

# Lack of response to selection for lower fluctuating asymmetry of mutant eyespots in the butterfly *Bicyclus anynana*

CJ Breuker and PM Brakefield

*Evolutionary Biology, Institute of Evolutionary and Ecological Sciences, Leiden University, PO Box 9516, 2300 RA Leiden, The Netherlands*

Fluctuating asymmetry (FA) is claimed both to provide a means of evaluating developmental stability, and to reflect an individual's quality or the stress experienced during development. FA refers to the nondirectional variation between left and right sides, whereas directional asymmetry (DA) refers to a significant directional variation between the sides. We studied four eyespots on the dorsal forewing of the tropical butterfly, *Bicyclus anynana*. Two of the eyespots were specified by a mutant allele, *Spotty*, that was fixed in the stock. These eyespots showed higher FA than the two flanking, wild-type eyespots, although they are all formed by the same developmental pathway. We applied artificial selection for lower FA of the novel eyespots in an attempt

to increase their developmental stability. There was significant variation present in individual FA in our study. However, this did not change as a result of the artificial selection. Most of the variation in FA can be accounted for by individual differences in developmental stability rather than by the applied selection or by environmental variation. Thus, it was not possible to produce any increased developmental stability of the novel eyespots by selecting for low FA. The estimates of realized heritability for both FA and DA of each eyespot were not significantly different from zero. The results suggest that FA provides little, if any, potential for exploring the mechanistic basis of developmental stability.

*Heredity* (2003) 91, 17–27. doi:10.1038/sj.hdy.6800279

**Keywords:** developmental stability; fluctuating asymmetry; directional asymmetry; mutant traits; artificial selection; realized heritability

## Introduction

Fluctuating asymmetry (FA) is thought to provide a means of evaluating developmental stability (Van Valen, 1962; Palmer and Strobeck, 1986, 1992; Palmer, 1994). FA is defined as small, directionally random deviations from perfect bilateral symmetry (ie distribution of left–right differences is normal with a mean of zero). Developmental stability (DS) refers to the ability of an individual to buffer its development against random perturbations. Developmental instability (DI) refers to the inability to do so and hence to developmental mistakes. Low FA has been suggested to be associated with high fitness and/or low stress during development (eg, reviews in Møller and Swaddle, 1997). As the fitness of an organism is likely to depend on the ability to produce the ideal phenotype in a given environment, it can be assumed that there is a genetic basis for DS (Møller and Thornhill, 1997).

The heritability of FA might be used to estimate the genetic basis of DS (Whitlock, 1996; Fuller and Houle, 2003). Based on a metaanalysis of published estimates of heritabilities of FA, there seems to be a moderate heritability of FA with a small (0.19) additive genetic

component (Møller and Thornhill, 1997). Many other studies argue, however, that the heritability is much smaller or even nonsignificantly different from zero (Fowler and Whitlock, 1994; Brakefield and Breuker, 1996; Leamy, 1997, 1999; Windig, 1998).

No quantitative trait loci (QTLs) have been found to affect FA of mandible characters in house mice (Leamy *et al*, 1997, 2000). Three QTLs, however, significantly affected directional asymmetry (where one side consistently differs from the other side) in several mandible characters. It was concluded that there was a very low genetical variability for directional asymmetry (DA) in these characters, but none for FA. For six discrete skeletal traits in house mice, however, the opposite seems to be true: significant numbers of QTLs for FA, but not DA (Leamy *et al*, 1998). In general, it has been argued that DA is not useful as an indicator of DS, because of a presumed heritable basis (Palmer, 1994). Some studies, therefore, exclude those traits exhibiting DA (eg, Hutcherson and Oliver Jr (1998) discarded 52 out of the 57 traits). This procedure is, however, highly debatable, since the few studies which estimated the heritability of DA report values comparable to those of FA, and moreover selection experiments on DA have always been unsuccessful (Palmer, 1996; Auffray *et al*, 1999). Some other (theoretical) studies have shown that FA, DA and antisymmetry (AS) (distribution of the signed differences between the sides is bimodal) may be dynamically related (Graham *et al*, 1993). FA may turn

into DA during artificial selection experiments (Graham *et al.*, 1993). Insect wings usually exhibit DA of size and shape, and have been shown to provide information on DS as a significant amount of phenotypic variation in DA is of environmental, rather than genetic origin (Brakefield and Breuker, 1996, 2003; Klingenberg and McIntyre, 1998; Windig and Nylin, 1999; Klingenberg *et al.*, 2001). The studies of the insecticide-resistant strains of the Australian sheep blowfly *Lucilia cuprina* have shown that antisymmetry may also arise from increased FA (McKenzie and Clarke, 1988). In summary, under certain circumstances, DA (and AS) may be useful as an indicator of DS and information on traits exhibiting DA should not be discarded too quickly.

The increased FA in resistant strains of the sheep blowfly was associated with phenotypes specified by mutant alleles, and subsequent (natural) selection decreased the FA (McKenzie and Clarke, 1988; McKenzie and O'Farrell, 1993; McKenzie and Yen, 1995). Studying the heritability of the FA of a trait specified by novel mutant alleles may yield higher and significant estimates of heritability than are usually found, as selection for DS of these traits has not yet taken place. In other words, the DS of these traits does not have a history of selection and comparatively high genetic variance may still be present.

Brakefield and Breuker (1996) showed from parent-offspring analyses that there was no significant heritable variation for FA of a series of eyespots on the dorsal and ventral sides of the wings of wild-type individuals of a tropical butterfly, *Bicyclus anynana*. However, the FA of the two novel forewing eyespots specified by the mutant allele *Spotty* was higher than for the two wild-type flanking eyespots. In the present study, we examine whether it is possible to reduce the FA of these novel eyespots by means of strong artificial selection over several generations. We also examine how the selection affected the variation in individual DS and DA (Graham *et al.*, 1993).

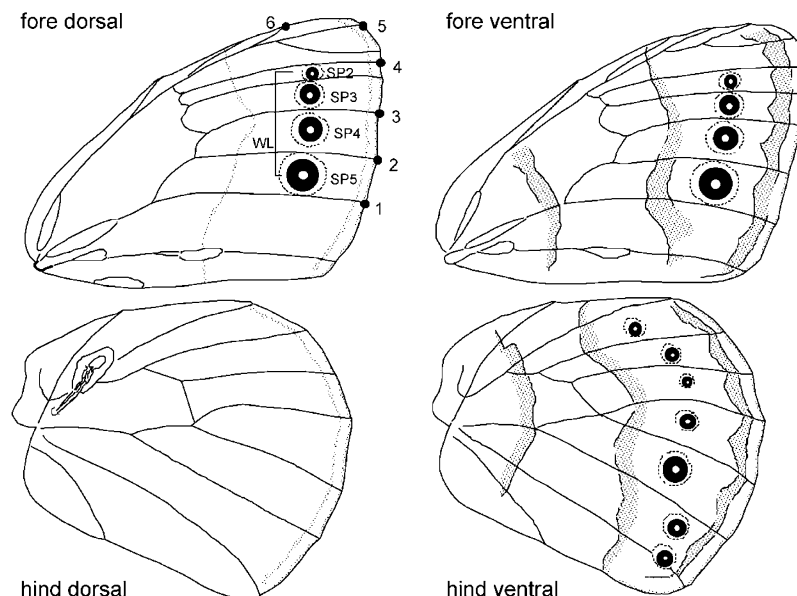
Individuals with lowered FA (because of selection) are predicted to exhibit higher fitness (Møller, 1997). Various studies suggest that this effect might be enhanced after exposure to stressful conditions during development (Watson and Thornhill, 1994; but see Bjorksten *et al.*, 2000a). Furthermore, it is generally assumed that variation in the expression of phenotypes increases as environmental conditions deteriorate and deviate from the optimum (Møller and Swaddle, 1997). Therefore, at the end of artificial selection we also reared both unselected control, and selected lines under a stress environment. We predicted that the stress might increase both the mean and variance of FA of the novel eyespots more in the unselected butterflies than in the selected butterflies.

## Materials and methods

### Experimental animals

The butterflies used in this experiment were derived from a pure breeding laboratory *Spotty* line (Figure 1). The *Spotty* allele almost certainly arose as a spontaneous mutant in our stock although, since it was the first mutant phenotype we established, it might have been present as a rare allele in the original founder population of 80 or so gravid females collected in Malawi in 1988 (Brakefield *et al.*, 1996). The latter scenario is unlikely because our *Spotty* line was not established until at least 25 generations after obtaining the founder population, and *Spotty* shows some incomplete dominance making it unlikely that we would have overlooked its presence in earlier generations. Here, we refer to *Spotty* as a mutant allele implying a recent origin and no history of selection. Early in establishment of the *Spotty* line, it was extensively outcrossed to the laboratory stock to ensure a heterogeneous genetic background.

To establish the selection lines we measured 552 butterflies from the *Spotty* line. These butterflies were



**Figure 1** A spotty *Bicyclus anynana*. The four measured eyespots on the dorsal forewing are indicated by SP2–SP5. The area of the white pupil was measured. The distance between the midpoints of SP2 and SP5 is called the interfoveal distance and was used as a measure of wing length (WL). The numbers 1–6 refer to the measured landmarks.

then used to establish the control, RELFA and SP4FA lines. We used 35 males and 35 females as selected founders of each line. The same numbers of founders for the control lines (UC) were chosen at random. We applied selection in two different ways to examine firstly whether it was possible to obtain lower FA of the pair of novel eyespots (SP3 and SP4, see Figure 1) combined relative to the pair of wild-type flanking eyespots (SP2 and SP5, see Figure 1), and secondly whether a response occurs when selection is targeted only on the novel eyespot 4. These options are referred to here as RELFA and SP4FA, respectively. We choose eyespot 4 for the single target eyespot as pilot experiments indicated that this eyespot showed the highest levels of FA, especially in males. Selected butterflies had mated and spent at least 5 days at 27°C before eggs were collected. Two batches of eggs per line were used to establish two replicates of each line (cohorts 1 and 2). The lines thus established will be designated as the UC1, UC2, RELFA1, RELFA2, SP4FA1 and SP4FA2. Larvae of each generation were raised on young maize plants with each line in a separate cage in the same climate room at 27°C, 12:12 L:D, and high relative humidity. Larval densities were kept uniform.

Over the next five generations, we selected butterflies in each of the six lines in the manner described above. In the final two generations (ie generations 6 and 7), we continued the lines without selection and examined environmental contributions to the variation in FA. Firstly, to investigate the extent to which small (random) differences in environmental conditions across rearing cages contributed to variation in FA, we divided each line among two replicates reared in separate cages. The cages were divided into two randomized blocks of six reared on different sides of the same climate room. To investigate the effects of a stress environment (ie a large environmental difference) and to examine whether the selection could cause differences among lines in the response to stress, we applied a heat shock to pupae of each line reared in the next generation. Again replicates were used, in one of which pupae now received a heat shock 1 h after pupation whereas pupae of the other 'control' group remained at 27°C. One hour after pupation coincides with a critical signalling phase in eyespot determination (Beldade and Brakefield, 2002). The heat-shock (5 h at 39.5°C) was chosen to be severe. It caused a slight increase in mortality of up to 5.0%. In addition, heat-shock protein (HSP) production is known to increase at this temperature (Breuker and Brakefield, 2003).

Adult eclosion, mating and egg laying all occurred in cylindrical hanging cages. Egg hatching in *B. anynana* is highly sensitive to inbreeding. Egg hatching averaged 65% over generations 4–7, with no significant differences between lines or generations. About two-thirds of females laid fertile eggs, of which an average of 95.6% hatched. These observations considered together with direct estimates from other studies (Brakefield *et al.*, 2001) indicate that under our rearing conditions, effective population size is likely to have been maintained at above 30 individuals.

#### Obtaining data and measurements

The (virgin) female and male butterflies were separated about 3 h after emergence and placed at  $\pm 19^\circ\text{C}$  to slow

down their activity and thus wing wear. One day after emergence both the left and right dorsal side of the butterfly wings were digitally photographed with a Leica DC 2000 mounted on a Wild binocular (magnification 8.82) by carefully holding the butterfly down with a forceps, so as not to damage the wings. The order of the photography was random with respect to line, sex and side. In each generation, replicate photographs were taken of a random subset of 30 butterflies, to check the measurement error (ME). The butterflies were released and repositioned between replicate photos. Measurements on a total of 12516 butterflies (on average 145 females and 135 males per line per generation) were carried out using SCION IMAGE (freeware from NIH, USA, 1998). The first few and last butterflies to emerge were not measured or used for selection.

Eyespot size was assessed by measuring the area (in  $\text{mm}^2$ ) of the white 'pupil' in each of the four dorsal eyespots (Figure 1). The pupil is the portion of the eyespot, which can be measured most accurately. Measurement repeatability and position on the wing were used as criteria for selecting six landmarks to analyze wing size and shape in generations 6 and 7 (cf. Klingenberg and McIntyre, 1998). Repeatabilities of the trait values were high (eyespots: 92–98%; landmarks: 95–99.5%) and comparable between generations. There was no significant heterogeneity across lines, traits, generations and sexes in measurement error (using methods described by Van Dongen *et al.*, 1999a, b), while measurement error itself was very low (using the two-factor ANOVAs of Palmer (1994): measurement error variance significantly lower than the side  $\times$  individual interaction variance). The coordinates of the landmarks were also used to calculate the centroid size, as a measure of wing size (Klingenberg *et al.*, 2001). This is the square root of the sum of squared distances from a set of landmarks to their centroid (references and details in Klingenberg and McIntyre, 1998).

#### Statistical analyses and selection procedure

Analyses of FA were carried out according to Palmer (1994). In each generation, males, females and lines were analyzed separately. As all eyespots are formed by the same developmental mechanisms, the sequential Bonferroni procedure was applied to each set of related tests to avoid making type I errors (Leamy, 1999; Klingenberg *et al.*, 2001).

FA did not significantly correlate with trait size in any generation. Part of the FA variation could, however, be explained by size differences among traits, but not among lines (data not shown). In this case, we recommend dividing FA1 (the absolute differences between the sides) by the overall mean trait size, that is the mean trait size of all lines. We will call this measure FA19, which is a multisample variant of FA3 (Palmer and Strobeck, 2003). This standardization enables the comparison of traits differing significantly in size. The combined measure of FA, that is combining FA19 across more than one trait, is CFA19. Distributions of the signed differences between left and right were tested for departures from normality and a mean of zero, that is to test for departures from ideal FA, using kurtosis, skewness and a *t*-test (Palmer, 1994). When DA was present, the mean signed difference was subtracted from

the individual signed difference between the sides before calculating the FA19 values.

The repeatability ( $\mathfrak{R}$ ) of a trait measures the reliability with which it can be measured.  $\mathfrak{R}$  sets an upper limit to the heritability of FA and thereby DI, such that

$$h_{DI}^2 = h_{FA}^2 / \mathfrak{R} \text{ (Houle, 2000; Fuller and Houle, 2003)}$$

$\mathfrak{R}$  is estimated by using the CV of FA:

$$CV_{FA} = \sqrt{(V_{FA} / \text{mean}_{FA}^2)}$$

The estimation of  $CV_{FA}$  is most accurate with a low variance in DI and/or low ME, when  $CV_{FA}$  approaches a value of 75.5 ( $= 0.755 \times 100$ ) (Merila and Bjorklund, 1995; Bjorklund and Merila, 1997; Houle, 1997, 2000; Whitlock, 1998).  $CV_{FA}$  values deviating substantially from 75.5 usually suggest a serious bias because of ME ( $V_{FA}$  contains  $V_{ME}$ ), but when ME has been kept low such a deviation may indicate significant genetic variance in FA (Houle, 1997, 2000). The accuracy of our measurements was high and outliers have been excluded (see further below) so, it can be assumed that ME contributed a negligible amount of variation to  $CV_{FA}$  (ie  $V_{ME} \approx 0$ ). In each generation, we have calculated the  $CV_{FA}$  values for each of the sexes, eyespots and lines. These values were used to examine (1) the amount of variation present for FA; (2) whether the different eyespots differed in (genetic) variation for FA; (3) whether either the applied selection or stress altered the  $CV_{FA}$  values; and (4)  $\mathfrak{R}$ , which indicates the reliability of estimating the genetic basis of DS by means of FA in this selection experiment. In each generation, we discarded on average 7.6% ( $\pm 0.68\%$ ) of the butterflies, the outliers, prior to analysis. Outliers were identified using Grubb's test statistic ( $t_G$ ; Palmer and Strobeck, 2003) and, in general, they differed by more than 3 SD's from the mean signed asymmetry value. Outliers significantly increase  $CV_{FA}$  and thereby inflate heritability estimates of DS based on FA (Whitlock, 1996; Houle, 2000).

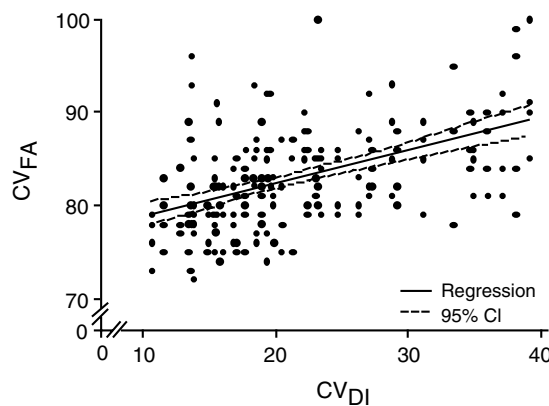
Whenever individual differences in DI exist, the signed asymmetry distribution should be leptokurtically distributed (Leung and Forbes, 1997; Gangestad and Thornhill, 1999). Figure 2 of Gangestad and Thornhill (1999) has been used to estimate the CV of DI ( $CV_{DI}$ ) and the concordance in DI between the four eyespots on the basis of the weighted kurtosis values of the signed differences between the sides of the four eyespots.

FA19 values were transformed prior to applying parametric tests that assume normality as they are half-normally distributed. In each generation, for both males and females, we applied Box-Cox power transformations of the form

$$(FA19 + 0.000001)^{0.33}$$

to achieve normality (Leamy, 1999; Van Dongen *et al.*, 1999b; Fuller and Houle, 2003). The  $\lambda$  did not differ among the four eyespots, and had a value of 0.33, yielding (standardized) asymmetry values for the different eyespots which remain comparable after transformation. Leamy (1999) reported the same value of 0.33 for mouse mandible traits. The  $\lambda$  of the size of the forewing (centroid size) was 0.45.

The butterflies to be used as control parents in each generation were chosen at random. In the first five generations, parents in the two RELFA lines were selected for minimal differences between the wild type



**Figure 2**  $CV_{FA}$  plotted against  $CV_{DI}$ . The equation for the regression line (with 95% CI) is  $CV_{FA} = 75.4 + 0.35 CV_{DI}$ , which was significant ( $P \ll 0.001$ ). The coordinates of each point are  $(CV_{DI_{i,j,k}}, CV_{FA_{i,j,k}})$ , with  $i$  = generation (0–7);  $j$  = line (1–3, except in generation 7 then 1–6);  $k$  = sex (females or males); 1 = eyespot (2–5). Data can be found in Table 2.

and novel eyespots by choosing the butterflies with the lowest values of

$$CFA19_{RELFA} = FA19_{spot2} + FA19_{spot5} \\ - FA19_{spot3} - FA19_{spot4}$$

In the final two generations, parents of all lines were chosen at random. Selection for parents in the two SP4FA lines in the first five generations consisted of taking the butterflies with the lowest FA19 values for spot 4. For each line in each generation, we selected the 35 most extreme individuals of each sex. This corresponded to selection intensities of 1.2–1.4 (Falconer and MacKay, 1996).

We estimated the realized heritability of FA of each eyespot per selection line by regressing the response of selection against the cumulated selection differential (Falconer and MacKay, 1996). The calculation of the response to selection and cumulated selection differential consisted of several steps. We first corrected for the presence of DA by subtracting the mean signed difference from the individual signed differences between the sides. This was done per sex, per line and per generation. We then standardized the absolute values of these differences (FA19) and transformed the data with a Box-Cox transformation ( $\lambda = 0.33$ ). The mean asymmetry values of the animals used as parents was then subtracted from the total mean, per generation and per line, and we then took the cumulative values across generations. This was done for the female parents, the male parents and both sexes pooled (cf. Leamy, 1999). These were the cumulated selection differentials. The response to selection was corrected by subtracting the mean FA of the controls from the selection lines. We selected for a reduced FA, which with a positive response would result in (significant) negative slopes of the regression lines. The absolute value of the slope is an approximation of the average (realized) heritability. We also carried out the calculations on heritability without correcting for DA to compare the results with those based on the corrected asymmetry values and to examine whether selection on FA had an effect on DA (see Graham *et al.*, 1993).

Factors used in ANOVAs are sex, line, individual, side and trait. To evaluate the significance of environmental variance on FA, which can confound interpretation of heritability estimates, we examined the effects of (small) random environmental differences and the impact of the heat stress by means of nested ANOVAs. Cohort was nested within line, and 'treatment' was nested within cohort and line. 'Line' can be UC, RELFA or SP4FA. 'Cohort' refers to either 1 or 2 (eg UC1, UC2, etc). 'Replicate' refers to the environmental replicate, A or B (eg UC1A, UC1B, UC2A, etc) (see also Polak and Starmer, 2001). Furthermore, modified Levene's tests for heterogeneity of variances were carried out to investigate the variability in individual FA and trait FA (Palmer, 1994).

The FA of shape was examined by using geometric morphometrics based on Procrustes methods, which were carried out with MORPHEUS (Slice, 1998). The analyses done here are based on the methods of Klingenberg and McIntyre (1998). The square root of the sum of the squared distances between corresponding landmarks of two optimally aligned configurations is an approximation of Procrustes Distance. This distance is similar to size-scaled unsigned asymmetry (FA19) for linear measurements, including the form of its distribution (related to the  $\chi^2$  distribution), as there are no

negative values. We will use the Procrustes distance as a measure of shape FA. For both males and females, we used log transformations of the form

$$(\log(\text{Procrustes distance}) + 5.0)$$

to achieve normality.

Significance is at the 0.05 level. All analyses were carried out with the statistical software package MINITAB 12.22.

## Results

### Directional and fluctuating asymmetry

All four eyespots showed significant DA for each sex in every line and generation ( $t$ -tests:  $P \ll 0.001$ ). None of the skewness values calculated was significant (range from  $-0.91$  to  $+0.82$ ). There were fluctuations from generation to generation in the level of DA, but as there were no consistent differences between the generations for a particular line (detailed later in Table 4) the values of DA reported in Table 1a are based on pooled data over all generations. In each sex, generation and line, the levels and direction of DA of each of the eyespots differed significantly (one-way ANOVA, trait as factor,  $P \ll 0.001$ ). The mean area of the white 'pupils' of each of the

**Table 1** Mean  $\pm$  standard errors for each sex, line and eyespot

Sex	Line	N Eyespots	Eyespot 2	Eyespot 3	Eyespot 4	Eyespot 5	N wing size	Forewing size	
<i>(a) Directional asymmetry</i>									
Females	UC1	899	$-0.109 \pm 0.008$	$-0.047 \pm 0.010$	$-0.045 \pm 0.010$	$0.060 \pm 0.006$	291	$-0.213 \pm 0.005$	
	UC2	1026	$-0.099 \pm 0.008$	$-0.036 \pm 0.010$	$-0.050 \pm 0.009$	$0.056 \pm 0.006$	284	$-0.227 \pm 0.005$	
	RELFA1	1068	$-0.088 \pm 0.007$	$-0.039 \pm 0.008$	$-0.073 \pm 0.007$	$0.047 \pm 0.004$	291	$-0.225 \pm 0.004$	
	RELFA2	946	$-0.095 \pm 0.008$	$-0.017 \pm 0.009$	$-0.036 \pm 0.009$	$0.043 \pm 0.005$	277	$-0.244 \pm 0.005$	
	SP4FA1	908	$-0.102 \pm 0.008$	$-0.042 \pm 0.010$	$-0.052 \pm 0.008$	$0.058 \pm 0.005$	279	$-0.240 \pm 0.005$	
	SP4FA2	1058	$-0.112 \pm 0.008$	$-0.035 \pm 0.009$	$-0.061 \pm 0.009$	$0.036 \pm 0.005$	270	$-0.233 \pm 0.005$	
Males	UC1	791	$-0.150 \pm 0.012$	$-0.099 \pm 0.014$	$-0.089 \pm 0.019$	$0.062 \pm 0.009$	232	$-0.222 \pm 0.007$	
	UC2	923	$-0.111 \pm 0.011$	$-0.054 \pm 0.012$	$-0.072 \pm 0.016$	$0.040 \pm 0.008$	283	$-0.266 \pm 0.006$	
	RELFA1	941	$-0.138 \pm 0.011$	$-0.057 \pm 0.010$	$-0.090 \pm 0.014$	$0.053 \pm 0.007$	252	$-0.262 \pm 0.006$	
	RELFA2	842	$-0.123 \pm 0.011$	$-0.066 \pm 0.012$	$-0.042 \pm 0.018$	$0.047 \pm 0.008$	282	$-0.278 \pm 0.006$	
	SP4FA1	801	$-0.120 \pm 0.011$	$-0.085 \pm 0.012$	$-0.069 \pm 0.016$	$0.064 \pm 0.008$	299	$-0.262 \pm 0.005$	
	SP4FA2	867	$-0.180 \pm 0.011$	$-0.119 \pm 0.013$	$-0.099 \pm 0.016$	$0.070 \pm 0.008$	206	$-0.276 \pm 0.007$	
<i>(b) Fluctuating asymmetry</i>									
Females	UC1	899	$0.516 \pm 0.006$	$0.562 \pm 0.006$	$0.549 \pm 0.006$	$0.467 \pm 0.005$	291	$1.790 \pm 0.023$	$0.264 \pm 0.007$
	UC2	1026	$0.524 \pm 0.005$	$0.562 \pm 0.006$	$0.542 \pm 0.006$	$0.470 \pm 0.005$	284	$1.815 \pm 0.025$	$0.317 \pm 0.008$
	RELFA1	1068	$0.516 \pm 0.005$	$0.538 \pm 0.005$	$0.524 \pm 0.005$	$0.445 \pm 0.004$	291	$1.874 \pm 0.024$	$0.294 \pm 0.006$
	RELFA2	946	$0.522 \pm 0.005$	$0.552 \pm 0.006$	$0.543 \pm 0.006$	$0.455 \pm 0.005$	277	$1.819 \pm 0.024$	$0.310 \pm 0.008$
	SP4FA1	908	$0.516 \pm 0.006$	$0.555 \pm 0.006$	$0.532 \pm 0.005$	$0.460 \pm 0.005$	279	$1.781 \pm 0.025$	$0.288 \pm 0.007$
	SP4FA2	1058	$0.536 \pm 0.005$	$0.567 \pm 0.005$	$0.550 \pm 0.006$	$0.467 \pm 0.005$	270	$1.813 \pm 0.024$	$0.297 \pm 0.008$
Males	UC1	791	$0.584 \pm 0.007$	$0.614 \pm 0.007$	$0.673 \pm 0.008$	$0.541 \pm 0.006$	232	$1.912 \pm 0.031$	$0.297 \pm 0.008$
	UC2	923	$0.587 \pm 0.006$	$0.594 \pm 0.006$	$0.644 \pm 0.007$	$0.524 \pm 0.006$	283	$1.924 \pm 0.023$	$0.355 \pm 0.007$
	RELFA1	941	$0.587 \pm 0.006$	$0.573 \pm 0.006$	$0.638 \pm 0.007$	$0.499 \pm 0.005$	252	$1.766 \pm 0.024$	$0.328 \pm 0.007$
	RELFA2	842	$0.576 \pm 0.006$	$0.593 \pm 0.007$	$0.667 \pm 0.008$	$0.517 \pm 0.006$	282	$1.986 \pm 0.027$	$0.374 \pm 0.007$
	SP4FA1	801	$0.576 \pm 0.006$	$0.596 \pm 0.007$	$0.626 \pm 0.008$	$0.520 \pm 0.006$	299	$1.877 \pm 0.025$	$0.310 \pm 0.007$
	SP4FA2	867	$0.589 \pm 0.006$	$0.599 \pm 0.007$	$0.652 \pm 0.007$	$0.523 \pm 0.006$	206	$1.938 \pm 0.029$	$0.354 \pm 0.009$

Generations, except the parents, are pooled. Forewing size is the centroid size as based on generations 6 and 7. Number of individuals (N) is indicated for each trait. All values are highly significant ( $t$ -tests,  $P \ll 0.001$ ).

eyespot 2, 3 and 4 was always significantly larger on the left wing than the right. The pattern was reversed for eyespot 5. The left forewing was on average significantly larger than the right forewing in all lines and in both sexes (Table 1a). DA of wing size was higher for males and differed significantly among lines, with the selected lines usually showing the largest differences between the sides (nested ANOVA:  $F_{1,3222sex} = 98.5$ ,  $F_{22,3222line\ within\ sex} = 5.23$ ;  $P \leq 0.001$  in each case). The wing size DA and eyespot DA did not correlate, however, the signed differences between the sides for each eyespot was independent from those between the left and right forewing ( $2 \times 12 \times 4 \times 2 = 192$  regression analyses,  $R^2$  ranged from 0.0 and 1.7%). In other words, the relative area of the white 'pupil' of eyespots 2, 3 and 4 on the left wing is not larger than that on the right wing because the left wing is larger. Various tests (see above) have also shown that the DA and FA of a trait did not significantly correlate with its size or with the size of the individual.

The variation in the signed differences between the sides around the mean difference between the sides (ie FA corrected for DA) was highly significant for each eyespot for both sexes and in every generation and line ( $t$ -tests:  $P \leq 0.001$ ) (Table 2). In other words, there was significant FA. Furthermore, these levels differed significantly between the sexes and traits (two-way ANOVAs: for each factor,  $P \leq 0.001$ ). Males had significantly

higher FA, as did the two novel eyespots. The FA of each eyespot did not correlate with the FA of forewing shape or size (the log-transformed Procrustes distance). In both the sixth and seventh generation and for each line and sex ( $2 \times 2 \times 12 \times 4 \times 2 = 384$  regression analyses), the  $R^2$  was between 0.0 and 2.9% and never statistically significant. FA of the forewing size and shape was significantly higher in males and significantly different among lines (*size* nested ANOVA:  $F_{1,3222sex} = 91.04$ ,  $F_{22,3222line\ within\ sex} = 6.99$ ; *shape* nested ANOVA:  $F_{1,3222sex} = 33.52$ ,  $F_{22,3222line\ within\ sex} = 3.40$ ,  $P$ -values  $\leq 0.001$ ; see also Table 1b).

A modified Levene's test for heterogeneity of variances with individuals as a random effect and trait as a fixed effect (Palmer, 1994) showed that the  $F$ -values for both effects were highly significant in each generation (per sex and line;  $P \leq 0.001$ ). The novel eyespots were thus repeatably less stable than the others and the level of DI varied significantly among individuals. We will now examine the individual variation in DI further by means of the  $CV_{FA}$ .

#### Heterogeneity in individual levels of developmental stability

In each generation, we calculated the  $CV_{FA}$  values for each sex, eyespot and line to examine: (1) the amount of variation present for FA, (2) whether different eyespots differed in (genetic) variation for FA and (3) whether

**Table 2** Coefficients of variation for fluctuating asymmetry ( $CV_{FA}$ ) and developmental stability ( $CV_{DI}$ ) for each sex in each line

		Females					Males				
		$CV_{Fasp\ 2}$	$CV_{Fasp\ 3}$	$CV_{Fasp\ 4}$	$CV_{Fasp\ 5}$	$CV_{DI}$	$CV_{Fasp\ 2}$	$CV_{Fasp\ 3}$	$CV_{Fasp\ 4}$	$CV_{Fasp\ 5}$	$CV_{DI}$
Gen 0		79	84	85	86	23	81	80	76	83	20
Gen 1	UC	86	85	79	80	25	85	81	85	83	23
	RELFA	79	82	76	79	17	80	82	89	74	16
	SP4FA	86	87	77	76	19	79	81	84	86	18
Gen 2	UC	91	77	82	79	16	86	79	88	80	22
	RELFA	79	80	75	76	15	81	75	84	86	18
	SP4FA	85	76	75	81	17	80	85	100	82	23
Gen 3	UC	82	82	75	79	16	83	86	81	80	18
	RELFA	75	78	79	85	15	78	77	84	78	13
	SP4FA	82	75	78	87	20	84	78	84	82	17
Gen 4	UC	84	96	79	99	38	90	91	85	100	39
	RELFA	82	86	84	86	27	89	86	85	81	35
	SP4FA	85	81	89	93	29	88	88	95	78	33
Gen 5	UC	86	84	90	90	35	84	90	92	81	37
	RELFA	81	84	90	88	36	80	88	86	82	29
	SP4FA	79	87	89	85	31	82	92	81	87	27
Gen 6	UC	86	88	90	85	22	92	86	85	74	19
	RELFA	78	83	80	75	12	80	80	82	78	14
	SP4FA	81	83	79	82	19	87	75	85	85	21
Gen 7	UC1	82	86	82	92	20	93	82	85	83	18
	UC2	86	77	83	83	19	78	79	80	72	14
	RELFA1	82	78	89	81	19	78	75	79	89	13
	RELFA2	82	79	78	83	15	73	78	76	79	11
	SP4FA1	78	85	77	77	14	80	86	75	81	18
	SP4FA2	79	88	83	80	26	96	87	93	73	14

Generation 0 is the parental generation. The replicates per line (Cohorts 1 and 2) were pooled, as the CVs did not differ significantly. In the last generation (1) refers to controls with no heat shock, and (2) refers to the heat-shock treatment.

selection or stress altered the  $CV_{FA}$  values. The results are presented in Table 2. The average  $CV_{FA}$  of the eyespots for both sexes varied between 81 and 87, differing significantly from 75.5 (Table 2). For each eyespot, there were no significant or consistent differences between the generations or the lines in either sex (Kruskal–Wallis tests:  $P > 0.05$ ). There are thus no indications that the selection led to changes in the amount of (genetic) variation in any eyespot. Furthermore, there were also no consistent differences between lines within the final generation (7) for each eyespot. The heat stress thus did not result in differences in the individual variation in FA between the different selection lines and the unselected controls (Table 2).

We have also examined whether the (significant) variation in FA reflects genetic variation in the underlying DS (Houle, 1997, 2000; Gangestad and Thornhill, 1999). Figure 2 shows the relationship between  $CV_{FA}$  and  $CV_{DI}$  for the values indicated in Table 2 (Pearson regression analysis:  $CV_{FA} = 75.4 + 0.35CV_{DI}$ ,  $R^2 = 24.5\%$ ,  $F_{1,198} = 64.3$ ,  $P \leq 0.001$ ). Five conclusions can be drawn. First, although the relationship is statistically significant, it is unreliable to estimate individual DI using FA of one trait only, as is apparent from the wide scatter around the regression line caused by traits exhibiting different levels of  $CV_{FA}$  but sharing the same factors affecting DI. Second, according to Gangestad and Thornhill (1999), the range of observed  $CV_{DI}$  values indicate that the individual difference factors underlying the DI of the different eyespots were completely shared. We have seen that individual traits may nevertheless differ in FA (and DA). Third, when  $CV_{DI}$  approaches the value of 0, the  $CV_{FA}$  approaches the value of 75.5 (the intercept of the

regression line is 75.4),  $\mathfrak{R}$  also becomes 0, and it is thus no longer possible to reliably estimate the underlying genetic basis of DS. Fourth, the average  $CV_{FA}$  of the eyespots for both males and females was between 81 and 87, which corresponded to  $\mathfrak{R}$ -values of between 0.082 and 0.16. This implies that the  $h^2_{DI}$  estimates are 6–12 times those of  $h^2_{FA}$  (cf. Leamy, 1999). Fifth, the  $CV_{DI}$  values given in Table 2 are comparable to those reported for traits under directional selection. This corresponds with the observation that DS is assumed to be under directional selection (Gangestad and Thornhill, 1999).

### Environmental variance

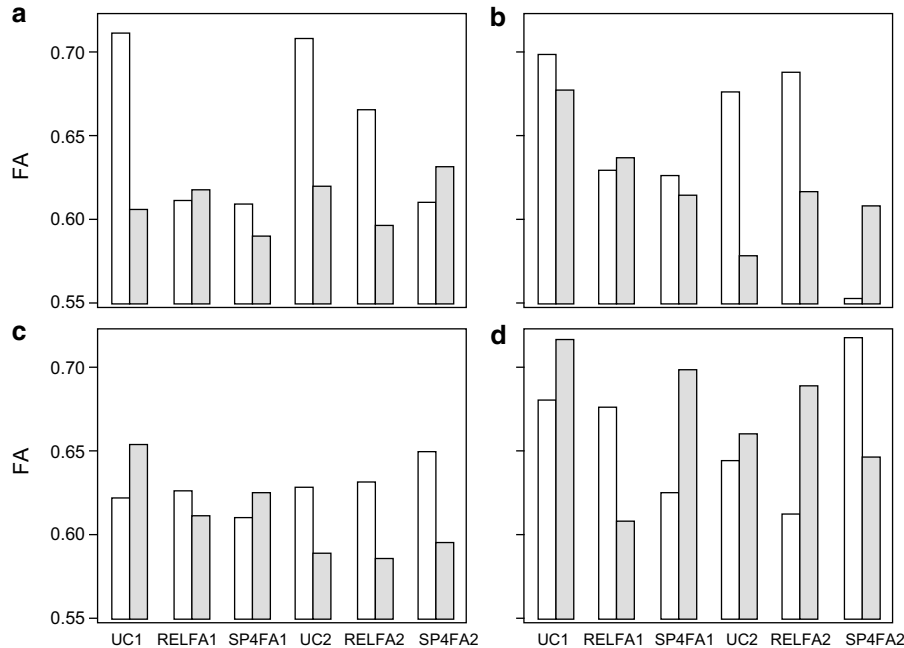
Individuals differ in DS because of environmental factors as well as genetic ones. Therefore, prior to estimating the response to selection and  $h^2_{FA}$  and  $h^2_{DI}$  for the various eyespots, we need to determine how much of the variation in FA can be accounted for by environmental variance (ie differences, albeit small, in rearing conditions between lines or cages).

In the sixth generation, we compared the differences in FA of lines with a different genetic background and those with the same (ie the environmental replicates). The 48 possible *t*-tested pairwise comparisons of the environmental replicates (six lines replicated, four eyespots and two sexes) of the FA (corrected for DA and Box–Cox transformed with  $\lambda = 0.33$ ) were all nonsignificant. From Table 3a, we can conclude that environmental differences accounted for extremely low proportions of overall variation in FA. However, replicates within lines were more similar to each other than to the samples from different lines. Thus when differences between lines

**Table 3** Three-level nested ANOVA results for each sex

Sex	Source	Eyespot 2			Eyespot 3			Eyespot 4			Eyespot 5		
		df	F	% of var	df	F	% of var	df	F	% of var	df	F	% of var
<i>(a) Sixth generation</i>													
Females	Line	2	1.26	0.15	2	8.81*	2.02	2	1.85	0.45	2	9.83*	2.87
	Cohort (line)	3	0.78	0.07	3	2.83	1.59	3	0.63	0.00	3	0.59	0.00
	Replicate (line cohort)	6	0.70	0.00	6	0.39	0.00	6	0.75	0.00	6	2.06	1.37
	Within replicate (=error)	873		99.79	873		96.36	873		99.55	873		99.75
Males	Line	2	0.47	0.00	2	1.22	0.00	2	0.69	0.00	2	3.48	0.68
	Cohort (line)	3	0.85	0.00	3	5.88*	3.49	3	3.19	1.95	3	3.06	0.02
	Replicate (line cohort)	6	1.37	0.57	6	1.50	0.74	6	0.52	0.00	6	1.13	0.19
	Within replicate (=error)	772		99.43	772		95.77	772		98.05	772		99.11
<i>(b) Seventh generation</i>													
Females	Line	2	7.75*	1.11	2	4.93*	1.34	2	5.13*	0.86	2	0.68	0.00
	Cohort (line)	3	3.87	1.78	3	0.46	0.00	3	2.79	0.29	3	1.69	0.00
	Treatment (line cohort)	6	0.97	0.00	6	3.46	3.51	6	2.31	1.91	6	2.43	2.11
	Within treatment (=error)	795		97.11	795		95.15	795		96.94	795		97.89
Males	Line	2	1.15	0.00	2	0.16	0.00	2	1.64	0.00	2	2.25	0.63
	Cohort (line)	3	2.95	1.61	3	0.53	0.00	3	1.47	0.00	3	0.61	0.00
	Treatment (line cohort)	6	1.24	0.37	6	1.07	0.11	6	2.92	2.92	6	1.53	0.81
	Within treatment (=error)	758		98.02	758		99.89	758		97.08	758		98.56

The df, *F*-values and percentage of variation that can be explained by the (nested) factors are given. ‘Line’ can be UC, RELFA or SP4FA. ‘Cohort’ refers to either 1 or 2 (eg, UC1, UC2, etc). In (a) ‘Replicate’ refers to the environmental replicate, A or B (eg, UC1A, UC1B, and UC2A, etc), in (b) ‘Treatment’ refers to the applied stress, A (=no heat shock) or B (=heat shock) (eg, UC1A, UCA1B, and UCB2A, etc). \* $P < 0.01$  in sequential Bonferroni tests of significance.



**Figure 3** Heat stress experiment in generation 7. Histograms depicting average FA19 values per line and treatment of (a) SP3 in females, (b) SP4 in females, (c) SP3 in males, and (d) SP4 in males. See text for line abbreviations. Open bars indicate FA19 values of untreated animals, the solid bars the FA19 values of heat-stressed animals.

occurred, they were more likely to be because of different genetic backgrounds as a result of the applied selection. These differences, however, were usually not consistent across generations as is examined further below in the analyses of heritability.

The heat-shock treatment in the final generation did not increase the  $CV_{FA}$  values and, furthermore, the pattern of FA variation across lines and replicates (in this case treated versus nontreated) is similar to that in the previous generation (Table 3b). Most of the variation in FA resides at the individual level with (very) little explained by differences between lines. Inspection of the 48 (nonsignificant) pairwise comparisons between the stressed and nonstressed replicates within a line revealed that the heat shocked females, but not the males, tended to show lower, rather than higher FA (16 of the 24  $t$ -test comparisons), although this is not a significant trend (one-proportion test:  $\hat{P} = 0.15$ ) (Figure 3). Such a trend is contrary to the expectation that a stress will result in elevated levels of FA and that selection lines will differ in their response from the unselected controls.

#### Response to selection and heritability estimates of both DA and FA

The heritability estimates of the Box–Cox transformed differences between the sides, with or without correction for DA, are the absolute values of the regression coefficients of the relation between the response to selection and cumulated selection differentials given in Table 4. All heritability estimates were nonsignificantly different from 0, in both the selected and unselected eyespots. In general, FA decreased or increased from generation to generation independently of the selection differentials. Differences in FA for the same eyespot

existed between the lines when averaged over all generations, but these were independent of the applied selection (Table 1b). The SP4FA lines behaved most consistently for eyespot 4 and the combined selection criterion relative to the unselected controls throughout the experiment (the SE values of the  $h^2_{FA}$  estimates are the lowest and the slopes are negative, Table 4b), perhaps because only a single target trait was involved.

#### Discussion

Although the novel eyespots significantly differed from the wild-type eyespots in FA and there was significant variation among individuals in DS, we were nevertheless unable to select for increased DS by means of FA. The selection intensities were high, ranging from 1.2 to 1.4. The heritability estimates were nonsignificantly different from 0, which is consistent with the low or nonsignificant estimates reported by other studies (Fuller and Houle, 2003).

Individuals differ in FA because of environmental factors as well as genetic ones. We have shown that although  $CV_{FA}$  correlates significantly with  $CV_{DI}$ , the  $CV_{FA}$  value of a single trait is likely to be a poor indicator of the underlying  $CV_{DI}$  (Houle, 2000; Leung *et al*, 2000). In a similar way we can conclude that an individual asymmetry value for any trait may be a poor indicator of its DS. The error involved in estimating the variance because of developmental noise at the individual level with only two data points (left and right side) is probably too high (Whitlock, 1996, 1998; Whitlock and Fowler, 1997; Lens and Van Dongen, 1999). Environmental variance did not cause differences between lines, but may nevertheless have caused differences between individual asymmetry values within a line. This result,



**Table 4** Regression coefficients  $\pm$  standard errors describing the response to selection to the cumulated selection differential for both parents, or male or female parents

Eyespot	Line	Midparents	Male parents	Female parents
<i>(a) Data not corrected for DA</i>				
2	RELFAA	$-0.24 \pm 0.19$	$0.07 \pm 0.11$	$0.08 \pm 0.11$
	RELFAB	$0.39 \pm 0.17$	$-0.05 \pm 0.07$	$-0.06 \pm 0.09$
	SP4FAA	$0.41 \pm 0.35$	$0.03 \pm 0.32$	$0.42 \pm 0.34$
	SP4FAB	$0.34 \pm 0.16$	$0.20 \pm 0.56$	$-0.26 \pm 0.32$
3	RELFAA	$-0.70 \pm 0.53$	$-0.26 \pm 0.57$	$-0.55 \pm 0.41$
	RELFAB	$0.08 \pm 0.65$	$-2.13 \pm 0.86$	$-0.34 \pm 0.41$
	SP4FAA	$-0.21 \pm 0.64$	$-0.32 \pm 0.35$	$-0.76 \pm 0.46$
	SP4FAB	$-0.29 \pm 0.65$	$-0.06 \pm 0.33$	$-0.03 \pm 0.44$
4	RELFAA	$0.14 \pm 0.94$	$1.15 \pm 0.78$	$-0.43 \pm 1.41$
	RELFAB	$-0.13 \pm 0.29$	$-0.16 \pm 0.12$	$0.13 \pm 0.34$
	SP4FAA	$-0.02 \pm 0.03$	$-0.01 \pm 0.04$	$-0.01 \pm 0.04$
	SP4FAB	$-0.02 \pm 0.02$	$-0.03 \pm 0.02$	$-0.04 \pm 0.02$
5	RELFAA	$-0.19 \pm 0.63$	$-0.36 \pm 0.08$	$-0.21 \pm 0.10$
	RELFAB	$0.02 \pm 0.53$	$-0.22 \pm 0.28$	$-0.45 \pm 0.33$
	SP4FAA	$0.22 \pm 0.20$	$-0.15 \pm 0.31$	$0.32 \pm 0.33$
	SP4FAB	$2.29 \pm 0.98$	$-0.44 \pm 0.89$	$-0.12 \pm 0.43$
<i>(b) Data corrected for DA</i>				
2	RELFAA	$0.08 \pm 0.26$	$-0.07 \pm 0.10$	$-0.07 \pm 0.10$
	RELFAB	$0.08 \pm 0.24$	$0.03 \pm 0.05$	$0.07 \pm 0.07$
	SP4FAA	$0.30 \pm 0.32$	$-0.07 \pm 0.37$	$0.40 \pm 0.41$
	SP4FAB	$-0.06 \pm 0.09$	$0.08 \pm 0.06$	$-0.04 \pm 0.07$
3	RELFAA	$-0.70 \pm 0.62$	$-0.19 \pm 0.31$	$-0.60 \pm 0.32$
	RELFAB	$-0.06 \pm 0.68$	$-2.01 \pm 1.39$	$-0.35 \pm 0.39$
	SP4FAA	$-0.05 \pm 0.77$	$0.15 \pm 0.81$	$-0.44 \pm 0.45$
	SP4FAB	$0.54 \pm 0.73$	$0.06 \pm 0.28$	$0.05 \pm 0.35$
4	RELFAA	$0.25 \pm 0.89$	$0.87 \pm 0.92$	$-1.28 \pm 1.04$
	RELFAB	$-0.21 \pm 0.27$	$-0.12 \pm 0.13$	$0.17 \pm 0.35$
	SP4FAA	$-0.01 \pm 0.04$	$-0.02 \pm 0.03$	$-0.02 \pm 0.03$
	SP4FAB	$-0.04 \pm 0.02$	$-0.02 \pm 0.02$	$-0.02 \pm 0.02$
5	RELFAA	$-0.11 \pm 0.57$	$-0.32 \pm 0.15$	$-0.22 \pm 0.12$
	RELFAB	$-0.21 \pm 0.62$	$-0.18 \pm 0.25$	$-0.45 \pm 0.28$
	SP4FAA	$0.12 \pm 0.16$	$0.27 \pm 0.41$	$0.56 \pm 0.43$
	SP4FAB	$2.47 \pm 1.17$	$-0.41 \pm 0.80$	$-1.11 \pm 0.41$
CFA19RELFA	RELFAA	$0.71 \pm 0.60$	$0.27 \pm 0.27$	$0.37 \pm 0.30$
	RELFAB	$-0.68 \pm 0.68$	$0.23 \pm 0.34$	$-0.27 \pm 0.40$
	SP4FAA	$-0.04 \pm 0.06$	$-0.04 \pm 0.06$	$-0.04 \pm 0.07$
	SP4FAB	$-0.05 \pm 0.06$	$-0.06 \pm 0.07$	$-0.06 \pm 0.06$

The absolute value of the regression coefficient is the realized heritability.

however, is paradoxical. On the one hand the very low (nonsignificant) heritability estimates of FA seem to suggest an overwhelmingly environmental origin for the phenotypic variation in FA, which is taken as an argument for using FA as an indicator of DS. FA itself does not have a genetic basis (Palmer, 1994). In fact, as the heritability estimates of DA are not different from those of FA, DA may also be an indicator of DS (Leamy, 1999). On the other hand it is clear that part of the variation in eyespot FA must have been of genetic origin because of the high  $CV_{FA}$  values correlating positively with  $CV_{DI}$  and the results of the Levene's test. There was nevertheless no response to selection. This may be because individual FA values, on which selection was targeted, did not reliably reflect development stability, even when based on several eyespots. This argues against FA as a reliable indicator of DS.

Furthermore, our experimental design assumes that the majority of genetic variance present for DS is purely additive, which may not be the case (Fuller and Houle,

2003). It is quite likely that epistasis and dominance effects play a significant role (reviews in Alibert and Auffray, 2003; Klingenberg, 2003; Leamy, 2003). At least five genes, for example, contribute to eyespot size on the ventral surface of *B. anynana* butterflies and dominance and epistasis effects play a major role (Wijngaarden and Brakefield, 2000).

Houle (2000) has shown that the underlying variance in DS must be (unrealistically) high for FA to reliably estimate DS. Houle (2000) indicates that  $CV_{DI}$  values must lie between 40 and 100 for  $CV_{FA}$  of a trait to reliably reflect developmental variance and the  $R$  to differ significantly from 0. In the present study,  $CV_{FA}$  correlated significantly with  $CV_{DI}$ , with  $CV_{DI}$  values between 11 and 38, so outside the range indicated by Houle (2000). The values of  $CV_{DI}$  reported in this study are consistent with those characteristic of a trait under directional selection (Gangestad and Thornhill, 1999). Our results differ quantitatively from the predictions made under the model presented by Houle (compare

Figure 2 of the present study with Figure 4 of Houle, 2000), but not qualitatively. In both cases FA is a poor estimator of DS. A likely explanation for the observed differences may be that DS is not a single (measurable) trait, but rather an epiphenomenon of the parameters of the developmental system. Variation in the expression of these parameters, themselves under genetic control, may have nonlinear effects on morphology, which could readily lead to non-normal distributions of sizes, and hence cause higher  $CV_{FA}$  values (Klingenberg and Nijhout, 1999; Fuller and Houle, 2003). Klingenberg and Nijhout (1999) developed, for example, a diffusion-threshold developmental model that simulated random noise in a bilateral trait. Although the developmental parameters in the model were set to act additively, the model was nonadditive and generated dominance and epistasis for FA.

The artificial selection in our study did not result in changes in the level and direction of DA of any eyespot and the heritability estimates were not significantly different from those of FA. This finding suggests that DA itself may be a potential indicator of DS (Leamy, 1999). In all eyespots, except the fifth, the area of the white 'pupil' on the left wing was significantly larger than on the right wing. The left wing itself was also significantly larger than the right wing and differently shaped, but this did not explain the DA of the eyespots. Furthermore, it is unclear why the fifth eyespot shows the opposite pattern, especially as the four eyespots seem to share factors affecting the DS. Perhaps there is an interaction between overall wing size or shape and different regions of the wing. It has been argued that when a significant DA exists there must have been a selective advantage for it and that it has a genetic basis (eg, Palmer, 1994). In our study there is no evidence for this. Few studies have been carried out on the heritability of DA, but the reported values are actually comparable to those of FA. Furthermore, selection experiments on DA have always been unsuccessful (Palmer, 1996; Leamy *et al*, 1997, 1998; Auffray *et al*, 1999).

Stressed animals are predicted to show higher FA and a greater variance in values of FA (higher  $CV_{FA}$ ) as the stability of development has been weakened due to the applied stress. Various studies have shown, however, that the relation between FA and stress is usually trait specific (Dufour and Weatherhead, 1996; Leung and Forbes, 1996; Bjorksten *et al*, 2000a, b). Since the FA of any given trait may only poorly estimate the underlying DS, it follows that the likelihood of finding a response to stress in terms of FA may actually be quite small. In our study, a heat stress applied to pupae during the process of eyespot pattern determination did not result in differences between and within the selected and unselected lines in levels of FA, DA and  $CV_{FA}$ . There are several possible explanations. First, perhaps a heat shock of 39°C for 5 h, 1 h after pupation is not a stress, or at least not with respect to eyespot development. However, in another study (Breuker and Brakefield, 2003) we have demonstrated that this treatment results in the increased expression of heat shock proteins. Second, the DS of an individual, whether an epiphenomenon of the parameters of the developmental system or because of whole genome characteristics, may have been sufficient to buffer the adverse effects of the heat shock, thereby averting an increase in FA. Third, once again eyespot FA

may not reliably reflect their DS. We conclude that either FA is unlikely to be a reliable measure of DS or that the majority of genetic variation present for FA, and hence DS, is nonadditive. In either case, we consider that FA gives little, if any, potential for providing insights about the mechanisms of developmental stability.

## Acknowledgements

We thank Niels Wurzer and colleagues for cultivating maize for caterpillars, M Brittijn for his help with the figures, and CP Klingenberg for his advice on the shape analyses. CJ Breuker was funded by the Dutch Organization for Scientific Research, on a NWO Grant: 33.360 – P.

## References

- Alibert P, Auffray JC (2003). Genomic coadaptation, outbreeding depression and developmental instability. In: Polak M (ed) *Developmental Instability: Causes and Consequences*, Oxford University Press: New York. pp 000–000.
- Auffray JC, Debat V, Alibert P (1999). Shape asymmetry and developmental stability. In: Chaplain MAJ, Singh GD, McLachlan JC (eds) *On Growth and Form: Spatio-temporal Pattern Formation in Biology*, Wiley: Chichester. pp 309–324.
- Beldade P, Brakefield PM (2002). The genetics and evo-devo of butterfly wing patterns. *Nat Rev Genet* 3: 442–452.
- Bjorklund M, Merila J (1997). Why some measures of fluctuating asymmetry are so sensitive to measurement error. *Ann Zool Fenn* 34: 133–137.
- Bjorksten TA, Fowler K, Pomiankowski A (2000a). What does sexual trait FA tell us about stress? *TREE* 15: 163–166.
- Bjorksten TA, Fowler K, Pomiankowski A (2000b). Symmetry, size and stress – reply from T Bjorksten, K Fowler and A Pomiankowski. *TREE* 15: 331.
- 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.
- Brakefield PM, El Filali E, Van der Laan R, Breuker CJ, Saccheri IJ, Zwaan B (2001). Effective population size, reproductive success and sperm precedence in the butterfly, *Bicyclus anynana*, in captivity. *J Evol Biol* 14: 148–156.
- Brakefield PM, Gates J, Keys D, Kesbeke F, Wijngaarden PJ, Monteiro A, *et al* (1996). Development, plasticity and evolution of butterfly eyespot patterns. *Nature* 384: 236–242.
- Breuker CJ, Brakefield PM (2003). Heat stress does not increase eyespot fluctuating asymmetry in the tropical butterfly *Bicyclus anynana*. *Evolution and Development* 5: 1–9.
- Dufour KW, Weatherhead PJ (1996). Estimation of organism-wide asymmetry in red-winged blackbirds and its relation to studies of mate selection. *Proc R Soc Lond B* 263: 769–775.
- Falconer DS, MacKay TFC (1996). *Introduction to Quantitative Genetics*, 4th edn. Longman: Burnt Mill, Essex, UK.
- Fowler K, Whitlock MC (1994). Fluctuating asymmetry does not increase with moderate inbreeding in *Drosophila melanogaster*. *Heredity* 73: 373–376.
- Fuller RC, Houle D (2003). Inheritance of developmental stability. In: Polak M (ed) *Developmental Instability: Causes and Consequences*, Oxford University Press: New York. pp 000–000.
- 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, Freeman DC, Emlen JM (1993). Antisymmetry, directional asymmetry, and dynamic morphogenesis. *Genetica* 89: 121–137.
- Houle D (1997). A meta-analysis of the heritability of developmental stability – comment. *J Evol Biol* 10: 17–20.

- Houle D (2000). A simple model of the relationship between asymmetry and developmental stability. *J Evol Biol* **13**: 720–730.
- Hutcheson HJ, Oliver Jr JH (1998). The question of fluctuating asymmetry in the blacklegged tick *Ixodes scapularis* (Acari: Ixodidae). *Exp Appl Acarol* **22**: 51–60.
- Klingenberg CP (2003). A developmental perspective on developmental instability: theory, models and mechanisms. In: Polak M (ed) *Developmental Instability: Causes and Consequences*, Oxford University Press: New York. pp 000–000.
- Klingenberg CP, McIntyre GS (1998). Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with procrustes methods. *Evolution* **52**: 1363–1375.
- Klingenberg CP, Nijhout HF (1999). Genetics of fluctuating asymmetry: a developmental model of developmental instability. *Evolution* **53**: 358–375.
- Klingenberg CP, Badyaev AV, Sowry SM, Beckwith NJ (2001). Inferring developmental modularity from morphological integration: analysis of individual variation and asymmetry in bumblebee wings. *Am Nat* **157**: 11–23.
- Leamy LJ (1997). Is developmental stability heritable? *J Evol Biol* **10**: 21–29.
- Leamy LJ (1999). Heritability of directional and fluctuating asymmetry for mandibular characters in random bred mice. *J Evol Biol* **12**: 146–155.
- Leamy LJ (2003). Dominance, epistasis, and fluctuating asymmetry. In: Polak M (ed) *Developmental Instability: Causes and Consequences*, Oxford University Press: New York. pp 000–000.
- Leamy LJ, Pomp D, Eisen EJ, Cheverud JM (2000). Quantitative trait loci for directional but not fluctuating asymmetry of mandible characters in mice. *Gen Res* **76**: 27–40.
- Leamy LJ, Routman EJ, Cheverud JM (1997). A search for quantitative trait loci affecting asymmetry of mandibular characters in mice. *Evolution* **51**: 957–969.
- Leamy LJ, Routman EJ, Cheverud JM (1998). Quantitative trait loci for fluctuating asymmetry of discrete skeletal characters in mice. *Heredity* **80**: 509–518.
- Lens L, Van Dongen S (1999). Evidence for organism-wide asymmetry in five bird species of a fragmented afro-tropical forest. *Proc R Soc Lond B* **266**: 1055–1060.
- Leung B, Forbes MR (1996). Fluctuating asymmetry in relation to stress and fitness: effects of trait type as revealed by meta-analysis. *Ecoscience* **3**: 400–413.
- Leung B, Forbes MR (1997). Modelling fluctuating asymmetry in relation to stress and fitness. *Oikos* **78**: 397–405.
- Leung B, Forbes MR, Houle D (2000). Fluctuating asymmetry as a bioindicator of stress: comparing efficacy of analyses involving multiple traits. *Am Nat* **155**: 101–115.
- McKenzie JA, Clarke GM (1988). Diazinon resistance, fluctuating asymmetry and fitness in the Australian sheep blowfly, *Lucilia cuprina*. *Genetics* **120**: 213–220.
- McKenzie JA, O'Farrell K (1993). Modification of developmental instability and fitness: malathion-resistance in the Australian sheep blowfly, *Lucilia cuprina*. *Genetica* **89**: 67–76.
- McKenzie JA, Yen JL (1995). Genotype, environment and the asymmetry phenotype. Dieldrin-resistance in *Lucilia cuprina* (the Australian sheep blowfly). *Heredity* **2**: 181–187.
- Merila J, Bjorklund M (1995). Fluctuating asymmetry and measurement error. *Syst Biol* **44**: 97–101.
- Møller AP (1997). Developmental stability and fitness: a review. *Am Nat* **149**: 916–932.
- Møller AP, Swaddle JP (1997). *Asymmetry, Developmental Stability, and Evolution*. Oxford University Press: Oxford.
- Møller AP, Thornhill R (1997). A meta-analysis of the heritability of developmental stability. *J Evol Biol* **10**: 1–16.
- Palmer AR (1994). Fluctuating asymmetry analyses: a primer. In: Markow TA (ed) *Developmental Instability: Its Origins and Evolutionary Implications*, Kluwer Academic Publishers: Dordrecht. pp 355–364.
- Palmer AR (1996). Waltzing with asymmetry. *Bioscience* **46**: 518–532.
- Palmer AR, Strobeck C (1986). Fluctuating asymmetry: measurement, analysis, patterns. *Annu Rev Ecol Syst* **17**: 391–421.
- Palmer AR, 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.
- Palmer AR, Strobeck C (2003). Fluctuating asymmetry analyses revisited. In: Polak M (ed) *Developmental Instability: Causes and Consequences*, Oxford University Press: New York. pp 000–000.
- Polak M, Starmer WT (2001). The quantitative genetics of fluctuating asymmetry. *Evolution* **55**: 498–511.
- Slice D (1998). *Morpheus et al*: software for morphometric research, Department of Ecology and Evolution, State University of New York, Stony Brook, New York.
- Van Dongen S, Molenberghs G, Matthysen E (1999a). The statistical analysis of fluctuating asymmetry: REML estimation of a mixed regression model. *J Evol Biol* **12**: 94–102.
- Van Dongen S, Sprengers E, Lofstedt 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.
- Watson PJ, Thornhill R (1994). Fluctuating asymmetry and sexual selection. *Trends Ecol Evol* **9**: 21–25.
- Whitlock MC (1996). The heritability of fluctuating asymmetry and the genetic control of developmental stability. *Proc R Soc Lond B* **263**: 849–853.
- Whitlock MC (1998). The repeatability of fluctuating asymmetry: a revision and extension. *Proc R Soc Lond B* **265**: 1429–1431.
- Whitlock MC, Fowler K (1997). The instability of studies of instability. *J Evol Biol* **10**: 63–67.
- Wijngaarden PJ, Brakefield PM (2000). The genetic basis of eyespot size in the butterfly *Bicyclus anynana*: an analysis of line crosses. *Heredity* **85**: 471–479.
- Windig JJ (1998). Evolutionary genetics of fluctuating asymmetry in the peacock butterfly (*Inachis io*). *Heredity* **80**: 382–392.
- Windig JJ, Nylin S (1999). Adaptive wing asymmetry in males of the speckled wood butterfly (*Pararge aegeria*)? *Proc R Soc Lond B* **266**: 1413–1418.