Table 3. — Observed pollen frequencies with partial credit for gametes which could have been produced by more than one pollen parent.

Pollens	Female Parent BA1-2	BA3F10-20	S2PT12	S3PT7
5-30	5.833 (5)[1]	14.250 (11)	7.834 (6)	16.333 (12)
7-34	4.000 (4)	6.000 (4)	11.000 (9)	5.500 (5)
11-45	8.000 (8)	7.250 (6)	8.000 (7)	6.500 (6)
12-8	4.833 (3)	9.933 (5)	13.750 (8)	13.167 (10)
14-20	7.000 (5)	16.650 (4)	12.250 (9)	11.000 (7)
15-39	2.000 (2)	5.733 (4)	6.416 (4)	7.500 (4)
S1PT10	2.000 (2)	6.433 (2)	6.750 (3)	7.666 (3)
S6PT2	0.834 (0)	7.584 (5)	4.500 (3)	6.334 (3)
S6PT3	4.500 (4)	6.167 (4)	9.500 (7)	6.000 (5)
Total	39.000 (33)	80.000 (65)	80.000 (56)	80.000 (54)

[1]) Number of positively identified pollen gametes.

Taking this into account, having only one significant chi-square test at the .05 level out of 56 indicates that the observed allozyme frequencies do not deviate significantly from the equal probability of mating model. If differential fertilization were occurring and a particular pollen parent contributed a disproportional number of gametes to the polymix mating, all of its isozymes would be expected to be over-represented resulting in significant chi-square values for several loci.

Frequency of fertilizations by pollen parents

The pollens used in this study were not chosen based on unique allozyme phenotypes, thus, several pollen parents could produce gametes which were indistinguishable from each other (*Table 2*). The method used to deal with this problem, discussed in the Materials and Methods section, could have biased the results of the pollen frequency chi-squares in favor of the null hypothesis, equal probability of mating, by causing a more even distribution of pollen contribution than that which actually occurred.

The observed frequencies of fertilization attributed to each pollen parent are given in *Table 3*. No pollen parent was estimated to have fertilized more than 21% of the seed of any female. This was determined by dividing the number of seed produce by a mating with a particular pollen, (*Table 3*) by the total number of seed sampled for that cross. The chi-square test results on these fertilization frequency data supports the conclusions derived from the allozyme frequency chi-square test. None of the four polymix pollinations produced evidence to reject the equal probability of mating hypothesis. None of the fertilization frequency chi-square tests are significant at the .05 level. The allozyme frequency chi-square test results, which are not affected by the aforementioned source of bias support the results of the fertilization frequency chi-square analysis.

When deviations from equal mating occur the GCA of some female parents may be over-or underestimated, because the mixture of male gametes effecting fertilization may contain an excess of superior or inferior genotypes. The situation would be particularly damaging if the overly successful male parents, differed from cross to cross.

The chi-square analyses of this study, indicate that no significant deviation from the equal probability of mating hypothesis occurred in the polymix matings which lends credence to the GCAs and other statistics produced from polymix progeny test data.

Literature Cited

Barnett, J. P.: Sterilizing southern pine seeds with hydrogen peroxide. Tree Planters Notes **27** (3): 1—17 (1976). — Conkle, M. T., Hodgskiss, P. D., Nunnally, L. B. and Hunter, S. C.: Starch gel electrophoresis of conifer seeds: a laboratory manual. Pacific Southwest Forest and Range Exp. Sta. USDA Gen. Tech. Rep. PSW-64. Berkeley, Calif. pp. 11-17 (1982). — Jech, K. S. and Wheeler, N. C.: Laboratory manual for horizontal starch gel electrophoresis. Weyerhaeuser Tech. Rep. 050-3210/7. pp. 12-61. Centralia, WA (1984). — Hartl, D. L.: Principles of Population Genetics. 1st ed. Sinauer Associates Inc. Publishers Sunderland, Mass. pp. 99—100 (1980). — Mitton, J. B.: Conifers. *In:* Isozymes in Plant Genetics and Breeding, Part B. eds. S. D. Tanksley and T. J. Orton. Elsevier Science Publishers, Amsterdam. pp. 443—472 (1983). — Moran, G. F. and Griffin, A. R.: Non-random contribution of pollen in polymix crosses of *Pinus radiata* D. Don. Silvae Genetica **34** (4—5): 117—121 (1986). — van Buijtenen, J. P.: Genetic improvement of forest trees through selection and breeding. *In:* SAF Forestry Handbook. Ed. C. Wenger. Roland Press Co., New York. pp. 457—488 (1984). — Wiselogel, A. E.: Testing the occurrence of random mating in polymix pollinations of loblolly pine (*Pinus taeda* L.) Ph. D. Dissertation Texas A & M University, College Station, TX. pp. 55—61 (1985). — Yeh, F. C. and Layton, C.: The organization of genetic variability in central and marginal populations of lodgepole pine. Can. J. Genet. Cytol. **24**: 487—503 (1979).

Analysis of Diallel Matings with Missing Values

By C. A. Dean[1]) and R. L. Correll[2])

(Received 29th June 1987)

Summary

A computer program (called Diallel Analysis by GENSTAT, DAG) is presented to analyse data from half-diallel matings without selfs. The program is flexible, coping with mixed models, varying number of traits and missing values (in the form of both missing trees and missing crosses). The program, based on analysis of variance and regression techniques, is easy to use and provides opportunities to pool sums of squares in the analysis of disconnected half-diallel matings. The output is directed towards generating variance components of particular interest to tree breeding and population genetics.

Key words: Computer program, half-diallel matings, missing values, variance components.

Zusammenfassung

Ein Computerprogramm (bezeichnet als Diallel Analyse mit Hilfe von GENSTAT, DAG) zur Auswertung von Daten aus halb-diallelen Kreuzungen ohne Selbstungen wird vorgestellt. Das Programm ist flexibel und setzt sich mit gemischten Modellen auseinander sowie mit variierenden Zahlen von Merkmalen und fehlenden Daten sowohl durch fehlende Bäume als auch durch fehlende Kreuzungen. Das Programm basiert auf der Varianz- und Regressionsanalyse, ist leicht anwendbar und bietet Möglichkeiten, die

[1]) CSIRO, Division of Forest Research, 306 Carmody Road, St. Lucia, Brisbane, 4067, Australia.

[2]) CSIRO, Division of Mathematics and Statistics, Private Bag No. 2, Glen Osmond, Adelaide, 5064, Australia.

Quadratsummen in der Analyse von nicht verbundenen, halb-diallelen Kreuzungen zu verarbeiten. Damit können Varianzkomponenten von besonderem Interesse für die Pflanzenzüchtung und Populationsgenetik erarbeitet werden.

Introduction

Diallel matings are widely used in plant breeding, and sometimes in animal breeding, to generate estimates of additive and non-additive genetic variance components. One of the simplest forms of diallel mating is the half-diallel which does not include selfed individuals or reciprocal crosses (Method 4 of GRIFFING 1956). In tree breeding the use of a disconnected series of such half-diallel matings has been widely recommended for regenerating the breeding population after each generation of selection (SQUILLACE, 1973; GRIFFIN, 1976; BURDON *et al.*, 1977; MATHESON, 1978; TALBERT, 1979). The half-diallels are disconnected in the sense that a group of usually five or six parents are mated in one half-diallel, a different five or six parents are mated in another half-diallel, and so on. The practical advantage of disconnected half-diallels compared with one large diallel is that the former involves far fewer crosses per parent.

The theory of analysis of diallels received considerable attention in the 1950's (notably HAYMAN, 1954; GRIFFING, 1956; KEMPTHORNE, 1956) and more recent developments have been reviewed by MAYO (1980). DE LACY and BASFORD (1988) show the relationships between the various methods of analysis of diallels and describe the genetic interpretation of the statistical partitioning of sources of variation in the diallel model.

In practise, diallel matings are often unbalanced due to missing crosses or because of unequal representation of crosses across blocks (missing plots). Accomodating such imbalance complicates the analysis. Analysing disconnected half-diallels imposes an additional requirement, that of combining sums of squares and products from each of the separate diallels to provide pooled estimates of genetic variance and covariance components.

There does not appear to be a computer program cited either in other crop or animal breeding literature which is capable of analysing diallels with missing crosses or missing plots. In tree breeding the program of SCHAFFER and USANIS (1969) (called DIALL) is available and can accommodate such missing values. The method of analysis employed by the DIALL program does not, however, allow analysis of fixed models or mixed models where say blocks or sites are considered fixed and parents (crosses) considered random. The mixed model is common in analysis of diallel experiments across breeding regions where sites may be considered fixed. Another practical problem with DIALL is that the program does not provide easy access to the sums of squares and products for pooling in the analysis of disconnected mating designs.

The present article outlines a computer program (called DAG: Diallel Analysis by GENSTAT) which is written in GENSTAT (ALVEY *et al.*, 1977), a widely-used statistical programming language. DAG is a flexible program which can be modified to fit any diallel design and can accommodate large numbers of missing values. The program is based on the combining ability analysis outlined by GRIFFING (1956).

Method of Analysis

(1) Models Fitted

The DAG program employs both analysis of variance and regression techniques to complete the analysis of diallel matings. The standard assumptions regarding additivity and normal distribution are made. We have also assumed that competition between trees is not influencing individual performance. In situations where single tree plots have been used it will be necessary to make a further assumption of independence of individual trees. In this special case additional correction terms may need to be added to the model (for example PAPADAKIS 1937 correction for local trends or DRAPER-GUTTMAN 1980 correction to remove competition effects).

An important feature of DAG is that effects in the model can be classified as either fixed or random. In the version of DAG presented here (*Figure 1*) the effects of blocks are assumed fixed while those of crosses are assumed random. Appendix 1 contains a description of the program code given in *Figure 1*. More detailed descriptions and worked examples of this and other versions of DAG are available on request from the senior author. The model for the analysis of individual tree data is — (1)

$$Y_{ijkl} = \mu + B_i + C_{jk} + BC_{ijk} + e_{ijkl} \qquad (1)$$

where μ is the mean, B_i the ith block, C_{jk} the cross between the jth female parent and the kth male parent, BC_{ijk} the interaction between the ith block and the cross between the jth female parent and the kth male parent, and e_{ijkl} the within-plot error. There are b blocks and p parents. The effect of crosses (C_{jk}) can be further partitioned according to the model — (2)

$$C_{jk} = G_j + G_k + S_{jk} \qquad (2)$$

where G_j and G_k are the general combining effects of the jth and kth parents respectively and S_{jk} the specific combining effect of these two parents.

In DAG, analysis is achieved by reducing these models into four sub-models which are then fitted to the least-squares adjusted data from the diallel experiment. The four submodels and their reduction notation are given in *Table 1* and are discussed later.

(2) Technical Details of the Analyses

In the diallel analysis presented here four separate analyses are required to partition the appropriate sums of squares and products and they are described below. Each analysis, based on a reduction of the models given in equations 1 and 2 (see *Table 1*), produces a residual which can be stored and used in subsequent calculations.

Analysis One: The first analysis is used to create residuals following the removal of block effects (R1 (μ, B); *Table 1*). The analysis consists of a least-squares analysis of covariance where GENSTAT dummy variates representing block effects are used as covariates. In DAG each individual block is identified as a separate effect. In the case where an experiment is planted on more than one site each block within each site would similarly be identified as a separate effect. The contrasts (or differences) between these blocks are calculated in such a way that they have zero mean and are stored as the dummy variates representing block effects. In DAG regression for blocks actually uses these contrasts, via the dummy variates, to eliminate as far as possible the influence of block effects on the estimation of other effects.

Analysis Two: The second analysis is used to create residuals following the removal of the block and parental effects (R2 (μ, B, G); *Table 1*). A linear regression is fitted with the dependent variables being the dummy variates for

```
'refe / nunn = 200, nid = 500' plot2

" ——— step 1 ———

basic data declarations and description"
'unit' $336
'scalar' bc
: nparent = 5
: nblock = 6
: bsize = 56
: ncross
: nattributes
: nplot
'fact' block $nblock = bsize!(1...nblock)
'set' attributes = ht, dia
'pointer' p = attributes
'calc' nattributes = nval (p)
: ncross = nparent * (nparent - 1) / 2
"form structures for holding residuals, sums of squares and
products, etc"
'dssp'r(1...4) $ attributes
: ssp_xs, ssp_gca, ssp_sca, ssp_plot, ssp_tree $ attribute
: msp_xs, msp_gca, msp_sca, msp_plot, msp_tree $ attribute
: vcv_xs, vcv_gca, vcv_sca, vcv_plot $ attribute
: cor_xs, cor_gca, cor_sca, cor_plot, cor_tree $ attribute
: var_mgca, var_msca $ attribute
'symmat' mspg_2, msps_2, mspp_2, var_g_var; var_s_var,
var_p_var $p
: mat_mspg, mat_msps, mat_mspp $p
'macro' find_max $
'scalar' maxlevel
'calc ' maxlevel = max (float ( fhartley ))
'endmacro' " find_max "
'macro' mhartley $
" The method requires the sum of the mean squares for each
of the dummy variates associated with the factor in question
(fhartley). The method used here finds the treatment sum of
squares by differencing the residual sum of squares with and
without the factor fhartley declared as a treatment. Note
that the relevant block and covariate declarations must be
made prior to the macro being called. The macro requires a
factor fhartley to be supplied and the effective replication is
returned in the scalar khartley."
'calc' khartley = 0

      'for' dummyw = 1...maxlevel 'calc' dummyy =
      float (fhartley) .eq. dummyw
      'treat'
      'anov / acon = acon1, pr = 0, prx = 0, pryu
      = 0'dummyy ; res = dummyz 'calc'khartley =
      khartley + sum(dummyz*dummyz)
      'treatment' cross_f 'anov / acon = acon2,
      pr = 0, prx = 0, pryu = 0'dummyy ;
      res = dummyz 'calc'khartley = khartley -
      sum(dummyz*dummyz)
      'repeat'
      'calc' khartley = khartley / ( ncross - 1 )
      'print' khartley $8.3

'endmacro' " ***** end of mhartley macro "
'macro' hartley $
" find the number of levels in the treatment factor "
'use / r' find_max $
" perform main hartley procedure "
'use / r' mhartley $
'endmacro' " hartley "
'run'

" ——— step 2 ——— "

'inpu' 2
'read / nun = q' xr, mal, fem, attributes
'inpu' 1
'run'

" ——— step 3 ——— "

" formation of factors and variates required in subsequent
analyses"
" form dummy block variates for covariate analyses. these
are chosen to have zero mean and to provide contrasts be-
tween the blocks."
'calc' dummy(2...nblock) = (float(block) .eq. 1) /
sum(float(block) .eq. 1) - (float(block) .eq. 2...nblock) /
sum(float(block) .eq. 2...nblock)
'covariates' dummy(2...nblock)
"form parent variates"
'calc' parent (1...nparent) = ((mal .eq. 1...nparent) + (fem
.eq. 1...nparent)) / 2
" form dummy variates for crosses "
'group' cross = intpt (xr)
'calc' dummy_x = (mal * 10 + fem) * (mal + 10 * fem)
'group' cross_f = intpt (dummy_x)

" ——— step 4 ——— "

"set missing value indicators"
'calc' missing = (ht .gt. 2) / (ht .gt. 2)
: attributes = attributes*missing
'calc' dummyxr(1...ncross) = (float(cross_f) .eq. 1...ncross )
* missing
" form a dummy factor to remove all cross_f, block and block
x cross_f effects "
'calc' dummygb = (float(block) - 1)*bsize + float(cross_f)
'group' fdummygb = intpt(dummygb)

" ——— step 5 ——— "

" restrict out controls"
'vari'limit = 60.5
'name'cnt = cross, control
'fact'wanted $cnt
'group'wanted = limits(xr ; limit)
'rest'attributes $wanted = 1
'run'

" ——— step 6 ——— "

"form a spare copy of data"
'set' sparecopy = sparecopy (1...nattributes)
'equa' sparecopy = attributes
```

blocks described above and another set of dummy variates set up to represent parental (general combining) effects. The dummy variates for parents are set up to account for the various contributions of parents to any given cross. A matrix of genetic relationship is created using zeros and halves to represent the gametic contribution of a parent to

```
'rest'attributes, sparecopy $wanted = 1

" ——— step 7 ——— "

" This section involves four analyses which become progressively more precise. Their main aim is to provide residuals from which subsequent calculations are made"
'anov / pryu = 0, prx = 0'sparecopy ; res = attributes
" note restricted printing to avoid a printout of the analyses of the covariates"
'calc' attribute = attribute * missing
'ssp' r(1)
" r(1) now holds sums of squares and products from enviromental model"
'run'
'regr' parent(1...nparent), dummy(2...nblock), sparecopy
'y' sparecopy
'fit / int = n' parent (1...nparent) , dummy(2...nblock) ; res = attributes
" because of the way dummy() has been defined, forcing the regression through the origin results in the first nparent coefficients holding estimates of family means "
'calc' attributes = attribute*missing
'ssp' r(2)
"compute residual sums of squares and products"
'treat' cross_f
'anov' sparecopy ; res = attribute
'calc' attributes = attribute*missing
'ssp' r(3)
" analysis 4 to find within plot error "
'block'
'cova'
'treat' fdummygb
'anov'sparecopy ; res = attribute
'calc' attribute = attribute * missing
'ssp' r(4)
'run'
'print' r(1...4)

" ——— step 8 ——— "

" in this section the coefficients for determining expected mean squares are calculated or estimated "
" form geometric mean of trees per cross per block "
'table' tab $fdummygb
'tabulate'missing ; tab
'scal' harmonic
'calc' harmonic = 1 / mean(1 / tab)
" note that harmonic could also be calculated using the hartley macro but in general it would be slower. However if there are covariates present the hartley method should be used. Calculation of the harmonic mean assumes that all plots considered have at least one entry. The hartley macro is robust against this. To incorporate into the code replace the harmonic 'calc' with 'for' hartleyk = harmonic ; hartleyf = fdummygb 'use' hartley 'repeat'"
'print' harmonic
'run' "(to allow compilation of design descriptors)"

" ——— step 9 ——— "

" in this section genetic parameters are calculated"
" ****** A ***** sums of squares and products "
'calc' ssp_xs = r(1) - r(3)
: ssp_gca = r(1) - r(2)
: ssp_sca = ssp_xs - ssp_gca
: ssp_plot = r(3) - r(4)
'print' ssp_xs, ssp_gca, ssp_sca, ssp_plot
" ***** B ***** calculate mean squares and products"
'calc' msp_xs = ssp_xs / (nparent * (nparent - 1) / 2 - 1)
: msp_gca = ssp_gca / (nparent - 1 )
: msp_sca = ssp_sca / (nparent * (nparent - 3) / 2)
: msp_plot = ssp_plot / (nval(tab) - nmv(tab / tab))
: ssp_tree = r(4)
: msp_tree = ssp_tree / (nval(missing) - nmv(missing) - nval(tab) + nmv(tab / tab))
'capt' " mean squares "
'print' msp_xs, msp_gca, msp_sca, msp_plot, msp_tree
" ***** C ***** form variances"
"use the hartley method to find the bc of griffing "
'for' khartley = bc ; fhartley = cross_f 'use / r' hartley $ 'repeat'
'calc' vcv_xs = (msp_xs - msp_plot) / bc
: vcv_gca = ( msp_gca - msp_sca) / bc / (nparent - 2 )
: vcv_sca = ( msp_sca - msp_plot) / bc
: vcv_plot = ( msp_plot - msp_tree ) / harmonic
'print' vcv_xs, vcv_gca, vcv_sca, vcv_plot
'run'
" ***** D ***** form correlations"
'calc' cor_xs, cor_gca, cor_sca, cor_plot, cor_tree = cormat(vcv_xs, vcv_gca, vcv_sca, vcv_plot, msp_tree)
'prin' cor_xs, cor_gca, cor_sca, cor_plot, cor_tree $8.3
" ***** E ***** approximate variances of mean squares "
" transfer dssp structures to matrices to allow matrix multiplication "
'equate' mat_mspg = msp_gca
: mat_msps = msp_sca
: mat_mspp = msp_plot
" form matrix products "
'calc' mspg_2 = pdtt(mat_mspg ; mat_mspg) / (nparent + 1)
: msps_2 = pdtt(mat_msps ; mat_msps) / (2 + nparent*(nparent - 3) / 2)
: mspp_2 = pdtt(mat_mspp ; mat_mspp) / (nplot - ncross + 2)
: var_g_var = 2 / (nparent - 2) / (nparent - 2) * (mspg_2 + msps_2)
: var_s_var = 2 * (msps_2 + mspp_2)
: var_p_var = 2 * mspp_2
'caption' " variances of genetic variance components (after Becker) "
'print' var_g_var, var_s_var, var_p_var
'run'
'close'
'stop'
```

Figure 1. — The GENSTAT code for DAG.

a particular cross. Thus, the rows of the matrix represent individual trees and the columns represent the genetic contribution of each parent to that tree. The total of each row is therefore one and any two full-sib individuals will be represented by identical rows in the matrix. The columns of this matrix can then be used as covariates for estimating

Table 1. — Models fitted and notation of the reductions made in the analyses of variance and regression carried out in step 7 of the DAG program.

Analysis	Model	Reduction denoted
1	$Y_{ijkl} = \mu + B_i + e_{ijkl}$	R1 (μ, B)
2	$Y_{ijkl} = \mu + B_i + G_j + G_k + e_{ijkl}$	R2 (μ, B, G)
3	$Y_{ijkl} = \mu + B_i + C_{jk} + e_{ijkl}$	R3 (μ, B, C)
4	$Y_{ijkl} = \mu + B_i + C_{jk} + BC_{ijk} + e_{ijkl}$	R4 (μ, B, C, BC)

between parent effects (general combining effects). Note that this dummy variate for parents cannot be used to represent selfed individuals.

The model for the regression analysis is defined in such a way that the regression line is forced through the origin. This definition means that the first P (where P is the number of parents) regression coefficients will be estimates of parental means. These parental mean values are thus accessible in the program.

Analysis Three: The third analysis is used to create residuals following the removal of block and cross effects (R3 (μ, B, C); *Table 1*). In this analysis the dummy variates for block effects are again used as covariates and a dummy factor for crosses is set up as a classifying factor in the analysis of covariance. The dummy factor for crosses is set up to identify each level of the crosses term so that even when there are missing crosses the model can still be represented in a simple form. Removing the effect of crosses in this way leads to the covariate means being estimates of the means of the crosses. Thus, the mean value of each cross (or family mean value) is also accessible in the program.

Analysis Four: The fourth analysis is used to create residuals following the removal of block, cross and block × cross interaction effects (R4 (μ, B, C, BC); *Table 1*). A dummy factor (similar to that described for crosses in Analysis Three) is used to index each of the plots measured in the experiment. Thus, even when there are missing plots the model for this analysis can be represented simply and analysis of variance carried out using this dummy factor for plot effects as the classifying factor.

(3) *Coping with Missing Values*

According to the general theory of least squares a model for a complete non-orthogonal analysis can be written for a data set with any number of missing values and solved by matrix inversion. This approach is the method used by Schaffer and Usanis (1969) in DIALL. Where there are large numbers of missing crosses in a diallel experiment the computations for this method become cumbersome. For ease of programming and flexibility of statistical analysis a more efficient method of coping with missing values in data from diallels may be to simplify the model for the analysis as much as possible within the constraints of the experimental design. This method is used in DAG and involves exploiting available orthogonality in the model. In effect, this means partitioning the X'X matrix into block and genetic components. After the removal of the block component as a covariate, the genetic effects are orthogonal even if the replication is uneven. The dummy variates and factors described in the preceding section are used to represent the effects in the model. These "dummies" are set up to minimise the computations required in the analyses. The size of the matrices to be inverted by DAG are equal to b in Analyses One and Three and b + p in Analysis Two (where b is the number of blocks and p the number of parents). In the regression approach, used in the complete non-orthogonal analysis, the order of the matrix for Analysis Three alone would be b + ½ p (p-1) which, as p increases, would rapidly become large.

The GENSTAT analysis of variance and regression commands routinely estimate missing values that make no contribution to the residual sums of squares and products (see Alvey *et al.*, 1977). These estimated values may lead to some inflation of the treatment sums of squares and products produced by these analyses. To avoid error from such inflation the DAG program uses only the unbiased residual sums of squares and products from the various analyses for further calculations.

(4) *Sums of Squares and Products Structures*

Sums of squares and products matrices required to estimate the desired genetic variance and covariance components are obtained by differencing sums of squares and products and residuals formed in the analyses outlined above. For instance, sums of squares and products for crosses (SSP_c) are obtained by subtracting the residual sums of squares for Analysis Three (that is, R3 (μ, B, C); *Table 2*) from the residual sums of squares for Analysis One (that is, R1 (μ, B); *Table 2*). The exact subtraction required to obtain each of the effects of interest (gca, sca, block × cross interaction and within-plot error) is given in *Table 2*. As these residuals are not biased by the estimates of missing values (discussed above) they provide an accurate partitioning of the sums of squares and products.

The residuals formed in the four analyses above and the sums of squares and products matrices formed by the subtractions given in *Table 2* are maintained in separate storage allocations throughout the DAG program. Thus, these structures are available for, say, accrual across a disconnected series of diallels and also for use in further computations within the diallel being considered.

Table 2. — Differences between sums of squares and products and residuals from four analyses relevant to the estimation of genetic variance and covariance components in half-diallel matings.

Effect	Sums of squares and products	Difference
Total	SSP_{total}	
Crosses	SSP_c	R1(μ,B) − R3(μ,B,C)
gca	SSP_{gca}	R1(μ,B) − R2(μ,B,G)
sca	SSP_{sca}	$SSP_c - SSP_{gca}$
Block x cross	SSP_{BC}	R3(μ,B,C) − R4(μ,B,C,BC)
Within-plot error	SSP_w	R4(μ,B,C,BC)

(5) *Variance Components*

Equating mean squares with their expectations (given in *Table 3*) provides estimates of the genetic variance due to crosses (σ^2_c), general combining ability (σ^2_{gca}), specific combining ability (σ^2_{sca}), block × cross interaction (σ^2_{bc}) and within-plot error (σ^2_w). The two k coefficients required to estimate these genetic variance and covariance components (see *Table 3*) are estimated as the harmonic mean of the number of full-sib individuals per plot in the case of k_1 and using the synthesis procedure of Hartley (1967) in the case of k_2. The Hartley procedure

Table 3. — Expectations of mean squares relevant to estimating genetic parameters from analyses of covariance and regression in the DAG program. There are p parents and N trees overall in the experiment: The variance component σ^2_c is due to crosses, σ^2_{gca} general combining ability, σ^2_{sca} specific combining ability, σ^2_{bc} block × cross interaction and σ^2_w within plot error.

Source of variation	d.f.	Mean squares	Expectations of mean squares
Crosses	(p(p-1)/2)-1	MSP_c	$\sigma^2_w + k_1\sigma^2_{bc} + k_2\sigma^2_c$
gca	p-1	MSP_{gca}	$\sigma^2_w + k_1\sigma^2_{bc} + k_2\sigma^2_{sca} + k_2(p-2)\sigma^2_{gca}$
sca	(p(p-3))/2	MSP_{sca}	$\sigma^2_w + k_1\sigma^2_{bc} + k_2\sigma^2_{sca}$
Block x cross	(b-1)((p(p-1)/2)-1)	MSP_{BC}	$\sigma^2_w + k_1\sigma^2_{bc}$
Within plot	N-b(p(p-1)/2)	MSP_w	σ^2_w

calculates the sum of mean squares for each of the dummy variates involving crosses. The difference between the residual sum of squares from the dummy variates with and without the crosses effect gives the treatment sum of squares which is a direct estimate of k_2 (the effective number of individuals per cross). Standard errors of the genetic variance and covariance components are obtained following Becker (1985).

Conclusions

Analysis of diallel matings is not simple even in the case of balanced data and complications increase when there are missing plots or missing crosses. A computer program called DAG has been written to analyse diallel experiments which have missing values. The program is written in GENSTAT, a widely used general statistics program, which allows data manipulation as well as flexibility of statistical analyses.

The analysis of diallels is achieved in DAG by reducing the overall model for the analysis into sub-models relating to the desired genetic variance and covariance components. The nature of the reduction analyses means that the program can easily be modified to cope with a range of diallel models. The data manipulation feature of GENSTAT provides DAG with the capability of accessing information calculated by the program, such as means of parents, means of crosses and sums of squares and products.

Acknowledgements

We gratefully acknowledge the valuable comments of Drs. Graham Constantine, Paul Cotterill, Neville Jackson and Peter Kanowski in review. This paper represents a portion of work by the senior author towards a Ph. D. program in the Botany Department of the University of Queensland.

Literature Cited

Alvey, N. G., Banfield, C. F., Baxter, R. I., Gower, J. C., Krzanowski, W. J., Lane, P. W., Leech, P. K., Nelder, J. A., Payne, R. W., Phelps, K. M., Rogers, C. E., Ross, G. J. S., Simpson, H. R., Todd, A. D., Wedderburn, R. W. M. and Wilkinson, G. N.: GENSTAT - A general statistics program. Rothamsted Experimental Station, Harpenden, Hertfordshire (1977). — Becker, W. A.: Manual of quantitative genetics. Fourth edition. Academic Enterprises, Pullman, Washington, U.S.A. (1985). — Burdon, R. D., Shelbourne, C. J. A. and Wilcox, M. D.: Advanced selection strategies. Proc. 3rd Wrld. Consult. For. Tree Breed., Canberra. pp. 1137—1147 (1977). — De Lacy, I. H. and Basford, K. E.: Analysis of diallels. Biometrics **44**: in press (1988). — Draper, N. R. and Guttman, I.: Incorporating overlap effects from neighbouring units into response surface models. Appl. Statist. **29**: 128—134 (1980). — Griffing, B.: Concept of general and specific combining ability in relation to diallel crossing systems. Aust. J. Biol. Sci. **9**: 463—493 (1956). — Griffin, A. R.: Mating designs for *Pinus radiata* breeding. Proc., Fifth Meet. Res. Work. Group No. 1, Aust. For. Council, Canberra and Tumut. pp 16—20 (1976). — Hartley, H. O.: Expectations, variances and covariances of ANOVA mean squares by synthesis. Biometrics **23**: 105—114 (1967). — Hayman, B. I.: The theory and analysis of diallel crosses. Genet. **39**: 789—809 (1954). — Kempthorne, O.: The theory of the diallel cross. Genet. **41**: 451—459 (1956). — Matheson, A. C.: Review of the breeding strategy adopted at the Research Working Group Meeting 1976. Proc., Sixth Meet. Res. Work. Group No. 1, Aust, For. Council, Coffs Harbour, Australia. pp. 17—21 (1978). — Mayo, O.: The theory of plant breeding. Oxford Uni. Press, Oxford (1980). — Papadakis, J. S.: Methode statistique pour des experimences sur champ. Bull. Inst. Amel. Plants a Salonique. **23**: 13—29 (1937). — Schaffer, H. E. and Usanis, R. A.: General least squares analysis of diallel experiments. A computer program — DIALL. Genetics Dept. Research Report No. 1. N. C. State University. Raleigh N.C. (1969). — Squillace, A. E.: Comparison of some alternative second-generation breeding programmes for slash pine. Proc., 12th South. For. Tree Improv. Conf., Baton Rouge, USA. pp. 2—13 (1973). — Talbert, J. T.: An advanced generation breeding plan for the N.C. State University — Industry pine tree improvement cooperative. Silvae Genet. **28**: 72—75 (1979).

Appendix One: Program Description

The following description gives a brief outline of each step of the DAG program detailed in *Figure 1.* Readers not familiar with the GENSTAT program should consult the manuals produced by Alvey *et al.* (1977) which give exact specifications of the GENSTAT commands. A test example including input data and the analysis produced by DAG is available on request from the senior author. A more detailed description of the mechanics of the program is also available.

Step 1. Initialisation: In step 1 the program compiles design descriptors and prepares data structures for subsequent analyses. GENSTAT macros are defined which will calculate effective replication for use in analyses.

Step 2. Data Input: In step 2 data are read in free field format and placed directly into the storage space allocated in step 1. Provision is made for data to be supplied from a separate file.

Step 3. Form Dummy Variates and Factors: The purpose of step 3 is to set up a series of identifiers (called dummies in GENSTAT) which are used to represent various model terms in subsequent analyses. Dummy variates are formed to represent (a) the contrast between block effects and (b) the various contributions of parents to each cross. These variates are used to remove the effects of blocks or parents respectively during any analysis. Dummy factors are set up to identify each individual level of each of the block, cross and block × cross interaction effects. These dummy factors account for any missing observations and allow these effects to be represented simply during the analyses.

Step 4. Missing Trees: The GENSTAT analysis of variance and regression commands routinely estimate missing values that have no contribution to the residual sum of squares. In the DAG program these substitute values are flagged so that they can be removed from any subsequent analysis, if desired.

Step 5. Selection of Units: This step of the program is included as an example of data manipulation and provides an opportunity for restricting the data set to be used in the analyses. For example, "unimproved control" seed lots are often included in genetics experiments but would be excluded from analyses for estimating genetic parameters.

Step 6. Working Copy of the Data: The original data has been read and is stored by the program. In step 6 provision is made for a "working copy" of this data to be produced for each analysis by copying the original data. This provision allows the original data to be modified separately

for each analysis (if required) and also to be resubmitted for any number of different analyses.

Step 7. Analyses of Data: In step 7 the analyses of the data are carried out. In the version of DAG presented in Figure 1 the four analyses outlined in this article are carried out. Each analysis is based on a reduction of the overall model and produces a residual which is used in subsequent calculations to obtain information about the desired model terms.

Step 8. Coefficients for Expected Mean Squares: Relevant k coefficients can be estimated in two ways by DAG-either by using the harmonic mean (used here to estimate the number of individuals per plot) or using the synthesis procedure of HARTLEY (1967) (used here to estimate the number of individuals per cross).

Step 9. Genetic Parameters: In this final step the total sums of squares and products for these effects. Mean sums of squares and products and residuals formed following the analyses in step 7 are differenced to obtain the squares are calculated and equated with their expectations to form the variance and covariance components. Standard errors of these components are also computed.

Genetics of Growth Rate in Quercus rubra: Provenance and Family Effects by the Early Third Decade in the North Central U.S.A.[1])

By H. B. KRIEBEL, C. MERRITT and TH. STADT[2])

(Received 21st July 1987)

Abstract

Northern red oak trees from seed origins at latitudes 43° to 46° N from the Mississippi River to western Maine were consistently high in growth rate in relation to trees of other origins in 3 provenance tests in the North Central region of the U.S.A. Seed sources from farther north, south or west were less promising, although the trees from some of these sources were above average in relation to test means. About 16 percent of measured variance in height growth was attributable to provenance and 7 percent to family. Age-age correlations in the Ohio test showed that selection for growth rate was not possible at age 8, probably partly because of early mortality and rabbit injury. We could have selected the best provenances from the standpoint of growth rate at age 14, but not the best individual trees. Early selection might be more feasible now, considering improvements in technology for red oak plantation establishment.

Key words: Quercus rubra, northern red oak, growth rate, provenance, family, juvenile selection.

Zusammenfassung

Aus Samen gezogene Bäume von *Quercus rubra* (amerik. Roteiche) aus dem Ursprungsgebiet von 43° bis 46° nördl. Breite vom Mississipi bis zum westl. Maine, zeigten übereinstimmend hohe Wachstumsraten im Vergleich zu anderen Herkünften bei drei Provenienzversuchen im mittleren Norden der USA. Herkünfte, die aus entfernteren Regionen in nördlicher, südlicher oder westlicher Richtung stammten, waren weniger versprechend, obgleich einige davon überdurchschnittlich erschienen. Die Varianz im Höhenwachstum wurde zu ca. 16 Prozent von Herkünften und zu 7 Prozent von Familien verursacht. Die Alters-Alters-Korrelationen im Ohio-Versuch waren für die 8jährigen Bäume nicht möglich, teilweise wegen Mortalität und Verbiß durch Kaninchen. Wir hätten die besten Herkünfte im Höhenwachstum im Alter von 14 Jahren selektieren können, jedoch nicht die besten Bäume. Frühselektion wäre jetzt besser durchführbar, wenn die technischen Verbesserungen bei der Begründung von Roteichenplantagen berücksichtigt werden.

Introduction

The economic importance of northern red oak (*Quercus rubra* L.) in the eastern United States gave the species a high priority for genetic research when a committee for cooperative regional research was organized by the state experiment stations of the North Central region in 1959. Seven provenance tests were established in six states in the early 1960s. The objective was to evaluate range-wide genetic variation to provide a basis for selecting seed sources for high vigor and other desirable traits. Three other previously-established smaller-scale provenance tests of *Quercus rubra* evaluated at an early age demonstrated significant provenance-related variation in vigor. The number of samples was small in each case and the nature of the variation patterns could not be defined (SCHREINER and SANTAMOUR, 1961; KRAHL-URBAN, 1966; GALL and TAFT, 1973). A nursery evaluation of our own material at age 1 showed that there were both parental and provenance effects on early vigor (KRIEBEL, 1965). Analysis of all seven of our experiments at ages 12 to 14 indicated that there was significant seed source variation in growth rate, even within limited geographic areas. There was also some indication of a regional pattern of moderately high vigor among samples from southern Illinois, southern Indiana, and western Ohio (KRIEBEL *et al.*, 1976). DENEKE (1975), SCHLARBAUM and BAGLEY (1981) and SCHLARBAUM *et al.*, (1982) also found source-related differences in height growth, survival and other traits in their early reports on our experiments in Kansas and Nebraska.

In this paper, we report early third decade results of analysis of height growth differences attributable to provenance and family, from observations made in the Ohio, Indiana and Michigan plantations. We describe an emerging pattern of geographic variation, appraise the effect of seed tree relative to that of provenance, and estimate the pos-

[1]) North Central Regional Project NC-99, Cooperative State Research Service, U.S. Deparment of Agriculture. Salaries and research support provided by State and Federal funds appropriated to the Ohio Agricultural Research and Development Center of the Ohio State University, to Purdue University and to Michigan State University. Journal article No. 171-86 of the OARDC.

[2]) Professor, The Ohio State University, Wooster, OH 44691, USA, Professor Emeritus, Purdue University, West Lafayette, IN 47907, USA, and Senior Resident Forester, Kellogg Forest, Michigan State University, Augusta, MI 49102, USA.