The struggle to exploit non-additive variation

Bruce Walsh

Departments of Ecology and Evolutionary Biology, Molecular and Cellular Biology, Animal Science, Plant Science,
Epidemiology and Biostatistics, University of Arizona, Tucson, AZ 85721, USA. Email: jbwalsh@u.arizona.edu

Abstract.   Whereas animal breeders largely focus on improvement using additive genetic variance, inbreeding and asexual reproduction allow plant breeders to at least partially exploit non-additive genetic variance as well. We briefly review various approaches used by breeders to exploit dominance and epistatic variance, discuss their constraints and limitations, and examine what (if anything) can be done to improve our ability to further use often untapped genetic variation.

Additional keywords: dominance, epistasis, epistatic genetic variance, inbreeding.

Introduction

Plant and animal breeding are sister disciplines united by the common framework of quantitative genetics but separated by historical differences in notation (although this is lessening) and (more centrally) by biological constraints imposed by the organism under selection. Animal breeders have the luxury of large individuals in which traits are (generally) relatively easy to measure and which often have at least modest individual (narrow-sense) heritabilities. These advantages are offset by low reproductive potential and strict outcrossing. Plant breeders, on the other hand, are blessed with much higher reproductive potentials and often the ability to easily self. Countering these reproductive advantages, it is generally more difficult to measure traits in single individuals and such traits often have lower individual heritabilities due to (among other things) higher sensitivity to environmental influences and competitive effects.

As a result of these strengths and weaknesses of each discipline, animal breeding is largely an additive enterprise, with the focus on keeping detailed pedigree information to maximally exploit additive variation through BLUP (best linear unbiased predictors). Conversely, certain crossing schemes allow plant breeders to potentially tap into at least some of the non-additive genetic variation in a population.

The Basic Model: additive and non-additive genetic contributions

Both plant and animal breeders start with Fisher's (1918) decomposition of the phenotypic value into genetic and environmental components, where the observed phenotypic value P of an individual is the sum of its genetic and environmental values:

$$P = G + E \tag{1}$$

Fisher had 2 fundamental insights in his 1918 paper (beyond coining the term variance and inventing the analysis of variance machinery). His first was that, for any given locus, (diploid) parents pass along only 1 of their 2 alleles to their offspring. As a result, parents do not pass along their entire genotypic value G to their offspring, but rather pass along only components of G that arise from the average contributions of single alleles at each locus. For outcrossing populations, Fisher showed (and was later extended by both Kempthorne and Cockerham) that we can further decompose an individual's genotypic value into additive (A), dominant (D), and epistatic (I) contributions:

$$G = A + D + I \tag{2a}$$

Letting α_i be the average contribution to the genotypic value in an individual that inherits a copy of allele i, then the sum of the α_i over all loci is the additive component A (often called an individual's breeding value). The α_i are obtained by a least-squares regression weighted by genotypic frequencies and hence change as allele (and genotype) frequencies change. Since A is the average contribution from single alleles, it should not be surprising that the expected offspring mean from 2 parents is simply the average of their breeding values. Thus A measures the portion of a parent's genotypic value that is passed onto their offspring under random mating. The non-additive genetic components (D and I) measure departures of particular combinations of alleles over the genotypic value expected simply from summing the average value of each allele. For a single locus, the dominance contribution from a locus with alleles i and j is $\delta_{ij} = G_{ij} - \alpha_i - \alpha_j$, the genotypic value for that locus minus the average contributions of both alleles. Thus dominance contributions are due to the interaction between 2 alleles resulting in the genotype having a different genotypic value

than predicted from summing the average effects of both individual alleles. In order for such a contribution to be passed along, a parent must pass along both its alleles to an offspring. Under random mating, a single (diploid) parent only contributes one allele per locus and hence cannot pass along any of its dominance component to its offspring. As we will see, this is not the case with either inbreeding or with autoploids. The epistatic contribution I arises when the value of a multilocus genotype is not simply the sum of the average single locus genotypic values. For example, with 2 loci, epistasis arises when $G_{ijkl} \neq G_{ij} + G_{kl}$. As shown by Cockerham (1954) and Kempthorne (1954), the epistatic contributions can be further subdivided as:

$$I = AA + AD + DD + AAA + AAD + \cdots A^u D^v + \quad (2b)$$

For example, the additive $\times$ additive (AA) component arises when epistatic contributions are generated by interactions between 2 alleles at different loci, whereas AD arises from the interactions of a single allele at one locus with the genotype at another locus. In general, $A^u D^v$ epistasis implies interaction of u single alleles with the diploid genotypes at v other loci, or interactions involving $u + 2v$ alleles over $u + v$ loci. The faithful transmission of an epistatic component to its offspring thus requires the transmission of all the interacting alleles in a gamete. Since parents pass along single alleles from each locus, fractions of additive epistasis (e.g. AA, AAA, etc.) can be passed from (diploid, random-mating) parents to their offspring, but no contributions involving 2 alleles from the same locus case (e.g. any term in I containing D). For example, AA contributions require alleles from 2 different loci to be jointly passed along to offspring. Although this can certainly happen under random mating, these alleles must remain associated for the extra AA contribution to persist. As these alleles at different loci are randomised by recombination with respect to each other, this contribution decays away to zero (see below).

In random mating populations, only a fraction of favourable genotypic values is passed onto the offspring of selected parents, and many (perhaps most) are destroyed by reproduction. Thus, one struggle to exploit non-additive genetic effects is to somehow allow selected parents to contribute more to their offspring than just their A values. This means the ability to pass along not just single alleles, but coordinated groups of alleles at different loci (for AA and higher additive epistatic effects) and whole genotypes at single (or more) loci (for D and epistatic terms involving D). In particular, to pass along a term of $A^n D^m$ requires that a parent pass along particular alleles at n loci and both alleles at another m loci.

Fisher's second fundamental insight was that the genetic covariances between different sets of relatives can be expressed in terms of the genetic variance components associated with the various terms in Eqn 2. By construction (under random mating and linkage equilibrium) the genetic components are uncorrelated with each other, and hence the genetic variance is simply the sum of the component variances:

$$\begin{aligned}
\sigma_G^2 &= \sigma_A^2 + \sigma_D^2 + \sigma_I^2 \\
&= \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 \\
&\quad + \sigma_{AD}^2 + \sigma_{DD}^2 + \cdots + \sigma_{A^u D^v}^2 + \cdots \quad (3)
\end{aligned}$$

The genetic covariance between 2 relatives x and y (in an outbred population under linkage equilibrium and random mating) is a function of 2 genetic identity coefficients: the coefficient of co-ancestry r_{xy} (the probability that 2 random alleles, one drawn from x, the other from y, are identical by descent) and the coefficient of fraternity Δ_{xy} (the probability that x and y share both alleles identical by descent at a locus), with:

$$\begin{aligned}
\sigma_G(x, y) &= 2r_{xy}\sigma_A^2 + \Delta_{xy}\sigma_D^2 + (2r_{xy})^2\sigma_{AA}^2 + 2r_{xy}\Delta_{xy}\sigma_{AD}^2 \\
&\quad + \Delta_{xy}^2\sigma_{DD}^2 + (2r_{xy})^3\sigma_{AAA}^2 + \cdots \\
&= \sum (2r_{xy})^u \Delta_{xy}^v \, \sigma_{A^u D^v}^2 \quad (4)
\end{aligned}$$

The coefficient of fraternity can be expressed in terms of the coefficient of co-ancestry of the parents of x (f_x, m_x for the coefficients of the father and mother of x) and y (f_y, m_y):

$$\Delta_{xy} = r_{(f_x f_y)}r_{(m_x m_y)} + r_{(f_x m_y)}r_{(m_x f_y)} \quad (5)$$

In order for Δ_{xy} to be non-zero, both sets of parents for x and y must be related. The critical feature for Eqn 3 for exploiting non-additive variance is that to pass along any contribution involving dominance (either straight dominance or epistatic terms with a dominance contribution), the relatives used from the selected parents must have a non-zero coefficient of fraternity with the selected individuals. Operationally, this means that both parents of x and y must be related. Whereas full sibs is one such case, parents and their offspring are not unless the parents are inbred. Hence, inbreeding is a key practical component to using non-additive variance to enhance the selection response.

The Breeders' Equation and beyond

Equation 4, and its extensions (e.g. Eqn 17), provide the key to exploiting non-additive variance and (more importantly) to predicting response for a given selection scheme. As the first step of development, consider the simple case of exploiting the additive variance. With a random mating population, the expected response in the next generation (i.e. the change in the population mean) due to selection on parents is given (in the simplest case) by the Breeders' Equation:

$$R = h^2 S = \frac{\sigma_A^2}{\sigma_P^2} S \quad (6)$$

where S is the within-generation change in the trait value between the selected parents and the trait mean before selection and h^2 is the narrow-sense heritability (the fraction of total phenotypic variance σ_P^2 due to variance in additive effects σ_A^2). The Breeders' Equation follows from the slope of a midparent–offspring regression, which is simply h^2 in the absence of epistasis. The Breeders' Equation offers the simplest situation for selection in the additive world. Another way to obtain Eqn 6 is that if we only know the phenotypic value P of an individual, the best estimate of its breeding value is:

$$\hat{A} = h^2(P - \mu)$$

where μ is the population mean. The expected mean of offspring from 2 parents is simply the average of their breeding values, which recovers the Breeders' Equation noting that $S = \bar{P} - \mu$, where $\bar{P}$ is the mean of the selected parents. The narrow-sense heritability h^2 is a measure of the accuracy in predicting parental breeding values if we only know their phenotypic values. We can increase the accuracy of estimating an individual's breeding value by also incorporating phenotypic information from its relatives. Animal breeders do this by keeping detailed pedigrees and using linear models (in particular Henderson's BLUP methodology, reviewed in Lynch and Walsh 1998) that fully account (and exploit) the covariances in breeding values between sets of known relatives. Hence, much of animal breeding is simply a quest to maximise the A values of selected parents, as this is the only genetic component that is passed on intact to offspring under random mating.

When additive epistasis (i.e. *AA*, *AAA*, etc.) is present, this inflates the selection response. Under random mating, the response to a single generation of selection becomes:

$$R = h^2 S + \frac{S}{\sigma_P^2}\left(\frac{\sigma_{AA}^2}{2} + \frac{\sigma_{AAA}^2}{4} + \frac{\sigma_{AAAA}^2}{8} + \cdots\right) \quad (7)$$

which can deviate significantly from $h^2 S$. The reason for the additional terms is that the parent–offspring genetic covariance involves not only additive values, but additive epistasis as well, as parents can pass along combinations of single alleles from many loci to their offspring:

$$\frac{\sigma(G_p, G_o)}{\sigma_P^2} = \frac{h^2}{2} + \frac{1}{\sigma_P^2}\left(\frac{\sigma_{AA}^2}{4} + \frac{\sigma_{AAA}^2}{8}\right.$$
$$\left. + \frac{\sigma_{AAAA}^2}{16} + \cdots + \frac{\sigma_{A^u}^2}{2^u}\right) \quad (8)$$

Although the presence of additive epistasis can inflate the selection response over what is expected from the additive variance alone, this is a transient increase, as recombination breaks down favourable combinations of alleles. After one generation, only 1/2 of pairwise combinations (*AA*) of favourable alleles generated by selection remain intact, and only $1/2^u$ of *u*-wise combinations (A^u) of alleles are present. Hence, the contribution to the current response from one generation of selection τ generations ago is just:

$$S\left(h^2 + \frac{1}{2^\tau}\frac{\sigma_{AA}^2}{2\sigma_P^2}\right) \quad (9)$$

which converges to $h^2 S$. Contributions from higher-order additive epistasis decay as $(1/2)^{u\tau}$ and hence are removed even quicker. Hence, under random mating, additive contributions are stable, whereas additive epistatic contributions quickly decay away.

An analogous situation occurs with selection on autotetraploids (Gallais 1975, 2003). With autotetraploids, the parent–offspring covariance when dominance (but no epistasis) is present is:

$$\sigma(G_p, G_o) = \frac{\sigma_A^2}{2} + \frac{\sigma_D^2}{6} \quad (10)$$

The inclusion of the dominance term arises because autotetraploids pass along pairs of alleles from each locus (rather than singletons) and hence D as well as A. This parent–offspring covariance results in a single-generation response to selection of:

$$R = S\left(h^2 + \frac{\sigma_D^2}{3\sigma_P^2}\right) \quad (11a)$$

The response after t generations of selection with constant differential S can be written as the sum of a stable additive component and a transient dominance component:

$$R(t) = th^2 S + R_D(t) \quad (11b)$$

where

$$R_D(t) = S\frac{3}{2}[1 - (1/3)^t]\frac{\sigma_D^2}{3\sigma_P^2} \quad (11c)$$

which converges to $S(\sigma_D^2/2\sigma_P^2)$. Segregation reduces the departure from tetraploid Hardy–Weinberg proportions generated by the selection of favourable combinations of allelic pairs, reducing their contribution to response. The response for t generations of selection followed by τ generations of no selection is:

$$th^2 S + (1/3)^\tau R_D(t)$$

which again converges to $th^2 S$, the additive component of response. With pair-wise epistasis:

$$\sigma(G_p, G_o) = \frac{\sigma_A^2}{2} + \frac{\sigma_D^2}{6} + \frac{\sigma_{AA}^2}{4} + \frac{\sigma_{AD}^2}{12} + \frac{\sigma_{DD}^2}{36}$$

Thus, with autotetraploids, dominance epistasis can be passed from parent to offspring. As was the case for diploids, the contribution from the epistatic terms decays at rate

$(1/2)^t$ (for unlinked loci), plus corrections for the decay of homozygotes passed along intact (i.e. D). As a reminder, with autotetraploids, there are 4 single locus genetic parameters, terms for 3- and 4-allele effects in addition to the standard single (A) and 2 (D) allele effects (Lynch and Walsh 1998; Gallais 2003). The epistatic terms for even a pair of loci represent all combinations of these 4 parameters, and hence only a fraction of the possible epistatic terms is included in the parent–offspring covariance.

Given this decay in response, one scheme for improving ultimate response is to minimise the number of generations of recombination between the selected lines and the material that is ultimately sold as seed. This scheme is well appreciated by breeders working with polyploids, such as Lucerne (alfalfa).

Predicting the genetic component of response: ancestral regressions

A general approach for examining which components of the response are transient is to consider the expected value of an offspring as a function of all its direct relatives that have been under selection. If this ancestral regression is linear (as would occur if the joint distribution of the phenotypic values of all relatives is multivariate normal), response can be described by specifying the regression coefficients by an obvious extension of the biparental regression to now include all selected relatives back to the original unselected base population (Bulmer 1980). For example, if selection starts in generation 0, the response in the first generation is:

$$R(1) = 2\beta_{1,0}S_0$$

where $\beta_{1,0}$ is the regression of offspring at generation 1 on a parent from generation 0 (this assumes that both parents have the same regression coefficients and selection differentials, an assumption that will be relaxed shortly). Likewise, the total response after 2 generations:

$$R(2) = 4\beta_{2,0}S_0 + 2\beta_{2,1}S_1$$

depends on the nature of selection on the 4 grandparents and both parents. Note that this formulation allows the parent–offspring regression to change through time (e.g. $\beta_{2,1}$ need not equal $\beta_{1,0}$), as might happen with inbreeding. Similarly, the response following 3 generations of selection depends upon the nature of selection on that individual's 8 great-grandparents, 4 grandparents, and 2 parents:

$$R(3) = 8\beta_{3,0}S_0 + 4\beta_{3,1}S_1 + 2\beta_{3,2}S_2$$

Proceeding in this fashion gives the response for generation T as:

$$R(T) = \sum_{t=0}^{T-1} \beta_{T,t}\,(2^{T-t}S_t) \qquad (12a)$$

where $\beta_{T,t}$ is the regression coefficient for the phenotype of an individual in generation T on one of its relatives in generation $t < T$. With pure selfing, each individual has only a single relative in each previous generation, giving the ancestral regression as:

$$R(T) = \sum_{t=0}^{T-1} \beta_{T,t}S_t \qquad (12b)$$

Recall from standard regression theory that the vector of partial regression coefficients $\boldsymbol{\beta} = \mathbf{V}^{-1}\boldsymbol{\sigma}$, where $\boldsymbol{\sigma}$ is a vector of covariances between the individuals in generation T with all relatives in previous generations and $\mathbf{V}$ is the phenotypic covariance matrix for the entire collection of individuals. The key here is that the regression coefficients are entirely determined by the covariances between relatives. If we have independence so that the partial regression coefficients reduce to univariate regression coefficients (i.e. $\beta_i = \sigma(y, x_i)/\sigma_{x_i}^2$), then we have:

$$R(T) = \sum_{t=0}^{T-1} \frac{\sigma_G(T, t)}{\sigma^2(P_t)} \,(2^{T-t}S_t) \qquad (12c)$$

where $\sigma_G(T, t) = \sigma(P_T, P_t)$ is the cross-generation covariance, the phenotypic covariance between an individual in generation t and its descendent in generation $T > t$. With selection under pure selfing, each individual has a single ancestor and the 2^{T-t} term in Eqn 12c is absent.

If different relatives in the same generation experience different amounts of selection, with $S_{t,i}$ being the selection differential on relative i in generation t, then:

$$R(T) = \sum_{t=0}^{T-1} \left[\beta_{T,t}\left(\sum_{i=1}^{n(t,T)} S_{t,i} \right) \right] \qquad (13)$$

where $n(t, T)$ is the number of relatives in generation t that contribute to response in generation T. Note for the case of pure selfing $n(t, T) = 1$. Finally, we can also allow for different regression coefficients on each relative to completely generalise this approach:

$$R(T) = \sum_{t=0}^{T-1} \left(\sum_{i=1}^{n(t,T)} \beta_{T,t,i}S_{t,i} \right) \qquad (14)$$

where $\beta_{T,t,i}$ is the regression coefficient of the phenotype on an individual in generation T on its i-th relative in generation t.

To apply ancestral regression for predicting response, we require that the regression remains linear and that selection-induced changes in the variance and covariances are negligible. Thus, although we allow changes in $\beta_{T,t}$ due to the particular genetic system being considered (e.g. selfing wherein the additive genetic variance decreases by a predictable amount each generation in the absence of

selection) we assume that selection does not confound these changes. Bulmer (1980) shows that under the infinitesimal model the joint distribution of an offspring and all its direct ancestors is multivariate normal and hence the ancestral regression is linear. Since selection does not change allele frequencies under the infinitesimal model (reviewed in Chapters 5 and 15 of Walsh and Lynch 2005), this might suggest that the regression coefficients $\beta_{T,t}$ are unaffected by selection. The problem, however, is that selection generates gametic-phase disequilibrium that can significantly alter the genotypic moments (reviewed in Walsh and Lynch 2005).

As an application of ancestral regressions, consider additive × additive epistasis. In this case, Cockerham (1984*b*) found that for 2 linked loci (recombination frequency *c*), the cross-generation covariance between a parent in generation τ and its descendants in generation $\tau + t$ is:

$$\sigma_G(\tau + t, \tau) = \frac{\sigma_A^2(\tau)}{2^t} + \frac{\sigma_{AA}^2(\tau)}{2}\left(\frac{1-c}{2}\right)^t$$

giving:

$$2^t\sigma_G(\tau + t, \tau) = \sigma_A^2(\tau) + (1-c)^t\frac{\sigma_{AA}^2(\tau)}{2}$$

If the genetic variances remain constant, then applying Eqn 12*a* we recover, and generalise, Eqn 9.

The behaviour of the regression coefficients over time thus informs us about the permanency of response. Note from Eqn 12*a* that unless $2^t\beta_{\tau+t,\tau}$ remains constant as *t* increases, the contribution to cumulative response from selection on adults in generation *t* changes over time. For example, when loci are strictly additive (no dominance or epistasis), $\sigma_G(\tau + t, \tau) = 2^{-t}\sigma_A^2(\tau)$ and thus $2^t\beta_{\tau+t,\tau} = h_\tau^2$, the standard result from the Breeders' Equation. Note that unless $2^t\sigma_G(\tau + t, \tau)$ remains constant, any response contributed decays. Hence, any term of $\sigma_G(\tau + t, \tau)$ that decreases by more than 1/2 each generation, contributes only to the transient response. An exception is with pure selfing where the total contribution in generation *t* from an ancestor τ previous generations is $\sigma_G(\tau + t, \tau)$, so that any components that decline as τ increases will contribute to the transient response.

Covariances between relatives with non-additive variance and inbreeding

As discussed above, one approach to exploiting non-additive variance is to use inbred individuals. The method of ancestral regression can be used to predict selection response in this case, provided we have a general expression for the genetic covariance between inbred relatives. Even with just simple dominance, this is a complex problem (Harris 1964; Cockerham 1983; Cornelius 1988; Edwards and Lamkey 2002). Consider the genotypic value at a single locus, where

p_i is the frequency of allele *i*. Fisher's genetic decomposition in a random mating population is:

$$G_{ij} = \mu + \alpha_i + \alpha_j + \delta_{ij}$$

where the average effects (α) and dominance deviations (δ) are obtained from a least-squares fit of the underlying gene effect and hence functions of the allele frequencies. The additive and dominance variances are just:

$$\sigma_A^2 = 2\sum_i p_i\alpha_i^2, \qquad \sigma_D^2 = \sum_i\sum_j p_i p_j\delta_{ij}^2 \qquad (15)$$

In a random-mating population, α and δ are uncorrelated and hence σ_A^2 and σ_D^2 are sufficient (in the absence of epistasis) to compute any covariance between sets of relatives.

With inbreeding, σ_A^2 and σ_D^2 are no longer sufficient and at least 3 additional terms are required to describe the covariances of inbred relatives for the simple case of a single locus with dominance. The first new term, σ_{ADI}, is the covariance between an additive effect and the dominance deviation for its corresponding homozygote:

$$\sigma(\alpha_i, \delta_{ii}) = \sigma_{ADI} = \sum_i p_i\alpha_i\delta_{ii} \qquad (16a)$$

This term enters because even though an inbred parent passes along only a single allele to its offspring, in an inbred individual a disproportionate fraction (relative to random mating) of allele *i* is in homozygotes and σ_{ADI} corrects for this excess. The second term ι^*, the squared sum of the homozygous dominance deviations, is the square of the inbreeding depression effect:

$$\iota^* = \left(\sum_i p_i\delta_{ii}\right)^2 \qquad (16b)$$

The final term, σ_{DI}^2, is the variance of dominance deviations in fully inbred individuals:

$$\sigma_{DI}^2 = \sum_i p_i\delta_{ii}^2 - \left(\sum_i p_i\delta_{ii}\right)^2 = \sum_i p_i\delta_{ii}^2 - \iota^* \quad (16c)$$

The correction of ι^* is required as the expected value of the δ (when restricted to just homozygotes) is not necessarily zero. Whereas σ_{DI}^2 and ι^* are non-negative, σ_{ADI} (being a covariance) can be negative, and indeed in the few cases where this has been estimated, it is often substantially negative (reviewed in Edwards and Lamkey 2002).

Simplification is possible under certain circumstances. For example, with only 2 alleles per locus, $\iota^* = \sigma_D^2$, and if all alleles have frequency 0.5, as in a cross between 2 pure lines, $\sigma_{DI}^2 = \sigma_{ADI} = 0$. Adding additive × additive epistasis is straightforward, but a full accounting of all possible pair-wise epistatic components (e.g. *AD* and *DD*) requires 12 components (Wright 1987), and this does

not even consider linkage disequilibrium. Hence, dealing with simple dominance alone is complex enough, and essentially no extensive work has been done allowing for dominance epistasis.

Recalling the above definitions for σ_{ADI}, σ_{DI}^2, and ι^*, the covariance between inbred relatives with dominance and additive epistasis (but no linkage disequilibrium) is:

$$\sigma_G(x, y) = 2r_{xy}\sigma_A^2 + \Delta_7\sigma_D^2 + \Delta_1\sigma_{DI}^2$$
$$+ (4\Delta_1 + \Delta_3 + \Delta_5)\sigma_{ADI} + (\Delta_2 - f_x f_y)\iota^*$$
$$+ (\tilde{f}_{xy} + \tilde{\gamma}_{\tilde{x}y} + \tilde{\gamma}_{x\tilde{y}} + \tilde{\Delta}_{\tilde{x}\tilde{y}})\sigma_{AA}^2 \qquad (17a)$$

Here, $2r_{xy}$ is the coancestry coefficient, f_x the inbreeding coefficient for individual x, the Δ_i are various pair-wise identity coefficients (see fig. 7.2 in Lynch and Walsh 1998), and the other higher-order identity coefficients ($\tilde{f}, \tilde{\gamma}, \tilde{\Delta}$) are defined in detail in Ch. 3 of Walsh and Lynch (2005). Note that the genetic variance in the population at any time is just the covariance of an individual with itself, $\sigma_G(x, x)$, and hence:

$$\sigma_{P_x}^2 = \sigma_G(x, x) + \sigma_e^2 \qquad (17b)$$

Equations 17a and b, given the appropriate identity coefficients, provide the required variances and covariances to apply ancestral regressions to predict selection response.

Drift and epistasis

One immediate application of Eqn 17a is the within-line additive genetic variance under inbreeding. Although it is well known that the contribution from additive variance within an inbred line declines by $(1 - f)\sigma_A^2$, it is less well appreciated that when non-additive gene action is present, inbreeding can actually increase the additive genetic variance. This was first pointed out (for dominance) by Alan Robertson (1952). More recent extensions have included additive × additive epistasis (Cockerham and Tachida 1988; Tachida and Cockerham 1989; Cheverud and Routman 1996). Modest levels of inbreeding can 'convert' some of the non-additive variance components in a random mating population into additive variance within an inbred line. In particular, the within-line additive genetic variance can be expressed as:

$$(1 - f)\sigma_A^2 + 2[f - \gamma - 2(\Delta - \delta)]\sigma_D^2 + 2(f - \gamma)\sigma_{ADI}$$
$$+ 2(\gamma - \delta)\sigma_{DI}^2 + 2(\gamma - \Delta)\iota^*$$
$$+ (4f - \tilde{f} - 2\tilde{\gamma} - \tilde{\Delta})\sigma_{AA}^2$$

Recursion equations for γ, Δ, and the other identity parameters can be found in Cockerham (1984a) and Walsh and Lynch (2005). Although inbreeding does indeed reduce the contribution from σ_A^2, this can be offset by positive contributions for σ_D^2, σ_{DI}^2, ι^*, and σ_{AA}^2. Contributions also arise from σ_{ADI}, but these can be positive or negative.

Thus, at least in theory, one can exploit some non-additive variance by modest inbreeding and within-line selection. The rub is that the optimal level of inbreeding to maximise the within-line additive variance depends on knowledge of difficult-to-estimate parameters (σ_D^2, σ_{DI}^2, ι^*, σ_{ADI}, σ_{AA}^2).

Prediction of response under selfing

For a parent x in generation t and its selfed descendant y in generation j, the identity coefficients in Eqn 17a reduce to simple functions of the inbreeding coefficient f:

$$\sigma_G(y_j, z_t) = (1 + f_t)\sigma_A^2 + (1 - f_j)(\sigma_D^2 + f_t\iota^*)$$
$$+ \frac{f_j + 3f_t}{2}\sigma_{ADI} + f_t\sigma_{DI}^2 + (1 + f_t)^2\sigma_{AA}^2 \qquad (18a)$$

where $f_t = 1 - (1/2)^t$. Likewise, the genetic variance in the population (the collection of lines) is just:

$$\sigma_G(x_t, x_t) = (1 + f_t)\sigma_A^2 + (1 - f_t)(\sigma_D^2 + f_t\iota^*)$$
$$+ 2f_t\sigma_{ADI} + f_t\sigma_{DI}^2 + (1 + f_t)^2\sigma_{AA}^2 \qquad (18b)$$

As an application of Eqns 18a and b, consider the selection response when non-inbred parents are selected and then selfed to fixation to form a new series of lines. Two measures of response are of concern here: the response after a single generation of selfing and the response remaining after selfing is complete. The genetic covariance between an outbred parent and a selfing offspring ($f_t = f_0 = 0$, $f_j = f_t = 1/2$) is:

$$\sigma_G(y_1, x_0) = \sigma_A^2 + \frac{1}{2}\sigma_D^2 + \frac{1}{4}\sigma_{ADI} + \sigma_{AA}^2 \qquad (19a)$$

and the eventual genetic covariance between the selected non-inbred parent and its completely selfed great-offspring ($f_t = 0$, $f_j = 1$) is:

$$\sigma_G(y_\infty, x_0) = \sigma_A^2 + \frac{1}{2}\sigma_{ADI} + \sigma_{AA}^2 \qquad (19b)$$

Note that the additive contributions (σ_A^2, σ_{AA}^2) remain unchanged, but that the initial contribution from the dominance variance (σ_D^2) decays to zero, whereas the contribution from σ_{ADI} actually doubles. Hence, the permanent response when selected (outbred) parents are selfed to fixation is:

$$R = \left(\frac{\sigma_A^2 + (1/2)\sigma_{ADI} + \sigma_{AA}^2}{\sigma_A^2 + \sigma_D^2 + \sigma_E^2}\right) S$$
$$= \left(h^2 + \frac{(1/2)\sigma_{ADI} + \sigma_{AA}^2}{\sigma_A^2 + \sigma_D^2 + \sigma_E^2}\right) S \qquad (20)$$

Under this scheme, response is inflated by $(\sigma_{ADI}/2 + \sigma_{AA}^2)/\sigma_P^2$, but at the cost of selfing the progeny

to fixation. One might use double-haploids (DH) to speed this process along, but the use of DH lines locks in any linkage disequilibrium generated by selection, lowering the response.

To put some numbers on these equations, consider the results of Edwards and Lamkey (2002) who obtained the following estimates for grain yield in a synthetic maize population: $\sigma_A^2 = 0.29$, $\sigma_D^2 = 0.32$, $\sigma_{ADI} = -0.18$, $\sigma_{DI}^2 = 0.85$, $\iota^* = 1.55$, and (approx.) $\sigma_e^2 = 1.1$. Their crossing scheme was not designed to detect epistasis. With these parameter values, the permanent response from a single generation of selection with outbred offspring is $h^2 S = 0.17 \cdot S$, whereas the permanent response if offspring are instead inbred to fixation is significantly less, $0.117 \cdot S$, due to the negative value of σ_{ADI}. Of course, the presence of additive × additive epistasis may inflate the inbred response. Nonetheless, the important message is that inbreeding to exploit non-additive variance can actually reduce the response.

More generally, the covariance for a parent with inbreeding level f and a fully selfed descendant is:

$$\sigma_G(y_\infty, x_f) = (1 + f)\sigma_A^2 + \frac{1 + 3f}{2}\sigma_{ADI}$$
$$+ f\sigma_{DI}^2 + (1 + f)^2 \sigma_{AA}^2 \qquad (21a)$$

Thus, when the parent is itself already inbred, σ_{DI}^2 enters as a permanent contribution. Higher order additive epistasis enters as $(1 + f)^u \sigma_{A^u}^2$, and potentially can be significant. Eqn 21a implies that the genetic variance between a group of completely selfed lines approaches:

$$\sigma_G(x_t, x_t) = 2\sigma_A^2 + 2\sigma_{ADI} + \sigma_{DI}^2 + 4\sigma_{AA}^2 + \cdots + 2^u \sigma_{A^u}^2 \qquad (21b)$$

This is in contrast to the standing genetic variance in a random-mating population, $\sigma_A^2 + \sigma_D^2$. Using the Edwards and Lamkey estimates, the between-line variance when $f = 1$ is 175% the variance under random mating (1.07 *v.* 0.6), again ignoring any additive epistasis.

The response for multiple generations of joint selection and selfing is a little more complex, as the covariances change each generation, giving the response after T generations as:

$$R(T) = \sum_{t=0}^{T-1} S_t \frac{\sigma_G(T, t)}{\sigma_G(t, t) + \sigma_e^2} \qquad (22)$$

For example, the response after 2 generations of inbreeding and selection is:

$$R(2) = S_0 \frac{\sigma_G(2, 0)}{\sigma^2(z_0)} + S_1 \frac{\sigma_G(2, 1)}{\sigma^2(z_1)} \qquad (23a)$$

The first term represents the response that carries over to the second generation from selection in generation 0, and the second term is the response to selection from generation 1. Using the maize yield estimates of Edwards and Lamkey, the response after 1 generation is:

$$R(1) = \frac{\sigma_G(1, 0)}{\sigma^2(z_0)} S_0 = 0.237 \cdot S_0$$

and after 2 generations:

$$R(2) = S_0 0.177 + S_1 0.400$$

If we stop selection after 2 generations, but continue to inbreed the population to complete homozygosity, the final response (after correcting for any inbreeding depression) is:

$$R_\infty(2) = S_0 \frac{\sigma_G(\infty, 0)}{\sigma^2(z_0)} + S_1 \frac{\sigma_G(\infty, 1)}{\sigma^2(z_1)} \qquad (23b)$$

Using the Edwards and Lamkey estimates:

$$R_\infty(2) = S_0 0.117 + S_1 0.273$$

or only 50% of the initial response from the first generation of selection and 68% of the initial second-generation response. Inspection of Eqn 23a and b points out a key feature of response with inbreeding. In most cases, these covariances change, so that it is generally the case that $\sigma_G(i, t) \neq \sigma_G(j, t)$. Thus, the relative contribution to response from selection in any particular generation t changes over time, so that there is both a transient and permanent component to response. Prediction of this response requires detailed knowledge of difficult-to-estimate components $(\sigma_{DI}^2, \sigma_{ADI}, \iota^*)$.

Line crosses and heterosis

Perhaps the most successful exploitation of non-additive effects occurs with hybrid breeding, where heterosis is often seen, with the F_1 exceeding the average parental value $\bar{P}$ for the crossed lines. Both dominance and epistasis can generate heterosis. In the simplest setting with only dominance present, and 2 alleles per locus (as would occur in a cross of pure lines), if the genotypes $Q_i Q_i : Q_i q_i : q_i q_i$ at a particular locus have average values of $-a_i : d_i : a_i$, then the F_1 heterosis H is given by:

$$H = F_1 - \bar{P} = 2 \sum_i d_i \Delta p_i^2 \qquad (24)$$

where Δp_i^2 is the difference in the frequency of allele Q_i between the crossed lines. Heterosis thus requires both directional dominance (the d_i tending to be positive) and significant differences between the lines allele frequencies at these loci (Δp_i^2 non-zero). When epistasis is present, its contribution to H is a complex function of the pairwise epistatic gene action parameters (*aa*, *ad*, and *dd*) and the appropriate product of differences in allele frequencies between the loci being considered (see Holland 2001). A more compact way to represent the epistatic contribution

to heterosis is to use composite cross parameters, which can be expressed as functions of the differences in allele frequencies (with no contribution from loci whose alleles have the same frequency in the 2 lines being crossed) and the least-squares estimates of dominance and epistatic effects. For these least-squares estimates, we need a hypothetical reference population, namely the population resulting from randomly mating the F_1 until both Hardy-Weinberg and linkage equilibrium are reached. In this case, we can express the heterosis in terms of dominance deviations and additive × additive effects:

$$H = \sum_i \left(\delta_{i,12} - \frac{\delta_{i,11} + \delta_{i,22}}{2} \right) \Delta p_i^2$$
$$+ \sum_{i \neq j} \Delta p_i \Delta p_j (\alpha\alpha_{ii} + \alpha\alpha_{jj} - \alpha\alpha_{ij} - \alpha\alpha_{ji})$$
$$(25a)$$

In terms of the composite dominance (δ^c) and additive × additive ($\alpha\alpha^c$) effects, this is simply:

$$H = F_1 - \bar{P} = 2\delta^c - \alpha\alpha^c \qquad (25b)$$

(see Lynch and Walsh 1998 for details). Hence, the F_1 heterosis is a function of both dominance effects and additive × additive epistasis. The gain from dominance (due to an excess of heterozygotes) is balanced by loss of favourable additive × additive combinations that have developed within each population.

In the F_2, the well-known result (at least for dominance) is that half of the F_1 heterotic advantage is lost. With epistasis:

$$F_2 - \bar{P} = \frac{H}{2} - \bar{c}[\alpha\alpha^c + 4(1 - \bar{c})\delta\delta^c] \qquad (26a)$$

where $\bar{c}$ is the average pairwise recombination between loci (Lynch and Walsh 1998). The second term represents Dickerson's (1969) recombinational loss, and is restricted to the epistatic contributions. In particular, to the loss of favourable additive × additive and dominant × dominant interactions via recombination in the F_2 and subsequent generations. Ultimately, when the line is intermated for sufficient generations to reach linkage equilibrium, we have:

$$F_\infty - \bar{P} = \delta^c - \alpha\alpha^c - \delta\delta^c$$

Note that both $\alpha\alpha^c$ and $\delta\delta^c$ can be either positive or negative. Thus, the loss in performance between the F_1 and the resulting equilibrium population is entirely a function of the dominance and dominance × dominance effects:

$$F_1 - F_\infty = \delta^c + \delta\delta^c \qquad (26b)$$

Equations 24 and 26 illustrate the 2 central problems of hybrid breeding (aside from obvious reproductive issues such as the ease of crossing): choosing the optimal lines

to cross and mitigating the loss of heterosis in subsequent generations beyond the F_1. Eqn 24 points out that differences between the lines are key, indeed the effects are largest when crossing two fixed lines (so that $\Delta p_i^2 = 1$), but attempts to base crossing decisions on genetic distances estimated using molecular markers have been largely unsatisfactory (Melchinger 1999). This is a potentially interesting observation in that if molecular markers reflect the amount of neutral divergence, and these are poor predictors of divergence at loci contributing to heterosis; the latter loci may have had their allele frequencies more influenced by selection.

Hope for the future? Of markers and E(NK)

The take-home message for exploiting non-additive variance is that we require breeding schemes with inbreeding or crossing designs that generate populations with substantial departures from Hardy-Weinberg. As we have detailed, with appropriate knowledge of the non-additive variance or composite cross components, we can predict the expected behaviour of the population under selection (assuming the infinitesimal model). With genes of large effect, selection changes the genetic variance components in a very different fashion from that expected for unselected loci under the same mating system. Standard 1- and 2-locus population genetics theory (e.g. Bürger 2000; Walsh and Lynch 2005) can be used in such cases to predict response. The practical problem in applying either variance component methods or explicit population genetic models is insufficient information on the appropriate population parameters. The experimental designs to estimate inbreeding variance components (e.g. σ_{DI}^2 and σ_{ADI}) are rather involved, and (even if successful) require the cost of several generations in which selection is not practiced. There is also the very non-trivial issue (especially in line crosses) of choosing the appropriate initial lines (for reviews of various strategies see Bernardo 2002; Walsh and Lynch 2005).

Two new tools (one empirical, the other theoretical) have recently been added to the arsenal of applied quantitative geneticists, namely molecular markers and E(NK) models. Might these be of some aid in exploiting non-additive variation? Considering molecular markers first, the issue of estimation of effects is still of concern. One has to find suitable markers and make sure that the marker–trait associations are sufficiently stable to be of use in selection schemes (Walsh 2002). With such information in hand, selection on combinations of markers can certainly improve our ability to exploit additive epistasis in random-mating populations. Since (under random mating) parents only pass along single alleles for each locus, marker information will have little effect on exploiting dominance or dominance epistasis unless lines are also inbred. Although no one disputes the power of molecular markers for introgression of known genomic regions/major genes into a line, their

effect is debatable when the infinitesimal model is a better descriptor of the genetic architecture for the traits being selected.

A second potentially useful tool is E(NK) models (Cooper *et al*. 2005). Starting with a framework of known parameter estimates, this class of models generates a universe of potential epistatic and genotype × environment interactions consistent with the initial framework. The utility of this approach, which in the appropriate situation can be considerable, is that it allows one to leverage known genetic details. This allows one to explore what breeding approaches are the most robust given the known information coupled with the space of possible genetic parameters values generated by incorporating the E(NK) model.

References

Bernardo R (2002) 'Breeding for quantitative traits in plants.' (Stemma Press: Woodbury, MI)

Bulmer MG (1980) 'The mathematical theory of quantitative genetics.' (Oxford University Press: New York)

Bürger R (2000) 'The mathematical theory of selection, recombination, and mutation.' (Wiley: New York)

Cheverud JM, Routman E (1996) Epistasis as a source of increased additive genetic variance at population bottlenecks. *Evolution; International Journal of Organic Evolution* **50**, 1042–1051.

Cockerham CC (1954) An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**, 859–882.

Cockerham CC (1983) Covariances of relatives from self-fertilization. *Crop Science* **23**, 1177–1180.

Cockerham CC (1984*a*) Covariances of relatives for quantitative characters with drift. In 'Human population genetics'. (Ed. A Charkravarti) pp. 195–208. (Van Nostrand Reinhold Co., Inc.: New York)

Cockerham CC (1984*b*) Additive by additve variance with inbreeding and linkage. *Genetics* **108**, 487–500.

Cockerham CC, Tachida H (1988) Permanency of response to selection for quantitative characters in finite populations. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 1563–1565.

Cooper M, Podlich DW, Smith OS (2005) Gene-to-phenotype models and complex trait genetics. *Australian Journal of Agricultural Research* **56**, 895–918.

Cornelius PL (1988) Properties of components of covariances of inbred relatives and their estimates in a maize population. *Theoretical and Applied Genetics* **75**, 701–711. doi: 10.1007/BF00265590

Dickerson GE (1969) Techniques for research in quantitative animal genetics. In 'Techniques and procedures in animal science research'. pp. 36–79. (American Society of Animal Science: Albany, NY)

Edwards JW, Lamkey KR (2002) Quantitative genetics of inbreeding in a synthetic maize population. *Crop Science* **42**, 1094–1104.

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399–433.

Gallais A (1975) Selection with truncation in autotetraploids: comparison with diploids. *Theoretical and Applied Genetics* **46**, 386–394.

Gallais A (2003) 'Quantitative genetics and breeding methods in autopolyploid plants.' (INRA: Paris)

Harris DL (1964) Genotypic covariances between inbred relatives. *Genetics* **50**, 1319–1348.

Holland JB (2001) Epistasis and plant breeding. *Plant Breeding Reviews* **21**, 27–92.

Kempthorne O (1954) The correlation between relatives in a random mating population. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **143**, 103–113.

Lynch M, Walsh B (1998) 'Genetics and analysis of quantitative traits.' (Sinauer Associates: Sunderland, MA)

Melchinger AE (1999) Genetic diversity and heterosis. In 'The genetics and exploitation of heterosis in crops'. (Eds JG Coors, S Pandey) pp. 99–118. (American Society of Agronomy: Madison, WI)

Robertson A (1952) The effect of inbreeding on the variation due to recessive genes. *Genetics* **37**, 189–207.

Tachida H, Cockerham CC (1989) Effects of identity disequilibrium and linkage on quantitative variation in finite populations. *Genetical Research* **53**, 63–70.

Walsh B (2002) Quantitative genetics, genomics and the future of plant breeding. In 'Quantitative genetics, genomics and plant breeding'. (Ed. MS Kang) pp. 23–32. (CABI Publishing: New York)

Walsh B, Lynch M (2005) 'Selection and evolution of quantitative traits.' (In prep.) Draft chapters can be found on the web at http://nitro.biosci.arizona.edu/zbook/volume_2/vol2.html (Sinauer Associates: Sunderland, MA)

Wright AJ (1987) Covariance of inbred relatives with special reference to selfing. *Genetics* **115**, 545–552.