J. evol. biol. 10 (1997) 31–37
1010-061X/97/010031-07 $1.50 + 0.20/0

Journal of Evolutionary Biology

Commentary

Meta-analysis of the heritability of developmental stability: a giant step backward

T. A. Markow[1,*] and G. M. Clarke[2]

[1]*Department of Zoology, Arizona State University, Tempe, AZ 85287-1501, USA*
e-mail: attam@asuvm.inre.asu.edu
[2]*CSIRO Division of Entomology, P.O. Box 1700, Canberra, ACT 2601, Australia*

Key words: Meta-analysis; heritability; asymmetry.

Introduction

Measuring developmental instability is becoming a popular activity among behavioral ecologists and conservation biologists. Because of the potential implications of this phenomenon, it is imperative that we have a clear understanding of its basis. The question of whether or not it is heritable, and if so, to what degree, is clearly relevant to the implementation of its measures in addressing a range of evolutionary questions. Because of this, we discuss below our concerns about the meta-analysis of Møller and Thornhill. Our commentary will focus upon two general points. One is the issue of appropriateness of meta-analysis for studies of heritability. The second deals specifically with the handling of the material retrieved from the literature and included in the tables and conclusions of the meta-analysis.

Meta-analysis and heritability of FA

Meta-analysis is rapidly gaining acceptance as a powerful statistical method for providing a quantitative review of results from variety of scientific studies addressing a common issue (Arnqvist and Wooster, 1995). Its roots lie in the social and medical sciences and has only more recently been applied to the field of evolution and ecology. Put simply, the procedure involves standardizing the statistical results from a number of studies into a common index (effect size), weighting the index by the sample size, averaging effect size across all studies and testing if this average effect size differs significantly from zero. Its main attraction is that it appears to be

* Author for correspondence.

able to resolve a number of issues where individual results are conflicting or inconclusive. At face value it would appear to be highly applicable to analysis of heritability of fluctuating asymmetry where results from individual studies have been varied (Møller and Thornhill *ibid*). The major issue is whether or not heritability estimates, and in particular heritability of FA, readily lend themselves to meta-analysis from both statistical and biological perspectives.

Arnqvist and Wooster (1995) suggest that meta-analysis may prove useful for comparing the degree of genetic variation among traits, i.e., comparing heritability values. The major problems with heritability estimates for this type of analysis are: 1) the fact that heritability estimates are character, population and environment specific, and 2) that they cannot be calculated with any great degree of precision (Falconer, 1981). Falconer states (p. 149) "It is important to realize that the heritability is a property not only of a character but also of the population and of the environmental circumstance to which the individuals are subjected. . . . So, whenever a value is stated for the heritability of a given character it must be understood to refer to a particular population under particular conditions". This quote is as applicable to estimates of developmental stability as to estimates of heritabilities of individual characters.

Increasingly, developmental stability is being shown to be character, population and taxon specific (Palmer and Strobeck, 1986; Clarke, 1995a) and influenced by a variety of genetic and environmental conditions (Leary and Allendorf, 1989; Parsons, 1990; Graham et al., 1993). In general, asymmetry estimates do not correlate across characters within the same individual (Leamy, 1993) although there is a tendency for significant correlation in FA estimates within populations (Soulé and Baker, 1968). Character specificity may be further confounded by a relationship between developmental stability and fitness. As with estimates of heritability for a range of characters (Mousseau and Roff, 1987), FA estimates appear to be dependent on the relationship between character symmetry and fitness (Palmer and Strobeck, 1986; Clarke, 1995a; Gummer and Brigham, 1995), with characters for which deviations from symmetry impinge on individual fitness showing lower levels of FA than characters for which symmetry seems less important. Given the observed heterogeneity of effect size among individual studies, it is unfortunate that Møller and Thornhill did not test for differences in effect size between putative fitness and non-fitness characters. Bearing such character and population specificity in mind, calculating an overall heritabilty estimate for FA across characters and taxa (which is essentially what Møller and Thornhill have done) is biologically meaningless. If developmental stability were some type of universal character across taxa, such as survival or fecundity, then perhaps the approach has some merit. Møller and Thornhill's approach is analogous to combining heritabilities for leg length in elephants, stem length in petunias, and wing length in *Drosophila* and testing to see whether there is a significant heritability for "length" as a character in its own right. Even though the wonders of meta-analysis could enable such an analysis, any attempt to do so would be little more than foolishness. This is clearly not the intended use of meta-analysis techniques.

The other major problem with employing meta-analysis to heritability values is the lack of precision that characterizes estimates of heritability. Estimating herita-

bilities from selection experiments is even less precise than using various breeding designs because it will not accurately reflect the heritability of the character in the base population (Falconer, 1981; Mather and Jinks, 1982). The lack of precision of calculated heritability values is illustrated by the typically large standard errors ($>10\%$) in most studies, and this is particularly true for estimates of FA heritability as shown in Møller and Thornhill's Table 1. Such large errors tend to homogenize estimates thus obscuring any real underlying pattern.

Another interesting feature of the heritability values reported in Table 1 is that a large number of them are actually means of a number of individual heritability values from the same study. Perusal of the large standard errors in the referenced studies illustrate how imprecise the heritability estimates actually are and how meaningless a standard error is for their average. For example, of the heritabilities averaged from Evans et al. (1996), the highest was 0.40 ± 0.19 and the lowest was 0.10 ± 0.14! In addition, it is unclear as to the test statistic used to calculate effect size when such mean heritability estimates are used. On these grounds alone, combining heritability values is clearly a biological nonsense.

A closer look at the referenced examples

Even if the issues raised above about the appropriateness of a meta-analysis were non existent, there are serious problems in the use of the values reported in the references used by Møller and Thornhill.

For one thing, the authors include many studies in which the phenomenon reported in the original paper was not even developmental instability. Not all asymmetries are fluctuating and can be used to indicate developmental instability. This is definitely true of the seven papers on human dermatoglyphics. Ridge counts are well known for their directional asymmetries (Holt, 1968), and because these were uncontrolled for in those studies, it is unlikely that the asymmetry for which the heritabilities were estimated were for developmental instability. Most of the studies of other characters did not test for directionality, and thus we cannot be certain what the heritability estimates were actually estimates of. Even if we could be confident that directionality was only a problem in the dermatoglyphic studies, the issue of measurement error, discussed in detail by Palmer and Strobeck (1986) and by Palmer (1994), and capable of creating serious problems of interpretation of fluctuating asymmetry data, is not accounted for in the majority of the studies.

In Table 1, four of the studies are unpublished, preventing the general reader from evaluating them. Of the nine studies that Møller and Thornhill found to have significant heritabilities of developmental instability, two are unpublished. Of the remaining seven, three (Singh, 1970; Mi and Rashad, 1977; Martin et al., 1982) measure directional, not fluctuating asymmetry or developmental instability. In three others reported to have significant developmental instability heritabilities, the asymmetries were not analyzed controlling for all of the error sources described by Palmer and Strobeck (1986) and Palmer (1994). These are not the only problems with the analyses.

These issues become even more problematic when we find that it is many of these same studies that are used in calculating genetic coefficients of variation in Table 2. There is not one study in Table 2 that allows for a meaningful calculation of a coefficient of genetic variation. Three of the seven values involve unpublished material. The heritability value and the coefficient of genetic variation value (Table 2) for the Scheiner et al. 1991 study, is something that neither we nor the senior author (Scheiner, personal communication) were able to reconstruct by any method of calculation from the published reference. In the Hagen 1973 paper, the heritability value, such that it was, was not for gill rakers but for lateral plates. Besides the lack of control for measurement error, scaling, or directionality, asymmetry was treated as a plus or minus character, since the differences in plate number between the right and left sides were never off by more than one. Corruccini and Potter (1981) reported seven heritabilities for dental characters. Six were zero, and one was 0.21. The authors themselves felt that their estimate of genetic variance for that character was "spuriously high". And finally, the paper from which the largest coefficient was derived was one of the dermatoglyphic studies in which directional asymmetry was never partitioned out (Martin et al., (1982).

Interpretation of the results – a hidden agenda?

In the discussion of their paper Møller and Thornhill state "The small, average heritability of developmental stability suggests that this trait is closely associated with fitness . . . Even small levels of developmental stability will have important implications for the overall performance of individuals . . . Given that symmetrically and developmentally stable phenotypes perform best, . . .". The first statement might have foundation if developmental stability was a trait in itself, which it isn't, and if the results of the meta-analysis were in any way meaningful, which they aren't for reasons given above. The latter two statements reflect a growing body of data (most of it contributed by these same two authors) of an association between developmental stability and fitness such that character symmetry provides reliable indication of overall individual fitness. This is no doubt the crux of the whole paper. Why is it so important? Both Møller and Thornhill are strong proponents of "good genes" or indicator models of sexual selection and the role that FA plays in female mate choice (Møller and Pomiankowski, 1993; Møller, 1993; Møller et al., 1995; Thornhill, 1992, 1993; Watson and Thornhill, 1994). Inherent in such models is that there be a close association between the indicator and fitness and that both indicator and fitness are heritable (Andersson, 1994). Thus a low but significant heritability value for FA provides much needed support for the hypothesis that females choose symmetrical males over asymmetrical males because FA is a reliable indicator of overall male fitness. In their current paper Møller and Thornhill cite no empirical evidence for any association between developmental stability and fitness other than a book by Møller and Swaddle, depicted as published, but at the time of this writing, remains uncompleted. Published reviews on this topic show that any association between FA and fitness is tenuous, speculative and character

specific (Markow, 1995; Clarke 1995b,c). In fact, the actual lack of data on such an association was underscored by discussions at the 1993 conference on developmental instability that both authors attended (transcribed in Markow 1994), so it is surprising that they would make such a blanket statement. Although Møller and Thornhill do not specifically address the sexual selection implications of the meta-analysis results in the current paper, there is a fear that this will happen in the future, without reference to published criticisms. We have already uncovered six papers in which the results of their unpublished meta-analysis of FA heritability have been cited and used to support their indicator hypotheses (Møller, 1995; Møller and Eriksson, 1995; Møller et al., 1995; Thornhill and Gangestad, 1994; Thornhill et al., 1995; Swaddle, 1996) and it is likely others are in press.

We are concerned about the number of inaccurate representations in the meta-analysis. For example, contrary to the statement in the introduction, neither Markow and Gottesman (1989) nor Coyne (1987) made any attempt to test for the heritability of developmental instability. In one case, where we were able to obtain the unpublished manuscript (Evans et al., 1996) the paper was found to state clearly that the only heritability value that differed significantly from zero was confounded by a maternal effect. Such misreading of the literature is very misleading, even if it is purely accidental.

Concluding remarks

The analysis of patterns of developmental stability from genetic and developmental perspectives has much to offer in the field of developmental and ecological evolutionary studies. An understanding of the genetic basis of developmental stability is fundamental to an understanding of developmental processes and is critical if stability, and in particular fluctuating asymmetry, are to be used in defining genetic and environmental stress in natural populations (Zakharov, 1989; Parsons, 1990; Graham et al., 1993; Clarke, 1994). Analyses and interpretations such as those conducted by Møller and Thornhill do little to further such an understanding, but rather downgrade not only the field of developmental stability but also that of meta-analysis which undoubtedly has considerable potential for use within evolutionary biology.

We conclude our review by emphasizing that despite the inappropriateness of the meta-analysis and the flawed interpretation of the referenced material, that Møller and Thornhill have provided a very valuable contribution to the literature: they have elegantly illustrated that there is a serious lack of conclusive studies to indicate any significant heritability to developmental instability, regardless of how it is measured.

References

Andersson, M. B. 1994. Sexual Selection. Princeton University Press, Princeton, NJ.

Arnqvist, G. and D. Wooster. 1995. Meta-analysis: synthesizing research findings in ecology and evolution. Trends in Ecology and Evolution 10: 236–240.

Clarke, G. M. 1994. Developmental stability analysis: an early-warning system for biological monitoring of water quality. Australian Biologist 7: 94–104.

Clarke, G. M. 1995a. The genetic basis of developmental stability. II. Asymmetry of extreme phenotypes revisited. The American Naturalist 146: 708–725.

Clarke, G. M. 1995b. Relationships between developmental stability and fitness: application for conservation biology. Conservation Biology 9: 18–24.

Clarke, G. M. 1995c. Relationships between fluctuating asyymetry and fitness: how good is the evidence? Pacific Conservation Biology 2: 146–149.

Falconer, D. S. 1981. Introduction to Quantitative Genetics. Longman, London.

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. Landia, J. S. Hughes, and M. A. Lewis (Eds.), Environmental Toxicology and Risk Assessment, ASTM STP 1179. American Society for Testing and Materials, Philadelphia.

Gummer, D. L. and R. M. Brigham. 1995. Does fluctuating asymmetry reflect the importance of traits in little brown bats (*Myotis lucifugus*)? Canadian Journal of Zoology 73: 990–992.

Holt, S. B. The Genetics of Dermal Ridges. Springfield, IL, Charles C. Thomas, 1968.

Houle, D. 1992. Comparing evolvability and variability of quantitative traits. Genetics 130: 195–205.

Leamy, L. 1993. Morphological integration of fluctuating asymmetry in the mouse mandible. Genetica 89: 139–153.

Leary, R. F. and F. W. Allendorf. 1989. Fluctuating asymmetry as an indicator of stress: Implications for conservation biology. Trends Ecology Evolution 4: 214–217.

Markow, T. A. 1994. Developmental Instability: Its Origins and Evolutionary Implications. Kluwer Academic Publisher, Dordrecht.

Markow, T. A. 1995. Evolutionary ecology and developmental instability. Annual Review Entomology 40: 105–120.

Mather, K. and J. L. Jinks. 1982. Biometrical Genetics. Chapman and Hall, London.

Møller, A. P. 1993. Developmental stability, sexual selection and speciation. Journal of Evolutionary Biology 6: 493–509.

Møller, A. P. 1995. Developmental stability and ideal despotic distribution of blackbirds in a patchy environment. Oikos 72: 228–234.

Møller, A. P. and M. Eriksson. 1995. Pollinator preference for symmetrical flowers and sexual selection in plants. Oikos 73: 15–22.

Møller, A. P. and A. Pomiankowski. 1993. Fluctuating asymmetry and sexual selection. Genetica 89: 267–279.

Møller, A. P., M. Soler and R. Thornhill. 1995. Breast asymmetry, sexual selection, and human reproductive success. Ethology and Sociobiology 16: 207–219.

Mousseau, T. A. and D. A. Roff. 1987. Natural selection and the heritability of fitness. Heredity 59: 181–197.

Palmer, A. R. 1994. Fluctuating asymmetry analysis: a primer, pp. 335–364. *In* T. A. Markow (Ed.), Developmental Instability: Its Origins and Evolutionary Implications. Kluwer Academic Publisher, Dordrecht.

Palmer, A. R. and C. Strobeck. 1986. Fluctuating asymmetry: measurement, analysis, patterns. Annual Review of Ecology and Systematics 17: 391–421.

Parsons, P. A. 1990. Fluctuating asymmetry: An epigenetic measure of stress. Biological Review 65: 131–145.

Soulé, M. E. and B. Baker. 1968. Phenetics of natural populations IV. The population asymmetry parameter in the butterfly *Coenonympha tullia*. Heredity 23: 611–614.

Thornhill, R. 1992. Female preference for the pheromone of males with low fluctuating asymmetry in the Japanese scorpionfly (*Panorpa japonica*: Mecoptera). Behavioral Ecology 3: 277–283.

Thornhill, R. 1993. The allure of symmetry. Natural History 9/93: 30–36.

Thornhill, R. and S. W. Gangestad. 1994. Human fluctuating asymmetry and sexual behavior. Psychology and Science 5: 297–302.

Thornhill, R., S. W. Gangestad and R. Comer. 1995. Human female orgasm and mate fluctuating asymmetry. Animal Behavior 50: 1601–1615.

Watson, P. J. and R. Thornhill. 1994. Fluctuating asymmetry and sexual selection. Trends in Ecology and Evolution 9: 21–25.

Zakharov, V. M. 1989. Future prospects for population phenogenetics. Soviet Scientific Review. F. Physiology General Biology 4: 1–79.