Theor Appl Genet (1999) 99:1207–1215

M. Shepherd · J.X. Chaparro · R. Teasdale

Genetic mapping of monoterpene composition in an interspecific eucalypt hybrid

Received: 20 November 1998 / Accepted: 16 June 1999

Abstract Genetic control of foliar oil composition was investigated amongst half-sib progeny of an interspecific eucalypt hybrid. The oil was found to be largely composed of the monoterpenes, limonene, α–pinene, γ–terpinene, 1,8 cineole and *p*-cymene. Due to difficulties in the interpretation of the compositional data based on raw proportions, further analysis was conducted using log-ratio variables. A high degree of intercorrelation amongst log-ratios was thought to be a consequence of commonality in the biosynthetic origins of the monoterpenes. Quantitative trait locus (QTL) analysis of log-ratio variables indicated that a significant (68–81%) proportion of the variation in four out of the ten possible log-ratios were controlled by a single genomic region of the maternal *Eucalyptus grandis* parent. The impact of this genomic region upon oil composition was thought to be a consequence of a gene, or genes, controlling the production of limonene, as limonene was the predominant oil constituent in many hybrid individuals and was common to all log-ratios associated with the identified genomic region.

Key words Monoterpenes quantitative trait loci analysis · *Eucalyptus* intercorrelated variables

Communicated by M.A. Saghai Marooff

M. Shepherd (✉) · J.X. Chaparro · R. Teasdale
Forbio Inc., 52 Douglas St, Milton, Brisbane, Queensland 4064,
PO Box 2256, Milton BC 4064, Australia

Present addresses:
M. Shepherd, Cooperative Research Centre for Sustainable Production Forestry, Centre for Plant Conservation Genetics, Southern Cross University, PO Box 157,
Lismore, New South Wales 2480, Australia
e-mail: mshepher@scu.edu.au
Fax: +61 2 66222080

J.X. Chaparro, USDA Horticulture Research Laboratories,
2120 Camden Rd Orlando, Florida, USA

Introduction

In the past decade, the introduction of PCR-based markers and the invention of techniques to reveal polymorphism without prior sequence knowledge, such as randomly amplified polymorphic DNA (RAPD), have facilitated the generation of genetic maps for a number of previously unmapped forest species (e.g. Grattapaglia and Sederoff 1994). Often the motivation for the production of genetic maps in domestic animals and crops has been to allow complex-trait analysis. Many traits of agronomic importance are quantitative in their expression and believed to be controlled by many genes (Tanksley et al. 1989). Quantitative trait locus (QTL) analysis is providing new insights into the genetic composition of many characters by allowing dissection into discrete genetic entities, the study of their genomic distribution, and the relative magnitude of their effects (Stuber 1992).

Monoterpenes are widely known components of plant leaf oils and resins as they comprise a major proportion of the volatile fraction which imparts the scents and odours of higher plants (Whittaker 1972). They have attracted research attention due to their ecological roles in the interactions between plants and their environment, acting as attractants for pollination and as herbivore repellents and toxins (reviewed in Charlwood and Charlwood 1991). Monoterpenes and other essential oil and resin components are also of taxonomic and commercial significance in a number of plant groups including conifers, mints (*Mentha* sp.), eucalypts and *Melaleuca* sp. (White 1983; Boland et al. 1991; Croteau and Gershenzon 1994; Southwell et al. 1997).

The most-detailed understanding of monoterpene genetics is available in mints from segregation analysis and more-recent molecular analysis (Croteau and Gershenzon 1994). In woody plants, genetic analysis of monoterpene composition has progressed more slowly. The studies that have been carried out have indicated a strong genetic basis to variation in terpene composition, major gene effects, and high heritabilities for individual components (White 1984; Barton et al. 1991). These findings

are consistent with observations made in herbaceous species and, generally, it appears that a major gene, or relatively few genes, control monoterpene levels in plants (reviewed in Birks and Kanowski 1988; Croteau and Gershenzon 1994). One QTL study which has been carried out to-date on monoterpenes also supported major gene control of the proportion of delta-3-carene in the foliar oil of maritime pine (*Pinus pinaster*) (Plomion et al. 1996).

Many of the aforementioned studies have used proportional measurements of oil or resin constituents. Proportions have been preferred for genetic studies of resin composition as they are independent of yield, which is often responsive to environmental effects (Birks and Kanowski 1988). Proportional data may not be validly analysed using standard statistical methods, a point overlooked in many previous investigations on the genetics of monoterpene composition (Birks and Kanowski 1988).

In the present study, genetic control of oil composition was investigated for an interspecific eucalypt hybrid. We present an analysis based on log-ratios which have been put forward as one approach to allow valid analysis of proportional data using standard statistical methodology (Birks and Kanowski 1993). Marker-trait co-segregation analysis was performed on the major monoterpenes using a genetic map for the maternal parent of a half-sib population. An interpretation of genetic associations between markers and oil-variables is given in the light of known metabolic processes affecting monoterpene composition in plants.

Materials and methods

Population and genetic map for *Eucalyptus grandis* selection G44

A single population of half-sibs was used for genetic mapping and marker-trait co-segregation analysis. The population originated from an open-pollinated (OP) seed orchard (Aracruz Florestal S.A., Brazil) which consisted of a single *E. grandis* G44 selection surrounded by 25 *Eucalyptus urophylla* selects. The origins of the seed-orchard material are described elsewhere (Grattapaglia et al. 1996). A half-sib population was achieved by excluding self-progeny from analysis by screening for the maternal genotype using co-dominant RAPD markers (Grattapaglia et al. 1995).

Foliage samples were collected from three sources for marker genotyping: two field trials (Site 1: a 5 year-old clone bank at Dongmen State Forest Farm, Dongmen, Guangxi Provenance, China, and Site 2: a Queensland Forest Service eucalypt species trial at Gympie, Queensland, Australia), as well as Site 3: a shade-house containing potted seedlings at Forbio Research, Indooroopilly, Queensland

A pseudo-testcross strategy (Grattapaglia et al. 1995) was used to construct a genetic map for the maternal parent by genotyping 165 F_1 progeny for 382 RAPDs (Shepherd 1998). The genetic map was constructed using Mapmaker v3.0 for DOS (Lander and Green 1987). Of the 382 RAPD markers, 162 fitted the expected 1:1 ratio for a testcross and were used in linkage analysis. Briefly, the map consisted of 91 RAPD markers in 28 linkage groups and covered a total distance of 964 cM (Haldane) with a mean distance and standard deviation of 15 ± 7.2 cM between markers. The genetic map and methodology are given in Shepherd (1998).

Markers are identified by the manufacturer's (Operon Technology, Alameda Calif.) code name for RAPD primers and the size of the RAPD in base pairs. That is, n140_711 is the RAPD band of 711 bp generated from Operon primer N14. A "r" between the prefix and suffix indicates the marker mapped in the reverse linkage phase (see Shepherd 1998).

Leaf-oil analysis

Foliage for leaf-oil analysis was collected from a single representative of 86 genotypes at Site 1. Unblemished mature leaves of the current growth season were selected from the 4th to the 9th position on the shoot tip from a branch 9–10 m above the ground and with a northerly aspect.

Extraction of oil from the foliage and gas-chromatography (GC) analysis were based on Ammon et al. (1985). Approximately equal masses of foliage were added to a set volume (usually 20 ml) of 99% ethanol in the field. The exact mass of foliage added was determined by weighing the vials with ethanol prior to and following the addition of foliage. A time-course experiment, in which aliquots from one vial were removed periodically, confirmed that the major monoterpenes were fully extracted after 18 h (data not shown). An aliquot of either 1.5 ml or 200 ml was removed and an internal standard, n-tetradecane, added to the aliquot at the rate of 200 mg/l prior to compositional analysis. The proportions of oil components were calculated as the ratio of the integrated area for each compound over the sum of the total integrator area for all isolated components. Yields (mg/10 g fresh weight) were determined by reference to a known amount of internal standard added to each sample and calculated according to Birks and Kanowski (1988). Total oil yield was the sum of the yields of all isolated components (i.e. excluding the internal standard).

To investigate the representativeness of the sampling technique, samples were also taken from juvenile foliage, a second aspect, and at a second height in the canopy of 11 trees. Additionally, samples were also collected from the same branch both in the morning and in the afternoon to test for diurnal effects. Repeatability of sampling within a branch was investigated by taking five samples from the same branch for six trees. None of these factors were found to significantly effect the yield or proportions of the major monoterpenes (Shepherd 1998).

Proportional data were converted into log-ratios according to Birks and Kanowski (1993). For the five major monoterpenes there was a total of ten possible log-ratio combinations. As some components were absent from some individuals it was not possible to calculate log-ratios, therefore sample sizes for log-ratios ranged from 52–74. Distributions of log-ratios were tested for normality using the Shapiro and Wilk W statistic. Pearson's correlations were used to examine the relationships between oil log-ratio variables. All test statistics and frequency distributions were generated using Statistica for Windows v 4 (Statsoft, Tulsa Okla.).

Marker-trait co-segregation analysis

Single marker tests

Two genotypic classes were identifiable amongst hybrid progeny based on RAPD markers: those individuals where a RAPD band was present and hence the individual was heterozygous and individuals which were homozygous null. Individuals were partitioned into the two RAPD marker classes for all 162 framework, accessory or unlinked maternal markers and tested for linkage with log-ratio variables using ANOVA (equivalent to a *t* test in this case) using the Basic Statistics module of Statistica. A sequential Bonferroni test was applied to identify those comparisons which were significant at an experiment-wise error rate of $\alpha \leq 0.05$ (Rice 1988).

Interval and composite-interval mapping

Interval and composite-interval mapping were carried out for each log-ratio variable using QTL Cartographer software V1.13 for a PC (Basten et al. 1994, 1999). A backcross model was used to estimate the additive effect of QTLs by comparing heterozygotes to a homozygous null background.

Interval mapping (IM) was conducted using model 3 of QTL Cartographer. Likelihood ratios (LRs) were calculated at 2-cM intervals along the mapped genome. Experiment-wise significance levels for each trait were applied to LR test statistics by deriving thresholds from a permutation test based on 20 permutations (Churchill and Doerge 1994).

Composite interval mapping (CIM) was carried out utilising model 6 in a QTL Cartographer. Markers used to control for background were first identified using stepwise regression implemented using the FB option in the SRmapqtl module of QTL Cartographer. Experiment-wise significance levels for CIM were determined as for IM. Approximate 95% confidence intervals on QTL position were determined for IM and CIM peak positions using one-LOD support interval tests (Conneally et al. 1985; Lander and Botstein 1989).

Results

Hybrid foliar-oil composition

The foliage of 86 trees was analysed for oil composition. Five compounds were found to exceed 1% in most samples and were identified as the monoterpenes limonene, 1,8-cineole, *p*-cymene, γ–terpinene and α–pinene. Due to the difficulty of interpreting distributions and correlations of compositional data (proportions), raw proportional data for the five major monoterpenes were transformed into log-ratios for further analysis as they are amenable to analysis by standard statistical methods (see Discussion). Of the ten possible ratios, the most variable were based on either limonene or γ–terpinene (Table 1). Testing of the frequency distributions for normality indicated that only the ratios of α–pinene to 1,8-cineole and α–pinene to *p*-cymene were approximately normal (Fig. 1 a-j). The remaining non-normal distributions appeared to be mono-, bi- or tri-modal.

Relationships amongst the oil components

Log-ratios of the proportions of the major monoterpenes were tested for intercorrelations (Table 2). The data set was highly intercorrelated with 28 of a possible 45 correlations significant. For example, the log-ratio of α–pinene to limonene was negatively correlated to the log-ratios of limonene to 1,8-cineole, *p*-cymene and γ–terpinene. This indicated that those individuals in which the proportion of limonene was greater than the proportion of α–pinene, tended to have lower amounts of the other three monoterpenes. Other relationships may be deduced from the matrix; for example, the log-ratio of α–pinene to 1,8-cineole is positively correlated with the log-ratio of α–pinene to *p*-cymene. This indicates that individuals with a high proportion of α–pinene with respect to the proportion of 1,8-cineole also tend to have lower proportions of *p*-cymene. The high degree of intercorrelation evident amongst oil variables suggested that difficulties may arise in the interpretation of associations with genetic markers as a consequence of epistatic or other genetic interactions.

Table 1 Descriptive statistics for ten log-ratio variables derived from the proportions of five major monoterpenes identified in the eucalypt hybrid

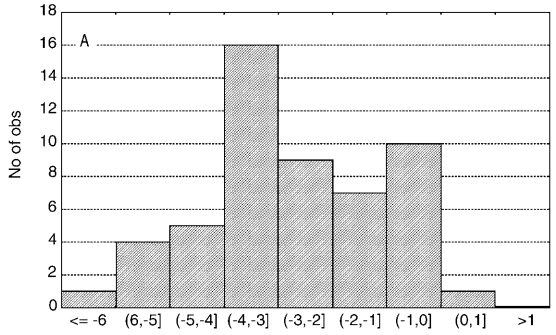
Log-ratio[a]	*n*	Mean	Minimum	Maximum	Range	SD
P/C	53	–2.67	–6.08	0.51	6.59	1.56
P/CY	53	–3.49	–5.96	–0.57	5.39	1.21
P/T	52	–2.97	–6.07	2.40	8.46	2.32
L/C	74	1.69	–2.14	5.81	7.95	2.39
L/CY	74	0.83	–3.44	3.77	7.21	1.98
L/T	73	1.46	–3.78	6.66	10.43	2.72
C/CY	74	–0.85	–3.47	2.79	6.25	1.59
C/T	73	–0.23	–4.25	6.16	10.41	2.66
CY/T	73	0.65	–1.48	5.90	7.38	1.87
P/L	53	–4.38	–6.58	–0.15	6.43	1.80

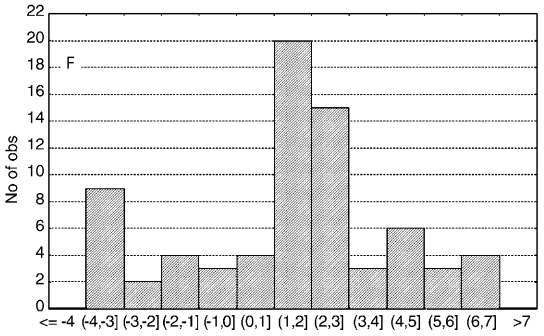
[a] P/L is the log-ratio of pinene to limonene; P = α-pinene; C = 1,8-cineole; CY = *p*-cymene; L = limonene; T = γ-terpinene

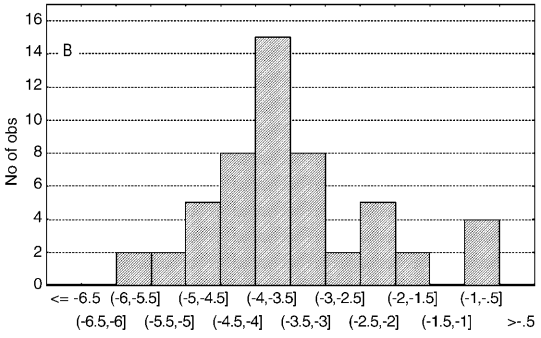
Table 2 Correlations between log-ratios of major monoterpenes (n ≥ 52). Correlations with *P* value < 0.05 in bold

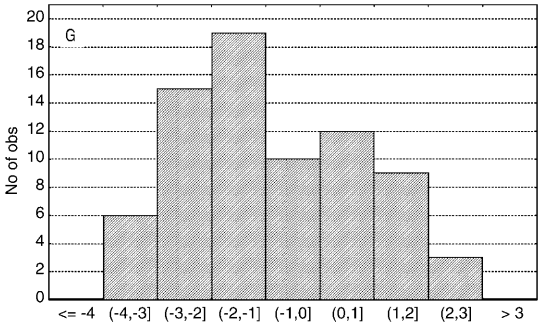
Log-ratio[a]	P/L	P/C	P/CY	P/T	L/C	L/CY	L/T	C/CY	C/T
P/C	0.07								
P/CY	0.23	**0.38**							
P/T	0.13	0.15	**0.69**						
L/C	**–0.75**	**0.60**	0.06	–0.01					
L/CY	**–0.71**	0.22	**0.52**	**0.38**	**0.71**				
L/T	**–0.49**	0.09	**0.46**	**0.80**	**0.45**	**0.77**			
C/CY	0.15	**–0.55**	**0.56**	**0.48**	**–0.48**	**0.27**	**0.34**		
C/T	0.09	**–0.38**	**0.44**	**0.85**	**–0.32**	0.24	**0.70**	**0.74**	
CY/T	0.00	–0.09	0.17	**0.83**	–0.06	0.12	**0.73**	0.23	**0.82**

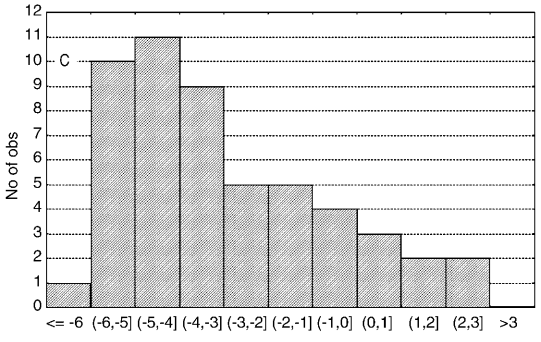
[a] Refer to Table 1

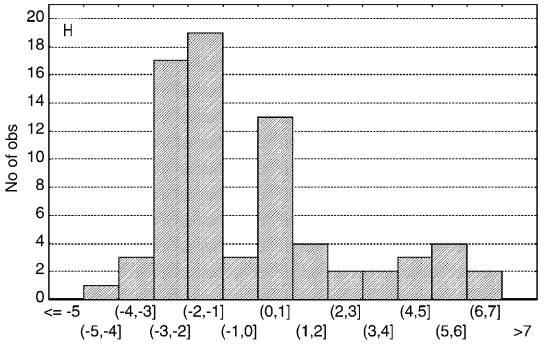

A
No of obs
<= -6 (6,-5] (-5,-4] (-4,-3] (-3,-2] (-2,-1] (-1,0] (0,1] >1

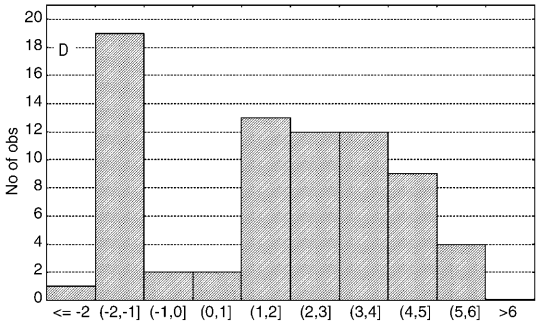

F
No of obs
<= -4 (-4,-3] (-3,-2] (-2,-1] (-1,0] (0,1] (1,2] (2,3] (3,4] (4,5] (5,6] (6,7] >7

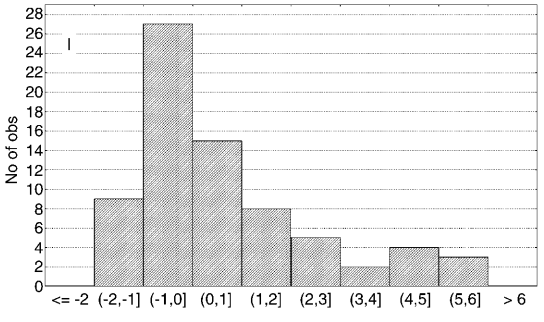

B
No of obs
<= -6.5 (-6,-5.5] (-5,-4.5] (-4,-3.5] (-3,-2.5] (-2,-1.5] (-1,-.5]
(-6.5,-6] (-5.5,-5] (-4.5,-4] (-3.5,-3] (-2.5,-2] (-1.5,-1] >-.5

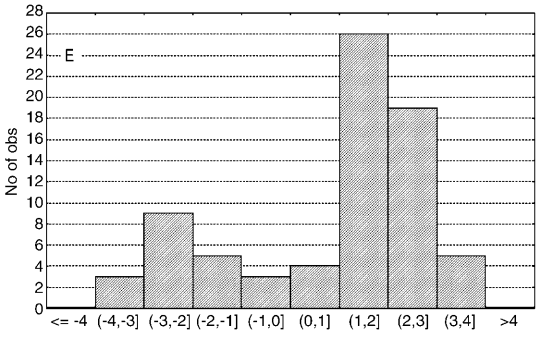

G
No of obs
<= -4 (-4,-3] (-3,-2] (-2,-1] (-1,0] (0,1] (1,2] (2,3] > 3

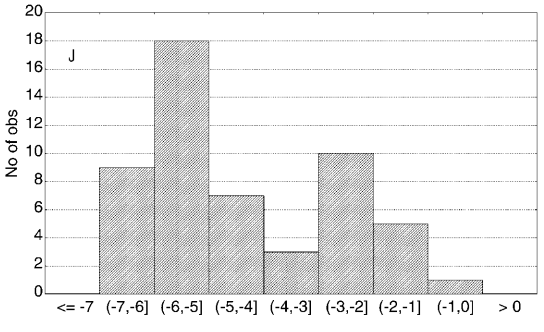

C
No of obs
<= -6 (-6,-5] (-5,-4] (-4,-3] (-3,-2] (-2,-1] (-1,0] (0,1] (1,2] (2,3] >3


H
No of obs
<= -5 (-5,-4] (-4,-3] (-3,-2] (-2,-1] (-1,0] (0,1] (1,2] (2,3] (3,4] (4,5] (5,6] (6,7] >7


D
No of obs
<= -2 (-2,-1] (-1,0] (0,1] (1,2] (2,3] (3,4] (4,5] (5,6] >6


I
No of obs
<= -2 (-2,-1] (-1,0] (0,1] (1,2] (2,3] (3,4] (4,5] (5,6] > 6


E
No of obs
<= -4 (-4,-3] (-3,-2] (-2,-1] (-1,0] (0,1] (1,2] (2,3] (3,4] >4


J
No of obs
<= -7 (-7,-6] (-6,-5] (-5,-4] (-4,-3] (-3,-2] (-2,-1] (-1,0] > 0

◀ **Fig. 1a–j** Frequency distributions for log-ratios of the major monoterpenes components. **a** α–pinene to 1,8-cineole; **b** α–pinene to *p*-cymene; **c** α–pinene to γ–terpinene; **d** limonene to 1,8-cineole; **e** limonene to *p*-cymene; **f** limonene to γ–terpinene; **g** 1,8-cineole to *p*-cymene **h** 1,8-cineole to γ–terpinene; **i** *p*-cymene to γ–terpinene; **j** α–pinene to limonene

Marker-trait associations

Single marker tests

Analysis of variance tests were used to examine linkage between 162 maternal markers and log-ratio variables based on 86 genotypes from Site 1. Thirteen tests were significant at an experiment-wise error rate of $\alpha \leq 0.05$ (Table 3). Four framework or accessory markers which mapped to linkage group 5 were variously associated with log-ratio variables based upon limonene. This included the two framework markers o020_711 and F100r550, and two accessory markers linked to o020_711, o020_521 and W020_910.

Table 3 Linkages between markers and log-ratio variables as determined by single-marker ANOVA tests and following the application of sequential Bonferroni test (experiment-wise significance $\alpha \leq 0.05$)

Marker cat.[a]	Marker ID	Log-ratio[b]	*df* error	*F*[c]
AC5	O020_52	L/C	68.00	92.95
AC5	O020_52	L/CY	68.00	79.85
AC5	O020_52	L/T	67.00	22.90
AC5	O020_52	P/L	48.00	114.27
AC5	W020_910	L/C	52.00	48.45
AC5	W020_910	L/CY	52.00	27.94
AC5	W020_910	P/L	41.00	61.21
FW5	F100r550	L/C	72.00	21.54
FW5	F100r550	P/L	52.00	22.05
FW5	o020_711	L/C	72.00	186.13
FW5	o020_711	L/CY	72.00	132.63
FW5	o020_711	L/T	71.00	29.61
FW5	o020_711	P/L	52.00	195.11

[a] Marker cat. (Marker category) where FW5 is a framework marker on linkage group 5, AC5 is an accessory marker on linkage group 5
[b] Refer to Table 1
[c] F test (1 *df* for all effects, *df* error as specified in column 4). All *p*-values for comparison-wise tests were < 2.16E-05

Interval and composite-interval mapping

Log-ratio variables were tested for associations with the framework map using interval mapping and composite-interval mapping implemented in QTL Cartographer using a previously derived linkage map (Shepherd 1998). Peak LR tests that were significant at the 0.05 level for each trait are reported in Table 4. Out of the ten log-ratios tested, four variables based on limonene were significantly associated with a region on linkage group 5. A high degree of phenotypic variation in all four ratios was explained by this region, ranging from 68 to 81%. For all four log-ratios, individuals heterozygous for the QTL were linked to individuals with ratios with higher amounts of limonene, indicating that heterozygotes, on average, had a higher amount of limonene and lower amounts of the other four major monoterpenes.

Group 5 is delineated by markers Q173_360 and F100r550 and covers a mapped distance of 46.2 cM. A third marker, o200_711, is located 14.2-cM from Q173_360. The peaks for three of the variables appear centred on the marker o200_711. Two variables, the log-ratio of limonene to 1,8-cineole and limonene to *p*-cymene, appear to have two peaks close by on group 5 centred on marker o200_711. The 95% confidence intervals

Table 4 Summary of QTLs for log-ratio variables of oil composition traits as determined by interval and composite-interval mapping using QTL Cartographer

Log-ratio[a]	Linkage group	Position[b]	CI[c]		LR[d]	Add[e]	R_2[f]	TR_2[g]	S[h]	EW LR 0.05[i]	EW LR 0.01[i]
Interval mapping											
P/L	5	14.89	6.01	24.89	84.16	3.33	0.78	0.79	4.13	27.91	28.01
L/C	5	14.01	8.01	22.89	97.03	–4.30	0.75	0.77	1.11	25.86	29.97
L/C	5	16.89	14.01	22.89	98.93	–4.30	0.76	0.77	1.10	25.86	29.97
L/CY	5	10.01	0.00	18.89	97.51	–3.88	0.80	0.82	1.65	39.68	63.28
L/CY	5	16.89	14.89	24.89	92.66	–3.73	0.77	0.78	2.27	39.68	63.28
L/T	5	2.01	0.00	10.01	36.85	–5.17	0.68	0.69	16.03	12.57	16.77
Composite-interval mapping											
P/L	5	14.01	10.01	14.89	89.54	3.33	0.78	0.82	2.16	30.06	35.48
L/C	5	14.01	8.01	14.89	115.92	–4.38	0.75	0.81	1.79	24.39	32.97
L/CY	5	12.01	2.01	14.89	95.51	–3.56	0.69	0.79	1.54	41.35	47.47
L/T	5	0.01	0.00	10.01	42.67	–4.77	0.60	0.77	4.12	13.70	18.22

[a] Refer to Table 1
[b] The distance in Haldane cM from the left-most marker on the linkage group
[c] 95% Confidence interval for QTL position in distance from left-most marker as determined by the one-LOD support interval
[d] Likelihood-ratio test statistic (see Basten et al. 1999)
[e] Estimate of a (additive effect) under H_1
[f] r^2 Proportion of variance explained by a QTL at the test site
[g] r_t^2 Proportion of variance explained by a QTL and any other explanitory variables in the model
[h] Test statistic S for the normality of the residuals under H_1
[i] Experiment-wise error rate at 0.05 and 0.01 levels determined by permutation testing

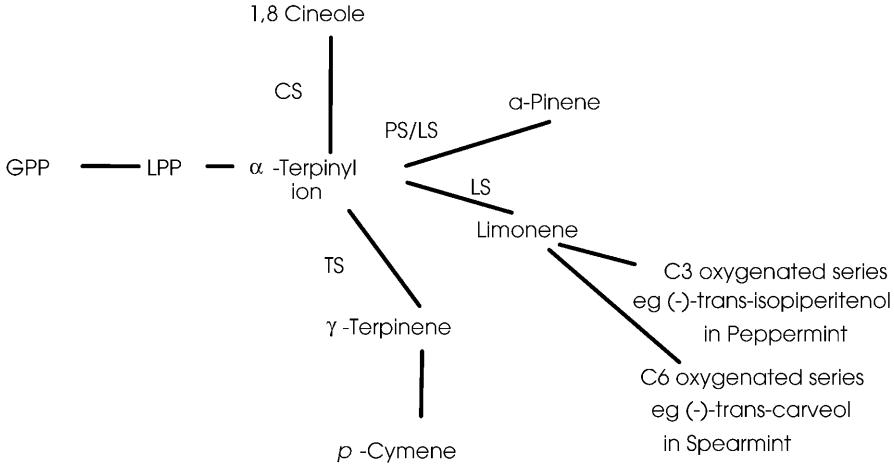


Fig. 2 Biosynthetic relationships for the major monoterpenes found in the eucalypt hybrid, based on studies on sage [1,8-cineole synthase (CS) and pinene synthase (PS)], peppermint [limonene synthase (LS)], and thyme [γ–terpinene synthase (TS)]. Derived from LaFever and Croteau (1993) scheme 1; Croteau (1987) scheme 10; Rajaonarivony et al. (1992) scheme 1; Croteau et al. (1994) and Gambliel and Croteau (1984) [Geranyl pyrophosphate (*GPP*), Linalyl pyrophosphate (*LPP*)]

for these two peaks largely overlap and it was not clear whether there were two QTLs close together or a single QTL. The tests at o200_711 (position 14.2 cM) for both traits were highly significant and hence consistent with results from the single marker tests. The two other log-ratio variables had a single peak each, the peak for the log-ratio of limonene to γ-terpinene was located 2 cM from Q173_360 rather than at o200_711, although there were large overlaps in the confidence intervals between these and the other traits.

Composite-interval mapping results were largely consistent with those of interval mapping (Table 4). The same four variables based on limonene were linked with group 5. Both IM and CIM positioned the peak for the log-ratio of limonene to r-terpinene closer to Q170_1 than to o200_711. The confidence intervals for this trait by both methods, however, still overlap the confidence intervals for the other traits. One significant difference between the methods was the determination of a single peak for both the log-ratio of limonene to 1,8-cineole and limonene to *p*-cymene positioned close to o0200_711. A further difference was the rapid decay in the LR significance upon all traits in the interval between o0200_711 and F100r550. Composite-interval mapping appears to have more-precisely positioned these effects with regard to the right boundary.

Discussion

Genetic control of monoterpene composition in an interspecific eucalypt hybrid

A mapping approach was used to investigate the genetics of monoterpene composition in an interspecific *Eucalyptus* hybrid. To obtain a suitable population for segregation analysis, a half-sib family was derived from open-pollinated progeny by eliminating selfs based on genetic-marker phenotypes. Segregation analysis of markers and traits followed a pseudo-testcross model which in a maternal half-sib family permits analysis of those markers in a testcross configuration, heterozygous in the common maternal parent and homozygous null in the paternal parents (Grattapaglia and Sederoff 1994). Analogous to QTL analysis in a backcross, this model contrasts a single gene dose at a QTL in heterozygous individuals with homozygous-absent individuals and hence does not allow a distinction between additive and dominant intralocus gene action.

Application of this approach in the hybrid eucalypt family revealed that a single gene dose at one region in the mapped genome profoundly influenced monoterpene composition. The correspondence of single marker tests with interval and composite-interval tests provides strong evidence for a major QTL affecting monoterpene composition located near o0200_711. Our results also indicate that the majority of variation in mapped traits was attributed to the *E. grandis* parent. To be detected in this family, the major-effect QTL must have been absent or present at a low frequency amongst paternal *E. urophylla* parents, suggesting that paternal, environmental or interaction factors have minor roles in determining phenotypic variation in the proportions of the major monoterpenes.

Genetics and biosynthesis of monoterpenes in an interspecific eucalypt

The formulation of a genetic model and definition of the traits affected by the QTL require a careful evaluation of the biochemical, physiological and statistical factors concerned with monoterpene compositional data sets (Birks and Kanowski 1988). The choice of measurement basis determines how valid analysis can proceed. We have used proportions as they have been preferred for genetic studies since they are independent of oil yield (Birks and Kanowski 1988). This means that it is unlikely that genes controlling physiological factors, such as oil-gland density, which are believed to be largely responsible for oil yield, are located at the QTL. Raw proportional data, however, can not be validly analysed using standard statistical methods due to a “summation to one restriction” (Birks and Kanowski 1988). A log-ratio

transformation was recommended to allow valid analysis of proportions using standard statistics (Birks and Kanowski 1993). This transformation addressed interdependence as a consequence of the summation to one restriction; however, the log-ratio data set still exhibited a high degree of intercorrelation. A high degree of intercorrelation is typical of compositional data for foliar oils or resins, has been thought to be due to physiological or biochemical relationships (Zavarin 1970), and shown to be true for resin components in conifers (White 1983). It is likely that the correlations evident amongst our log-ratio data are also the result of a commonality in biosynthetic origins.

Log-ratios based on limonene were highly variable and all log-ratios in which limonene was a factor were linked to the identified genomic region of *E. grandis*. The predominance of limonene in the composition of some hybrids suggested that it influenced the proportions of the other monoterpene constituents. The predominance of limonene and the hypothesised biochemical relationships amongst monoterpenes, together suggest that the identified region contains a gene or genes directly affecting the proportion of limonene. The effect of the region on the other monoterpenes was probably a consequence of the biosynthetic relationships to limonene.

One model for the biosynthesis of monoterpenes that is consistent with many of the observed relationships in the present study, as well as the known biochemical pathways in herbaceous species, is a branch-point model (Fig. 2). In such a model, a branch in the pathway occurs where multiple enzymes compete for a common substrate. Such a divergence into an array of products from a single precursor is known at a key regulatory point in monoterpene biosynthesis, where geranyl pyrophosphate undergoes cyclisation and isomerisation, by a class of enzymes known as cyclases, into limonene and other monoterpenes [LaFever and Croteau (1993) scheme 1 (γ–terpinene synthase); Croteau (1987) scheme 10; Rajaonarivony et al. (1992) scheme 1 (limonene synthase); Croteau et al. (1994) (1,8-cineole synthase); Gambliel and Croteau (1984) (pinene synthase)]. The agreement between observed relationships and those predicted from this model support the conclusion that the observed correlations are a consequence of biosynthetic origins.

In accounting for the phenotypic interactions evident amongst the log-ratio variables, the proposed biosynthetic model also suggests possible genetic mechanisms for the cause of the correlation. One possibility is a non-additive interaction between genetic loci (epistasis). For example, this may occur if an allele for a highly active limonene synthase segregated in a background of alleles for low limonene production. As a consequence of substrate competition, productivity of gene products at loci for other monoterpene synthases may be reduced, yielding low levels of these compounds in high-limonene individuals. Further multiple-loci models may be hypothesised based on the additive action of several linked, segregating loci. Alone, these two models may not provide a full explanation, either, since pleiotrophy could also account from some of the observed relationships; for example, where a synthase generates multiple products. Formal testing by the fitting of multi-locus models with and without non-additive interaction may clarify the nature of the detected phenotypic correlations.

Further elucidation of the role of the gene(s) located on group 5 may arise through co-localisation studies with candidate genes. A number of monoterpene synthases have now been cloned, including a limonene synthase (reviewed in McGarvey and Croteau 1995). Other genes fundamental in terpene biosynthesis have also been cloned and could be investigated through co-localisation studies with the mapped major-gene effect.

Another approach to characterise the function of the gene(s) located on group 5 may be to examine higher-order products derived from limonene. Limonene is a key intermediate in the biosynthesis of terpenes of the *p*-menthane series (Croteau and Gershenzon 1994). *p*-Menthane series compounds are present in the foliar oils of many eucalypts and in some species may make up 40–55% of the oil composition (Boland et al. 1991). In *Mentha* sp., high levels of limonene were believed to result from an inability to oxidise limonene (Croteau and Gerhenzon 1994; Lincoln et al. 1971) and, as a consequence, these individuals had reduced amounts of the *p*-menthane series products. Identification and quantification of *p*-methane series monoterpenes in the eucalypt hybrid may help designate a function to a gene, or genes, in the region affecting the proportion of limonene.

Factors influencing the detection of gene effects and their interpretation in an interspecific eucalypt hybrid

The detection of a major gene effect was somewhat surprising given the complexity of monoterpene metabolism. One explanation may be that, although many gene products contribute to the overall expression of the trait, the magnitude of the effect of one gene product far exceeds the effect of all other minor genes. Whilst, a single-gene model is the simplest most model parsimonious with our data we feel that caution is required as several aspects of the study may have favoured the detection of large effects.

Several factors may also have favoured an over-simplification of a genetic model for the control of monoterpene composition. A small sample size for marker-trait co-segregation analysis tends to bias the detection of genes of large effect (Kearsey and Farquahr 1998). A significant overestimate of the gene(s) effect may occur where the test power is low and the number of QTLs is large (Lynch and Walsh 1997). Hence, as yet undetected QTLs may be discovered by analysis of a larger sample, or with more powerful designs, leading to a reduction in the significance of the detected genomic region. A second source of bias may be present in estimates of the phenotypic effect obtained from interval mapping. This is because interval mapping assumes an approximately normally distributed trait (Lincoln et al. 1993). Most of

the log-ratio variables used in this analysis were multimodal; therefore, non-normality of the distributions may have led to an overestimate of the region's effect on limonene expression. Single marker tests provide some protection in this area as they are more robust with regard to the non-normality of distributions. A third factor was map coverage. The genetic map was estimated to cover approximately 75% of the genome (Shepherd 1998), hence additional QTLs may yet be found in unmapped regions.

Genetic studies in *Mentha* sp. suggest that interspecific hybridisation leads to the enhanced detection of some loci. Rare individuals, yielding limonene grossly in excess of either parent, were found in an F_1 generation between *Mentha citrata* and *Mentha spicata* L var. *crispata* or *M. citrata* by *Mentha aquatica* (Lincoln et al. 1971). The rarity of these individuals was believed to be due to dominant gene action and a requirement for a particular genotype combination at three loci which arose due to crossing-over between homoeologous chromosomes in unusual quadrivalent pairing of the octaploid *M. citrata* parent (2n = 96) (Lincoln et al. 1971; Hefendehl and Murray 1973, 1976). Parental strains were considered to be true-breeding with compositions low in limonene. It seems that whilst the rare recombination events leading to high limonene hybrids may have occurred in pure-strain crosses, segregation was probably not detected because there was not an appropriate null background. The use of interspecific hybrids for genetic studies appears to provide a method for the detection of genes which may not be detected in crosses between more-closely related individuals and hence not otherwise amenable to study. Large dominance effects evident in a wide cross may be misleading for an interpretation of the trait complexity, however, as they may overshadow the contribution of other genes.

Conclusions

Studies of potentially intercorrelated biological parameters present a challenge for the identification of valid linkage with genetic markers. Genetic linkage must be distinguished from interdependence amongst variables which may arise due to biochemical relationships. Interdependence inherent in the measurement basis may also confound the analysis of proportional data. Log-ratios allow a valid testing of proportions but do not overcome the interdependence that is a consequence of commonality in biosynthetic origins. The profound influence upon monoterpene composition by a single genomic region inherited from the *E. grandis* parent was thought to be the consequence of an investigation of the effect in a wide cross. The impact of the QTL upon compositional variables was believed to be due to a gene, or genes, affecting limonene metabolism and, although our data is parsimonious with a single-gene mode, further clarification of the genetic model involved is required.

Acknowledgements The authors thank Dongmen State Forest Farms and the Queensland Department of Primary Industries (Forestry) for providing access to their populations. We are in debt to Dr. G. Jones and Ms. K. Shepherd for GC analysis of the foliage samples. We are also grateful for helpful communications with Prof. P. Kanowski and Dr. J. Birks during data analysis and to Dr. K. Mitchelson for comments during preparation of the manuscript. M. Shepherd was supported by an Australian Postgraduate Research Award.

References

Ammon DG, Barton AFM, Clarke DA, Tjandra J (1985) Rapid and accurate chemical determination of terpenes in the leaves of *Eucalyptus* species. Analyst 110:921–924

Barton AFM, Cotterill PP, Brooker MIH (1991) Heritability of cineole yield in *Eucalyptus kochii*. Silvae Genet 40:37–38

Basten CJ, Weir BS, Zeng ZB (1994) Zmap-a QTL cartographer. In: Smith C. et al. (eds) Proc of the 5th World Congr on Genetics Applied to Livestock Production: Computing Strategies and Software, vol 22, pp 65–66

Basten CJ, Weir BS, Zeng Z-B (1999) QTL Cartographer V1.13. A reference manual and tutorial for QTL mapping. Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, USA

Birks JS, Kanowski PJ (1988) Interpretation of the composition of coniferous resin. Silvae Genet 37:29–39

Birks JS, Kanowski PJ (1993) Analysis of resin compositional data. Silvae Genet 42:340–350

Boland DJ, Brophy JJ, House APN (1991) *Eucalyptus* leaf oils, use, chemistry, distillation and marketing. In: *Eucalyptus* leaf oils, use, chemistry, distillation and marketing. Inkarta Press, Melbourne, Australia

Charlwood BV, Charlwood KA (1991) Monoterpenoids. In: Charlwood BV, Banthorpe DV (eds) Terpenoids, vol 7, Academic Press, London, pp 43–98

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative-trait mapping. Genetics 142:285–294

Conneally PM, Edwards JH, Kidd KK Lalouel JM Morton NE, Ort J, White R (1985) Report of the committee on methods of linkage analysis and reporting. Cytogenet Cell Genet 40: 356–359

Croteau R (1987) Biosynthesis and catabolism of monoterpenoids. Chem Rev 87:929–954

Croteau R, Gerahenzon J (1994) Genetic control of monoterpene biosynthesis in mints (*Mentha*: Laminaceae). In: Ellis BE, Kuroki GW, Stafford HA (eds) Genetic engineering of plant secondary metabolism. vol 28, Plenum Press, New York, pp 193–229

Croteau R, Alonso WR, Koepp AE, Johnson MA (1994) Biosynthesis of monoterpenes: partial purification, characterization and mechanism of action of 1,8-Cineole synthase. Arch Biochem Biophys 309:184–192

Gambliel H, Croteau R (1984) Pinene cyclases I and II: two enzymes from sage (*Salvia officinalis*) which catalyse stereospecific cyclisation of geranyl pyrophosphate to monoterpene olefins of opposite configuration. J Biol Chem 259:740–748

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

Grattapaglia D, Bertolucci FLG, Sederoff RR (1995) Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD Markers. Theor Appl Genet 90:933–947

Grattapaglia D, Bertolucci FLG, Penchel R, Sederoff RR (1996) Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. Genetics 144:1205–1214

Hefendehl FW, Murray MJ (1973) Monoterpene composition of a chemotype of *Mentha piperita* having high limonene. Planta Med 23:101–109

Hefendehl FW, Murray MJ (1976) Genetic aspects of the biosynthesis of natural odors. Lloydia 39:39–52

Kearsey MJ, Farquhar GL (1998) QTL analysis in plants: where are we now? Heredity 80:137–142

LaFever RE, Croteau R (1993) Hydride shifts in the biosynthesis of the *p*-menthane monoterpenes α–terpinene, γ–terpinene, and β-phellandrene. Arch Biochem Biophys 301:361–366

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lander E, Green P (1987) Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA 84:2363–2367

Lincoln DE, Marble PM, Cramer FJ, Murray MJ (1971) Genetic basis for high limonene-cineole content of exceptional *Mentha citrata* hybrids. Theor Appl Genet 41:365—370

Lincoln SE, Daly MJ, Lander ES (1993) Mapping genes controlling quantitative traits using MAPMAKER/QTL version 1.1: A totorial and reference manual. (A Whitehead Institute for Biomedical Research Technical Report, Second Edition, Jan 1993)

Lynch M, Walsh B (1997) Genetics and analysis of quantitative traits. Sinauer Associates Inc., Sunderland, USA

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

Rice WR (1988) Analyzing tables of statistical tests. Evolution 43:223–225

Plomion C, Yani A, Marpeau A (1996) Genetic determinism of δ-3-carene in maritime pine using random amplified polymorphic DNA (RAPD) markers. Genome 39:1123–1127

Rajaonarivony JIM, Gershenzon J, Croteau R (1992) Characterization and mechanism of (4 S)-Limonene synthase, a monoterpene cyclase from the glandular trichomes of peppermint *Mentha piperita*. Arch Biochem Biophys 296:49–57

Shepherd M (1998) Genetic mapping in a *Eucalyptus* hybrid. PhD dissertation, Botany Department, University of Queensland, Australia

Southwell IA, Freeman S, Rubel D (1997) Skin irritancy of tea tree oil. J Essent Oil Res 9:47–52

Stuber CW (1992) Biochemical and molecular markers in plant breeding. In: Plant breeding reviews, vol 9. Wiley, New York, pp 37–61

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Bio/Technology 7:257–264

White EE (1983) Biosynthetic implications of terpene correlations in *Pinus contorta*. Phytochemistry 22:1399–1405

White EE (1984) Mode of genetic control of monoterpenes in foliage of controlled crosses of *Pinus contorta*. Silvae Genet 33:115–119

Whittaker D (1972) The monoterpenes. In: Newman AA (ed) Chemistry of terpene and terpenoids. Academic Press, London, pp 11–67

Zavarin E (1970) Qualitative and quantitative co-occurrence of terpenoids as a tool for elucidation of their biosynthesis. Phytochemistry 9:1049–1063