11

Inheritance of Developmental Instability

REBECCA C. FULLER
DAVID HOULE

Fluctuating asymmetries (FAs) are nondirectional departures from perfect symmetry. FA is often used as a measure of developmental instability (DI), a hypothetical construct that we use to refer to the idea, that in some cases, variance in morphological traits may be maladaptive. Asymmetry is a logical way to approach this issue, because we can usually assume that the optimum phenotype is the same on either side of the body. This connection of FA and DI is speculative, as it is not clear that DI is a real property of an individual. In this chapter, we are concerned with the issue of whether FA is heritable, and what this might tell us about the inheritance of DI, assuming that DI does in fact exist. To study this relationship, we assume a standard model of the relationship between FA and DI (Van Valen 1962; Palmer and Strobeck 1986; Whitlock 1996; Gangestad and Thornhill 1999; Houle 2000).

We assume that FA and DI are affected by both environment and genetics. Many studies have demonstrated that FA is affected by environmental conditions (Parsons 1990; Rettig et al. 1997; Lens et al. 1999; Bjorksten et al. 2000a). However, the question of whether or not FA is heritable has been controversial (Houle 1997; Møller and Thornhill 1997; Pomiankowski 1997; Whitlock and Fowler 1997). One early meta-analysis suggested that the amount of genetic variation in FA (and therefore DI) was high (Møller and Thornhill 1997), but later reviews found much lower levels of additive genetic variation in FA (Whitlock and Fowler 1997; Gangestad and Thornhill 1999; Van Dongen 2000a; Van Dongen and Lens 2000). This problem has become more complex with the realization that FA is an imprecise measure of DI (Houle 1997, 2000; Van Dongen 1998; Whitlock 1996, 1998).

Inferring developmental instability from fluctuating asymmetry is problematic due to the difficulties of estimating a variance precisely. When we use FA to infer DI, we are attempting to measure the variance of a trait. In other words, we want to know the potential range of values a trait can assume for an individual. However, FA has usually been used to measure this variance with only two data points (e.g., left and right trait values). Measuring a variance with two data points is highly inaccurate and introduces a large amount of error into the relationship between FA and DI.

The question of whether differences in asymmetry and developmental instability are heritable can be of considerable importance when using individual asymmetry to infer the quality of individuals. For example, there has been a good deal of speculation that asymmetry is an important cue for mate choice because it reflects good genes (Møller 1994; Swaddle and Cuthill 1994; Watson and Thornhill 1994). In order for asymmetry to be a reliable indicator of the quality of genes passed to offspring, there must be genetic variation in FA in the population. Similarly, the use of asymmetry as an indicator of inbreeding depression again demands that genetic differences account for differ-

ences in asymmetry (Sheridan and Pomiankowski 1997), although, in this case, the quality of the genes passed to outcrossed offspring might not be predicted by parental asymmetry. In contrast, the use of asymmetry to diagnose the relative health of populations does not depend on a genetic basis to differences in asymmetry.

In this chapter, we consider the basics of inferring genetic differences among individuals in asymmetry and developmental instability in outbreeding populations. First, we review some principles from quantitative genetics to frame our discussion. Second, we discuss the special challenges of quantitative genetic studies of asymmetry and developmental stability. We then discuss a simple model of the relationship between asymmetry and developmental instability. The model parameters are chosen based on our review of the literature on the quantitative genetics of asymmetry. Based on this model, we then investigate which statistical approaches to the inheritance of asymmetry are likely to have the greatest power.

Quantitative Genetics

The inheritance of asymmetry and developmental instability is usually quantitative in nature; that is, observed differences among individuals are quantitative rather than qualitative. This puts asymmetry in the company of the vast majority of phenotypic traits whose variation is not primarily determined by a few genes of large effect. This offers some challenges when investigating inheritance and evolution. An excellent introduction to quantitative genetics is available (Falconer and Mackay 1996), and a comprehensive guide to quantitative genetic theory and practice has recently appeared (Lynch and Walsh 1998).

Genetic variation in quantitative traits depends on alleles with effects too small to stand out against a background of other kinds of variation. These other sources of variation include variation in the environment, nongenetic parental effects, or variation at other genetic loci. Detection of quantitative genetic variation, therefore, depends on statistically partitioning variation among individuals into that due to genetic and nongenetic causes. Quantitative genetic experiments, therefore, need to be designed to allow estimation of effects likely to occur in that species. For example, in a species with parental care, separation of genetic and environmental effects requires separating some parents and offspring during development, for example by cross-fostering. Designs that fail to partition such effects cannot necessarily be interpreted in genetic terms. In this chapter, we confine our attention to the question of whether there is additive genetic variance for asymmetry and developmental instability (i.e., whether offspring resemble their parents). Leamy (chapter 10, this volume) takes up the important issue of the actual mechanism of gene action.

Because of the central importance to partitioning phenotypic variance into its components, it is crucial to keep in mind what can be concluded based on a particular experimental design. Our review of the literature on the inheritance of asymmetry (see section "Literature Review" below) shows that three experimental designs predominate: parent–offspring regression, a nested half-sib design where male parents are mated to several females, and studies that examine sets of full sibs. Two studies have used twin data in humans. Of these designs, the parent–offspring and nested half-sib design enable additive genetic variance to be estimated, provided that critical assumptions are met (Falconer and Mackay 1996).

Parent–offspring regression studies include what we want to know to predict evolutionary potential (i.e., the degree to which offspring resemble their parents). However, parents and offspring may also resemble each other for a variety of nongenetic reasons, such as a shared environment. For example, in organisms with limited gene flow, parents and offspring will probably have matured in similar environments. Similarly, offspring can resemble their parents due to maternal effects from provisioning offspring during development. In species with parental care, parents may transmit both material and genetic benefits to their offspring. A simple check for maternal effects is to test whether the relationship between mothers and offspring is equal to that between fathers and offspring. Observational studies are particularly susceptible to confounding of environmental, parental and additive genetic effects.

The nested half sib design generates three levels of relationship between the offspring. Individuals may be unrelated, half sibs, or full sibs. This allows a more detailed partitioning of variation than in the parent–offspring design. The ability to perform a nested half-sib design implies a level of experimental control in the system that is not necessarily pre-

sent in parent–offspring regression experiments, and a good deal of the advantage of this design is attributable to this. For example, the half sibs share some genes, but may not share a rearing environment. In most cases, a male is mated to several females, so maternal effects show up at the level of full-sib families, and do not affect the resemblance of half sibs. In species where both sexes care for offspring, half sibs may share some parental-care effects. The main disadvantage of the nested half-sib design is that in some species it is not possible to control or identify matings necessary to create these different levels of relatedness. When it is possible to do the necessary crosses, this often requires displacing individuals to an artificial environment that may introduce artifactual results.

A common type of study in the asymmetry literature contains comparisons between full sibs versus unrelated individuals. Full-sib designs are not proper quantitative genetic designs. In addition to confounding all of the environmental components that affect families with all of the genetic effects, full sibs resemble each other genetically in ways that parents and offspring do not. Full sibs resemble each other due to dominance variation, as well as additive genetic variation. Thus, even if we could assume that the problems of common rearing and parental effects were negligible, this design simply does not estimate the proper genetic parameter. This design is, presumably, so widely used because it is frequently convenient to use in free-ranging organisms with a minimum of manipulation. A failure to find a significant family effect in a full-sib design would be good evidence against the importance of a genetic component to asymmetry. Unfortunately full-sib designs cannot be used to conclude any more than this about inheritance.

Inheritance of Fluctuating Asymmetry and Developmental Instability

Developmental instability is morphological variation around an assumed optimum state. Fluctuating asymmetry is used as a measure of DI because we often assume that perfect symmetry is optimal. Many different measures of fluctuating asymmetry have been proposed and used in the literature (reviewed in Palmer and Strobeck 1986; Palmer 1994). The most commonly used measure in genetic studies is the unsigned difference between realizations of the same structure across an axis of symmetry, such as the left versus right sides of the body. Fluctuating asymmetry of the ith individual is then

$$FA_i = |L_i - R_i| \quad (1)$$

where L_i and R_i are the sizes of the relevant part on the left and right sides of the body of individual i. Given that what we seek to measure with FA is the variance in development of a structure, it would be preferable to use the estimated variance between sides. However, genetic studies have not made use of variances, so we will not consider this statistic.

What we seek to measure with FA is the variance of the structures measured in a situation where the expected value is the same and, thus, the expected difference is zero. Normally, we cannot replicate individuals and repeatedly rear them in the same environment. The two sides of a basically symmetrical individual do represent such a situation, as the genotype of the two sides is the same (barring somatic mutation), and the whole-organism environment is also the same. In the ideal case, there would be more than two replicates to this natural experiment (Leung et al. 2000). Plants, for example, usually have many leaves, and may have many petals within the same flower. Unfortunately, data on asymmetry of multiply replicated parts has only been used in a single genetic analysis (Wilsey and Saloniemi 1999).

The fact that the within-individual sample size is two in all but one genetic study is an important limitation. Few of us would ever design an experiment where the sample size is two, even if we were only estimating the mean of a trait. Variances are much harder to estimate accurately than means, so the limited sample size is even more critical with FA. The result is that any particular FA value will be a very unreliable indicator of the development variance or DI for an individual (Whitlock 1996; Houle 1997). This suggests that unusually large experiments will be necessary to detect significant genetic variation in FA and DI.

This challenging situation is represented schematically in figure 11.1, where we show data from four simulated populations with different levels of variation in DI. Each pair of panels represents a popula-

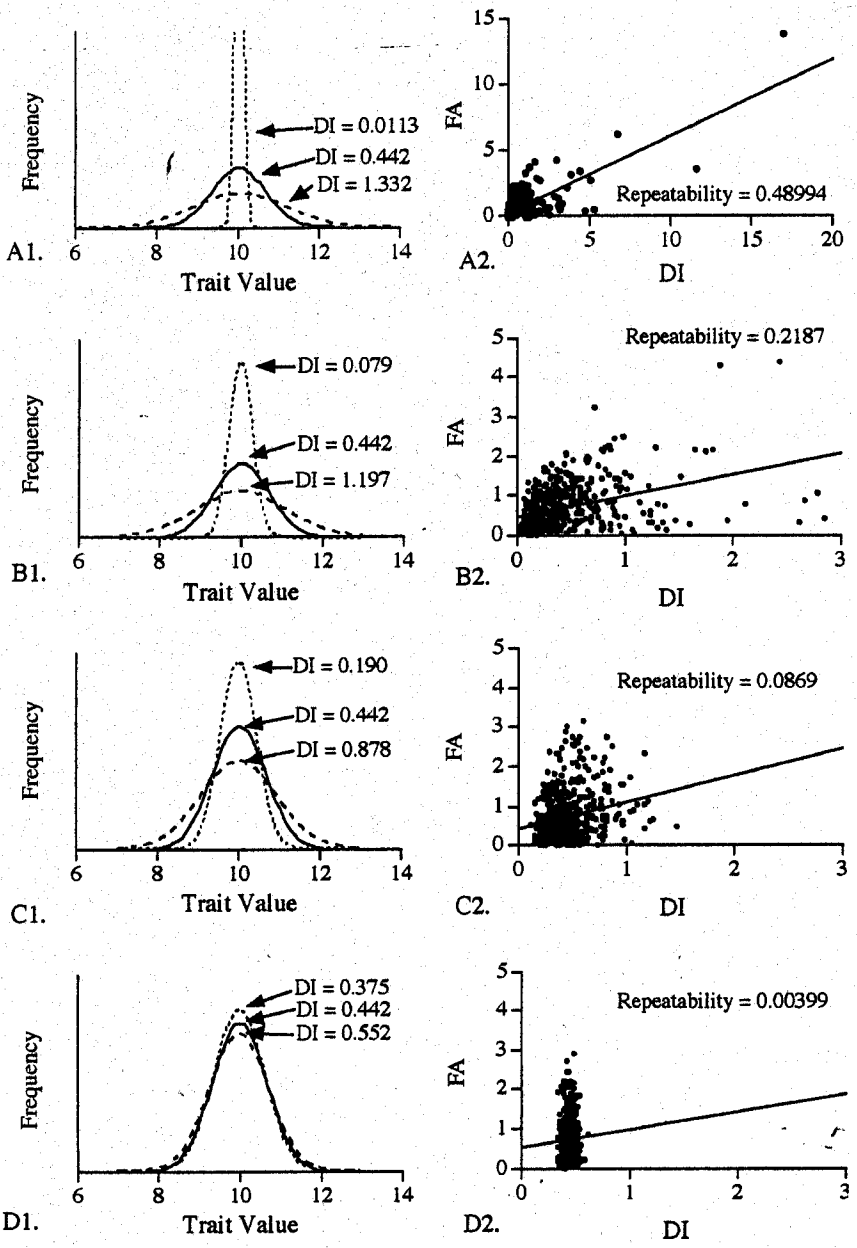

Figure 11.1 The relationship between trait value, developmental instability (DI), and fluctuating asymmetry (FA). The four sets of graphs depict populations with (A) very high, (B) high, (C) medium, and (D) low variation in developmental instability. The left-hand panel (A1, B1, C1, D1) shows the frequency of expected trait values for individuals exhibiting the mean level of DI (solid line), individuals at the 5th percentile of DI values (small dashed line), and individuals at the 95th percentile of DI values (long dashed line). An individual with a low value for DI has a high probability for expressing trait values close to the mean. Conversely, an individual with a high value for DI is more likely to express trait values that deviate from the mean. The right-hand panel (A2, B2, C2, D2) shows the relationship between FA and DI. Each graph shows 500 independent individuals generated using our model with 500 sires, 1 dam/sire, and 1 offspring/clutch (i.e., there are no related individuals). Repeatability describes the degree to which differences in FA reflect differences in developmental instability among individuals. Note on A2 that the scale for both the x- and y-axes differs from B2, C2, and D2. The relationship between DI and FA (i.e., the slope) is approximately equal across the four graphs. Due to the high realization error in FA, a large amount of variation in DI is required for high repeatability.

tion. The three curves in each of the left-hand graphs represent the mean, 5th, and 95th percentile values for DI for a population with a given level of variation in DI. Each individual has a unique value for DI that corresponds to a unique curve of expected trait values most likely occurring between the 5th and 95th percentiles. The range of different curve shapes represents the variation in DI among individuals in a population. The right-hand panels show the relationship between DI and many realizations of FA. A realization entails two expressions of the trait, corresponding to right and left trait values. The probability of a trait assuming any one value is determined by DI, the variance in potential trait values. An individual with a low value for DI has a small variance of expected trait values, and therefore, a low average value of FA. An individual with a large value for DI has larger variance of expected trait values and a large average value for FA. Note that, while there is some relationship between DI and FA, individuals with all levels of DI are still fairly likely to be nearly symmetrical. Thus, the "realization" of DI as FA in an individual is highly variable.

The realization variance in FA adds substantially to the nongenetic sources of variance when we partition FA variance during a quantitative genetic experiment. This means that heritability estimates of FA drastically underestimate the heritability of DI (Whitlock 1996, 1998; Houle 1997; Van Dongen 1998). One way to see this is to examine the relationship between FA and DI in the simulated data in Figure 11.1. A regression of FA on DI shows that DI explains rather little of the true variance in FA, with repeatability values or R^2 of less than 49% for even the best-case scenario in figure 11.1.

The repeatability of FA, which Van Dongen (1998) refers to as the hypothetical repeatability, is the expected proportion of variation due to real differences in DI among individuals. It excludes both measurement error and the variance introduced by differences in development. It thus differs from the usual measures of repeatability that only take measurement error into account (Swaddle et al. 1994; Van Dongen 2000b). The repeatability associated with measurement error reflects differences in FA when the same individual is measured more than once. The repeatability associated with DI deals with the degree to which differences in FA reflect differences in DI, as depicted in figure 11.2. For the remainder of this chapter, we will use the latter definition when referring to repeatability.

As stated above, the repeatability cannot be estimated from real data, because DI cannot be directly measured. Whitlock (1996) suggested a potential solution to this problem. He pointed out that the mean and variance of FA are closely related and that variation in DI increases the variance in FA faster than it increases the mean of FA. If one can assume the shape of the distribution of the traits that are used to calculate FA, then one can calculate the expected coefficient of variation of FA (CV_{FA}) under the assumption that DI does not vary. Any CV_{FA} higher than this minimal value then has to reflect real differences in DI among individuals. The repeatability, or proportion of variance in FA that is due to real differences in developmental variance among individuals, can be calculated on this basis, and the realization error of FA partitioned out, allowing a less biased estimate of the heritability of DI.

Accurate formulas for calculating repeatability of FA under the assumption that the trait values are normally distributed have been derived by Van Dongen (1998) and Whitlock (1998). Van Dongen (1998) showed that the repeatability is

$$\Re = [V_{FA} - V_{L-R}(\pi - 2)/\pi)/V_{FA}] \qquad (2)$$

where V_{FA} is the total variance in FA among individuals, and V_{L-R} is the variance in the difference between left and right sides, or signed FA. Alternatively, Whitlock (1998) showed that the approximate repeatability of FA can be calculated as

$$\Re = \frac{2}{\pi} - \frac{(\pi - 2)}{\pi CV_{FA}^2} \qquad (3)$$

In our simulations, these two formulas for the repeatability give nearly identical results (see section "Modeling Fluctuating Asymmetry and Developmental Instability"). In either case, the heritability (h^2) of DI may be estimated by dividing the heritability of FA by the repeatability

$$h_{DI}^2 = h_{FA}^2/\Re \qquad (4)$$

An alternative method of estimating repeatability from kurtosis, rather than the lower moments, has

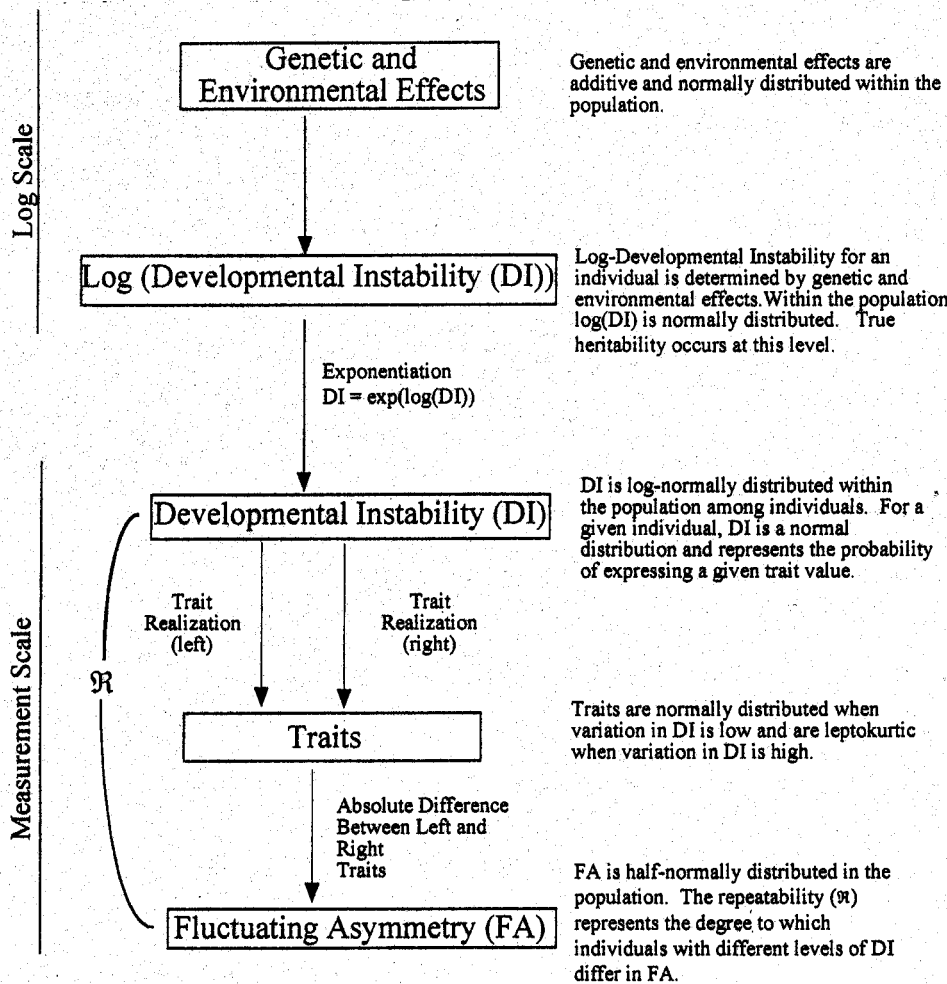

Figure 11.2 Schematic diagram for the expression of DI and FA. Genetics and environment are additive and affect DI on the log scale. The expression of left and right traits are two independent realizations of developmental instability. Fluctuating asymmetry is the absolute difference between these two traits.

also been proposed (Gangestad and Thornhill 1999, and chapter 5, this volume).

Literature Review

We reviewed the literature for studies reporting heritability estimates of FA (h^2_{FA}). The heritability of FA has been the subject of a number of recent reviews (Møller and Thornhill 1997; Whitlock and Fowler 1997; Gangestad and Thornhill 1999; Van Dongen 2000a,b; Van Dongen and Lens 2000). Each of these reviews used different criteria for inclusion of studies that we did not think entirely appropriate. For example, both Whitlock and Fowler (1997) and Gangestad and Thornhill (1999) misreported the heritability of FA for bristle number in female *Drosophila melanogaster* as 0.05. It should be 0.005 (Reeve 1960). Gangestad and Thornhill (1999) correctly report three heritability estimates from Scheiner et al. (1991) (0.039, 0.027, 0.005), but exclude a negative heritability estimate (−0.026). While this estimate was excluded from their table, it apparently was used in their calculation for mean FA heritability (Gangestad, pers. comm.). Gangestad and Thornhill (1999) also included an unusually high heritability estimate (0.630) calculated for the FA of lateral plates in

sticklebacks (Hagen 1973). This study calculates the heritability from animals pooled across two different rearing conditions (21°C and 16°C) and six different populations. Furthermore, their analysis results in pseudoreplication of parental observations. Van Dongen (2000a) uses a Bayesian analysis to estimate the mean and standard deviation of the heritability of FA. However, he modeled a beta distribution that constrains the estimate to lie between 0 and 1. This may result in an estimate biased away from zero. This practice results in an overestimation of additive genetic variation (Lynch and Walsh 1998). Finally, Van Dongen and Lens (2000) excluded studies with measurement error larger than real asymmetry despite the fact that repeatability-corrected estimates of h^2_{DI} are expected to be unbiased. Furthermore, they included some studies where measurement error was not calculated.

Here, we report heritability estimates of FA. We concentrated on finding recently published studies and relied on previous reviews for references to older literature (Gangestad and Thornhill 1999; Van Dongen and Lens 2000). In addition, we performed a literature search using Cambridge Scientific Abstracts employing the search words fluctuating asymmetry, heritability, and genetic variation. We included studies that used half-sib, full-sib, parent–offspring, and twin methods to measure the heritability of FA. We do not report the results for studies examining variation among clones. We only report data from studies examining differences between right and left traits as opposed to the total number of asymmetrical traits (see Leary et al. 1985). Estimates that showed significant directional symmetry or antisymmetry were excluded. Studies that did not test for such a phenomenon were included. Finally, we only included studies that reported the heritability estimates of FA for all measured traits. Studies that reported only heritability estimates that differed significantly from zero were excluded (see Parker and Leamy 1991; Cadeé 2000).

For every study in our sample, we recorded each estimate of the heritability of FA, the method used to calculate heritability, the overall sample size of the experiment, and, if available, the coefficient of variation of FA (CV_{FA}), CV of trait values, and $\Re$ (hypothetical repeatability) when available. For many studies, we calculated $\Re$ from the CV_{FA} (or mean and standard deviation of FA) reported in the essay. For one study (No. 18), the $\Re$ reported did not agree with the $\Re$ we calculated based on the means and standard deviations of FA. For this study, we used the $\Re$ provided by the authors. Electronic Appendix 1 (see p. xxiii for details) gives the raw values for each estimate in each study. Most recent reviews have taken the form of meta-analyses (Møller and Thornhill 1997; Whitlock and Fowler 1997; Gangestad and Thornhill 1999; Van Dongen and Lens 2000). Our goal here is primarily to provide an exhaustive list of heritability estimates for FA in order to ascertain the strength of evidence for inheritance in particular cases. We also compare the proportion of estimates that differ from 0 with the expected Type 1 error rate.

We found a total of 183 estimates of heritability for FA, in 20 published studies. Four studies used nested half-sib breeding designs (yielding 24 estimates), seven studies used full-sib breeding designs (55 estimates), ten used parent–offspring regression (91 estimates), and two studies used analyses of monozygotic and dizygotic twins (13 estimates). Two studies used estimated values based on both full-sib and parent–offspring regression (see Electronic Appendix 1). The mean number of offspring used in breeding designs was 740.3, but the median number of offspring was 233.9. Two studies (No. 6, No. 7) with large sample sizes created this disparity between mean and median in sample size.

The average hypothetical repeatability was 0.294. Gangestad and Thornhill (chapter 5, this volume) show that $\Re$ is sensitive to the scale of measurement relative to the degree of asymmetry. They adjusted many $\Re$-values for this phenomenon and found $\Re$ typically between 0.08 and 0.10. This phenomenon has probably led us to overestimate the hypothetical repeatability somewhat.

Few studies found a significant amount of genetic variation for FA (see Electronic Appendix 1 for individual tests). Of the 55 estimates based on full-sib breeding designs, only four differed significantly from zero. Three of these estimates were problematic because FA was corrected for size (i.e., measured as the absolute difference between left and right trait values divided by the mean trait value). The practice of dividing FA by trait size potentially confounds genetic variation in trait size and asymmetry. Residuals from a regression of FA on trait size provide a much better estimate of size-corrected FA. Of the 13 estimates based on twin studies, four heritability estimates differed from zero. However, all four of these estimates came from one study (No. 10) that may have inflated

the heritability estimate by a factor of 2 (see Pechenkina et al. 2000 for details). Of the 90 estimates based on parent–offspring studies, only six were significantly different from zero. However, five of these significant estimates came from two studies of primates (macaques and humans) where mother FA was regressed on daughter FA. These studies could not eliminate environmental effects due to maternal effects and extended parental care. Finally, of the 21 estimates based on nested half-sib designs, only one estimate differed significantly from zero. This study was a Herculean effort that employed a massive sample size (table 11.1).

Unweighted mean h^2_{FA} values arrived at by the different methods are 0.062 (half-sib), 0.043 (parent–offspring), 0.032 (full-sib), and 0.147 (twin). These are nearly significantly different in a one-way analysis of variance (ANOVA) ($P = 0.067$) on the unaveraged values, due primarily to the large discrepancy between the twin estimates and the others.

For each study, we calculated the unweighted mean for each of these parameters and median values for the heritability of DI and FA from these raw estimates. These values are shown in table 11.1. Overall, the average values in table 11.1 are not quite significantly different from 0 by an unweighted t-test (mean 0.026 ± 0.015 SE, $P = 0.092$). However, these numbers are biased upwards because many studies reported negative heritabilities as 0. Study No. 20 (Arnqvist and Thornhill 1998) found that 16 out of 32 h^2_{FA} estimates were negative and rounded them to zero. Similarly, study No. 11 (Corruccini and Potter 1981) rounded three out of four h^2_{FA} estimates to zero. Study No. 10 (Pechenkina et al. 2000) rounded three of 18 h^2_{FA} estimates to zero.

We estimated the heritability of DI (h^2_{DI}) for the above studies by dividing h^2_{FA} by the hypothetical repeatability, $\Re$ (see section "Inheritance of Fluctuating Asymmetry and Developmental Instability"). In cases where h^2_{FA} was negative, we set $h^2_{DI} = 0$, so average h^2_{DI} will be highly biased. The relationship between the coefficient of variation of FA, which is used to calculate $\Re$, and h^2_{DI} is shown in figure 11.3. Estimates of h^2_{DI} were highly variable and often did not make sense biologically (i.e., negative or greater than 1). Two studies had average h^2_{DI} above 1.0 (No. 18, No. 14). In study No. 18 (Bjorksten et al. 2000b), only one of 18 estimates of h^2_{DI} was a value between 0 and 1. Two heritability estimates were negative, eight were zero, and seven were above 1.0 (see Electronic Appendix 1; see p. xxiii for details). In study No. 14 (Leamy 1999), four of ten estimates were above 1.0. In one study (No. 15, Evans and Marshall 1996), the average h^2_{DI} was negative despite the fact that all negative h^2_{FA} values were set to zero when calculating h^2_{DI}. Although the average $\Re$ for study No. 14 is positive in table 1.11, many $\Re$-values for individual traits were less than zero, and this had a large effect on h^2_{DI} (see Electronic Appendix 1).

The bizarre nature of these heritability estimates suggests that something is amiss about the use of $\Re$ to correct for the relationship between variation in FA and variation in DI. One possible reason for this is that heritability corrected by $\Re$ has too high a variance to be useful. In contrast to Gangestad and Thornhill (chapter 5, this volume), who concentrated on the problems associated when $\Re$ is large, these bizarre results occur when $\Re$ is too low. This interpretation is supported by the fact that the extreme values of h^2_{DI} we have noted arise when the coefficient of variation in FA is low, causing $\Re$ to be a small positive or negative value, as shown in figure 11.3. We next explore this possibility using a simulation model. Another possible reason for this poor performance is that the model used to determine the relationship between FA and DI is not correct. For example, the model assumes that the distribution of the parts is normal, which may not be correct (see Klingenberg, chapter 2, this volume).

Modeling Fluctuating Asymmetry and Developmental Instability

We developed a model to simulate the expression and inheritance of fluctuating asymmetry and developmental instability (figure 11.2). We envision developmental instability (DI) as a variance around a trait mean (figure 11.1; Houle 2000). DI is equivalent to V_D in Houle (2000). A large value for DI means that there is a wide possibility of trait values that an individual can assume. A small value for DI means that development is more canalized, and an individual is more likely to develop trait values closer to its expected mean. In these simulations, all individuals have the same expected mean trait value.

Variances must be positive, and it is important to take this into account when modeling or simulating developmental instabilities. In this study, we

Table 11.1 Summary Statistics from Reviewed Studies.

For each study, we list the organism, the average number of sires, average number of dams/sire, and average total number of offspring used. We also list the average coefficient of variation of FA, the average $\mathfrak{R}$, the average heritability of FA, the average heritability of DI, the median heritability of FA, and the median heritability of DI. Below each column we list the average and standard deviation of all the studies. For the medians, we list the median and range of all the studies. Note that all negative FA heritability estimates were rounded to zero when calculating the raw heritability DI values. See Electronic Appendix 1 for raw data; see p. xxiii for details.

Study	Organism	Avg. no. sires	Avg. no. dams/sire	Totals	CV_{FA}	$\mathfrak{R}$	h^2_{FA}	h^2_{DI}	Median h^2_{FA}	Median h^2_{DI}
1	Birch	9	4	108	—	—	0.0001	—	0.0001	—
2	Crickets	21	1	179	—	—	−0.0218	—	−0.0218	—
3	Dung flies	38	1	354.8	1.55	0.476	0.0392	0.108	0.05	0.104
4	Dung flies	41.5	1	196.9	0.903	0.173	−0.0315	0.336	−0.045	0
5	Earwigs	22	2.1	233	2.255	0.565	−0.09	0	−0.09	0
6	Fruit fly	588	2.7	7200.5	2.39	0.423	0.0113	0.033	0.016	—
7	Fruit fly	31	1	1373.3	—	—	0.0225	—	0.0225	—
8	Fruit fly	130.5	1	233.9	—	0.516	−0.058	0.0436	−0.035	0
9	House fly	23	1	474	0.98	0.260	−0.00082	0	−0.00082	−0.0032
10	Humans	—	—	492	—	—	0.156	—	0.17	—
11	Humans	—	—	154	1.07	0.282	0.0525	0.135	0	0.034
12	Macaques	133	1	442	—	—	0.161	—	0.108	—
13	Mice	200	1	—	—	—	0.0325	—	0.03	—
14	Mice	200	1	1003	0.781	0.0394	0.01	1.19	0.025	0.406
15	Mustard	13	3	148.8	0.873	0.0719	0.113	−0.298	0.09	0.323
16	Parasitoid	100	1	200	—	—	−0.0817	—	−0.09	—
17	Peacock butterfly	15	1	—	—	—	0.0208	—	0.003	—
18	Stalk-eyed flies	12	2.4	116.9	0.781	0.0307	0.085	1.17	0.09	0
19	Trout	10	1	383.3	—	—	0.0167	—	0.02	—
20	Water strider	47	1	188	—	—	0.08	—	0.03	—
21	Winter moth	31	1	585	1.33	0.394	0.0267	0.0820	0.03	0.083
	Average or median [M]	87.63	1.48	740.34	1.29	0.294	0.026	0.255	[M] 0.020	[M] 0.017
	(SD or range [R])	(136.2)	(0.90)	(1597)	(0.596)	(0.195)	(0.067)	(0.481)	[R] (0.29)	[R] (0.406)

The study number refers to following references: 1. Wilsey and Saloniemi 1999; 2. Eggert and Sakaluk 1994; 3. Blanckenhorn et al. 1998; 4. Blanckenhorn and Hosken unpublished manuscript; 5. Tomkins and Simmons 1999; 6. Scheiner et al. 1991; 7. Reeve 1960; 8. Woods et al. 1998; 9. Chapman and Goulson 2000; 10. Pechenkina et al. 2000; 11. Corruccini and Potter 1981; 12. McGrath et al. 1984; 13. Leamy 1997; 14. Leamy 1999; 15. Evans and Marshall 1996; 16. Bennett and Hoffmann 1998; 17. Windig 1998; 18. Bjorksten et al. 2000b; 19. Leary et al. 1992; 20. Arnqvist and Thornhill 1998; 21. Van Dongen et al. 1999.

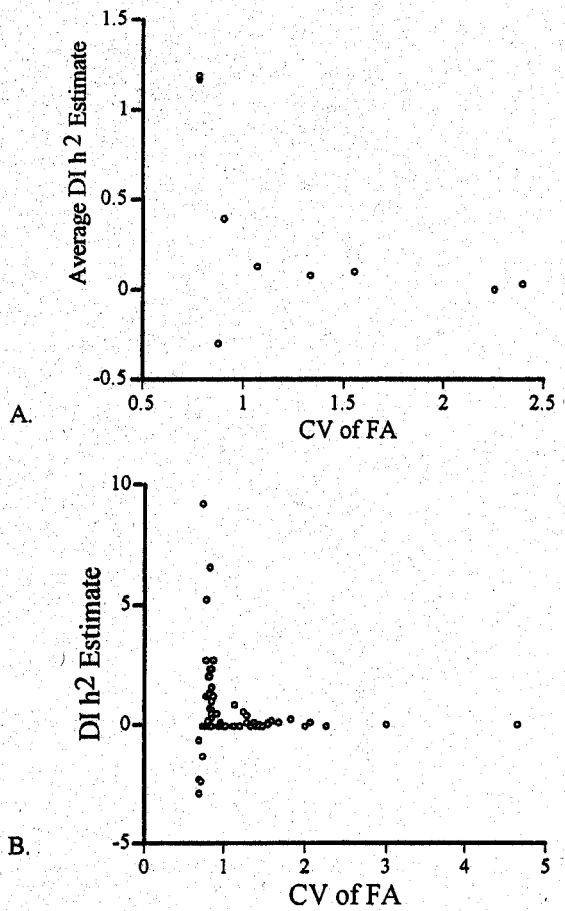

Figure 11.3 The relationship between the coefficient of variation of FA (CV of FA) and estimates of DI heritability. (A) Each data point represents the average of one study. (B) All estimates pooled.

assume that DI is log-normally distributed, in contrast to Houle (2000), who assumed the gamma distribution. The log-normal distribution accommodates the use of additive genetic and environmental variances while allowing for a lower bound for FA values (i.e., zero). Throughout this chapter, log refers to the natural log.

The log-normal distribution describes a variable whose log-transformed values are normally distributed. The log-normal is commonly applied to data where factors influencing a distribution are assumed to act multiplicatively, that is they increase or decrease the value of a variable by a certain proportion, rather than by a given amount. By log transforming, one then puts the variable on a scale where multiplicative effects are additive. On this new scale, the central limit theorem holds, suggesting that the log-transformed variables will be normally distributed. Since proportional changes do not change the sign of the variable, if is appropriate for variables such as sizes or variances where values less than 0 are not possible. Figure 11.2 shows the relationship between the log-scale and measurement scale.

In our simulations, we have held the mean developmental instability (DI) constant, while allowing the variation in DI to vary. When the mean is held constant, a convenient way to express the relative variability is with the coefficient of variation. The four distributions of DI values used in our simulations are shown in figure 11.4. The most important thing to realize about this figure is that each distribution has the same mean DI value. When the variance is small relative to the mean, as represented here by the curve labeled $CV_{DI} = 0.1$, log-normal distributions are nearly symmetrical, and closely

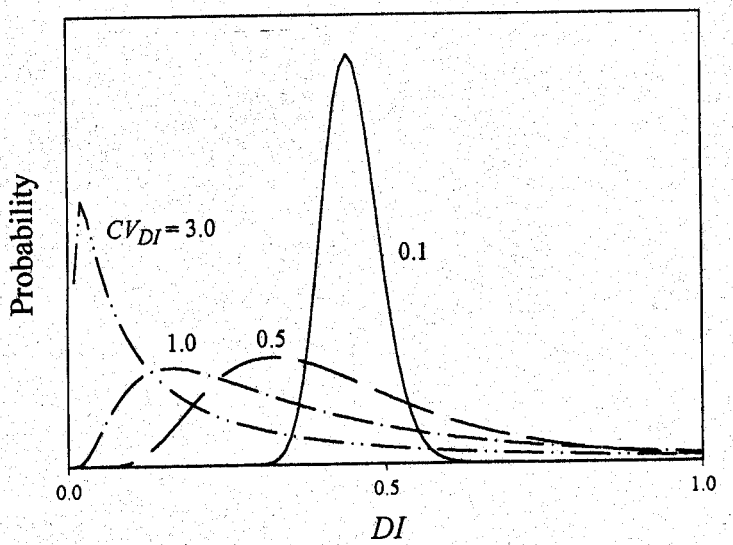

Figure 11.4 Distribution of developmental instability (DI) for different levels of population variation in DI. The four curves (CV_{DI} = 3.0, 1.0, 0.5, and 0.1) represent very high, high, medium, and low variation in DI respectively. Note that the mean DI is kept constant for all four curves.

approximate normal distributions. However, as the variance in DI becomes larger, the fact that observations cannot be less than 0 forces the distribution to become more asymmetrical. In each curve shown in figure 11.4, the mass of observations constrained to lie between the mean and 0 must be larger than those to the right of the mean, whose values may become larger and larger. Extremely large values of DI are rare, but they have a large influence on the mean when they occur.

To simulate fluctuating asymmetry and the underlying developmental variances, we assume that individual values of log-transformed developmental instabilities (ln(DI)) are determined as

$$\ln(\mathrm{DI}) = \mu + E_i S \qquad (5)$$

where μ is the mean DI on the log scale, S is the scaling constant that determines the standard deviation of DI, and E_i is a normally distributed random variable with mean 0 and variance 1. The mean of the developmental variance on the measurement scale will depend on the distribution of ln(DI), and not just the mean (Crow and Shimizu 1988; Lynch and Walsh 1998, pp. 294–5). Therefore, to generate data with an expected value $\widehat{\mathrm{DI}}$ on the measurement scale we set

$$\mu = \ln \widehat{\mathrm{DI}} - S^2/2 \qquad (6)$$

This equation corrects the mean on the measurement scale for the effects of variance on a log scale. Similarly, the variance on the log scale, S^2, is related to variation on the measurement scale as

$$S^2 = \ln(1 + \widehat{CV}_{DI}) \qquad (7)$$

where $\widehat{CV}_{DI}$ is the expected coefficient of variation of developmental instability, expressed as a proportion. The expected level of $\widehat{\mathrm{DI}}$ is the product of the trait mean and coefficient of variation, $\widehat{\mathrm{DI}} = (\omega \mathrm{CV}_{trait})^2$, where CV_{trait} is the coefficient of variation in trait values and ω is the mean of the trait values on the measurement scale (figure 11.2). Both CV_{trait} and ω are constant in all our simulations.

E_i is determined by three different kinds of effects, additive genetic effects of dam, sire, and segregation within families (a), maternal effects of family (m) (which may be interpreted as any combination of nonadditive genetic effects, maternal effects, or common environment effects), and individual deviations due to environment or measurement error (e). Then the deviation of individual i as the offspring of the jth sire and the kth dam is

$$E_i = \frac{a_j + a_k}{2} + \frac{a_i}{\sqrt{2}} + m_k + e_i \qquad (8)$$

Each of the a, m, and e are independent normally distributed random variables with 0 mean. The variances of the variables a, m, and e are chosen such that their total variance is 1. The as have variance equal to the heritability of log developmental variance, with the variance of m and e making up the balance of the variance. The a_i term represents the effect of genetic segregation on individuals within families (Quinton et al. 1992; Lynch and Walsh 1998, p. 758).

We assume that the two sides of an individual are drawn from a normal distribution with variance that equals DI, and that the mean does not differ among individuals. FA is the absolute value of the difference between two such trait values (i.e., left and right sides). For the purpose of these simulations we set ω to an arbitrary value of 10. We set CV_{trait} as 0.0665, a value that is close to the mean value found in our literature review.

We used our model to generate a series of half-sib families to address the following question: How well do estimates of the heritability of DI ($h^2_{DI} = h^2_{FA}/\Re$) approximate the true heritability of DI? We know these values, of course, because we have simulated them, a situation that does not arise in experimental studies.

To do this, we estimated h^2_{DI} for our simulated nested half-sib design. For each simulation we generated 100 sires, each mated to five dams with 100 offspring per clutch, resulting in a total of 50,000 offspring (see Appendix 1 for SAS code). We simulated four levels of variation in DI, labeled low variation ($CV_{DI} = 0.1$), medium variation ($CV_{DI} = 0.5$), high variation ($CV_{DI} = 1.0$), and very high variation ($CV_{DI} = 3.0$). The $\Re$ produced in the high variation treatment (0.2187) approximates the average $\Re$ found in the literature review. The very high and low variation treatments produce $\Re$-values that are approximately one standard deviation higher and lower than the mean (0.49, 0.004). The medium variation treatment represents an intermediate value of $\Re$ (0.0869). In addition, we also independently varied the heritability of DI on the log scale (hereafter referred to as the true heritability) from 0 to 0.8 by increments of 0.2 plus a value of 0.1. For each combination of DI variation and true heritability, we conducted 15 simulations each producing 50,000 offspring. From each simulation, we calculated heritabilities from the variance components (Lynch and Walsh 1998, p. 573) of the simulated data for log(DI), DI on the measurement scale, and FA. In addition, we calculated the estimated h^2_{DI}.

In this chapter, there are five separate heritabilities (four response variables, one model parameter). The heritability of log(DI) and of DI on the measurement scale are response variables and are calculated based on simulated data. These variables cannot be measured using actual organisms. The heritability of FA (h^2_{FA}) is also a response variable and is similarly calculated using variance components. We calculated the estimated heritability of DI ($h^2_{DI} = h^2_{FA}/\Re$). These latter two values (h^2_{DI} and h^2_{FA}) are estimable in real organisms. In contrast to these four estimates, the true heritability on the log scale is a model parameter (i.e., an independent variable).

Finally, we also calculated the hypothetical repeatability ($\Re$) using both Whitlock's and Van Dongen's method (Whitlock 1998; Van Dongen 1998). The two methods produced nearly equal, but not identical results. In this chapter, we present results based on Whitlock's method. All simulations and statistical analyses were performed with SAS V8 (SAS Institute 1999–2000).

Model Performance

To test the validity of the trait expression component of our model, we examined DI and FA across treatments, and compared trait variation and average DI. As expected, average developmental instability did not change with DI variation nor with the true heritability (Kruskal–Wallis tests; $P > 0.150$ in all tests). Average FA did differ with DI variation and true heritability (DI variation by true heritability interaction: Kruskal–Wallis test = 337.24, $P = 0.001$). Average FA in the medium, high, and very high DI variation treatments were 97% , 90%, and 75% of that in the low DI variation treatment. Presumably the cause of this discrepancy is the large departure of the distribution of DI from normal when the variation in log(DI) is large relative to its mean. The extreme skew of the distribution for the highest level of DI variation is evident in figure 11.4 ($CV_{DI} = 3$). With such a skewed distribution, individuals with extremely large DI values occur occasionally, and have an extremely large effect when they do.

For a subsample of simulations, we examined the relationship between average DI, trait variation, and the coefficient of variation of trait values. As

expected, average DI equaled trait variation (0.44). In addition, the measured coefficient of variation of trait value matched the model parameter (0.0665).

To test the validity of the breeding equation, we compared the heritability of DI with the true heritability (model parameter). The heritability of log (DI) did not differ from the true heritability (figure 11.5B, Wilcoxon signed ranks test, $z = 0.3102$, $P = 0.7564$). The heritability of DI on the measurement scale slightly underestimates the true heritability at the higher levels of overall variation in DI (Wilcoxon signed ranks test, $z = -9.940$, $P = 0.0001$).

Our treatments had large effects on the heritability of FA and DI (figure 11.5A–B). Both DI variation and the true heritability affected FA h^2 (figure 11.5A, overall DI variation: Kruskal–Wallis test = 153.92, $P = 0.0001$; True heritability: Kruskal–Wallis test = 129.44, $P = 0.0001$). In contrast, only the true heritability had a significant effect on the heritability of log (DI) (figure 11.5B, true heritability: Kruskal–Wallis test = 324.29, $P = 0.0001$; overall DI variation: Kruskal–Wallis test = 0.271, $P = 0.9655$). Together, these results support the validity of our model.

Measuring Heritability of Developmental Instability

With our simulated data we can determine how well the estimate of the heritability of DI from FA data ($h^2_{DI} = h^2_{FA}/\Re$) approximates the true heritability. Under some circumstances, the estimate performs reasonably well (figure 11.6A–B, figure 11.7B). The estimated h^2_{DI} closely approximates the true heritability under medium to very high levels of variation in DI. The estimate h^2_{DI} is very inaccurate when there is low variation in DI and when there is no real genetic variation in DI (figure 11.6B, figure 11.7B). When overall variation in DI was low, the median absolute difference between h^2_{DI} and true h^2 was 0.345, and the mean was 1.374. When there was no genetic variation, the median absolute difference was 0.072, and the mean was 0.275.

How can a researcher working with real organisms know when his/her estimate of the heritability of DI is unreliable, given that DI is not directly measurable? We investigated this by graphing the relationship between CV_{FA} and the difference in true and estimated heritabilites for all of our simulated data in figure 11.7B. Examination of the figure shows that scientists should be highly skeptical of DI heritability estimates when the hypothetical repeatability of FA is low. Repeatability is negative when the coefficient of variation in FA is lower than 0.75 (figure 11.7A). As CV of FA increases above 0.75, the difference between the estimate and true DI h^2 decreases and the estimate becomes more reliable (figure 11.7B).

The above results were obtained from the simulation of 50,000 offspring per replicate. Is the problem worse with a more realistic number of offspring? To address this question, we simulated a smaller nested half-sib breeding design with ten sires, five dams/sire, and ten offspring/clutch (500 offspring total). Again, we ran 15 simulations for each combination of DI variation and true heritability, and h^2_{FA} estimated and $h^2_{DI} = h^2_{FA}/\Re$. The absolute difference between true and estimated heritabilities frequently approached 1 for all levels of CV_{FA}. In general, we found the same qualitative pattern (figure 11.7C) as in the unrealistically large samples. Estimation of the heritability of DI from the heritability of FA is reasonably accurate only when the coefficient of variation of FA is high.

Building a Better Mousetrap for Asymmetrical Mice

Fluctuating asymmetry is an imprecise measure of developmental instability because it tries to measure a variance with two data points. This may have important implications for the design and analysis of breeding experiments, but the question of the power and optimal design of such experiments has not previously been addressed. In the majority of the breeding studies, there were no significant effects of the terms that would indicate the presence of genetic variation. Thus the heritability of FA did not differ significantly from zero in the majority of cases. This can be explained by an absence of genetic effects on FA and/or by a lack of power of these studies. In this section of the chapter, we take up questions of power and optimal design for data sets with realistic sample sizes.

As outlined earlier, the most informative designs for the estimation of h^2_{FA} in wide use are the nested half-sib experiment and parent–offspring regression. The full-sib and twin designs that have been used in many studies never allow the estimation of the additive genetic variance without the confound-

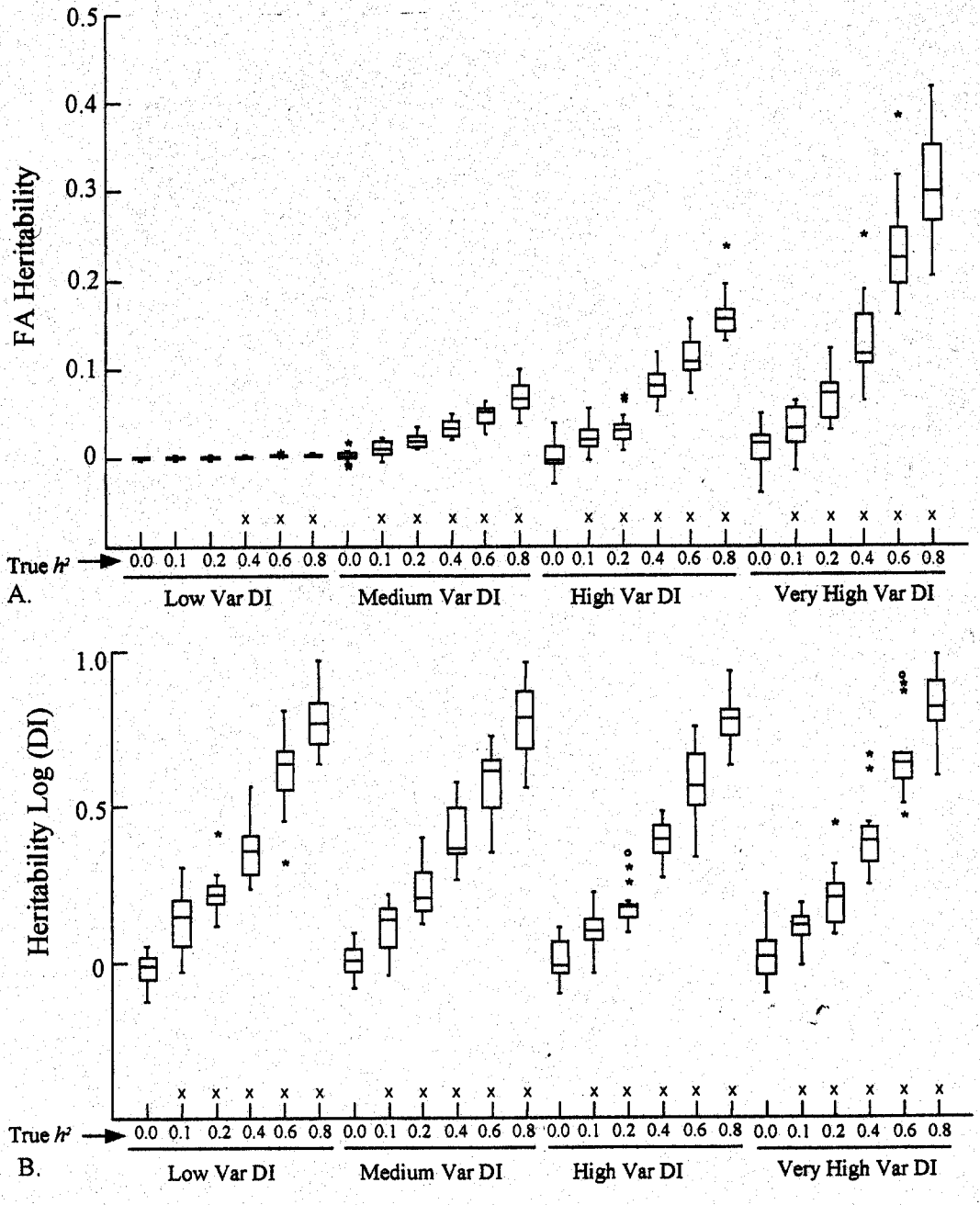

Figure 11.5 Box-plots of (A) FA heritability and (B) heritability log(DI) across treatments. Each data point is the heritability calculated from the 50,000 offspring generated from a single simulation. $N = 15$ for each combination of true heritability and DI variation. × indicates treatments in which heritability differs significantly from zero according to Wilcoxon Signed ranks test. The middle line of the box-plot represents the median. The edges of the boxes represent the first and third quartiles. * Denotes outliers; ○ denotes far outliers.

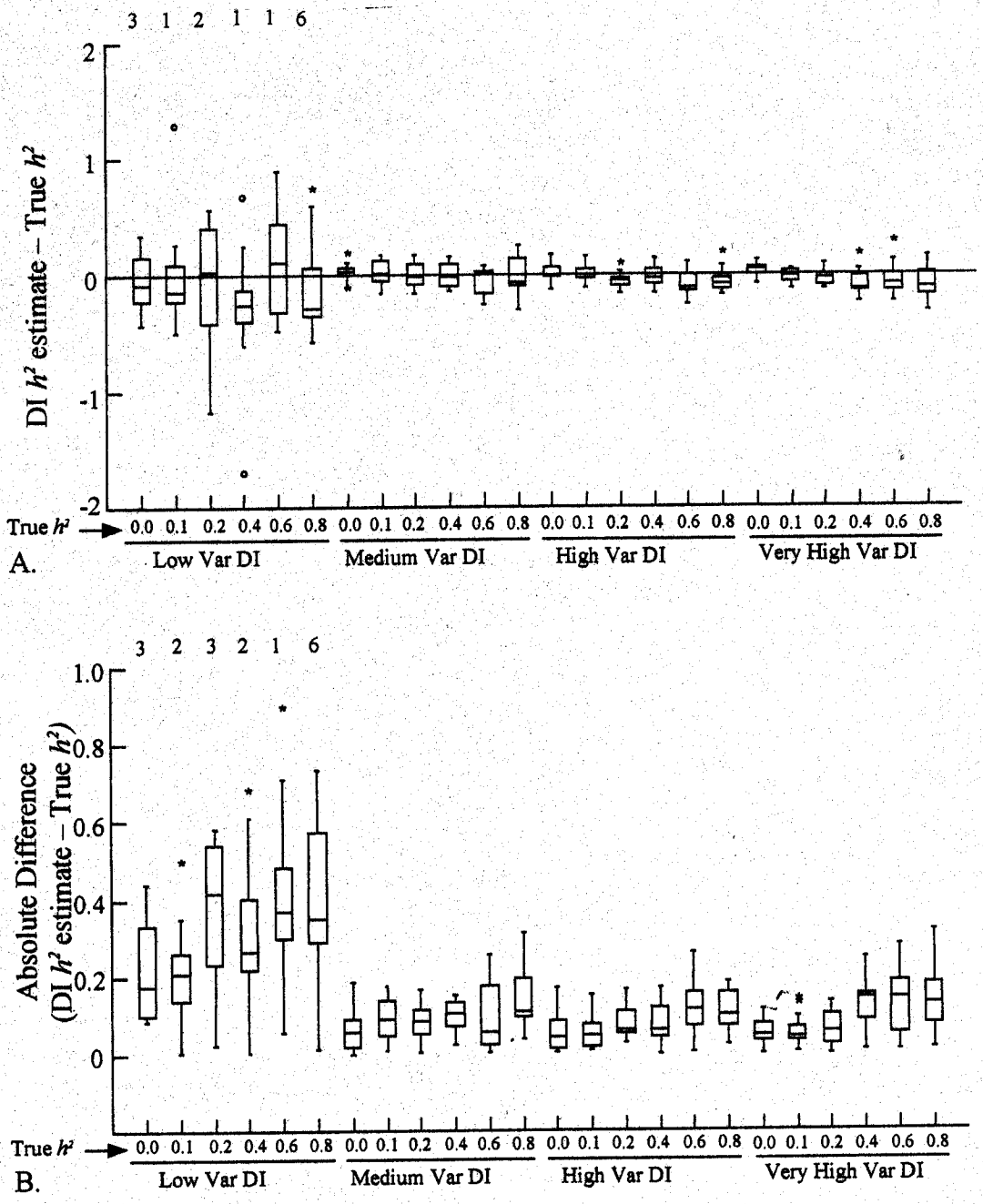

Figure 11.6 (A) Signed difference between DI h^2 estimate and the true h^2. (B) Absolute difference between DI h^2 estimate and the true h^2. Box plots are shown for all combinations of DI Variation and true heritability. Numbers above each graph refer to the data points that lay off the scale of the graph. $N = 15$ for each treatment. * Denotes outliers; ○ denotes far outliers.

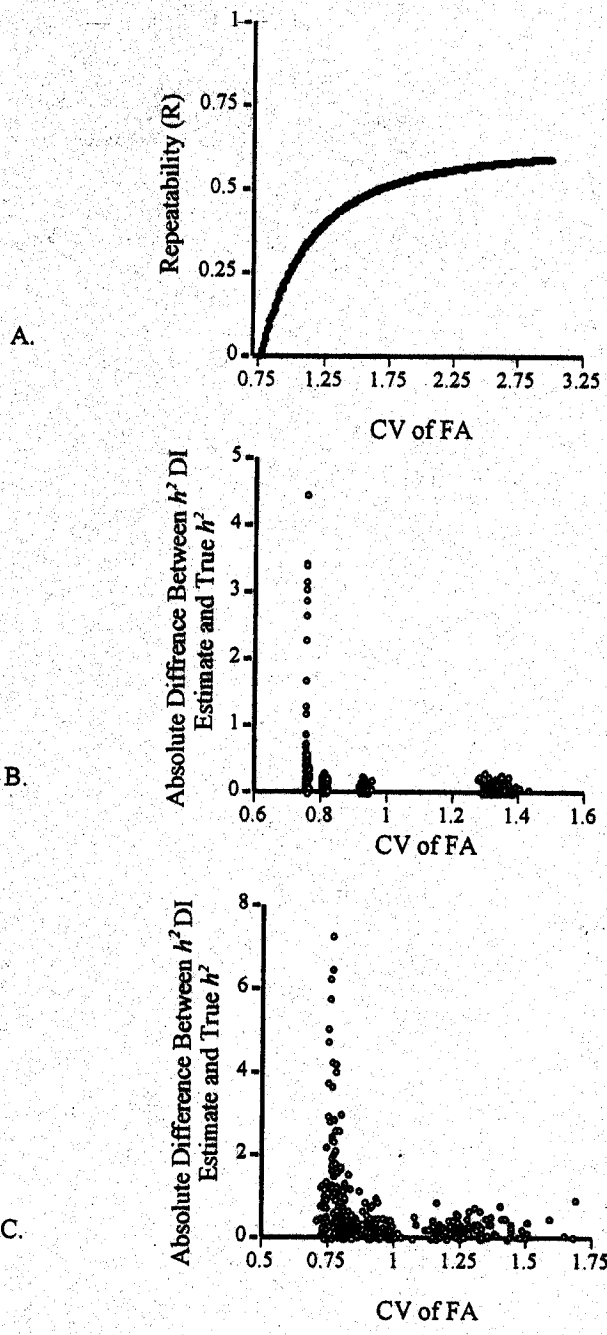

Figure 11.7 (A) The relationship between the coefficient of variation of FA (CV of FA) and the hypothetical repeatability (see Whitlock 1998). (B and C) The relationship between CV of FA and the absolute difference of the true and estimated heritability of DI. (B) Data from 360 simulations with 50,000 offspring per simulation. Six data points occur beyond the scale of the graph. (C) Small simulation: data from 360 simulations with 500 offspring per simulation. Fifteen data points occur beyond the scale of the graph. Note the change in scaling on on the *y*-axes in graphs B and C.

ing influence of other likely sources of covariance among relatives. Consequently these designs are not considered.

General Approach

Simulated DI and FA data were generated using the above model assuming that the proportion of variation due to additive genetic variance was 0.1, 0.4, or 0.8. Unless otherwise noted, the maternal family effect was assumed to be absent. We compared breeding designs by manipulating the numbers of sires, dams, and clutches. For the nested half-sib design, we simulated every possible combination of sires, dams, and clutch size that resulted in a total sample size of 500 and had a minimum of two individuals in a category (e.g., 2 sires, 5 dams/sire, 50 offspring/clutch; 5 sires, 2 dams/sire, 50 offspring/clutch, etc.), for a total of 24 different breeding combinations. In nested half-sib designs, we tested the effect of sire using the mean square of dams nested within sires as the error term unless otherwise noted. For parent–offspring regression, unrelated full-sib families were simulated for every combination of sires and clutch size that resulted in a total sample size of 500 (e.g., 250 sires, 1 dam/sire, 2 offspring/clutch; 125 sires, 1 dam/sire, 4 offspring/clutch, etc.). For each breeding combination, we performed 100 replicate simulations. We first generated parents, then generated their offspring based on the parental breeding values. The effects of maternal environment (when present) and individual variances were assumed to be independent of the parental values.

Overall Power and Breeding Design

The average probability values generated by the three methods (nested half-sib ANOVA averaged over all 24 combinations of sire and dam numbers, midparent–offspring regression, and one parent–offspring regressions) are shown in figure 11.8. Clearly, breeding designs with only 500 offspring can only detect an effect of sire when total variation in DI and the true heritability of that variation are high. Note that the average *P*-value never approaches 0.05, as expected if power were high.

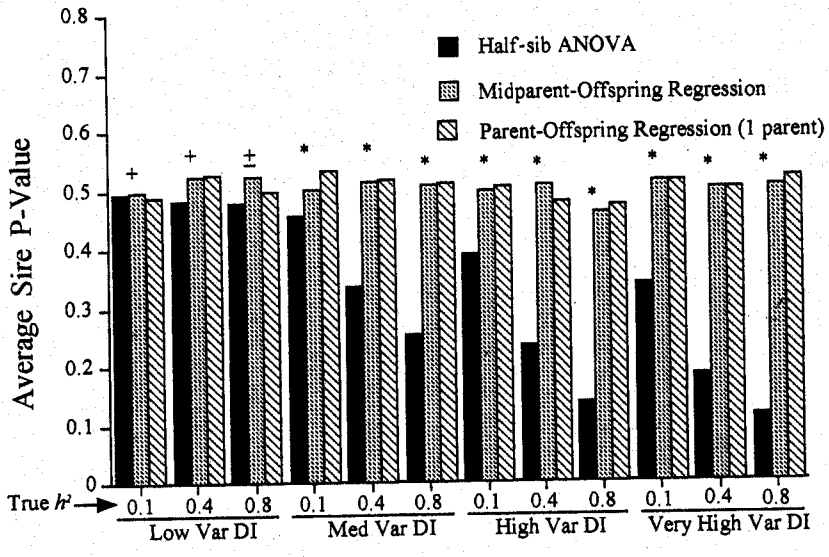

Figure 11.8 Average significance of sire effect for different methods across different levels of variation in DI and true heritability. Means are shown. $N = 800$ simulations for each mean for both midparent–offspring regression and parent–offspring regression based on one parental value. $N = 2400$ for half-sib ANOVA. Data are pooled across different breeding combinations. *P*-value of sire effect for half-sib ANOVA is calculated using the dam mean square as the error term. *Half-sib ANOVA significantly different from both midparent–offspring and single parent–offspring regression at $P < 0.0001$. + Half-sib ANOVA significantly different from midparent–offspring regression at $P < 0.05$. ± Half-sib ANOVA significantly different from parent–offspring regression at $P < 0.05$.

The second result made clear in figure 11.8 is that the nested half-sib design is far more effective than parent–offspring regression at detecting genetic variation in FA (overall t-test, $t = 51.38$, $P < 0.0001$, mean difference = 0.176) and mid-parent–offspring regression (overall t-test $t = 51.83$, $P < 0.0001$, mean difference = 0.178). This enhanced performance of the nested half-sib design occurs despite the fact that we have lumped all the simulated combinations of sire and dam numbers together for comparison, and some of these combinations have greater power than others. The power of the regressions are not better than the type I error rate, even when the variation in DI and the true heritability were high. Regressions are poor methods for determining the genetic basis of DI. This suggests that all of the parent–offspring regressions in the literature offer no information about the inheritance of FA or DI.

The FA of a parent is only a single realization of DI, and single realizations of DI provide little information concerning the true value of DI, and thus very little predictive information about DI of its offspring. Nested half-sib designs outperform the regression methods because they use information on family means of FA.

Optimizing the Nested Half-Sib Design

The performance of each of the 24 combinations of dam and sire numbers is compared in figure 11.9. Clearly, there is a large effect of family size on the ability to detect a sire effect, with its additive component ($F_{23,21384} = 9.77$, $P = 0.0001$). Power is highest when there are relatively few sires (4) and a large number of dams per sire (25). As we would expect, sire effects are also more easily detected when there is high variation in DI and high heritability of DI (DI variation: $F_{2,21599} = 1189.48$, $P = 0.0001$; heritability: $F_{2,21599} = 551.35$, $P = 0.0001$). The effect of breeding design also varies with DI variation and the true heritability (breeding design by DI variation: $F_{46,21599} = 3.36$, $P = 0.0001$; breeding design by true heritability: $F_{46,21599} = 1.52$, $P = 0.0132$; breeding design by DI variation by true heritability: $F_{92,21599} = 1.37$, $P = 0.0110$). When DI variation and true heritability are high, it is easiest to detect an effect of sire with small clutches (2 offspring/clutch), and with an intermediate number of sires and dams per sire (10) (figure 11.9A). When overall variation is low, and it is difficult to detect an effect of sire, there is little difference between the breeding designs because none of the designs have any power. For intermediate conditions, breeding design has a dramatic effect on P-value.

These results should not be taken as a definitive indication of the optimal design. In this simulation, we used minimization of the P-value of the sire effect as our criterion for an effective breeding design. Other criteria, such as accuracy, could produce slightly different results. Therefore we would take the numerical recommendation of four sires with a healthy dose of skepticism, but we do suggest that empiricists maximize the number of dams per sire. Family means are essentially the units of measurement. Because of the loose relationship between DI and FA, many family means (i.e., replications) are needed to precisely measure the effect of an individual sire.

To Pool or Not To Pool

Another issue that affects the analysis of nested half-sib designs is the decision about whether to pool dam and error terms when the dam-level effect is not significant. Normally, the effect of sire is tested using the mean square of dams nested within sires as the error term. However, some studies (e.g., Tomkins and Simmons 1999) have tested the sire effect by pooling dams and testing over the mean square error, provided that the dam effect was not significant. In doing so, number of degrees of freedom in the error term is increased, which in turn increases the power of the test. Unfortunately, this procedure is always inappropriate in a nested half-sib design, as the dam component includes an additive genetic causal component, as does the sire component. Assuming that the dam component is equal to the error component is equivalent to assuming that there is no additive genetic variance in the trait.

Pooled tests have been used in several published studies as a last attempt to detect additive genetic variation. Despite the fact that we believe pooling to be inappropriate, it is useful to investigate the effect of pooling on power. We first tested the effect of pooling when there was no additional effect of maternal family above that provided by additive genetic effects for low, medium and high levels of variation in DI. Here, we investigate the effects of the practice of pooling on the ability to detect an effect of sire.

We first tested the effect of pooling when there was no additional effect of maternal family above

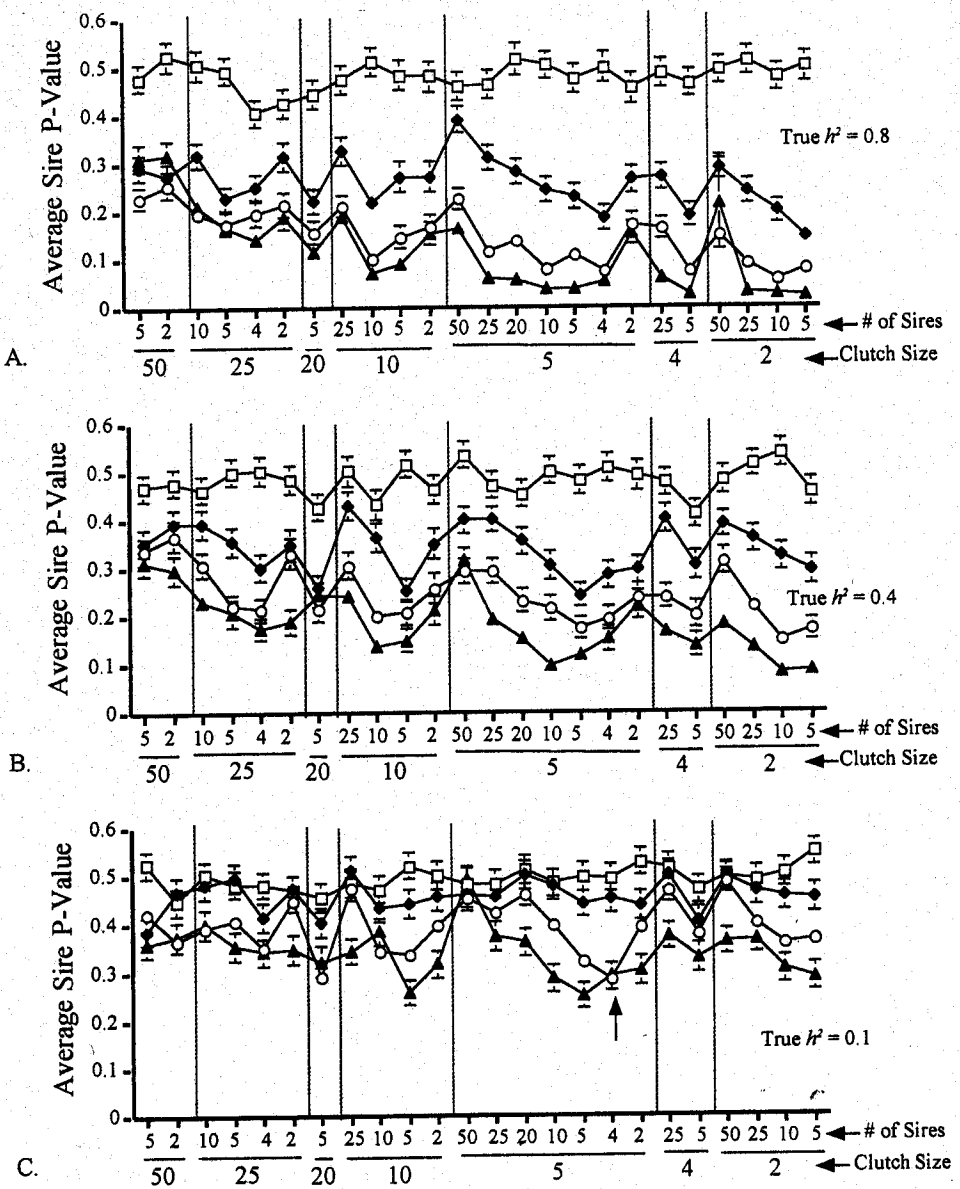

Figure 11.9 Average *P*-values of sire-effect from a half-sib ANOVA of variance in relation to the breeding design (sire-dam-clutch combination) across treatments for a half-sib breeding design using dam mean-square as the error term. The *x*-axis shows the combination of clutch and sire. (A) True heritability = 0.8. (B) True heritability = 0.4. (C) True heritability = 0.1. (C) The arrow indicates the best breeding combination for average levels of variation in DI and average DI heritability as indicated by our literature review. The total number of offspring generated is 500 for each simulation. For example, if clutch size = 50 and sire = 5, then the number of dams must be 2 (5 sires, 2 dams per sire, 50 offspring/clutch = 500 offspring total). Means and standard errors are shown. $N = 100$ simulations for each data point. Open squares, black diamonds, open circles, and black triangles denote low, medium, high, and very high variation in DI, respectively.

that provided by additive genetic effects for low, medium, and high levels of variation in DI. In these simulations we first tested the dam term over the error term. If the *P*-value for this test was greater than 0.15, we pooled, otherwise we did not. These results are shown in figure 11.10a. The average *P*-value of the sire effect over all treatments was lowest when tested using the nested dam mean square as the error term (paired t-test, $t = 8.09$, $N = 21{,}600$, $P < 0.0001$). However, the effect of sire tested over a pooled error term had a slightly lower *P*-value when there was a high variation in DI and a high true heritability. These differences were quite small. The largest difference in *P*-values where pooling outperformed using dams as the error term was 0.011.

Next, we repeated this analysis assuming that the maternal family effect was equal to the individual variance, as above (results not shown). In this case, a paired t-test showed no significant difference in power overall ($t = 0.58$, $N = 21{,}600$, $P = 0.564$). As above, pooling gave rise to less power when variation in DI was small, and greater power when variation in DI was large. This increased power is spurious, as pooling gives a biased estimate of the appropriate mean square, even in cases where the dam effect is not significant.

Box–Cox Transformations

One difficulty of analyzing FA is that in many cases it has a half-normal distribution, rather than the normal distribution assumed in parametric statistical tests. Many authors have suggested using the Box–Cox algorithm for finding the optimal transformation of the data to normality (Palmer and Strobeck 1986; Swaddle et al. 1994). The Box–Cox algorithm chooses the parameter λ in the function $Y' = (Y^{\lambda} - 1)/\lambda$ to optimize the fit of Y' to a normal distribution. If the true distribution of the FA is a half-normal, as in our simulations, then this seems a very dubious undertaking. In order to create a left tail to the distribution, the transformation must magnify tiny differences very near 0. This suggests that the Box–Cox transformation will magnify error variance of small FA values, while compressing the potentially real differences in the right tail of the raw data. This should decrease the overall power of any statistical tests.

As expected from this argument, Box–Cox transformation decreases the probability of detecting an effect of sire regardless of the error term employed (figure 11.10b, paired *t*-tests, dams as error term: $t = -10.19$; pooling allowed: $t = -10.09$, $P = 0.001$, $N = 21{,}600$). The difference in average *P*-values between the two methods is small (−0.008 for both methods). Still, Box–Cox transforming the data rarely increased the ability to detect an effect due to sire, and for most treatments, slightly decreased our ability to detect an effect of sire.

The modest impact of Box–Cox on power may be due to the fact that the sire term is tested over the dam term, which is determined from multiple offspring in all of our simulated experiments. Family means will more closely approach a normal distribution than the individual FA values. When other measures of asymmetry besides FA = |L−R| are used (Palmer and Strobeck 1986; Palmer 1994), such as size-adjusted FA, the distribution of asymmetry values may not depart from normality so drastically, in which case the Box–Cox transformation could still prove useful.

Sample Size and Power

Our review of the literature suggests that the repeatability of FA is fairly high, but the heritability of FA is low. Examination of figure 11.8C suggests that a nested half-sib analysis of variance with 25 dams per sire, and five offspring per clutch maximizes power in such cases. Figure 11.11 shows the total sample size necessary to detect an effect of sire with probability 90%, for various levels of variation in DI and genetic variation. Even when variation in DI and true heritability are high, a sample size of approximately 1000 offspring is required to achieve high power. When variation in DI is low, 50,000 to 100,000 offspring are needed to achieve high power according to our model. There is a large discrepancy between the sample sizes required to achieve high power and the actual sample sizes employed in most previous studies (table 11.1).

Discussion

In this chapter we have explored the consequences of the "standard model" of the relationship between asymmetry and developmental instability (DI) for studies of the inheritance of DI. Our main conclusion is that the question of whether there is additive genetic variation in DI and hence whether DI is capable of evolving in the face of selection is

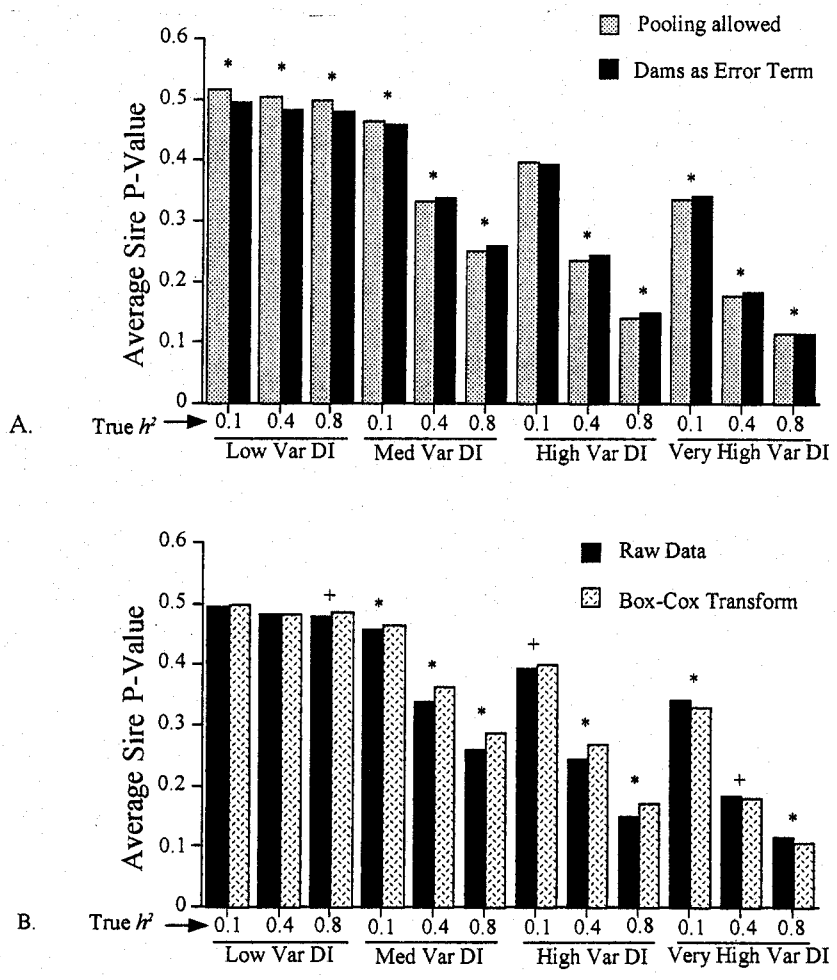

Figure 11.10 The ability to detect a significant effect due to sire. (A) Analyses where a pooled error term is allowed (provided the P-value for the dam effect is greater than 0.15) versus analyses where the dam mean square serves as the error term. (B) Analyses on raw versus Box–Cox transformed data. These analyses were performed using the dam effect as the error term. Data is shown for each combination of DI variation and true heritability. The data are pooled across breeding combinations of sire–dam–clutch. Means are shown. $N = 2400$ for all means. + Denotes statistically significant differences ($P < 0.05$) based on paired t-tests. $^*P < 0.01$.

completely unresolved. This profound uncertainty has multiple sources, each with serious implications for the study of the evolution of developmental instability.

The first source of uncertainty is the experimental data. Our review of the literature makes clear that the heritability of FA is, on the average, quite low, as made clear in a number of recent reviews (Whitlock and Fowler 1997; Gangestad and Thornhill 1999; Van Dongen 2000a; Van Dongen and Lens 2000). Furthermore, the vast majority of estimates of the heritability of FA are not significantly different from 0. In fact, the proportion of estimates that are significantly different from 0 is very close to the expected type I error rate (0.082). Of the 15 significant estimates in our sample, eight are from full-sib or twin designs that cannot separate additive genetic variance from other genetic and nongenetic sources of variation. Similarly, five of the six parent–offspring regressions cannot separate additive genetic variance from environmental effects. The most reliable sig-

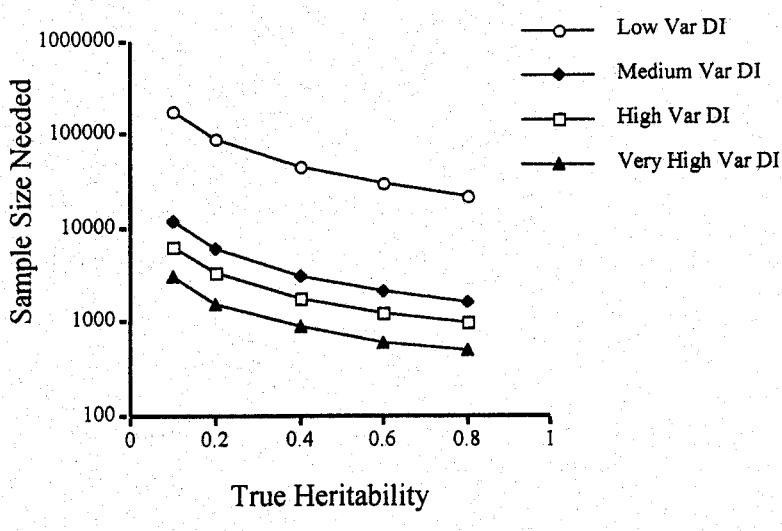

Figure 11.11 Total sample size needed to achieve power = 0.90 for a breeding design with 25 dams per sire and 5 offspring per clutch (sire number varies). We chose this breeding design because it is the most powerful for detecting additive genetic variation for average values of $\mathfrak{R}$ and average heritability of DI. Total sample size is equal to the product of dams per sire, offspring per clutch, and number of sires. Note that the *y*-axis is on a log-10 scale.

nificant heritability estimate is from an exemplary half-sib experiment on sternopleural bristle number in *Drosophila melanogaster*, with a total sample size of over 10,000 individuals (Scheiner et al. 1991). Two artificial selection experiments on sternopleural bristle fluctuating asymmetry in *D. melanogaster* have also provided significant evidence for heritability of FA (Mather 1953; Reeve 1960). Thus, there is some consistent evidence for the heritability of FA in this one character in one species. This picture is not significantly altered when we include estimates from three studies that did not give numerical results for nonsignificant estimates. Parker and Leamy (1991) found two significant relationships out of nine. Cadée (2000) found one out of 12 (the number of estimates in this study is not explicit). Woods et al. (1999) found no significant additive genetic variation (after a correction for multiple comparisons) for 12 different estimates. There is little evidence for an additive genetic component to DI in the vast majority of traits.

Conversely, our simulation results make it clear that most of the experiments that have been carried out had designs ill-suited to the detection of genetic variation in FA and therefore DI. A prime example is the common use of parent–offspring regressions. Our simulation results indicate that regressions have literally no power to detect parent–offspring resemblance in asymmetry in the range of sample sizes commonly employed. This large segment of the literature on inheritance of FA is simply not informative. Full-sib designs cannot estimate the proper causal components of variance, and are therefore not informative about the inheritance of FA and developmental instability for a different reason.

These problems pose a challenge to field studies where the assignment of parentage is difficult. Many of the studies that have used full-sib or parent–offspring designs do so because it is possible to identify maternal parentage in unmanipulated field populations, but extremely difficult to identify paternal parentage. Studying inheritance in the natural environment is a laudable goal. However, since neither of these popular designs is capable of answering the question posed, it is difficult to justify carrying out such experiments. The question of whether there is additive genetic variance for fluctuating asymmetry is one of many cases in evolutionary biology where we will be forced to rely on model systems or populations where investigators have developed extensive pedigree information.

It is also important to bear in mind that in the "standard model" of the relationship between fluctuating asymmetry and developmental instability, the heritability of asymmetry may be much less

than that of developmental instability (Whitlock 1996; Houle 1997). The degree to which heritability of FA underestimates that for DI is the inverse of the repeatability, which ranges from 1.7 to 32 times for the data in table 11.1, with an average of 5.4 times. Thus even though the average heritability of FA is certainly less than 10%, the heritability of developmental instability might still be rather high, at least for some traits. It is quite disturbing that, after almost 50 years of studies of the inheritance of asymmetry and DI, we still cannot say whether either is heritable in general, and we cannot set a meaningful upper bound on the heritability of DI.

Houle (2000) noted that the high repeatabilities of FA in some traits pose some challenges to the "standard model" of independently developing, normally distributed body parts, controlled by one overall developmental variance or developmental instability. If repeatabilities are as high as the average or highest values found in the studies reviewed here, the overall distribution of DI values must be extremely skewed, with most individuals having very low DI, with the variance due to a few individuals with very large DI values. Example distributions having such extremely skewed distributions of DI values are shown in figures 11.1 and 11.4. It is unclear what processes could give rise to such extreme distributions, particularly in genetic terms.

One explanation of this is that the "standard model" itself may be incorrect. For example, the sizes of body parts may not be normally distributed. This could arise either from developmental processes per se (Klingenberg and Nijhout 1998, 1999; Klingenberg, chapter 2, this volume) or from environmental insults such as physical traumas, starvation, or asymmetical use (Houle 2000). Organisms with serial development of homologous parts, such as plants (Wilsey and Saloniemi 1999; Freeman et al., chapter 20, this volume), could provide a useful way to estimate the underlying distributions of parts and a partial test of the standard model.

Another fundamental uncertainty is whether each individual possesses any overall ability to resist or correct the development of asymmetrical parts. While it is clear that overall asymmetry can be affected by the environment (Parsons 1990; Lens and Van Dongen 1999), it is not clear whether variation among individuals following such manipulations is due to chance, degree of exposure to the manipulation, or to some property of the individuals themselves. It is possible that the details of the development of each part of the body may render it more or less susceptible to environmental effects, but that these differences in susceptibility may be independent among traits. The inheritance of developmental instability could be of fundamental importance if there were a generalized buffering capacity that applied to all or even many morphological traits simultaneously. However, if what is evolvable is the independent susceptibility of traits to insult, the arguments that FA and DI indicate overall quality of the individual must be false. Until such time as we reliably measure the additive genetic component of variation in FA for multiple traits in the same species, and therefore detect any genetic relationship among their asymmetries it will be impossible to answer these questions through studies of whole-organism asymmetry.

In summary, we recommend that future quantitative genetic studies of asymmetry and developmental instability be carried out in species where environmental effects can be partitioned, and the genetic components of variation can be unambiguously partitioned into additive and nonadditive parts. Furthermore, the high realization variance of fluctuating asymmetry precludes the use of parent–offspring regression as an effective experimental strategy. The simplest design that meets these criteria is the nested half-sib design. More complicated designs such as diallels or pedigree analyses might also be suitable. To get the most power from a nested half-sib design, our analysis suggests that one should maximize the number of full-sib families within each half-sib family. Finally, it is clear that enormous sample sizes are usually required to detect any signal at all. The detection of additive genetic effects on asymmetry is a challenging experimental task.

Acknowledgments

We thank Jean Richardson and Alice Winn for help developing the breeding equations and SAS code. Lisa Horth, Thomas Hansen, Alice Winn, and two anonymous reviewers provided very helpful comments on an early version of the manuscript. We thank Wolf Blanckenhorn, David Hosken, Joseph Tomkins, and Leigh Simmons for giving us access to unpublished data. R.C. Fuller was supported by a University Fellowship from Florida State University. The SAS macro for the Box–Cox

transformation was created by Dr. Michael Friendly and can be found at http://www.math.yorku.ca/SCS/sasmac/boxglm.html.

References

Arnqvist G, Thornhill, R (1998) Evolution of animal genitalia: patterns of phenotypic and genotypic variation and condition dependence of genital and non-genital morphology in water strider (Heteroptera: Gerridae: Insecta). Genet Res 71:193–212.

Bennett DM, Hoffmann, AA (1998) Effects of size and fluctuating asymmetry on field fitness of the parasitoid *Trichogramma carverae* (Hymenoptera: Trichogrammatidae). J Anim Ecol 67:580–591.

Bjorksten T, Fowler K , Pomiankowski A (2000a) What does sexual trait FA tell us about stress? Trends Ecol Evol 15:163–166.

Bjorksten T, David P, Pomiankowski A, Fowler K (2000b) Fluctuating asymmetry of sexual and nonsexual traits in stalk-eyed flies: a poor indicator of developmental stress and genetic quality. J Evol Biol 13:89–97.

Blanckenhorn WU, Reusch T, Mühlhäuser C (1998) Fluctuating asymmetry, body size and sexual selection in the dung fly *Sepsis cynipsea*—testing the good genes assumptions and predictions. J Evol Biol 11:735–753.

Cadée N (2000) Genetic and environmental effects on morphology and fluctuating asymmetry in nestling barn swallows. J Evol Biol 13:359–370.

Chapman JW, Goulson D (2000) Environmental versus genetic influences on fluctuating asymmetry in the house fly, *Musca domestica*. Biol J Linn Soc 70:403–413.

Corruccini RS, Potter, RHY (1981) Developmental correlates of crown component asymmetry and occlusal discrepancy. Am J Physio Anthropol 55:21–31.

Crow EL, Shimizu K (1988) Lognormal Distributions: Theory and Applications. M Dekker, New York.

Eggert A, Sakaluk SK (1994) Fluctuating asymmetry and variation in the size of courtship food gifts in decorated crickets. Am Nat 144:708–716.

Evans AS, Marshall M (1996) Developmental instability in *Brassica campestris* (Crucifera): fluctuating asymmetry of foliar and floral traits. J Evol Biol 9:717–736.

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics, 4th Edn. Longman, Harlow.

Gangestad SW, Thornhill R (1999) Individual differences in developmental precision and fluctuating asymmetry: a model and its implications. J Evol Biol 12:402–416.

Hagen DW (1973) Inheritance of numbers of lateral plates and gill rakers in *Gasterosteus aculeatus*. Heredity 30:313–312.

Houle D (1997) Comment on "A meta-analysis of the heritability of developmental stability" by Møller and Thornhill. J Evol Biol 10:17–20.

Houle D (2000) A simple model of the relationship between asymmetry and developmental stability. J Evol Biol 13:720–730.

Klingenberg CP, Nijhout HF (1998) Competition among growing organs and developmental control of morphological asymmetry. Proc R Soc Lond B Biol Sci 265:1135–1139.

Klingenberg CP, Nijhout HF (1999) Genetics and fluctuating asymmetry: a developmental model of developmental instability. Evolution 53:358–375.

Leamy L (1997) Genetic analysis of fluctuating asymmetry for skeletal characters in mice. J Hered 88:85–92.

Leamy L (1999) Heritability of directional and fluctuating asymmetry for mandibular characters in random-bred mice. J Evol Biol 12:146–155.

Leary RF, Allendorf FW, Knudsen KL (1985) Inheritance of meristic variation and the evolution of developmental stability in rainbow trout. Evolution 39:308–314.

Leary RF, Allendorf FW, Knudsen KL (1992) Genetic, environmental, and developmental causes of meristic variation in rainbow trout. Acta Zool Fenn 191:79–95.

Lens L, Van Dongen S (1999) Evidence for organism-wide asymmetry in five bird species of a fragmented afrotropical forest. Proc R Soc Lond B Biol Sci 266:1055–1060.

Lens L, Van Dongen S, Wilder CM, Brooks TM, Matthysen E (1999) Fluctuating asymmetry increases with habitat disturbance in seven bird species of a fragmented afrotropical forest. Proc R Soc Lond B Biol Sci 266:1241–1246.

Leung B, Forbes MR, Houle D (2000) Fluctuating asymmetry as a bio-indicator of stress: comparing efficacy of analyses involving multiple traits. Am Nat 155:101–115.

Lynch M, Walsh B (1998) Genetics and Analysis of Quantitative Traits. Sinauer, Sunderland, Mass.

Mather K (1953) Genetic control of stability in development. Heredity 7:297–336.

McGrath JW, Cheverud JM, Buikstra JE (1984) Genetic correlations between sides and heritability of asymmetry for nonmetric traits in rhesus macaques on Cayo Santiago. Am J Phys Anthropol 64:401–412.

Møller AP (1994) Sexual Selection and the Barn Swallow. Oxford University Press, Oxford.

Møller AP, Thornhill R (1997) A meta-analysis of the heritability of developmental stability. J Evol Biol 10:1–16.

Palmer AR (1994) Fluctuating asymmetry analyses: a primer. In: Markow TA (ed) Developmental Instability: Its Origins and Evolutionary Implications. Kluwer, Dordrecht, pp 335–364.

Palmer AR, Strobeck C (1986) Fluctuating asymmetry: measurement, analysis, patterns. Annu Rev Ecol Syst 17:391–421.

Parker LT, Leamy L (1991) Fluctuating asymmetry of morphometric characters in house mice: the effects of age, sex, and phenotypical extremeness in a randombred population. J Hered 82:145–150.

Parsons PA (1990) Fluctuating asymmetry: an epigenetic measure of stress. Biol Rev 65: 131–145.

Pechenkina EA, Benfer RA, Vershoubskaya GG, Kozlov AI (2000) Genetic and environmental influence on the asymmetry of dermatoglyphic traits. Am J Phys Anthropol 111:531–543.

Pomiankowski A (1997) Genetic variation in fluctuating asymmetry. J Evol Biol 10:51–55.

Quinton M, Smith C, Goddard ME (1992) Comparison of selection methods at the same level of inbreeding. J Anim Sci 70:1060–1067.

Reeve ECR (1960) Some genetic tests on asymmetry of sternopleural chaeta number in *Drosophila*. Genet Res 1:151–172.

Rettig JE, Fuller RC, Corbett AL, Getty T (1997) Fluctuating asymmetry indicates levels of competition in an even-aged poplar clone. Oikos 80:123–127.

Scheiner SM, Caplan RL, Lyman RF (1991) The genetics of phenotypic plasticity. III. Genetic correlations and fluctuating asymmetries. J Evol Biol 4:51–68.

Sheridan L, Pomiankowski A (1997) Fluctuating asymmetry, spot asymmetry and inbreeding depression in the sexual coloration of male guppy fish. Heredity 79:515–523.

Swaddle JP, Cuthill IC (1994) Preference for symmetric males by female zebra finches. Nature 367:165–166.

Swaddle JP, Witter MS, Cuthill IC (1994) The analysis of fluctuating asymmetry. Anim Behav 48:986–989.

Tomkins JL, Simmons LW (1999) Heritability of size but not symmetry in a sexually selected trait chosen by female earwigs. Heredity 82:151–157.

Van Dongen S (1998) How repeatable is the estimation of developmental stability by fluctuating asymmetry? Proc R Soc Lond B Biol Sci 265:1423–1427.

Van Dongen S (2000a) The heritability of fluctuating asymmetry: a Bayesian hierarchical model. Ann Zool Fenn 37:15–23.

Van Dongen S (2000b) Unbiased estimation of individual asymmetry. J Evol Biol 35:79–85.

Van Dongen S, Sprengers E, Loefstedt C, Matthysen E (1999) Heritability of tibia fluctuating asymmetry and developmental instability in the winter moth (Operophtera brumata L.) (Lepidoptera, Geometridae). Heredity 82:535–542.

Van Dongen S, Lens L (2000) The evolutionary potential of developmental instability. J Evol Biol 13:326–335.

Van Valen L (1962) A study of fluctuating asymmetry. Evolution 16:125–142.

Watson PJ, Thornhill R (1994) Fluctuating asymmetry and sexual selection. Trends Ecol Evol 9: 21–25.

Whitlock M (1996) The heritability of fluctuating asymmetry and the genetic control of developmental stability. Proc R Soc Lond B Biol Sci 263:849–854.

Whitlock M (1998) The repeatability of fluctuating asymmetry: a revision and extension. Proc R Soc Lond B Biol Sci 265:1429–1431.

Whitlock MC, Fowler K (1997) The instability of studies of instability. J Evol Biol 10: 63–67.

Wilsey BJ, Saloniemi I (1999) Leaf fluctuating asymmetry in tree-line mountain birches, *Betula pubescens* ssp. *tortuosa*: genetic or environmentally influenced? Oikos 87:341–345.

Windig JJ (1998) Evolutionary genetics of fluctuating asymmetry in the peacock butterfly (*Inachis io*). Heredity 80:382–392.

Woods RE, Hercus MJ, Hoffmann AA (1998) Estimating the heritability of fluctuating asymmetry in field *Drosophila*. Evolution 52:816–824.

Woods RE, Srgó CM, Hercus MJ, Hoffmann AA (1999) The association between fluctuating asymmetry, trait variability, trait heritability, and stress: a multiple replicated experiment on combined stresses in *Drosophila melanogaster*. Evolution 53:493–505.