Quantitative genetics of allogamous F_2 : an origin of randomly fertilized populations

IAN L. GORDON*

Institute of Molecular Biosciences, Massey University, Private Bag 11222, Palmerston North, New Zealand

The quantitative genetic properties are derived for the bulk F_2 originating from random fertilization (RF) amongst hybrid (F_1) individuals. Only its mean appears to have been derived previously, and that definition is confirmed (by another method). New general equations are found also for all genotype frequencies, allele frequencies, inbreeding coefficient, the genotypic, additive-genetic and dominance variances, and broad-sense and narrow-sense heritabilities. The assumption that such an F_2 is a classical RF population is shown to be correct. Indeed, the allogamous F_2 is a natural origin for the RF population. The relationships are given between precedent RF populations (parents) and subsequent RF populations following hybridization (allogamous F_2). The allogamous F_2 is generally inbred with respect to its parental F_1 , the degree depending on the hybrid's parents' allele frequencies. At the same time, it is *outbred* with respect to those original parents, and not inbred at all with respect to the equivalent RF population. The genotypic variance is generally more than in the F_1 , and likewise for heritabilities. These findings make it possible to evaluate the genetic advance from selection and hybridization. The results depend on the allele frequencies of the original parents and the degree of overdominance, but generally, selection is more advantageous than hybrid vigour.

Keywords: F_2 , genetic advance, genetic variances, heritabilities, inbreeding, random fertilization.

Introduction

Many episodes in evolution and plant breeding are initiated by hybridizing between parent populations or individuals. Meiosis and syngamy within the hybrid (F_1) lead to a population with a different frequency structure (the F_2), i.e. the F_1 is in disequilibrium and has an ephemeral population structure. The quantitative genetic properties of the F_1 population have been examined (Gordon, 1999), but how do the properties of the emergent F_2 relate to these? Intuitively, because of meiosis, we would expect more variance: but we need to explore the quantitative relationships of this matter. Intuitively, also, we would expect to have to account for effects of inbreeding because of shared ancestral origins (F_1 individuals) amongst members of the F_2 population. This needs to be reconciled with the assumption that an allogamous F_2 can be considered as a single panmictic random fertilizing gamodeme with zero inbreeding. It appears that these, and related, matters have not been explored rigorously nor quantitatively.

These omissions became strongly apparent when the author sought to estimate the genetic advance arising from selection from an F_2 . This is a fundamental phase in important protocols of plant breeding, such as Line Breeding and Pedigree Breeding (Allard, 1960; Poehlmann, 1979; Moore and Janik, 1983). Such selection from the F_2 is essential also to the ongoing contribution of a hybrid in natural selection. In order to evaluate genetic advance from such selection, we need more knowledge on the properties of the F_2 .

The present paper examines the quantitative genetic properties of the bulk F_2 population generated by random fertilization amongst F_1 individuals. Some earlier focus has examined the mean of such an F_2 (Falconer, 1981), in order to demonstrate the drop in hybrid vigour compared with the F_1 . Here, we derive as well the allogamous F_2 genotypic variance (σ_G^2), genic variance (σ_A^2), dominance variance (σ_D^2), broad-sense and narrow-sense heritabilities (h_B^2 and h_N^2 , respectively), genotypic and allelic frequencies, and levels of inbreeding. Following this, we will be able to estimate selection genetic advance from the F_2 and we will examine the relationship of the allogamous F_2 to random fertilizing (RF) populations.

*Correspondence. E-mail: genovir@massey.ac.nz

General method

A population consists of genotypes arising from alleles (or allele Expectations) A_1 and A_2 present with frequencies p and q respectively, with no omissions (i.e. $p + q = 1$). Its individual phenotypic Expectations are its genotype effects defined as deviates from the homozygote midpoint, namely $\mathbf{g}' = [a, d, -a]$ for $\{A_1A_1, A_1A_2, A_2A_2\}$, respectively. This is the familiar single-factor gene model used to define classical RF populations (Falconer, 1981). It has been used also to define hybrids (Gordon, 1999), where further discussion is presented on its utility. In examples given in this paper, we will use $a = 10$, and $d = 7.5$ (partial dominance). The population properties also depend on the genotype frequencies; it is one of the tasks of this paper to discover these for the allogamous F_2 .

The original parent populations set the parameters not only for the hybrid (F_1), but also for this F_2 . The focal parent (P_1) is a population of individuals, with allele frequencies p_1 and q_1 . As $p_1 \rightarrow 1$ (or 0), we are dealing with a pure-line population (or even a homozygous individual); but a focal-parent population may have any p_1 within the range $0 \leq p_1 \leq 1$. The other parent (P_2) has frequencies defined as *offsets* from the P_1 frequencies, following Falconer (1981). This leads to $p_2 = p_1 - y$, and $q_2 = q_1 + y$, where $y = p_1 - p_2$. These are the same definitions as used for hybrids (Falconer, 1981; Gordon, 1999).

The RF population analogous to the allogamous F_2 is of special relevance, and is based on $p_{F_1} = p_1 - \frac{1}{2}y$ and $q_{F_1} = q_1 + \frac{1}{2}y$ (Gordon, 1999). It has the classical properties (Falconer, 1981):

$$\bar{g}_{RF} = a(p_{F_1} - q_{F_1}) + 2p_{F_1}q_{F_1}d \quad (\text{gene model mean}),$$

$$\mathbf{f}'_{RF} = [q_{F_1}, 2p_{F_1}q_{F_1}, p_{F_1}^2] \quad (\text{genotype frequencies})$$

(for $\{A_2A_2, A_2A_1, A_1A_1\}$, respectively),

$$\sigma_{ARF}^2 = 2p_{F_1}q_{F_1}\alpha_{F_1}^2 \quad (\text{genic variance})$$

(where $\alpha_{F_1} = a + (q_{F_1} - p_{F_1})d$),

$$\sigma_{DRF}^2 = (2p_{F_1}q_{F_1})^2 d^2 \quad (\text{dominance variance}).$$

A major task of this paper is to redefine these in terms of the F_2 'offset' parameters, and to compare the two sets of results.

All other Methods form an integral part of the derivations which constitute the Results of this paper, and will be presented there.

Results

Genotype frequencies

Allogamous F_2 We adopt a fundamental means of deriving the genotype frequencies: we use the F_1 genotype frequencies (Gordon, 1999) as parental frequencies, and the cross-products of these form crossing (mating) frequencies. Next, we note the segregation ratios of each family of cross which, when multiplied by the crossing frequencies, gives the family frequencies of segregating progeny genotypes. When these terms are accumulated across all crossing families, we obtain the allogamous F_2 genotype frequencies. The F_1 parent frequencies and the crossing frequencies are presented in Table 1.

Gathering the outcomes for A_1A_1 progenies leads to:

$$\mathbf{f}_{11} = p_1^4 + 2p_1^3q_1 + p_1^2q_1^2 + y[p_1^3 - 2p_1^2q_1 - p_1q_1^2] \\ + y^2[\frac{1}{2}p_1q_1 + \frac{1}{4}p_1^2 + \frac{1}{4}q_1^2] = p_1^2 - yp_1 + \frac{1}{4}y^2.$$

Similarly, for A_1A_2 progenies:

$$\mathbf{f}_{12} = 2p_1q_1[p_1^2 + 2p_1q_1 + q_1^2] + y[p_1^3 + p_1^2q_1 - p_1q_1^2 - q_1^3] \\ + \frac{1}{2}y^2[p_1^2 + 2p_1q_1 + q_1^2] \\ = 2p_1q_1 + y(p_1 - q_1) - \frac{1}{2}y^2.$$

And lastly, for A_2A_2 progenies:

$$\mathbf{f}_{22} = [q_1(p_1 + q_1)]^2 + yq_1[p_1^2 + 2p_1q_1 + q_1^2] \\ + \frac{1}{4}y^2[p_1^2 + 2p_1q_1 + q_1^2] = q_1^2 + yq_1 + \frac{1}{4}y^2.$$

We define the row-vector containing these results as $\mathbf{f}'$.

These results reveal that for any allogamous F_2 from crosses with *complementary opposite* parents (i.e. $p_2 = 1 - p_1$), the heterozygote frequency is constant ($\mathbf{f}_{12} = 0.5$) across all values of p_1 . Whenever the *other* parent has a *fixed* p_2 , the general levels of heterozygosity are often lower than for the *complementary opposites* cases. The heterozygosity of *any* F_2 is generally less than that in its corresponding F_1 ; but the two are equal when $p_1 = p_2$.

Equivalent RF and panmictic Here, we use the classical RF genotype frequencies, and substitute $(p_1 - \frac{1}{2}y)$ for p_{F_1} and $(q_1 + \frac{1}{2}y)$ for q_{F_1} (Gordon, 1999). Thus:

$$\mathbf{f}_{11} = (p_1 - \frac{1}{2}y)^2 = p_1^2 - yp_1 + \frac{1}{4}y^2, \\ \mathbf{f}_{12} = 2(p_1 - \frac{1}{2}y)(q_1 + \frac{1}{2}y) \\ = 2p_1q_1 - y(q_1 - p_1) - \frac{1}{2}y^2, \\ \mathbf{f}_{22} = (q_1 + \frac{1}{2}y)^2 = q_1^2 + yq_1 + \frac{1}{4}y^2.$$

Table 1 Crossing frequencies of F₁ genotypes (♀ and ♂) to produce allogamous F₂, together with segregations ($A_1A_1 : A_1A_2 : A_2A_2$) in the resultant progeny lines. Allele frequencies p_1 and q_1 refer to original focal parent, and $y = p_1 - p_2$, the offset between the two parent populations

♂ \ ♀	$A_1A_1 (p_1^2 \quad p_1y)$	$A_1A_2 (2p_1q_1 + p_1y \quad q_1y)$	$A_2A_2 (q_1^2 + q_1y)$
$A_1A_1 (p_1^2 \quad p_1y)$	$p_1^4 \quad 2p_1^3y + p_1^2y^2$ 1:0:0	$2p_1^3q_1 + p_1^3y \quad 3p_1^2q_1y$ $p_1^2y^2 + p_1q_1y^2$ 1:1:0	$p_1^2q_1^2 + p_1^2q_1y \quad p_1q_1^2y$ $p_1q_1y^2$ 0:1:0
$A_1A_2 (2p_1q_1 + p_1y \quad q_1y)$	$2p_1^3q_1 + p_1^3y \quad 3p_1^2q_1y$ $p_1^2y^2 + p_1q_1y^2$ 1:1:0	$4p_1^2q_1^2 + 4p_1^2q_1y \quad 4p_1q_1^2y$ $+ p_1^2y^2 \quad 2p_1q_1y^2 + q_1^2y^2$ 1:2:1	$2p_1q_1^3 \quad q_1^3y + 3p_1q_1^2y$ $q_1^2y^2 + p_1q_1y^2$ 0:1:1
$A_2A_2 (q_1^2 + q_1y)$	$p_1^2q_1^2 + p_1^2q_1y \quad p_1q_1^2y$ $p_1q_1y^2$ 0:1:0	$2p_1q_1^3 \quad q_1^3y + 3p_1q_1^2y$ $q_1^2y^2 + p_1q_1y^2$ 0:1:1	$q_1^4 + 2q_1^3 + q_1^2y^2$ 0:0:1

*Genotype frequency in the F₁.

We notice that these are the same results as obtained for the allogamous F₂, proving that the genotype frequencies at least are the same for the two situations. Note that these two results were derived in different but relevant ways, and obtaining the same outcome establishes their equivalence.

Allele frequencies

We need to check the allele frequencies of the F₂ using the genotype frequencies $\mathbf{f}$. This can be done as weighted means using the allele structures of genotypes as weights. Thus, for the A_1 allele, the weights are $\mathbf{w}'_1 = [1, 1/2, 0]$ for $\{A_1A_1, A_1A_2, A_2A_2\}$, respectively; and for the A_2 allele, the respective weights are $\mathbf{w}'_2 = [0, 1/2, 1]$. The allele frequencies are then found as follows:

$$p_{F_2} = \mathbf{w}'_1 \mathbf{f} = p_1^2 \quad yp_1 + \frac{1}{4}y^2 + p_1q_1 + \frac{1}{2}yp_1 \quad \frac{1}{2}yq_1 \quad \frac{1}{4}y^2$$

$$= p_1 \quad \frac{1}{2}y;$$

and

$$q_{F_2} = \mathbf{w}'_2 \mathbf{f} = q_1^2 + yq_1 + \frac{1}{4}y^2 + p_1q_1 + \frac{1}{2}yp_1 \quad \frac{1}{2}yq_1 \quad \frac{1}{4}y^2$$

$$= q_1 + \frac{1}{2}y.$$

Note that these are the same as the F₁ frequencies (Gordon, 1999), which we would expect. (The RF frequencies were defined to these values at the outset).

Inbreeding coefficient

The natural inbreeding reference population for the F₂ is the F₁, as the former is sexually derived from the latter. The F₂ bulk population is a mixture of many different

full-sib lines and half-sib lines, with varying degrees of coancestry amongst them. The simplest approach to defining inbreeding in this bulk mixture is the net drop in heterozygosity of the F₂ relative to the F₁. This approach precludes any need to unravel the breeding system mixture. Furthermore, it automatically reflects the dynamics of the bulk which depend on the parental inputs (p_1). Therefore, the inbreeding coefficient is defined as:

$$\phi = \left(1 \quad \frac{f_{12F_2}}{f_{12F_1}} \right),$$

where f_{12} of the F₂ is defined in this paper, and f_{12} of the F₁ is defined in Gordon (1999). The degree of inbreeding in the allogamous F₂, for all possible focus-parent inputs (p_1), is visualized in Fig. 1, where the other-parent is *complementary opposite* or *fixed* $p_2 = 0.01, 0.5$ or 0.8 as examples.

We observe that the level of inbreeding depends on the relatives-mixture present in the F₁ bulk, which changes as does p_1 . Where $p_1 \rightarrow 0$ (or 1), for *complementary opposite crosses*, the hybrid is virtually entirely heterozygous, and the F₂ is equivalent to selfing a heterozygote; hence the inbreeding is $\phi \rightarrow \frac{1}{2}$. Again, for this same crossing system, when $p_1 \rightarrow p_2 \rightarrow \frac{1}{2}$, the 'hybrid' is virtually a RF population, and the F₂ is equivalent to a second-generation RF in equilibrium with the 'hybrid'. Therefore, there is no change in heterozygosity, and $\phi \rightarrow 0$.

Now, we will compare this F₂ heterozygosity against that of the equivalent RF, as an alternative reference. This is achieved simply by replacing f_{12} of the F₁ with f_{12} of the RF in the equation for ϕ . We recall that we have shown that both the allogamous F₂ and the equivalent

Fig. 1 Inbreeding coefficient (ϕ) for the allogamous F_2 from crosses of the *focus* parent (A_1 frequency p_1), and the *other* parent either with *complementary-opposite* p_2 , or with *fixed* $p_2 = 0.5$ or 0.8 .

RF (based on p_{F_1}) have the same level of heterozygosity. Therefore, the inbreeding of the allogamous F_2 is zero when referenced against the equivalent RF population. This conforms with widespread notions already extant. But it is of doubtful relevance, because the *natural reference* population is the parental origin, viz. the F_1 . The whole issue becomes more intriguing still if we use the *mean heterozygote frequency of the two parents* as the reference basis. We then find that nearly all allogamous F_2 are *outbred* (negative ϕ) with respect to the parents, except at $p_1 \rightarrow \frac{1}{2}$. This dependence of the value of inbreeding on an appropriate reference base was noted by its pioneer, Sewall Wright (Wright, 1921).

Population mean

Allogamous F_2 The F_2 mean ($\bar{F}_2$) is $\mu + \bar{G}_{F_2}$, where μ is the attribute background mean, and $\bar{G}_{F_2}$ ($=\gamma$ also) is the gene-model mean. The latter is obtained as follows:

$$\begin{aligned}\bar{G}_{F_2} &= \mathbf{f}'\mathbf{g} \\ &= (p_1^2 - yp_1 + \frac{1}{4}y^2)a + (2p_1q_1 + y(p_1 - q_1) - \frac{1}{2}y^2)d \\ &\quad + (q_1 + yq_1 + \frac{1}{4}y^2)(-a) \\ &= [(p_1 - q_1)a + 2p_1q_1d] - y[a + (q_1 - p_1)d] - \frac{1}{2}y^2d \\ &= \bar{G}_1 - y\alpha_1 - \frac{1}{2}y^2d.\end{aligned}$$

Fig. 2 Population mean of the allogamous F_2 from crosses between the *focus* parent (A_1 frequency p_1), and the *other* parent either with *complementary-opposite* p_2 , or with *fixed* $p_2 = 0.5, 0.8$, or 0.01 . The corresponding F_1 means are given also in the first two cases, for comparison.

The α_1 is the usual *average allele substitution effect* for the focal parent population (e.g. Falconer, 1981), and equals $a + (q_1 - p_1)d$. This is the same result as presented by Falconer (1981) using another method of derivation, and the $\frac{1}{2}y^2d$ represents the drop in hybrid vigour, which was his focus. Several examples of this mean are shown in Fig. 2, all for partial dominance ($a = 10, d = 7.5, \mu = 10$). Figure 2 shows both *complementary opposite parents* and *fixed other-parent* $p_2 = 0.5$, and compares both the F_2 and F_1 means. The loss of vigour is obvious, especially for $p_1 \rightarrow 0$ (or 1). Two other F_2 -mean examples are shown to illustrate extremes, namely *fixed* $p_2 = 0.01$ and *fixed* $p_2 = 0.8$.

It is interesting to note the patterns of increase in the mean with increasing overdominance. The effect of *parental complementarity* (i.e. $p_{F_1} = \frac{1}{2}$ always) leads to the mean being constant across all p_1 , but rising in steps with increasing overdominance. The patterns with other forms of parental relationship (with respect to p_1 and p_2) are more varied, but there is still a general increase in the mean with increasing overdominance.

Equivalent RF and panmictic We use at once the classical RF definition of the gene-model mean (see General Method), and substitute into it the values for p_{F_1} in *offset* form. Thus:

$$\begin{aligned} G_{RF(F_1)} &= (p_{F_1} - q_{F_1})a + 2p_{F_1}q_{F_1}d \\ &= (p_1 - \frac{1}{2}y - q_1 - \frac{1}{2}y)a + 2(p_1 - \frac{1}{2}y) \\ &\quad \times (q_1 + \frac{1}{2}y)d \\ &= \bar{G}_{F_2}. \end{aligned}$$

Through this appropriate independent derivation, we find that the RF mean is the same as the F₂ mean. We now have proved that in two properties the allogamous F₂ and the RF population are identical, viz. *genotype frequencies* and *gene-model mean*.

Genotypic variance

Allogamous F₂ The allogamous F₂ genotypic variance has the Expectation

$$\sigma_G^2 = \sum_{i \leq j}^3 f_{ij} g_{ij}^2 - \gamma^2,$$

where the right-hand terms are an unadjusted sum-of-squares (USS) and a correction factor (CF), respectively. Using the previously defined **f'** and **g'** vectors and the scalar γ , and after subtracting the CF from the USS followed by gathering and factoring several terms*, the SS (i.e. σ_G^2 the genotypic variance†), becomes:

$$\begin{aligned} \sigma_G^2(SS) &= 2p_1q_1a^2 + (q_1 - p_1)4p_1q_1ad + 2p_1q_1d^2 \\ &\quad \times (1 - 2p_1q_1) - y(q_1 - p_1)a^2 \\ &\quad - y2ad(p_1^2 + q_1^2 - 4p_1q_1) - yd^2(q_1 - p_1) \\ &\quad - 4p_1q_1^2 + 4p_1^2q_1 - \frac{1}{2}y^2a^2 - y^23ad(q_1 - p_1) \\ &\quad - \frac{1}{2}y^2d^2[1 + 2p_1^2 - 8p_1q_1 + 2q_1^2] \\ &\quad - y^3[(q_1 - p_1)d^2 + ad] - \frac{1}{4}y^4d^2. \\ &= 2p_1q_1\alpha_1^2 + (2p_1q_1)^2d^2 + y4p_1q_1ad \\ &\quad - y(q_1 - p_1)[a^2 + 2(q_1 - p_1)ad + (q_1 - p_1)^2d^2] \\ &\quad - \frac{1}{2}y^2[a^2 + 2(q_1 - p_1)ad + (q_1 - p_1)^2d^2] \\ &\quad - \frac{1}{2}y^2[4(q_1 - p_1)ad - 4(q_1 - p_1)^2d^2] \\ &\quad - y^3[(q_1 - p_1)d^2 + ad] - \frac{1}{4}y^4d^2. \end{aligned} \quad (1)$$

*The following relations have been used to resolve these equations:

$$\begin{aligned} (1 - 2pq) &= p^2 - 2pq + q^2 + 2pq = (q - p)^2 + 2pq; \\ (p^2 + q^2 - 4pq) &= p^2 - 2pq + q^2 - 2pq = (q - p)^2 - 2pq; \\ [1 + 2p^2 - 8pq + 2q^2] &= (1 - 4pq) + 2(p^2 - 2pq + q^2) \{= 3(q - p)^2\}; \\ [a^2 + 2(q - p)ad + (q - p)^2d^2] &= \alpha^2; \\ (1 - 4pq) &= p^2 + 2pq + q^2 - 4pq = (q - p)^2; \\ [2(q - p)ad + (q - p)^2d^2] &= \alpha^2 - a^2; \text{ and} \\ [(q - p)d^2 + ad] &= d\alpha. \end{aligned}$$

†Because these SS are based on frequencies rather than counts, they equate immediately to mean-squares in fact.

$$\begin{aligned} &= 2p_1q_1\alpha_1^2 + (2p_1q_1)^2d^2 + y4p_1q_1ad - y(q_1 - p_1)\alpha_1^2 \\ &\quad - \frac{1}{2}y^2\alpha_1^2 - y^2(\alpha_1^2 - a^2) - y^3(d\alpha_1) - \frac{1}{4}y^4d^2. \end{aligned} \quad (2)$$

Now, recalling the classical definitions of σ_A^2 , and of σ_D^2 (Falconer, 1981), and of $\text{cov}(ad)$ and σ_d^2 , (Gordon, 1999), this whole expression resolves into:

$$\begin{aligned} \sigma_{G_{F_2}}^2 &= \sigma_{A_1}^2 \left[1 - \frac{y}{2p_1q_1}(q_1 - p_1 + \frac{1}{2}y) \right] + \sigma_{D_1}^2 + y\text{cov}_1(ad) \\ &\quad - \sigma_{d_1}^2 \left(\frac{y^4}{8p_1q_1} \right) - y^2(\alpha_1^2 - a^2) - y^3(\alpha_1 d). \end{aligned} \quad (3)$$

The previous equation (eqn 2) is simpler, but this one provides some link with the traditional nomenclature. However, there is some danger in it, for these components are *not* the genic and dominance variances of the F₂ σ_G^2 . We will explore that issue in due course.

The variance can be visualized for our partial dominance and *complementary opposites* and *fixed other-parent* $p_2 = \frac{1}{2}$ crosses in Fig. 3. In each case, it is compared also with the σ_G^2 of its F₁.

Notice from Fig. 3 that this F₂ variance is generally greater than that of its F₁, except for the region $p_1 \rightarrow \frac{1}{2}$ where they are virtually equal. This was one of the questions we posed at the beginning. For the *complementary opposites* cross, the F₂ variance greatly exceeds that of its F₁ as $p_1 \rightarrow 0$ (or 1). Another conspicuous result for the *complementary opposites* is that the variance is constant across all p_1 ; this results from the facts that p_{F_1} is always $\frac{1}{2}$, and F₂ genotype frequencies are constant for such crosses. The results for the *fixed* $p_2 = \frac{1}{2}$ crosses are conspicuously different. These general differences between the F₂ and F₁ variances have interesting consequences regarding ongoing selection efficiency, and the balances between selection and hybrid vigour, which we will discuss later.

Equivalent RF and panmictic As before, we establish an independent derivation by using the classical RF approach, and substitute into it the offset p_{F_1} structures. Therefore, for the RF base $\sigma_G^2 = \sigma_A^2 + \sigma_D^2$, where σ_A^2 is the genic variance and σ_D^2 is the dominance variance. We define each of these separately, and accumulate them at the end.

(i) Genic variance

$$\begin{aligned} \sigma_A^2(RF) &= [2p_{F_1}q_{F_1}][\alpha_{F_1}^2] \\ &= 2p_1q_1[a^2 + 2(q_1 - p_1)ad + (q_1 - p_1)^2d^2] \\ &\quad - y(q_1 - p_1)[a^2 + 2(q_1 - p_1)ad + (q_1 - p_1)^2d^2] \end{aligned}$$

Fig. 3 Genotypic variance (VG) of the allogamous F_2 from crosses between the *focus* parent (A_1 frequency p_1), and the *other* parent either with *complementary-opposite* p_2 , or with *fixed* $p_2=0.5$. Partial dominance is depicted ($a=10$, $d=7.5$). The corresponding F_1 variances are given also for comparison.

$$\begin{aligned}
 &+ y4p_1q_1ad + 2y(q_1 - p_1)2p_1q_1d^2 \\
 &\quad \frac{1}{2}y^2[a^2 + 2(q_1 - p_1)ad + (q_1 - p_1)^2d^2] \\
 &\quad y^2[2(q_1 - p_1)ad + (q_1 - p_1)^2d^2] \\
 &\quad y^2d^2[(q_1 - p_1)^2 - 2p_1q_1] \\
 &\quad y^3[ad + 2(q_1 - p_1)d^2] - \frac{1}{2}y^4d^2. \quad (4)
 \end{aligned}$$

$$\begin{aligned}
 &= 2p_1q_1\alpha_1^2 - y(q_1 - p_1)\alpha_1^2 + y4p_1q_1ad - \frac{1}{2}y^2\alpha_1^2 \\
 &\quad y^2(\alpha_1^2 - a^2) - y^3\alpha_1d + 2y(q_1 - p_1)2p_1q_1d^2 \\
 &\quad y^2(q_1 - p_1)^2d^2 + y^22p_1q_1d^2 \\
 &\quad y^3(q_1 - p_1)d^2 - \frac{1}{2}y^4d^2. \quad (5)
 \end{aligned}$$

Substituting previous definitions of σ_A^2 , α , σ_d^2 , $\text{cov}(a,d)$ into this, it becomes:

$$\begin{aligned}
 \sigma_{A_{RF}}^2 &= \sigma_{A_1}^2 \left[1 - \frac{y(q_1 - p_1)}{2p_1q_1} - \frac{y^2}{4p_1q_1} \right] - y^2(\alpha_1^2 - a^2) \\
 &\quad + y\text{cov}_1(ad) - y^3(d\alpha_1) + 2y(q_1 - p_1)\sigma_{d_1}^2 \\
 &\quad + y^2 \left[1 - \frac{(q_1 - p_1)^2}{2p_1q_1} \right] \sigma_{d_1}^2 - y^3 \left(\frac{q_1 - p_1}{2p_1q_1} \right) \sigma_{d_1}^2 \\
 &\quad \left(\frac{y^4}{4p_1q_1} \right) \sigma_{d_1}^2. \quad (6)
 \end{aligned}$$

It is convenient subsequently to leave the several σ_d^2 terms separated rather than gathered together.

We can note immediately that this equation converts the genic variance of one RF population (the focal parent) into the genic variance of another RF population (the p_{F_1} equivalent), which presages things to come. Notice that this genic variance absorbs several genetical terms, involving not only the *focal parent genic variance*, but also *cov(a,d)*, *average allele substitution effect* (α), *gene-model homozygote effect* (a), and much of the *gene-model heterozygote variance* (σ_d^2)*. Definitely, this is not simple in terms of the gene-model interpretation; but this difficulty extends to the widely accepted classical (RF) genic variance as well.

(ii) Dominance variance

$$\begin{aligned}
 \sigma_D^2(\text{RF}) &= [2p_{F_1}q_{F_1}]^2 d^2 \\
 &= 4p_1^2q_1^2d^2 - 2y(q_1 - p_1)2p_1q_1d^2 + y^2(q_1 - p_1)^2d^2 \\
 &\quad y^22p_1q_1d^2 + y^3(q_1 - p_1)d^2 + \frac{1}{4}y^4d^2. \quad (7)
 \end{aligned}$$

After substituting previous definitions of σ_D^2 and σ_d^2 this becomes:

$$\begin{aligned}
 \sigma_{D_{RF}}^2 &= \sigma_{D_1}^2 - 2y(q_1 - p_1)\sigma_{d_1}^2 - y^2\sigma_{d_1}^2 \\
 &\quad + y^2 \left[\frac{(q_1 - p_1)^2}{2p_1q_1} \right] \sigma_{d_1}^2 + y^3 \left(\frac{q_1 - p_1}{2p_1q_1} \right) \sigma_{d_1}^2 \\
 &\quad + \left(\frac{y^4}{8p_1q_1} \right) \sigma_{d_1}^2. \quad (8)
 \end{aligned}$$

As previously, it is convenient to leave the σ_d^2 terms separated.

Once again, we have a conversion equation from one RF (parental) to another RF (hybrid equivalent). Notice that this dominance variance collects together all of the *gene-model heterozygote variance* (σ_d^2) not absorbed into the genic variance. It can be shown that this occurs also in the classical definition of dominance variance.

(iii) Genotypic variance

Upon accumulating eqns (5) and (7), four cancellations occur (terms 2, 3, 4 and 5 of eqn 7 with respective terms 7, 8, 9 and 10 of eqn 5) and one half-cancellation occurs (the two last terms), leading to the RF genotypic variance:

*Define the gene-model heterozygote variance as $\sigma_d^2 = 2pqd^2$. Also define the gene-model homozygote variance as $\sigma_a^2 = 2pqa^2$.

$$\begin{aligned} \sigma_G^2(\text{RF}) = & 2p_1q_1\alpha_1^2 + y4p_1q_1ad - y(q_1 - p_1)\alpha_1^2 \\ & \frac{1}{2}y^2\alpha_1^2 - y^2(\alpha_1^2 - a^2) - y^3(d\alpha_1) \\ & + (2p_1q_1)^2d^2 - \frac{1}{4}y^4d^2. \end{aligned} \quad (9)$$

Comparison of eqn (9) with eqn (2) reveals that the genotypic variance (σ_G^2) of this RF population is *identical* with that of the allogamous F₂. That is, we have now proved that these two populations are the same for yet another key property. An implication is that the allogamous F₂ is an *origin* of RF populations. In the next section we will also make a comparison against *conventionally* obtained RF variances based on p_{F_1} , which will affirm this result.

Genic and dominance variances

Having shown that the allogamous F₂ and the equivalent RF are the same in several key properties, including the genotypic variance, it is obvious that the genic (σ_A^2) and dominance (σ_D^2) components of the RF population are also those of the allogamous F₂. Furthermore, these two components account for all of the genotypic variance, as expected from the *classical* RF variances (e.g. Falconer, 1981). Even some forms of epistasis may be included as part of the dominance/overdominance specifications. Therefore, eqn (5) or eqn (3) defines the *genic* variance of the allogamous F₂ (as well as of the equivalent RF), and eqn (7) or eqn (8) defines its *dominance* variance. These are visualized in Fig. 4 for our partial dominance example, where they are compared with the overall *genotypic* variance as well. Both the *complementary opposite* and the *fixed* $p_2 = \frac{1}{2}$ parental examples are depicted.

We noted earlier that *complementary opposites* genotypic variance was constant across all p_1 , owing to the result that $p_{F_1} = \frac{1}{2}$ in this situation. We see that the same outcome applies to the genic (σ_A^2) and dominance (σ_D^2) variances as well. The *fixed* $p_2 = \frac{1}{2}$ case reveals that the genic and dominance variances curve-shapes are reminiscent of a segment of the curves for classical RF populations. Indeed, this is exactly the situation. If we remember that the x -axis in these graphs is p_1 and not p_{F_1} , and that only a segment of the possible p_{F_1} spectrum can result from our example parentage, we then can observe the *fixed* $p_2 = \frac{1}{2}$ curves as a segment of the entire RF distribution based on its own p . This is shown in Fig. 5, where classical RF variances for our examples are graphed. The x -axis is now p_{RF} (i.e. p_{F_1} in our present context). The 'window' between the two vertical lines shows the same variances as in Fig. 4 for the *fixed* $p_2 = \frac{1}{2}$ case; and the vertical dashes at $p = 0.5$ show where lie the variances for the *complementary opposites* case. This brings all of our present variances into line with the more familiar RF variances.

Fig. 4 Genic variance (VA) and dominance variance (VD) components of the genotypic variance (VG) of the allogamous F₂ from crosses between a *focus* parent (A_1 frequency p_1), and *other* parent either with *complementary-opposite* p_2 , or with *fixed* $p_2 = 0.5$. Partial dominance is illustrated ($d = 7.5$ with $a = 10$).

Overdominance ($d > a$) has a marked effect on both genic and dominance variances. The dominance variance generally increases with increasing overdominance. However, the effect on the genic variance is more dynamic. For *complementary opposites* crosses, overdominance does not alter the genic variance, which is constant for all p_1 and d . For *other* kinds of parental relationships, the genic variance rises with increasing overdominance for $p_1 < \frac{1}{2}$, and drops with increasing overdominance for $p_1 > \frac{1}{2}$. With extreme overdominance, the genic variance is zero for some p_1 .

Heritabilities

As all the genetical variances of the allogamous F₂ and the RF population are identical, we can discuss heritabilities for both. The broad-sense heritability ($h_B^2 = \sigma_G^2/\sigma_P^2$, where $\sigma_P^2 = \sigma_G^2 + \sigma_E^2$, and $\sigma_E^2 = 25$ in the examples) generally increases with increasing overdominance. For *other parent fixed* p_2 crosses (as opposed to *complementary opposites*), there is a decline in broad-sense heritability as p_1 increases.

Narrow-sense heritabilities ($h_N^2 = \sigma_A^2/\sigma_P^2$) are of greater interest than broad-sense because of their prominence in selection theory, and because they exist for this F₂ case. Those for the *complementary opposites* crosses are

Fig. 5 Genotypic (VG), genic (VA) and dominance (VD) variances of a classical random fertilizing (RF) population across the Universe of allele frequency p , for partial dominance ($d = 7.5$ and $a = 10$). Allogamous F_2 form subsets of this Universe in which p is equivalent to p_{F_1} from a cross between a focal parent (with A_1 frequency p_1), and an other parent (with p_2 defined as an offset y from p_1). The values marked at $p = 0.5$ are those of the F_2 in which the other parent is a complementary opposite to the focal parent (over all p_1); while the window between the vertical lines is the sub-Universe for all p_1 crossed with fixed $p_2 = \frac{1}{2}$. These are actually the same variances as shown in Fig. 4, where, however, the x -axis is p_1 rather than p_{F_1} .

constant across all levels of p_1 , as expected from the constancy of the genic variances. Notice, however, that the level of this heritability drops with increasing dominance, because of the rising genotypic variance against the constancy of the genic variance (see earlier paragraph). For cases involving fixed other-parent p_2 , the levels of heritability have various patterns which reflect the underlying patterns of the genic variance (see earlier paragraph). Generally for these F_2 s, h_N^2 is higher for $p_1 < \frac{1}{2}$, indicating that genetic advance will be greater in such parent combinations. As noted earlier, some combinations of a , d and p_1 lead to zero genic variance, and hence zero h_N^2 and zero genetic advance from selection.

Discussion

Randomly fertilized populations

Quantitative genetics is built largely around the randomly fertilized (RF) population, or its inbred

derivatives. Also, it is often an added implication that the population is one RF gamodeme in which the frequencies of alleles are uniform over the entire area of the population (i.e. panmixia). Research on gene-flow (Levin and Kerster, 1974) has revealed that such a uniform one-gamodeme view is often unreasonable. For instance, of the examples tabulated in Richards (1986) (Table 5.10, p. 179), five-eighths of the examples definitely were *not* panmictic. However, an alternative realistic approach exists, i.e. to consider the population as a bulk of small dispersing RF gamodemes (Wright, 1943, 1946). This fits well with the natural evidence cited above. Nevertheless, we have shown here that the allogamous F_2 is a RF population, and that the assumption to this effect is valid! Furthermore, we have discovered a plausible origin for RF populations in the allogamous F_2 from hybrid bulks. We should realize that this RF may not necessarily be panmictic, however, and certainly not with the passage of subsequent generations when it will almost certainly become dispersed into smaller RF gamodemes.

Also, we now have the equations for converting one RF population (the focus parent) into another RF population (the derivative allogamous F_2). Such re-construction of RF populations via hybridization would be expected to occur commonly in natural genetics; and it is a frequently planned activity in plant breeding.

Inbreeding

It is widely assumed that a RF population has zero inbreeding. This is largely because of its being the philosophical base of Quantitative Genetics for which it may be defined as the "natural" reference for a definition of inbreeding. The problem is that, until now, we had no clear view of a natural history of which RF populations were a part. However, this paper has demonstrated that an allogamous F_2 is an origin for RF populations, thereby providing an external reference by which its inbreeding can be measured. The sexual pathway from F_1 to F_2 unequivocally shows that the natural reference population for the F_2 is its own F_1 ; which, in turn, has its properties set by the parents which crossed to produce it. This paper has derived the inbreeding of the allogamous F_2 (and equivalent RF population) with respect to that F_1 . On that basis it has considerable inbreeding (see Fig. 1), the value of which depends on the relatives mixture arising from the values of p_1 and p_2 in the original parents. We have already noted that, if we adopt the RF population as the reference base, then inbreeding is zero for the allogamous F_2 ; conversely, if we adopt the parental mean heterozygosity as the base, then this F_2 is outbred (with negative ϕ).

Another issue is that it is difficult to define inbreeding in terms of *full-sib*, *half-sib* or other relationships. These terms strictly apply only to the progeny of *individual* parents, and not to the bulk progeny of parental *populations*. The proportions of full-sib, half-sib, cousin-like and backcross-like progeny outcomes within the F_2 will vary considerably according to the original parent's p_1 or p_2 , making it difficult to resolve the inbreeding by this approach. The method used here (based on the net level of heterozygosity) is immediately exact and straightforward, and permits ready comparison with other kinds of populations, including all the classical situations mentioned above. For the *complementary opposites*, we have already noted that $\phi \rightarrow \frac{1}{2}$, the level for selfing (generation 1), as $p_1 \rightarrow 1$ (or 0). And for most crossing systems, there is a p_1 region for which $\phi \rightarrow 0$. The equivalent of full-sib (generation 1) ($\phi \rightarrow 0.25$) and of half-sib (generation 1) ($\phi \rightarrow 0.125$) inbreeding can also be observed at different p_1 depending on the crossing system (i.e. the relativity of p_1 and p_2).

Selection vs. hybrid vigour

Genetic advance (Δ_G) from forward selection increases with larger phenotypic variance and higher narrow-sense heritability (e.g. Falconer, 1981), as well as with stronger selection pressure. We now have the equations for these quantities in the allogamous F_2 , which enables us to estimate the relative effects of forward selection, hybrid vigour and inbreeding depression. In an earlier paper, similar properties were presented for the F_1 (Gordon, 1999): where it was noted that maximum hybrid vigour arose from crosses with extreme p_1 , whereas maximum selection advance arose from central values of p_1 . With *complementary opposites* crosses, all F_2 , irrespective of p_1 , will have the same genetic advance, as discussed earlier for genic variances. With F_2 from *fixed* p_2 crosses, greater selection advance occurs asymmetrically, for noncentral p_1 . The mean, cumulative genetic advance, and vigour for selection pressure of $P=0.1$ are shown in Fig. 6, using individual selection with both sexes equally selected, and with partial dominance ($a=10$, $d=7.5$). A *complementary opposites* cross, with $p_1=0.05$ is shown in Fig. 6(a), while a similar cross with $p_1=0.45$ is shown in Fig. 6(b). Concatenating the previous F_1 results with the present results, we would expect considerable hybrid vigour followed by some selection advance in the first case. In the second case, we expect minimum vigour followed by optimum selection advance.

In (a), hybrid vigour is observed, but it is followed by noticeable inbreeding depression from $F_1 \rightarrow F_2$. Selection begins also at this stage, but it barely offsets the effects of the depression. Finally, in selecting $F_2 \rightarrow F_3$,

Fig. 6 Generation performance from *focus parent* (P_1), through *hybrid* (F_1) and *allogamous* F_2 (F_2) to the allogamous F_3 , with individual selection ($P=0.1$) at $F_1 \rightarrow F_2$ and $F_2 \rightarrow F_3$, under the partial dominance and environment variance defined previously. The mean and its components [accumulated selection advance (delta-G), and hybrid vigour from the mid-parent (vigour)] are given for two parental situations: (a) *focal* $p_1=0.05$, and (b) $p_1=0.45$. The other parent in each case is a *complementary opposite*.

some real advance is made from the selection. With the case in (b), both vigour and depression are trivial, and both stages of selection yield relatively large advances. The final result at F_3 is much better in case (b) than in case (a). These results are caused *solely* by the different

properties of F_1 and/or F_2 for different regions of the p_1 range. It appears that, in these examples, selection was more effective than hybrid vigour.

Where overdominance occurs, hybrid vigour is more conspicuous, but so is subsequent depression. Selection advance also is greater with overdominance, and always overcomes the depression to make subsequent substantial advance. In addition, if we *utilize* dispersion by employing selection strategies such as Combined Selection (the best individuals from the best lines), the advantage to selection becomes much greater still.

References

- ALLARD, R. W. 1960. *Principles of Plant Breeding*. Wiley & Sons, New York.
- FALCONER, D. S. 1981. *Introduction of Quantitative Genetics*, 2nd edn. Longman, New York.
- GORDON, I. L. 1999. Quantitative genetics of intraspecies hybrids. *Heredity*, **83**, 757–764.
- LEVIN, D. A. AND KERSTER, H. W. 1974. Gene flow in seed plants. *Evol. Biol.*, **7**, 139–220.
- MOORE, J. N. AND JANIK, J. (eds). 1983. *Methods in Fruit Breeding*. Purdue University Press, West Lafayette, IN.
- POEHLMAN, J. M. 1979. *Breeding Field Crops*, 2nd edn. AVI Publishing, Westport.
- RICHARDS, A. J. 1986. *Plant Breeding Systems*. Allen & Unwin, London.
- WRIGHT, S. 1921. Systems of mating. I. The biometrical relations between parent and offspring. *Genetics*, **6**, 111–123.
- WRIGHT, S. 1943. Isolation by distance. *Genetics*, **28**, 114–138.
- WRIGHT, S. 1946. Isolation by distance under diverse systems of mating. *Genetics*, **31**, 39–59.