ecology and how the effects of individual genes might extend beyond the phenotype of individual plants. Although significant progress is being made in developing the concepts of community genetics, there are few data on the relevant gene variants and little knowledge of the selective pressures on those genes and their evolutionary trajectory.

Many PSMs, including most terpenes, phenolic compounds and formylated phloroglucinol compounds (FPCs) are normally distributed in natural populations (Lawler *et al.*, 2000; Wallis *et al.*, 2003). This suggests that the

concentrations of PSMs are regulated by multiple genes, and indeed where biosynthetic pathways have been elucidated, this has proven to be the case (e.g. glucosinolates: Grubb & Abel, 2006; Halkier & Gershenzon, 2006; terpenes: Wildung & Croteau, 2005; Keeling & Bohlmann, 2006). Allelic variants in these genes potentially affect the concentrations of secondary compounds. There are many selective pressures, including variable responses of herbivores (Iason *et al.*, 2011) and pathogens, that maintain such wide variation in PSM concentrations, but understanding these processes requires first discovering the gene variants that influence these quantitative traits.

Although several studies have identified the genes that account for differences in the profile of secondary compounds within a plant (Albinsky *et al.*, 2010; Padovan *et al.*, 2010), few have addressed the molecular basis of quantitative differences (Wentzell *et al.*, 2007; Chan *et al.*, 2010), with just one recent publication in any species of woody plant (Hall *et al.*, 2011). There are several ways to identify genetic elements that explain variations in the concentration of secondary compounds. For example, several studies (Henery *et al.*, 2007; Freeman *et al.*, 2008; O'Reilly-Wapstra *et al.*, 2011) identified significant quantitative trait loci (QTLs) for terpenes and FPCs in eucalypts – some of the traits studied here. The QTLs span large genomic regions and probably many hundreds of genes (Henery *et al.*, 2007; O'Reilly-Wapstra *et al.*, 2011). In addition, because the QTL mapping population depends on a controlled cross between just two parents, the resulting population represents very little of the variation that occurs in the whole species. By contrast, association or linkage disequilibrium mapping identifies the specific SNPs that correlate with a particular trait, and incorporates a much greater range of the allelic diversity of the target species. Association mapping has become a standard tool to dissect quantitative traits in humans, farmed animals and, recently, forest trees (Thumma *et al.*, 2005; Gonzalez-Martinez *et al.*, 2007, 2008; Eckert *et al.*, 2009). Linkage disequilibrium declines slowly in humans and often extends over 50 kb (Reich *et al.*, 2001), but outcrossing plant species, such as forest trees, including *Eucalyptus*, have low linkage that typically extends only over 500 bp (Thumma *et al.*, 2005; Ingvarsson, 2008). This makes it feasible to attempt to identify quantitative trait nucleotides using association mapping – something which would not be possible in humans or *Arabidopsis*. Recent studies in forest trees have revealed associations of two linked SNPs with a wood property, the microfibril angle in *Eucalyptus nitens* (Thumma *et al.*, 2005). Several studies of conifers (Gonzalez-Martinez *et al.*, 2007, 2008; Eckert *et al.*, 2009) and of poplars

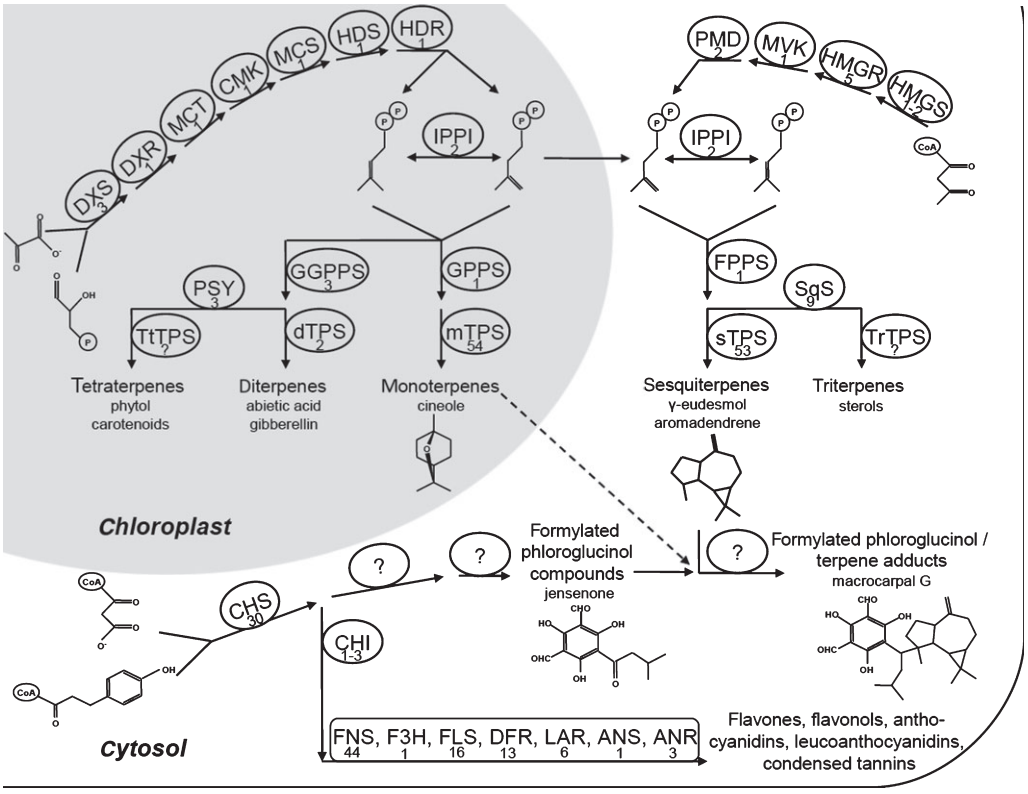

Fig. 1 Schematic of the terpenoid, flavonoid and formylated phloroglucinol compound (FPC) biosynthesis pathway. Starting substrates and examples of end products are shown. Below each enzyme are estimates of gene copy number from the draft genome of *Eucalyptus grandis*. Compartmentalization of pathways is indicated by shading of the chloroplast, and location of synthesis of FPCs is hypothetical.

(Ingvarsson *et al.*, 2008) followed and examined a range of cold tolerance or wood traits.

In this study, we used candidate genes from known biosynthetic pathways (Fig. 1) to identify SNPs that explain variation in the concentrations of PSMs that defend eucalypt foliage against both vertebrate and invertebrate herbivores (Lawler *et al.*, 1998, 2000; Moore *et al.*, 2005; Andrew *et al.*, 2007). In particular, we focus on foliar mono- and sesquiterpenes, because the pathways of terpene synthesis have been well described in other plants (Bohlmann *et al.*, 1998; Rohmer, 1999; McGarvey & Croteau, 1995) and because we know that terpenes in *Eucalyptus* affect ecological processes at both small and large scales; tannin and phenolic compounds produced via the flavonoid pathway and, in particular, a functional assay of those tannins that reduce the availability of plant protein to vertebrate herbivores; and a group of secondary metabolites that are unique to *Eucalyptus*, the FPCs (Fig. 1). A set of 195 SNPs was genotyped in 475 individuals of *Eucalyptus globulus* and the genotypes were tested for associations with 33 traits.

Materials and Methods

Plant material, trait quantification and DNA extraction

Samples of fully expanded, mature foliage of *Eucalyptus globulus* (Labill.) were collected at Latrobe in Tasmania, Australia (S41°14′ E146°25′), from a common-garden experiment derived from open-pollinated seed collected from 46 populations over the geographic range of *E. globulus* (Gardiner & Crawford, 1987, 1988). The experiment was designed as a randomized incomplete block design of five complete replicates and two trees per plot (see Jordan *et al.*, 1994 for further details), and we sampled foliage from the canopy of 511 trees (half-sib families) in September 2006. The samples were frozen immediately at −20°C and stored at −80°C until further use.

Genomic DNA (gDNA) was extracted using a modified CTAB method (Glaubitz *et al.*, 2001) as in (Külheim *et al.*, 2009). DNA concentrations were quantified using a NanoDrop ND-1000 (Thermo Scientific, Wilmington, DE, USA) and each sample was diluted to 50 ng μl^{-1}. We separated DNA on an agarose gel to check its quality.

Sixty chemical compositional traits were considered from amongst three broad classes of PSMs (FPCs, terpenes and various measures of tannins) that have been widely studied in relation to their effect on vertebrate and invertebrate herbivores of *Eucalyptus*. These data are described separately by Wallis *et al.* (I. R. Wallis, unpublished). For the present work, each trait was tested for normal distribution using a probability distribution plot with 95% confidence intervals and a maximum of three outliers as implemented in Genstat, 12th edn (VSN International, Hemel Hempstead, UK) and those that were not normally distributed were transformed ($\log_{10}$) after any values of zero were removed. Three new traits were created arithmetically: the sum of the concentration of all monoterpenes, the sum of all sesquiterpenes and the ratio of the concentration of mono- to sesquiterpenes (Tables S1, S2).

SNP discovery and genotyping

The candidate genes included five from the 2-C-methyl-D-erythritol 4-phosphate pathway (MEP), three from the mevalonate pathway (MVA), 11 further genes involved in terpene biosynthesis and seven genes involved in flavonoid biosynthesis. The SNPs from these 23 loci were discovered by pyrosequencing loci that were amplified from pools of DNA of all 511 individuals of *E. globulus* (Külheim *et al.*, 2009); three more loci were discovered specifically for this work (C Külheim, unpublished). We aimed to select on average one SNP every 100 bp with a minor allele frequency of at least 0.1. Assays were designed for 449 SNPs using the Sequenom MassARRAY platform at the Australian Genome Research Facility (AGRF) and analysed in 475 individuals. Assay primers were designed within 150 bp on each side of a SNP. Problems arose from the high frequency of SNPs in eucalypts with an average of one SNP per 31 bp in *E. globulus* (Külheim *et al.*, 2009), which led to 90 out of 475 designed assays failing *in silico*, and another 90 failing in practice. We only used assays for which at least 75% of individuals had usable data. This resulted in 195 successful SNP assays, which was, relative to other studies, a low success rate (< 50%) (Tables 1, S3).

Genetic structure of the population

Genetic structure in populations can lead to false positives in association mapping, unless it is incorporated into the experimental design. Genetic structure was determined by genotyping 16 microsatellite loci that were distributed across the *Eucalyptus* genome from a subset of 444 individuals that enabled us to determine structural variables for each individual. This dataset has been used to study the spatial genetic structure of *E. globulus* and full details are reported separately by Yeoh *et al.* (S. H. Yeoh, unpublished). Briefly, population membership estimates were derived using a Bayesian model-based method implemented in STRUCTURE, version 2.3.1 (Falush *et al.*, 2003). We used the correlated allele frequencies and admixture model as they give the most meaningful estimates. We therefore conducted the analyses using a discarded burn-in of 100 000 followed by 1 000 000 Markov chain Monte Carlo (MCMC) steps for $K = 1$ to $K = 20$ with 10 replicates for each K-value. Maximal mean posterior probability across replicates and second rate of change in log probability of data according to Evanno *et al.* (2005) showed that five

Table 1 Single nucleotide polymorphism (SNP) genotype assay design across 26 loci, their biosynthetic pathway and assay success rate

Gene	Pathway	Locus length (bp)	Designed SNPs	SNPs/ 100 bp	Assayed SNPs	SNPs/ 100 bp
dxs1	MEP	1826	18	0.99	9	0.49
dxs2	MEP	772	6	0.78	5	0.65
dxr	MEP	2092	16	0.76	13	0.62
hds	MEP	4998	53	1.06	12	0.24
hdr	MEP	3200	15	0.47	10	0.31
ippi	TPS	1736	11	0.63	9	0.52
hmgs	MVA	2463	18	0.73	12	0.49
mvk	MVA	3039	41	1.35	12	0.39
pmd	MVA	1956	24	1.23	16	0.82
ggpps	TPS	1179	4	0.34	4	0.34
gpps	TPS	4691	23	0.49	12	0.26
fpps	TPS	3473	19	0.55	10	0.29
ans	FLAV	1168	7	0.60	1	0.09
anr	FLAV	1079	9	0.83	9	0.83
lar	FLAV	2666	11	0.41	9	0.34
dfr	FLAV	2360	20	0.85	8	0.34
chi	FLAV	1165	3	0.26	2	0.17
chs	FLAV	1488	24	1.61	6	0.40
f3h	FLAV	2018	27	1.34	11	0.55
iso	TPS	3675	23	0.63	3	0.08
psy1	TPS	1561	8	0.51	5	0.32
psy2	TPS	1499	14	0.93	7	0.47
psy3	TPS	1156	9	0.78	7	0.61
smo	TPS	2347	14	0.60	3	0.13
ts2	TPS	2479	13	0.52	0	0.00
ts3	TPS	2648	19	0.72	0	0.00
		58734	449	0.77	195	0.37

MEP, methylerythriol phosphate pathway; MVA, mevalonate pathway; TPS, terpene synthesis pathway; FLAV, flavonoid biosynthesis pathway.

is the optimal number of cluster (K). We therefore selected the admixture proportion for each individual (Q) from the replicate with the highest log probability of data from $K = 5$ as covariates in the association tests (Table S4). The sum of each individual's covariate is one and therefore, by omitting one of the five covariates in the association analysis, we could obtain valid F-statistics.

Test for trait association and linkage

Association analysis was performed using a least-squares fixed effects linear model using the software package 'Trait Analysis by aSSociation, Evolution and Linkage' (TASSEL version 2.1) (Bradbury *et al.*, 2007). There were 416 individuals in all three datasets (genotype, phenotype and genetic structure, Tables S2–S4). The data were analysed using a general linear model (GLM) embedded in TASSEL with four covariates (Q-matrix) and 1000 permutations in a statistical F-test. We used a false discovery rate (FDR)-corrected P-value in our analysis with a cutoff value of $Q < 0.05$. Although population structure is widely known to lead to an increased rate of false positives, our preliminary analysis showed that chemical structure represented by the presence of distort chemotypes (I. R. Wallis *et al.* unpublished) can have a similar effect. Our initial analysis resulted in a high number of positive associations with traits that were absent in individuals of certain chemotypes. Therefore, we removed individuals of chemotype 2 and chemotype 3 (as defined by I. R. Wallis *et al.*, unpublished) for association analysis for traits related to sesquiterpenes and FPCs. For these traits, 332 individuals remained in the statistical analysis. By contrast, phenotypes based around monoterpenes and flavonoids, which did not display chemical structuring, used 416 individuals.

We investigated linkage disequilibrium (LD) between all 195 SNPs using TASSEL. For the calculations of P-values, 1000 permutations were run. Patterns of LD were quantified using the squared allelic correlation coefficient (r^2) and tested for significance using a two-sided Fisher's Exact test.

The geographic patterns of association across the sampling area were assessed using spatial statistics, implemented within a geographic information system (GIS). Specifically, we assessed the trends between populations and regions for two traits (cineole and the ratio of mono- to sesquiterpenes) and two allele frequency data sets (hds2099 and ggpps103). We used the Getis-Ord Gi* hotspot statistic (Ord & Getis, 1995; Laffan, 2002). This statistic assessed the degree to which the values of a spatial subset of samples is greater or less than the average of the data set, and is expressed as a z-score. Those Gi* values that are > 1.96 represent subsets that are significantly greater than would be expected ($\alpha = 0.05$), representing a 'hotspot' of high values. Those Gi* values < −1.96 are significantly less than expected, representing a 'coldspot'. The Gi* statistic was assessed using a moving window approach with two window sizes. Each used a bisquare weighting kernel (see Laffan, 2006), with the sizes extending to 50 and 200 km.

Results

Metabolite data

We used 30 chemical traits studied by Wallis *et al.* (I. R. Wallis, unpublished), as well as three traits that were created arithmetically from the data (Table S1). Of those, 11 were related to FPCs, nine belonged to the flavonoid/tannin group, including various measures of the effect of tannins on the availability of foliar N, and 13 were related to mono- and sesquiterpenes. Compared with the study by Wallis *et al.* (I. R. Wallis, unpublished), 30 chemical traits were not studied here. These were mostly minor terpene traits, with foliar concentrations of < 0.2% dry matter (DM) or traits with discontinuous distributions.

SNP selection and genotyping

Of the 449 SNP assays that were designed to be evenly spaced among 26 genes, 195 assays passed all quality controls and were used in all downstream analysis (Table 1). The success rate within loci varied widely. While six loci had < 25% successful SNP assays and two loci had none, seven loci had > 75% successful SNP assays. We aimed to have a spacing of approximately one SNP in every 100 bp, but the successful assays showed a maximum of 0.83 SNPs per 100 bp in anthocyanidin reductase (*anr*) and 0.82 SNPs per 100 bp in phosphomevalonate kinase (*pmd*). The lowest assayed SNP density was in loci from the terpene synthase family with no successful assays in two monoterpene synthases, *ts2* and *ts3*, and 0.08 SNPs per 100 bp in *iso*, a hemiterpene synthase (Table 1). This was likely because of the high degree of sequence similarity in the very large terpene synthase gene family in *Eucalyptus*.

Trait association analysis

A total of 6435 association tests were performed (195 SNPs and 33 traits) and resulted in 497 positive associations at the significance level of $P = 0.05$. Using 1000 permutations in an *F*-statistic test for correcting the rate of false discoveries

Table 2 List of significant marker–trait pairs, including allele frequency, type of single nucleotide polymorphism (SNP) and association statistics

Trait	Marker	SNP	Frequency	Type	*n*	*F*	*P*	*Q*	r^2
TotFPCs	hmgs1461	C/T	0.12	Exon s	316	7.62	0.0006	0.0010	3.8
TotFPCs	anr338	G/A	0.15	Intron	315	7.26	0.0008	0.0010	3.5
MacA	hds4746	T/A	0.33	Exon ns	314	7.23	0.0009	0.0010	3.1
P27min	hds4216	G/A	0.45	Intron	317	6.63	0.0015	0.0350	3.0
P27min	hds2099	G/T	0.23	Exon ns	312	7.06	0.0010	0.0310	3.3
P27min	hds4746	T/A	0.33	Exon ns	314	7.35	0.0008	0.0010	3.4
P28min	hds1843	A/T	0.37	Intron	277	6.95	0.0011	0.0340	3.4
P28min	hmgs1461	C/T	0.12	Exon s	316	7.51	0.0006	0.0010	3.3
P39min	hds4746	T/A	0.33	Exon ns	314	7.09	0.0010	0.0010	2.9
P48min	dfr1843	A/T	0.37	Intron	277	7.07	0.0010	0.0320	4.2
P48min	anr338	G/A	0.15	Intron	315	7.79	0.0005	0.0010	4.1
aPin	f3h878	G/T	0.13	Intron	385	7.76	0.0005	0.0010	3.7
Cin	hds1181	G/T	0.28	Intron	392	7.25	0.0008	0.0010	2.8
Cin	hds2099	G/T	0.23	Exon ns	390	9.82	0.0001	0.0010	3.8
Cin	hds4216	G/A	0.45	Intron	399	6.52	0.0016	0.0190	2.5
Cin	hds4631	C/T	0.33	Intron	391	8.61	0.0002	0.0010	3.3
Cin	hds4746	T/A	0.33	Exon ns	395	9.75	0.0001	0.0010	3.7
Cin	hds4907	A/G	0.33	Exon s	398	9.26	0.0001	0.0010	3.6
Cin	hdr484	C/T	0.20	Intron	312	7.07	0.0010	0.0010	3.3
SumM	f3h878	G/T	0.13	Intron	385	6.61	0.0015	0.0360	2.9
mon/sesq	ggpps103	T/C	0.17	Exon ns	314	11.93	0.0001	0.0010	6.1
gEudesmol	hmgs1816	C/T	0.47	Intron	311	8.23	0.0003	0.0010	4.4
sumchem1	hmgs1816	C/T	0.47	Intron	311	7.16	0.0009	0.0010	3.5
Sumsesq	hmgs1816	C/T	0.47	Intron	311	7.47	0.0007	0.0010	3.7
AvlN	chs1168	A/G	0.38	Exon ns	391	7.76	0.0005	0.0010	2.7
AvlN	chs570	G/T	0.30	Exon ns	395	6.71	0.0014	0.0140	2.3
AvlN	hds4746	T/A	0.33	Exon ns	395	6.43	0.0018	0.0280	2.2
digNPEG	hds2099	G/T	0.23	Exon ns	390	7.63	0.0006	0.0010	2.3
digNPEG	hds4746	T/A	0.33	Exon ns	395	8.39	0.0003	0.0010	2.5
digNPEG	hds4907	A/G	0.33	Exon s	398	5.96	0.0028	0.0400	1.8
digNPEG	psyB155	G/A	0.46	Exon s	381	8.81	0.0002	0.0010	2.7
digNmPEG	lar1338	C/A	0.31	Intron	383	7.33	0.0007	0.0010	1.8
digNmPEG	psyB155	G/A	0.46	Exon s	381	8.19	0.0003	0.0010	2.1
DMDPEG	mvk2230	A/G	0.33	Exon ns	354	6.74	0.0013	0.0130	2.2
PEG	hmgs123	T/C	0.14	Intron	394	6.2	0.0022	0.0480	2.8
PEG	hmgs577	T/A	0.15	Intron	368	7.66	0.0005	0.0010	3.8
Paqueb	psyB155	G/A	0.46	Exon s	381	8.55	0.0002	0.0010	2.0

s, synonymous; ns, nonsynonymous.

reduced the number of positive associations to 37. Of these, 13 were related to terpene biosynthesis, 13 were associated with flavonoids and the effect of tannins on the availability of foliar N and 11 were associated with FPC (Table 2). The amount of phenotypic variation that could be explained by each polymorphism (r^2) varied between 1.8 and 6.1%. Traits with a more direct link to the biosynthetic pathways, such as the foliar concentration of 1,8-cineole or γ-eudesmol, generally had higher r^2 values than those traits that are more complex, such as the effect of tannins on N availability for mammals. The average amount of phenotypic variation that could be explained by a single allelic variant for FPC-related traits was 3.4%, for terpenes it was 3.8%, and for traits related to the effect of tannins on N availability to mammals it was 2.4%. Several SNPs were associated with more than one trait, giving a total of 20 unique SNPs that were present in the association analysis. Of these, 11 were in introns, three were synonymous SNPs in the exons and six were nonsynonymous (Table 2).

Examples of the effect of different alleles on the concentration of a particular trait are shown in Fig. 2. Nine associations were selected to represent FPCs, terpenes and flavonoid/available N traits from this study. The concentration of total foliar FPCs was affected by a synonymous SNP in the exon of 3-hydroxy-3-methylglutaryl-CoA synthase (hmgs1461) with the heterozygous allele having the lowest concentration and the homozygous CC allele having the highest concentration of FPCs. One specific FPC which elutes at 48 min (FPC 48 min; but which remains uncharacterized) was affected by a SNP in the first intron of anthocyanidin reductase (anr338), where the homozygous allele of AA leads to lower concentrations of the metabolite. The strongest association was between a nonsynonymous SNP in gerany-geranyl pyrophosphate synthase (ggpps103) and the ratio of mono- to sesquiterpenes, where the homozygous CC allele had the highest ratio, homozygous TT the lowest and the heterozygous allele fell between these two extremes. A SNP in the intron of 3-hydroxy-3-methylglutaryl-CoA synthase (hmgs1816) significantly associated with three phenotypes related to sesquiterpenes. The effect on the sum of all sesquiterpenes is shown in Fig. 2, with the homozygous TT allele associated with the highest concentrations of sesquiterpenes.

Two effects of associations between the traits related to the effect of tannins on N availability in leaves are shown in Fig. 2. A nonsynonymous SNP in the exon of chalcone

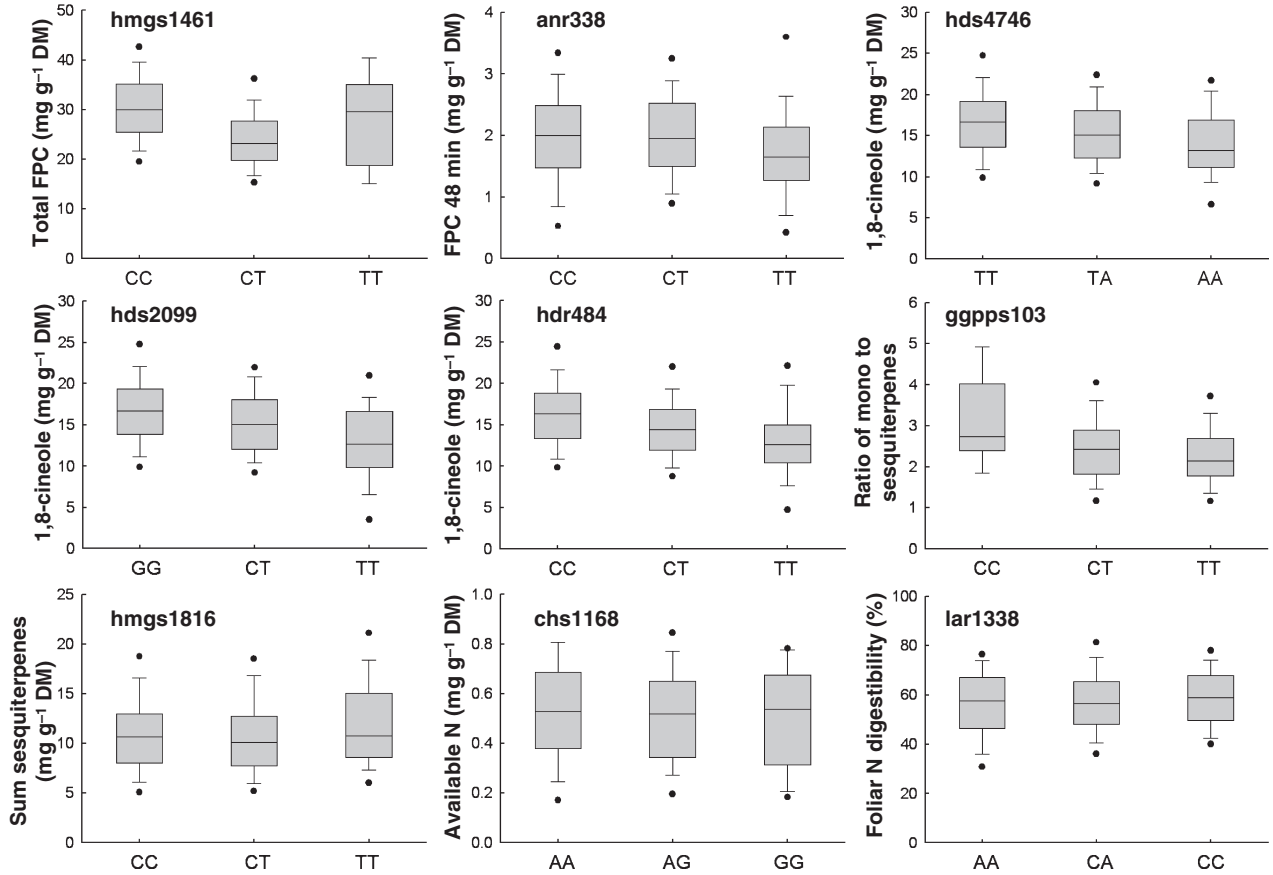

Fig. 2 Examples of marker effects on selected traits, including all three groups of secondary metabolites of *Eucalyptus globulus*. Boxplots of nine selected associations to show the marker–trait effect range. Dots indicate the 95% confidence interval for each population of alleles. DM, dry matter; FPC, formylated phloroglucinol compound.

synthase (chs1168) is associated with the effects of tannins on foliar N, and a SNP in the intron of leucoanthocyanidin reductase (lar1338) is associated with the overall *in vitro* digestibility of N in leaves. In both cases, these traits are dependent on foliar tannins, although the effects are smaller than those of terpene and FPC traits.

A much larger effect was apparent from associations between 1,8-cineole and two genes in the MEP pathway, (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (*hds*) and (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate reductase (*hdr*): hds2099, hds4746 and hdr484. This prompted us to investigate whether the effect of two unlinked variants from different genes could be combined and, if so, what effect this would have on the phenotype. Therefore, allele data for hds2099 and hdr484 were combined by sorting both alleles for each individual into categories (Fig. 3). For hds2099 and hdr484, homozygous GG allele individuals and homozygous CC allele individuals, respectively, had the highest foliar concentrations of 1,8-cineole. Individuals that contain both homozygous alleles have the highest foliar 1,8-cineole concentrations, with a tendency towards the two homozygous TT alleles having lower concentrations (Fig. 3).

Linkage disequilibrium within (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase

Linkage disequilibrium typically decays in forest trees to below r^2 = 0.3 within 500 bp. While our data confirm this (data not shown), the locus *hds* was an exception. There were six significant associations between polymorphisms in *hds* and the foliar concentrations of the monoterpene 1,8-cineole (Table 2). Our investigation of LD within this gene shows that linkage exists over nearly 4 kb with r^2 > 0.3. Between hds1183 and hds4476, the r^2 is 0.53 (Fig. 4). The presence of this degree of linkage makes it more difficult to determine which of the SNPs is most probably causing the variation in phenotype. Two of the six significantly associated SNPs in this gene cause a change in the amino acid (nonsynonymous) of the enzyme, one of the SNPs is synonymous (does not change the amino acid), and three more are found in the introns (Table 2, Fig. 4).

Geographic patterns of traits and associations

There was a significant relationship (r^2 = 0.428; P < 0.001) between variation in foliar 1,8-cineole concentration across its geographic range and the allele frequency for the SNP hds2099 (Fig. 5a). Other trait associations (e.g. ggpps103

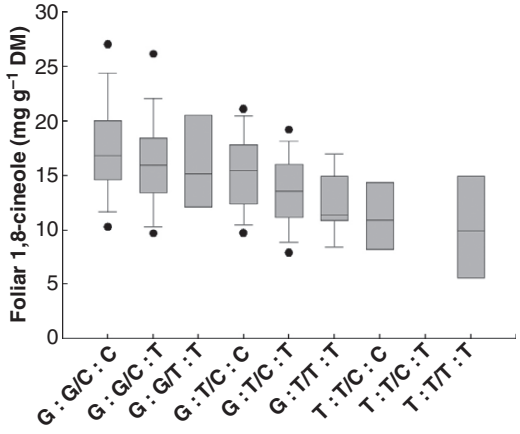

Fig. 3 Boxplot that combines the effects of two allelic variants from (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (*hds*) and (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate reductase (*hdr*). There are nine possible combinations from the two *Eucalyptus globulus* alleles of hds2099 and hdr484. For one combination no boxplot could be produced because the sample size was two. Dots indicate the 95% confidence interval for each allele combination. DM, dry matter.

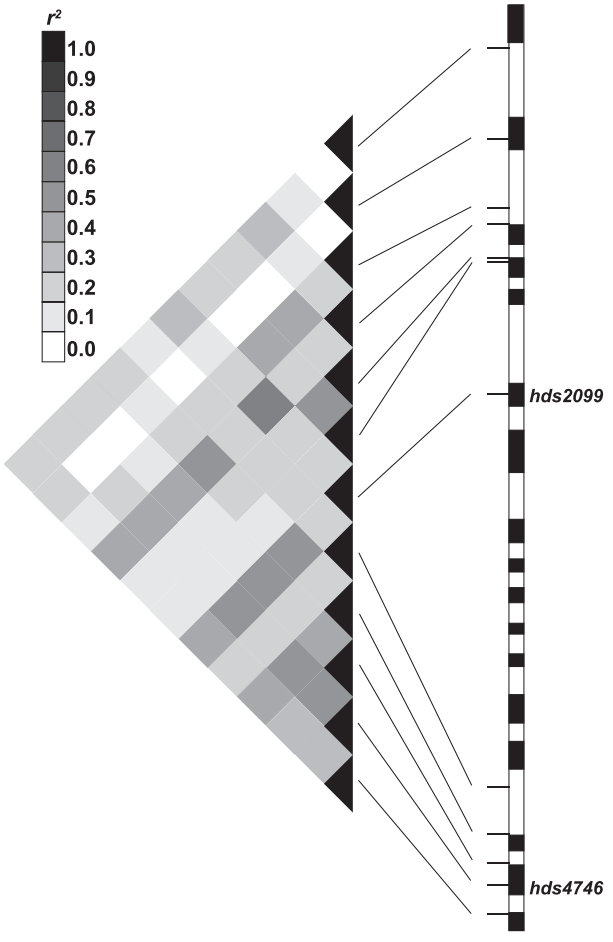

Fig. 4 Exon/intron structure and linkage of single nucleotide polymorphisms (SNPs) within the *Eucalyptus globulus hds* locus. Illustrated are 12 SNPs genotyped in this locus. Linkage across (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (*hds*) is shown as well as the locations of exons (closed bars) and introns (open bars). The locations of the two nonsynonymous SNPs are shown in exons 6 and 16.

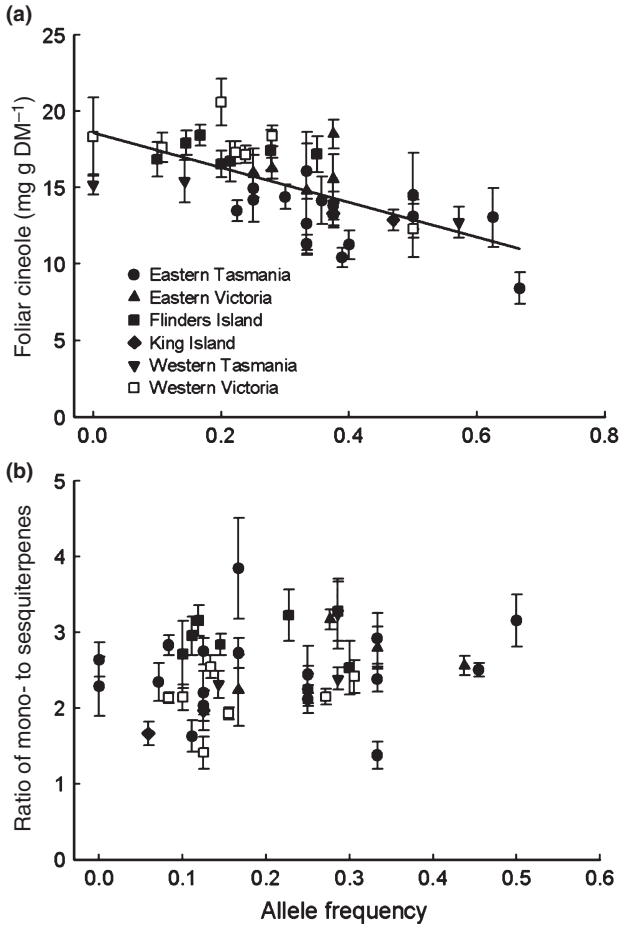

Fig. 5 Relationship between variation in foliar 1,8-cineole concentration and the allele frequency for hds2099. (a) Cineole content (average per *Eucalyptus globulus* population) vs the allele frequency of the minor allele of hds2099 (average within each population); (b) ratio of mono- to sesquiterpenes vs the minor allele frequency of ggpps103. DM, dry matter.

and the ratio of mono- to sesquiterpenes) showed no relationship with allele frequency across the geographic range (Fig. 5b). Patterns were visible for regional variation; for example, eastern Tasmania tended to have lower foliar 1,8-cineole concentration and higher allele frequencies, while western Victoria tended to have higher foliar 1,8-cineole and lower allele frequencies. To further analyse the relationship between these two allelic variants and traits, we employed Gi* hotspot statistics.

Fig. 6 shows the Gi* hotspot statistic results. Foliar concentrations of 1,8-cineole have a clear north–south cline from hotspot to coldspot, with highest above-average populations in western Victoria and the Furneaux group. The east coast of Tasmania has the lowest values, with the exception of some outliers on North Maria Island, South Bruny Island and in Taranna. Most of these are not significant for the 50 km window. These tendencies are mirrored by the allele frequency of hds2099, with below-average allele frequency in areas of above-average foliar cineole concentration and *vice versa*. However, only two of these associations are significant. One coldspot from Cape Patton to Lorne in Victoria is significant at a window size of 50 km and corresponds to a cineole hotspot, and one hotspot in Mayfield, Triabunna and North Maria Island (east Tasmania) at the 200 km window size falls within the cineole coldspot.

The ratio of mono- to sesquiterpenes shows less geographic structure, but does have two significant clusters for the 200 km window size (west Victoria and King Island is a coldspot and the Furneaux group is a hotspot). Similar to Fig. 5(b), the allele frequency Gi* results of ggpps103 do not correspond with the ratio of mono- to sesquiterpenes. The significant clusters are all at the smaller window size assessed and do not correspond to any geographic clines, suggesting that individuals play a larger role in this trait–genotype association.

Discussion

This is one of the very few studies to identify specific allelic variants correlated with changes in the concentration of ecologically significant secondary metabolites. It provides an opportunity to examine how these SNPs and their associated traits vary across landscapes and so influence community organization in *Eucalyptus* forests. Quantitative variation in secondary metabolites in *Eucalyptus* is of major ecological importance and earlier studies have demonstrated that fine-scale genetic variation, particularly in dominant trees such as *E. globulus*, can have extended consequences on associated foliar insect, fungal and litter communities (Barbour *et al.*, 2009a,b). In this work, we have shown that specific variants of genes in the biosynthetic pathways of PSMs are associated with quantitative variation in PSMs and that, in some cases, these variants occur predictably across the landscape. *E. globulus* has a high degree of spatial genetic structure (Steane *et al.*, 2006), and tests for associations with phenotypic traits require this structure to be taken into account. Tests without incorporation of spatial structure resulted in hundreds of false-positive associations. After accounting for genetic structure, we discovered 37 significant associations between 11 candidate genes and 19 quantitative traits of foliar PSMs.

Terpenes

Although several studies have identified variations in terpene synthase genes that lead to changes in the profile of foliar essential oils (e.g. Kollner *et al.*, 2004; Keeling *et al.*, 2008; Padovan *et al.*, 2010), there has been little attention given to the molecular differences that lead to changes in their concentration (Hall *et al.*, 2011). Monoterpenes are synthesized in the chloroplast from D-glyceraldehyde-3-phospate pyruvate via the MEP pathway. The first two steps

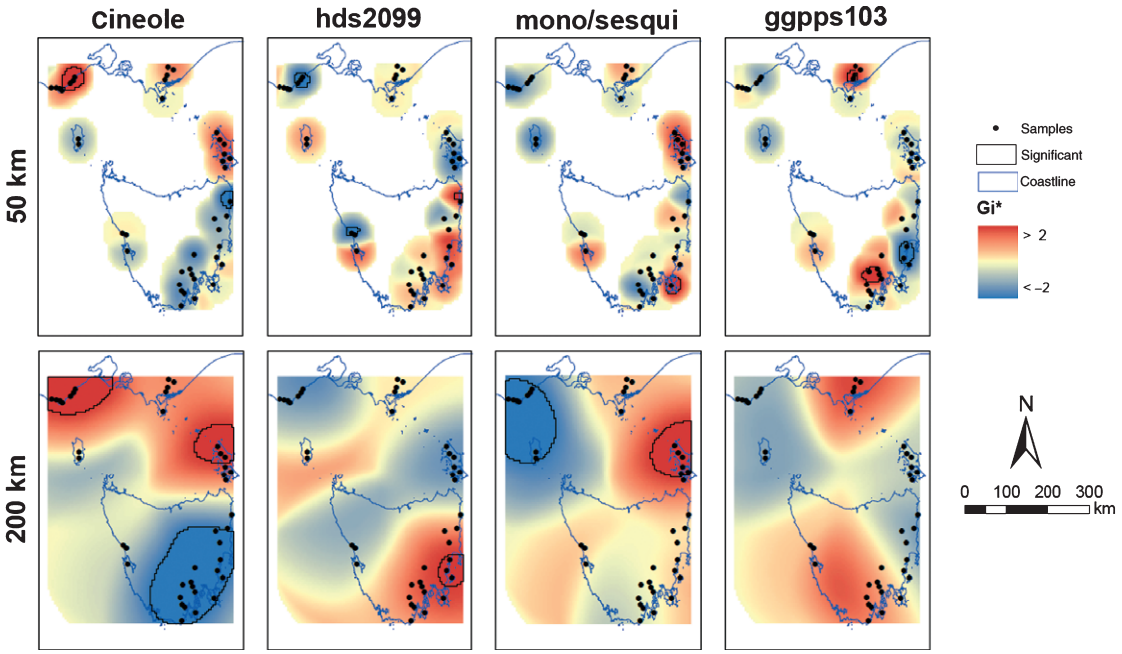

Fig. 6 Graphic depiction of Gi* hotspot statistics. The upper panel has a 50 km radius around each *Eucalyptus globulus* population; the lower 200 km. Lines around areas represent significant deviations from the average for the area. Cineole and hds2099 have opposing directions (high cineole and low allele frequency), ratio of mono- to sesquiterpenes (mono/sesqui) and ggpps103 have the same direction. The 200 km analysis gets rid of outliers, while they are still visible in the 50 km analysis (e.g. Pepper Hill in hds2099, North Maria Island in cineole).

of the pathway, 1-deoxyxylulose-5-phosphate synthase (*dxs*) and 1-deoxy-D-xylulose 5-phosphate reductoisomerase (*dxr*), have been shown to influence foliar terpene yield in several other plants (Wildung & Croteau, 2005; Battilana *et al.*, 2009). Based on these data, we expected that variants in *dxr* and *dxs* would be important loci for monoterpene concentration in *Eucalyptus* as well, but this was not the case. We tested associations between measures of foliar terpenes and 27 SNPs of two *dxs* homologues and *dxr* and found no significant associations ($Q < 0.05$) to any trait. Surprisingly, in eucalypts, the last two steps of the MEP pathway appear to influence the foliar concentration of monoterpenes. Multiple SNPs in *hds* and one SNP in *hdr* were strongly associated with variations in the foliar concentration of 1,8-cineole, which is the major monoterpene in *E. globulus*. The high degree of linkage between SNPs in *hds* meant that it was not possible to identify which SNP directly influences the foliar cineole concentration. It is possible that the two nonsynonymous SNPs (hds2099 and hds4746) directly change the flux of metabolites through the MEP pathway, but in the absence of a crystal structure of HDS we are unable to speculate on the positions of the amino acid changes. We can, however, compare the occurrence of these SNPs with other species with a known *hds* sequence. Seventeen of 18 higher plant species for which there is publicly available sequence data for *hds* have a methionine at the hds2099 position, as do most of the green algae, with only one each having isoleucine or valine (Fig. S1). Four other *Eucalyptus* species that we have investigated are also homozygous for methionine (Fig. S1). This suggests that the mutation leading to an isoleucine has occurred relatively recently, especially as > 300 individuals of *Eucalyptus nitens*, a close relative of *E. globulus*, which separated < 4 Ma from *E. globulus* (Crisp *et al.*, 2004), also lack this mutation. The mutation could have arisen in an isolated population, possibly where selective pressures from herbivores were lower, and then been carried by wide-ranging vectors such as the swift parrot (*Lathamus discolor*) (Hingston *et al.*, 2004) to other populations where it was maintained because of balanced selection towards growth and against biotic defence. The second nonsynonymous SNP at hds4746 results in an amino acid that is less conserved than that seen in hds2099 and the majority of higher plants have phenylalanine, four other *Eucalyptus* species have a leucine and green algae have phenylalanine, leucine or threonine and only *E. globulus* has an allele that leads to methionine (Fig. S2).

In contrast to the monoterpenes, sesquiterpenes are synthesized from acetyl-CoA in the cytosol via the MVA pathway. Additional to the isopentyl pyrophosphate (IPP) produced through this pathway, it is believed that some IPP is exported from the chloroplast (Laule *et al.*, 2003). A SNP in the enzyme of the first committed step of the MVA pathway (hydroxymethylglutaryl CoA synthase, *hmgs*) (hmgs1816) associated with variations in the concentration of three sesquiterpene traits, including the sum of all

sesquiterpenes and the concentration of γ-eudesmol, the major foliar sesquiterpene in *E. globulus*. Often metabolite flux through a biosynthetic pathway is regulated in the early steps of the pathway, as is the case for the MEP pathway in other plant species (Enfissi *et al.*, 2005; Xie *et al.*, 2008; Battilana *et al.*, 2009). In *Eucalyptus*, *hmgs* appears to influence metabolite flux through the MVA pathway, with hmgs1816 playing an important regulatory role. Exactly what that role is (e.g. an influence on gene expression) remains to be confirmed by specific characterization of the SNP.

The ratio of mono- to sesquiterpenes in leaves is influenced by a variety of factors, including the availability of substrate for geranyl pyrophosphate synthase (*gpps*) in the chloroplast and farnesyl pyrophosphate synthesis (*fpps*) in the cytosol. Both enzymes use IPP and its isomer, dimethylallyl pyrophosphate (DMAPP), at varying ratios (Hugueney & Camara, 1990; Bouvier *et al.*, 2000). In the chloroplast, two enzymes compete for the available IPP and DMAPP, *gpps* and geranylgeranyl pyrophosphate synthase (*ggpps*) (Bouvier *et al.*, 2000; Allen & Banthorpe, 1981). A change in the kinetics of *ggpps* could lead to a change in the availability of the substrate pool for *gpps* and therefore indirectly influence the ratio of mono- to sesquiterpenes. SNP ggpps103, which is a nonsynonymous SNP in exon 1 of *ggpps*, has a strong effect on the ratio of mono- to sesquiterpenes (Table 2) and we speculate that this is a result of changes in enzyme kinetics of GGPPS. A recent QTL study in *E. globulus* found a major QTL between a small region on linkage group 6 and the foliar concentration of six sesquiterpenes, one monoterpene and the sum of mono- and sesquiterpenes (O'Reilly-Wapstra *et al.*, 2011). We found that one genomic copy of *ggpps* is in close proximity to that QTL (data not shown) and speculate that it may be the cause of the QTL. However, it was not the same genomic copy as the one genotyped in this study, which is located on linkage group 8. Other candidate genes investigated in this study did not map to QTLs that have been published (Henery *et al.*, 2007; Freeman *et al.*, 2008; O'Reilly-Wapstra *et al.*, 2011) for eucalypts. This is not surprising, however, as these QTL studies are based on the cross of two individuals and therefore only represent the phenotypic and genetic diversity of the parents. Our study instead investigates the genetic and phenotypic variation of the species.

Formylated phloroglucinol compounds

Variations in concentrations of FPCs are major factors affecting the feeding behaviour of marsupials and some species of insect herbivores on *Eucalyptus* (Moore *et al.*, 2005; Andrew *et al.*, 2007) and so play an important role in Australian ecosystems. FPCs are formed by a Diels–Alder condensation of a terpene and phloroglucinol, which in turn is synthesized from phloretin, a dihydrochalcone (Ghisalberti, 1996; Singh *et al.*, 2010). The majority of FPCs that occur in *E. globulus* have a sesquiterpene moiety as part of their structure and so genes from the MVA pathway, as well as *fpps*, are suitable candidates in an association study. Two associations were discovered between *hmgs* and the sum of FPCs as well as the uncharacterized FPC that elutes at 28 min. Although we expected to find associations with genes from the early steps of the flavonoid pathway, such as chalcone synthase (*chs*), this was not the case. Chalcone synthases form a large gene family in rice (Goff *et al.*, 2002) and preliminary analysis of the *Eucalyptus grandis* genome (C Külheim, unpublished) has identified > 30 copies of this gene. This may explain why we found no associations of the one copy of *chs* analysed here with any FPCs.

Compositional and functional traits of tannins and flavonoids

The effect of *Eucalyptus* tannins in decreasing the availability of foliar N for mammalian herbivores has major effects on animal populations (DeGabriel *et al.*, 2008). A variety of direct, chemical measures of tannins and their effectiveness at binding some proteins have not proven to be robust measures of ecological processes in *Eucalyptus* forests (Cork & Catling, 1996). By contrast, a new integrative measure called 'available nitrogen' (DeGabriel *et al.*, 2008) has proven to be ecologically informative (DeGabriel *et al.*, 2009; Wallis *et al.*, 2010). Available N integrates the effects of variations in foliar N, the digestibility of the overall dry matter of a leaf and the effect of tannins in binding some of the protein. It is thus a widely applicable and ecologically relevant measure. Nonetheless, while integrative measures make sense ecologically, the more removed these are from single compounds whose biosynthetic pathway is known, the harder it is to choose appropriate candidate genes and to interpret the associations between leaf traits and gene variants. Eucalypts have a complex mixture of flavonoids and we believe that this is why the observed effects of single polymorphisms were low. For example, two SNPs in chalcone synthase (*chs*) associated with foliar concentrations of available N, but these explained only 2.7% and 2.3% of the phenotypic variation, respectively.

Associations with *hds* are found throughout all three metabolite groups that were investigated in this study. This may be the result of a strong influence of *hds* on the concentration of monoterpenes and particularly 1,8-cineole, which leads to different allocations of carbon in the cell. Monoterpenes in *Eucalyptus* leaves can make up to 20% of dry matter, and the 2.5-fold variation of foliar 1,8-cineole found in *E. globulus* could lead to changes in carbon allocation across all groups of secondary metabolites, thereby explaining why *hds* is found to associate with traits as diverse as available N and FPCs.

To date, much discussion involving 'genes to ecosystems' has not been strongly linked to specific genes or gene variants (Bailey *et al.*, 2009). Clearly, the significant genomic resources available for *Populus* and, to a lesser extent, *Eucalyptus*, as well as their keystone role in many ecosystems, make these species the best places to pursue these questions. Although association genetics is clearly a powerful approach through which to identify key allelic variants, relying entirely on candidate genes will quickly become limiting. While in this study, structural genes from the biosynthetic pathways of secondary metabolites were investigated, this does not give a complete picture and could be expanded to transcriptional and post-transcriptional regulators. Current developments in next-generation sequencing allow for the genotyping of whole, small, genomes, or enriched parts of medium to large genomes, but this would limit the number of individuals that can be used. Nucleic acid tags for multiplexing of next-generation sequencing have been developed (e.g. Meyer *et al.*, 2008) and will enable the sequencing of dozens of individuals simultaneously. The second challenge is to develop ways to combine the many small contributions of individual SNPs to explain a larger proportion of a particular phenotype. The largest effect that we identified explained only 6% of the phenotypic variation. Although we are certain that there are other as yet unknown genes that contribute to variation in the traits that we have studied, combining many small-effect SNPs is difficult. Combining two SNPs, as we did in Fig. 3, was helpful in increasing the amount of variation explained, but it is likely that there are many SNPs of small effect that contribute to variation in quantitative traits. Recently, Yang *et al.* (2010) developed a method to combine the effects of thousands of markers to explain a much greater proportion of the variation in human height than had previously been possible and these approaches have potential in plant science as well.

Acknowledgements

The samples used in this work were collected jointly with CSIRO and we thank John Owen, Fred Ford and Frances Marsh for their help in the field. We are grateful to Gunns Ltd for permission to sample the experimental plantation at Latrobe. The work was funded by an Australian Research Council Linkage Grant to W.J.F. (LP0667708) with the active partnership of Oji Forests and Forests NSW and supplementary support of the Australian National University.

References

Albinsky D, Sawada Y, Kuwahara A, Nagano M, Hirai A, Saito K, Hirai MY. 2010. Widely targeted metabolomics and coexpression analysis as tools to identify genes involved in the side-chain elongation steps of aliphatic glucosinolate biosynthesis. *Amino Acids* **39**: 1067–1075.

Allen BE, Banthorpe DV. 1981. Terpene biosynthesis. 28. Partial-purification and properties of prenyltransferase from *Pisum sativum*. *Phytochemistry* **20**: 35–40.

Andrew RL, Wallis IR, Harwood CE, Foley WJ. 2010. Genetic and environmental contributions to variation and population divergence in a broad-spectrum foliar defence of *Eucalyptus tricarpa*. *Annals of Botany* **105**: 707–717.

Andrew RL, Wallis IR, Harwood CE, Henson M, Foley WJ. 2007. Heritable variation in the foliar secondary metabolite sideroxylonal in *Eucalyptus* confers cross-resistance to herbivores. *Oecologia* **153**: 891–901.

Bailey JK, Hendry AP, Kinnison MT, Post DM, Palkovacs EP, Pelletier F, Harmon LJ, Schweitzer JA. 2009. From genes to ecosystems: an emerging synthesis of eco-evolutionary dynamics. *New Phytologist* **184**: 746–749.

Barbour RC, Baker SC, O'Reilly-Wapstra JM, Harvest TM, Potts BM. 2009a. A footprint of tree-genetics on the biota of the forest floor. *Oikos* **118**: 1917–1923.

Barbour RC, O'Reilly-Wapstra JM, De Little DW, Jordan GJ, Steane DA, Humphreys JR, Bailey JK, Whitham TG, Potts BM. 2009b. A geographic mosaic of genetic variation within a foundation tree species and its community-level consequences. *Ecology* **90**: 1762–1772.

Battilana J, Costantini L, Emanuelli F, Sevini F, Segala C, Moser S, Velasco R, Versini G, Grando M. 2009. The 1-deoxy-d -xylulose 5-phosphate synthase gene co-localizes with a major QTL affecting monoterpene content in grapevine. *Theoretical and Applied Genetics* **118**: 653–669.

Bohlmann J, Meyer-Gauen G, Croteau R. 1998. Plant terpenoid synthases: molecular biology and phytogenetic analysis. *Proceedings of the National Academy of Sciences, USA* **95**: 4126.

Bouvier F, Suire C, d'Harlingue A, Backhaus RA, Camara B. 2000. Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. *Plant Journal* **24**: 241–252.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. Tassel: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**: 2633–2635.

Bryant JP, Chapin FS, Klein DR. 1983. Carbon nutrient balance of boreal plants in relation to vertebrate herbivory. *Oikos* **40**: 357–368.

Chan EKF, Rowe HC, Kliebenstein DJ. 2010. Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics* **185**: 991–1007.

Coley PD, Bryant JP, Chapin FS. 1985. Resource availability and plant antiherbivore defense. *Science* **230**: 895–899.

Cork SJ, Catling PC. 1996. Modelling distributions of arboreal and ground-dwelling mammals in relation to climate, nutrients, plant chemical defences and vegetation structure in the eucalypt forests of southeastern Australia. *Forest Ecology and Management* **85**: 163–175.

Crisp M, Cook L, Steane D. 2004. Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present day communities? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **359**: 1551–1571.

DeGabriel JL, Moore BD, Foley WJ, Johnson CN. 2009. The effects of plant defensive chemistry on nutrient availability predict reproductive success in a mammal. *Ecology* **90**: 711–719.

DeGabriel JL, Wallis IR, Moore BD, Foley WJ. 2008. A simple, integrative assay to quantify nutritional quality of browses for herbivores. *Oecologia* **156**: 107–116.

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV, Clair JBS, Neale DB. 2009. Asssociation genetics of coastal douglas fir (*Pseudotsuga menziesu* var. Menziesii, Pinaceae). I. Cold-hardiness related traits. *Genetics* **182**: 1289–1302.

Enfissi EMA, Fraser PD, Lois L-M, Boronat A, Schuch W, Bramley PM. 2005. Metabolic engineering of the mevalonate and non-mevalonate isopentenyl diphosphate-forming pathways for the production of health-promoting isoprenoids in tomato. *Plant Biotechnology Journal* **3**: 17–27.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**: 2611–2620.

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.

Freeman JS, O'Reilly-Wapstra JM, Vaillancourt RE, Wiggins N, Potts BM. 2008. Quantitative trait loci for key defensive compounds affecting herbivory of eucalypts in Australia. *New Phytologist* **178**: 846–851.

Gardiner CA, Crawford DA. 1987. *1987 seed collections of* Eucalyptus globulus *subsp.* globulus *for tree improvement purposes*. Canberra, Australia: CSIRO Division of Forest Research.

Gardiner CA, Crawford DA. 1988. *1988 seed collections of* Eucalyptus globulus *subsp.* globulus *for tree improvement purposes*. Canberra, Australia: CSIRO Division of Forest research.

Ghisalberti EL. 1996. Bioactive acylphloroglucinol derivatives from *Eucalyptus* species. *Phytochemistry* **41**: 7–22.

Glaubitz JC, Emebiri LC, Moran GF. 2001. Dinucleotide microsatelites from *Eucalyptus sieberi*: inheritance, diversity, and improved scoring of single-based differences. *Genome* **44**: 1041.

Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H *et al.* 2002. A draft sequence of the rice genome (*Oryza sativa* l. ssp. *japonica*). *Science* **296**: 92–100.

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.

Grubb CD, Abel S. 2006. Glucosinolate metabolism and its control. *Trends in Plant Science* **11**: 89–100.

Halkier BA, Gershenzon J. 2006. Biology and biochemistry of glucosinolates. *Annual Review of Plant Biology* **57**: 303–333.

Hall DE, Robert JA, Keeling CI, Domanski D, Quesada AL, Jancsik S, Kuzyk MA, Hamberger B, Borchers CH, Bohlmann J. 2011. An integrated genomic, proteomic and biochemical analysis of (+)-3-carene biosynthesis in Sitka spruce (*Picea sitchensis*) genotypes that are resistant or susceptible to white pine weevil. *Plant Journal* **65**: 936–948.

Hamilton JG, Zangerl AR, DeLucia EH, Berenbaum MR. 2001. The carbon-nutrient balance hypothesis: its rise and fall. *Ecology Letters* **4**: 86–95.

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

Hingston AB, Gartrell BD, Pinchbeck G. 2004. How specialized is the plant-pollinator association between *Eucalyptus globulus* ssp. *globulus* and the swift parrot *Lathamus discolor*? *Austral Ecology* **29**: 624–630.

Hugueney P, Camara B. 1990. Purification and characterization of farnesyl pyrophosphate synthase from *Capsicum annuum*. *FEBS Letters* **273**: 235–238.

Iason GR, O'Reilly-Wapstra J, Brewer M, Summers RW, Moore BD. 2011. Do multiple herbivores maintain chemical diversity of Scots pine monoterpenes? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **366**: 1337–1345.

Ingvarsson PK. 2008. Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics* **180**: 329–340.

Ingvarsson PK, Garcia MV, Luquez V, Hall D, Jansson S. 2008. Nucleotide polymorphism and phenotypic associations within and around the phytochrome b2 locus in European aspen (*Populus tremula*, Salicaceae). *Genetics* **178**: 2217–2226.

Jordan GJB, Nolan MF, Tilyard P, Potts BM. 1994. Identification of races in *Eucalyptus globulus* ssp. *globulus* based on growth traits in Tasmania and geographic distribution. *Silvae Genetica* **43**: 292–298.

Keeling CI, Bohlmann J. 2006. Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. *New Phytologist* **170**: 657–675.

Keeling CI, Weisshaar S, Lin RPC, Bohlmann J. 2008. Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proceedings of the National Academy of Sciences, USA* **105**: 1085–1090.

Kollner TG, Schnee C, Gershenzon J, Degenhardt J. 2004. The variability of sesquiterpenes cultivars is controlled by allelic emitted from two *Zea mays* variation of two terpene synthase genes encoding stereoselective multiple product enzymes. *Plant Cell* **16**: 1115–1131.

Külheim C, Yeoh SH, Maintz J, Foley WJ, Moran GF. 2009. Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. *BMC Genomics* **10**: 11.

Laffan SW. 2002. Using process models to improve spatial analysis. *International Journal of Geographical Information Science* **16**: 245–257.

Laffan SW. 2006. Assessing regional scale weed distributions, with an Australian example using *Nassella trichotoma*. *Weed Research* **46**: 194–206.

Laule O, Furholz A, Chang HS, Zhu T, Wang X, Heifetz PB, Gruissem W, Lange BM. 2003. Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* **100**: 6866–6871.

Lawler IR, Foley WJ, Eschler BM. 2000. Foliar concentration of a single toxin creates habitat patchiness for a marsupial folivore. *Ecology* **81**: 1327–1338.

Lawler IR, Foley WJ, Eschler BM, Pass DM, Handasyde K. 1998. Intraspecific variation in *Eucalyptus* secondary metabolites determines food intake by folivorous marsupials. *Oecologia* **116**: 160–169.

LeRoy CJ, Whitham TG, Keim P, Marks JC. 2006. Plant genes link forests and streams. *Ecology* **87**: 255–261.

McGarvey DJ, Croteau R. 1995. Terpenoid metabolism. *Plant Cell* **7**: 1015–1026.

Meyer M, Stenzel U, Hofreiter M. 2008. Parallel tagged sequencing on the 454 platform. *Nature Protocols* **3**: 267–278.

Moore BD, Foley WJ, Wallis IR, Cowling A, Handasyde KA. 2005. *Eucalyptus* foliar chemistry explains selective feeding by koalas. *Biology Letters* **1**: 64–67.

O'Reilly-Wapstra JM, Freeman JS, Davies NW, Vaillancourt RE, Fitzgerald H, Potts BM. 2011. Quantitative trait loci for foliar terpenes in a global eucalypt species. *Tree Genetics & Genomes*. doi: 10.1007/s11295-010-0350-6.

Ord JK, Getis A. 1995. Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis* **27**: 286–306.

Padovan A, Keszei A, Kollner TG, Degenhardt J, Foley WJ. 2010. The molecular basis of host plant selection in *Melaleuca quinquenervia* by a successful biological control agent. *Phytochemistry* **71**: 1237–1244.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R *et al.* 2001. Linkage disequilibrium in the human genome. *Nature* **411**: 199–204.

Rohmer M. 1999. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Natural Product Reports* **16**: 565–574.

Singh IP, Sidana J, Bharate SB, Foley WJ. 2010. Phloroglucinol compounds of natural origin: synthetic aspects. *Natural Product Reports* **27**: 393–416.

Steane DA, Conod N, Jones RC, Vaillancourt RE, Potts BM. 2006. A comparative analysis of population structure of a forest tree, *Eucalyptus*

globulus (Myrtaceae), using microsatellite markers and quantitative traits. *Tree Genetics and Genomes* **2**: 30–38.

Thumma BR, Nolan MF, Evans R, Moran GF. 2005. Polymorphisms in cinnamoyl CoA reductase (ccr) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* **171**: 1257–1265.

Tobler M, Carson EW. 2010. Environmental variation, hybridization, and phenotypic diversification in *Cuatro cienegas* pupfishes. *Journal of Evolutionary Biology* **23**: 1475–1489.

Wallis IR, Herlt AJ, Eschler BM, Takasaki M, Foley WJ. 2003. Quantification of sideroxylonals in *Eucalyptus* foliage by high-performance liquid chromatography. *Phytochemical Analysis* **14**: 360–365.

Wallis IR, Nicolle D, Foley WJ. 2010. Available and not total nitrogen in leaves explains key chemical differences between the eucalypt subgenera. *Forest Ecology and Management* **260**: 814–821.

Wentzell AM, Rowe HC, Hansen BG, Ticconi C, Halkier BA, Kliebenstein DJ. 2007. Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. *PLoS Genetics* **3**: 1687–1701.

Whitham TG, Bailey JK, Schweitzer JA, Shuster SM, Bangert RK, Leroy CJ, Lonsdorf EV, Allan GJ, DiFazio SP, Potts BM *et al.* 2006. A framework for community and ecosystem genetics: from genes to ecosystems. *Nature Reviews Genetics* **7**: 510–523.

Wildung MR, Croteau RB. 2005. Genetic engineering of peppermint for improved essential oil composition and yield. *Transgenic Research* **14**: 365–372.

Xie Z, Kapteyn J, Gang DR. 2008. A systems biology investigation of the MEP/terpenoid and shikimate/phenylpropanoid pathways points to multiple levels of metabolic control in sweet basil glandular trichomes. *Plant Journal* **54**: 349–361.

Yang JA, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW *et al.* 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**: 565–569.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Alignment of hds region including single nucleotide polymorphism hds2099 in comparison to other plant species.

Fig. S2 Alignment of hds region including single nucleotide polymorphism hds4746 in comparison to other plant species.

Table S1 Trait description

Table S2 Raw phenotype data

Table S3 Allele data for all 195 genotyped single nucleotide polymorphisms

Table S4 All genetic structure data (*Q*-values)

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.