HANOVER, J. W.: Inheritance of isozymes in seed and bud tissues of blue and engelmann spruce Genome 29: 239—246 (1987) —— FARJON, A. and RUSHFORTH. K. D.: A classification of *Abies* MILLER Pinaceae. Notes from the Royal Botanic Garden Edinburgh 46: 56—79 (1989). —— FINNERTY, V. and JOHNSON, G.: Post-translational modification as a potential explanation of high levels of enzyme polymorphism, xanthine dehydrogenase and aldehyde oxidase in *Drosophila melangoaster*. Genetics 91: 695—722 (1979). —— FRANCO, J. A.: Abetos. An. Inst. Super. Agron. 17. Lisbon, Portugal (1950). —— FURNIER, G. R., KNOWLES, P., ALFRSIUK, M. A. and DANCIK, B. P.: Inheritance and linkage of allozymes in seed tissues of white-bark pine. Can. J. Genet. Cytol. 28: 601—604 (1986). —— GURIES, R. P. and LEDIG, F. T.: Inheritance of some polymorphic isoenzymes in Pitch pine (*Pinus rigida* MILL.). Heredity 40: 27—32 (1978). —— GURIES, R. P. and LEDIG, F. T.: Genetic diversity and population structure in Pitch pine (*Pinus rigida* MILL.). Evolution 36: 387—402 (1982). —— HARRY, D. E.: Inheritance and linkage of isozyme variants in incense-cedar. J. Hered. 77: 261—266 (1986). —— JACOBS, B. F., WERTH, C. R. and GUTTMAN, S. I.: Genetic relationships in *Abies* (fir) of eastern United States, an electrophoretic study. Can. J. Bot. 62: 609—616 (1984). —— KING, J. N. and DANCIK, B. P.: Inheritance and linkage of isozymes in white spruce (*Picea glauca*). Can. J. Genet. Cytol. 25: 430—436 (1983). —— LEDIG, F. T. and CONKLE, M. T.: Gene diversity and genetic structure in a narrow endemic, Torrey pine (*Pinus torreyana* PERRY ex CARR). Evolution 37: 79—85 (1983). —— LIU, T. S.: A monograph of the genus Abies. Dep. For. Nat. Taiwan Univ., Taipei, Taiwan. (1971). —— LOUKAS, M., VERGINI, Y. and KRIMBAS, B.: Isozyme variation and heterozygosity in *Pinus halepensis* L. Biochem. Genet. 21: 497—509 (1983). —— MEJNARTOWICZ, L.: Polymorphic of the LAP and GOT in *Abies alba* MILL. Bulletin Académie Polonaise Sciences. Serie des Sciences Biologiques XXVII 12: 1063—1069 (1979). —— MILLAR, C. I.: Inheritance of allozyme variants in Bishop pine (*Pinus muricuta* D. DON). Biochem. Genet. 23: 993—946 (1985). —— MITTON, J. B., LINHART, Y. B., STURGEON, K. B. and HAMRICK, J. L.: Allozyme polymorphisms detected in mature needle tissue of ponderosa pine. J. Hered. 70: 86—89 (1979). —— MORAN, G. F., BELL, J. C. and MATHESON, A. C.: The genetic structure and levels of inbreeding in a *Pinus radiata* D. DON Seed Orchard. Silvae Genet. 29: 5—6 (1980). —— MORETTI, J., BROUSSIER, G. and JAYLE, M. F.: Réalization techniques et premières applications de l'electrophorèse sur gel d'amidon. Bull. Soc. Chim. Biol. 39: 593—605 (1957). NEALE, D. B. and ADAMS, W. T.: Inheritance of isozyme variants in seed tissues of balsam fir (*Abies balsamea*). Can. J. Bot. 59: 1285—1291 (1981). —— NEALE, D. B., WEBWE, J. C. and ADAMS, T.: Inheritance of needle tissue isozymes in Douglas-fir. Can. J. Genet. Cytol. 26: 459—468 (1984). —— NEWTON, K. J.: A gene which alters the electrophoretic mobilities of maize mitochondrial malate dehydrogenase isozymes. Genetics 91 (Suppl.): 88—89 (1979). —— O'MALLEY, D. M., ALLENDORF, F. W. and BLAKE, G. M.: Inheritance of isozyme variation and heterozygosity in *Pinus ponderosa*. Biochem. Genet. 17: 233—250 (1979). —— PASCUAL, L., MARQUEZ, I. A. and LOPEZ-ALONSO, D.: Evidence for the duplication of PGI genes in *Dipcadi serotinum* L. (Liliaceae). Heredity 60: 247—252 (1988). —— PITEL, J. A. and CHELIAK, W. M.: Effect of extraction buffers on characterization of isoenzymes from vegetative tissues of five conifer species. A user's manual. Information Report PI-X-34. Petawawa National Institute. Canadian Forestry Service (1984). —— RUDIN, D.: Leucine-amino-peptidases (LAP) from needles and macrogametophytes of *Pinus sylvestris* L. Inheritance of allozymes. Hereditas 85: 219—226 (1977). —— RUDIN, D. and EKBERG, I.: Linkage studies in *Pinus sylvestris* L. using macro gametophyte allozymes. Silvae Genet. 27: 1—12 (1978). —— STEINHOFF, R. J., JOYCE, D. G. and FINS, L.: Isozyme variation in *Pinus monticola*. Can. J. For. Res. 13: 1122—1132 (1983). —— STEWART, S. C. and SCHOEN, D. J.: Segregation at enzyme loci in megagametophytes of white spruce, *Picea glauca*. Can. J. Genet. Cytol. 28: 149—153 (1986). —— STRAUSS, S. H. and CONKLE, M. T.: Segregation, linkage, and diversity of allozymes in knobcone pine Theor. Appl. Genet 72: 483—493 (1986). —— VALLEJOS, E.: Enzyme activity staining. In: Isozymes in Plant Genetics and Breeding. Part A. S. D. TANKSLEY and T. L. ORTON (eds.). Elsevier, Amsterdam. pp 469—516 (1983). WHEELER, N. C. and GURIES, R. P.: Populations structure, genetic diversity, and morphological variation in *Pinus contorta* DOUGL. Can. J. For. Res. 12: 595—606 (1982). —— YEH, F. C. and EL-KASSABY, Y. A.: Enzyme variation in natural populations of sitka spruce (*Picea sitchensis*). I. Genetic variation patterns among trees from IUFRO provenances. Can. J. For. Res. 10: 415—422 (1980).

Analysis of Resin Compositional Data

By J. S. BIRKS and P. J. KANOWSKI

Oxford Forestry Institute, Department of Plant Sciences,
University of Oxford, South Parks Road, Oxford OX1 3RB,
United Kingdom

(Received 15 July 1993)

Summary

Resin composition can be quantified with relative ease but, because of the properties of proportional data, analysis and interpretation of patterns of variation in the resultant compositional data sets are rather less easy. Proportions of a composition are constrained to sum to unity; individual proportions are not, therefore, independent variables, and cannot be analyzed as such. Resin data shares with other compositional data sets various analytical and interpretative difficulties: a major limitation for taxonomic, genetic and biosynthetic studies is the absence of any interpretable covariance structure for proportional data sets.

A valid approach to the interpretation of resin compositional data is through the analysis of ratios of proportions of a composition, which have the essential property of being invariant under rescaling. Transformation of the proportional data to a logratio data set allows the properties of the lognormal distribution to be employed; the logratio transformation removes the constraint of summation to unity, and provides an interpretable covariance structure. The logratios form a multivariate data set amenable to analysis using standard techniques, which we demonstrate for a sample set of resin compositional data.

Key words: Compositional data, genetic markers, monoterpenes, oleoresin.

1. Introduction

In our earlier critique of the interpretation of resin compositional data (BIRKS and KANOWSKI, 1988), we reviewed the relevant literature over the 2 decades preceding 1986. It was in this era, subsequent to HANOVER's (1966a and b) reports of monogenic control of monoterpene levels in *Pinus monticola* DOUGL., that there was most enthusiasm for genetic investigations of resin composition. The subsequent advent of isozyme (BROWN and MORAN, 1981; YEH, 1989; MUONA, 1990) and molecular (CHELIAK and ROGERS, 1990; NEALE and WILLIAMS, 1991) techniques, with their apparently more direct association with the genome, has

reduced the level of interest in resin composition in taxonomic and genetic studies. Nevertheless, as evident from Hanover's (1992) recent review, our increasing understanding of the biosynthetic pathways leading to resin production (BERNARD-DAGAN, 1988), the apparent utility of terpene data in assisting taxonomic studies (ADAMS, 1989; HUNT et al., 1990; LOCKHART, 1990; FADY et al., 1992; COOL and ZAVARIN, 1992) — where it may play a role complementary to other techniques (VON RUDLOFF and LAPP, 1991), implications of associations between terpene composition and pest or disease resistance (BROOKS et al., 1987; MICHELOZZI et al., 1991), and the tantalising prospect of establishing the mode of genetic control of resin composition (BARADAT and YAZDANI, 1988; ZAVARIN et al., 1990), suggest that it is worth attempting to resolve the difficulties inherent in the acquisition and analysis of such data. While issues relevant to data acquisition have been acknowledged and addressed (RAFA and STEFFECK, 1988; MUZIKA et al., 1990), those inherent in analysis and interpretation remain, with few exceptions (BARADAT and YAZDANI, 1988; MUZIKA et al., 1990), unacknowledged. In our 1988 paper, we suggested a variety of quantitative methods which might be developed to assist the interpretation of compositional data. The purpose of this paper is to remind those working on the topic that use of "standard" analyses will seldom be valid (AITCHISON, 1984), and to demonstrate quantitative methods that are appropriate to the analysis and interpretation of data describing resin composition.

2. The Nature of Resin Data

As most work has been concerned with the composition of coniferous resin, we shall use the term "resin data" as a rubric, to refer to any which describes the composition of plant resins or essential oils, including those from non-coniferous species (SMITH et al., 1988; BARTON et al., 1991). The results of gas-chromatographic analysis, on which these studies are based, can be reported in a number of forms (SQUILLACE, 1976; WHITE and NILSSON, 1983; BIRKS and KANOWSKI, 1988); the quantity of each constituent can be expressed as:

1. a proportion of total resin (SCHÖNWITZ et al., 1990);
2. a proportion of comparable constituents (of all mono- or sesqui-terpenes; FADY et al., 1992);
3. weight per unit volume of resin (mg/ml; CHANG and HANOVER, 1991);
4. weight per unit weight of tissue (μg/g fresh weight; SCHÖNWITZ et al., 1990).

Appropriate analyses and interpretation of resin data are *not* independent of the form in which they are reported (SQUILLACE, 1976; WHITE, 1983; BARADAT and YAZDANI, 1988; BIRKS and KANOWSKI, 1988; SCHÖNWITZ et al., 1990); 1 of the limitations of too many papers in the subject area is their imprecision about the form in which the data were expressed, with consequent ambiguity about the results reported.

Most studies have been based on proportions of comparable constituents or on weight per weight of tissue, generally described as proportions and absolute amounts, respectively. These data have been used for a number of purposes:

1. to investigate biosynthetic processes (BERNARD-DAGAN, 1988; COOL and ZAVARIN, 1992);
2. to assist taxonomic study (FORREST, 1987; VON RUDLOFF and LAPP, 1991; COOL and ZAVARIN, 1992; FADY et al., 1992; SQUILLACE and PERRY, 1992);
3. to infer mechanisms of genetic control and frequencies of genotypes (BARADAT and YAZDANI, 1988; ZAVARIN et al., 1991; review by BIRKS and KANOWSKI, 1988).

The utility of the different forms of resin data varies with the purpose of analysis. The prevailing paradigm, expressed by WHITE and NILSSON (1983), HARBONE and TURNER (1984), HALL and LANGENHEIM (1987), or SCHÖNWITZ et al. (1990), is summarized in the words of the latter: "terpene variability based on relative amounts is probably due to genetic variation, while the variability of absolute amounts will reflect environmental factors". For this reason, taxonomic and genetic studies have been based almost exclusively on proportional data, although CHANG and HANOVER (1991) have suggested that "the absolute amount of monoterpenes is a more sensitive method to analyze species with a homogenous terpene composition pattern". However, the constraint, Σ(proportions) = 1, inherent in proportional data, limits its utility for other purposes — in the investigation of biosynthetic processes, for which WHITE (1983) found that the effect of "expressing results as percentages . . . was to obscure possible biosynthetic relationships", and in quantitative genetic studies where, as BARADAT and YAZDANI (1988) noted, "autocorrelations between terpenes may cause some bias in estimates of genetic parameters". The considerable difficulties inherent in defining segregating characters from resin compositional data, especially in the case of proportional data, have also been reviewed, by WHITE and NILSSON (1983) and BIRKS and KANOWSKI (1988). The cautionary words of MUZIKA et al. (1990), that "researchers studying monoterpenes should closely assess the technique appropriate for a given species as well as direct the analysis towards the questions addressed", are particularly apposite, but there is little evidence that they have been acknowledged, much less heeded, by some workers.

We assume here that proportional (or, in AITCHISON's (1984) terminology, compositional) data will continue to be that of most interest for forest taxonomists and geneticists; those more concerned with biosynthetic pathways will doubtless favour data expressed in "absolute" terms. Recent taxonomic and genetic studies based on resin data, almost invariably expressed as proportions, have used a variety of the standard techniques of numerical taxonomy: various discriminant analysis (SCHILLER and GRUNWALD, 1987; VON RUDLOFF et al., 1988; ZAVARIN et al., 1989, 1990, 1991; COOL et al., 1991; COOL and ZAVARIN, 1992; FADY et al., 1992), principal component analysis (BARADAT and YAZDANI, 1988; VON RUDLOFF et al., 1988; COOL et al., 1991; FADY et al., 1992; SQUILLICE and PERRY, 1992), cluster analyses (SCHILLER and GRUNWALD, 1987; BARADAT and YAZDANI, 1988; SCHÖNWITZ et al., 1990; CHANG and HANOVER, 1991; VON RUDLOFF and LAPP, 1991), and uni- and multi-variate analyses of variance (BARADAT and YAZDANI, 1988; MICHELOZZI et al., 1990; CHANG and HANOVER, 1991; ZAVARIN et al., 1991; FADY et al., 1992). In AITCHISON's (1984) terminology, these are "standard" methods which are "improper and inadequate" if applied to compositional date. AITCHISON's (1984) exposition of the nature of compositional data and of the dangers of its simplistic interpretation is so clear and concise that we can not do better than to refer the reader to it. He noted 3 major interpretative difficulties in such data — their high dimensionality, the absence of an interpretable covariance structure, and the limits of parametric modelling applied to them — and suggested how these might be addressed. He noted that the work of geo-

logical researchers, at whom his entreaties were directed, had taken little cognizance of the nature of their compositional data sets; regrettably, it appears that the same criticism applies to most scientists working with resin data. In the following sections, we explain AITCHISON's (1984, 1986) proposals for more appropriate analytical methods, and apply them to examples of resin compositional data.

3. Issues in the Analysis of Compositional Data

Compositional data with n observations of an m-part composition are subject to the constraint:

$$x_{il} + \ldots + x_{im} = 1 \qquad (1)$$

where x_{ij} is the proportion of constituent j in sample i. Each data vector is completely specified by m-1 components. The appropriate sample space for the m-part composition is the (m-1) dimensional simplex embedded in m-dimensional Euclidean space. A simple example of a simplex is the 1-dimensional simplex, the line $x_1 + x_2 = 1$, on which would be found all possible compositions with only 2 components. This line is embedded in 2-dimensional Euclidean space, but x_1 and x_2 are not free to roam over this space.

The effect of this constraint is to limit severely the utility of standard statistical procedures applied to these raw proportions.

3.1 Correlations between proportions

The correlations between the proportions, which are the basis of inferences about biosynthetic pathways and their genetic control (LAPP and VON RUDLOFF, 1982; BERNARD-DAGAN, 1988; CHANG and HANOVER, 1991; COOL and ZAVARIN, 1992), do not have a simple interpretation. The constraint places restrictions on the correlations, giving a bias towards negative values. This is most easily appreciated in the 2 component case where the correlation can only be —1 instead of being free to range from —1 to +1.

3.2 Subcompositions

It is possible to extract, from all terpenes that constitute the resin, subsets which are of particular interest. There are many justifiable reasons for studying subcompositions, and indeed the monoterpenes themselves are a subcomposition of the resin, but the use of subcompositional data further complicates the analysis and interpretation of compositional data, for 2 reasons.

The first is the form in which subcompositional data are expressed. The proportion of each constituent of a subcomposition may be expressed as that of the whole composition (ie x_{ij} as defined in Equation 1; SCHÖNWITZ et al., 1990; COOL and ZAVARIN, 1992), or the original proportions may be rescaled so that the components of the subcomposition sum to unity (LAPP and VON RUDLOFF, 1982). There is no functional relationship between the 2 subcompositional data sets, nor is there any presumptive reason for giving preference to 1 form.

A related issue concerns the absence of any functional relationships between the variances and the covariances of subcompositional data sets expressed on different bases. The 2 subcompositional data sets described in the paragraph above would have different, functionally unrelated, covariance matrices, although each purports to describe the same subcomposition.

3.3 The lognormal distribution

Valid methods of analysis of compositional data must resolve the difficulties described above. The simplex, within which proportional data are constrained, is difficult to work with. Although no classes of distributions have been discovered which describe the variability observed in untransformed compositional data, the class of additive-logistic normal distributions has been found to be useful and tractable (AITCHISON, 1986).

In order to simplify notation subsequently, we redefine x_i to be the proportion of constituent i in any given sample. As above, $\Sigma x_i = 1$. The data matrix x has a lognormal distribution if the logratio vector:

$$y_{ij} = \log_e(x_i/x_j), \quad i \neq j \qquad (2)$$

has a m-1 dimensional normal distribution, $N^d(\mu,\Sigma)$. The logratio transformation removes the constraint $\Sigma x_i = 1$, but imposes the additional restriction that $x_i > 0$, i = 1, ... m.

The logratio transformation confers the important property on constituents of independence of subcomposition, because the ratio of 2 components is invariant. This property is essential if robust inferences are to be drawn regarding, for example, biosynthetic pathways or the nature of the genetic control of the resin composition.

The 2nd relevant property of logratios concerns the set of variances of and covariances between logratios, termed the covariance matrix. The covariance matrix is important as it has a central role in the multivariate procedures appropriate for the interpretation of logratio data sets.

3.4 The covariance structure of a composition

It is possible to define m(m-1)/2 logratios for an m-part composition, and the covariance structure is based on the set of covariances:

$$\text{cov}\{\log(x_i/x_k), \log(x_j/x_l)\} \quad i,j,k,l = 1,\ldots m. \qquad (3)$$

It appears from equation (3) that there are m^4 covariances, but it can be shown that they are not independent of each other and that the complete set can be constructed from the set of logratio variances (AITCHISON, 1986):

$$v_{ij} = \text{var}\{\log(x_i/x_j)\} \qquad (4)$$

There are other covariances that can be defined; AITCHISON (1986) explored the range of possibilities and set out the relationships between the various forms. One form is the logratio covariance matrix, Σ, which completely defines the covariance structure of a composition:

$$\Sigma = [\sigma_{ij}] = [\text{cov}\{\log(x_i/x_m, \log(x_j/x_m)\}], \qquad (5)$$

$$i,j = 1,\ldots(m-1).$$

Any variance or covariance of logratios can be constructed from Σ by using the relationship:

$$\text{cov}\{\log(x_i/x_j),\log(x_k/x_l)\} = \sigma_{ik} - \sigma_{il} - \sigma_{jk} + \sigma_{jl}. \qquad (6)$$

A second form is described as the logcentred covariance matrix:

$$\Gamma = \mathrm{cov}\{\log(x_i/g(x)),\log(x_j/g(x))\}, \quad i,j = 1,...m. \tag{7}$$

where g(x) is the geometric mean of the m proportions of the composition:

$$g(x) = (x_1...x_m)^{1/m}. \tag{8}$$

It may appear from the above that x_m has been given a special place in the definition of the logratio covariance matrix, but it can be shown that the results of any multivariate procedure used in this paper are invariant under any permutation of the components (AITCHISON, 1986). Σ and Γ are functionally related and, therefore, can be used variously, depending on the purpose of the analysis.

3.5 Zero proportions

Logratios involving a particular constituent cannot be calculated if the proportion of that constituent is zero. Zero values may be due to amounts so small that they are not detected, or to a compound not being present; both cases may have important implications, and are common in resin data sets.

It is necessary to restrict analyses based on logratios to subcompositions with no zero values. The information from data sets which include zero values should not be discarded, but different forms of analysis are required for those data sets in which not all components are present in all samples. A data set recording the presence or absence of components could be analyzed, for example, using procedures based on contingency tables to investigate associations between the absences (EVERITT, 1977).

3.6 The analysis of a compositional data set

If the set of logratios has a multivariate normal distribution (AITCHISON, 1986), the whole range of procedures based on multivariate normality becomes available.

The first step in the analysis of compositional data is to calculate the set of means and variances of the set of logratios that define a group of individuals, be that group a provenance, family, or clone. This is a descriptive device that allows patterns of variability to be explored. The next step is to reformulate the problem to be addressed in terms of the logratios, and to make use of standard multivariate procedures. AITCHISON (1986) recommended the use of the logratio covariance matrix for this purpose.

4. Illustrative Examples

These methods are demonstrated below using data from 2 experiments of *Pinus elliottii* var *elliottii* LITTLE and DORMAN growing in Zimbabwe. The number of observations is not extensive, but these data sets were chosen to provide clear and simple examples of the appropriate methodology for the analysis of resin compositional data.

4.1 Experimental material and analytical procedures

Samples of xylem resin were collected from 2 sets of material, a progeny trial and a clonal seed orchard, in the Zimbabwean *P. elliotii* breeding programme (BARNES, 1986). Three full-sib families were sampled in the progeny trial; by chance, they had the same female parent, and were identified locally as families 22 x 49, 22 x 130, and 22 x 173. The families had been established 15 years previously in an experiment replicated at 2 sites, identified as A and B, in the Eastern Highlands of Zimbabwe. The experiments had previously been thinned silviculturally; at the time of sampling, 3 trees remained in each of 3 replications at each site. Samples were collected at a height of 0.5 m.

Xylem resin of ramets of the male parents of the full-sib families (parents 49, 130 and 173) was also sampled, from 2 seed orchards in Zimbabwe's Eastern Highlands. Samples were collected at breast height from each of 10 ramets in line plots of each clone, in each of the 2 orchards, also identified as A and B, 13 years after grafting.

Samples were taken on the north side of all trees. The procedures and equipment used to collect, store and analyze samples were those described by LOCKHART (1990). Data were reported as proportions of all terpenes detected.

The following terpenes were detected in measurable quantity in some or all samples: α-pinene, camphene, β-pinene, sabinene, myrcene, α-phellandrene, Δ-3-carene, limonene, β-phellandrene, terpinolene, estragole, and caryophyllene. Other terpenes — sabinene and terpinolene — were present in trace amounts, but not in every sample. Analyses of the clonal and family data sets were restricted to the subcompositions of monoterpenes present in samples of that data set: these subcompositions, which are identified in *table 1*, comprised 8 components in the case of the

Table 1. — Monoterpene constituents used in analysis of subcompositions.

Monoterpene	Component	
	Progeny trial	Clonal orchard
α-pinene	x_1	x_1
camphene	x_2	x_2
β-pinene	x_3	x_3
myrcene	x_4	x_4
limonene	x_5	x_5
β-phellandrene	x_6	x_6
estragole	x_7	x_7
α-phellandrene	-	x_8

Table 2. — Percentages of 8 monoterpenes in the terpene component of the oleoresin of 3 clones at 2 sites (A and B).

site & clone	α-pinene	camphene	β-pinene	myrcene	limonene	β-phellandrene	estragole	α-phellandrene
A 49	54.74	0.89	5.69	1.34	1.13	33.13	1.50	0.78
	53.28	0.86	7.12	1.33	1.12	33.24	1.67	0.78
	50.74	0.86	6.06	1.42	1.22	36.10	1.78	0.86
	50.59	0.90	7.23	1.40	1.21	34.57	2.57	0.86
	60.72	0.84	6.21	1.19	0.95	27.95	1.03	0.62
	36.47	0.71	32.92	1.38	1.12	25.70	0.40	0.67
	55.64	0.76	7.28	1.13	0.98	25.68	1.35	0.63
	73.13	2.06	11.36	1.56	1.01	7.79	1.02	0.18
	54.27	0.80	4.53	1.23	0.95	31.12	1.74	0.72
	53.79	0.82	5.83	1.24	0.94	30.94	1.22	0.72
A 130	18.22	0.71	23.04	1.69	1.39	49.93	2.29	1.20
	15.10	0.57	35.83	1.51	1.21	38.79	1.33	0.94
	15.84	0.64	24.65	1.57	1.36	45.16	1.89	1.07
	16.13	0.65	24.90	1.58	1.91	45.09	1.81	1.08
	17.81	0.66	22.47	1.61	1.31	46.49	2.51	1.12
	19.33	0.69	21.11	1.65	1.52	49.26	1.98	1.19
	19.61	0.69	24.65	1.65	1.49	47.10	1.09	1.15
	19.68	0.72	19.73	1.66	1.54	50.86	2.02	1.18
	17.28	0.63	27.63	1.55	1.42	43.59	1.04	1.06
	17.88	0.68	25.50	1.64	1.37	46.45	1.43	1.15
A 173	29.55	0.66	42.44	1.21	1.17	19.66	2.05	0.42
	20.04	0.40	38.48	1.54	1.30	33.43	0.72	0.85
	26.63	0.65	42.05	1.30	1.01	24.12	2.19	0.58
	31.40	0.67	36.93	1.29	0.94	22.45	1.51	0.52
	22.67	0.55	42.53	1.24	1.11	23.16	1.54	0.55
	20.82	0.59	38.33	1.26	1.28	27.58	1.48	0.60
	33.62	0.64	41.98	1.03	1.03	12.67	1.17	0.28
	33.59	0.63	42.30	1.05	0.73	14.31	1.25	0.31
	35.76	0.68	43.18	1.14	1.13	14.05	2.75	0.33
	31.18	0.67	36.84	1.25	1.10	22.07	2.27	0.51
B 49	50.87	0.95	3.96	1.41	1.13	34.69	2.85	0.81
	51.21	0.96	4.88	1.46	1.19	34.79	2.59	0.85
	56.75	0.89	6.35	1.28	1.06	29.51	2.23	0.67
	50.79	1.00	5.60	1.49	1.15	34.35	3.28	0.83
	51.66	0.96	6.40	1.50	1.12	33.96	1.82	0.84
	52.34	0.96	10.16	1.38	1.14	29.13	2.89	0.67
	51.60	0.93	6.28	1.40	1.02	31.03	3.91	0.75
	51.35	0.93	6.01	1.48	1.24	33.28	3.26	0.84
	47.01	0.94	8.07	1.60	1.20	34.31	1.73	0.93
B 130	24.35	0.76	24.52	1.75	1.37	41.52	2.72	1.15
	25.76	0.76	26.86	1.69	1.32	38.81	1.90	1.07
	25.15	0.80	25.13	1.77	1.36	39.98	1.92	1.17
	26.59	0.80	22.45	1.73	1.30	41.93	2.28	1.15
	21.48	0.73	27.65	1.68	1.40	43.30	1.47	1.12
	22.35	0.75	32.99	1.77	1.31	37.34	1.75	1.05
	26.75	0.83	25.51	1.79	1.18	39.69	2.03	1.11
	27.03	0.87	22.74	1.87	1.32	41.53	1.75	1.22
	23.55	0.83	23.45	1.92	1.32	44.26	2.11	1.31
	25.52	0.83	25.37	1.86	1.44	40.93	1.74	1.22
B 173	33.25	0.73	35.35	1.39	0.98	22.18	3.47	0.55
	30.07	0.66	42.46	1.36	1.54	19.30	2.10	0.48
	32.38	0.81	38.92	1.52	1.28	19.28	3.20	0.49
	31.78	0.71	39.72	1.39	1.30	19.38	2.38	0.51
	29.09	0.69	35.49	1.36	1.24	23.12	2.75	0.59
	28.39	0.73	38.85	1.55	1.23	23.05	2.78	0.63
	30.79	0.79	37.40	1.50	1.19	21.69	3.95	0.57
	25.87	0.64	37.73	1.35	1.22	23.28	2.94	0.59
	29.91	0.79	34.27	1.56	1.04	25.27	3.28	0.71
	40.67	0.79	37.45	1.23	1.40	13.13	3.71	0.39

seed orchard, and 7 in the case of the progeny trial. These data sets are presented as *tables 2* and *3*, respectively.

The mean and variances of logratios between the 8 or 7 components in the subcomposition formed the summary data sets, and are presented in *tables 4* to *7*. Examination of *tables 4* to *7* shows some features of the compositions that might command our attention. There are logratios, for example those of α-pinene to camphene, that are reasonably constant across site and clone, or across site and family. This consistency would not have been apparent from the compositional data, as the proportions of α-pinene and camphene vary quite widely in the data sets, as evident from *tables 2* and *3*.

It is also apparent from *tables 4* to *7* that the variation within families is greater than that within clones. The mean logratios are more consistent across the families, which are all half-sibs, than across the clones, which are unrelated. The logratios appear to be similar across sites for the same family or clone.

Table 3. — Percentages of 7 monoterpenes in the terpene component of the oleoresin of 3 families at 2 sites (A and B).

site & family	rep	tree	α-pinene	camphene	β-pinene	myrcene	limonene	β-phellandrene	estragole
A	1	1	81.54	1.07	8.80	1.11	0.72	0.81	0.70
22x49		2	68.29	1.14	25.36	1.12	0.88	1.86	1.03
		3	41.14	0.70	23.53	1.38	0.81	26.18	4.20
	2	1	49.36	0.76	22.72	1.01	0.59	19.18	0.97
		2	75.23	1.13	17.00	0.73	0.78	1.72	0.67
		3	56.59	1.01	30.28	0.92	0.80	8.44	1.28
	3	1	71.36	1.70	21.29	1.56	1.40	1.69	0.46
		2	71.90	1.14	19.84	1.08	0.92	2.06	1.12
		3	65.16	1.34	24.26	1.38	1.05	2.29	2.43
A	1	1	59.39	0.95	43.26	1.38	1.14	8.97	2.13
22x130		2	68.27	1.32	25.28	1.20	1.10	2.03	0.48
		3	77.32	1.87	13.49	1.60	0.63	1.07	1.86
	2	1	74.52	1.43	20.02	1.32	0.58	1.29	0.51
		2	62.97	1.16	28.89	0.44	0.55	0.58	1.05
		3	77.62	1.28	11.19	0.92	0.50	2.93	4.62
	3	1	78.17	1.16	13.74	0.87	0.94	1.75	2.87
		2	67.67	1.43	24.89	1.32	1.18	1.76	0.70
		3	57.37	1.58	22.32	1.88	1.19	11.80	1.78
A	1	1	51.59	0.99	39.34	1.13	1.07	3.89	1.66
22x173		2	64.53	1.17	19.16	1.19	0.99	11.28	0.95
		3	61.91	0.87	25.46	0.81	1.06	1.60	1.76
	2	1	62.52	1.24	30.57	0.87	0.90	1.09	1.27
		2	58.66	1.23	34.28	1.20	1.23	1.42	1.37
		3	69.64	1.58	22.75	1.30	0.77	0.81	2.15
	3	1	87.22	1.61	6.50	1.46	0.69	1.03	0.75
		2	77.91	1.34	14.94	1.23	1.92	0.63	0.58
		3	67.10	1.29	24.60	1.14	1.08	1.70	1.03
B	1	1	80.54	0.98	9.95	0.83	0.66	0.44	4.90
22x49		2	80.54	1.60	9.55	1.44	1.05	1.46	3.20
		3	68.66	0.94	25.06	0.86	1.53	0.80	2.15
	2	1	70.33	1.05	14.27	1.09	11.96	0.94	0.10
		2	92.38	0.83	3.89	0.66	0.55	0.51	0.78
		3	90.34	1.26	4.81	0.90	0.85	1.01	0.84
	3	1	70.47	0.86	24.06	0.83	0.88	0.86	2.03
		2	88.77	1.07	5.66	0.82	0.90	0.60	2.18
		3	71.93	0.87	23.59	0.67	0.98	0.62	1.40
B	1	1	77.33	0.10	18.52	0.10	2.49	1.66	0.10
22x130		2	83.66	1.17	9.02	1.07	2.63	0.93	1.52
		3	75.59	1.14	19.59	0.87	1.39	1.06	0.36
	2	1	76.97	1.17	16.51	0.92	1.58	0.96	1.89
		2	78.58	1.07	9.78	0.91	5.45	0.79	2.86
		3	69.02	0.76	19.47	0.88	6.25	0.92	2.70
	3	1	72.73	0.98	20.52	0.10	1.56	1.30	2.91
		2	71.10	1.04	16.65	0.94	2.20	1.09	2.60
		3	75.04	1.01	17.87	0.95	2.69	1.01	1.43
B	1	1	70.08	1.12	24.76	0.95	1.40	0.80	0.89
22x173		2	74.17	1.08	20.83	0.89	1.03	1.10	0.91
		3	65.06	1.00	29.39	0.83	1.41	1.24	1.08
	2	1	83.64	1.12	7.12	0.91	1.42	0.95	1.37
		2	65.03	0.95	29.55	0.95	1.67	1.03	0.82
		3	73.51	1.01	21.41	0.88	1.13	0.90	1.17
	3	1	69.22	0.99	25.77	0.88	1.59	0.87	0.51
		2	83.26	1.43	10.87	1.23	1.61	1.04	0.45
		3	69.73	1.04	24.16	0.68	1.13	1.01	0.74

By investigating different subcompositions it is possible to discover those responsible for the variation observed in the total composition. The percentage variation of three-component subcompositions compared to the 8-component composition for the 3 clones is shown in *table 8.* These results suggest that dimension-reducing procedures, such as principal component analysis or canonical variates analysis, could be useful.

The next stage is to redefine the particular problems in terms of the logratio covariance structure and to apply the appropriate multivariate procedure.

4.2 An example of discriminant analysis

As we noted in 2. above, many resin compositional data sets are used in taxonomic studies, in which standard multivariate methods, such as principal component or canonical variate analysis, are used to discriminate between samples. These multivariate methods depend on the assumption of multivariate normality, although they are fairly robust to small departures from it. There is also an assumption of homogeneity of covariance across groups. When these procedures involve the comparison of means and covariance matrices then significance levels are only valid when the distributional assumptions apply. These assumptions have not been tested in most analyses of resin compositional data, and are unlikely to hold considering the variety of distributions evident in resin data sets, and the prevalence of multimodal distributions (from which single gene control has been inferred).

Table 4. — Mean logratios between 8 components of 3 clones at 2 sites (A and B).

Site	A	B	A	B	A	B
Clone	49	49	130	130	173	173
$\log(x_1/x_2)$	4.08	4.00	3.28	3.44	3.83	3.75
$\log(x_1/x_3)$	1.93	2.12	-0.34	-0.03	-0.37	-0.20
$\log(x_1/x_4)$	3.71	3.58	2.39	2.63	3.13	3.09
$\log(x_1/x_5)$	3.93	3.81	2.50	2.92	3.27	3.22
$\log(x_1/x_6)$	0.69	0.45	-0.96	-0.50	0.31	0.40
$\log(x_1/x_7)$	3.72	2.97	2.36	2.55	2.87	2.33
$\log(x_1/x_8)$	4.43	4.17	2.76	3.07	4.09	4.04
$\log(x_2/x_3)$	-2.15	-1.88	-3.62	-3.47	-4.20	-3.94
$\log(x_2/x_4)$	-0.38	-0.42	-0.89	-0.81	-0.70	-0.66
$\log(x_2/x_5)$	-0.16	-0.18	-0.78	-0.52	-0.56	-0.52
$\log(x_2/x_6)$	-3.39	-3.54	-4.24	-3.94	-3.52	-3.34
$\log(x_2/x_7)$	-0.36	-1.03	-0.92	-0.89	-0.96	-1.41
$\log(x_2/x_8)$	0.35	0.18	-0.52	-0.37	0.26	0.30
$\log(x_3/x_4)$	1.77	1.46	2.73	2.66	3.50	3.28
$\log(x_3/x_5)$	1.99	1.70	2.84	2.95	3.64	3.42
$\log(x_3/x_6)$	-1.24	-1.66	-0.63	-0.47	0.68	0.60
$\log(x_3/x_7)$	1.79	0.85	2.69	2.58	3.24	2.53
$\log(x_3/x_8)$	2.50	2.06	3.10	3.09	4.46	4.24
$\log(x_4/x_5)$	0.22	0.24	0.11	0.29	0.14	0.14
$\log(x_4/x_6)$	-3.01	-3.12	-3.36	-3.13	-2.82	-2.68
$\log(x_4/x_7)$	0.01	-0.60	-0.04	-0.09	-0.26	-0.75
$\log(x_4/x_8)$	0.73	0.60	0.37	0.43	0.96	0.96
$\log(x_5/x_6)$	-3.23	-3.36	-3.47	-3.42	-2.95	-2.82
$\log(x_5/x_7)$	-0.21	-0.84	-0.15	-0.38	-0.40	-0.89
$\log(x_5/x_8)$	0.51	0.36	0.26	0.14	0.82	0.82
$\log(x_6/x_7)$	3.03	2.52	3.32	3.05	2.56	1.93
$\log(x_6/x_8)$	-0.51	-0.36	-0.26	-0.14	-0.82	-0.82
$\log(x_7/x_8)$	0.71	1.20	0.41	0.52	1.22	1.71

Table 5. — Variances of logratios between 8 components of 3 clones at 2 sites (A and B).

Site	A	B	A	B	A	B
Clone	49	49	130	130	173	173
$\log(x_1/x_2)$	.045	.005	.002	.003	.018	.009
$\log(x_1/x_3)$	.447	.077	.058	.030	.041	.019
$\log(x_1/x_4)$	.037	.012	.004	.006	.093	.028
$\log(x_1/x_5)$	.056	.009	.019	.012	.115	.026
$\log(x_1/x_6)$	.319	.012	.003	.010	.258	.085
$\log(x_1/x_7)$	.212	.073	.095	.026	.119	.024
$\log(x_1/x_8)$	.339	.023	.002	.009	.288	.072
$\log(x_2/x_3)$	.379	.076	.052	.024	.026	.014
$\log(x_2/x_4)$	.060	.003	.001	.001	.064	.008
$\log(x_2/x_5)$	.104	.003	.013	.008	.077	.029
$\log(x_2/x_6)$	.528	.004	.000	.004	.187	.049
$\log(x_2/x_7)$	.331	.074	.081	.030	.080	.023
$\log(x_2/x_8)$	.545	.011	.000	.002	.219	.037
$\log(x_3/x_4)$	.281	.075	.037	.018	.025	.011
$\log(x_3/x_5)$	.299	.077	.059	.015	.037	.008
$\log(x_3/x_6)$	.702	.101	.057	.023	.126	.046
$\log(x_3/x_7)$	1.025	.183	.180	.059	.146	.058
$\log(x_3/x_8)$	.675	.105	.054	.028	.148	.042
$\log(x_4/x_5)$	.008	.003	.014	.005	.019	.032
$\log(x_4/x_6)$	.252	.002	.002	.003	.044	.021
$\log(x_4/x_7)$	.263	.091	.085	.030	.192	.042
$\log(x_4/x_8)$	.259	.004	.002	.002	.056	.016
$\log(x_5/x_6)$	.185	.004	.014	.004	.068	.080
$\log(x_5/x_7)$	.237	.087	.102	.036	.185	.083
$\log(x_5/x_8)$	.188	.008	.014	.005	.083	.073
$\log(x_6/x_7)$	.290	.088	.074	.030	.324	.082
$\log(x_6/x_8)$	.002	.003	.000	.002	.003	.003
$\log(x_7/x_8)$	.313	.107	.080	.030	.354	.066
total variance	8.381	1.320	1.104	.452	3.395	1.096

There is little point in attempting to use a dimension reducing technique unless there are high correlations between some of the original variates. As discussed in 3.2 above, further difficulties should have been evident to those attempting to calculate the correlation matrices using the raw proportions, as illustrated empirically in the

Table 6. — Mean logratios between 7 components of 3 families at 2 sites (A and B).

Site	A	B	A	B	A	B
Family	22x49	22x49	22x130	22x130	22x173	22x173
$\log(x_1/x_2)$	4.07	4.34	3.94	4.58	3.99	4.21
$\log(x_1/x_3)$	1.23	1.98	1.24	1.60	1.15	1.30
$\log(x_1/x_4)$	4.04	4.50	4.15	4.68	4.09	4.38
$\log(x_1/x_5)$	4.30	4.20	4.45	3.33	4.19	3.98
$\log(x_1/x_6)$	2.88	4.65	3.596	4.300	3.73	4.30
$\log(x_1/x_7)$	4.02	4.08	4.08	4.19	4.12	4.46
$\log(x_2/x_3)$	-2.95	-2.35	-2.70	-2.98	-2.84	-2.91
$\log(x_2/x_4)$	-0.04	0.16	0.22	0.10	0.10	0.18
$\log(x_2/x_5)$	0.22	-0.14	0.52	-1.25	0.20	-0.23
$\log(x_2/x_6)$	-1.19	0.31	-0.34	-0.28	-0.27	0.09
$\log(x_2/x_7)$	-0.06	-0.26	0.14	-0.39	0.13	0.25
$\log(x_3/x_4)$	2.91	2.51	2.92	3.08	2.94	3.09
$\log(x_3/x_5)$	3.17	2.21	3.22	1.73	3.04	2.68
$\log(x_3/x_6)$	1.75	2.66	2.36	2.70	2.57	3.00
$\log(x_3/x_7)$	2.89	2.09	2.84	2.59	2.97	3.17
$\log(x_4/x_5)$	0.26	-0.30	0.30	-1.35	0.10	-0.41
$\log(x_4/x_6)$	-1.15	0.15	-0.56	-0.38	-0.37	-0.09
$\log(x_4/x_7)$	-0.02	-0.42	-0.08	-0.48	0.03	0.08
$\log(x_5/x_6)$	-1.42	0.45	-0.86	0.972	-0.47	0.32
$\log(x_5/x_7)$	-0.28	-0.12	-0.38	0.86	-0.07	0.49
$\log(x_6/x_7)$	1.14	-0.57	0.48	-0.11	0.39	0.17

Table 7. — Variances of logratios between 7 components of 3 families at 2 sites (A and B).

Site	A	B	A	B	A	B
Family	22x49	22x49	22x130	22x130	22x173	22x173
$\log(x_1/x_2)$	.279	.045	.036	.801	.018	.009
$\log(x_1/x_3)$	.261	.702	.216	.144	.405	.324
$\log(x_1/x_4)$	.117	.081	.198	.720	.018	.018
$\log(x_1/x_5)$	.072	.963	.180	.342	.108	.045
$\log(x_1/x_6)$	2.043	.171	.765	.054	.855	.027
$\log(x_1/x_7)$	.720	1.269	.621	1.629	.252	.144
$\log(x_2/x_3)$	.216	.657	.207	1.080	.414	.324
$\log(x_2/x_4)$	.090	.018	.099	.018	.009	.009
$\log(x_2/x_5)$	.027	.837	.153	1.170	.153	.036
$\log(x_2/x_6)$	2.052	.072	.648	1.206	.927	.036
$\log(x_2/x_7)$	.729	1.287	.684	.639	.243	.189
$\log(x_3/x_4)$	.162	.567	.378	.981	.423	.324
$\log(x_3/x_5)$	.144	.918	144	.531	.297	.279
$\log(x_3/x_6)$	1.107	.576	.873	.081	.837	.243
$\log(x_3/x_7)$	.387	1.701	1.323	2.016	.234	.405
$\log(x_4/x_5)$	.045	.702	.171	.972	.135	.027
$\log(x_4/x_6)$	1.503	.054	.432	1.107	.810	.045
$\log(x_4/x_7)$	.423	1.386	.909	.531	.270	.198
$\log(x_5/x_6)$	1.773	.720	.504	.468	.918	.054
$\log(x_5/x_7)$	.558	3.735	1.008	1.305	.360	.216
$\log(x_6/x_7)$	.936	1.593	.999	2.142	.990	.144
total variance	13.644	18.054	10.548	17.937	8.676	3.096

Table 8. — Percentage variation, estimated from the variance of logratios, of 3-component subcompositions compared to the 8-component terpene composition of 3 clones at 2 sites.

Site	A	B	A	B	A	B
Clone	49	49	130	130	173	173
Subcomposition of β-pinene,estragole, and α-phellandrene (x_3, x_7, x_8)						
	24.0	29.9	28.4	25.9	19.1	15.1
Subcomposition of α-pinene, camphene, and myrcene (x_1, x_2, x_4)						
	1.7	1.5	0.6	2.2	5.2	4.1
Subcomposition of α-pinene, camphene, and estragole (x_1, x_2, x_7)						
	7.0	11.5	16.1	13.1	6.4	5.1

following example. The correlation matrices for 4 different 4-component subcompositions are shown in *table 9*. In each case, the proportions have been expressed on 2 different bases, the total terpene basis and the subcomposition basis. There are other possible bases on which proportions could be calculated; none of them can claim to be definitive. The variability in the correlations presented in *table 9* illustrates the severe limitations of basing inferences on the results of analyses of raw proportions, whatever their basis. In order to define a useful and robust correlation structure ,the logratio covariance matrix, outlined in 3.4 above, should be used.

Although small, the clonal data set (*Table 2*) has sufficient in common with the data sets used in discriminant studies to be of value in demonstrating an appropriate ap-

Table 9. — Comparison of pooled within-clone correlations between proportions of 4-part subcompositions, estimated with respect to 2 bases, total terpene and subcomposition.

Subcomposition of α-pinene, camphene, myrcene and limonene (x_1, x_2, x_4, x_5)						
x_1, x_2, x_4, x_5 are proportions of total terpene				$x_1 + x_2 + x_4 + x_5 = 1$		
	x_1	x_2	x_4	x_1	x_2	x_4
x_2	.53			-.73		
x_4	-.15	.45		-.98	.66	
x_5	-.35	.06	.67	-.89	.52	.85
Subcomposition of β-pinene, β-phellandrene, estragole, α-phellanarene (x_3, x_6, x_7, x_8)						
x_3, x_6, x_7, x_8 are proportions of total terpene				$x_3 + x_6 + x_7 + x_8 = 1$		
	x_3	x_6	x_7	x_3	x_6	x_7
x_6	.07			.03		
x_7	-.58	.25		-.98	-.04	
x_8	-.37	-.07	.09	-.23	.07	.05
Subcomposition of α-pinene, β-pinene, limonene, estragole (x_1, x_3, x_5, x_7)						
x_1, x_3, x_5, x_7 are proportions of total terpene				$x_1 + x_3 + x_5 + x_7 = 1$		
	x_1	x_3	x_5	x_1	x_3	x_5
x_3	-.29			-.39		
x_5	-.35	-.52		-.35	-.54	
x_7	-.39	-.58	.92	-.39	-.59	.93
Subcomposition of camphene, myrcene, β-phellandrene and α-phellandrene (x_2, x_4, x_6, x_8)						
x_2, x_4, x_6, x_8 are proportions of total terpene				$x_2 + x_4 + x_6 + x_8 = 1$		
	x_2	x_4	x_6	x_2	x_4	x_6
x_4	.45			.55		
x_6	-.08	.25		.22	.65	
x_8	.31	.01	-.07	-.63	-.89	.82

proach to such data. The logratio data set was calculated with respect to the proportion of α-pinene. Standard statistics (ANDREWS et al., 1973) were calculated to test for multivariate normality of the logratios, and no significant deviations from normality were detected. There was, however, significant heterogeneity of covariance between sites for the same clone, and between clones within the same site. This is not expected to have a large effect on the following analyses.

The canonical variate analyses was carried out using the seven logratios. This technique is an extension of discriminant analysis, and is used for the investigation of differences among a number of populations. The plot of the first 2 canonical variates is shown as *figure 1*, from which a clear separation of the 3 clones, but only a partial separation between the same clones on different sites, is evident.

In conclusion, the raw proportions are neither normally distributed nor of homogenous covariance, and therefore fail both assumptions inherent in standard multivariate analyses. Further, results of analyses of the raw proportions depend on which subcomposition is used. In contrast, analyses based on logratio data are more likely to satisfy the assumption of multivariate normality and are independent of the basis on which proportions are defined.

4.3 An example of analysis of variance

Resin compositional data from genetic trials allows the opportunity for estimation of genetic parameters, for those metric traits whose distribution approximates the Normal. Neither assumption applies for the case of raw proportional data; the latter has often been addressed by use of the arcsine transformation of the proportions of individual components (SCHILLER and GRUNWALD, 1987; BARADAT and YAZDANI, 1988; SMITH et al., 1988). However, the arcsine transformation is the specific variance-stabilizing transformation applicable to proportional data that result from counts that follow a binomial distribution; it is entirely inappropriate to apply it to resin compositional data.

An appropriate analysis, acknowledging the non-normality of and the constraint acting on resin compositional data, is to calculate logratios, which overcomes these two limitations. The genetic parameters of the logratios can then be estimated and interpreted, perhaps in the light of our increasing knowledge of biosynthetic pathways.

The data set available from the *P. elliottii* progeny trial (*Table 3*) comprised only 3 related families, and is therefore an inadequate basis for the estimation of genetic parameters. Nevertheless, the 1st step in parameter estimation, a univariate analysis of variance of each logratio, could be conducted. In this case, 6 logratios for the components were calculated with the proportion of β-pinene as the divisor, and the analysis revealed significant differences between sites, families, and replicates for some logratios. A multivariate analysis of variance also revealed significant differences between sites, families and replicates. Given the very limited sample size, one would not want to generalise from this result, but it nevertheless demonstrates the applicability of "standard" analytical methods to logratio data.

Site	Clone	Code
A	49	1
A	130	2
A	173	3
B	49	4
B	130	5
B	173	6
clone/site mean		*

5. Conclusions

Most researchers working with resin compositional data have acknowledged, implicity at least, the complications such proportional data pose for analyses and interpretation, but few appear to appreciate quite how fundamental these difficulties are. The major limitation of proportional data, from the point of view of taxonomists, geneticists and biochemists, is the lack of an interpretable covariance structure for such data. Those researchers who have attempted to address the difficulties inherent in proportional data have generally done so through inappropriate transformation of the data, or by marginal modification of standard statistical techniques. Neither addresses the fundamental problem. There are few statistically valid approaches to such data sets, some of which we described in an earlier paper (Birks and Kanowski, 1988); that which appears to offer most promise was developed by Aitchison (1984, 1986), and applied in this study to resin data sets. It is derived from properties of the lognormal distribution, is based on expression of the compositional data as sets of logratios of constituents, and allows the application of standard multivariate procedures to logratio data sets. We look forward to the valid reanalysis and reinterpretation of the many valuable resin data sets already in existence, and to the taxonomic, genetic and biochemical information these studies will provide.

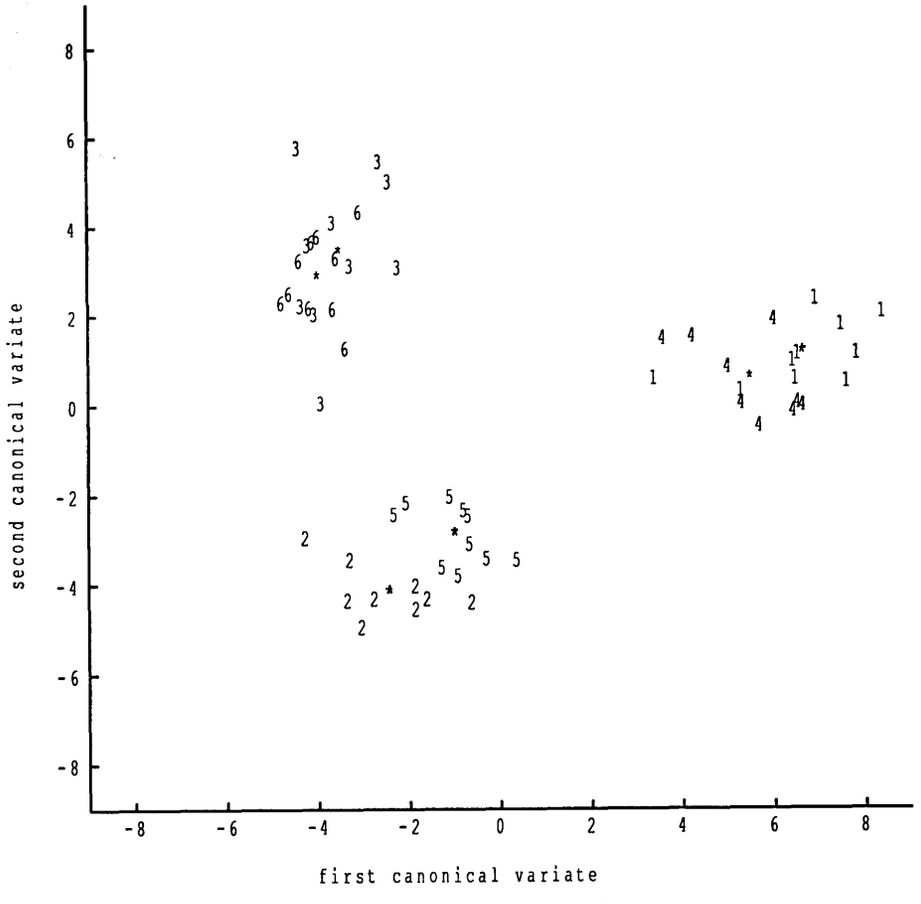

Figure 1. — Canonical variate analysis of 7 logratios of 3 clones at 2 sites.

6. Acknowledgements

The resin samples used in this study were provided by the Zimbabwe Forestry Commission; their collection was kindly organised by Dr. R. D. Barnes. Chromatographic analyses were conducted at the Oxford Forestry Institute by Ms L. A. Lockhart. The study was partially funded by the UK Overseas Development Administration. An earlier and less complete version of the paper was presented at the IUFRO Symposium "Population genetics and gene conservation of forest trees", Bordeaux, France, 24 to 28 August 1992. Drafts of this paper benefitted from review by Drs R. D. Barnes and J. Burley.

7. References

Adams, W. T.: Range-wide patterns of allozyme variation in Douglas-fir (*Pseudotsuga menziesii*). Canadian J. Forest Research **19**: 149—161 (1989). —— Aitchison, J.: The statistical analysis of geochemical compositions. Mathematical Geology **16**: 531—564 (1984). —— Aitchison, J.: The statistical analysis of compositional data. Chapman and Hall, London. 416 p. (1986). —— Andrews, J. A., Gnanadesikan, R. and Warner, J. L.: Methods for assessing multivariate normality. In: P. R. Krishneiah (Ed.). Multivariate analysis. III. Academic Press, New York. 95—116 (1973). —— Baradat, Ph. and Yazdani, R.: Genetic expression for monoterpenes in clones of *Pinus sylvestris* grown on different sites. Scandinavian J. Forest Research **3**: 25—36 (1988). —— Barnes, R. D.: The tree breeding programme in Zimbabwe. Plan of work for 1986/1987. Oxford Forestry Institute. 303 p. (1986). —— Barton, A. F. M., Cotterill, P. P. and Brooker, M. I. H.: Heritability of cineole yield in *Eucalyptus kochii*. Silvae Genetica **40**: 37—38 (1991). —— Bernard-Dagan, C.: Biosynthesis of lower terpenoids: genetic and physiological controls in woody plants. In: J. W. Hanover and D. E. Keathley (eds): Genetic manipulation of woody plants. Plenum Press, New York. 329—351 (1988). —— Birks, J. S. and Kanowski, P. J.: Interpretation of the composition of coniferous resin. Silvae Genetica **37**: 29—39 (1988). —— Brooks, J. E., Borden, J. H. and Pierce, H. D.: Jr.: Foliar and cortial monoterpenes in Sitka Spruce: potential indicators of resistance to the white pine weevil, *Pissodes strobi* Peck (Coleoptera: Curculionidae). Canadian J. Forest Research **17**: 740—745 (1987). —— Brown, A. H. D. and Moran, G. F.: Isozymes and genetic resources of forest trees. In: M. T. Conkle (Ed): Isozymes of North American forest trees and forest insects. USDA, Berkeley, CA, USA. 1—10 (1981). —— Chang, J. and Hanover, J. W.: Geographic variation in the monoterpene composition of black spruce. Canadian J. Forest Research **21**: 1796—1800 (1991). —— Cheliak, W. M. and Rogers, D. L.: Integrating biotechnology into tree improvement programs. Canadian J. Forest Research **21**: 1796—1800 (1991). —— Cool, L. G., Power, A. B. and Zavarin, E.: Variability of foliage terpenes of *Fitzroya cupressoides*. Biochemical Systematics and Ecology **19**: 421—432 (1991). —— Cool, L. G. and Zavarin, E.: Terpene variability of mainland *Pinus radiata*. Biochemical Systematics and Ecology **20**: 133—144 (1992). —— Everitt, B. S.: The analysis of contingency tables. Chapman and Hall, London. 128 p. (1977). —— Fady, B., Arbez, M. and Marpeau, A.: Geographic variability of terpene composition in *Abies cephalonica* Loudon and *Abies* species around the Aegean: hypotheses for their possible phylogeny from the Miocene. Trees **6**: 162—171 (1992). —— Forrest, G. I.: A rangewide comparison of outlying and central lodgepole pine populations based on oleoresin monoterpene analysis. Biochemical Systematics and Ecology **15**: 19—30 (1987). —— Hall, G. D. and Langenheim, J. H.: Geographic variation in leaf monoterpenes of *Sequoia sempervirens*. Biochemical Systematics and Ecology **15**: 31—43 (1987). —— Hanover, J. W.: Genetics of terpenes. I. Gene control of monoterpene levels in *Pinus monticola* Dougl.. Heredity **21**: 73—84 (1966a). —— Hanover, J. W.: Environmental variation in the monoterpenes of *Pinus monticola* Dougl. Phytochemistry **5**: 713—717 (1966b). —— Hanover, J. W.: Applications of terpene analysis in forest genetics. New Forests **6**: 159—178 (1992). —— Harbone, J. B. and Turner, B. L.: Plant chemosystematics. Academic Press, London. 562 p. (1984). —— Hunt, R. S., Meagher, M. D. and Manville, J. F.: Morphological and foliar terpene characters to distinguish between western and eastern white pine. Canadian J. Botany **68**: 2525—2530 (1990). —— Lockhart, L. A.: Chemotaxonomic relationships within the Central American closed-cone pines. Silvae Genetica **39**: 173—184 (1990). —— Michelozzi, M., Squillace, A. E. and White, T. L.: Monoterpene composition and fusiform rust resistance in slash pine. Forest Science **30**: 470—475 (1990). —— Muona, O.: Population genetics in forest tree improvement. Chapter 16 in: A. H. D. Brown, M. T. Clegg, A. L. Kahler and B. S. Weir (Eds): Plant population genetics, breeding, and genetic resources. Sinauer Associates, Sunderland. 282—298 (1989). —— Muzika, R. M., Campbell, C. L., Hanover, J. W. and Smith, A. L.: Comparison of techniques for extracting volatile compounds from conifer needles. J. Chemical Ecology **16**: 2713—2722 (1990). —— Neale, D. B. and Williams, C. G.: Restriction fragment length polymorphism mapping in conifers and applications to forest genetics and tree improvement. Canadian J. Forest Research **21**: 545—554 (1991). —— Raffa, K. F. and Steffeck, R. J.: Computation of response factors for quantitative analyses of monoterpenes by gas-liquid chromatography. J. Chemical Ecology **14**: 1385—1390 (1988). —— Schiller, G. and Grunwald, C.: Cortex resin monoterpene composition in *Pinus brutia* provenances grown in Israel. Biochemical Systematics and Ecology **15**: 389—394 (1987). —— Schönwitz, R., Kloos, M., Merk, L. and Zeigler, H.: Patterns of monoterpenes stored in the needles of *Picea abies* (L.) Karst. from several locations in mountainous regions of southern Germany. Trees **4**: 27—33 (1990). —— Smith, A. L., Campbell, C. L., Walker, D. B., Hanover, J. W. and Miller, R. O.: Geographic variation in the essential oil monoterpenes of *Liriodendron tulipifera* L. Biochemical Systematics and Ecology **16**: 627—630 (1988). —— Squillace, A. E.: Analysis of monoterpenes of conifers by gas-liquid chromatography. In: J. P. Miksche (ed.): Modern methods in forest genetics. Springer-Verlag, Berlin. 120—137 (1976). —— Squillace, A. E. and Perry Jr., J. P.: Classification of *Pinus patula, P. tecunumanii, P. oocarpa, P. caribaea* var *hondurensis*, and related taxonomic entities. USDA Forest Service Research Paper SE-285. 23 p. (1992). —— von Rudloff, E. and Lapp, M. S.: Chemosystematic studies in the genus *Pinus*. IV. The leaf oil terpene composition of ponderosa pine, *Pinus ponderosa*. Canadian J. Botany **70**: 374—378 (1991). —— von Rudloff, E., Lapp, M. S., and Yeh, F. C.: Chemosystematic study of *Thuja plicata:* multivariate analysis of leaf oil terpene compostion. Biochemical Systematics and Ecology **16**: 119—125 (1988). —— White, E. E.: Biosynthetic implications of terpene correlations in *Pinus contorta*. Phytochemistry **22**: 1399—1405 (1983). —— White, E. E. and Nilsson, J. E.: Foliar terpene heritability in *Pinus contorta*. Silvae Genetica **33**: 16—22 (1983). —— Yeh, F. C.: Isozyme analysis for revealing population structure for use in breeding strategies. In: G. L. Gibson, A. R. Griffin and A. C. Matheson (Eds.): Breeding tropical trees-population structure and genetic improvement strategy in clonal and seedling forestry. Oxford Forestry Institute, Oxford, and Winrock International, Arlington. 119—131 (1989). —— Zavarin, E., Cool, L. G. and Snajberk, K.: Geographic monoterpene variability of *Pinus albicaulis*. Biochemical Systematics and Ecology **19**: 147—156 (1991). —— Zavarin, E., Snajberk, K. and Cool, L.: Monoterpenoid differentiation in relation to the morphology of *Pinus edulis*. Biochemical Systemtics and Ecology **17**: 271—282 (1989). —— Zavarin, E., Snajberk, K. and Cool, L.: Monoterpene variability of *Pinus monticola* wood. Biochemical Systematics and Ecology **18**: 117—124 (1990).