Analyses of Half-Diallel Mating Designs with Missing Crosses: Theory and SAS Program for Testing and Estimating GCA and SCA Variance Components

By H. Wu and A. C. Matheson[1])

(Received 16th February 2001)

Abstract

The half-diallel mating design, particularly a series of disconnected half-diallel mating design has been widely adopted as a mating design in many commercially important tree species for estimating the variances of general (GCA) and specific (SCA) combining abilities, heritability, and genetic correlation. Standard commercial statistical packages do not allow direct specification of the half-diallel model and therefore are not able to analyse the diallel mating design even for a balanced diallel mating structure (no missing crosses). Published special computer programs for diallel analyses do not provide an adequate solution for GCA and SCA variance components with hypothesis testing for half-diallel mating designs with missing crosses. This paper presents the theory of the method of fitting constants for analysing half-diallel mating designs with or without missing crosses. A SAS computer program (DIARANDM.SAS) for testing hypotheses and estimation of GCA and SCA variance components is developed. The program accommodates multiple variables, multiple environments, environment by GCA and environment by SCA interactions as well as data imbalance (e.g. unequal number of observations among cells and missing cells). The DIARANDM.SAS output includes (1) hypothesis testing for variance components of GCA, SCA and all interaction effects, (2) estimate of GCA, SCA and all interaction variances, (3) estimate of standard error for GCA and SCA and all interaction variance components, and (4) variance and covariance matrix for computing genetic correlations and other genetic and environmental parameters. Two examples are given to demonstrate the use of the program.

Key words: Mating design, half-diallel, random effect, variance components, radiata pine.

Introduction

Diallel mating designs are frequently deployed in tree breeding programs to maximize opportunity for managing coancestry in breeding populations and to maximize the selection differential (van Buijtenen, 1976; Bridgwater, 1992). A salient feature of diallel mating design is that it provides a pedigreed breeding population, which can be used to estimate fixed parental genetic effects for backward selection. Variance components may also be estimated from diallels for general (GCA) and specific (SCA) combining abilities as well as heritability, and genetic correlation (Griffing, 1956). In the former, parental effects (GCA and SCA) are regarded as sole interests and fixed effects (or model 1, Eisenhart, 1947) are estimated to rank the parents for selection of future breeding and deployment material. For variance component estimation, parents are regarded as random selections from a population (random effects or model 2), and genetic variances and population parameters are estimated to provide prediction of genetic gain and other essential information assisting in development of optimal breeding strategies.

Whether estimating fixed effects or variance components, diallels are the most complex mating design to analyse relative to most others used in plant breeding programs (e.g. open-pollinated, polycrossed, single-paired, nested, and factorial). Because the same parent acts as both male and female in the mating structure, available commercial statistical packages do not allow direct specification of the diallel mating structure in the linear model procedure and therefore are not able to analyse the design directly, even for a balanced structure with no missing crosses. Analysing diallel matings with missing crosses is even more demanding. To circumvent the difficult of analysing diallel crosses, some specialised programs were developed for diallel mating design with no missing crosses (Magari and Kang, 1994; Burow and Coors, 1994; Linda, 1993; Zhang and Kang, 1997) and with missing crosses (Schaffer and Usanis, 1969; Snyder, 1975; Dean and Corell, 1990; Johnson and King, 1998). However, they all suffer from incompleteness (lack of hypothesis testing or limited model specification) (Wu and Matheson, 2000). In a previous paper, we presented a more complete method and algorithm to estimate GCA and SCA fixed effects for half-diallels with missing and non-missing crosses, along with statistical testing of GCA and SCA fixed effects and a SAS program to implement the algorithm (Wu and Matheson, 2000). In this paper, a general linear model method for fitting constants (Henderson method III) to estimate GCA and SCA variance components is presented for half-diallel mating designs with missing crosses. The fitting constant method is used because it can apply to random models as well as mixed models and provide unbiased estimates of variance components for unbalanced data (missing crosses, missing cells, and unequal number of trees among cells). A comprehensive SAS program for computing unbiased estimates of GCA and SCA variance components is developed along with hypothesis testing and sampling variances and covariances (The code of the program can be downloaded from the Web site www.ffp.csiro.au). The theory of the fitting constant method for analysing diallel random models has not been presented before in the literature. A brief description of the theory will assist tree breeders to understand assumptions and the algorithm underlying the analysis. Through understanding of these matrix operations, data analysts and tree breeders could implement the algorithm into any other statistical packages equipped with matrix manipulation.

Theory of Fitting Constants Method for Estimating Variance Components in Half-Diallel Mating Designs

Assuming a half-diallel mating structure of 5 parents with crosses C13 and C25 missing, as illustrated in following diagram:

	Female (J)			
Male (I)	2	3	4	5
1	C12		C14	C15
2		C23	C24	
3			C34	C35
4				C45

[1]) CSIRO Division of Forestry and Forest Products, PO Box E4008, Kingston, Canbera, ACT 2604, Australia,
E-mail: Harry.wu@ffp.csiro.au.

the standard scalar linear model for estimating variance is usually expressed as

$$Y_{ijk} = \mu + g_i + g_j + s_{ij} + e_{ijk} \quad (1)$$

where Y_{ijk} is the k^{th} observation for cross between the i^{th} and the j^{th} parent, μ is the grand mean, g_i and g_j are the general combining abilities for the i^{th} and j^{th} parents, respectively, s_{ij} is the specific combining ability between the i^{th} and j^{th} parents and e_{ijk} is the residual. In matrix notation, this is expressed as

$$\mathbf{Y} = \mu\mathbf{1}_n + \mathbf{Z}_g\,\mathbf{g} + \mathbf{Z}_s\,\mathbf{s} + \mathbf{e} \quad (1a)$$

where **Y** and **e** are vectors of individual observations and residuals, respectively, $\mathbf{1}_n' = \{1\ 1\ 1\ \cdots\}$ corresponding to the n individual observations, $\mathbf{Z}_g$ and $\mathbf{Z}_s$ are matrices arranged according to following array of crosses:

$$\begin{bmatrix} c12 \\ c14 \\ c15 \\ c23 \\ c24 \\ c34 \\ c35 \\ c45 \end{bmatrix} \quad \mathbf{Z}_g = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \quad \mathbf{Z}_s = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $\mathbf{g}' = [g_1\ g_2\ g_3\ g_4\ g_5]$, and $\mathbf{s}' = [s_{12}\ s_{14}\ s_{15}\ s_{23}\ s_{24}\ s_{34}\ s_{35}\ s_{45}]$.

Estimation of variance components for the fitting constant method starts with estimation of sum of squares and expected sum of squares for each effect or a set of effects in linear model 1 or 1a and finishes with equating estimated sum of squares to their expected values (method of moments). The sum of squares for a particular effect or a set of effects is estimated by fitting a linear model with the corresponding effect or a set of effects as the model elements. For example, with linear model 1 or 1a, sums of squares could be estimated for grand mean μ, combinations of μ and GCA effects, μ and SCA effects, and μ, GCA, and SCA effects. These sums of squares are derived by fitting the model $\mathbf{Y} = \mu\mathbf{1}_n + \mathbf{e}$, $\mathbf{Y} = \mu\mathbf{1}_n + \mathbf{Z}_g\,\mathbf{g} + \mathbf{e}$, $\mathbf{Y} = \mu\mathbf{1}_n + \mathbf{Z}_s\,\mathbf{s} + \mathbf{e}$, **and** $\mathbf{Y} = \mu\mathbf{1}_n + \mathbf{Z}_g\,\mathbf{g} + \mathbf{Z}_s\,\mathbf{s} + \mathbf{e}$, respectively. The values of the sums of squares are estimated by $\mathbf{Y}'\,\mathbf{X}(\mathbf{X}'\,\mathbf{X})^{-}\,\mathbf{X}'\,\mathbf{Y}$ with **X** equal to $\{\mathbf{1}_n\}$, $\{\mathbf{1}_n\ \mathbf{Z}_g\}$, $\{\mathbf{1}_n\ \mathbf{Z}_s\}$, $\{\mathbf{1}_n\ \mathbf{Z}_g\ \mathbf{Z}_s\}$, respectively, for grand mean and three combinations (p. 443, SEARLE, 1971). Sums of squares derived in this way are often referred as the reductions of sum of squares of fitting a model and usually denoted as $\mathbf{R}(\mu)$, $\mathbf{R}(\mu\ g)$, $\mathbf{R}(\mu\ s)$, $\mathbf{R}(\mu\ g\ s)$, respectively. They are called reductions because they are true reductions of sum of squares from the total sum of squares after fitting the corresponding models.

Not all sums of squares are equally informative in estimating variance components. According to theory of the fitting constant method, expected sums of squares for $\mathbf{R}(\mu\ g)$, $\mathbf{R}(\mu\ s)$ and $\mathbf{R}(\mu\ g\ s)$ include a function of the grand mean μ, a fixed term, which is not very useful (nussiance parameter) in estimating variance components and it is often excluded. To eliminate μ, a new set of sum of squares is derived by subtracting $\mathbf{R}(u)$ from $\mathbf{R}(\mu\ g)$, $\mathbf{R}(\mu\ s)$, and $\mathbf{R}(\mu\ g\ s)$, respectively, which results in a new set of sum of squares denoted as $\mathbf{R}(g \mid \mu) = \mathbf{R}(\mu\ g) - \mathbf{R}(\mu)$, $\mathbf{R}(s \mid \mu) = \mathbf{R}(\mu\ s) - \mathbf{R}(\mu)$, $\mathbf{R}(g\ s \mid \mu) = \mathbf{R}(\mu\ g\ s) - \mathbf{R}(\mu)$. In the same way, sums of squares only involving GCA and SCA effect could also be obtained. For example, we estimated $\mathbf{R}(g \mid \mu\ s) = \mathbf{R}(\mu\ g\ s) - \mathbf{R}(\mu\ s)$, $\mathbf{R}(s \mid \mu\ g) = \mathbf{R}(\mu\ g\ s) - \mathbf{R}(\mu\ g)$. These $\mathbf{R}(\alpha \mid \beta)$ notations are often referred to as the sum of squares due to fitting α after fitting β (SEARLE, 1971; SEARLE *et al.*, 1992).

In the half-diallel mating model 1 or 1a, one useful set of sum of squares (or reduction of sum of squares) is $\mathbf{R}(g \mid \mu)$, $\mathbf{R}(s \mid \mu\ g)$, $\mathbf{R}(g \mid \mu\ s)$, and $\mathbf{R}(g\ s \mid \mu)$, all obtained by subtraction as above, plus the residual sum of squares ($\mathbf{SSE} = \Sigma\, Y_{ijk}^2 - \mathbf{R}(\mu\ g\ s)$. Among these six sums of squares, $\mathbf{R}(\mu\ s) = \mathbf{R}(\mu\ g\ s)$ because SCA is the interaction between two GCAs. Thus, the sum of squares for GCA in model 1 or 1a can not be estimated by subtracting the sum of squares for SCA from sum of squares for the whole model since the reduction $\mathbf{R}(g \mid \mu\ s) = \mathbf{R}(\mu\ g\ s) - \mathbf{R}(\mu\ s) = 0$. Furthermore, the reduction of $\mathbf{R}(s \mid \mu) = \mathbf{R}(\mu\ s) - \mathbf{R}(\mu)$ is derived from fitting a submodel of family effects (e.g. $Y_{ijk} = \mu + s_{ij} + e_{ijk}$), which is equivalent to $\mathbf{R}(g\ s \mid \mu)$ ($= \mathbf{R}(\mu\ g\ s) - \mathbf{R}(\mu)$). Hence, only one set of three sums of squares (e.g. $\mathbf{R}(g \mid \mu)$, $\mathbf{R}(g\ s \mid \mu)$, $\mathbf{R}(s \mid \mu\ g)$ plus the sum of squares for the residual (**SSE**) is useful in estimating GCA and SCA variance components for a half-diallel mating structure (*Table 1*).

Table 1. – Possible sums of squares, mean squares and their expected sum of squares, mean squares for estimating GCA (σ_g^2) and SCA (σ_s^2) variance components in a half-diallel mating design.

Sum of squares	Expected sum of squares	Mean squares	Expected mean squares
$R(g \mid \mu) = R(\mu\ g) - R(\mu)$	$k_7\,\sigma_e^2 + k_8\,\sigma_g^2$	MSG	$\sigma_e^2 + \gamma_8\,\sigma_g^2$
$R(g\ s \mid \mu) = R(\mu\ g\ s) - R(\mu)$	$k_4\,\sigma_e^2 + k_5\,\sigma_s^2 + k_6\,\sigma_g^2$	MSGS	$\sigma_e^2 + \gamma_5\,\sigma_s^2 + \gamma_6\,\sigma_g^2$
$R(s \mid \mu\ g) = R(\mu\ g\ s) - R(\mu\ g)$	$k_2\,\sigma_e^2 + k_3\,\sigma_s^2$	MSS	$\sigma_e^2 + \gamma_3\,\sigma_s^2$
$SSE = \Sigma\, Y_{ijk}^2 - R(\mu\ g\ s)$	$k_1\,\sigma_e^2$	MSE	σ_e^2

Now, only three sums of squares (e.g. $\mathbf{R}(g \mid \mu)$ and $\mathbf{R}(s \mid \mu\ g)$ or $\mathbf{R}(g\ s \mid \mu)$ and $\mathbf{R}(s \mid \mu\ g)$, plus **SSE**) are sufficient for estimating GCA and SCA variance components. The combination of $\mathbf{R}(g\ s \mid \mu)$ and $\mathbf{R}(s \mid \mu\ g)$ is preferable because computing the expected sum of squares for $\mathbf{R}(g \mid \mu) = \mathbf{R}(\mu\ g) - \mathbf{R}(\mu)$ is more cumbersome than for $\mathbf{R}(g\ s \mid \mu) = \mathbf{R}(\mu\ g\ s) - \mathbf{R}(\mu)$. This is because the former is the difference of fitting two submodels and the later is the difference of fitting a full model and submodel. According to theory, estimating the expected sum of squares for the difference of two submodels requires their sums of squares to be converted into the difference between two reductions involving the full model (SEARLE, 1971). For the half-diallel mating design, if the sum of squares $\mathbf{R}(g \mid \mu)$ was used instead of $\mathbf{R}(g\ s \mid \mu)$, then $\mathbf{R}(g \mid \mu)$ ($= \mathbf{R}(\mu\ g) - \mathbf{R}(\mu)$) must be converted into the difference $\mathbf{R}(g\ s \mid \mu) - \mathbf{R}(s \mid \mu\ g)$ first. The expected sum of squares would be estimated for this new difference. It is apparent that the sum of squares for $\mathbf{R}(g \mid \mu)$ is a function of sum of squares of $\mathbf{R}(g\ s \mid \mu)$ and $\mathbf{R}(s \mid \mu\ g)$. Thus, use of the sums of squares $\mathbf{R}(g\ s \mid \mu)$ and $\mathbf{R}(s \mid \mu\ g)$ in estimating variance components is simpler. This is the basis of our development of the algorithm and the SAS program.

Applying the theory of fitting constant method to model 1 or 1a, coefficients in the expected sum of squares for $\mathbf{R}(s \mid \mu\ g)$ are given (p. 444, SEARLE, 1971) by

$$\text{Expected } R(s \mid \mu\ g) = \text{Expected } [R(\mu\ g\ s) - R(\mu\ g)]$$
$$= trace\{(I_n - [1_n\ Z_g]\left(\begin{bmatrix} 1_n' \\ Z_g' \end{bmatrix}[1_n\ Z_g]\right)^{-}\begin{bmatrix} 1_n' \\ Z_g' \end{bmatrix})\ Z_s\ Z_s'\ \sigma_s^2\} + (r[1_n\ Z_g\ Z_s] - r[1_n\ Z_g])\ \sigma_e^2.$$

and for the sum of squares $\mathbf{R}(g\ s \mid \mu)$, coefficients are computed by

$$\text{Expected } R(g\ s \mid \mu) = \text{Expected } [R(\mu\ g\ s) - R(\mu)]$$
$$= trace\{(I_n - \frac{1}{n}\ 1_n\ 1_n')\ [Z_g\ Z_s]\begin{bmatrix} Z_g' \\ Z_s' \end{bmatrix}\begin{bmatrix} \sigma_g^2 & 0 \\ 0 & \sigma_s^2 \end{bmatrix}\} + (r[1_n\ Z_g\ Z_s] - r[1_n])\ \sigma_e^2$$
$$= trace\{C_n\ Z_g\ Z_g'\}\ \sigma_g^2 + trace\{C_n\ Z_s\ Z_s'\}\ \sigma_s^2 + (r[1_n\ Z_g\ Z_s] - 1])\ \sigma_e^2.$$

where $\mathbf{I}_n$ is an identity matrix of dimension n, r $[\mathbf{1}_n\ \mathbf{Z}_g\ \mathbf{Z}_s]$, r $[\mathbf{1}_n\ \mathbf{Z}_g]$, and r $[\mathbf{1}_n]$ are ranks for matrices $[\mathbf{1}_n\ \mathbf{Z}_g\ \mathbf{Z}_s]$, $[\mathbf{1}_n\ \mathbf{Z}_g]$, and $[\mathbf{1}_n]$, respectively, and $\mathbf{C}_n$ represents the matrix ($\mathbf{I}_n$ -1/n $\mathbf{1}_n\mathbf{1}_n'$).

When the coefficients of expected sums of squares are estimated, the variance components are usually derived by solving following equation (see *Table 1* and SEARLE, 1971)

$$\begin{bmatrix} \hat{\sigma}^2_g \\ \hat{\sigma}^2_s \\ \hat{\sigma}^2_e \end{bmatrix} = \begin{bmatrix} k_6 & k_5 & k_4 \\ 0 & k_3 & k_2 \\ 0 & 0 & k_1 \end{bmatrix}^{-1} \begin{bmatrix} R(g\ s \mid \mu) \\ R(s \mid \mu\ g) \\ SSE \end{bmatrix}$$

Standard errors for estimated GCA and SCA variance components are approximately estimated using a TAYLOR's expansion (NAMKOONG, 1979). For example, the standard error for the GCA variance of model 1 is estimated approximately by

$$S.E.(\sigma^2_g) = \sqrt{\frac{1}{\gamma_6^2}\left[\frac{2*MSGS^2}{k_4+2} + \frac{2\,(\frac{\gamma_5}{\gamma_3})^2*MSS^2}{k_2+2} + \frac{2*(\frac{\gamma_5-\gamma_3}{\gamma_3})^2*MSE^2}{k_1+2}\right]}$$

where MSGS, MSS, and MSE are the mean squares for the joint GCA and SCA effects, GCA effect, and residual effect, respectively, and γ_3, γ_5 and γ_6 are coefficients for expected mean squares (*Table 1*). Similarly, sampling variances can be estimated for SCA and other variance component estimates.

With missing crosses, significance of variance component σ_g^2 can not be tested directly from the computed sum of squares since appropriate mean squares for the denominator of an *F* test do not exist for model 1 or 1a. To derive the proper denominator for testing σ_g^2, a SATTERTHWAITE synthesis (1946) is used as an approximate approach to synthesize for the denominator and its degree of freedom. For the half-diallel mating model, to synthesize mean squares to test σ_g^2 , we need a term to have an expected value of $\sigma_e^2 + \gamma_5\,\sigma_s^2$ since $E(MSGS) - \gamma_6\,\sigma_g^2 = \sigma_e^2 + \gamma_5\,\sigma_s^2$ (*Table 1*). It is obvious that $\gamma_5 / \gamma_3\ E(MSS) + (1- \gamma_5 / \gamma_3)\ E(MSE)$ has expected value of $\sigma_e^2 + \gamma_5\,\sigma_s^2$ and can be used for testing σ_g^2. Therefore an *F* statistic can be constructed as $F = MSGS/(\gamma_5 / \gamma_3 *(MSS) + (1- \gamma_5 / \gamma_3)*MSE)$ and with degree of freedom for the denominator estimated as $q=(\gamma_5 / \gamma_3 *(MSS) + (1- \gamma_5 / \gamma_3)*MSE)^2 / (\gamma_5 / \gamma_3 *(MSS)/k_2 +(1- \gamma_5 / \gamma_3)*MSE/k_1)$. A similar method is used for testing significance of other variance components.

Now, if there are S sets of disconnected half-diallel crosses and they are planted on L sites with M replications at each site, the full scalar linear model including site (E), replication (R), set (T) effects and all interactions is expressed as

$$Y_{mij(l)k} = \mu + E_l + R_{m(l)} + T_s + ET_{ls} + RT_{ms(l)} + G_{i(s)} + G_{j(s)} + S_{ij(s)} + EG_{li(s)} + EG_{lj(s)} + ES_{lij(s)} + RG_{mi(ls)} + RG_{mj(ls)} + RS_{mij(ls)} + e_{mij(l)k} \quad (2)$$

where ET_{ls} is the interaction effect between l^{th} site and s^{th} set, $RT_{ms(l)}$ is the interaction between m^{th} replication within the l^{th} site and s^{th} set effect, $G_{i(s)}$ is the i^{th} GCA effect within s^{th} set, $S_{ij(s)}$ is the ij^{th} SCA effect within the s^{th} set, EG_{li} is the interactions between the l^{th} site and the i^{th} GCA effects within s^{th} set, $ES_{lij(s)}$ is the interaction between the l^{th} site and the ij^{th} SCA effect within the s^{th} set, $RG_{mi(ls)}$ and $RG_{mj(ls)}$ are the interactions between the m^{th} replication in the l^{th} site and the i^{th} plus j^{th} GCA effects, respectively within the s^{th} set, and $RS_{mij(ls)}$ is the interaction between the m^{th} replication in the l^{th} site and the ij^{th} SCA effect within the s^{th} set. There are many possible sets of reductions of sum of squares for estimating variance components for this extended model, one sequential set of reduction of special importance is:

R(E | μ) = R(E R T ET RT G S EG ES RG RS | μ) – R(R T ET RT G S EG ES RG RS | μ E)
R(R | μ E) = R(R T ET RT G S EG ES RG RS | μ E) – R(T ET RT G S EG ES RG RS | μ E R)
R(T | μ E R) = R(T ET RT G S EG ES RG RS | μ E R) – R(ET RT G S EG ES RG RS | μ E R T)
R(ET | μ E R T) = R(ET RT G S EG ES RG RS | μ E R T) – R(RT G S EG ES RG RS | μ E R T ET)
R(RT | μ E R T ET) = R(RT G S EG ES RG RS | μ E R T ET) – R(G S EG ES RG RS | μ E R T ET RT)
R(G | μ E R T ET RT) = R(G S EG ES RG RS | μ E R T ET RT) – R(S EG ES RG RS | μ E R T ET RT G)
R(S | μ E R T ET RT G) = R(S EG ES RG RS | μ E R T ET RT G) – R(EG ES RG RS | μ E R T ET RT G S)
R(EG |μ E R T ET RT G S) = R(EG ES RG RS |μ E R T ET RT G S) – R(ES RG RS |μ E R T ET RT G S EG)
R(ES | μ E R T ET RT G S EG) = R(ES RG RS | μ E R T ET RT G S EG) – R(RG RS | μ E R T ET RT G S EG ES)
R(RG | μ E R T ET RT G S EG ES) = R(RG RS | μ E R T ET RT G S EG ES) – R(RS | μ E R T ET RT G S EG ES RG)
R(RS | μ E R T ET RT G S EG ES RG) = R(μ E R T ET RT G S EG ES RG RS) – R(μ E R T ET RT G S EG ES RG)
SSE = $\Sigma\, Y_{mij(l)k}^2$ – R(μ E R T ET RT G S EG ES R RG RS)= Y'IY – R(μ E R T ET RT G S EG ES RG RS).

The above set is simpler for computing and estimation of all model components relative to other sets of deductions, the expected sums of squares for the sequence of the set are listed in *Table 2*.

Table 2. – A preferable sequential set of sums of squares and expected sum of squares for estimating GCA (σ_g^2) and SCA (σ_s^2) and other variance components in a series of disconnected half-diallel mating designs planted in L sites with M replications each site.

Sum of Squares	Expected Sum of Squares
R(E | μ)	$k_{67}\,\sigma_e^2 + k_{68}\,\sigma_{rs}^2 + k_{69}\,\sigma_{rg}^2 + k_{70}\,\sigma_{es}^2 + k_{71}\,\sigma_{eg}^2 + k_{72}\,\sigma_s^2 + k_{73}\,\sigma_g^2 + k_{74}\,\sigma_{rt}^2 + k_{75}\,\sigma_{et}^2 + k_{76}\,\sigma_t^2 + k_{77}\,\sigma r^2 + k_{78}\,\sigma_E^2$
R(R | μE)	$k_{56}\,\sigma_e^2 + k_{57}\,\sigma_{rs}^2 + k_{58}\,\sigma_{rg}^2 + k_{59}\,\sigma_{es}^2 + k_{60}\,\sigma_{eg}^2 + k_{61}\,\sigma_s^2 + k_{62}\,\sigma_g^2 + k_{63}\,\sigma_{rt}^2 + k_{64}\,\sigma_{et}^2 + k_{65}\,\sigma_t^2 + k_{66}\,\sigma r^2$
R(T | μE R)	$k_{46}\,\sigma_e^2 + k_{47}\,\sigma_{rs}^2 + k_{48}\,\sigma_{rg}^2 + k_{49}\,\sigma_{es}^2 + k_{50}\,\sigma_{eg}^2 + k_{51}\,\sigma_s^2 + k_{52}\,\sigma_g^2 + k_{53}\,\sigma_{rt}^2 + k_{54}\,\sigma_{et}^2 + k_{55}\,\sigma_t^2$
R(ET | μE R T)	$k_{37}\,\sigma_e^2 + k_{38}\,\sigma_{rs}^2 + k_{39}\,\sigma_{rg}^2 + k_{40}\,\sigma_{es}^2 + k_{41}\,\sigma_{eg}^2 + k_{42}\,\sigma_s^2 + k_{43}\,\sigma_g^2 + k_{44}\,\sigma_{rt}^2 + k_{45}\,\sigma_{et}^2$
R(RT | μ E R T ET)	$k_{29}\,\sigma_e^2 + k_{30}\,\sigma_{rs}^2 + k_{31}\,\sigma_{rg}^2 + k_{32}\,\sigma_{es}^2 + k_{33}\,\sigma_{eg}^2 + k_{34}\,\sigma_s^2 + k_{35}\,\sigma_g^2 + k_{36}\,\sigma_{rt}^2$
R(G | μ E R T ET RT)	$k_{22}\,\sigma_e^2 + k_{23}\,\sigma_{rs}^2 + k_{24}\,\sigma_{rg}^2 + k_{25}\,\sigma_{es}^2 + k_{26}\,\sigma_{eg}^2 + k_{27}\,\sigma_s^2 + k_{28}\,\sigma_g^2$
R(S | μ E R T ET RT G)	$k_{16}\,\sigma_e^2 + k_{17}\,\sigma_{rs}^2 + k_{18}\,\sigma_{rg}^2 + k_{19}\,\sigma_{es}^2 + k_{20}\,\sigma_{eg}^2 + k_{21}\,\sigma_s^2$
R(EG | μ E R T ET RT G S)	$k_{11}\,\sigma_e^2 + k_{12}\,\sigma_{rs}^2 + k_{13}\,\sigma_{rg}^2 + k_{14}\,\sigma_{es}^2 + k_{15}\,\sigma_{eg}^2$
R(ES | μ E R T ET RT G S EG)	$k_7\,\sigma_e^2 + k_8\,\sigma_{rs}^2 + k_9\,\sigma_{rg}^2 + k_{10}\,\sigma_{es}^2$
R(RG | μ E R T ET RT G S EG ES)	$k_4\,\sigma_e^2 + k_5\,\sigma_{rs}^2 + k_6\,\sigma_{rg}^2$
R(RS | μ E R T ET RT G S EG ES RG)	$k_2\,\sigma_e^2 + k_3\,\sigma_{rs}^2$
SSE	$k_1\,\sigma_e^2$

To estimate sum of squares, mean squares and expected sum or mean of squares for this extended model, the appropriate design matrix is first established. The principle illustrated for model 1 or 1a is then applied to this extended model for estimating variance components, along with sampling variance and hypothesis testing. This formulation for sum of squares is similar to Type I sum of squares used by SAS GLM and VARCOMP Procedures. The only difference is that the required design matrix for the half diallel mating structure and all the interactions are established through a SAS Macro program.

If the experiment involves only a single half-diallel set and the interactions between replication and GCA and SCA effects are not interested, then the model is simpler:

$$Y_{mij(l)k} = \mu + E_l + R_{m(l)} + G_i + G_j + S_{ij} + EG_{li} + EG_{lj} + ES_{lij} + e_{mij(l)k.} \quad (3)$$

Furthermore, if the experiment is planted only on a single site, then the model reduces to

$$Y_{mijk} = \mu + R_m + G_i + G_j + S_{ij} + RG_{gmi} + RG_{mj} + RS_{mij} + e_{mijk} \quad (4)$$

These models (model 1 to 4) are the basis for the following SAS program.

Detail of SAS Program

A comprehensive SAS program (DIARANDM.SAS, available from the senior author upon request) for analysing half-diallels at multiple sites with multiple replications for balanced and unbalanced data (missing crosses, missing cells and unequal number of observations) has been developed using SAS/Macros and the IML Procedure (SAS Institute 1987, 1989, 1990). The program can also analyse multiple disconnected diallel sets simultaneously and perform multivariate analysis.

Before running the program, the user must specify four basic SAS Macro variables, the desired linear model for analysis and variable names if multivariate analysis is required. The four Macro variables are:

N_DIALL: Number of diallel set (e.g. N_DIALL=3 for 3 sets of diallels);

N_PARENT: Number of parent in the half-diallel (e.g. N_PARENT=5 for a 5 by 5 half-diallel);

N_ENV: Number of environments (sites) (e.g. N_ENV=2 for 2 sites);

N_REP: Number of replication each environment (e.g. N_REP=2 for 2 replications at each site).

If there are missing crosses in some sets, these missing crosses must be listed as macro variables in the program. For example, Macro variable MC lists missing crosses such as:

MC1: List of all missing crosses for set 1 (e.g. MC1=S13 S15 for indicating missing crosses C13 and C15);

MC2: List of all missing crosses for set 2, and so on.

The linear model specification is similar to SAS GLM and VARCOMP Procedures: e. g. for extended model 2, the MODEL statement is specified as

MODEL= ENV REP SET ESET RSET GCAS SCAS EGCA ESCA RGCA RSCA;

where ENV is the environment (site) effect, REP is the designation for replication or replication within environment effect, SET is the diallel set effect, ESET is for environment by set interaction, RSET is for replication by set interaction, GCAS, and SCAS are GCA and SCA effects within each set, respectively. EGCA and ESCA denote interaction effects between environment and within-set GCA, and between environment and within-set SCA, respectively. RGCA and RSCA represents interactions between replication and within-set GCA, and between replication and within-set SCA, respectively. Thus, the only differences from SAS GLM are that;

(1) the symbol REP is always used to represent either the replication within environment for a multiple sites experiment (cf REP(ENV) in SAS GLM) or the replication effect in a single site experiment (cf REP in SAS GLM),

(2) the environment and replication by set interaction are simplified as ESET, RSET (cf ENV*SET and REP*SET(ENV) in GLM),

(3) GCAS and SCAS are the same as GCA(SET) and SCA(SET) in SAS GLM,

(4) EGCA, ESCA, RGCA and RSCA are same as ENV*GCA(SET), ENV*SCA(SET), REP*GCA(ENV SET), and REP*SCA(ENV SET) in SAS GLM, respectively.

There are some simple format requirements for the diallel raw data to be used for analysis. First, variables for environment (ENV), replication (REP), parent (male and female) should be coded as continuous numeric data and with values starting from 1 (e.g. ENV=1, 2, 3 for a three sites experiment; I or J=1, 2, 3, 4, 5 for variable I, J for an 5 parent half-diallel). Second, variables I and J are used for labelling parents in the diallel, I is used to designate the female parent (I=1, 2, 3, 4 for an 5 parent diallel), J is used to designate the male parent (J=2, 3, 4, 5 for an 5 parent diallel). Third, raw data should be sorted by I first, then by J before merging with diallel design matrix. Fourth, all missing values in variables for analysis should be deleted before entering into the Macro MODEL since multiplication and general inverse operations in PROC IML do not recognise missing values. All these can be accomplished within the SAS data steps and procedures.

The core of DIARANDM.SAS is divided into three major steps:

Step 1 generates the design matrix for the half-diallel mating structure with or without missing crosses and the design matrices for the experimental design. The missing crosses are flagged in the design matrix.

Step 2 combines raw data with the mating design and experimental design matrices. Missing plots (cells) and unequal numbers of observations for each plot (cell) are also accounted for by this approach. Thus, the whole design matrix deals with dual data unbalances (mating design and experiment).

Step 3 computes sum of squares, mean squares, expected sum of squares and mean squares and degree of freedom for each reduction of sum of squares in *Table 2* or its variations. Variance components are estimated from sums of squares and expected sums of squares. Step 3 is mainly made by the Macro program MODEL. Within MODEL, two subroutines (sub-Macros) estimate reduction of sum of squares, conduct hypothesis-testing for each variance component in the model and estimate the sampling variances of the estimated variance components. The sub-Macro REDUCT estimate the reduction of sum of squares and coefficients of expected sum of squares for each effect in the model. The sub-Macro HTEST conducts hypothesis-testing for each variance component and estimate unbiased sampling variance for variance estimates.

The program can also partially deal with mixed models. For a multiple-sites experiment, site or/and replication can be assumed as fixed or random effects. If they are assumed fixed, the calculated expected coefficients for sites or/and replications under a random assumption are no longer valid. However, these do not affect estimates of other variance components since site or/and replication effects are not in the expected mean squares for effects after them. For a single site experiment, the testing for replication is still valid, if replication is assumed fixed. For a multiple-site experiment, the testing for site is valid only if site is assumed fixed. If both site and replication are assumed fixed, then testing for site is valid only if there is no significant replication effect. Whether sites or/and replication are assumed fixed or random, the interaction effects between site and GCA and SCA as well as between replication and GCA and SCA are random.

Examples

Two examples were presented to demonstrate the use of the program and the features of output. The first example is taken from GRIFFING's original data used in his 1956 paper. SCHAFFER and USANIS (1969) have also used this data in their Fortran program. The second example used data from a radiata pine genetic trial to demonstrate multivariate analysis for a series

of disconnected sets of half-diallel with missing crosses planted on multiple sites.

GRIFFING's Example

Corn yield data from GRIFFING's half-diallel mating example (GRIFFING, 1956, p. 482) was used to compare results between SCHAFFER and USANIS's DIALL program with the results from DIARANDM.SAS program. The design is an 9x9 half-diallel structure without missing crosses. Before running the DIARANDM.SAS program, we need specify following five macro variables, the desired linear model, and the variables for the analysis as following

```
%LET N_DIAL=1                 *NUMBER OF DIALLEL SET;
%LET N_PARENT=9;              *NUMBER OF PARENT IN
                               DIALLEL CROSSES;
%LET MC1=;                    *LIST ALL MISSING
                               CROSSES WITH INITIAL S:
                               e.g. S13 S25;
%LET N_ENV=1;                 *TOTAL NUMBER OF SITE:
                               e.g 1, 2,...;
%LET N_REP=1;                 *NUMBER OF REPLICATION
                               AT EACH SITE;
%LET MODEL= GCA SCA;          *LINEAR MODEL;
%LET VARIABLEL=YIELD;         *VARIABLES TO BE
                               ANALYSED;
```

The output result is listed below:

SAS Program 'DIARANDM.SAS' to Estimate Variance Components for Half-diallel Mating Design

Expected MS for 9 by 9 Half-diallel planted in 1 Sites with 1 Replications, and 0 Missing Crosses

DEGREE OF FREEDOM AND MEAN SQUARES

SOURCES	DF	MS
GCA	35.00	793.45
SCA	27.00	339.44

DEGREE OF FREEDOM AND COEFFICIENT OF EXPECTED MEAN SQUARES

SOURCES	DF	GCA	SCA
GCA	35.00	56.00	35.00
SCA	27.00	0.00	27.00

HYPOTHESIS TEST FOR: GCA

ERROR TERM FOR GCA +1.000*MS_SCA+ 0.000*MS_ERR

DF	MS	DF_Denom	MS_Denom	F-Value	Pr>F
35.000	793.452	27.000	339.439	2.338	0.013

HYPOTHESIS TEST FOR: SCA

ERROR TERM FOR SCA + 1.000*MS_ERR

DF	MS	DF_Denom	MS_Denom	F-Value	Pr>F
27.000	339.439	0.000	0.000	0.000	.

VARIANCE COMPONENTS

VARIABLE	VARIANCE	STANDARD ERROR
GCA	283.758	128.051
SCA	339.439	89.141

Since there is no error term, the SCA variance canot be tested. The GCA and SCA variances (283.8, 339.4) are exactly the same as estimates from DIALL program, but the standard error for GCA variance (128.1) is smaller than the DIALL estimate (149.1) by SCHAFFER and USANIS (1969). This is because different sums of squares were used (p. 208, SEARLE et al., 1992). Since there is no error term with the full model for this data set, we could use the alternative model MODEL=GCA for the analysis. When MODEL=GCA is used, result showed that both variance estimates for GCA and SCA and their standard errors were same as estimates from DIALL's program (e.g. standard error 149.1 and 89.1, respectively).

Example of Multivariate and Multiple Sites for Radiata Pine (Three Disconnectd Diallel Sets Planted in Two Sites)

This example presents an analysis of radiata pine data at two sites each with three replications and 4 tree plots. There are three sets of 6 by 6 half-diallel with two missing crosses (C16, C23) for the first set, one missing cross (C14) for the second set and two missing crosses (C35, C56) for the third set (part of a Australia-wide diallel mating experiment). Three variables (DBH, STEM-stem straightness scores and CLUST-cluster number) are analysed. For demonstration purpose, we test all effects and interactions except for interactions between replication and GCA and SCA within set (i.e. testing effects of site, replication, set, site*set, replication*set(site), GCA(set), SCA(set), site*GCA(set), site*SCA(set). These effects and interactions were included in the MODEL statement. The following seven macro variables are specified before running the DIARANDM.SAS program according to the experiment, linear model and three variables,

```
%LET N_DIALL=3;                 *NUMBER OF DIALLEL SET;
%LET N_PARENT=6;                *NUMBER OF PARENT IN EACH DIALLEL
                                 SET;
%LET MC1=S16 S23;               *LIST ALL MISSING CROSSESS IN SET 1
                                 WITH INITIAL S: e.g. S13 S25;
%LET MC2=S14;                   *LIST ALL MISSING CROSSESS IN SET 2;
%LET MC3=S35 S56;               *LIST ALL MISSING CROSSESS IN SET 3;
%LET N_ENV=2;                   *TOTAL NUMBER OF SITE: e.g 1, 2,...;
%LET N_REP=3;                   *NUMBER OF REPLICATION AT EACH SITE;
%LET MODEL=ENV REP SET ESET RSET GCAS SCAS EGCA ESCA;
                                 *SPECIFY MODEL;
%LET VARIABLE=DBH STEM CLUST; *SPECIFY VARIABLE;
```

The output is listed as below

SAS Program 'DIARANDM.SAS' to Estimate Variance Components for Half-diallel Mating Design

Expected MS for 3 Sets of 6 by 6 Half-diallel Planted in 2 Sites

DEGREE OF FREEDOM AND MEAN SQUARES (MEAN CROSS-PRODUCTS)

SOURCES	DF	DBH*DBH	DBH*STEM	DBH*CLUST	STEM*STEM	STEM*CLUST	CLUST*CLUST
ENV	1.00	33417.6	355.76	-2958.3	3.79	-31.49	261.88
REP	4.00	2357.21	-58.03	-4.54	4.62	-2.60	9.64
SET	2.00	4213.38	-12.42	121.39	3.11	1.70	4.88
ESET	2.00	4139.26	-64.83	30.39	2.28	0.98	1.90
RSET	8.00	831.62	7.84	-4.57	0.97	0.72	3.66
GCAS	15.00	2042.23	-22.33	55.46	4.64	8.50	32.35
SCAS	22.00	878.54	13.28	20.30	1.10	0.82	4.99
EGCA	15.00	1548.79	14.47	-9.77	0.99	0.43	1.46
ESCA	22.00	999.37	3.28	-0.56	1.53	1.15	1.74
ERR	699.00	621.90	2.73	3.52	0.90	0.18	2.30

DEGREE OF FREEDOM AND COEFFICENT OF EXPECTED MEAN SQUARES (MEAN CROSS-PRODUCTS)

SOURCESE	DF	ENV	REP	SET	ESET	RSET	GCAS	SCAS	EGCA	ESCA	ERR
ENV	1.00	394.10	133.39	0.54	131.84	44.82	0.60	0.18	89.46	10.01	1.00
REP	4.00	0.00	130.87	0.35	0.35	44.04	0.56	0.25	0.56	0.25	1.00
SET	2.00	0.00	0.00	262.48	131.83	44.68	177.89	19.84	89.45	10.03	1.00
ESET	2.00	0.00	0.00	0.00	130.65	44.39	0.22	0.14	88.66	9.94	1.00
RSET	8.00	0.00	0.00	0.00	0.00	43.35	0.36	0.25	0.36	0.25	1.00
GCAS	15.00	0.00	0.00	0.00	0.00	0.00	69.40	19.77	34.98	10.00	1.00
SCAS	22.00	0.00	0.00	0.00	0.00	0.00	0.00	19.61	0.11	9.94	1.00
EGCA	15.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	34.26	9.76	1.00
ESCA	22.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.68	1.00
ERR	699.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

HYPOTHESIS TEST FOR: ENV
ERROR TERM FOR ENV

+1.019*MS_REP+ 0.001*MS_SET+1.006*MS_ESET+ -1.032*MS_RSET+ 0.001*MS_GCAS+ 0.000*MS_SCAS+ 0.000*MS_EGCA+ -0.001*MS_ESCA+ 0.006*MS_ERR

	DF	MS	DF_Denom	MS_Denom	F-Value	Pr>F
DBH	1.000	33417.6	3.202	5714.87	5.847	0.089
STEM	1.000	3.787	4.350	6.009	0.630	0.468
CLUST	1.000	261.879	2.310	8.007	32.704	0.021

HYPOTHESIS TEST FOR: REP
ERROR TERM FOR REP

+ 0.001*MS_SET+ 0.001*MS_ESET+ 1.013*MS_RSET+ -0.001*MS_GCAS+ -0.001*MS_SCAS+ -0.001*MS_EGCA+ -0.001*MS_ESCA+ -0.013*MS_ERR

	DF	MS	DF_Denom	MS_Denom	F-Value	Pr>F
DBH	4.000	2357.21	7.989	842.053	2.799	0.101
STEM	4.000	4.625	7.843	0.976	4.739	0.030
CLUST	4.000	9.645	7.805	3.661	2.634	0.116

HYPOTHESIS TEST FOR: SET
ERROR TERM FOR SET

+ 1.009*MS_ESET+ -0.002*MS_RSET+ 2.560*MS_GCAS+ -1.575*MS_SCAS+ -2.609*MS_EGCA+ 1.603*MS_ESCA+ 0.015*MS_ERR

	DF	MS	DF_Denom	MS_Denom	F-Value	Pr>F
DBH	2.000	4213.38	2.639	5588.79	0.754	0.551
STEM	2.000	3.106	11.771	12.326	0.252	0.781
CLUST	2.000	4.883	12.435	75.899	0.064	0.938

HYPOTHESIS TEST FOR: ESET
ERROR TERM FOR ESET

+ 1.024*MS_RSET+ -0.002*MS_GCAS+ -0.004*MS_SCAS+ 2.579*MS_EGCA+ -1.595*MS_ESCA+ -1.003*MS_ERR

	DF	MS	DF_Denom	MS_Denom	F-Value	Pr>F
DBH	2.000	4139.26	5.409	2621.33	1.579	0.288
STEM	2.000	2.276	0.051	0.205	11.093	0.857
CLUST	2.000	1.898	1.774	2.328	0.816	0.561

HYPOTHESIS TEST FOR: RSET

ERROR TERM FOR RSET + 0.005*MS_GCAS+ 0.007*MS_SCAS+ 0.005*MS_EGCA+ 0.007*MS_ESCA+ 0.975*MS_ERR

	DF	MS	DF_Denom	MS_Denom	F-Value	Pr>F
DBH	8.000	831.624	752.538	638.779	1.302	0.239
STEM	8.000	0.973	744.948	0.924	1.053	0.395
CLUST	8.000	3.659	664.069	2.468	1.483	0.160

HYPOTHESIS TEST FOR: GCAS

ERROR TERM FOR GCAS + 1.008*MS_SCAS+ 1.018*MS_EGCA+ -1.028*MS_ESCA+ 0.002*MS_ERR

	DF	MS	DF_Denom	MS_Denom	F-Value	Pr>F
DBH	15.000	2042.23	8.273	1436.05	1.422	0.311
STEM	15.000	4.642	1.267	0.546	8.502	0.205
CLUST	15.000	32.354	15.479	4.724	6.849	0.000

HYPOTHESIS TEST FOR: SCAS

ERROR TERM FOR SCAS + 0.003*MS_EGCA+ 1.024*MS_ESCA+ -0.027*MS_ERR

	DF	MS	DF_Denom	MS_Denom	F-Value	Pr>F
DBH	22.000	878.539	21.498	1011.22	0.869	0.628
STEM	22.000	1.095	21.415	1.542	0.710	0.785
CLUST	22.000	4.990	20.617	1.727	2.888	0.009

HYPOTHESIS TEST FOR: EGCA

ERROR TERM FOR EGCA + 1.009*MS_ESCA+ -0.009*MS_ERR

	DF	MS	DF_Denom	MS_Denom	F-Value	Pr>F
DBH	15.000	1548.79	21.759	1002.73	1.545	0.173
STEM	15.000	0.992	21.772	1.533	0.647	0.805
CLUST	15.000	1.456	21.491	1.738	0.838	0.632

HYPOTHESIS TEST FOR: ESCA

ERROR TERM FOR ESCA + 1.000*MS_ERR

	DF	MS	DF_Denom	MS_Denom	F-Value	Pr>F
DBH	22.000	999.367	699.000	621.897	1.607	0.039
STEM	22.000	1.527	699.000	0.898	1.700	0.024
CLUST	22.000	1.743	699.000	2.299	0.758	0.779

ESTIMATES OF VARIANCE AND COVARIANCE COMPONENTS

VARCOMP	DBH*DBH	DBH*STEM	DBH*CLUST	STEM*STEM	STEM*CLUST	CLUST*CLUST
ENV	70.293	1.239	-7.584	-0.006	-0.074	0.644
REP	11.577	-0.503	-0.000	0.028	-0.025	0.046
SET	-5.240	0.623	-0.167	-0.035	-0.078	-0.271
ESET	11.618	-0.782	0.483	0.016	0.009	-0.003
RSET	4.448	0.118	-0.193	0.001	0.011	0.027
GCAS	8.734	-0.678	0.639	0.059	0.121	0.398
SCAS	-6.765	0.507	1.071	-0.023	-0.018	0.166
EGCA	15.939	0.327	-0.268	-0.016	-0.021	-0.008
ESCA	39.009	0.057	-0.422	0.065	0.100	-0.057
ERR	621.897	2.726	3.521	0.898	0.180	2.299

STANDARD ERRORS OF VARIANCE AND COVARIANCE COMPONENTS

SE_VAR	DBH*DBH	DBH*STEM	DBH*CLUST	STEM*STEM	STEM*CLUST	CLUST*CLUST
ENV	69.731	0.751	6.129	0.011	0.065	0.543
REP	10.790	0.257	0.026	0.021	0.012	0.044
SET	18.314	0.202	0.388	0.019	0.029	0.110
ESET	25.155	0.365	0.178	0.015	0.008	0.020
RSET	8.612	0.081	0.047	0.010	0.007	0.038
GCAS	13.942	0.144	0.291	0.025	0.043	0.162
SCAS	19.847	0.202	0.299	0.028	0.021	0.078
EGCA	17.680	0.148	0.098	0.016	0.011	0.021
ESCA	30.011	0.099	0.026	0.046	0.034	0.054
ERR	33.218	0.146	0.188	0.048	0.010	0.123

where DF_DENOM and MS_DENOM are degrees of freedom and mean squares for the denominator of the *F* test, respectively. DBH*DBH column is for variance of DBH, and DBH*STEM column is for covariance between DBH and STEM. Since there were no significant set, environment by set, replication by set, and environment by GCA interaction effects, model could be simplified as MODEL=ENV REP GCAS SCAS ESCA. The raw data set is available from the senior author upon request so that reader may use them to verify their computation.

Acknowledgements

This paper resulted from a cooperative research project between CSIRO, STBA (Southern Tree Breeding Association Inc.), and State Forests of New South Wales. Financial support from STBA and its members are gladly acknowledged.

Reference

BRIDGWATER, F. E.: Mating designs. In: Handbook of quantitative forest genetics. Edited by FINS, L., FRIEDMAN, S.T., BROTSCHOL, J.V.. pp. 69–95. Kluwer Academic Publishers (1992). — BUROW, M. D. and CORRS, J. G.: Diallel: A microcomputer program for the simulation and analysis of diallel crosses. Agron. J. **86**: 154–158 (1994). — DEAN, C. A. and CORRELL, R. L.: Analysis of diallel matings with missing values. Silvae Genet. **37**: 187–193 (1988). — EISENHART, C.: The assumptions underlying the analysis of variance. Biometrics **3**: 1–21 (1947). — GRIFFING, B.: Concept of general and specific combining ability in relation to diallel crossing system. Aust. J. Biol. Sci. **9**: 463–493 (1956). — JOHNSON, G. R. and KING, J. N.: Analysis of half diallel mating designs. Silvae Genetica **47**: 74–79 (1998). — JONSSON, A., DORMLING, I., ERIKSSON, G. and NORELL, L.: GCA variance components in 36 *Pinus sylvestris* L. Full-sib families cultivated at five nutrient levels in a growth chamber. For. Sci. **38**: 575–593 (1992). — LINDA, S. B.: GRIFFING – a SAS macro implementing GRIFFING's analysis of diallel crossing systems. HortScience **28**: 61, (1993). — MAGARI, R. and KANG, M. S.: Interactive BASIC program for GRIFFING's diallel analysis. J. Hered. **85**: 336 (1994). — NAMKOONG, G.: Introduction to quantitative genetics in forestry. USDA Forest Service, Tech. Bull. No. 1588 (1979). — SAS Institute Inc.: SAS Guide to Macro processing, Version 6, First Edition, Cary, NC. SAS Institute Inc., 233 pp. (1987). — SAS Institute Inc.: SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1&2, Cary, NC. SAS Institute Inc. 943 pp, 846 pp. (1989). — SAS Institute Inc.: SAS/LANGUAGE: Reference, Version 6, First Edition, Cary, NC. SAS Institute Inc., 1042 pp. (1990). — SATTERTHWAITE, F. E.: An approximate distribution of estimates of variance components. Biometrics Bulletin **2**: 110–114 (1946). — SCHAFFER, H. E. and USANIS, R. A.: General least squares analysis of diallel experiments. North Carolina State University, Genetics Department Research Report Number 1. 61 pp. (1969). — SEARLE, S. R.: Linear models. John Wiley and Sons, New York, NY. 532 pp. (1971). — SEARLE,

S. R., CASELLA, G. and MCCULLOCH, C. E.: Variance components. John Wiley and Sons, New York, NY. 501 pp. (1992). — SNYDER, E. B.: Combining-ability determinations for incomplete mating designs. USDA Forest Service, Southern Forest Experiment Station, General Technical Report SO-9. (1975). — VAN BUIJTENEN, J. P.: Mating Designs. In: Proc. Of the IUFRO Joint Meeting of Genetic Working Parties on Advanced Generation Breeding. Bordeaux, pp. 11–27 (1976). — WU, H. X. and MATHESON, A. C.: Analysis of half-diallel mating design with missing crosses: Theory and SAS program for testing and estimating GCA and SCA fixed effects. Silvae Genet. **49**(3): 130–137 (2000). — ZHANG, Y. and KANG, M. S.: Diallel-SAS: A SAS program for GRIFFING's diallel analyses. Agron. J. **89**: 176–182 (1997).

A Fast Method for Checking the Genetic Identity of Ramets in a Clonal Seed Orchard by RAPD Analysis with a Bulking Procedure

By S. GOTO[1]), F. MIYAHARA[2]) and Y. IDE[3])

(Received 30th July 2001)

Summary

In this study we demonstrate a fast method for checking the genetic identity of ramets in a Japanese black pine (*Pinus thunbergii* PARL.) clonal seed orchard, using random amplified polymorphic DNA (RAPD) analysis with a bulking procedure. We used six different artificial mixtures consisting of needle samples from two clones that were bulked in the proportion three to one to test the sensitivity of RAPD markers. We compared the RAPD patterns of the bulked samples with those of the single clones used for the artificial mixtures. Out of 20 markers, 18 markers were present in the bulked samples, when one of the clones possessed the marker. However, two markers were absent in the bulked samples, even though one of the clones possessed the marker. Using the 18 markers, RAPD patterns of the bulked samples were different from those of single clones. Out of 18 markers selected in this study, we used 15 markers for checking the genetic identity of ramets in the seed orchard. First, we collected the needles of 157 trees from the seed orchard, individually. Second, we mixed an equal amount of needle samples from a maximum of four individuals of the same clone, depending on the planting map. Third, we compared the RAPD patterns of the bulked samples with those of their standard individuals of the clone. Out of 42 bulked samples (14 clones x 3 bulked samples) investigated, we found the RAPD patterns of 3 bulked samples to be different from those of the standard individual of the diagnostic clone. Subsequently, we fingerprinted a total of 12 trees comprised of 3 suspicious bulked samples with RAPD markers individually, and detected one mislabeled tree per bulked sample. We were able to check the genetic identity of 157 trees by making a RAPD analysis of 42 bulked samples and 12 individuals. The workload was only about one-third of the workload when making the individual RAPD analyses. We concluded that RAPD analysis with a bulking procedure would be useful for rapidly checking the genetic identity of ramets in clonal seed orchards.

Key words: Bulked samples, genetic identity, mislabeling, *Pinus thunbergii*, RAPD, seed orchard.

Introduction

Tree improvement strategies include the control of natural seed sources and the establishment of orchards of selected genotypes (ZOBEL and TALBERT, 1984). When properly performed, the vegetative propagation method is a powerful means of making the clonal materials consisting of a seed orchard and capturing the genetic superiority of selected individuals. However, mis-plantings and mis-labelings are unfortunately common during the establishment of seed orchards (ADAMS, 1983; HARJU and MUONA, 1989; WHEELER and JECH, 1992). Additionally, the grafted materials are often used for ramets in the seed orchard (HONG, 1975), and sometimes the rootstocks overtake the graft. As it is difficult to detect the genetic identity of ramets in clonal seed orchards by visual inspections, tools are needed for this purpose. Molecular markers have proven to be very useful in distinguishing among related genotypes. Recently developed random amplified polymorphic DNA (RAPD) markers (WELSH and MCCLLELAND, 1990; WILLIAMS *et al.*, 1990) are polymorphic within-species levels (e.g. CARLSON *et al.*, 1991; KEIL and GRIFFIN, 1994; SCHEEPERS *et al.*, 1997), and have been successfully used for distinguishing among orchard clones (VAN DE VEN and MCNICOL, 1995; KAWAUCHI and GOTO, 1999).

Despite the fact that the RAPD procedure is relatively simple and fast, its practical application is still limited in cases where large numbers of individuals need to be examined. The number of individuals in seed orchards must be large enough to allow for the desired spacing, maximum seed production, adequate pollination, and minimum of relatedness among individuals (ZOBEL and TALBERT, 1984), so a fast method for checking the genetic identity of ramets in clonal seed orchards is needed. One approach is to use a bulking procedure. DNA extractions and polymerase chain reaction (PCR) amplifications for several plants can occur in a single step with bulked

[1]) Corresponding author: SUSUMU GOTO, University Forest in Hokkaido, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Yamabe, Furano, Hokkaido 079-1561, Japan.
Ph. +81-167-42-2111; Fax. +81-167-42-2689.
E-mail: gotos@uf.a.u-tokyo.ac.jp

[2]) Fukuoka Prefecture Forest Research and Extension Center, Toyoda 1438-2, Yamamoto-machi, Kurume, Fukuoka 839-0827, Japan.
Ph. +81-942-45-7983; Fax. +81-942-45-7901.

[3]) Laboratory of Forest Ecosystem Studies, Department of Ecosystem Studies, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan.
Ph. +81-3-5841-5490; Fax. +81-3-5841-5494