J. evol. biol. 10 (1997) 39–49
1010–061X/97/010039–11 $1.50 + 0.20/0

Journal of Evolutionary Biology

Commentary

Fluctuating asymmetry and developmental stability: heritability of observable variation vs. heritability of inferred cause

A. R. Palmer* and C. Strobeck

Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada, e-mail: Rich.Palmer@UAlberta.ca

Key words: Developmental noise; directional asymmetry; antisymmetry; statistics; measurement error; kurtosis; methods; analysis; sexual selection; genetic; phenotypic.

Introduction

Fluctuating asymmetry – subtle random deviations from perfect bilateral symmetry – is an appealing measure of developmental precision because of the apparent ease with which it may be measured and because its developmental origins seem so straightforward (Palmer, 1996). This appeal has led to its wide application as a measure of developmental stability in studies of inbreeding and outbreeding depression (reviewed in Palmer and Strobeck, 1986; Graham, 1992), as a measure of genetic or environmental stress in biomonitoring studies (Leary and Allendorf, 1989; Graham et al., 1993), and most recently as a measure of fitness or mate quality in studies of sexual selection (e.g., see Møller, 1994; Tomkins and Simmons, 1995). Criticisms of the uses of fluctuating asymmetry (FA) have focused mainly on methodological issues (Palmer and Strobeck, 1992; Fields et al., 1995), but doubts have also been raised about the developmental origins of subtle bilateral variation (reviewed in Markow, 1994; Palmer, 1996). If proponents of FA are not more reflective about these methodological and conceptual issues, the whole approach may become tarnished. This point was made emphatically by Phil Hedrick in his concluding comments to the symposium on developmental instability at which both Møller and Thornhill were present (Markow, 1994, pp. 434–435).

For bilateral variation to serve as a credible measure of developmental precision, three essential statistical tests of procedures must be applied (Palmer, 1994, 1996; Swaddle et al., 1994). First, because true FA (the result of real differences between

* Author for correspondence.

sides) is indistinguishable from measurement error (biologically meaningless differences between sides), the between-sides variation must be significantly greater than the variation due to measurement error. Second, because measurement error actually inflates and therefore biases measures of FA upwards, the variation due to measurement error must be factored out to yield credible *quantitative* measures of bilateral variation. Third, because departures from 'ideal' FA may be caused by factors other than developmental noise, the between-sides variation must meet the statistical criteria for 'ideal' FA. Although other factors may also be important (Palmer, 1994), if these three issues are not addressed, then quantitative measures of bilateral variation and their heritability can not be interpreted with much confidence at all.

Møller and Thornhill (this issue) provide a valuable service by tabulating the many studies that contain information on the heritability of bilateral variation. For this they are certainly to be commended. Such information is often buried deeply, particularly in the older literature, and is not readily accessible. In addition, their *qualitative* conclusion may, in fact, be correct: factors that influence the magnitude of subtle bilateral variation may very well have a weakly heritable basis, at least in some traits or species.

So why criticise such an extensive and seemingly worthwhile survey? Unfortunately, it suffers from two serious shortcomings elaborated upon below: I) the quantitative data are tabulated with very little critical evaluation, and II) the interpretation sidesteps or ignores some of the legitimate and widely acknowledged questions about the developmental origins of bilateral variation. Given their visibility, Møller and Thornhill could do the community a great service if they were to lead by example and temper their enthusiasm with greater caution and discrimination. By not subjecting the studies they tabulate to the three key criteria outlined above, and by ignoring or obscuring acknowledged conceptual difficulties, they appear to rely on the size of their table and the magic of meta-analysis to give a false confidence to both their qualitative and quantitative result and therefore an inflated credibility to their interpretation.

Quantifying and interpreting heritability estimates

I) Validity of published heritability estimates

Confounding effects of measurement error

The problems that measurement error creates for studies of FA are widely recognized (Palmer and Strobeck, 1986; Graham et al., 1993; Swaddle et al., 1994; Fields et al., 1995; Merilä and Björklund, 1995). Unlike the impact of measurement error on conventional morphological traits, where it merely reduces confidence in the data, measurement error weakens FA analyses in two ways: it reduces confidence in the data and it artificially inflates quantitative measures of FA. Unless measurement error has been factored out using established statistical techniques (e.g., Palmer, 1994; Fields et al., 1995; Merilä and Björklund, 1995), quantitative measures of FA will represent a nearly uninterpretable mix of true bilateral variation and measurement error.

What are the implications of this for Møller and Thornhill's meta-analysis of heritability? First, studies that reported no significant heritability to FA variation and that did not conduct an error analysis should have been excluded from any meta-analysis because these heritability estimates may be completely meaningless biologically (e.g., Bailit et al., 1970; Eggert and Sakaluk, 1994; and possibly others). Including such studies artificially biases the estimates of average heritability downwards. Unfortunately, Table 1 of Møller and Thornhill does not distinguish heritability estimates based on sound data from those where measurement error may have obscured the true underlying FA variation.

Second, the effects of measurement error on estimates of the heritability of FA are unpredictable. Where it is constant and truly random (i.e., unaffected by among-observer differences, or within-observer differences over time), measurement error should generally decrease estimates of heritability of FA as it does for conventional traits, because it increases the apparent environmental component of variation. However, because it biases measures of FA upwards, measurement error can, under certain conditions, actually yield a significant heritability estimate that is entirely an artifact of the measurement protocol.

For example, if different observers take measurements on different sets of parents and their offspring, among-observer differences in measurement error yield an artificial correlation between the apparent bilateral variation of parents and their offspring when the data from several observers are pooled in a single analysis. Thus, Livshits and Kobyliansky (1989) report a heritability of 0.224 for a suite of eight anthropometric traits in Israeli parents and their children (see Table 1 of Møller and Thornhill). They report "reliability coefficients approximating 0.98" (p. 122), but these reliability coefficients were for the trait dimensions measured, not for the asymmetry measurements. However, in a follow-up error analysis using seven of the same traits in a second sample of Israeli infants, Fields et al. (1995, their Tab. 3) found that even though the mean reliability coefficient for trait dimensions was 94.9 ($N = 7$, range 92.4–97.8), the measurement error variation for these traits was 2.2 times greater than the between-sides variation [FA = var($R - L$)]. Thus, because the measurements used in the analysis by Livshits and Kobyliansky (1989) may have been taken by different nurses, where each nurse would likely measure both the parents and offspring in a single family, some unknown and perhaps large fraction of the mid-parent/child heritability could be due to measurement error differences among nurses. Unfortunately, insufficient information is given in Livshits and Kobyliansky (1989) to determine just how large this bias may have been. Furthermore, FA measurements in other studies cited by Møller and Thornhill (e.g., Mason et al., 1967) were also collected by more than one individual. How many other studies in their Table 1 might be similarly compromised?

For the quantitative measures of heritability to be credible and comparable, Møller and Thornhill should have indicated in their Table 1 a) which studies confirmed that bilateral variation was greater than measurement error, b) how large the measurement error was relative to the bilateral variation (e.g., as SD of repeat measurements compared to the SD of bilateral variation), and c) which studies may

have been compromised by among-observer differences, or within-observer differences over time. Without such information, the tabulated heritability estimates are exceedingly difficult to interpret. Because the differences between sides for most bilateral traits are exceedingly small ($< 1\%$ of trait size; Palmer, 1996), and because of the bias that it introduces into quantitative measures of FA and the heritability of FA, measurement error simply can not be dismissed as unimportant.

Confounding effects of direction asymmetry and antisymmetry

A central and critical question underlies all attempts to use bilateral variation to infer the level of developmental stability: What is its biological origin? Causes of bilateral variation *other* than developmental noise will not yield valid inferences about levels of developmental stability. Møller and Thornhill claim they are using published measures of the heritability of bilateral variation to infer the heritability of "developmental stability". However, differences between sides may not always serve as valid proxies for developmental stability (Graham et al. 1993; Palmer, 1994, 1996; see also *heritability of phenotypic variation vs. heritability of causes* below).

The two most common patterns of bilateral variation that suggest causes other than developmental noise are directional asymmetry (repeatable asymmetry towards the same side) and antisymmetry (repeatable asymmetry that is random with respect to side). Although Møller and Thornhill acknowledge these two kinds of bilateral variation are a potential problem (see *Data set* section of *Materials and methods*), their treatment of such deviations from ideal FA seems unforgivable. Although requested to do so by reviewers, they still do not indicate clearly which studies tested for directional asymmetry (DA) *and* which also tested for antisymmetry. If either DA or antisymmetry are present, or if tests for DA or antisymmetry are not reported, then the bilateral variation may not yield a valid measure of developmental stability (Graham et al., 1993; Palmer, 1994; Swaddle et al., 1994) and we cannot determine whether the published heritability estimates relate to developmental stability or to some other predisposition towards asymmetry.

Although formal tests for the presence of antisymmetry are an essential requirement when using deviations from symmetry as a measure of developmental precision, such tests if conducted at all are rarely reported. To complicate matters further, some studies claiming to describe patterns of FA variation actually show clear evidence of antisymmetry. Thus in the case of sheep blowflies, increased bilateral variation following the initial evolution of pesticide resistance (Clarke and McKenzie, 1987) was later shown to be antisymmetry, not FA (McKenzie and Clarke, 1988). In addition, when such data are presented so that the variation may be examined in more detail, the outer tail feathers of barn swallows, and the petals of flowers, exhibit a most peculiar pattern: less well-developed feathers or petals exhibit antisymmetry and more well-developed feathers or petals exhibit FA [see, for example, Fig. 1 of Møller (1990), Fig. 2a of Møller (1992), Fig. 1a of Møller and Eriksson (1994), and commentary in Rowe et al. (submitted). Thus, even where antisymmetry is conspicuous, it may be either overlooked or ignored.

Unfortunately, the summary of tests for DA and antisymmetry in earlier studies (Tab. 1 of Møller and Thornhill) is deceptive and misleading because, in the column under 'FA test', the entry '1' has multiple meanings. Møller and Thornhill presumably intend a '1' in this column to mean that the study tested for and confirmed the presence of ideal FA (mean = 0, normally distributed variation), since they use this information to exclude studies coded as 0 from one of their meta-analyses. However, a '1' actually can mean any one of three possibilities: i) the original study tested *only* for DA and found none, or ii) the original study tested *only* for antisymmetry and found none, or iii) the original study tested for *both* DA and antisymmetry and found neither.

Clearly, only if a '1' indicates case (iii) can Møller and Thornhill conclude that the original study measured the heritability of FA and not the heritability of some other type of bilateral variation. Yet Mason et al. (1967), Martin et al. (1982), Livshits and Kobyliansky (1989), Tuinstra et al. (1990) and Thornhill and Sauer (1992) *only* report tests for DA and, so far as we can tell, either did not test for antisymmetry or did not report the results of their tests. How many other studies did not test for antisymmetry? More seriously, Mason et al. (1967; p. 88) and Livshits and Kobyliansky (1989; p. 123) clearly state that they detected significant DA in their samples, and antisymmetry is clearly evident in the tail feathers of shorter-tailed male barn swallows (Fig. 1 of Møller, 1990, and Fig. 2a of Møller, 1992). However, all of these studies are still coded as a '1' under 'FA Test' in Table 1, giving the mistaken impression that the original studies were actually quantifying ideal FA. How can readers possibly judge the content of this table without referring back to the original sources?

If Møller and Thornhill wished to present these published data in a judicial and critical manner, they should have indicated clearly what is known about the bilateral variation examined in each of the original studies. This could have been accomplished by including two separate columns indicating a) which studies specifically tested for DA and b) which specifically tested for antisymmetry. In fact, as noted above, many studies did not explicitly test for antisymmetry, so Møller and Thornhill should have acknowledged this and been frank about how it affected their meta-analysis and conclusions. Since reviewers alerted them to this problem, why were these potentially confounding factors ignored?

Finally, statistical tests for departures from normality in the direction of antisymmetry (i.e., negative or platykurtosis) are distressingly low in power, and hence rather large sample sizes are required to detect antisymmetry reliably (Palmer and Strobeck, in preparation). Thus, for example, a sample size of at least 100 is required to detect a significant departure from normality ($\alpha = 5\%$) as little as 50% of the time even when the difference between the two peaks of a bimodal distribution seems large (e.g., is equal to twice the SD of the variation about one peak), and even when using the proper SE to test for platykurtosis. Although Møller and Thornhill can be excused for not being aware of this additional problem, it raises questions about the true form of the bilateral variation in earlier studies, and thus about whether the heritability estimates reported in their Table 1 really represent heritability of FA as opposed to subtle antisymmetry.

Confounding effects of overall size variation

Overall size variation can introduce unwanted sources of error in studies of FA in two ways. First, because measures of FA are measures of variability, and because most measures of variability increase with trait size (Lande, 1977), non-random differences in overall size among groups being compared may generate artificial differences in measures of variability such as FA. Typically, such potentially confounding effects are removed by scaling the differences between sides by the average trait size or by some other measure of body size (Palmer, 1994). Second, where the bilateral variation exhibits *no* association with trait size, an uncritical application of size scaling can also introduce unwanted effects where non-random differences in overall size among groups are present. Both of these sources of error may confound estimates of heritability of developmental stability with heritable variation in overall body size.

Thus, for example, Thornhill and Sauer (1992) report the results of a regression analysis of FA of forewing lengths of sons and daughters against that of the fathers. Remarkably, in light of the small size of most deviations from symmetry, and in light of all the attendant problems associated with measurement error outlined above, their analysis yields nearly perfect heritability of asymmetry variation (1.072), even though, in theory, such values should never exceed 1.0. However, the index of FA they used was: $(R-L)/[(R+L)/2]$. Because they divided the difference between sides by the mean, without showing that this 'size-correction' was necessary, the apparent high slope to the parent-offspring regression of FA might really reflect heritable variation in some 'size-factor' (as measured by $[(R+L)/2]$) rather than heritable variation in $(R-L)$, since body size variation is often highly heritable (Futuyma, 1986). Without a test for the potentially confounding effects of overall size, we cannot know what fraction of the heritable variation Thornhill and Sauer report is in 'general size' as opposed to developmental stability, or to what extent the parent offspring correlation was influenced by a paternal nutrition effect (Eggert and Sakaluk, 1994).

Curiously, Møller and Thornhill only acknowledge one case of size-scaled variation: "one estimate of the heritability of individual fluctuating asymmetry was based on asymmetries corrected for character size . . . (Møller, 1994)". Yet size corrections were also applied in at least four other studies in Table 1 either via correlation coefficients of R vs. L (Bailit et al., 1970; Townsend and Brown, 1980) or via ratios of $(R-L)/[(R+L)/2]$ (Livshits and Kobyliansky, 1989; Thornhill and Sauer, 1992).

Here again, for the contents of Møller and Thornhill's Table 1 to be judged fairly as measures of the heritability of developmental stability that are not confounded by heritability in overall body size variation, they should have indicated for each study: a) whether FA was correlated with overall size or not and b) whether the index for FA incorporated size-scaling (e.g., by indicating which index was used). As they did for other possible confounding factors in their meta-analysis, they should have then indicated to what extent their conclusions depended on studies in which size-scaling was applied compared to those in which it was not. In view of

the past debate over the validity of size-scaling methods in which Møller participated (e.g., see the exchange between Harvey et al., Cuthill et al., and Møller, 1993, and references therein), why was more care not taken to assess these potentially confounding factors in this review?

Studies overlooked in the analysis

Although Møller and Thornhill tabulate the results from a commendably broad range of studies, a few others that report estimates of the heritability of deviations from symmetry were overlooked: one for *Drosophila* (Fowler and Whitlock, 1994), one for sticklebacks (Blouw and Boyd, 1992), and one for bird feathers (Price et al., 1991). Would the inclusion of these in their meta-analysis have affected their conclusions?

II) Interpreting patterns of subtle bilateral variation

Pattern vs. processes

To draw reliable inferences about the processes that influence developmental precision based on patterns of bilateral variation, the pattern of bilateral variation should conform to that expected if deviations from symmetry arose due to developmental noise. In other words, the bilateral variation should reflect the cumulative effect of small errors during development that a) are random and b) affect the right and left sides independently (Palmer and Strobeck, 1992; Palmer, 1994). The only distribution of bilateral variation that truly meets these *a priori* criteria is FA. In the case of DA, deviations from symmetry are neither random nor independent – overdevelopment of one side occurs consistently towards the same side of the body. In the case of antisymmetry, deviations from symmetry are random with respect to side but they are not independent, as required of a noise-like process – although overdevelopment occurs consistently on one side, the side that is larger varies at random. Thus for patterns of bilateral variation like DA or antisymmetry, the loss of either randomness or independence between sides strongly implies that the processes giving rise to them are not noise-like.

Although some have argued, using a complex and debatable model of development, that DA or antisymmetry can arise via a noise-like process (Graham et al., 1993c), it should be obvious that not *all* forms of DA or antisymmetry are evidence of reduced developmental stability. Are adult flatfish, where most species are exclusively right- or left-sided and thus directionally asymmetrical (Neville, 1976), more developmentally unstable than rockfish, which are effectively symmetrical? Are male fiddler crabs, which exhibit pronounced antisymmetry in their master claws, more developmentally unstable than females, whose master claws are effectively symmetrical (Crane, 1975)? Clearly, anyone wishing to infer the level of developmental stability based on the level of DA or antisymmetry must somehow

confirm that this variation has arisen via a noise-like process and not via a deterministic developmental program. Such arguments are greatly complicated, however, by well-documented examples of single-gene effects that yield antisymmetry (Shapiro, 1970; McKenzie and Clarke, 1988; Socha et al., 1993). If antisymmetry can sometimes be induced by one or a few genes, how can antisymmetry due to developmental noise be distinguished from antisymmetry due to direct genetic effects?

Clearly, without additional information about developmental processes, only FA meets the *a priori* criteria of a pattern of bilateral variation that is hard to explain except as the result of a noise-like process.

Heritability of phenotypic variation vs. heritability of cause

Møller and Thornhill fall into a trap by mixing up the heritability of observable phenotypic variation with the heritability of causes of that variation. They are not alone here, but as noted at the conference in which both of them participated (Markow, 1994), if proponents of the use of FA as a measure of development precision are not careful to distinguish patterns of variation from the processes *inferred* to give rise to those patterns, much confusion will result. They could have avoided much of the confusion in the present review by rigidly adhering to phrases referring explicitly to patterns of variation – 'heritability of FA', 'heritability of antisymmetry' or 'heritability of variability' (in the case of the plant studies) – to make it absolutely clear in the methods and results that they were focusing on the heritabilities of phenotypically distinct phenomena. Once average heritabilities for each of these different phenomena were computed from the literature, and a clear case made for which ones yield defendable inferences about developmental stability and which do not, then a logical discussion of the heritability of developmental stability is possible.

Unfortunately, the review is written so loosely we cannot determine what Møller and Thornhill actually believe about pattern and process, because so many different patterns of variation are pooled as if they represented a single biological phenomenon. Do they believe that any pattern of bilateral variation is a valid proxy for developmental stability? If not, which patterns of bilateral variation are most useful for inferring differences in developmental stability and which are potentially misleading? If tests for departures from ideal FA are absent or inadequate (e.g., Mason et al., 1967; Martin et al., 1982; Livshits and Kobyliansky, 1989; Tuinstra et al., 1990; Thornhill and Sauer, 1992; and possibly others), do Møller and Thornhill assume the pattern must have been FA or do they believe such studies should be excluded from the meta-analysis? In the plant studies, how is the variability due to developmental noise distinguished from that due to developmental plasticity in response to micro-environmental variation in growth conditions (Solangaarachchi and Harper, 1989)? In the five (or perhaps more) studies where deviations from symmetry were corrected by overall size (Bailit et al., 1970; Townsend and Brown, 1980; Livshits and Kobyliansky, 1989; Thornhill and Sauer,

1992; Møller, 1994), how much of the apparent heritable variation in asymmetry is due to variation in overall body size, which itself is often highly heritable (Futuyma, 1986)?

Several examples illustrate how pattern and inferred process are jumbled together so that the quantitative conclusions of Møller and Thornhill's review are almost meaningless. First, the abstract clearly emphasizes a focus on FA "Here we review . . . published estimates of one measure of developmental stability, the degree of individual fluctuating asymmetry . . . The overall mean effect size of heritabilities of individual fluctuating asymmetry was 0.19 from 34 studies of 17 species", yet 12 of these studies are coded as '0' under 'FA test' in Table 1, indicating either that they did not meet the criteria of FA (e.g., Mather, 1953) or that no test was conducted (most remaining studies). At least two of the remaining 22 studies report significant DA (Mason et al., 1967; Livshits and Kobyliansky, 1989) but are nonetheless coded as '1' (ideal FA) in Table 1. In addition, one of the three plant studies (Bagchi and Iyama, 1983) quantifies "days to first flowering and plant height", yet it too is coded as '1' (ideal FA) in Table 1. Do Møller and Thornhill wish to broaden the definition of FA to include days to first flowering and plant height?

Second, the results section also places great emphasis on the pooled estimates from all 34 studies in Table 1, with little regard for the different patterns being measured and their likely different developmental origins. Thus, "Nine out of 34 estimates were statistically significant which is five times the number of studies predicted to reach statistical significance by chance", and "It is clear from the graph [cumulative frequency distribution of mean heritability estimates for the 14 species] that most estimates are small with a median value of 0.21". How do these patterns change when: a) only those studies that confirmed deviations from symmetry were greater than measurement error are included (see *Confounding effects of measurement error* above)? b) only studies that meet the rigid criteria for ideal FA are included (see *Confounding effects of directional asymmetry and antisymmetry* above)? or c) only studies in which potentially confounding effects of overall size variation are included (see *Confounding effects of overall size variation* above)? The attempts to assess some confounding factors in the meta-analysis (e.g., FA test, internal validity, and type of heritability study) are commendable, but surely these other factors warrant attention as well.

An average value versus an universal value

It is not clear whether Møller and Thornhill appreciate the difference between an average value of a variable process and a variable estimate of a universal one. In their abstract they emphasize: "The mean heritability for 14 species was 0.27. This indicates that there is a significant additive genetic component to developmental stability". What do they intend by these statements? Do they wish to imply that developmental stability variation is heritable to roughly the same extent in all traits and all species? Do they wish to assure others, or perhaps themselves, that anyone

can safely conclude that developmental stability is heritable in their own system without conducting a heritability study? If not, then why focus so much attention on an overall average value as opposed to the potentially far more informative patterns of variation among traits or species? Other more reflective reviews of variation in heritability estimates have arrived at more satisfying conclusions (e.g., Mousseau and Roff, 1987).

Conclusion

Unfortunately, Møller and Thornhill provide only enough information in their review to make the *qualitative* point they wished to make at the outset: FA appears to have a heritable component more frequently than expected by chance. Yet for the reasons outlined above, even this qualitative conclusion is open to doubt.

More seriously, when they have ignored or obscured the many potentially confounding effects of widely acknowledged difficulties with the original data and analyses, their use of a meta-analysis to buttress claims for a robust *quantitative* estimate seems misleading at best or deceptive at worst. Since several of the concerns outlined above were raised by reviewers, why were Møller and Thornhill not more judicious in their presentation and analysis? Such a seemingly dismissive attitude towards acknowledged difficulties with analyses and interpretation runs the risk of tarnishing all those who use FA as a tool, because it suggests that proponents of FA are too quick to accept uncritically the answer they expect and too willing to build a conceptual ediface on a foundation of sand. Furthermore, given the high visibility of Møller and Thornhill, newcomers to the field, or those outside of it, may be inclined by believe that such dismissiveness is considered acceptable. This seems a most unflattering message to send.

Acknowledgements

We thank L. Leamy and L. Hammond for comments on the MS, and the Natural Sciences and Engineering Research Council of Canada for continued support of our research (operating grants GP7542 to ARP and GP0502 to CS).

References

Blouw, D. M. and G. J. Boyd. 1992. Inheritance of reduction, loss, and asymmetry of the pelvis in *Pungitius pungitius* (ninespine stickleback). Heredity 68: 33–42.

Crane, J. 1975. Fiddler Crabs of the World. Ocypodidae: Genus *Uca*. Princeton Univ. Pr., Princeton.

Cuthill, I. C., J. P. Swaddle and M. S. Witter. 1993. Fluctuating asymmetry. Nature 363: 217–218.

Fields, S. J., M. Spiers, I. Herschkovitz and G. Livshits. 1995. Reliability of reliability coefficients in the estimation of asymmetry. Amer. J. Phys. Anth. 96: 83–87.

Fowler, K. and M. C. Whitlock. 1994. Fluctuating asymmetry does not increase with moderate inbreeding in *Drosophila melanogaster*. Heredity 73: 373–376.

Futuyma, D. J. 1986. Evolutionary Biology. Sinauer, Sunderland, MA.

Graham, J. H. 1992. Genomic coadaption and development stability in hybrid zones. Acta Zool. Fenn. 191: 121–132.

Graham, J. H., D. C. Freeman and J. M. Emlen. 1993. Developmental stability: A sensitive indicator of populations under stress, pp. 136–158. *In* W. G. Landis, J. S. Hughes and M. A. Lewis (Eds.), Environmental Toxicology and Risk Assessment. American Society for Testing and Materials, Philadelphia, PA.

Harvey, P. H., S. Nee and A. F. Read. 1993. Fluctuating asymmetry. Nature 363: 217–218.

Lande, R. 1977. On comparing coefficients of variation. Syst. Zool. 26: 214–217.

Leary, R. F. and F. W. Allendorf. 1989. Fluctuating asymmetry as an indicator of stress in conservation biology. Trends Ecol. Evol. 4: 214–217.

Markow, T. A. (Ed.), 1994. Developmental Instability: Its Original and Evolutionary Implications. Kluwer, Dordrecht.

McKenzie, J. A. and G. M. Clarke. 1988. Diazinon resistance, fluctuating asymmetry and fitness in the Australian sheep blowfly, *Lucilia cuprina*. Genetics 120: 213–220.

Merilä, J. and M. Björklund. 1995. Fluctuating asymmetry and measurement error. Systematic Biology 44: 97–101.

Møller, A. P. 1990. Fluctuating asymmetry in male sexual ornaments may reliably reveal male quality. Anim. Behav. 40: 1185–1187.

Møller, A. P. 1992. Parasites differentially increase the degree of fluctuating asymmetry in secondary sexual characters. J. Evol. Biol. 5: 691–699.

Møller, A. P. 1993. Fluctuating asymmetry. Nature 363: 217.

Møller, A. P. 1994. Sexual Selection and the Barn Swallow. Oxford University Press, Oxford.

Møller, A. P. and M. Eriksson. 1994. Patterns of fluctuating asymmetry in flowers: Implications for sexual selection in plants. J. Evol. Biol. 7: 97–113.

Neville, A. C. 1976. Animal Asymmetry. Arnold (Inst. Biol. Stud. Biol.), London.

Palmer, A. R. 1996. Waltzing with asymmetry. BioScience 46: 518–532.

Palmer, A. R. and C. Strobeck. 1986. Fluctuating asymmetry: measurement, analysis, patterns. Ann. Rev. Ecol. Syst. 17: 391–421.

Palmer, A. R. and C. Strobeck. 1992. Fluctuating asymmetry as a measure of developmental stability: Implications of non-normal distributions and power of statistical tests. Acta Zool. Fenn. 191: 57–72.

Palmer, A. R. and C. Strobeck. Cryptic antisymmetry: Detection and relevance to studies of developmental stability. (in preparation).

Price, T., E. Chi, M. Pavelka and M. Hack. 1991. Population and developmental variation in the feather tip. Evolution 45: 518–533.

Rowe, L., R. Repasky and A. R. Palmer. Condition-dependent antisymmetry evolution (submitted).

Shapiro, A. M. 1970. Inheritance of three genes affecting wing pattern of *Pieris protodice* Boisduval and Le Conte (Lepidoptera: Pieridae). Wasmann J. Biol. 28: 245–257.

Socha, R., O. Nedved and J. Zrzavy. 1993. Unstable forewing polymorphism in a strain of *Pyrrhocoris apterus* (Hemiptera: Pyrrhocoridae). Ann. Entomol. Soc. Amer. 86: 484–489.

Solangaarachchi, S. M. and J. L. Harper. 1989. The growth and asymmetry of neighbouring plants of white clover (*Trifolium repens* L.). Oecologia 78: 208–213.

Swaddle, J. P., M. S. Witter and I. C. Cuthill. 1994. The analysis of fluctuating asymmetry. Anim. Behav. 48: 986–989.

Tomkins, J. L. and L. W. Simmons. 1995. Patterns of fluctuating asymmetry in earwig forceps: No evidence for reliable signalling. Proc. Roy. Soc. Lond. B 259: 89–96.