

Estimating quantitative genetic parameters in haplodiploid organisms

FU-HUA LIU & SANDY M. SMITH*

Faculty of Forestry, University of Toronto, 33 Willcocks Street, Toronto, Ontario, Canada, M5S 3B3

Haplodiploid organisms (diploid females and haploid males) are genetically distinct in that they have asymmetrical genetic segregation, gene dosage compensation in the haploid male, maternal effects, and a general propensity to inbreed. Taking these characteristics into account, we demonstrate how conventional experimental designs and statistics derived to measure quantitative genetic parameters for diploid organisms can be applied to haplodiploid ones. First, the *C* matrix of the joint-scaling test is modified to test the additive–dominance model and then the calculations for fractions in the *C* matrices (for males and females) are shown for an infinite number of generations with both random and brother–sister mating. Second, analytical adjustments are outlined for the derivation of covariance between relatives using populations at either Hardy–Weinberg equilibrium or those that prefer to inbreed. Finally, four of the conventional designs (sib-analysis, offspring–parental regression, North Carolina III (*NC III*), and diallel cross) are modified and then compared according to their ability to deal with the experimental conditions encountered in haplodiploid systems. Although all designs can be used with caution, our analysis suggests that the *NC III* design is the most broadly applicable because it consistently meets the theoretical assumptions. The parameters we derive here for inbred populations are recommended for analysing the typically skewed genotypic distributions found in many natural populations of haplodiploid organisms. The theoretical and applied aspects of our work as well as the difference between our work and that for X-linked genes are discussed.

Keywords: experimental design, genetic parameters, haplodiploid species.

Introduction

Seven experimental designs and their statistical analyses have been developed in conventional quantitative genetics to estimate inheritance parameters for continuously varying traits in diploid organisms (Mather & Jinks, 1982). Three have been transferred directly to haplodiploid systems including: sib-analysis for honeybees (Rinderer, 1977), offspring–parental regression for the wasp, *Nasonia vitripennis* (Orzack & Gladstone, 1994), and diallel cross for the wasp, *Muscidifurax raptor* (Antolin, 1992). Unfortunately, the statistics used in these analyses have not been shown to be valid for haplodiploid organisms and statistics for the remaining four approaches have not been derived. The lack of appropriate techniques for analysing genetic parameters in haplodiploid organisms limits our understanding about the evolutionary mechanisms which underlie their speciation as well as their poten-

tial use as beneficial species in applied pest management programmes.

Haplodiploid organisms are found in seven animal orders including rotifers (Monogonata), arachnids (Acarina), pinworms (Oxyurida), bees, ants, wasps and sawflies (Hymenoptera), thrips (Thysanoptera), white flies and scale insects (Homoptera), bark and ambrosia beetles (Coleoptera) (Adamson, 1989). Aside from being haplodiploid, these organisms often display a female-biased sex ratio (Hamilton, 1967), low heterozygosity (Graur, 1985), more than one reproductive mode (Pinto *et al.*, 1991), and some degrees of inbreeding (Hardy, 1994). Based on such characteristics, haplodiploid populations are unlikely to be in Hardy–Weinberg equilibrium, particularly as the offspring inherit parental genomes asymmetrically depending on their sex (a mother contributes to both sons and daughters, whereas a father contributes uniquely to daughters). Thus, we would expect that data collected from haplodiploid organisms cannot be analysed genetically using the conventional statistics developed for diploid ones.

*Correspondence. E-mail: smith@larva.forestry.utoronto.ca

Theoretically, haplodiploid systems operate under the same genetic principles as X-linked genes in diploid systems (Crozier, 1970). The genetic components for X-linked genes are the same as those for haplodiploid ones if only the sex chromosomes are considered, under the assumption of no crossover between the X and Y chromosomes. However, all previous work on X-linked genes have treated sex chromosomes and autosomes simultaneously with sex limitation (e.g. Mather & Jinks (1982) for a panmictic population, Killick (1971) for biparental progeny and North Carolina I, II, III and Bohidar (1964) for random mating). Thus, the formulas they derive for X-linked genes are much more complicated than those for haplodiploid ones, making it impossible to use them directly for the latter.

The present study examines the use of conventional genetic techniques for haplodiploid organisms. We first test the assumption of the additive–dominance model using a modified joint-scaling test, and then examine the effects of this breeding system on the calculation of covariance among relatives. Finally, formulas for the quantitative genetic components of four classic experimental designs (sib-analysis, offspring–parental regression, *NC III*, and diallel cross) are derived and the relative application and merits of each as well as the differences between our work and that for X-linked genes are discussed.

Additive–dominance model and epistasis

The additive–dominance model underlies quantitative genetic components estimated by conventional genetic experimental designs. Before a genetic component can be estimated, one must determine whether the data fit an additive–dominance model using a joint (Cavalli, 1952) or *ABC* (Mather & Jinks, 1982) scaling test. Mather & Jinks (1982) have modified the *ABC* test for X-linked genes and we will modify the joint test for haplodiploid organisms.

In a haplodiploid system, it is assumed that we start with two pure-breeding strains from which a female of genotype A_1A_1 and a male of genotype A_2 are crossed to produce the first filial generation (F_1), and that random mating occurs from the F_1 generation onwards. In developing the C matrix for the joint test, the female and male populations must be treated separately to avoid any problems arising from the difference in gene effects between sexes. Thus, for females, the expected C matrix of the joint test (with t rows and k columns where t refers to the t th segregating generation and k denotes vectors of the genotypic fractions) from the parental females to the F_5 generation will be:

$$C_f = \begin{pmatrix} F_0 & 1 & 1 & 0 \\ F_1 & 1 & 0 & 1 \\ F_2 & 1 & \frac{1}{2} & \frac{1}{2} \\ F_3 & 1 & \frac{1}{4} & \frac{1}{2} \\ F_4 & 1 & \frac{3}{8} & \frac{7}{16} \\ F_5 & 1 & \frac{5}{16} & \frac{29}{64} \end{pmatrix}, \quad (1)$$

where C_f is the C matrix for females. In eqn 1, the first column contains legends for successive nonoverlapping generations, the second column a fraction of the overall population mean for a trait in a generation (μ), the third, the additive fractions from the parental females to females in the fifth generation (a), and the fourth, the dominant fractions (d).

The expected C matrix for males will be simpler because there are no dominance fractions and this allows the allelic interactions from only a few segregating generations to be tested. For example, if a male with a lower population mean value for a trait is crossed with a homozygous female having a higher mean value and random mating in each of the subsequent nonoverlapping generations is allowed, the C matrix of a digenic model for the males will become:

$$C_m = \begin{pmatrix} F_0 & 1 & -1 & 1 \\ F_1 & 1 & 1 & 1 \\ F_2 & 1 & 0 & 0 \\ F_3 & 1 & \frac{1}{2} & \frac{1}{4} \\ F_4 & 1 & \frac{1}{4} & \frac{1}{16} \\ F_5 & 1 & \frac{3}{8} & \frac{9}{64} \end{pmatrix}, \quad (2)$$

where C_m refers to the C matrix for the male population. The columns are the same as for females except the last two contain only additive fractions (no dominance); the last being the fractions for additive–additive interaction.

Now, we need to incorporate observed values into our model to estimate the genetic parameters. Thus, J , a parameter vector with k rows, can be calculated by letting W be a weight vector in which each element is the reciprocal of a squared standard error around the observed generational mean and Y is the observed generational mean vector with t rows:

$$J = [(C\#W)'C]^{-1}[C(Y\#W)], \quad (3)$$

where C' is the transposed C matrix and $\#$ is the elementwise multiplication (SAS/IML, 1990). The size of vector J depends on the type of genetic model used; in an additive–dominance model, it is composed of three elements; namely an estimated population mean ($\hat{\mu}$), an overall genetic additive deviation (\hat{a}), and a dominant

deviation (\hat{d}) from $\hat{\mu}$ over n loci. If a digenic interaction model for females is needed, then \mathbf{J} would contain six elements.

The acceptability of the additive–dominance model can be tested using a $\chi^2_{(t-k)}$ distribution, while estimates of the \mathbf{J} vector can be tested with a $t_{(t-k)}$ distribution (Mather & Jinks, 1982). Thus,

$$\chi^2 = \sum_1^t \{[(\mathbf{E} - \mathbf{Y})\#\#2]\#\mathbf{W}\}, \tag{4}$$

where \mathbf{E} is the expected generational mean vector and $\#\#$ is the elementwise power (SAS/IML, 1990). As in diploid systems, rejection of the additive–dominance hypothesis here means that further analysis can only proceed if the data are transformed to fit the model.

Alternatively, if the data do not fit the above additive–dominance model, then epistasis is implied. Epistasis can be explored further using a modified \mathbf{C} matrix from eqn 3. In this case, the \mathbf{J} vector from eqn 3 would have six elements (μ, a, d, i, j, l) corresponding to the six numerical columns in \mathbf{C}_f from left to right; namely grand mean (μ), additive deviation (a), dominance deviation (d), additive–additive interaction ($i = a \cdot a$), additive–dominance interaction ($j = a \cdot d$), and dominance–dominance interaction ($l = d \cdot d$). These six parameters can be estimated simultaneously using eqn 3 and the significance for each estimate can be tested based on the inversed \mathbf{C}_f matrix.

The fractions in the above \mathbf{C} matrices for females and males are given only for first five generations. In order to increase the statistical power or to explore more than two genic interactions, more successive generations need to be observed. Thus, it is necessary to work out the recurrent formulas for the genotypic array. Based on Jennings’s (1916) work in the first four generations for a random mating system originating from a single cross between two pure-breeding strains, we obtain the recurrent formulas of genotypic array at generation t for a population of equal numbers of A_1A_1 and A_2A_2 as:

$$\text{male proportion of genotype } A_1 = \frac{2}{3} + \frac{1}{3} \left(-\frac{1}{2}\right)^t, \tag{5}$$

$$\text{male proportion of genotype } A_2 = \frac{1}{3} - \frac{1}{3} \left(-\frac{1}{2}\right)^t, \tag{6}$$

female proportion of genotype

$$A_1A_1 = \frac{4}{9} - \frac{2}{9} \left[\left(-\frac{1}{2}\right)^t + \left(-\frac{1}{2}\right)^{2t} \right], \tag{7}$$

female heterozygote A_1A_2

$$= \frac{4}{9} + \frac{1}{9} \left[3 \left(-\frac{1}{2}\right)^t + 2 \left(-\frac{1}{2}\right)^{2t} + \left(-\frac{1}{2}\right)^{t-1} - \left(-\frac{1}{2}\right)^{2t-1} \right], \text{ and } \tag{8}$$

female homozygote A_2A_2

$$= \frac{1}{9} - \frac{1}{9} \left[\left(-\frac{1}{2}\right)^t + \left(-\frac{1}{2}\right)^{t-1} - \left(-\frac{1}{2}\right)^{2t-1} \right]. \tag{9}$$

If the beginning population has equal numbers of A_2A_2 and A_1 , then the results will be the same as above, except A_1 and A_2 will be interchanged and A_1A_1 and A_2A_2 will be reversed.

In contrast to a random mating population from F_1 onwards, if only brothers and sisters mate beginning with A_1A_1 and A_2 , then the genotypic array formulas for A_1 and A_2 males will be the same as eqns 5 and 6. However, the formulas for females will be different:

$$A_1A_1 = \frac{2}{3} + \frac{1}{3} \left(-\frac{1}{2}\right)^{t+1} - \frac{1}{2^{2(t+1)}\sqrt{5}} \left[(1 + \sqrt{5})^{t+1} - (1 - \sqrt{5})^{t+1} \right], \tag{10}$$

$$A_1A_2 = \frac{1}{2^{2t+1}\sqrt{5}} \left[(1 + \sqrt{5})^{t+1} - (1 - \sqrt{5})^{t+1} \right], \text{ and } \tag{11}$$

$$A_2A_2 = \frac{1}{3} - \frac{1}{3} \left(-\frac{1}{2}\right)^{t+1} - \frac{1}{2^{2(t+1)}\sqrt{5}} \left[(1 + \sqrt{5})^{t+1} - (1 - \sqrt{5})^{t+1} \right]. \tag{12}$$

For $A_2A_2 \times A_1$, the genotypic values of A_1A_1 and A_2A_2 would be interchanged.

Genetic components in quantitative experimental designs

The theory underlying four of the seven conventional designs can be applied to haplodiploid organisms, namely: sib-analysis (Turner & Young, 1969), offspring–parental regression (Fisher, 1941), North Carolina III (Comstock & Robinson, 1952), and diallel cross (Hayman, 1954a,b). We do not consider the North Carolina I, II, and the biparental progeny designs here because the formulas for these designs derived by Killick (1971) for X-linkage genes can be readily modified for

haplodiploid organisms. Before deriving the quantitative parameters for each design, however, we must first determine, more generally, the covariance between relatives from the additive–dominance model above.

Covariance between relatives

Covariance between relatives can be considered for two population types: (1) those at Hardy–Weinberg equilibrium and (2) those that prefer to inbreed (inbreeding coefficient (F)). Genotypic values for females can be assigned in both of these situations if the allelic frequencies for A_1 and A_2 are assumed to be p and q , respectively (Table 1). Similarly, allelic values can be derived for both such that α_1 represents A_1 and α_2 represents A_2 , and then these values can be expressed as gene values which minimize the difference between the gene and assigned genotypic values (Table 1) (Fisher, 1941; Crow & Kimura, 1970) by:

$$X = \varepsilon_i^2 = p^2(a - \mu_f - 2\alpha_1)^2 + 2pq(d - \mu_f - \alpha_1 - \alpha_2)^2 + q^2(-a - \mu_f - \alpha_2)^2, \tag{13}$$

When the partial derivatives for each of α_1 , α_2 and μ_f are solved, $\alpha_1 = pa + qd - \mu_f$ and $\alpha_2 = pd - qa - \mu_f$ are obtained. If $\alpha = \alpha_1 - \alpha_2$, then the average effect of gene substitution (Fisher, 1941) becomes

$$\alpha = a + (q - p)d. \tag{14}$$

For an inbreeding population, the allelic values for females, α_{1F} and α_{2F} , are

$$\alpha_{1F} = [(p + qF)a + q(1 - F)d]/(1 + F)$$

and

$$\alpha_{2F} = [p(1 - F)d - (q + pF)a]/(1 + F),$$

and this leads to:

$$\alpha_F = \frac{(1 - F)a + (1 - F)(q - p)d}{1 + F}. \tag{15}$$

Breeding, genotypic, and dominance values for inbreeding females can then be expressed in terms of α (Table 1), where α in the females is replaced by α_F , giving the corresponding values for an inbreeding population.

Using this same approach, the gene value for inbreeding males is $\alpha_1 = a - \mu_m$ and $\alpha_2 = -a - \mu_m$ which leads to $\alpha = \alpha_1 - \alpha_2 = 2a$ and is not related to the gene frequency. Because the population mean in the male is $\mu_m = (p - q)a$, the genotypic values in terms of α are $A_1 = q\alpha$ and $A_2 = -p\alpha$ (Table 1). There are no allelic frequencies or dominance values for the haploid males because they are not directly affected by inbreeding.

Given the values in Table 1, the covariance between relatives within a haplodiploid population, whether in Hardy–Weinberg equilibrium or as an inbreeding population, can now be deduced. Based on joint probability distributions between relatives (mother–daughter (M, D), mother–son (M, Son), father–daughter (F, D), sister–sister (S, S), sister–brother (S, B), brother–brother (B, B)) for X-linked genes (Li & Sacks, 1954), the products of the number of parental A_1 alleles, genotypic values of the offspring, and the probability of the corresponding joint distribution can be summed to obtain the covariance between relatives:

$$\text{Cov}(M, D) = \text{Cov}(M, Son) = \text{Cov}(F, D) = \frac{1}{2}V_A, \tag{16}$$

where the additive genetic variance, $V_A = 2pq\alpha^2$, is considered a heritable component (Falconer, 1989). Fathers do not contribute anything to their sons, so there is no father–son covariance. The covariance between sisters is:

$$\text{Cov}(S, S) = \frac{3}{4}V_A + \frac{1}{2}V_D, \tag{17}$$

where the dominance variance, $V_D = p^2q^2d^2$. The covariance between sister and brother or between brothers is:

Table 1 Genetic components by genotype for haplodiploid organisms

Genotype	A_1A_1	A_1A_2	A_2A_2	A_1	A_2
Frequency	p^2	$2pq$	q^2	p	q
Inbreeding frequency	$p^2 + pqF$	$2pq(1 - F)$	$q^2 + pqF$	—	—
Assigned value	a	d	$-a$	a	$-a$
Gene value	$\mu_f + 2\alpha_1$	$\mu_f + \alpha_1 + \alpha_2$	$\mu_f + \alpha_2$	$\mu_m + \alpha_1$	$\mu_m + \alpha_2$
Breeding value	$2q\alpha$	$(p - q)\alpha$	$-2p\alpha$	$q\alpha$	$-p\alpha$
Genotypic value	$2q(\alpha - qd)$	$(p - q)\alpha + 2pqd$	$-2p(\alpha + pd)$	$q\alpha$	$-p\alpha$
Dominant value	$-2q^2d$	$2pqd$	$-2p^2d$	—	—

Table 2 Mid-parental values and joint distribution probabilities between mid-parent and daughter for haplodiploid organisms

Mate in parents	Frequency	Mid-value	Daughter		
$A_1A_1 \times A_1$	p^3	$3q/2$	p^3	0	0
$A_1A_1 \times A_2$	p^2q	$q - p/2$	0	p^2q	0
$A_1A_2 \times A_1$	$2p^2q$	$q - p/2$	p^2q	p^2q	0
$A_1A_2 \times A_2$	$2pq^2$	$q/2 - p$	0	pq^2	pq^2
$A_2A_2 \times A_1$	pq^2	$q/2 - p$	0	pq^2	0
$A_2A_2 \times A_2$	q^3	$-p/2$	0	0	q^3

$$\text{Cov}(S, B) = \frac{1}{4}V_A. \tag{18}$$

Now the covariance between mid-parent and offspring is more complicated. There is no covariance for sons, but for daughters, the combined joint genotypic distribution can be derived from Penrose’s work on X-linked genes (Li, 1976). Mid-parental values are the averages of the deviations from the means of the number of alleles over the mother and the father (Table 2). When the daughter breeding values from Table 1 are multiplied by the mid-values and joint probabilities in Table 2, the covariance between the mid-parental values and daughters becomes:

$$\text{Cov}(Mid, D) = \frac{1}{2}V_A. \tag{19}$$

It is necessary to note here that when covariance between the sexes is deduced, it is assumed that the gene effects in the haploid male are equal to those in diploid homozygous females. However, if this is not the case, then estimates of covariance between the sexes may be biased using the above formulas.

Experimental designs

Sib-analysis In a sib-analysis design, n males and m females are chosen randomly from a panmictic population and then one male mates randomly with several females to generate half- and full-sib offspring. The variances for haplodiploid organisms are calculated the same as for diploid ones; between sires, the variance is the covariance between the means of the half-sibs in the females, and between dams, the variance is the difference between the sisters’ covariance (eqn 17) and sire’s variance (Table 3). V_{Ec} in Table 3 is the common environmental variance, and V_{Ew} is the within-group environmental variance (Falconer, 1989). Although generally in accordance with previous work on sib-

Table 3 Expected genetic variance and narrow-sense heritability in sib analysis for haplodiploid organisms

Source	Expected variance and heritability
Sires (σ_S^2)	$(V_A)/2$
Dams (σ_D^2)	$(V_A)/4 + (V_D)/2 + V_{Ec}$
Progenies	$(V_A)/4 + (V_D)/2 + V_{Ew}$
Total	$V_A + V_D + V_{Ec} + V_{Ew} = V_P$
Heritability in male	$h^2 = 2\sigma_S^2/V_P = V_A/V_P$
Heritability in female	$h^2 = 4\sigma_D^2/V_P = (3V_A + 2V_D)/V_P$

analysis for diploid organisms (Falconer, 1989), Table 3 shows that the formulas differ slightly.

Although the above sib-analysis is based on a population of haplodiploid organisms in Hardy–Weinberg equilibrium, the statistics can also be used to approximate a population with moderate inbreeding ($F \leq 0.5$, see discussion section). If $F > 0.5$, then other more appropriate designs such as biparental progeny and North Carolina *I*, *II*, and *III* should be used. The F -value for a haplodiploid population at generation t (F_t) can be calculated using eqn 20 which is modified from Falconer’s equation 312 (Falconer, 1989, p. 65):

$$F_t = 1 - \left(1 - \frac{1}{(1 + S)N_e}\right)^t, \tag{20}$$

where S is sex ratio (female fraction in a brood), N_e is an effective population size which can be calculated using Caballero’s equation 4 (1995) for X-linked loci. Sex ratio is included in the formula because a female has two genes and a male only one at any given locus. Equation 20 is an accumulated inbreeding coefficient after t breeding generations from the base haplodiploid population. This means that this formula can only be used for a population of known origin, such as artificial outcrosses ($F=0, t=0$) and breeding history (the t -value).

Offspring–parental regression Based on eqns 16 and 19, it can be shown that the slope of the regression line between offspring and parents in Hardy–Weinberg equilibrium represents a fraction of the average effect of gene substitution; this fraction coincides with the concept of narrow-sense heritability (h^2). Therefore, when independent observations are made in a haplodiploid population, the data can be analysed by regression of offspring on parents, assuming that there is gene dosage compensation in the haploid males.

Given this, when daughters are regressed on mothers, the slope of the regression is half of the average effect of

gene substitution. Hence, when the mean number of A_1 alleles in the mother ($M_P = 2p$) and its variance ($V_P = 2pq$) from Table 2 are substituted into eqn 16, the slope will become $b_{MD} = \alpha/2$, and heritability will become:

$$h^2 = 2b_{MD}. \quad (21)$$

This is also true for the regression of sons on mothers and only slightly modified for daughters on fathers where $b_{DF} = \alpha$ and thus,

$$h^2 = b_{DF}. \quad (22)$$

When the mid-parent mean ($M_{\bar{P}} = 0$) and variance ($V_{\bar{P}} = \frac{3}{4}pq$) from Table 2 are substituted into eqn 19, the slope of the regression line will become $b_{Mid-D} = \frac{4}{3}\alpha$ and heritability can be defined as:

$$h^2 = \frac{3}{4}b_{Mid-D}. \quad (23)$$

For an inbreeding population, the difference between α and α_F would be $[-2F(q-p)d]/(1+F)$, based on eqns 14 and 15. This suggests that the heritability derived for a population in Hardy-Weinberg equilibrium will be overestimated if $[-2F(q-p)d] > 0$, underestimated if $[-2F(q-p)d] < 0$ or unbiased if $[-2F(q-p)d] = 0$ for an inbred population. This last condition is probably rare in haplodiploid systems because it requires that $d=0$ or $p=q$, and the only work to date which has examined this has found that $|d| \gg 0$ for some life history traits (Antolin, 1992; Liu, 1998). In reality, the parameters F , d , p , and q are often unknown for most populations. Thus, the above conditions can only be determined by using estimates of F , d , p or q from other studies with the same species reared under similar conditions.

North Carolina design III (NC III) In a *NC III* design, male progeny from generation 2 (F_2) of two inbred strains (L_1 and L_2) are backcrossed to their grandmothers, and their progeny are arranged in a completely randomized block design (Comstock & Robinson, 1952). For diploid organisms, the sex of the F_2 progeny that are backcrossed does not affect the genetic analysis, however, it does for haplodiploid ones. For example, expectations differ when males in the F_2 are backcrossed to their grandmothers vs. when females in the F_2 are backcrossed to their grandfathers (Table 4) and this affects their respective variances relative to diploid organisms.

In haplodiploid organisms, male backcrosses result in doubled additive and dominance variances:

$$\begin{aligned} V_{\frac{1}{2}(L_1+L_2)} &= \sum_{i=1}^n \frac{1}{4}a_i^2 + \frac{1}{2}V_{Eb} \quad \text{and} \\ V_{\frac{1}{2}(L_1-L_2)} &= \sum_{i=1}^n \frac{1}{4}a_i^2 + \frac{1}{2}V_{Eb} \end{aligned} \quad (24)$$

where V_{Eb} is the environmental variance (Falconer, 1989) while female backcrosses result in halved additive or dominance variances:

$$\begin{aligned} V_{\frac{1}{2}(L_1+L_2)} &= \sum_{i=1}^n \frac{1}{16}a_i^2 + \frac{1}{2}V_{Eb} \quad \text{and} \\ V_{\frac{1}{2}(L_1-L_2)} &= \sum_{i=1}^n \frac{1}{16}a_i^2 + \frac{1}{2}V_{Eb}^2. \end{aligned} \quad (25)$$

Unlike male backcrosses in which the covariance between the sum and difference ($\text{Cov}(L_1 + L_2, L_1 - L_2) = a^2 - ad$) provides no information on the direction of dominance, female backcrosses do. In the latter case, the covariance between the sum and difference ($\text{Cov}(L_1 + L_2, L_1 - L_2) = -\sum_{i=1}^n \frac{1}{4}a_i d_i$) can show whether dominance is unidirectional using an F -test as Mather & Jinks (1982) have demonstrated for diploid organisms.

Diallel cross A full diallel cross means that n inbred strains are randomly crossed with each other to produce n^2 families (Hayman, 1954a,b). Analysis of the diallel cross is based on six assumptions: homozygous parents, two alleles at a single locus, diploid segregation, no reciprocal difference, no epistasis, and no correlated gene distribution between the parents. Some of these assumptions are almost always violated with haplodiploid organisms, e.g. diploid segregation and reciprocal difference, but fortunately, female offspring can usually be analysed if they are in Hardy-Weinberg equilibrium.

If the above assumptions are true, then all the marginal frequencies, array means, and offspring means for haplodiploid females are the same as those for diploid organisms (Table 5). Thus, genetic variances and covariance for the diploid female offspring can be estimated using Hayman's (1954a,b) method for diploid organisms. Hayman's assumptions can also be tested using array variances and covariance for female offspring taking the methods suggested by Mather & Jinks (1982) for diploid organisms. However, only covariance for the maternal array can be estimated with this approach because covariance between the two sexes is not readily obtained if there is no gene dosage compensation in the haploid males or if there are maternal effects. Therefore, it is inappropriate to use averages of reciprocal family means to calculate the covariance (Wr) in haplodiploid organisms as Crusio *et al.* (1984) have done for diploid ones.

Table 4 Mating scheme and segregation of a *NC III* design and the expected female observations when males and females sampled randomly from the F_2 s that arose from a cross between A_1A_1 and A_2 parents are backcrossed to their grandmothers and grandfathers, respectively

Backcrossed lineage	Grandmother genotype and frequency		Male genotype and frequency in F_2		Grandfather genotype and frequency		Female genotype and frequency in F_2		Offspring mean	
	A_1A_1	$\frac{1}{2}$	$A_1\frac{1}{2}$	$A_2\frac{1}{2}$	$\frac{1}{2}(a+d)$	A_1	$\frac{1}{2}$	$A_1A_1\frac{1}{2}$		$A_1A_2\frac{1}{2}$
L_1	A_1A_1	$\frac{1}{2}$	a	d	$\frac{1}{2}(a+d)$	A_1	$\frac{1}{2}$	a	$\frac{1}{2}(a+d)$	$\frac{3}{4}a + \frac{1}{4}d$
L_2	A_2A_2	$\frac{1}{2}$	d	$-a$	$\frac{1}{2}(d-a)$	A_2	$\frac{1}{2}$	d	$\frac{1}{2}(a+d)$	$\frac{3}{4}d - \frac{1}{4}a$
$L_1 + L_2$			$a+d$	$d-a$	d			$a+d$	d	$\frac{1}{2}a + d$
$L_1 - L_2$			$a-d$	$a+d$	a			$a-d$	a	$a - \frac{1}{2}d$

Hayman’s analysis of variance for a diallel table cannot be used to estimate genetic parameters for the males because of the unequal mean values between the paternal and maternal arrays (Table 5). Fortunately, males do not have a dominance effect, making it simple to estimate genetic variances and covariance for them by analysis of variance. Thus, the expected variances would be:

$$V_{A(\text{Parental males})} = 2pq \sum_i^k a_i^2, \tag{26}$$

$$V_{A(\text{Male offspring})} = [4pq - 2(p - q)] \sum_i^k a_i^2, \tag{27}$$

$$V_{A(\text{Maternal array mean of male offspring})} = 4pq \sum_i^k a_i^2. \tag{28}$$

There will be no genetic variation for the paternal array mean of male offspring, because fathers do not contribute genes to their sons. The covariance between the means of mothers and sons will be the same as the variance between the parental males if there is gene dosage compensation in the males.

Discussion

Our work shows that quantitative genetic parameters can be estimated for haplodiploid organisms by modifying the conventional diploid approaches. All four standard experimental designs can be used following modifications to the test for additive–dominance, sib-analysis, offspring–parental regression, *NC III*, and diallel cross. These adjusted designs are all appropriate under specific situations when populations are in Hardy–Weinberg equilibrium. However, under the further condition of inbreeding, additional changes are needed. Because of the underlying assumption of random mating, neither sib-analysis nor offspring–parental regression deal satisfactorily with inbreeding, even when the precautions outlined below are taken into consideration. Similarly, while the diallel cross is appropriate for inbred populations, many of the assumptions underlying this design are not met by haplodiploid organisms and it is only appropriate under restricted conditions. Thus, under the assumption of inbred populations, the *NC III* design is recommended over the others because pure breeding populations are required to initiate the experiment.

The literature suggests that genotypic distributions are skewed to the homozygous state for some haplodiploid populations and this will affect our ability to estimate quantitative genetic parameters with different

Table 5 Full diallel cross and genotypic values in parents and offspring for haplodiploid organisms

Genotype Frequency Genotypic value			Mother						
			A_1A_1		A_2A_2		Mean of paternal array		
			P	a	q	$-a$	♀	♂	
Father	A_1	p	a	p^2 a	p a	pq d	q $-a$	$(pa + qd)$	$(p - q)a$
	A_2	q	$-a$	pq d	p a	q^2 $-a$	q $-a$	$(pd - qa)$	$(p - q)a$
Mean of maternal array			$(pa + qd)$	$(2pa)$	$(pd - qa)$	$(-2qa)$	$a(p - q) + 2pqd$	$2(p - q)a$	

♀ column, female offspring; ♂ column, male offspring.

experimental designs. Avery (1984) has shown that the heterozygosity at soluble isoenzyme loci (from 15 to 20 loci per individual) for natural populations of hymenopteran parasitoids is very low (averaging 0.037 over 51 hymenopteran species, with a standard error of 0.004 and a standard deviation of 0.027). If such a population is at Hardy–Weinberg equilibrium, then its genotypic fractions are skewed (0.9629*P* 0.0367*Q* 0.0004*R* 0.9813*S* 0.0187*T*) when $2pq = 0.037$. The skewness of advantageous genotypes will be similar (90% of advantageous homozygote) even if $2pq = 0.1$ (i.e. adding two times of the standard deviation (0.027) to the average (0.037) is 0.091, which is less than 0.1). Under such a low deleterious allelic frequency, the additive variance in the females would be $V_A = 2pq\alpha^2 = 0.04(a - 0.96d)^2$, where V_A is hard to detect if $a \approx d > 0$. If a is positive and d is negative, V_A contains almost 50% of the dominance contribution when $|a| \approx |d|$, implying that the estimate of V_A for such a population may not help predict its response to selection. Therefore, to estimate a meaningful quantitative genetic parameter, sampling methods and experimental design must be taken into account. Because of the high proportion of homozygotes in such skewed populations, the *NC III* or diallel cross designs should be used to estimate the parameters, rather than sib-analysis or offspring–parental regression.

Moderate inbreeding ($F \leq 0.5$) in the skewed population does not substantially change its genotypic distribution. For example, when $2pq(1 - F) = 0.037$ where $F = 0.5$, the genotypic distribution at equilibrium becomes (0.9319*P* 0.0367*Q* 0.0089*R* 0.9615*S* 0.0385*T*), and this is not much different from that above. While the gene frequency in the male population is not directly affected by inbreeding, it must be the same as that for the female population when the whole haplodiploid population reaches equilibrium (Elandt-Johnson, 1971). Thus, the proportion of genotype *T* is changed from a

random mating population (0.0187) to a nonrandom mating one (0.0385). The genetic parameters estimated by sib or regression analysis for such a mixed (random/inbred) mating population may not be notably biased. However, if inbreeding is high ($F = 0.9$), then the skewed genotypic distribution is changed considerably (0.7364*P* 0.0367*Q* 0.2265*R* 0.75495*S* 0.24505*T*), from which $V_A = 0.37(0.09a + 0.05d)^2$. This suggests that the additive genetic variance for inbred skewed populations can be more easily detected than that for skewed populations with random mating or moderate inbreeding.

Theoretically, V_A within strains increases to a maximum at $F \approx 0.5$ and declines to zero at $F = 1$, whereas V_A between strains increases as the *F*-value increases (Robertson, 1952). Therefore, heritabilities estimated by sib-analysis or regression analysis should be corrected by adding or subtracting $[-2F(q - p)d]/(1 + F)$, depending on the sign of the numerator for the particular population. Although estimates for *F*, *d*, *p*, or *q* can be derived from various experiments, the effect and frequency of the multiple genes *per se* will vary. Thus, conventional quantitative analyses can provide only a general description of the probable behaviour for any given polygene. New techniques, such as combining statistics with molecular and physiological observations, are needed to give better precision to these estimates.

Most of the quantitative experimental designs for diploid organisms have also been modified for X-linked genes (Bohidar, 1964; Killick, 1971; Mather & Jinks, 1982); however, our estimates differ from theirs in two main ways. First, our solutions for estimating many of the formulas and parameters are simpler and provide more information for haplodiploid organisms than does the analysis of X-linked genes. For example, Mather & Jinks's (1982) *ABC* test can be used for haplodiploid organisms only when the autosome components are removed and the remaining X-linkage components are

averaged over the reciprocal crosses. This makes it more laborious and less informative than the modified joint test we have proposed. Similarly, our formulas for calculating genotypic fractions are more convenient than those proposed by Jennings (1916) for X-linked genes, because they do not require repeated reference back to the original sequence. In terms of the diallel cross, we have relied on Hayman's analysis which provides more information about the estimates, in contrast to Mather & Jinks (1982) who simply extend their analysis of a panmictic population to the diallel cross.

Second, our estimates are more appropriate than those developed for X-linked genes in that they are directly applicable to haplodiploid organisms. For example, Dickerson's (1969) regression coefficients for X-linked genes would be incorrectly applied to haplodiploid organisms because, unlike ours, his formulas do not maintain the assumption of gene compensation in the male (James, 1973). Our method for calculating the covariance between relatives provides a more straightforward solution than the X-linked work of Bohidar (1964), because, by necessity, his definition of the genetic components is broad and unclear (James, 1973). As well, while both Mather & Jinks (1982) and Bohidar (1964) have modified sib-analysis for X-linked genes, unlike us, the former did not distinguish offspring variance by parental sex, while the latter included interaction terms and other main effects that are impossible to separate for haplodiploid organisms. Finally, the North Carolina III analysis we propose recognizes that there are two different backcrosses possible (specifically, females backcrossed to their grandfathers) and that they have different expected variances and tests for the direction of dominance, in contrast to the X-linked work by Killick (1971).

The statistics derived in our study have both theoretical and applied significance. From a theoretical perspective, the genetic parameters provide a base from which further evolutionary analysis on haplodiploid populations should proceed. In particular, we have shown an effective means for dealing with the inherent problem of inbreeding and its associated skewed genotypic distribution commonly found in these populations. From the applied perspective, our modified estimations for genetic variance and covariance in haplodiploid populations can be used for future biological control programmes to monitor the quality of hymenopteran parasitoids, and to examine the maintenance of genetic variation in their populations during commercial mass production (Liu and Smith, 2000). Similarly, our modified heritability estimates will be valuable in designing optimal breeding plans and improving the quality of biological control agents for such programmes. Finally,

when used in combination with fitness data, these parameters can help predict the evolutionary consequences of haplodiploid pests (e.g. thrips, white flies, scale insects, and many species of bark beetles) over successive generations, and thus, provide invaluable information on their population dynamics and the need for suppression.

Acknowledgements

We thank K. Ritland and S. H. Orzack for comments on the original thesis work and two anonymous reviewers for their suggested revision. This research was supported financially by an NSERC grant and Premier's Council Technology Fund contract to S. M. S.

References

- ADAMSON, M. L. 1989. Evolutionary biology of the Oxyurida (Nematoda): Biofacies of a haplodiploid taxon. *Adv. Parasitol.*, **28**, 175–228.
- ANTOLIN, M. F. 1992. Sex ratio variation in a parasitic wasp. II. Diallel cross. *Evolution*, **46**, 1511–1524.
- AVERY, P. J. 1984. The population genetics of haplodiploids and X-linked genes. *Genet. Res.*, **44**, 321–341.
- BOHIDAR, N. R. 1964. Derivation and estimation of variance and covariance components associated with covariance between relatives under sex-linked transmission. *Biometrics*, **20**, 505–521.
- CABALLERO, A. 1995. On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics*, **139**, 1007–1011.
- CAVALLI, L. L. 1952. An analysis of linkage in quantitative inheritance. In: Reeve, E. C. R. and Waddington, C. H. (eds) *Quantitative Inheritance*, pp. 135–144. HMSO, London.
- COMSTOCK, R. E. AND ROBINSON, H. F. 1952. Estimation of average dominance of genes. In: Gowen, J. W. (ed.) *Heterosis*, pp. 494–516. Iowa State College Press, Ames IA.
- CROZIER, R. H. 1970. On the potential for genetic variability in haplo-diploidy. *Genetica*, **41**, 551–556.
- CROW, J. F. AND KIMURA, M. 1970. *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- CRUSIO, W. E. J., KERBUSCH, M. L. AND VAN ABEELLEN, J. H. F. 1984. The replicated diallel cross: a generalized method of analysis. *Behav. Genet.*, **14**, 85–103.
- DICKERSON, G. J. 1969. Techniques for research in quantitative animal genetics. In: *Techniques and Procedures in Animal Science Research*, pp. 36–79. Am. Soc. Animal Sci., Albany, NY.
- ELANDT-JOHNSON, R. C. 1971. *Probability Models and Statistical Methods in Genetics*. John Wiley & Sons, New York.
- FALCONER, D. S. 1989. *Introduction to Quantitative Genetics*, 3rd edn. Longman Scientific and Technical – copublished in the United States with John Wiley and Sons, Inc., New York.
- FISHER, R. A. 1941. Average excess and average effect of a gene substitution. *Ann. Eugen.*, **11**, 53–63.

- GRAUR, D. 1985. Gene diversity in Hymenoptera. *Evolution*, **39**, 190–199.
- HAMILTON, W. D. 1967. Extraordinary sex ratios. *Science*, **156**, 477–488.
- HARDY, I. C. W. 1994. Sex ratio and mating structure in the parasitoid Hymenoptera. *Oikos*, **69**, 3–20.
- HAYMAN, B. I. 1954a. The analysis of variance of diallel tables. *Biometrics*, **10**, 235–244.
- HAYMAN, B. I. 1954b. The theory and analysis of diallel crosses. *Genetics*, **39**, 789–809.
- JAMES, J. W. 1973. Covariances between relatives due to sex-linked genes. *Biometrics*, **29**, 584–588.
- JENNINGS, H. S. 1916. The numerical results of diverse systems of breeding. *Genetics*, **1**, 53–89.
- KILLICK, R. J. 1971. Sex-linkage and sex-limitation in quantitative inheritance. *Heredity*, **27**, 175–188.
- LI, C. C. 1976. *First Course in Population Genetics*. The Boxwood Press, Pacific Grove, CA.
- LI, C. C. AND SACKS, L. 1954. The derivation of joint distribution and correlation between relatives by the use of stochastic matrices. *Biometrics*, **10**, 347–360.
- LIU, F. H. 1998. *Quantitative Genetics of Male-Haploid Organisms Used for Biological Control*. PhD Dissertation. University of Toronto.
- LIU, F. H. AND SMITH, S. M. 2000. Measurement and selection of parasitoid quality for biological control. *Biocontrol Sci. Technol.*, **10**, 3–13.
- MATHER, K. AND JINKS, J. L. 1982. *Biometrical Genetics*, 3rd edn. Chapman & Hall Ltd, London, New York.
- ORZACK, S. H. AND GLADSTONE, J. 1994. Quantitative genetics of sex ratio traits in the parasitic wasp, *Nasonia vitripennis*. *Genetics*, **137**, 211–220.
- PINTO, J. D., STOUTHAMER, R., PLATNER, G. R. AND OATMAN, E. R. 1991. Variation in reproductive compatibility in *Trichogramma* and its taxonomic significance (Hymenoptera: Trichogrammatidae). *Ann. Entomol. Soc. Am.*, **84**, 37–46.
- RINDERER, T. E. 1977. Measuring the heritability of characters of honeybees. *J. Apic. Res.*, **16**, 95–98.
- ROBERTSON, A. 1952. The effect of inbreeding on the variation due to recessive genes. *Genetics*, **37**, 189–207.
- SAS/IML. Software. 1990. *Version 6*. 1st edn. SAS Institute, Cary.
- TURNER, H. N. AND YOUNG, S. S. Y. 1969. *Quantitative Genetics in Sheep Breeding*. Cornell University Press, Ithaca, NY.