npg

# Genetics of fluctuating asymmetry in pupal traits of the Speckled Wood butterfly (*Pararge aegeria*)

JJ Windig[1,2] and S Nylin[1]

[1]*Department of Zoology, University of Stockholm, 106 91 Stockholm, Sweden;* [2]*ID-Lelystad, Institute for Animal Science and Health, PO Box 65, 8200 AB Lelystad, The Netherlands*

Fluctuating asymmetry (FA), small random differences between left and right, has been extensively used as a measure of individual quality, though its usefulness in that respect is controversial. Whether FA is heritable has implications for sexual selection theory and for its usefulness as an indicator of stress. Heritability ($h^2$) of FA is, however, difficult to estimate precisely and reliably. Here we report $h^2$s of FA for two pupal traits in the speckled wood butterfly (*Pararge aegeria*). We used a restriction error maximum likelihood (REML) analysis in combination with a jackknife procedure to analyse a large mixed offspring–parent/half-sib/full sib data set. A five-generation selection experiment provided a second set of narrow sense $h^2$s. Narrow sense $h^2$s were not significant and on average −0.029 (REML-analysis) and 0.031 (selection experiment) for the pupal segment covering the fore leg (LEG) and 0.057 and 0.004 for a SPOT on that segment. Estimated percentage dominance variances were 0.057 (LEG) and 0.027 (SPOT) and not significantly different from 0. The $h^2$ estimates had been slightly increased by cage effects. Average FA for LEG after five generations of selection were higher in the high lines than in the low lines, and the control lines were in between. No difference in FA between lines was found for SPOT. Although differences between lines were not significant, a slight $h^2$ (<3%) for LEG could not be excluded. The genetic effect was, however, small compared with the effect of foodplant quality. Larvae grown on foodplants that were not watered enough for good growth showed significantly higher FA for LEG, but not for SPOT, compared with larvae grown on good foodplants.
*Heredity* (2002) **89,** 225–234. doi:10.1038/sj.hdy.6800130

## Introduction

Fluctuating asymmetry (FA) has been the focus of a wide range of evolutionary and ecological studies. It can be defined as small random departures from anticipated bilateral symmetry (Van Valen, 1962; Markow, 1995). Whether or not FA has a genetic basis has evolutionary implications. If for example FA is to be used as a measure of the quality of the environment individuals live in, genetic variation in FA may have caused individuals to become adapted to polluted environments resulting in decreased FA. In addition if females use FA to evaluate genetic quality, ie the presence of 'good genes' in potential mates, an additive genetic component would be expected.

Whether or not FA is heritable has led to considerable controversy. Møller and Thornhill (1997b) used a meta-analysis to evaluate the existing literature and concluded the mean heritability ($h^2$) of FA was 0.27, highly significantly different from 0. Severe criticism of this study was published in the same issue (*Journal of Evolutionary Biology* 10). Most common was the criticism that studies had been included that did not properly estimate $h^2$ of FA, for example studies that did not test whether asym-

metry was fluctuating or directional (eg, Mason *et al*, 1967; Martin *et al*, 1982). Whitlock and Fowler (1997) concluded that the average $h^2$ of FA was around 0.03 when the analysis was restricted to those studies that did estimate it in the proper way. Møller and Thornhill (1997a) replied that even the studies that Fowler and Whitlock included did not fulfil all criteria, but maintained that FA is heritable.

How to analyse FA and its $h^2$ has been studied by Palmer and co-workers (Palmer and Strobeck, 1986, 1997; Palmer, 1994; Rowe *et al*, 1997). They recommend that the following four criteria must be fulfilled to estimate $h^2$ of FA confidently:

(1) An assessment must be made of measurement error. FA is, by definition, small and it thus can have a similar size as measurement error, and since measurement error fluctuates between left and right it is easily confounded with FA.

(2) The amount of asymmetry may be related to trait size (eg, Windig, 1998). If so, then this effect must be controlled, otherwise a relationship between FA and quality may be entirely due to trait size.

(3) The type of asymmetry must be determined. In the case of directional and anti-symmetry a number of indices used to quantify FA are not reliable.

(4) Narrow sense $h^2$s must be estimated. When a broad sense $h^2$ is estimated in, for example, a full sib estimate, the additive genetic variance (that part of vari-

*Correspondence: JJ Windig, ID-Lelystad, Institute for Animal Science and Health, PO Box 65, 8200 AB Lelystad. The Netherlands.*
*E-mail: J.J.Windig@id.wag-ur.nl*

ance that can be directly utilized by selection) is confounded by dominance variance. Therefore dominance variance has to be estimated separately. Moreover, dominance variance tends to be higher in traits that are strongly correlated to fitness (Roff, 1997), and therefore it is interesting to know it.

Beside these criteria, large sample sizes are needed to estimate quantitative genetic parameters in general. For quantitative genetics of FA, large sample sizes are even more important since the expected $h^2$s are low and thus difficult to estimate reliably (Houle, 1997).

FA can be seen as an estimate of developmental stability (Møller and Swaddle, 1997), but in fact it is an attempt to estimate the variance generated by developmental instability by using only two data points (Whitlock, 1996). Consequently the correlation between FA of two different traits at the individual level, is expected to be low even if genome-wide differences exist in developmental stability among individuals. The $h^2$ of fluctuating asymmetry is expected to be low for the same reason (Houle, 1997). By making use of the variation between left and right of the whole population one can estimate a hypothetical repeatability of developmental stability (Whitlock, 1996; Van Dongen, 1998), which can be used to transform a $h^2$ of FA into a $h^2$ of developmental instability.

We explored the quantitative genetics of FA for two pupal traits in the speckled wood butterfly in a rigorous way. Two different methods, a restricted error maximum likelihood (REML) analysis of a large mixed offspring–parent/half-sib/full sib data set and a five generation selection experiment are used to estimate $h^2$ of FA, fulfilling all the criteria for reliable estimation of $h^2$ of FA.

## Methods

### Study species and traits
We studied the southern European subspecies of the speckled wood butterfly *Pararge aegeria aegeria* caught on the Atlantic island of Madeira (Portugal, 33°N, 17°W). Two pupal traits were chosen to measure before the emergence of butterflies, so we could select parents in the selection experiment (see below) without having to handle the more vulnerable adult butterflies. Different pupal traits were measured (several wing veins, width of the wing, width of the antennal segment, width of the segment covering the tongue, width of the spot on the tongue segment, width of the eyes) on 58 individuals twice, until two traits were found where FA exceeded measurement error. These two traits were the length of the segment covering the fore leg (LEG) and the width of the brown spot (SPOT) on this segment (Figure 1). LEG was measured from the point where the suture near the eye is clearly bent, until the point of the segment closest to the tip of the pupa (where it is normally attached to a surface). SPOT was measured as the greatest width of the spot, measured parallel to the suture that touches the spot (Figure 1).

The reliability of the FA measurements was tested following the procedure outlined in Palmer (1994). Each of the authors measured traits on the left and right side of the same 58 individuals without knowledge of the results of the other person. Measurements were made with a microscope fitted with a micrometer at a magnification
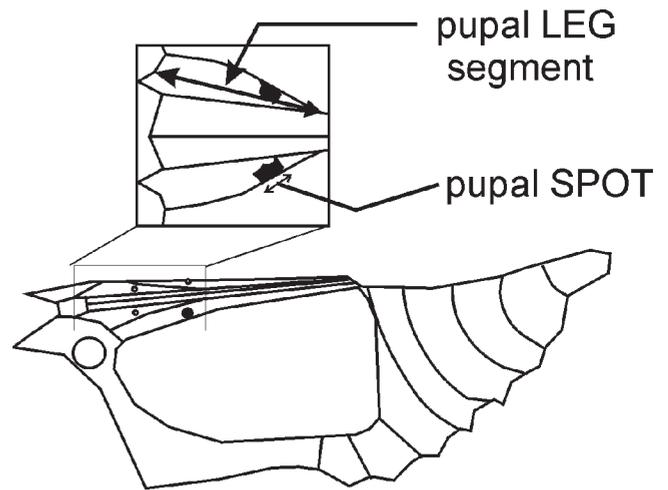


**Figure 1** Two pupal traits measured for analysis of fluctuating asymmetry : S = width of pupal SPOT, L = length of segment covering front LEG.

of 25 (LEG) or 50 (SPOT). Measurements were analysed with a repeated measures ANOVA, with individual, side and their interaction as factors. Besides estimating the magnitude of measurement error, this procedure also allows for testing for the presence of directional asymmetry. Dependence of FA on trait size was tested for by calculating correlation coefficients between unsigned asymmetry ($|L–R|$) and the mean trait size $[(L+R)/2]$. All measures other than those taken to asses reliability were taken by one of us (JJW).

### Genetics: mixed parent–offspring/sib analysis
Twenty-four wild caught females were allowed to oviposit on grass leaves in small containers at the end of October 1997. Females of the speckled wood butterfly have never been found to mate more than once (spermatophore counts) so their offspring can be considered full sibs (Wiklund and Forsberg, 1991). Upon emergence the offspring was transferred to grass tufts standing in water. Individuals were raised solitarily at a constant temperature of 23° and a day length of 14 h. Early frost considerably damaged the stock of grass, which had been kept outside, and consequently survival of the larvae was reduced. Roots were, however, well developed and after planting into soil the tufts developed into large plants. These plants were used to raise a second generation stemming from 24 pairs of unrelated, first-generation individuals. Twelve individuals were raised for each of these full sib families in pairs on the live plants.

The breeding resulted in a mixture of parents, partly being full sibs, and offspring consisting of full sibs and half cousins. A REML analysis was used to analyse this data set. Such a procedure takes into account all relationships in the data set and can be used to estimate additive and dominance variance components (Lynch and Walsh, 1998). We used the nf3 program developed by Shaw and Shaw (1992) adapted by one of us (JJW) for use on a PC and for the analysis of a three-generation data set to estimate genetic parameters. The adaptation gave identical results to the original program for pedigrees that could be analysed by both programs. Genetic parameters were

estimated for both trait size [(L+R)/2] and unsigned asymmetry (v|L–R|). Sexes were analysed separately since both traits varied over the sexes. The program was run in the unconstrained mode, allowing negative variance components to be estimated, because negative estimates do contain information. When the program did not converge, the program was run constraining the dominance variance, which always resulted in convergence. To reduce the effect of differences between the generations, generation was added as a fixed effect to the trait size.

Significance of variance components was tested using the likelihood ratio test (Lynch and Walsh, 1998). Standard errors estimated in a REML analysis are only asymptotically correct, therefore we estimated standard errors using a jackknife procedure. The jackknife evaluates variation of a parameter by calculating pseudovalues. For each pseudo-value one family is left out of the normal calculation. Jackknife estimates are calculated by subtracting n-1 times the pseudo-value from n times the value that was calculated for the complete data-set. The mean of the jackknife estimates tends to be similar to the parameter estimated from the complete data set only if the jackknife estimates are normally distributed. The standard error of the jackknife estimates divided by n gives the standard error of the parameter, but again a normal distribution of the jackknife estimates is assumed. The jackknife procedure is particularly sensitive to outliers among the jackknife estimates. When outliers occur it is best to remove families causing the outlying jackknife estimates altogether from the estimation. The same applies to families that cause a REML procedure not to converge.

We carried out the jackknife procedure by removing one full sib family, either in the first or in the second laboratory generation, in each iteration. Family members that were parents in other families were removed by replacing their observed value by a missing value, so that the program could still estimate the relationship of the individuals not removed from the data set. For the estimates for LEG (SPOT) 1 (2) jackknife estimates were considered as outliers (>3 standard deviations from the mean) in females, and removed from the calculation of the standard error. For males 1 (3) families resulted in outliers and a further 2 (2) families caused the REML procedure not to converge. The estimates for the $h^2$s of FA for males and females were combined to one estimate by taking their average. Standard errors were attached to these estimates by a jackknife procedure as described for the estimates of the sexes separately.

## Genetics: selection experiment

Selection lines were used to estimate realised $h^2$s of FA. LEG was selected in the upward and downward direction, SPOT only upward. For each trait and direction two lines were used. Two control lines were used to be able to correct for overall differences in the breeding environment. Shortage of space and manpower constrained the experiment to eight lines. We decided not to perform selection for low FA in SPOT because selection differentials would have been lowest in such lines (about 33% of individuals have an FA of 0 for spot). All lines were raised in the same climate room at approximately 26° and 14 h day length.

Each line was raised in its own cage on large grass plants raised in a greenhouse (winter) or collected at the university campus. Three lines (HIGH LEG1, LOW LEG1, CONTROL1) were started with offspring of the first laboratory generation (mixed parent–offspring–sib analysis), a further five lines (HIGH LEG2, LOW LEG2, CONTROL2, HIGH SPOT1 and HIGH SPOT2) with offspring of the second generation. Each generation after measurement of all individuals the five males and five females with the highest (lowest) FA were selected as parents for the next generation of the high (low) lines. If there were more than five individuals with the same amount of high or low FA, random numbers were used to determine the parents for the next generation. Random numbers also determined the 10 parents in the control lines. Parents were allowed to mate at random with each other. Invariably all females laid fertile eggs. One hundred of these eggs, all laid on the same day, were selected for the next generation, except for the control lines for which 25 eggs were used. About 80% of the eggs survived up to the adult stage. At the fifth generation 150 eggs were used for all lines, including control lines. In the sixth generation survival was very low (<10%) as was growth, probably as a result of a combination of lower food plant quality at the start of the winter and inbreeding effects, and the experiment was terminated. Combination of the remaining individuals of the different lines produced a generation with high survival and growth.

Differences in FA between the lines were tested in the fifth generation using both Levene's and Bartlett's test for homogeneity of variance (Palmer, 1994). We also used an ANOVA on unsigned asymmetry with line number (1 or 2) nested within type of line (high, low or control). Response to selection was calculated as the difference between the average unsigned FA of a line, to the average of the control lines. In this way the response was corrected for overall environmental influences, such as effects of different quality of foodplants over the generations, and for effects of inbreeding. Realised $h^2$'s were calculated by regressing the response on the cumulative selection differential (the difference in average phenotype of selected parents and the whole population (Falconer, 1989).

## Repeatability

It has been argued that low values of the $h^2$ of FA do not necessarily mean that the $h^2$ of developmental instability (DI) is low, because FA is an attempt to estimate variance caused by developmental instability with only two data points (Palmer, 1994; Whitlock, 1996; Van Dongen, 1998; Gangestad and Thornhill, 1999). These downward biases can be corrected for by the hypothetical repeatability $R$. The left and right trait value of an individual can be viewed as two data points from a normal distribution with mean $\mu$ and variance $\sigma^2_{DI}$. The signed asymmetry follows a normal distribution with mean zero and variance $2\sigma^2_{DI}$. If each individual has the same level of DI, all variation in the signed FA reflects sampling variation. This can be used to estimate the repeatability of DI. The repeatability was estimated in this study by using the following formula: $R = (\sigma^2_{|L-R|} - 0.363\sigma^2_{|L-R|}/\sigma^2_{|L-R|}$ (Van Dongen, 1998). The $h^2$ of FA was converted into $h^2$ of DI by division by $R$.

## Influence of host plant quality

Large differences in average FA were observed in all experiments more or less coinciding with apparent food

plant quality (judged by colour and water content of leaves). To determine whether it was really food plant quality causing these differences, and to determine the relative magnitude of foodplant and genetic causes of FA an additional experiment was performed. One hundred and twenty individuals were raised on grass tufts standing in water. For 80 of these the water was removed for 1 day twice a week, resulting in lower quality grass, at least to the human eye. Asymmetry of LEG and SPOT of pupated individuals was determined for the surviving individuals.

## Results

### Reliability of FA measurements

The difference between left and right was large relative to measurement error for both traits (side*individual term $P$ <0.001; Table 1). Directional asymmetry was absent (side term $P > 0.10$; Table 1). Distribution of signed asymmetry was somewhat leptokurtic for both traits (Kurtosis: LEG: 5.525; SPOT: 1.697; Figure 2). Such a distribution is expected if FA is a mixture of several normal distributions with different variances (Van Dongen, 1998). It indicates absence of strong anti-symmetry, although a mixture of FA and weak anti-symmetry cannot be excluded (Palmer and Strobeck, 1997).

Unsigned FA was significantly negatively correlated with average trait size for LEG ($r = -0.448$, $P < 0.0001$), but not significantly correlated in SPOT ($r = 0.0831$, $P = 0.090$). The strong negative correlation of FA with trait size for LEG reduced to $-0.213$ ($P = 0.003$) if the largest of the two sides was used instead of the average. This correlation was almost entirely due to a few extremely asymmetric individuals at the lowest trait sizes (Figure 2). Examination of these extremely asymmetric, small individuals suggested that something in the development of these pupae had gone wrong. Parts of the pupal cuticle were missing, some parts of the pupal case seemed not to have fused properly. We propose that bad condition was the cause of both the small size, and asymmetry of the LEG segment, and that the apparently antisymmetric small individuals formed part of a larger group showing fluctuating asymmetry. The members of this group that were symmetric had a larger size than members that were asymmetric. Our conclusion is that asymmetry of both SPOT and LEG can be considered to reflect developmental noise, and can be analysed accordingly.

### Phenotypic relations

Unsigned FA was rather high for SPOT (about 14% of trait size, Table 2), but smaller for LEG (around 4%).

Although LEG and SPOT were both significantly larger in females ($t$-test, LEG: $t = 2.812$, $P = 0.0052$; SPOT: $t = 15.166$, $P < 0.0001$), there was no significant difference in FA between the sexes ($F$-test on variance of signed FA, LEG: $F_{219,199} = 0.996$, $P = 0.976$; SPOT: $F_{218,197} = 1.287$, $P = 0.0715$). For LEG there was no significant difference between the first two laboratory generations in FA ($F_{120,298} = 1.25$, $P = 0.178$). For SPOT FA in the first generation was larger than in the second ($F_{119,296} = 2.12$, $P = < 0.001$). Survival of pupated individuals was high (94.5%). SPOT FA was slightly, but not significantly larger in pupae from which no adults emerged ($F_{28,404} = 1.34$, $P = 0.125$). LEG FA was larger, but not significantly, in pupae from which adults did emerge ($F_{28,404} = 0.89$, $P = 0.539$), so pupal survival was unrelated to FA. Both signed and unsigned FA were significantly correlated between LEG and SPOT (unsigned $r = 0.258$; signed $r = 0.2068$, both $P < 0.0001$).

### Genetics: mixed parent–offspring/sib analysis

Despite significant $h^2$s for trait size, none of the estimated heritabilities for FA in the mixed parent–offspring–sib analysis were significant. Two of the four $h^2$s estimated for the sexes separately were negative (Table 3). Averaged over sexes, the $h^2$ for LEG was $-2.9\%$, and for SPOT 5.7%. These values were similar to many published values for $h^2$ of FA. The range of estimates for the non-additive genetic variance components for FA was similar to the additive components. Their standard errors were considerably higher and none was thus significant. For trait size there was no dominance variance for SPOT, but considerable amounts for LEG.

Overall variation in signed asymmetry was used to estimate the repeatability of FA. This repeatability was used to convert $h^2$ of FA to $h^2$ of DI. The repeatability of FA is around 27% for SPOT and 37% for LEG, rather low compared to some morphological traits (Gangestad and Thornhill, 1999), but not unlike other estimates for FA (Van Dongen, 2000). Negative estimates of $h^2$ of FA (or $V_D/V_P$) cannot be converted to $h^2$ ($V_D/V_P$) of DI. This meant that only for the combined estimate for SPOT both $h^2$ and $V_D/V_P$ could be computed, and for the other estimates either an estimate for $h^2$ or an estimate for $V_D/V_P$ of DI had be computed. Broad sense $H^2$s ($h^2 + V_D/V_P$) were about 25–30% for SPOT and about 10–20% for LEG (Table 3).

### Selection experiment

The average FA varied widely over generations and was more or less similar in all lines (Figure 3). For LEG the order of FA of the different lines at the end of the experi-

**Table 1** Evaluation of the reliability of FA measurements by repeated measures ANOVA

| | Df | LEG | | | SPOT | | |
| | | MS | F-value | P | MS | F-value | P |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Individual | 1 | 64.836 | 5.01 | <0.001 | 12.555 | 7.21 | <0.0001 |
| Side | 57 | 28.280 | 2.18 | 0.145 | 0.211 | 0.12 | 0.737 |
| Ind*Side | 57 | 12.949 | 25.78 | <0.0001 | 1.742 | 4.23 | <0.0001 |
| Error | 116 | 0.5022 | – | – | 0.412 | – | – |

58 Individuals were measured twice. Significance of the side factor (either left or right) indicates presence of directional asymmetry, significance of the interaction between individual and side indicates that fluctuating asymmetry is larger than measurement error.

FLUCTUATING ASYMMETRY VS. TRAIT SIZE

LEG                    SPOT
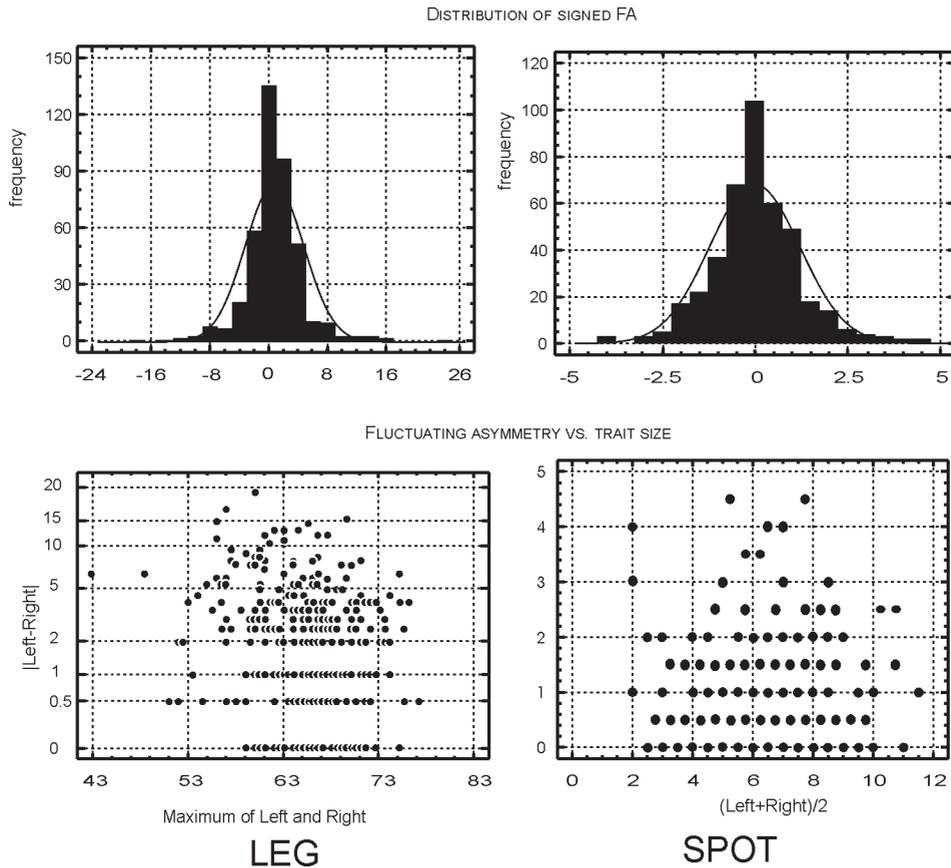
**Figure 2** Graphical representation of FA for two pupal traits: left graphs: LEG, right: SPOT. Top row frequency distribution of signed asymmetry follow a lepto-kurtic distribution. Drawn curves are normal frequency distributions based on mean and variance of the actual samples. Bottom Row FA (unsigned asymmetry) *vs* trait size is negatively related for LEG, but unrelated for SPOT.

**Table 2** Several indices of FA for both traits in the first two generations raised in the laboratory (420 individuals)

|  | LEG | | | SPOT | | |
|---|---|---|---|---|---|---|
|  | *All* | *Male* | *Female* | *All* | *Male* | *Female* |
| **FA1** (mean $|L–R|$) | 2.60 | 2.65 | 2.55 | 0.863 | 0.900 | 0.823 |
| **FA2** (mean $|L–R|/TS$) | 4.05% | 4.16% | 3.93% | 13.9% | 12.4% | 16.2% |
| **FA4** (Var.$(L—R)$) | 16.09 | 16.09 | 16.15 | 1.491 | 1.670 | 1.300 |
| **FA6** (Var.$(L–R)/TS$) | 0.250 | 0.253 | 0.248 | 0.240 | 0.231 | 0.256 |
| **FA9** ($1-r_{L–R}$) | 0.247 | 0.261 | 0.237 | 0.228 | 0.239 | 0.259 |
| **FA10** (FA4/2 - ME) | 6.223 | 6.191 | 6.291 | 0.665 | 0.777 | 0.622 |
| Mean trait size | 64.28 | 63.59 | 65.04 | 6.206 | 7.233 | 5.071 |
| Variance trait size | 28.38 | 26.77 | 29.19 | 3.277 | 2.072 | 2.159 |
| No. | 420 | 220 | 200 | 417 | 219 | 198 |

Numbers in bold refer to Palmer and Strobeck (1986). L = left side; R = right side; TS = trait size; ME = measurement error. **FA10** was estimated from repeated measurements ANOVA on 58 individuals (Table 1).

ment was as expected (Low < Control < High). For SPOT the control line ended between the two high lines (no low lines have been maintained for SPOT). FA between the lines, however, does not differ in the fifth generation for either LEG (Levene's test: $F = 1.420$, $P = 0.243$, Bartlett's test: $B = 1.007$, $P = 0.246$) or SPOT (Levene's test: $F = 0.787$, $P = 0.386$, Bartlett's test: $B = 1.004$, $P = 0.403$). There was no consistent response whatsoever for SPOT when it was corrected using the average of the control lines and plotted against the cumulative

selection differential (Figure 4). For LEG the response in the LOW lines was more or less downwards, for the HIGH lines more or less upwards.

Realised heritabilities estimated by the slope of the regression of response on cumulative selection differential assume a linear response (Falconer, 1989). Since responses were not linear in any of the lines, the estimated realised $h^2$s can only be seen as approximations. The estimated realised $h^2$ for the high lines of LEG were slightly lower (slope ± standard error for High 1: 0.020 ±

**Table 3** Genetic analysis of fluctuating asymmetry for two generations of SPOT and LEG. Estimates are based on unconstrained REML-analyses combined with a jackknife procedure. Families that caused outlying jackknife estimates (>3 standard deviations from the mean) were removed from the analysis. Standard errors were estimated in the Jackknife procedure. *P*-values are based on Log Likelihood ratios of the REML procedure, and **in bold** if significantly different from 0. For the calculation of $h^2$ of developmental instability a repeatability of 0.2738 (SPOT) and 0.372 (LEG) was used, computed from Table 2

| SPOT | Male | Female | Combined | LEG | | |
|---|---|---|---|---|---|---|
| | | | | Male | Female | Combined |
| *Trait size* | | | | | | |
| $h^2$ | 0.404 | 0.487 | | 0.208 | 0.139 | |
| *P*-value | **0.0003** | **0.0025** | | 0.290 | 0.448 | |
| $V_D/V_P$ | −0.086 | −0.176 | | 1.02 | 0.190 | |
| *P*-value | 0.800 | 0.731 | | **0.0028** | 0.645 | |
| $V_P$ | 7.81 | 8.44 | | 72.95 | 63.84 | |
| *Fluctuating asymmetry* | | | | | | |
| $h^2$ | −0.00015 | 0.114 | 0.057 | 0.0038 | −0.061 | −0.029 |
| St. error | 0.126 | 0.173 | 0.068 | 0.061 | 0.119 | 0.053 |
| *P*-value | 0.995 | 0.150 | | 0.115 | 0.557 | |
| $V_D/V_P$ | 0.078 | −0.024 | 0.027 | −0.032 | 0.082 | 0.057 |
| St. error | 0.266 | 0.320 | 0.153 | 0.184 | 0.288 | 0.125 |
| *P*-value | 0.876 | 0.416 | | – | 0.913 | |
| $V_P$ | 0.801 | 0.648 | | 9.34 | 10.21 | |
| *Developmental instability* | | | | | | |
| $h^2$ | – | 0.306 | 0.153 | 0.0102 | – | – |
| $V_D/V_P$ | 0.285 | – | 0.099 | – | 0.220 | 0.153 |

$V_D$ = dominance variance; $V_P$ = total phenotypic variance; $h^2$ = narrow sense heritability.
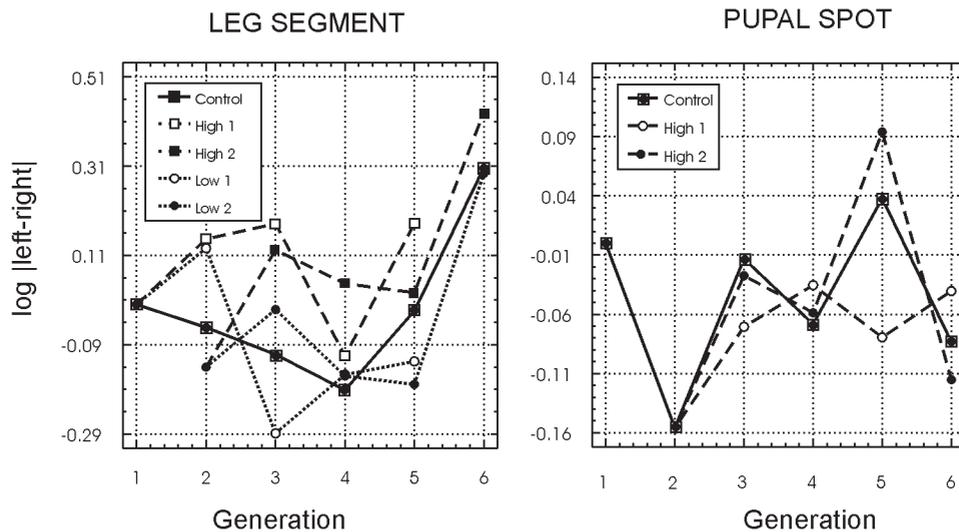


**Figure 3** Results of selection experiment uncorrected for generation differences. Average phenotype in generation one is set to 0. Points represent average FA (= log |left-right|) for each generation.

0.027; High 2: 0.004 ± 0.030) than for the low lines (Low 1: 0.060 ± 0.066; Low 2: 0.038 ± 0.044). For the SPOT lines the estimated realised heritabilities were very close to 0 (High 1: 0.0033 ± 0.035; High 2: 0.0037 ± 0.016). Averaged over four lines the realised $h^2$ for LEG is 0.031 ± 0.012 (standard error calculated from the four slopes of the lines). When tested with a *t*-test, this average was not significantly different from 0 ($t = 2.5432$, $P = 0.0844$). Using the repeatabilities of FA as estimated in the parent–offspring–sib experiment and combining the $h^2$s of the different lines the estimated $h^2$ of developmental stability for LEG was 0.0825 and for SPOT 0.0128.

### Host plant quality
Survival was low both for larvae grown on low water and on control plants (control: 37.5%; drought 28.45%), whereas survival in experiments following the same breeding procedure is normally between 80 and 100%. Pupal weight was not significantly lower in the drought group than in the control group (Table 4), nor was trait size for SPOT. For LEG, however, the drought group was significantly smaller than the control. FA was lower in the drought group for SPOT but not significantly so (Table 4), but was significantly higher in the drought group for LEG.

## LEG SEGMENT

## PUPAL SPOT
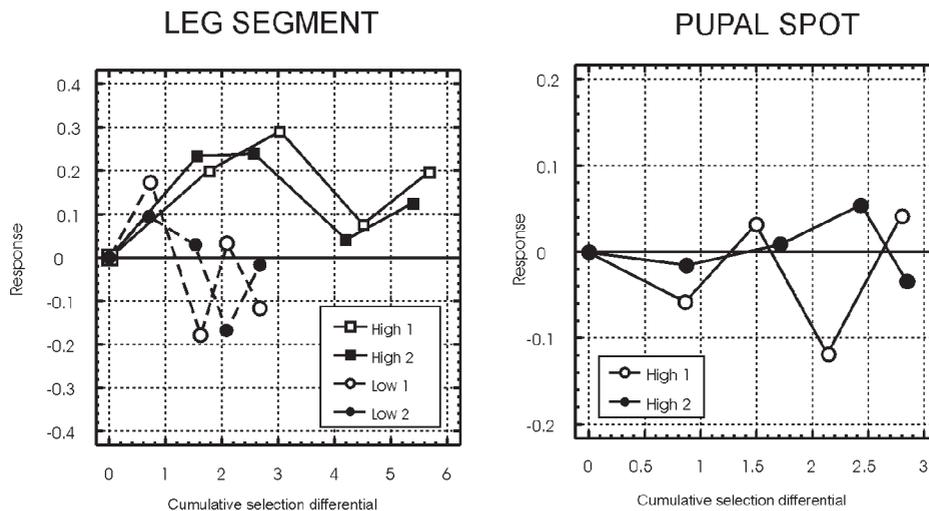


**Figure 4** Results of selection experiment relative to control lines *vs* cumulative selection differential. Points represent the difference of average FA for each selection line relative to the average of both control lines.

**Table 4** Influence of host plant quality on trait size (TS) and asymmetry. Host plants in control group received normal water. Water was removed for 1 day/week in drought group

|  | Control | Drought | F | P |
|---|---|---|---|---|
| Pupal weight | 0.129 | 0.134 | 1.14 | 0.29 |
| SPOT TS | 6.172 | 5.082 | 0.83 | 0.47 |
| LEG TS | 67.34 | 64.53 | 5.10 | 0.031 |
| SPOT FA1 | 1.08 | 0.86 | 0.77 | 0.23 |
| LEG FA1 | 1.51 | 2.57 | 3.20 | 0.018 |

FA1 = mean unsigned asymmetry. $F$ = $F$-value for treatment in two-way ANOVA with treatment and sex as factors (for trait size and weight) or $F$-value of an $F$-test for signed asymmetry (for FA).

## Discussion

Heritability of FA is theoretically expected to be low (Houle, 1997), but opinions differ on whether it is low in reality. Møller and Thornhill (1997b) concluded from studies available to them in 1996 that the average $h^2$ of FA was 0.27. Whitlock and Fowler (1997) concluded that FA was heritable but only about 0.03, while Markow and Clarke (1997) concluded that there was no evidence for heritable variation of FA. Since then a number of studies have appeared that have carefully analysed $h^2$ of FA. Windig (1998) found for four traits in a butterfly (*Inachis io*) an average $h^2$ of FA of 0.025, Woods *et al* (1998) an average of −0.07 for five traits in *Drosophila melanogaster*, Blanckenhorn *et al* (1998) an average of 0.04 for 13 traits in *Sepsis cynipsea*, Leamy (1999) an average of 0.03 for 10 traits in mice, Van Dongen *et al* (1999a) an average of 0.03 for three traits in *Operophtera brumata* and Tomkins and Simmons (1999) an average of −0.13 for forceps in earwigs. The results of this study conform to this general picture of low or negative $h^2$s of FA. When averaged over sexes and the selection and the mixed parent–offspring–sib experiments the $h^2$ for LEG is 0.001 and for SPOT 0.031.

All the criteria for reliable research on FA as formulated by Palmer and Strobeck (1997) have been unambiguously met in this study except one. When plotted against mean trait size the type of asymmetry was antisymmetric for small trait size values of LEG (Figure 2). We showed that this was, at least partly, an artefact; when plotted against the maximum trait size antisymmetry is less pronounced. If bad growth leading to a small size occurs on one side only, asymmetry will be high. At the same time average trait size has to be smaller, which may explain why small individuals tend to be asymmetric. Rowe *et al* (1997) discusses other possibilities why small individuals may show antisymmetry. Moreover, several authors have argued that both directional and antisymmetry may reflect developmental stability as well as FA (see discussions in Graham *et al*, 1998; Van Dongen *et al*, 1999b; Kark, 2001).

Asymmetry seemed rather high in this study, eg 14% of trait size for SPOT, where FA is generally less than 5% (Pomiankowski, 1997) or 1% (Palmer and Strobeck, 1997). Trait size may, however, not be the best way to standardise FA for comparisons across traits. One minus the correlation coefficient between left and right (FA9 of Palmer, 1994) may be a better way to compare FA across traits (Windig and Nylin, 2000). This index standardises FA not by trait size but by trait size variance. Values found for FA9 in this study (0.228–0.293) seem rather high, but not exceptionally high. A very interesting question for future research is what kind of traits have relatively high or low FA, for instance if pupal traits have higher FA than adult traits.

In this study there is no hint of genetic variation for SPOT, but for LEG a small amount of genetic variation may be present. If one ignores the control lines and combines replicates within high and low lines, FA is significantly higher in high lines than in low lines ($F$-test on variance of unsigned asymmetry: $F_{185,196}$ = 1.427, $P$ = 0.0072). One thus cannot distinguish between the possibilities that differences between lines are caused by chance, or that there is a difference but that this difference is too small to be significantly different from 0. It is, however, clear that if there is any genetically caused variation in FA it is small compared to environmentally induced variation in FA. The difference in unsigned FA of LEG between the high and low lines after five gener-

**Table 5** Comparison of $h^2$ of FA estimated in different ways (average of estimates for males and females). Methods differ with respect to components included in $V_A$

| | Components included in $V_A$ | | | | LEG | SPOT |
|---|---|---|---|---|---|---|
| | $V_A$ | $V_D$ | $V_{GE}$ | $V_{CE}$ | | |
| Selection | + | − | − | − | 0.031 | 0.004 |
| REML 2 generations | + | − | + | + | 0.018 | 0.073 |
| REML 2 generations $V_D$ constrained | + | + | + | + | 0.022 | 0.071 |
| REML 2 generations, jackknifed | + | − | + | + | −0.029 | 0.057 |
| REML 2 generations, cage effects | + | + | + | − | −0.022 | 0.032 |
| REML 1 generation | + | − | − | + | 0.043 | 0 (n.c.) |
| REML 1 generation $V_D$ constrained | + | + | − | + | −0.052 | 0 (n.c.) |
| Full Sib | + | + | − | + | 0.037 | 0.006 |
| REML 1 generation, cage effects | + | + | − | − | −0.041 | 0 (n.c.) |
| Full Sib, cage effects | + | + | − | − | −0.052 | −0.017 |

$V_A$ = Additive variance component; $V_D$ = Dominance variance component; $V_{GE}$ = Genotype by environment interaction component (due to environmental variation between generations); $V_{CE}$ = Common environmental variance component (due to cage effects). Selection = five generation selection experiment. REML 2 generations are mixed parent–offspring/full sib half sib data set estimated by REML analysis, 1 generation is the same but values of the parental generation treated as missing values. Full sib is conventional full sib analysis on 16 families in offspring generation that were not related. n.c. = not converged.

ations of intense selection was only 0.68 micrometer units while it was almost double (1.20) in the small scale host plant quality experiment. The large fluctuations over the generations in the selection experiments may well have been caused by differences in foodplant quality. All lines received foodplants from the same natural patch, but over the season differences in rainfall and temperature will have caused differences in quality of foodplants. These fluctuations may have slowed down progress in the selection experiment considerably. Another source of FA may have been inbreeding since artificial selection will cause inbreeding, especially if it is carried out at the rate in this experiment. Indeed after five generations all lines showed signs of inbreeding such as slow growth and low survival. Inbreeding may also increase FA. Since the control lines were subject to the same amount of inbreeding this effect is controlled for in the selection experiment. FA did not show a consistent increase in the control lines. Only LEG FA increased towards the end in all lines, which may have been due to inbreeding. So there is no clear indication of inbreeding effects on FA, and if there is any it is small compared to the influence of foodplant stress on FA.

Narrow sense $h^2$s are not easily estimated. Several authors (Swaddle *et al*, 1994; Palmer and Strobeck, 1997; Whitlock and Fowler, 1997) have pointed out that the additive variance is often confounded with other effects, such as dominance effects in a full sib analysis. Both the mixed parent–offspring–sib analysis and the selection analysis estimate additive genetic variance without confounding it with dominance variance. When a full-sib

analysis, that confounds $V_A$ and $V_D$, is employed on those families in the second generation that are not related through their grand-parents the estimates for $V_A$ are still very similar to those from the selection experiment (Table 5).

In the mixed parent–offspring–sib analysis there are two sources of variation that may influence estimated $V_A$. First both generations are not raised in exactly the same way which may influence FA, as is evident from variation between generations for SPOT. This effect was partly accounted for by adding the generation difference as a fixed effect in the model estimated in the REML analysis. If there are, however, genotype by generation effects (eg, the difference between the generations depends on the genotype), estimations are still influenced by the generation effect. The estimation of dominance variance will most seriously be influenced, since it depends to a large extent on differences between parents–offspring combinations (estimating $0.5\ V_A$) and full-sib combinations (estimating $0.5\ V_A + 0.25\ V_D$). The second source of variation is cage effects in the second generation. Here larvae were raised in pairs on food plants, these pairs sharing a common environment. Since six different cages were used for each family these cage effects will have only a limited influence on the final result. Fortunately, one can estimate the magnitude of these influences by comparing different analyses.

The estimated $V_A$ in the mixed parent–offspring–sib analysis contains variance caused by cage effects (=common environmental effects, further $V_{CE}$) and variance caused by the difference between generations

(genotype by generation (environment) effects, further called $V_{GE}$), while the selection experiment does not. When the estimates of the selection experiment are compared with the mixed parent–offspring–sib analysis, the result is ambiguous. For LEG the estimate of $h^2$ is larger in the selection experiment, suggesting no influence of $V_{CE}$ or $V_{GE}$ (Table 5). For SPOT the $h^2$ is about 7% lower in the selection experiment suggesting some influence. For the mixed parent–offspring–sib analysis the REML program used cannot estimate $V_A$, $V_D$ and $V_{CE}$ at the same time, but $V_{CE}$ can be estimated when $V_D$ is left out of the model, and thus confounded with $V_A$. Confounding $V_A$ and $V_D$ has little effect, estimates of $h^2$ for FA remain nearly identical (Table 5). Estimation of $V_{CE}$, however, reduces $h^2$ for both traits, although $V_{CE}$ is not significant for either trait. By restricting the analysis to one generation one can eliminate the influence of $V_{GE}$. For SPOT this has, however, the effect that none of the analyses converge, there is apparently too little variation left in order to estimate different variance components. For LEG $h^2$ increases slightly for the full model, and decreases when $V_A$ is confounded with $V_D$, both with and without estimation of $V_{CE}$. All estimates for $h^2$ of FA are not significantly different from 0, regardless of the method.

The general conclusion for this study is that the $h^2$ of FA is not significantly different from 0, and at most very small. The method employed did not influence this conclusion. For individual studies the gain of using more elaborate designs to estimate other components than $V_A$ for FA is limited. When the choice is to estimate either $V_D$ or $V_{CE}$ the latter seems more important. When different studies are combined to get an overall idea of $h^2$ of FA it remains important to estimate $h^2$ in the strict sense. The most reliable method is probably a long term selection experiment (eg, Mather, 1953; Reeve, 1960; see Brakefield and Breuker, 1996 for a short term selection experiment). In selection experiments $V_A$ is not confounded with other variances and even small $h^2$s can be detected. Replicate lines and control lines are an essential element of such experiments. They are however labour intensive and must be continued for a long time, preferably longer than the five generations possible in this study. Selection experiments can thus not be employed for all organisms.

## Acknowledgements

## References

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.

Brakefield PM, Breuker CJ (1996). The genetical basis of fluctuating asymmetry for developmentally integrated traits in a butterfly eyespot pattern. *Proc R Soc Lond B* **263**: 1557–1563.

Falconer DS (1989). *Introduction to Quantitative Genetics*. 3rd edn. Longman Scientific & Technical: 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.

Graham JH, Emlen JM, Freeman DC, Leamy LJ, Kieser JA (1998). Directional asymmetry and the measurement of developmental instability. *Biol J Linn Soc* **64**: 1–16.

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

Kark S (2001). Shifts in bilateral asymmetry within a distribution range: the case of the chukar partridge. *Evolution* **55**: 2088–2096.

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

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

Markow TA (1995). Evolutionary ecology and developmental stability. *Ann Rev Ent* **40**: 105–120.

Markow TA, Clarke GM (1997). Meta-analysis of the heritability of developmental stability: a giant step backward. *J Evol Biol* **10**: 31–37.

Martin NG, Jinks JL, Berry HS, Loesch DZ (1982). A genetical analysis of diversity and asymmetry in finger ridge counts. *Heredity* **48**: 393–405.

Mason LG, Ehrlich PR, Emmel TC (1967). The population biology of the butterfly, *Euphydryas editha*. V. Character clusters and asymmetry. *Evolution* **21**: 85–91.

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

Møller AP, Swaddle J (1997). *Developmental Stability and Evolution*. Oxford University Press: Oxford.

Møller AP, Thornhill R (1997a). Developmental stability is heritable. *J Evol Biol* **10**: 69–76.

Møller AP, Thornhill R (1997b). 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 Academic Publishers: Dordrecht, 335–364.

Palmer AR, Strobeck C (1986). Fluctuating asymmetry: measurements, analysis and pattern. *Ann Rev Ecol Syst* **17**: 391–421.

Palmer AR, Strobeck C (1997). Fluctuating asymmetry and developmental stability: heritability of observable variation vs. heritability of inferred cause. *J Evol Biol* **10**: 39–49.

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

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

Roff, DA (1997). *Evolutionary Quantitative Genetics*. Chapman and Hall: New York.

Rowe L, Repasky RR, Palmer AR (1997). Size-dependent asymmetry: fluctuating asymmetry versus antisymmetry and its relevance to condition-dependent signaling. *Evolution* **51**: 1401–1408.

Shaw RG, Shaw FH (1992). *Quercus: programs for quantitative genetic analysis using maximum likelihood*; published electronically on the Internet, available through anonymous ftp from ftp.bio.indiana.edu.

Swaddle JP, Witter MS, Cuthill IC (1994). The analysis of fluctuating asymmetry. *Anim Behav* **48**: 686–689.

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 individual fluctuating asymmetry? *Proc R Soc Lond B* **263**:1423–1427.

Van Dongen S (2000). The heritability of fluctuating asymmetry: a Bayesian hierarchical model. *Annales Zoologici Fennici* **37**: 15–23.

Van Dongen S, Lens L, Molenberghs G (1999a). Mixture analysis of asymmetry: modelling directional asymmetry, antisym-

234

metry and heterogeneity in fluctuating asymmetry. *Ecol Let* **2**: 387–396.

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

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

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

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

Wiklund C, Forsberg J (1991). Sexual size dimorphism in relation to female polygamy and protandry in butterflies: a comparative study of Swedish Pieridae and Satyridae. *Oikos* **60**: 373–381.

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

Windig JJ, Nylin S (2000). How to compare fluctuating asymmetry across traits? *J Evol Biol* **13**: 29–37.

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