How repeatable is the estimation of developmental stability by fluctuating asymmetry?

Stefan Van Dongen

Pheromone Group, Department of Ecology, University of Lund, S-223 62 Lund, Sweden (stefan.vandongen@zooekol.lu.se), and Department of Biology, University of Antwerp (UIA), Universiteitsplein 1, B-2610 Wilrijk, Belgium (svdongen@uia.ua.ac.be)

The estimation of individual fluctuating asymmetry (FA) is subject to large sampling variabilities. Heritability estimates, as well as correlations between developmental stability and any other individual character and/or between-trait correlations, are consequently biased downward if FA is used as an estimate of an individual's ability to buffer its development against developmental noise. The estimation of the hypothetical repeatability, defined as the ratio of the between-individual component of variation in the unsigned FA divided by the total variance, allows correction for these biases such that patterns observed for FA can be translated to make inferences about the presumed underlying developmental stability. In this paper I show that previous estimates of this repeatability are incorrect. I provide a new method and show by means of simulations that the hypothetical repeatability is in most cases even lower than previously thought. This has important consequences for the analysis of FA with respect to statistical power and the interpretation of patterns in FA.

Keywords: fluctuating asymmetry; developmental stability; repeatability; heritability

1. INTRODUCTION

Fluctuating asymmetry (FA, random deviations from perfect bilateral symmetry) has become a controversial measure of developmental stability (i.e. an individual's ability to buffer its development against developmental noise) (e.g. Palmer & Strobeck 1986, 1992; Palmer 1996; Pomiankowski 1997; Rowe *et al.* 1997). The suitability of FA as an estimate of homeostasis and the statistical analysis involved in the calculation of FA have both received much attention in the literature. Although several studies support a relationship between FA and fitness, environmental and/or genetic stress, others fail to support such correlations. This may, in part, be due to inadequate statistical procedures and/or experimental design and low sample sizes. Because the degree of FA is often small and sometimes of the same magnitude as measurement error (ME), within-subject repeats and mixed-model analysis are required to obtain FA estimates that are not confounded with ME (Palmer & Strobeck 1986). Furthermore, Whitlock (1996) recently pointed out that the estimation of individual single-trait FA is subject to large sampling variability, because it is an attempt to estimate a variance with two data points. As a result, individual single-trait FA only loosely reflects the presumed underlying developmental process. Therefore, heritability estimates, as well as correlations of developmental stability with other characters and/or between-trait correlations, are biased downwards if FA is applied as an estimate of developmental stability. To be able to correct for these biases, the hypothetical repeatability (R) can be used. This parameter represents the proportion of

variation in individual FA due to between-individual variation in developmental stability. Whitlock (1996) derived that $R = 1/[1 + (1 + 1/C_v^2)(\pi - 2)/2 + V_{me}/V_p]$ where CV , V_p and V_{me} represent the coefficient of variation, the variance and the measurement error of the unsigned FA [$\text{abs}(\text{left} - \text{right})$], respectively. If developmental stability, for example, would have heritability equal to 1, the heritability of FA can maximally be equal to R , or in other words, R represents the maximal between-individual component of variability in FA. Whitlock (1996) derived the above formula from a model where values for the left and right trait are drawn from the same normal distribution. However, if the signed FA follows a normal distribution, the sample is homogeneous, whereas leptokurtic distributions may indicate heterogeneity in developmental stability between individuals (e.g. Palmer & Strobeck 1992; Leung & Forbes 1997). In the case of homogeneity (normal distribution of the signed FA) there is no between-individual variability in developmental stability. All observed variation in FA is due to sampling variation, such that a repeatability of zero would be expected (figure 1; in addition, see Palmer (1996)). Yet, $R = 0.39$ (if $ME = 0$, e.g. Björklund & Merilä (1997)) is found. Thus, the above formula, as well as the slight modification presented in Björklund & Merilä (1997) (assuming C_v to be a constant approximately equal to 0.76), appears to overestimate R in this case, such that bias corrections in order to be able to translate patterns observed for FA to patterns in developmental stability are incorrect. This incorrectness finds its origin in the use of the variance of the unsigned FA (V_p in Whitlock (1996)) as the between-individual component of variation in FA,

whereas it is in fact the combined variability of between- and within-individual effects as well as ME. In this note I derive a new formula to estimate R and show by means of simulations that this hypothetical repeatability is in most cases even smaller than previously thought.

2. THE ESTIMATION OF THE HYPOTHETICAL REPEATABILITY

The repeatability of a trait can be defined as the variance due to between-individual variability divided by the total phenotypic variance (i.e. between+within, e.g. Falconer & Mackay (1996)). In this section I will show how the total variation in the unsigned FA can be partitioned into a within- and a between-individual component, leading to an estimate of R . I mainly follow the notations of Whitlock (1996), although I introduce several new parameters. Table 1 gives an overview of the used notations and their definitions. I also indicate differences with Whitlock (1996) as well as which of these parameters can be estimated from empirical data.

Assume a sample of N_t individuals with N unknown different degrees of developmental stability (the heterogeneity in developmental stability may have an environmental and/or genetic origin), which have frequencies w_k (sum of all $w_k = 1$, where $k=1, \dots, N$). The distribution of the signed FA (left-right) will be a leptokurtic distribution (Palmer & Strobeck 1992; see figure 1) (a mixture of N normal distributions) with variance V_{fa} . This variance is equal to the weighted average (with w_k as weight factor) of the N V_{fak} values, where the values of V_{fak} are the variances of the N normal distributions respectively (because all distributions have mean zero, i.e. no directional asymmetry). The values of V_{fak} (which cannot be estimated from classic FA data) represent the N different developmental stabilities. The distribution of the unsigned FA [abs(left-right)] will be a mixed distribution of N half-normal distributions. Each of these N distributions will have mean M_k (equal to twice the square root of $V_{fak}/2\pi$) and variance V_k (equal to $V_{fak}(\pi-2)/\pi$ (Whitlock 1996)) (obviously M_k and V_k cannot be estimated).

To be able to estimate R , the total variance of the unsigned FA (V_t) must be decomposed in two components representing the between-individual variability in developmental stability (V_{ind}), which corresponds to variation in the unobservable values of M_k , and the within-individual variability (V_{err}) due to sampling variation and measurement error for most data sets. V_{ind} is unknown in most practical cases, except for some controlled experimental conditions. Thus V_{ind} should be estimated indirectly as $V_t - V_{err}$. Following an approach comparable to the decomposition of sums of squares in an ANOVA model (e.g. Neter *et al.* 1990), it can be easily shown that for relatively large sample sizes, V_{err} equals the weighted average (with w_k as the weight factor) of the V_k values, each of which equal $V_{fak}(\pi-2)/\pi$. Thus, as the weighted average of the values of V_{fak} equals V_{fa} (see above), V_{err} can be shown to equal $V_{fa}(\pi-2)/\pi$. For small samples a bias correction of $N_t/(N_t-1)$ should be added. The between-individual component of variation V_{ind} can then be calculated as $V_t - V_{err}$, and thus does not require

Table 1. *Overview of the different notations*

symbol ^a	emp. ^b	explanation
R	Y	hypothetical repeatability
N_t	Y	total sample size
N	N	number of different degrees of developmental stability (DS)
w_k	N	frequency of the k^{th} degree of DS ($k=1, \dots, N$)
V_{fa}	Y	variance of the signed FA ($2V_a$ in Whitlock (1996))
V_{fak}	N	variance of the signed FA of the k^{th} degree of DS ($k=1, \dots, N$)
M_k	N	mean of the unsigned FA of the k^{th} degree of DS ($k=1, \dots, N$)
V_k	N	variance of the unsigned FA of the k^{th} degree of DS ($k=1, \dots, N$)
V_t	Y	total variance of the unsigned FA (V_p in Whitlock (1996))
V_{ind}	Y ^c	between-individual component of variation in the unsigned FA
V_{err}	Y	within-individual component of variation in the unsigned FA
V_{me}	Y	measurement error component of variation in the unsigned FA
V_{me-mix}	Y	estimate of measurement error as obtained from a mixed model
r	Y	number of within-subject repeats

^a Symbols are ordered according to their sequence in the text.
^b Indicates if (Y) or if not (N) this parameter can be estimated from empirical data.
^c V_{ind} can only be estimated indirectly as $V_t - V_{err}$ (see text for details).

a priori knowledge of the different degrees of the N developmental stabilities. The hypothetical repeatability can be easily estimated as $R = [(V_t - V_{fa}(\pi-2)/\pi)/V_t]$, where V_t represents the total variation in the unsigned FA, and V_{fa} is the variance of the signed FA (and ME, see below).

FA is usually not measured without ME. If the variance of the signed FA (V_{fa}) is estimated directly from the asymmetry values without making use of within-subject repeats and mixed-model analysis, the variability due to ME is already contained in V_{fa} as the latter estimate is biased (Palmer & Strobeck 1986). However, an unbiased estimate of V_{fa} can be obtained from a mixed model. In that case, it can be shown, in a similar way as for V_t , that the within-individual component can be further decomposed into two components: V_{err} (see above) and V_{me} , where V_{me} equals $(V_{me-mix}/r)(\pi-2)/\pi$. V_{me-mix} is the residual variance component obtained from the mixed model, whereas r is the number of within-subject repeats used to calculate the average individual unsigned FA values. Thus, the total variation in unsigned FA can be decomposed in a within- and a between-individual component. The within component ($V_{err} + V_{me}$) can be easily derived from the variance of the signed FA and that of its ME, whereas the between component V_{ind} equals $V_t - (V_{err} + V_{me})$. This procedure is, however, not required as R can be estimated immediately from the variance of the signed and unsigned FA without mixed-model analysis. In addition, note that the use of mixed-model analysis for leptokurtically distributed signed FA may have other implications (see Palmer & Strobeck 1992).

Table 2. *Hypothetical repeatability of fluctuating asymmetry*

1:2:3 ^a	V_{fa}^b	R_{sim}^c	V_{fa-obs}^d	V_{t-obs}^e	R_{SVD}^f	R_W^g	R_{BM}^h
1:0:0 ⁱ	0.25	-0.01	0.25	0.09	0	0.39	0.39
1:1:1 ⁱ	1.75	0.27	1.74	0.86	0.29	0.47	0.39
1:1:1 ^j	2.00	0.22	1.97	0.90	0.21	0.42	0.37
3:1:1	1.15	0.36	1.22	0.68	0.35	0.48	0.39
3:1:1 ^j	1.40	0.29	1.48	0.73	0.26	0.42	0.36
8:1:1	0.70	0.34	0.70	0.38	0.33	0.49	0.39
8:1:1 ^j	0.95	0.22	0.95	0.45	0.23	0.39	0.34
2:4:4	2.05	0.23	2.06	0.97	0.23	0.48	0.39
2:4:4 ^j	2.30	0.17	2.30	1.02	0.18	0.41	0.37
1:1:8	3.32	0.11	3.32	1.37	0.12	0.42	0.39
1:1:8 ^j	3.57	0.09	3.60	1.46	0.10	0.40	0.38
9:0:1 ^k	2.73	0.53	2.56	2.03	0.54	0.58	0.39
9:0:1 ^{j,k}	2.98	0.50	3.10	2.26	0.50	0.53	0.38

^aRatio of three different degrees of developmental stability (variances equal to: FA1 = 0.25; FA2 = 1; FA3 = 4; except for the last two simulations (see below)).

^bWeighed average of the degrees of developmental stability + measurement error (i.e. expected value of V_{fa}).

^cEstimate of the true repeatability (see text for details).

^dObserved variance of the signed FA.

^eObserved variance of the unsigned FA.

^fEstimate of R following this paper.

^gEstimate of R following Whitlock (1996).

^hEstimate of R following Björklund & Merilä (1997).

ⁱData from these simulations were plotted in figure 1.

^jMeasurement error of FA (normally distributed with zero mean and variance equal to 0.25) was added to the simulation model.

^kFA1 = 0.25; FA3 = 25.

3. A SIMULATION STUDY OF R

To illustrate the correctness of the above-presented estimation of R and to show the difference between this estimate and the previously derived ones, I performed a set of simulations. Samples of 10 000 individuals were generated in SAS (v. 6.12). Individuals were randomly assigned to two or three different degrees of developmental stability with different probabilities (i.e. w_k). For each combination of degrees of developmental stability, a data set was generated with and without some degree of measurement error (table 2). For each individual, FA was independently generated for two traits by random sampling from the underlying normal distributions with variances corresponding to the respective developmental stabilities. In empirical studies this is not possible as each trait develops only once under a set of environmental conditions. I then calculated the repeatability of the unsigned FA as the intraclass correlation of an ANOVA with the individual as the factor (two observations, i.e. the two traits, for each individual) (Sokal & Rohlf 1995). This repeatability can be assumed to be a very reliable estimate of the true repeatability because of the large sample sizes of the simulated data sets. The hypothetical repeatability, which can be estimated from empirical data, was estimated using the methods presented in this paper and by Whitlock (1996) and Björklund & Merilä (1997). In this way it was possible to compare the different estimation methods for the hypothetical repeatability with a reliable estimate of the true repeatability. Note that the simulation of this true repeatability assumes independent development of the two sides (see also Houle 1997).

As expected, the hypothetical repeatabilities as estimated by the method presented in this paper corre-

sponded closely to the values obtained from the ANOVA analysis, whereas the other procedures appeared to overestimate R in most cases. The method proposed by Björklund & Merilä (1997) appeared to underestimate R when developmentally stable individuals (variance = 0.25) dominate the population and a minority of extremely unstable individuals (variance = 25) is present (table 2), while no other parameter combinations appeared to result in such a pattern. However, this situation may be rather unlikely to occur under natural conditions.

Figure 1 clearly illustrates the absence of any repeatability for a homogeneous sample with only one degree of developmental stability and thus normally distributed signed FA. The leptokurtically distributed signed FA originated from a heterogeneous sample illustrating the low repeatability.

4. DISCUSSION AND CONCLUSIONS

Because individual FA is only loosely correlated with developmental stability, studies will in general underestimate the relationship between developmental stability and other characteristics of an individual, as well as between-trait correlations and the heritability of developmental stability (Whitlock 1996). The estimation of the hypothetical repeatability of FA allows us to correct for this downward bias and thus permits us to translate patterns in FA to patterns in the presumed underlying developmental stability. Furthermore, it allows us to evaluate how much variation in FA is due to between-individual heterogeneity (genetic and/or environmental in origin) in developmental stability. The results presented here show that R can be estimated from the total variation of the signed and the unsigned FA. R can be

Figure 1. Distribution of signed FA and the between-trait correlation of unsigned FA. Fluctuating asymmetry for two traits was generated for 10 000 individuals in SAS (v. 6.12). Degree of developmental stability equalled 0.25 for all individuals (*a*, *b*) and 0.25, 1 and 4 for 33.3% of the individuals, respectively (*c*, *d*) (see table 2 and text for more details). The kurtosis of the distribution of the signed FA, and the coefficient of determination for the between-trait correlation of the unsigned FA, are indicated.

estimated from the unbiased variance components of FA and ME, as obtained from mixed-model analysis of within-subject repeats as well. However, in practice this can be expected to be more prone to sampling variation as three variance components (instead of two) are involved in the calculations. The method of estimation presented here is consistent with the simulated results and shows that R is even smaller than previously thought, except for some unrealistic situations. As a result, correlations between FA and other individual characteristics, between-trait correlations of FA and heritabilities of FA are expected to be underestimates of patterns in developmental stability and need to be adjusted to make inferences about the underlying developmental stability. I refer the reader to Whitlock (1996) for details and computational formulas for the bias correction. The statistical power to determine patterns in developmental stability through the study of FA can be expected to be relatively low because of its weak correlation with FA (see Whitlock 1996; Houle 1997).

It is important to note that the presence of a small proportion of antisymmetrical individuals in a sample where most individuals exhibit 'ideal' FA (*sensu* Palmer & Strobeck 1992) may result in a leptokurtic distribution of the signed FA as well (Palmer & Strobeck 1992). As heterogeneity in developmental stability and the presence of antisymmetry both result in leptokurtosis, it is very difficult to differentiate between them (Palmer & Strobeck 1992), except in some specific situations (Rowe *et al.* 1997). Obviously, the above results apply only to FA, and not to other forms of asymmetry.

The author is a research assistant with the Fund for Scientific Research Flanders (FWO vlaanderen). I thank M. Whitlock and two anonymous referees for their comments on an earlier version of the manuscript.

REFERENCES

- Björklund, M. & Merilä, J. 1997 Why some measures of fluctuating asymmetry are so sensitive to measurement error. *Ann. Zool. Fenn.* **34**, 133–137.

- Falconer, D. S. & Mackay, T. F. C. 1996 *Introduction to quantitative genetics*. Harlow: Longman.
- Houle, D. 1997 Comment on 'A meta-analysis of the heritability of developmental stability' by Møller and Thornhill. *J. Evol. Biol.* **10**, 17–20.
- Leung, B. & Forbes, M. R. 1997 Modelling fluctuating asymmetry in relation to stress and fitness. *Oikos* **78**, 397–405.
- Neter, J., Wasserman, W. & Kutner, M. H. 1990 *Applied linear statistical models*. Boston, MA: Irwin.
- Palmer, A. R. 1996 Waltzing with asymmetry. *BioScience* **46**, 518–532.
- Palmer, A. R. & Strobeck, C. 1986 Fluctuating asymmetry: measurement, analysis, patterns. *A. Rev. Ecol. Syst.* **17**, 391–421.
- Palmer, A. R. & Strobeck, C. 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.
- Pomiankowski, A. 1997 Genetic variation in fluctuating asymmetry. *J. Evol. Biol.* **10**, 51–55.
- Rowe, L., Repasky, R. R. & Palmer, R. 1997 Size-dependent asymmetry: fluctuating asymmetry versus antisymmetry and its relevance to condition-dependent signalling. *Evolution* **51**, 1401–1408.
- Sokal, R. R. & Rohlf, F. J. 1995 *Biometry*. New York: Freeman & Co.
- Whitlock, M. 1996 The heritability of fluctuating asymmetry and the genetic control of developmental stability. *Proc. R. Soc. Lond. B* **263**, 849–854.