INBREEDING DEPRESSION IN SELFING EXPERIMENTS: STATISTICAL ISSUES

R. D. Burdon[1] & J. H. Russell[2]

[1)] New Zealand Forest Research Institute Ltd, Private Bag 3020, Rotorua, New Zealand (corresponding author)
[2)] Cowichan Lake Research Station, Research Branch, British Columbia Ministry of Forests, PO Box 335, Mesachie Lake, B.C., Canada V0R 2N0

Received January 8, 1997; accepted August 18, 1998

ABSTRACT

Variation among pedigrees in inbreeding depression (*ID*) can greatly reduce the value of close inbreeding as a breeding tool. However, estimation of *ID* for an individual pedigree is subject to various biases. These biases stem largely from the typical differences between outcrosses and inbreds in family coefficients of relationship, but can also arise from certain maternal effects. Consequently biases can arise in estimates of variation among pedigrees in *ID* and thence in tests for both the general presence of *ID* and individual cases of *ID*. The comparison of selfs and half-sib families is used as an illustration. A genetic model is presented, with associated forms of analysis of variance, to show theoretically how the biases arise and how they might be overcome. Three ways of addressing the biases, which can be used in combination, are presented and discussed: adjusting self- and/or outcross-family effects for differential additive genetic coefficients of relationship, use of analysis of covariance to adjust for putative maternal effects, and estimation of genetic correlations between self- and outcross performance. Each approach can help, but all have their limitations. These limitations arise largely from the complex nature of genetic parameters that are specifically associated with *ID*, which tends to entail far more parameters than can be satisfactorily estimated. Possibilities of more generalised forms of statistical analysis are also reviewed, with their potential advantages and limitations.

Key words: Inbreeding depression, self-fertilisation, genetic correlation, maternal effects, statistical analysis

INTRODUCTION

Inbreeding depression (*ID*) represents the inferiority of inbred material compared with outcrossed material of the same additive genetic value (breeding value), for the trait concerned. This inferiority is usually expressed as a proportion (or percentage):

$$ID = O - I/O = 1 - I/O \quad [1]$$

where I and O are the means for inbred and outcross material respectively.

ID is a very widespread phenomenon, which is typically severe in forest trees, particularly for seed viability and growth-rate traits (WILLIAMS & SAVOLAINEN 1996). It poses major constraints for the tree breeder. At the worst, *ID* can reduce the production of viable offspring to below acceptable numbers. Even without such viability losses, *ID* for growth rate, site tolerances or resistance to pests or diseases may vitiate genetic gains within a commercial crop. Within a breeding population *ID* can severely reduce the efficiency of selection, if the candidates show varying levels of inbreeding and/or varying responses to a given level of inbreeding (*e.g.* HARDNER & POTTS 1995). Variation among parents in susceptibility to *ID* can nullify both the theoretical advantages of self-fertilisation for evaluation of candidates and the potential advantages of close inbreeding for amplifying the expression of additive genetic variation.

In fact, *ID* typically shows obvious variation among pedigrees for a given level of inbreeding (*e.g.* ANDERSSON *et al.* 1974; WILCOX 1983; GEBUREK 1986; LUNDQVIST *et al.* 1987; GRIFFIN & COTTERILL 1988; SKRØPPA 1996), evidently being a feature of individual parents. It is desirable to quantify this within-population variation and to estimate *ID* at the level of the individual family, to help ascertain the errors of estimating the population mean for *ID*. Also of interest are the estimation errors and statistical significance of *ID* in the individual family. Of special value is a knowledge of the reliability of inbred performance as a guide to breeding value. At a more fundamental level, the variation in *ID* may afford clues to its underlying causes, which are still open to some debate (WILLIAMS & SAVOLAINEN 1996).

Estimating *ID*, however, can be surprisingly difficult. Even for the population (Equation 1) it can be hard to obtain unbiased comparisons of inbred and outcross material of equivalent breeding value (WILLIAMS & SAVOLAINEN 1996), largely through varying success in producing seed from closely related matings. Estimating *ID* for the individual family and making associated tests of statistical significance is not straightforward, even in principle. This is because the coefficients of additive genetic relationship are typically not the same for inbred and outcross families (cf FALCONER 1981; BURROWS & ASKEW 1982). This factor does not appear to have been taken into account by WILCOX (1983), either in calculating *ID* effects for individual parents or in testing for *ID* effects specific to individual parents (parent × mating type interaction). SKRØPPA (1996), while specifying the model correctly, described a calculation of individual-parent *ID* that apparently did not take account of the factor either. The complication resulting from the different coefficients of relationship can become much worse if certain epigenetic effects (*e.g.* maternal effects) occur. A further complication can arise because within-family variances can differ between self- and outcross material, although the nuisance value of this for statistical analysis may be less significant than what the behaviour of these variances, relative to straightforward expectations, tells about the genetic mechanisms of *ID*.

We develop a genetic model incorporating not only *ID* but also maternal effects, which represent a complication not addressed by BURROWS & ASKEW (1982). This is done for the specific comparison of a set of self-families with their corresponding maternal half-sib families, which we see as a case of special interest (despite the contention of BURROWS & ASKEW (1982)), but we give pointers to a more generalised treatment. The model is used to illustrate the principles of estimating *ID* for the individual family, and the problems that can arise from maternal effects. The model then is applied to develop analysis of variance (ANOVA), primarily for testing the significance of: mean *ID* in the population, seed-parent × inbreeding-level interaction, and *ID* shown by individual parents in their self families. We then outline the use of analysis of covariance (ANCOVA) to adjust for starting-size differences as putative maternal effects. As a different but complementary approach, we present the estimation of genetic correlations between self- and outcross performance, based on separate ANOVAs of self- and half-sib data, to give direct estimates of the reliability of self-family data for estimating parental breeding values. The advantages and limitations of these procedures are discussed. Possibilities of more generalised data analysis are reviewed, also with their advantages and limitations.

Probable causes of *ID* are considered here only insofar as they are likely to affect the properties of the data.

Practical illustration of our procedures is given by RUSSELL *et al.* (in prep.).

THEORETICAL BASIS FOR ANALYSING *ID* DATA

Basic conditions

For comparing inbred and outcross material we specify that the two categories should have essentially the same mean breeding value, for both male and female parentage. For female parentage this can be assured by using the same seed parents for both inbreeding and outcrossing. More difficult may to be to ensure that the pollen parents represent a sample of the same mean breeding value.

We specify two classes of families, selfs and female half-sibs, (noting that open-pollinated progenies may not equate closely to half-sib families, see *e.g.* HARDNER & POTTS 1995) assuming: (i) that the seed parents are a random sample of a large panmictic population and therefore have zero inbreeding (*i.e.* F = 0), F being the inbreeding statistic which represents the probability of any pair of alleles within the individual being identical by descent, (ii) that the half-sib families represent a large random sample of pollens, whose average breeding value therefore equals the population mean.

Also assumed are f seed parents with n individuals in each seed-parent/mating-type family, giving a total of $2nf$ individuals, in a fully random layout.

Implicit in the development of the model and the expected structures of variances and covariances among relatives is a large panmictic base population as the norm (cf COCKERHAM & WEIR 1984). Thus the coefficients of relationship correspond to the widely used Malécot coefficients only for the half-sib families. This simplifies the theory, in making the variance components more directly commensurable between the two classes of family, and is further justified on the grounds that very often $F \approx 0$ in the trees that reproduce (BUSH & SMOUSE 1992). Even if the assumption F = 0 in the parents is violated it generates expectations with which observed statistics may be compared.

Linear model expectations

We assume random effects unless otherwise stated. For the ***half-sib families*** (outcrosses) we can express the individual phenotype (P_o) as

$$P_o = \mu_o + f_{io} + w_{iol} \quad [2]$$

where μ_o = population mean under outcrossing, f_{io} = effect of the i^{th} seed parent under outcrossing, w_{iol} = effect of the l^{th} individual within the i^{th} family under outcrossing which is a composite of genotypic, microenvironmental and individual-seed maternal effects.

This formulation may be expanded, to include two additional classes of maternal effect, as

$$P_o = \mu_o + g_i + M_i + m_{io} + w_{iol} \quad [3]$$

where g_i = genotypic effect of the family, which for half-sibs corresponds to half the breeding value (as an effect) of the seed parent (plus some small fraction of the various orders of additive x additive epistasis which we assume can be ignored safely under panmixis), M_i = general maternal effect of the i^{th} parent, m_{ik} = specific maternal effect associated with the combination of the i^{th} parent and the k^{th} treatment, *i.e.* mating type, k = selfing (s) or outcrossing (o), w_{iol} = effect of the l^{th} individual within the i^{th} outcross family.

For the ***self-families*** we can express the individual phenotype (P_s) as (cf Equation 2)

$$P_s = \mu_s + f_{is} + w_{isl} \quad [4]$$

where μ_s = overall mean of self-family material, f_{is} = effect of the ith parent under selfing, w_{is} = effect of the l^{th} individual within the i^{th} family under selfing or, alternatively,

$$P_s = \mu_o - \Delta + f_{is} + w_{isl} \quad [5]$$

where Δ = mean *ID* effect (fixed), as an absolute value, not a proportion.

Equation 5 may be expanded, as in the expansion of Equation 2 to Equation 3, to

$$P_s = \mu_o - \Delta + 2g_i - d_i + M_i + m_{is} + w_{isl} \quad [6]$$

where d_i = *ID* effect specific to the i^{th} parent as expressed in the self-family with a mean expectation about Δ of zero, such that the total *ID* for the self-family is given by $\Delta + d_i$.

Thus *ID* for the ith self-family can be expressed as a proportion (cf Equation 1) as

$$ID_i = 1 - (\mu_o + g_i - (\Delta + d_i))/(\mu_o + g_i) \quad [7]$$

Such *ID* values for the various families will not average exactly to the population *ID* (Equation 1) because of the inherent behaviour of proportions (JOHNSTON & SCHOEN 1994).

The following may be noted:

(i) For outcrossing (Equations 2 and 3), accepting it as a norm, Δ and d_i are all treated as null effects, *i.e.* not present;

(ii) the sign convention used assigns positive values to *ID*, hence "$-\Delta$" and "$-d_i$" in Equations 5 and 6;

(iii) with outcrossing as specified, g_i, M_i and m_{io} are fully confounded;

(iv) with selfing the coefficient of 2 for g_i (Equation 6) relates to the fact that the full additive genetic effect, *i.e.* twice the general combining ability, contributes to the family effect, in contrast to the half-sib case;

(v) with selfing, $2g_i$, M_i and m_{is} are also confounded;

(vi) M_i is likely to represent a general maternal effect, *e.g.* of seed size or germination rate;

(vii) m_{ik} (m_{io} or m_{is}) is likely to represent, *inter alia*, an effect of the crossing-bag environment on seed size or quality (cf effects on seed set reported by ANDERSSON *et al.* 1974), which is likely to be reduced by replicating the crosses. While little understood, it potentially causes significant 'noise' variation;

(viii) with lesser degrees of inbreeding we would need to consider the effect of the pollen parentage g_j ($j \neq i$), which will have an additive genetic covariance with g_i (cf BURROWS & ASKEW 1982). For half-sib families, the expectation of g_j by definition equals zero. For related matings in general, in the absence of an actual estimate of g_j, the appropriate coefficient of g (U) is given by

$$U = (1 + r_{ij}) \quad [8]$$

r_{ij} being the additive genetic correlation between i and j, which with respect to the large panmictic population equals unity for self-families ($i = j$) and zero for half-sib families.

For accommodating ***both mating types simultaneously***, we can draw up the provisional model

$$P = \mu + f_i + \Delta_k + d_{ik} + w_{ikl} \quad [9]$$

where P is the individual phenotype, μ = overall mean for the two mating types = $(\mu_s + \mu_o)/2$, f_i = random seed-parent effect, across both mating types, which in terms of Equations 3 and 6 equals $1.5\, g_i + M_i + (m_{io} + m_{is})/2$, Δ_k = fixed effect of mating type ($\mu = \mu_o - \Delta/2$, $\Delta_s = -\Delta_o = \Delta/2$), d_{ik} = random seed-parent × mating type interaction effect ($d_{is} = -d_{io} = \frac{1}{2}\, d_i$) which, as is evident from considering Equations 3 and 6 (and see later), is not a straightforward genotype × mating type interaction, w_{ikl} = effect of individual phenotype within a seed-parent/mating type subclass.

This treatment, it may be noted, departs from the

original convention that panmixis is the norm, taking instead an intermediate level of inbreeding as the baseline.

Variance structures

In terms of variance components, assuming complete individual randomisation of layout and no additive × additive epistasis, we have the expectations for ***half-sib families:***

$$\sigma^2_{Po} = \sigma^2_{fo} + \sigma^2_{wo} \quad [10]$$

$$\sigma^2_{Po} = \sigma^2_g + \sigma^2_M + \sigma^2_m + \sigma^2_{wo} \quad [11]$$

and for ***self-families*** :

$$\sigma^2_{Ps} = \sigma^2_{fs} + \sigma^2_{wo} \quad [12]$$

$$\sigma^2_{Ps} = 4\sigma^2_g + Q + \sigma^2_M + \sigma^2_m + \sigma^2_{ws} \quad [13]$$

where σ^2_{Po} and σ^2_{Ps} are the phenotypic variances for outcrosses and selfs respectively, σ^2_{fo} and σ^2_{fs} are variances among half-sib and self-families respectively σ^2_g = general combining ability variance, which equals one-quarter the additive genetic variance (σ^2_A), Q = a composite genotypic variance, resulting mainly from inbreeding (cf COCKERHAM 1983; COCKERHAM & WEIR 1984), the nature of which is discussed later, which can be large and contribute strongly to differential *ID*, σ^2_M = general maternal effects variance, σ^2_m = specific maternal effects variance (assumed here to be the same for both mating types), σ^2_{wo} and σ^2_{ws} are the variances among individuals within the half-sib and self-families respectively, the coefficients for σ^2_g (Equations 11 and 13) being special cases of $(1 + r_{ij})^2$ (cf Equation 8).

The expectation for σ^2_{wo} is given by

$$\sigma^2_{wo} = 3\sigma^2_g + \sigma^2_D + \sigma^2_{eo} \quad [14]$$

and for σ^2_{ws},

$$\sigma^2_{ws} = 2\sigma^2_g + R + \tfrac{1}{2}\sigma^2_D + \sigma^2_{es} \quad [15]$$

where σ^2_D = allelic dominance variance as expressed under panmixis, *R* = another composite genotypic variance component, of poorly specified expectation, which can also be large, $\sigma^2_{eo}, \sigma^2_{es}$ = variances due to environmental effects specific to the individual, within the respective classes of family, which may include a further class of maternal effects, *e.g.*, carry-over from individual seed weights or germination times. Despite the lower coefficient of $\sigma^2_A, \sigma^2_{es}$ can often exceed σ^2_{eo} (*e.g.* WILCOX 1983; LUNDQVIST *et al.* 1987: MATHESON *et al.* 1995; HARDNER & POTTS 1995: SKRØPPA 1996).

If, contrary to the initial assumption, F > 0 in the parents through significant inbreeding and/or local differentiation of the population, the appropriate coefficients for σ^2_g and σ^2_D would be increased in Equations 11 and 13, but decreased in Equations 14 and 15.

Addressing ***both mating types simultaneously*** an expected composition of individual phenotypic variance follows from Equation 9 in a manner analogous to Equation 10 and 12 following from Equations 2 and 4 respectively. We thus have

$$\sigma^2_f = 9/4\sigma^2_g + Q/2 + \sigma^2_m + \tfrac{1}{2}\sigma^2_m \quad [16]$$

$$\sigma^2_w = \tfrac{1}{2}(\sigma^2_{wo} + \sigma^2_{ws}) \quad [17]$$

Estimation of effects

For estimating effects, based on seed-parent/mating-type family means, it is now straightforward to estimate *D*, g_i and d_i, applying Equations 3 and 6 (or Equation 9), provided *M* and *m* can be disregarded. Thus

$$\Delta = Y_{.o.} - Y_{.s.} \quad [18]$$

$$\hat{g}_i (= \hat{f}_{io}) = Y_{io.} - Y_{.o.} \quad [19]$$

$$\hat{d}_i = 2Y_{io.} - 2Y_{.o.} - Y_{.s.} + Y_{is.} \quad [20]$$

so the total *ID* for the ith family is given by

$$\Delta + \hat{d}_i = 2Y_{io.} - Y_{.o.} - Y_{is.} \quad [21]$$

where $Y_{io.}$, $Y_{is.}$ = means for the half-sib and self-families respectively of the i^{th} parent, $Y_{.o.}$, $Y_{.s.}$ = overall means of the half-sib and self-family material respectively, (Y... being the grand mean).

However, the occurrence of either or both of *M* and *m* (Equations 3, 6) can clearly lead to considerable bias in estimating g_i and d_i, and thence in ID_i (Equation 7), unless these maternal effects can be effectively removed (see later).

Analysis of variance models

Analysing data from each type of family separately and assuming balanced classifications, we have the ANOVA (Table 1). The variances can be designated according to the self-family or half-sib family cases (cf Equations 10, 12), *e.g.* σ^2_{fo}, σ^2_{fs}, etc. The significance tests are self-evident from Table 1.

Extending the analysis to embrace both mating types, applying the model represented in Equation 9, we in principle have the ANOVA in Table 2. The main complication in this analysis arises in deriving an appropriate sum of squares (SS) in order to test for both this effect and an overall difference between the mating types. The differential additive genetic contributions to self- and half-sib family variances respectively (Equations 10 and 12) will automatically tend to generate spurious interactions if SS are calculated straightforwardly (see Appendix 1). The following gives a more appropriate F × T interaction sum of squares (SS_{ft}), assuming σ^2_M and σ^2_m are zero:

$$SS_{ft} = \frac{n \sum_i [2^2(Y_{io.} - Y_{.o.})^2 + (Y_{is.} - Y_{.s.})^2 - 4(Y_{io.} - Y_{.o.})(Y_{is.} - Y_{.s.})]}{5} - (f-1)\ 0.3\ (\hat{\sigma}^2_{wo} - \hat{\sigma}^2_{ws}) \quad [22]$$

the second right hand side (r.h.s.) term being a correction for bias resulting from $\sigma^2_{wo} \neq \sigma^2_{ws}$ (see Appendix 2a).

A proposed alternative is:

$$SS_{ft} = n \sum_i [(\mu_o + 1.5\hat{f}_{io}) - (\mu_s + 0.75\hat{f}_{is})]^2 - \frac{f-1}{2} [(1.5^2 - 1)\hat{\sigma}^2_{wo} + (0.75^2 - 1)\ \hat{\sigma}^2_{ws}] - SS_t \quad [23]$$

where SS_t = mating type sum of squares, the second r.h.s. term being a correction for bias (see Appendix 2b), henceforth denoted *CB*, with the coefficients 1.5 and 0.75 both giving a coefficient of 1.5 for g_i (cf Equations 3, 6, 9). Equation 23 may be expanded to:

$$SS_{ft} = n \sum_i [[(Y_{.o.} + 1.5\ (Y_{io.} - Y_{.o.}) - Y_{i..} + Y_{...}]^2 + [(Y_{.s.} + 0.75\ (Y_{is.} - Y_{.s.}) - Y_{i..} + Y_{...})]^2] - CB \quad [24]$$

This formulation, while easily simplified, reflects the derivation. An alternative formulation is:

$$SS_{ft} = n \sum_i [(Y_{io.} + 1.5(Y_{oi.} - Y_{.o.})) - (Y_{.s.} + 0.75(Y_{si.} - Y_{.s.}))]^2 - CB \quad [25]$$

Given unbiased *SS* for F × T interaction, a proper (if conservative) test can be made for *ID* overall. However, a crude test for *ID*, which is essentially unbiased by the differential representation of additive gene effects in the two types of family, is a χ^2 test of H_o: $N(f_s > f_o) = N(f_s < f_o)$, namely whether self-families are either more or, less often than not, inferior to half-sib families.

We have noted that σ^2_w will effectively be an average of σ^2_{wo} and σ^2_{ws} (Equations 10 and 12), and how a difference between these values may bias tests and estimates of the variances unless corrections are made. In this connection, σ^2_{ws} may well be highly heterogeneous among self-families.

Testing of individual cases of ID

It may be of interest to test specific cases of *ID*, especially possible cases of negative *ID* or 'outbreeding depression.' There have been various reports, cited by JOHNSTON & SCHOEN (1994), of individual self-families outperforming their outcross counterparts. However, there appears to have been limited critical examination of these cases. True negative *ID* in our scenario entails $-(\Delta + d_i) > 0$ (cf Equation 7), not just $Y_{is.} > Y_{io..}$ In effect, this means that one must allow for the fact that with selfing a superior parent the pollen will likewise be of superior breeding value, but in a half-sib cross the pollen will be the population average.

A *t*-test can be made for whether $\Delta + \hat{d}_i$ differs significantly from zero, assuming $\sigma^2_m = 0$. The standard error, $SE(\Delta + \hat{d}_i)$, against which $\Delta + \hat{d}_i$ must be tested, is given by

$$SE(\Delta + \hat{d}_i) = [\frac{1.5^2 MS_E + 0.75^2 MS_E}{n}]^{0.5} \quad [26]$$

where MS_E = appropriate error mean square for the mean of an individual family (Table 2), and the coefficients of 1.5 and 0.75 correspond to the weightings in the pairwise contrasts involved in Equation 23. The degrees of freedom for *t* will be $2f(n-1)$. Each d_i can be tested on the same basis. However, if there is strong heterogeneity of within-family variances arising, say, in σ^2_{ws}, it may be appropriate to modify Equation 26 and the degrees of freedom. However, it seems unlikely that very large σ^2_{ws} in a particular self-family would be associated with negative *ID*.

The multiplicity of pairwise comparisons argues for setting an experimentwise Type 1 error rate. This entails adjusting the comparisonwise rate to αE, appropriate to the experimentwise Type I error rate α. Thus with α = 0.05 and 20 pairwise comparisons we have αE = 0.05/20, *i.e.*, *t* (p = 0.0025) for each comparison between the outcross and the inbred offspring of a single parent. This experimentwise test may be overstringent for 'outbreeding depression', since only in a few cases *ID* may be around zero, let alone negative. No such question, however, attaches to experimentwise α levels in testing for individual cases of 'positive' *ID*.

Table 1. Form of analysis of variance for a single mating type.

Item	df	Expected mean squares (MS_F, etc.)
Families (F)	f – 1	$\sigma_w^2 + n\sigma_f^2$
Within families	(n – 1) f	σ_w^2
Total	nf – 1	

where σ_f^2 = among-family variance and σ_w^2 = within-family variance

Table 2. Form of analysis of variance for data involving both mating types.

Item	df	Expected mean squares (MS_F, etc.)
Seed parents (F)	$f - 1$	$\sigma_w^2 + 2n\sigma_f^2$
Mating type (T)	*1*	$\sigma_w^2 + n\sigma_{ft}^2 + fn\Phi_t^2$
F × T (FT)	$f - 1$	$\sigma_w^2 + n\sigma_{ft}^2$
Within-subclass	$(n - 1)f$	σ_w^2

where σ_{ft}^2 = F × T interaction variance and Φ^2_t = fixed-effect mating type 'variance'

Analysis of covariance (ANCOVA)

If a variable, say, height *ex* nursery, is deemed to reflect maternal effects (*M* or *m*, Equations 3, 6) it may be used as a covariate in ANCOVA. Thus, for variables like subsequent heights, the sums of squares (Tables 1, 2) can be adjusted for individual covariance on nursery height. In principle, that should adjust for the maternal effects, but with any direct positive genotypic correlation between nursery and subsequent heights there will tend to be some over-adjustment. Nevertheless, if nursery height gives a good measure of effective starting size, covariance-adjusted family means may lead to essentially unbiased estimates of inbred-outcross genetic correlations (see later).

When covariance adjustment is indicated, alternative approaches to ANCOVA may be considered:

(i) Adjusting the dependent-variable (*Y*) SS for the covariate (*X*) and the sum of cross-products (CP) (applying Equation 22 or 25 for interaction), using these in conjunction with the corresponding within-subclass SS and CP to derive covariate-adjusted SS for the dependent variable, as follows

Item	Covariance-adjusted SS
Hypothesis + error	SS_Y – Regression *SS* (= *H* + *E*)
Individual error	SS_Y – Regression *SS* (= E)
Hypothesis (*i.e.* F × T interaction)	(*H* + *E*) – *E*

(ii) Adjusting all individual values of the Y variable for departures from the overall covariate mean. This will tend to give tests that are under-stringent, having no built-in allowance for errors of estimating the within-subclass regression. This approach, however, while not usually preferred, may be easier to implement and more robust with respect to imbalance, particularly if the within-subclass regression is well estimated.

Genetic correlation analysis

An alternative to studying possible F × T interaction as such is to estimate the genetic correlation (r_g) between performance under the two mating types, interaction representing departures from a value of +1. This gives a very direct measure of how well self-family performance reflects the parents' breeding values. The basic approach is straightforward, being analogous to that for 'Type B' genetic correlations, between performance in different environments (BURDON 1977). Hence, with zero non-genetic covariance between self- and half-sib family means,

$$r_g = r_{Pso}/(h_{fs} \times h_{fo}) \qquad [27]$$

where r_{Pso} is the correlation of self- and half-sib family means and h^2_{fo} and h^2_{fs} are the repeatabilities of self- and half-sib family means respectively. Alternatively,

$$r_g = cov_{Pso}/(\sigma_{fo} \times \sigma_{fs}) \qquad [28]$$

where cov_{Pso} is the covariance between self- and half-sib family means, which can be easily estimated as the mean cross-product of family means.

Several authors (*e.g.* WILCOX 1983; GRIFFIN & COTTERILL 1988; SKRØPPA 1996), while presenting r_{Pso}, did not calculate r_g.

That the additive genetic effect is expressed in differing degrees between the two types of family (Equations 3, 6) does not in itself create bias in estimating the genetic correlation, since within each mating type the family effect contains the same fraction of the additive genetic effect for the seed parent. The statistical significance of r_g is that of r_{Pso}, which is straightforward. Any test for F × T interaction, such as would tend to cause either rank changes or departures from linearity, is effectively a test for departures from $r_g = 1$ (assuming that $r_g < 0$ is unlikely). A test was proposed by BURDON (1977) and, not knowing of a better one, and in view of past typographical problems (*op. cit.*), we present it (Appendix 3).

The genetic correlation estimates will be biased upwards by general maternal effects (*M*) and downwards by specific maternal effects (*m*) (Equations 3, 6). Such effects, however, may sometimes be addressed appropriately by analysis of covariance (see earlier).The value of r_g is of course specific to the level of inbreeding represented in selfing.

SUGGESTED PLAN FOR ANALYSING SUCH DATA SETS USING ANOVA / ANCOVA

1. Preliminary study of data properties: possible non-normality (need for transformation?), heterogeneity of within-subclass variances, existence of within-subclass correlations.
2. Separate main analyses for selfs and outcrosses: means/effects (Equations 18–21), ANOVAs Table 1), and variance-component estimation.
3. Analysis of combined data for both mating types: ANOVA (Table 2), using Equations 22 or 23–25 for appropriate T × F *SS* (SS_{ft}), also testing individual cases of *ID* (Equation 26).
4. If ANCOVA indicated (from Step 1), repeat Steps 2 and 3, with individual covariance adjustment(*s*).
5. Estimate genetic correlation between self-family and half-sib family performance (Equation 27 or 28), using results from Step 2 and any repetition thereof in Step 4.

DISCUSSION

We have used a stringent criterion of what constitutes an individual case of negative *ID*, in that we exclude from it favourable additive gene effects that are potentially expressed fully in self-family means. This contrasts with the convention of SCHULTZ & WILLIS (1995, Table 1). However, it is still possible that additive × additive epistasis of various orders could generate individual cases of genuine negative *ID* (cf. COCKERHAM 1983), given the expectation that such effects can be strongly expressed under severe inbreeding.

With only two levels of F, namely 0 and 0.5, we have not had to consider any general relationship of *ID* to F. It is widely assumed (*e.g.* BURROWS & ASKEW 1982; DE BOER & HOESCHELE 1993) that there is generally a linear relationship, primarily for cross-referencing observed *ID* between populations that differ in F. Non-linearity of *ID* in relation to F, however, can arise in several ways: through epistasis increasing *ID* more than in proportion to F (CROW & KIMURA 1970); through multiplicative effects of genes having the opposite effect, particularly when initial *ID* is severe; and through effective purging of deleterious alleles producing a relationship like that for multiplicative effects (in effect through true F becoming progressively less than nominal F). Thus, even if a global linearity exists, compensating non-linearities could still exist among the various effects that are operating (WILLIAMS & SAVOLAINEN 1996).

Another consequence of using only half-sib and self-families is that the issue of discriminating between *ID* and any specific combining ability that arises under outcrossing (BURROWS & ASKEW 1982) does not arise.

We have shown how maternal effects, unless they can be eliminated, can generate major bias in estimating effects and variances in this situation. The value of the proposed use of ANCOVA will depend on having a covariate that is a good measure of any maternal effect. This may not be assured, which calls for cautious interpretation, particularly if there are complications such as variances within self-families being very heterogeneous. Analysis of covariance could also be applied to quantify better the timing of the expression of *ID*, which is of much biological interest (HUSBAND & SCHEMSKE 1996; RUSSELL *et al.* in prep.). In principle, the value for a trait at any preceding age could be used as a covariate in testing for whether *ID* has been expressed in the intervening period (cf JOHNSTON & SCHOEN 1994). Again, however, such analyses will need to be used and interpreted with caution.

The genetic correlation approach (Equation 28), apart from giving a direct indication of the inherent reliability of self-family information as a guide to parental breeding values, can be very flexible. It readily acccommodates differences in within-family variances, and even differences in experimental layout parameters, between the selfs and outcrosses respectively. Indeed, if the average *ID* is already known adequately, leaving

r_g as the item of prime interest, there is no need to have the selfs and the outcrosses together in the same statistical layout. However, if the two classes of family are planted on quite different sites r_g would be depressed by any genotype-site interaction (cf BURDON 1977).

If the genetic correlation is the parameter of prime interest, it can be expressed as a function of variance components that can be estimated in joint analysis of data from the two treatments (mating systems), analogous to the analysis of genotype-environment interaction specified by ROBERTSON (1959) (see also DICKERSON 1962: COCKERHAM 1963). However, unless there is major imbalance in the classification the approach outlined here, beginning with separate analyses of data from the two particular mating systems, would seem more straightforward and robust.

The analyses of variance that we have developed apply to a specific situation, but they serve to illustrate most of our points. While some extensions, *e.g.* accommodating a block layout and modest imbalance in the classification, are straightforward and are at least good approximations, there are several respects in which it would be desirable to achieve much greater generality in the data analysis. These include accommodating: complex imbalance in the classification (which may even allow one to relax the requirement for inbred and outcross material to be of equivalent mean breeding value); a virtually continuous range of F values; complex inter-relationships among various families; the estimation of a number of genotypic variances that contribute to *ID*; variations among different classes of family in within-family variance: and the presence of covariates for which adjustments may be needed. In principle, this can be achieved within the framework of Restricted Maximum Likelihood (REML) analysis, to arrive at mixed-model Best Linear Unbiased Predictor (BLUP) solutions. As set out by DE BOER & HOESCHELE (1993) (see also HARDNER *et al.* 1996) this can be addressed by using a linear weighting, in the ratio (1 - F) : F, of the genetic parameters assumed for a large panmictic population and those assumed for a population inbred to complete homozygosity. However, as discussed earlier, the assumption of linearity may not be satisfactory.

Such an approach, apart from being technically demanding, may depend on certain approximations being satisfactory unless the data set is quite small (HARDNER *et al.* 1996). However, for estimating the detailed genetic parameters, in order to elucidate the genetic mechanisms of *ID*, there may be no very satisfactory solution. Several categories of genotypic variance, which are likely to be unimportant (or can be harmlessly mis-specified) under panmixis, can become very important under inbreeding (*e.g.* COCKERHAM & WEIR 1984). These are denoted collectively in Q and R in our Equations 13 and 15 respectively. They include: various classes of epistasis, covariance between additive and dominance effects, and a distinct class of dominance effects (offset in some degree by a drop in the conventional dominance that is expressed under panmixis). Incorporating all these items into a single genetic model becomes "unmanageable" (COCKERHAM & WEIR 1984), in that all their respective roles can only be assigned very imprecisely or not at all. While a good statistical fit may be obtained using a manageable model, it may bear little relationship to the real mechanisms. Use of clonal material may in principle help to discriminate between some alternative models, but may do so only in restricted situations which would include minimal epigenetic effects (*e.g.* 'c-effects' or 'C-effects'; LIBBY 1962; LIBBY & JUND 1962; BURDON *et al.* 1992).

That the genetic model can become more complex under inbreeding is well illustrated by strong *ID* occurring when outcrosses show negligible Specific Combining Ability (SCA) (SKRØPPA, 1996). It is also illustrated by phenotypic variation often being greater within self-families than within outcross families (*e.g.* WILCOX 1983: LUNDQVIST *et al.* 1987; MATHESON *et al.* 1995; HARDNER & POTTS 1995; SKRØPPA 1996), although an additional cause of the latter could be a loss of developmental homoeostasis upon inbreeding ($\sigma^2_{es} < \sigma^2_{eo}$) (See LYNCH 1988 for review). Whatever the exact genetic mechanism, the underlying precondition for *ID* in the population as a whole is directional dominance of favourable alleles (CROW & KIMURA 1970). In outbreeding populations, with genetic load in the form of various deleterious recessive alleles, such alleles will each tend to be present at very low frequencies, even though they may involve numerous loci over the large population. They thus contribute essentally nothing to the genetic variation that is expressed under panmixis, but may be very strongly expressed under inbreeding, and some can even cause a breakdown of the infinitesimal (polygenic) model.

While a generalised REML-based solution may often be indicated as the definitive data analysis, piecemeal exploratory analysis of the data may still seems advisable, if only in order to arrive at the most appropriate REML model.

ACKNOWLEDGEMENTS

We thank Dr N. Borralho, Prof. W. Libby and Dr C. Raymond for helpful comments on the draft, Dr H. X. Wu for leading us towards Equation 22, and anonymous referees for further helpful suggestions.

REFERENCES

ANDERSSON, E., JANSSON, R. & LINDGREN, D. 1974: Some results from second generation crossings involving inbreeding in Norway spruce. *Silvae Genet.* **23**: 34–43.

BURDON, R. D. 1977: Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding. *Silvae Genetica* **26:** 168–175.

BURDON, R. D., BANNISTER, M. H. & LOW, C. B. 1992. Genetic survey of *Pinus radiata.* 4: Variance structures and heritabilities in juvenile clones. *New Zeal. J. For. Sci.* **22**: 187–210.

BURROWS, P. M. & ASKEW, G. R. 1982: Experimental inbreeding depression. *For. Sci.* **28**: 148–156.

BUSH, R. M. & SMOUSE, P. E. 1992: Evidence for the adaptive significance of allozymes in forest trees. *New Forests* **6:** 179–196.

COCKERHAM, C. C. 1963: Estimation of genetic variances. *In* Statistical Genetics and Plant Breeding (ed. Hanson, W. D. & Robinson, H. F.). Pp.53–94. NAS–NRC, Washington, DC.

COCKERHAM, C. C. 1983: Covariances of relatives from self-fertilization. *Crop. Sci.* **23**: 1177–1180.

COCKERHAM, C. C. & WEIR, B. S. 1984: Covariances of relatives stemming from a population undergoing mixed self and random mating. *Biometrics* **40**: 157–164.

DE BOER, I. J. M. & HOESCHELE, I. 1993: Genetic evaluation methods for populations with dominance and inbreeding. *Theor. Appl. Genet.* **86**: 245–258.

DICKERSON, G. E. 1962: Implications of genetic-environmental interactions in animal breeding. *Anim. Prod.* **4**: 47–64.

FALCONER, D. S. 1981: Introduction to Quantitative Genetics, 2nd Ed. Longman.

GEBUREK, T. 1986: Some results of inbreeding depression in Serbian spruce (*Picea omorika* (Panc.) Purk.). *Silvae Genet.* **35**: 169–172.

GRIFFIN, A. R & COTTERILL, P. P. 1988: Genetic variation in growth of outcrossed, selfed and open-pollinated progenies of *Eucalyptus regnans* and some implications for breeding strategy. *Silvae Genet.* **37:** 124–131.

HARDNER, C. M. & POTTS, B. M. 1995: Inbreeding depression and changes in variation after selfing in *Eucalyptus globulus* ssp. *globulus. Silvae Genet.***44**: 46–54.

HARDNER, C., BORRALHO, N., TIER, B., MILLER, S. & GODDARD, M. 1996: Accounting for dominance and inbreeding in genetic evaluations using individual tree mixed models. *In* Tree Improvement for Sustainable Tropical Forestry, Dieters, M. J. *et al.* (eds). Pp. 143–147. Proc. QFRI–IUFRO Conf., Caloundra, Queensland, Australia, 27 Oct.–1 Nov. (Queensland Forestry Research Institute, Gympie).

HUSBAND, B. C. & SCHEMSKE, D. W. 1996: Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution* **50**: 54–70.

JOHNSTON, M. O. & SCHOEN, D. J. 1994: On the measurement of inbreeding depression. *Evolution* **48:**1735 –1741.

LUNDQVIST, K, ERIKSSON, G., NORELL, L. & EKBERG, I. 1987: Inbreeding depression in two field tests of young *Pinus sylvestris* (L.). *Scand. J. For. Res.* **2**: 281–290.

LIBBY, W. J. 1962: Estimation of variance components in a clonal population of *Mimulus guttatus. Genetics* **47**: 769–777.

LIBBY, W. J. & JUND, E. E. 1962: Variances associated with cloning. *Heredity* **17**: 533–540.

LYNCH, M. 1988: Design and analysis of experiments on random drift and inbreeding depression. *Genetics* **120:** 791–807.

MATHESON, A. C., WHITE, T. L. & POWELL, G. R. 1995: Effects of inbreeding on growth, form and rust resistance in *Pinus elliottii. Silvae Genet.* **44:** 37–46.

ROBERTSON, A. 1959: The sampling variance of the genetic correlation coefficient. *Biometrics* **15**: 469–485.

RUSSELL, J. H., BURDON, R. D. & YANCHUK, A. D. (in prep.): Inbreeding depression in self-families of western redcedar.

SCHULTZ, S. T. & WILLIS, J. H. 1995: Individual variation in inbreeding depression: The roles of inbreeding history and mutation. *Genetics* **141:** 1209–1223.

SKRØPPA, T. 1996: Diallel crosses in *Picea abies.* II. Performance and inbreeding depression of selfed families. *For. Genet.* **3**: 69–79.

WILCOX, M. D. 1983: Inbreeding depression and genetic variances estimated from self- and cross-pollinated families of *Pinus radiata. Silvae Genet.* **32**: 89–96.

WILLIAMS, C. D. & SAVOLAINEN, O. 1995: Inbreeding depression in conifers: Implications for breeding. *Forest Science.* **42**: 102–117.

Appendix 1

Demonstration of spurious interaction that can arise in Table 2 unless interaction SS is recalculated as in Equations 22 or 24.

Seed-parent × mating type interaction (Table 2) arises from violation of the condition

$$(f_{io} - f_{is}) - (f_{jo} - f_{js}) \equiv 0 \qquad [A1]$$

where i and j denote different seed parents.
From Equations 3 and 6 we can derive

$$(f_{io} - f_{is}) - (f_{jo} - f_{js}) = -(g_i - g_j) + (d_i - d_j) + [(m_{io} - m_{is}) - (m_{jo} - m_{js})] \qquad [A2]$$

Even if $d_i = d_j$ which is the central condition for zero genotype x mating type interaction, we still have several right hand side (r.h.s.) terms that will only fortuitously equal zero. For instance, $g_i \equiv g_j$ does not hold if there is any additive genetic variation. The final r.h.s. term will also almost certainly be non-zero if specific maternal effects exist.

Adjusting coefficients of d_o and d_s (cf Equation 23 or 25) to give equal coefficients of g_i and g_j in Equation A2, will not eliminate the other non-zero terms in Equation A2 and is liable to introduce a non-zero term involving $M_i - M_j$, although ANCOVA is directed at eliminating terms involving M and m.

Appendix 2

Derivations of corrections for bias (Equations 22–25)

a. Equation 22 (if $\sigma^2_{wo} \neq \sigma^2_{ws}$)

$$SS_{ft} = nf \frac{f-1}{2f} \left[2\frac{(4\sigma^2_{wo} + \sigma^2_{ws})}{5n} + 2\sigma^2_{ft}\right]$$

Compared with the requisite expectation that may be written as

$$SS_{ft} = nf \frac{f-1}{2f} \left[2\frac{(\sigma^2_{wo} + \sigma^2_{ws})}{2n} + 2\sigma^2_{ft}\right]$$

$(f - 1) / f$ being the appropriate finite population correction for variances.

The difference between the two expressions can be simplified to $(f - 1)\ 0.3\ (\sigma^2_{wo} = \sigma^2_{ws})$.

b. Equations 23–25

$$SS_{ft} = nf \frac{f-1}{2f} \left[2\frac{(1.5^2\sigma^2_{wo} + 0.75^2\sigma^2_{ws})}{n} + 2\sigma^2_{ft}\right]$$

compared with the requisite expectation that may be rewritten as

$$SS_{ft} = nf \frac{f-1}{2f} \left[\frac{(\sigma^2_{wo} + \sigma^2_{ws})}{n} + 2\sigma^2_{ft}\right]$$

The difference between the two can be simplified to

$$\frac{f-1}{2}[(1.5^2 - 1)\ \sigma^2_{wo} + (0.75^2 - 1)\ \sigma^2_{wo}]$$

While further simplification is obvious this formulation reflects the derivation.

Appendix 3

Test for seed-parent × mating type interaction in the form of departures from linear genetic correlation (r_g) of + 1 between mating types. (Adapted and typographically corrected from BURDON 1977).

Consider the regression (Y on X) of half-sib family means on self-family means, the designation of Y and X being somewhat arbitrary but adopted because self-family performance might be used to estimate additive breeding values. The test, assuming balanced classification, is of the ANOVA form (cf Table 1).

Mean squares	Df	Expected mean square
Lack of fit (H_0: $r_g = +1$)	f–2	$\sigma^2_{\varepsilon'Y} + \sigma^2_L$
Residual error	f(n–1)	$\sigma^2_{\varepsilon'Y}$

where lack of fit MS = error MS after fitting regression, σ^2_L = variance due to true lack of fit, $\sigma^2_{\varepsilon'Y}$ = component of variance of family means that is purely due to independent sampling errors in the two types of family mean.

Now $\sigma^2_{\varepsilon'Y}$ can be formulated as follows:

$$\sigma^2_{\varepsilon'Y} = \frac{\sigma^2_{eX}}{\sigma^2_X - \sigma^2_{\varepsilon X}} (\sigma^2_Y - \sigma^2_{Y.X}) + \sigma^2_{\varepsilon Y}$$

where σ^2_X = variance of self-family means (estimated as MS_{fs}/n), σ^2_Y = variance of half-sib family means (estimated as MS_{fo}/n), $\sigma^2_{\varepsilon X}$ = sampling error variance of self-family means (estimated as M_{sws}/n), $\sigma^2_{\varepsilon Y}$ = sampling error variance of half-sib family means (estimated as M_{swo}/n), $\sigma^2_{Y.X}$ = lack of fit MS =

$$(MS_{fo}/n)(1 - r^2_{Pso}) \frac{f-1}{f-2}$$

Hence

$$\sigma^2_{\varepsilon'Y} = \frac{MS_{ws}}{MS_{fs} - MS_{ws}} \left(\frac{MS_{fo}}{n}(1 - (1 - r^2_{Pso})\frac{f-1}{f-2})\right) + MS_{wo}/n$$

MS_{fs}, MS_{fo} = family mean squares for selfs and half-sibs respectively, MS_{ws}, MS_{wo} = corresponding within-family mean squares.

The power of this test depends on which variable (half-sib or self-family performance in this case) is designated Y, being greater when family differences are more strongly resolved in the Y variable. However, the ratio σ^2_L/σ^2_{fY}, which is a measure of the degree of lack of fit, is essentially independent of which variable is designated Y.