

## Fluctuating asymmetry and sexual performance in the drumming wolf spider *Hygrolycosa rubrofasciata*

Jari J. Ahtiainen\*, Rauno V. Alatalo, Johanna Mappes & Laura Vertainen

*University of Jyväskylä, Department of Biological and Environmental Science, P.O. Box 35, FIN-40351 Jyväskylä, Finland (\*e-mail: jjahti@dodo.jyu.fi)*

*Received 21 Oct. 2002, revised version received 25 Nov. 2002, accepted 6 Dec. 2002*

Ahtiainen, J. A., Alatalo, R. V., Mappes, J. & Vertainen, L. 2003: Fluctuating asymmetry and sexual performance in the drumming wolf spider *Hygrolycosa rubrofasciata*. — *Ann. Zool. Fennici* 40: 281–292.

Recently, there has been much interest in estimating fluctuating asymmetry (FA) of morphological traits as a short-cut measure of individual quality. FA deals with small differences around the symmetry value of zero. Thus, measurement error is often relatively large. However, repeated measurements and large sample sizes allow reliable estimates of FA that can be corrected for errors. The purpose of this study was two-fold: at the biological level, we examined with a large sample size ( $N = 804$ ) whether pedipalp FA could be used as a short-cut measure of individual quality in the wolf spider *Hygrolycosa rubrofasciata*. This was done by estimating how strongly FA correlates with male sexual performance, i.e. drumming rate and mobility. At the statistical level, we examined how the deflating effect of measurement error could be eliminated to get unbiased correlations between FA and any repeatable trait. We also examined which is the more economical way to get accurate population-level estimates of FA–trait associations, to increase sample size or the number of within-subject repeats. Our results show that there was a very weak, but significant negative relationship between pedipalp FA and mobility (Kendall's partial  $r_k = -0.086$ ). However, this value inevitably underestimates the true relationship, given the large measurement error. It is possible to estimate the unbiased relationship by correcting the above correlation coefficient with effective reliability estimates of both FA and sexual performance. After the correction, the unbiased relationship between FA and mobility was  $r = -0.183$ . Our results indicate that FA is weakly related to male sexual performance in *H. rubrofasciata*. Our results also show that leptokurtosis, which is characteristic of signed FA distributions, was entirely caused by 17 outlier individuals. This indicates that large sample sizes are important to illustrate the true between-individual heterogeneity in FA. Furthermore, our power analysis indicates that it pays to measure more individuals than to increase the number of within-subject repeats to obtain accurate population-level estimates of FA–trait associations.

## Introduction

Fluctuating asymmetry (FA) of a bilateral trait is defined as subtle, random deviations from genetically controlled trajectory towards optimal phenotype, i.e. perfect symmetry (Van Valen 1962, *see also* Palmer & Strobeck 1986). FA has recently been proposed to indicate individual quality (review in Møller & Swaddle 1997, Møller & Thornhill 1998, Thornhill *et al.* 1999, but *see* Palmer 1999, Simmons *et al.* 1999). This proposal is based on the assumption that FA is a sensitive measure of environmental and genetic stress, and has led to the prediction that FA covaries negatively with condition (Møller 1990, Møller & Pomiankowski 1993, Watson & Thornhill 1994). It also has been predicted that there is condition-dependence in the expression of sexual traits: Individuals in good condition may be better able to bear costs of sexual traits, and therefore may produce larger sexual traits than individuals in poor condition (review in Jennions *et al.* 2001, Kotiaho 2001). Therefore, through condition dependence of both the size of the sexual trait and FA, size may be negatively related to FA at the population level.

The studies dealing with FA and individual quality can be divided into two cases, depending on whether FA is concerned as a sexual signal for choosy females, or as a short-cut stress indicator for researchers. Those studies that concern FA as a sexual signal have shown that females can detect subtle differences in FA of exaggerated male ornaments and use symmetry preference directly in mate choice (e.g. Møller 1992, Swaddle & Cuthill 1994, Bennett *et al.* 1996, Møller 1999, Uetz & Smith 1999). There is also an increasing amount of evidence that FA plays only a correlative role in sexual selection (e.g. Simmons 1995, Møller *et al.* 1996, Badyaev 1998, Badyaev *et al.* 1998, Roldan *et al.* 1998, Rantala *et al.* 2000). This means that FA can be negatively related to the size of the sexual trait, even if FA itself is not directly under mate choice. Also in this way, FA can work as a reliable measure of individual quality, which can be used by researchers as a short-cut measure for monitoring biological stress and fitness.

FA has attracted major attention, because bilateral traits are commonplace in nature. Also,

the measurement of FA appears to represent a fairly simple method of assessing biological stress at the individual and population levels. The precision of measuring subtle deviations from the symmetry value of zero is complicated by the fact that the magnitude of FA frequently approaches measurement error (e.g. Palmer 1994). This means that the measurement error often explains a great deal of the total variance in FA studies (e.g. Brakefield & Breuker 1996, Kokko *et al.* 1996). Mixed-model analysis is necessary to test whether the between-sides variance estimate of FA is confounded by measurement error (Palmer & Strobeck 1986, Van Dongen *et al.* 1999). If FA is not significantly larger than measurement error after the optimization of measurement method, one must make more within-subject repeats (to decrease the random error in individual data points), or increase sample size if possible (to decrease the random error and systematic bias in population-level estimates) (*see* Van Dongen 1999). By using a mixed-model analysis, one can obtain unbiased estimates of the between-sides variance component of FA. The effective reliability method can be used in correlative analyses to rule out the deflating effect of measurement error on observed correlation coefficients between FA and other repeatable traits (Furlow *et al.* 1997, Gangestad & Thornhill 1999). This idea is based on one of the standard methods in meta-analysis (Hedges & Olkin 1985, Rosenthal 1986), which corrects observed correlation coefficients with effective reliability estimates. The effective reliability estimate obtained from replicate measurements deals with the degree to how reliably individual trait can be measured. This method yields unbiased correlation estimates between any repeatable traits.

Male wolf spiders of *Hygrolycosa rubrofasciata* (Araneae: Lycosidae) (Ohlert 1865) have a drumming signal that is used for sexual communication (Kronstedt 1996). During the mating season, males produce drumming signals by hitting their abdomen on dry leaves or other suitable substrate to court females. One courtship drumming consists of ca. 30–40 separate pulses, lasts ca. 1 second (Rivero *et al.* 2000), and is audible to the human ear up to a distance of several meters. Shortly before and during the

emittance of a drumming produced by tapping the abdomen against the substrate, males of *H. rubrofasciata* also perform oscillations of the pedipalps. The tips of pedipalps are in contact with the drumming substrate producing a weak humming sound just prior to drumming and a weak rattling sound at the end. Spider pedipalps are used in food handling, but they also function as sexual organs in males. Male pedipalps do not actually produce the sperm, but, instead, they are used to transfer the sperm to the female, via the epigyne, for storage in the spermathecae prior to fertilization (Roberts 1996). In *H. rubrofasciata*, male courtship drumming has proved to be an honest viability indicator for choosy females (review in Ahtiainen *et al.* 2001). Females prefer more actively drumming males as mating partners (Kotiaho *et al.* 1996). There is considerable within-male repeatability and among-male variability in drumming rates (Kotiaho *et al.* 1996). Male drumming incurs both physiological (Mappes *et al.* 1996, Kotiaho *et al.* 1998a) and predation costs (Kotiaho *et al.* 1998b). The more actively drumming males have better viability (Kotiaho *et al.* 1996, Mappes *et al.* 1996, Kotiaho *et al.* 1999). Survival costs of male drumming are condition-dependent, being manifested in decreased viability of males in poorer condition (Mappes *et al.* 1996, Kotiaho 2000). By choosing males with the highest drumming rates females benefit through better offspring survival (Alatalo *et al.* 1998). Also, male mobility is positively associated with male mating success (Kotiaho *et al.* 1998b). Therefore, also mobility reflects male quality in *H. rubrofasciata*.

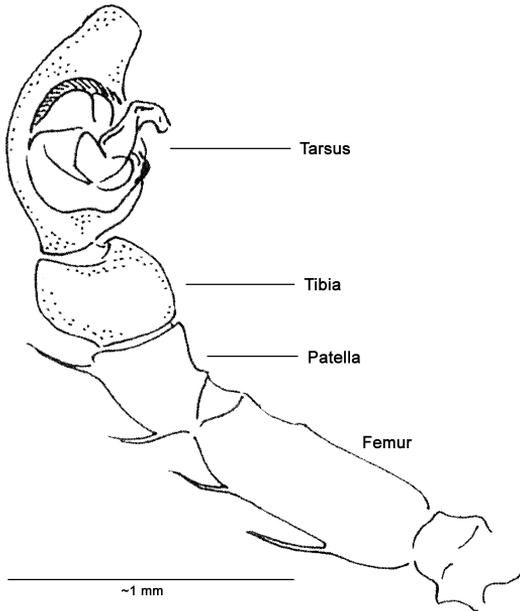
The purpose of this study is two-fold: at the biological level, we examine with a large sample size whether pedipalp FA could be used as a short-cut measure of individual quality in the wolf spider *H. rubrofasciata*. This is done by estimating how strongly FA correlates with male sexual performance, i.e. drumming rate and mobility. We acknowledge that increasing the number of traits is the best way to maximize the probability of detecting FA–fitness relations when they exist (*see* Leung *et al.* 2000). Our preliminary study showed that pedipalp asymmetry was the only trait from several bilateral traits measured in the wolf spider *H. rubrofasciata* that had adequate measurement repeatability

(unpublished data). Therefore, we measure only pedipalp FA in this study. At the statistical level, we examine how the deflating effect of measurement error could be eliminated to get unbiased correlations between FA and any repeatable trait. We also examine which is the more economical way to get accurate population-level estimates of FA–trait associations, to increase sample size or the number of within-subject repeats. This is done by applying the effective reliability method in the power analysis.

## Material and methods

Male wolf spiders (*Hygrolycosa rubrofasciata*) were collected using pitfall traps from 36 populations in Sipoo, Southern Finland (60°16'N, 25°14'E) at the peak of the mating season. Spiders were individually placed in small plastic film jars with some moss (*Sphagnum* spp.). They were kept in cool temperature (ca. +10 °C). In the laboratory, males were weighed to the nearest 0.1 mg with an analytical balance (Sauter AR 1014). Body masses were measured only once, because Kotiaho *et al.* (1996) have shown that individual body masses do not significantly change during the laboratory period under similar conditions ( $R = 0.994$ ). After body mass measurements, spiders were individually kept in plastic containers (diameter 40 mm) filled with some moss (*Sphagnum* spp.) at  $5 \pm 2$  °C in darkness to keep their activity levels low. Food (*Drosophila melanogaster* fruit flies) was provided *ad libitum*. These containers were moistened regularly.

For drumming rate and mobility measurements, we took each male randomly from the sample of all specimens, and placed them individually in plastic arenas (125 mm × 88 mm × 110 mm high). The bottom of plastic arenas was covered either with white paper (8 cm × 4 cm) and two dry even-sized birch leaves on the paper or with white paper (8 cm × 4 cm) without any leaves. After each trial day, papers were cleaned to avoid the possible confounding effects of pheromones. Drumming rate of each individual was measured twice, first with leaves followed by the second trial day without leaves. Since males of *H. rubrofasciata* have shown to exhibit moderate



**Fig. 1.** The structural diagram of the pedipalp measured in male wolf spider *Hygrolycosa rubrofasciata* (Lycosidae). Each bilateral pedipalp consists of six segments (beginning from the body): coxa, trochanter, femur, patella, tibia, and tarsus (see Roberts 1996). The length between femur and tarsus was measured from each pedipalp, as shown in this figure (with permission of Ms Katja Löytynoja).

repeatability in their mobility across trial days ( $R = 0.46$ ; Kotiaho *et al.* 1999), mobility was measured only once without leaves. Leaves were excluded from arenas during the first behavioral experiment (drumming rate + mobility) to monitor the mobility of spiders more accurately. Leaves were put onto arenas during the second behavioral experiment (drumming rate) in order to mimic the natural environment of *H. rubrofasciata*. For the mobility experiment, each arena was divided with a line into two equal rectangles. The laboratory was illuminated with fluorescent tubes and lamps with 40 W bulbs placed 30 cm above the floors of drumming arenas to give extra heat and light. On the day before the beginning of the behavioral measurements, males were kept in the laboratory temperature ( $31 \pm 1$  °C) for two hours to trigger their sexual activity. Drumming rate was measured as the number of separate drumming bouts, and mobility as the number of times the male crossed a line between the rectangles. Drumming rate and mobility

were measured five times for 2 min during the trial day. Between the trial days, males were fed with fruit flies (*D. melanogaster*) and kept in moistened containers filled with some moss (*Sphagnum* spp.) in cool temperature ( $+5 \pm 2$  °C) in darkness. After behavioral measurements, males were preserved individually in 70% ethanol. Because the distributions of drumming rate and mobility could not be normalized, we used ranked values in our data analysis. The repeatability for drumming rate across trial days was small, i.e.  $R = 0.269$  (ANOVA with ranked values:  $F_{809,810} = 1.736$ ,  $P < 0.001$ ) (for methods see Krebs 1989, Zar 1996). There was a moderate positive relationship between drumming rate and mobility ( $r_s = 0.450$ ,  $N = 812$ ,  $P < 0.001$ ). Because of this correlation, we used Kendall's partial correlations between FA and sexual performance.

Each pedipalp consists of six segments (beginning from the body): coxa, trochanter, femur, patella, tibia, and tarsus. Pedipalps were cut off from the cephalothorax (i.e. body) under a binocular microscope. After that, they were straightened between two object glasses. We used a stereomicroscope (Wild M5A) equipped with an ocular micrometer to measure the length between femur and tarsus to the nearest 0.01 mm from each pedipalp (Fig. 1). According to our preliminary study, this measure was the most repeatable of several bilateral traits measured in *H. rubrofasciata* (J. J. Ahtiainen, R. V. Alatalo & J. Mappes unpubl.). After each measurement, pedipalps were individually preserved in 70% ethanol. One of the authors (J. J. Ahtiainen) made all the measurements to eliminate the between-observer variation in measurements (see Hubert & Alexander 1995). A few individuals which clearly had injuries or deformities in either pedipalp were not measured (see Cuthill *et al.* 1993, Møller 1993). From all individuals each pedipalp was measured twice to calculate repeatabilities of its asymmetry and length. The repetition procedure was as follows: First, the left-hand side pedipalps (*L*) were measured once, followed by the measurements of the right-hand side pedipalps (*R*) in a random order without reference to the measurements of the left-hand side pedipalps (see Palmer 1994). Remeasuring the same individuals was

done with exactly the same procedure, and in a random order without reference to the first measurements. The mounting of pedipalps is likely to be a considerable source of measurement error. As each pedipalp had to be mounted twice, our repetition procedure gave an accurate estimate of the measurement error variance. Pedipalp asymmetry was defined as the mean length difference between left- and right-hand side pedipalps. Because the distribution of unsigned pedipalp asymmetry could not be normalized, we used ranked values in our data analysis. The repeatability for unsigned pedipalp asymmetry was moderate, i.e.  $R = 0.314$  (ANOVA with ranked values:  $F_{803,804} = 1.916$ ,  $P < 0.001$ ). The repeatability for pedipalp length was very high, i.e.  $R = 0.982$  ( $F_{803,804} = 113.131$ ,  $P < 0.001$ ). There was a strong positive relationship between pedipalp length and body mass ( $r = 0.828$ ,  $N = 782$ ,  $P < 0.001$ ). Because of this correlation, we used Kendall's partial correlations between FA and body size. The inclusion of body size as a covariate did not affect the relationship between FA and sexual performance.

It has been shown that the admixture of FA with antisymmetry results in leptokurtic or platykurtic distributions with zero mean, while the admixture of FA with directional asymmetry skews the distribution of the signed asymmetry (Palmer & Strobeck 1992, Rowe *et al.* 1997, Van Dongen 1998a). As antisymmetry and directional asymmetry can have a genetic basis, those asymmetries do not necessarily reflect developmental stability (Palmer & Strobeck 1992, Palmer 1994, but *see* Graham *et al.* 1993). FA and directional asymmetry can be separated statistically by a two-way mixed model ANOVA (Palmer & Strobeck 1986). Whereas FA and antisymmetry cannot be separated statistically with high power, they can be distinguished reliably through inspection of the relationship between trait size and unsigned FA (Rowe *et al.* 1997). According to the method described by Rowe *et al.* (1997), we divided data into two subsets using the midpoint of the range of pedipalp length as a boundary. We calculated the fifth percentile as a measure of the lower-bound of the distribution for both subsets. The fifth percentile of the right side subset was subtracted from that of the left side subset to estimate the difference in lower

bounds. To be detected, antisymmetry must be so pronounced that the lower bound of the unsigned asymmetry distribution departs from zero as asymmetry increases. Also, the lower bound of the unsigned asymmetry distribution must depart from zero near the midpoint of the range of pedipalp length so that the lower bound of the left-hand subset will be greater than that of the right-hand subset. Rowe *et al.* (1997) have found this method to be more powerful than other antisymmetry tests.

It is possible to estimate unbiased correlation coefficients by correcting observed correlation coefficients with effective reliability estimates. The unbiased Pearson's correlation coefficient between any two repeatable variables  $x$  and  $y$  is (Hedges & Olkin 1985):

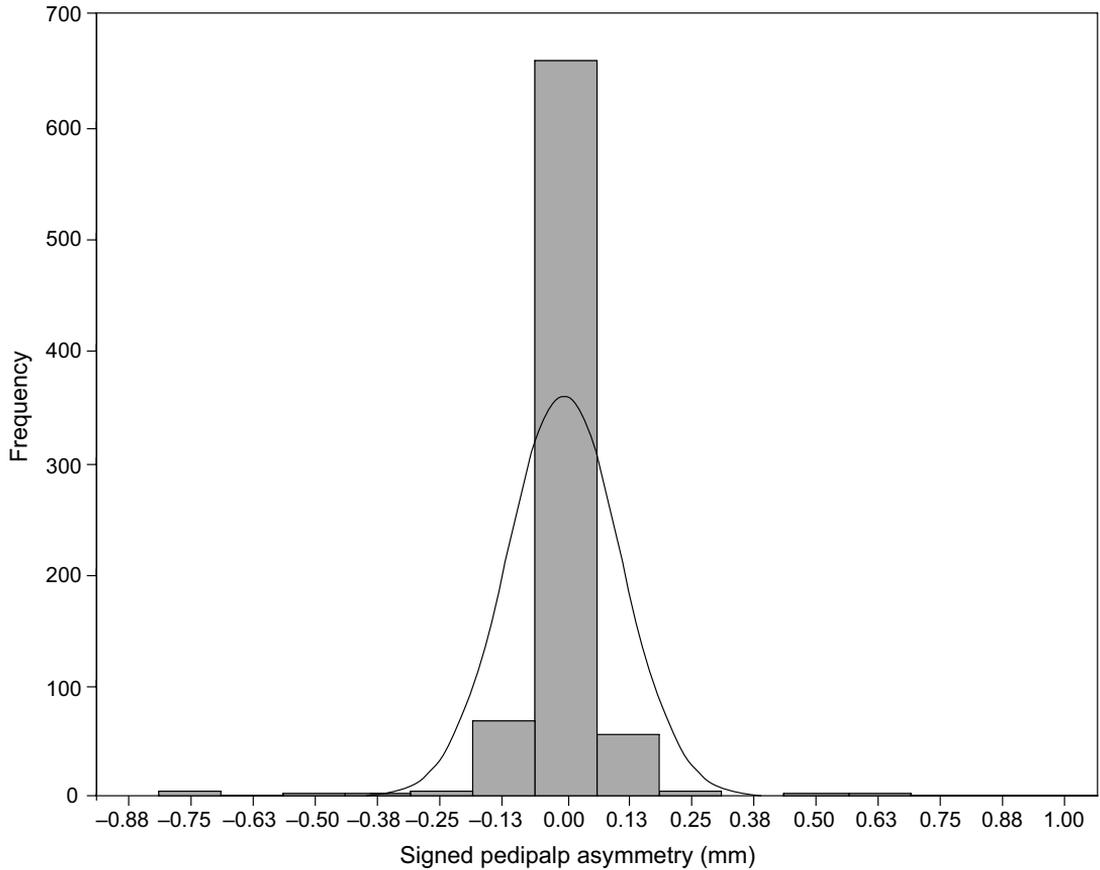
$$r = \frac{r_{xy}}{\sqrt{ER_x \times ER_y}}, \quad (1)$$

where  $r_{xy}$  = the observed correlation coefficient deflated by measurement error (*see* Rosenthal 1986), and  $ER_{x,y}$  = effective reliability (*see* below). Statistical significance of associations is not affected by corrections, because the above equation only corrects the deflating bias.

The repeatability of a trait measures the reliability with which it can be measured (Krebs 1989). It is the relationship between two randomly chosen replicate measurements of all  $J$  independent replicate measurements. Thus, it is not an appropriate estimate of the reliability of any trait that is based on several within-subject repeats. Obviously, the reliability of a trait increases with an increasing number of within-subject repeats. The effective reliability ER is developed to quantify the true reliability of a trait over all  $J$  replicate measurements. It can be derived by the following equation (Rosenthal 1986):

$$ER = \frac{WSR \times R}{1 + (WSR - 1) \times R}, \quad (2)$$

where WSR = the number of within-subject repeats, and  $R$  = the repeatability of a trait, which is the intraclass correlation coefficient equal to the among-groups variance component divided by the sum of the within-group and among-groups variance components (Sokal & Rohlf 1981, Lessells & Boag 1987).



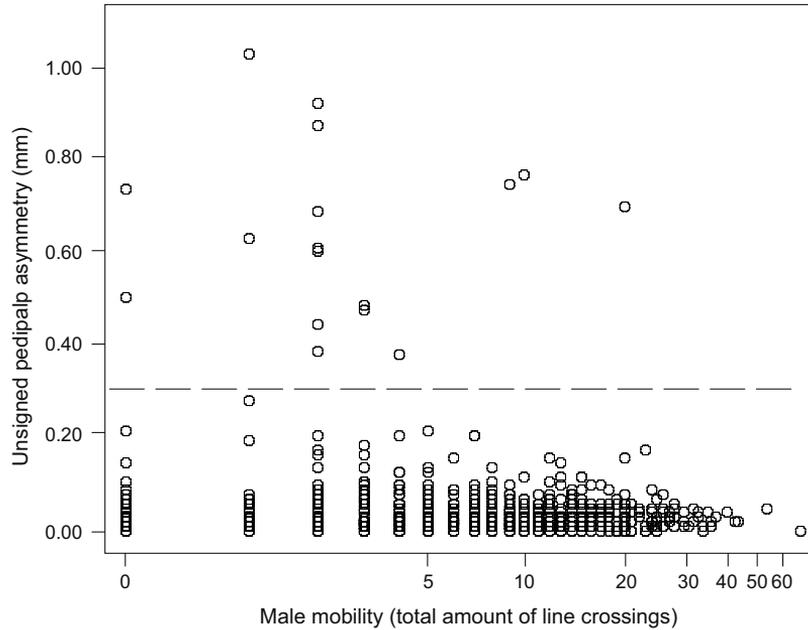
**Fig. 2.** The frequency distribution of signed pedipalp FA of male wolf spiders *Hygrolycosa rubrofasciata* ( $N = 804$ ). The histogram is not even close to a normal distribution, even if symmetric around zero ( $-0.0047 \pm 0.0039$  mm (mean  $\pm$  SE), skew  $g_1 = 0.14 \pm 0.086$  (SE), and kurtosis  $g_2 = 35.49 \pm 0.17$  (SE)).

To answer the statistical questions in the Introduction, we made two power analyses from the effective reliability method. In the first power analysis, we looked at the relationship between required sample size and the number of within-subject repeats that gives 80% power of attaining the 0.05 significance level at two different levels of true effect sizes (see Cohen 1988). The number of within-subject repeats  $WSR_{x,y}$  used was 1, 2, 3, 4, 5, 10, and 20, effect sizes were 0.20 and 0.30, and the repeatability of  $R_{x,y} = 0.50$  for both traits (i.e. FA and sexual performance). Effect sizes used in our power analysis closely match with overall effect sizes between single trait's FA and fitness computed in two broad meta-analyses (Leung & Forbes 1996, Møller & Thornhill 1998). First, we calculated effective reliabilities  $ER_x$  and  $ER_y$  at each level of effect

sizes by varying the number of within-subject repeats (Eq. 2). Then, observed correlation coefficients  $r_{xy}$  were calculated by standardizing unbiased correlation coefficients  $r$  as effect sizes (Eq. 1). Sample sizes needed for 80% power of attaining the 0.05 significance level at the given effect size were calculated from observed correlation coefficients with the power analysis program GPOWER (Faul & Erdfelder 1992). In the second power analysis, we looked at the relationship between required number of measurements ( $N \times WSR$ ) and observed correlation coefficient that gives 80% power of attaining the 0.05 significance level at the true effect size of 0.20, and the repeatability of  $R_{x,y} = 0.50$  for both traits by using the same equations 1 and 2.

In the Results, all significance levels are for two-tailed tests. Because of the few missing

**Fig. 3.** The relationship between unsigned pedipalp FA (mm) and mobility (total sum of line crossings per individual; log-scaled) in male wolf spiders *Hygrolycosa rubrofasciata* (Kendall's partial  $r_k = -0.086$  (controlling for drumming rate),  $N = 755$ ,  $P = 0.018$ ). There were 17 specimens that were classified as extreme outliers ( $|R - L| \geq 0.30$  mm) indicated by the reference line.



values, sample sizes are not equal. As populations did not differ significantly in the mean signed pedipalp asymmetry (Kruskal-Wallis test,  $\chi^2 = 45.151$ , d.f. = 35,  $P = 0.117$ ), or in the mean unsigned pedipalp asymmetry (Kruskal-Wallis test,  $\chi^2 = 46.300$ , d.f. = 35,  $P = 0.096$ ), the data was pooled over all 36 populations.

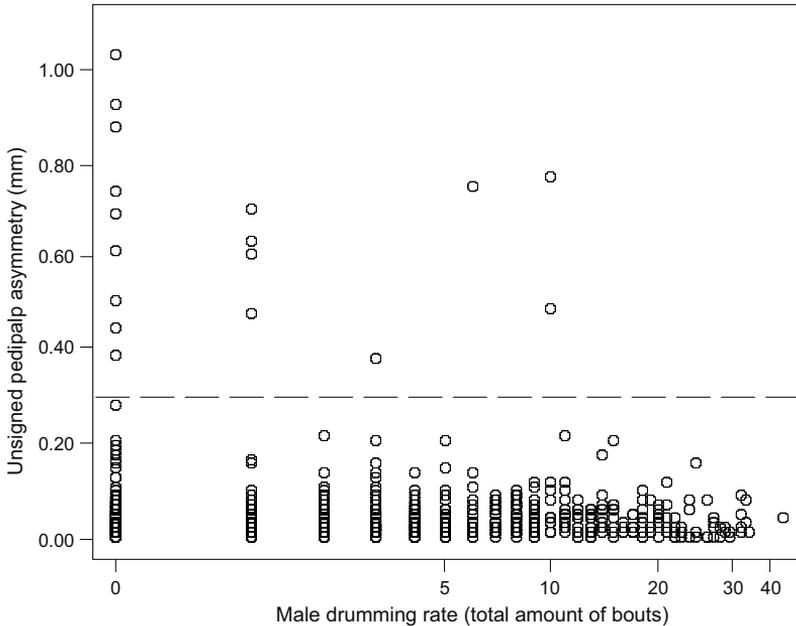
## Results

We tested whether signed FA ( $R - L$ ) of pedipalps was normally distributed in the wolf spider males *Hygrolycosa rubrofasciata*. The frequency distribution of signed FA (Fig. 2) was not even close to a normal distribution (Kolmogorov-Smirnov test,  $Z = 6.078$ , d.f. = 802,  $P < 0.001$ ). This result, together with skewness ( $g_1 = 0.14 \pm 0.086$  (SE)) and extreme kurtosis ( $g_2 = 35.49 \pm 0.17$  (SE)) values, indicated that the histogram of signed FA was very leptokurtic, even if symmetric around zero. The mean and standard error for signed FA, unsigned FA, and pedipalp length were  $-0.0047 \pm 0.0039$  mm,  $0.051 \pm 0.0035$  mm, and  $2.72 \pm 0.0050$  mm, respectively.

In order to test whether there was any true FA separable from measurement error, the ranked pedipalp asymmetry data was tested with the two-way mixed model ANOVA (Palmer &

Strobeck 1986, Zar 1996). The true between-sides variance, i.e. the FA variance, differed significantly from the measurement error variance (ranked values:  $F_{803,1608} = 4.703$ ,  $P = 0.03$ ), indicating that FA can be separated out from measurement error, even if measurement error is large (54% of the mean FA). There was no indication of directional asymmetry in the pedipalp asymmetry distribution (ranked values:  $F_{1,803} = 1.728$ ,  $P = 0.189$ ). As the lower bounds for both subsets were zero (0.00 mm), there was no antisymmetry in the pedipalp asymmetry distribution.

The outlier has to be more than 3 interquartile ranges from the 75% upper quartile to be statistically classified as the extreme outlier (Moore & McCabe 1989). The extreme outlier criterion for unsigned FA ( $|R - L|$ ) of pedipalps was  $\geq 0.30$  mm. There were 17 specimens (2.1% of the FA data) that were classified as extreme outliers (see Figs. 3 and 4). Outliers differed highly significantly from other specimens with respect to drumming rate and mobility (for drumming rate: outliers  $0.097 \pm 0.041$  drumming bouts per min; data without outliers  $0.31 \pm 0.013$  drumming bouts per min (mean  $\pm$  SE); Mann-Whitney  $U$ -test,  $z = -3.025$ ,  $N_1 = 17$ ,  $N_2 = 742$ ,  $P = 0.002$ ) (for mobility: outliers  $0.38 \pm 0.12$  line crossings per min; data without outliers  $0.90 \pm 0.031$  line



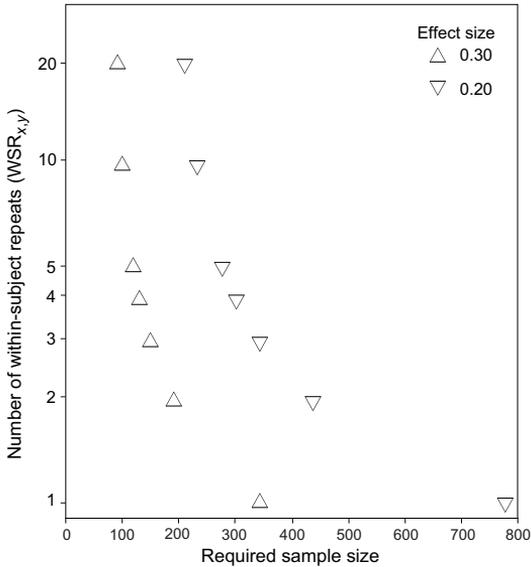
**Fig. 4.** The relationship between unsigned pedipalp FA (mm) and drumming rate (total sum of drums per individual; log-scaled) in male wolf spiders *Hygrolycosa rubrofasciata* (Kendall's partial  $r_k = 0.022$  (controlling for mobility),  $N = 755$ ,  $P = 0.542$ ). There were 17 specimens that were classified as extreme outliers ( $IR - LI \geq 0.30$  mm) indicated by the reference line.

crossings per min (mean  $\pm$  SE); Mann-Whitney  $U$ -test,  $z = -3.098$ ,  $N_1 = 17$ ,  $N_2 = 743$ ,  $P = 0.002$ ). Therefore, the outlier individuals can be regarded as biologically meaningful outliers with a very poor sexual performance. After the removal of outliers, the kurtosis value of signed FA was substantially reduced (kurtosis with outliers:  $35.49 \pm 0.17$  (SE); without outliers:  $3.42 \pm 0.17$  (SE)).

There was a very weak, but significant negative relationship between unsigned FA and mobility (Kendall's partial  $r_k = -0.086$  (controlling for drumming rate),  $N = 755$ ,  $P = 0.018$ ; Fig. 3). The unbiased correlation coefficient free of measurement error can be calculated by correcting the observed correlation coefficient with effective reliability estimates of variables which form that relationship (see Material and methods). The effective reliability estimate for unsigned FA with the ranked data was  $ER = 0.478$  (calculated from  $R = 0.314$ ). As we did not repeat the mobility experiment at this time, we used  $R = 0.46$  as the best effective reliability estimate for this mobility data. The unbiased correlation coefficient between unsigned FA and mobility was  $r = -0.086 / (0.478 \times 0.46)^{0.5} = -0.183$ . There was no significant relationship between unsigned FA and drumming rate (Kendall's partial  $r_k = 0.022$  (controlling for mobility),  $N = 755$ ,  $P = 0.542$ ;

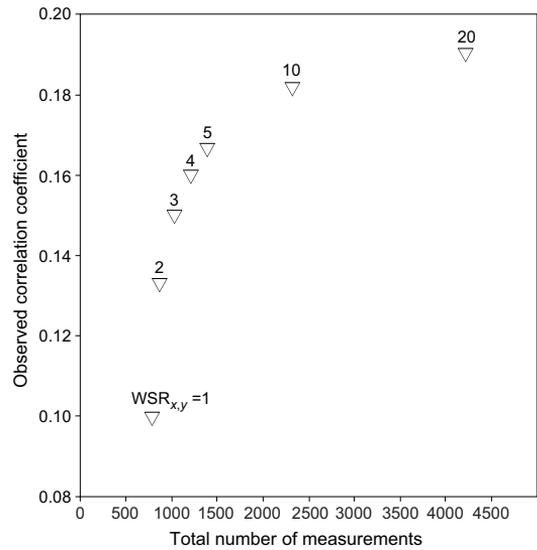
Fig. 4). There was a weak, but significant negative relationship between unsigned FA and pedipalp length (Kendall's partial  $r_k = -0.145$  (controlling for body mass),  $N = 779$ ,  $P < 0.001$ ). There was a weak, but significant positive relationship between unsigned FA and body mass (Kendall's partial  $r_k = 0.114$  (controlling for pedipalp length),  $N = 779$ ,  $P < 0.001$ ).

In the first power analysis, increasing the number of within-subject repeats decreased sample sizes needed for the true effect sizes at a decreasing rate, given 80% power of attaining the 0.05 significance level (Fig. 5). Required sample sizes decreased 44% from  $WSR_{x,y} = 1$  (effect size = 0.20:  $N = 779$ ; effect size = 0.30:  $N = 343$ ) to  $WSR_{x,y} = 2$  (effect size = 0.20:  $N = 436$ ; effect size = 0.30:  $N = 191$ ) in both effect size classes, while required sample sizes decreased 21% from  $WSR_{x,y} = 2$  to  $WSR_{x,y} = 3$  (effect size = 0.20:  $N = 343$ ; effect size = 0.30:  $N = 150$ ) in both effect size classes. From  $WSR_{x,y} = 10$  (effect size = 0.20:  $N = 232$ ; effect size = 0.30:  $N = 100$ ) to  $WSR_{x,y} = 20$  (effect size = 0.20:  $N = 211$ ; effect size = 0.30:  $N = 91$ ), required sample sizes decreased only 9% in both effect size classes. In the second power analysis, the total number of measurements ( $N \times WSR$ ) needed for 80% power of attaining the 0.05 significance level at the true effect size of 0.20 was plotted against



**Fig. 5.** The relationship between required sample size and the number of within-subject repeats (WSR) that gives 80% power of attaining the 0.05 significance level at two different levels of true effect sizes. The number of within-subject repeats is 1, 2, 3, 4, 5, 10, and 20, effect sizes are 0.20 and 0.30, and the repeatability of  $R_{x,y} = 0.50$  for both traits (i.e. FA and sexual performance).

observed correlation coefficients (Fig. 6). As the number of within-subject repeats increased, the increase in observed correlation coefficients decreased approaching to the asymptotic value of 0.20. In other words, the observed correlation coefficient  $r$  increased 33% from  $WSR_{x,y} = 1$  ( $N = 779$ ;  $r = 0.100$ ) to  $WSR_{x,y} = 2$  ( $N = 436$ ;  $r = 0.133$ ), while  $r$  increased 13% from  $WSR_{x,y} = 2$  to  $WSR_{x,y} = 3$  ( $N = 343$ ;  $r = 0.150$ ). From  $WSR_{x,y} = 10$  ( $N = 232$ ;  $r = 0.182$ ) to  $WSR_{x,y} = 20$  ( $N = 211$ ;  $r = 0.190$ ),  $r$  increased only 4%. At the same time, required number of measurements increased in a following fashion ( $WSR_{x,y} = 1$ :  $779 \times 1 = 779$  measurements;  $WSR_{x,y} = 2$ :  $436 \times 2 = 872$  measurements;  $WSR_{x,y} = 3$ :  $343 \times 3 = 1029$  measurements;  $WSR_{x,y} = 10$ :  $232 \times 10 = 2320$  measurements;  $WSR_{x,y} = 20$ :  $211 \times 20 = 4220$  measurements). Thus, even if the requirement for the number of individuals is reduced with increasing the number of within-subject repeats, the total number of measurements needed for 80% power of attaining the 0.05 significance level increases greatly.



**Fig. 6.** The relationship between the total number of measurements ( $N \times WSR$ ) and observed correlation coefficient that gives 80% power of attaining the 0.05 significance level at the true effect size of 0.20, and the repeatability of  $R_{x,y} = 0.50$  for both traits.

## Discussion

In this study, we examined with a large sample size ( $N = 804$ ) whether pedipalp FA can be used as a short-cut measure of individual quality in male wolf spiders (*Hygrolycosa rubrofasciata*). This was done by estimating how strongly FA correlates with two measures of male quality, i.e. drumming rate and mobility. In *H. rubrofasciata*, male courtship drumming has demonstrated to be an honest indicator of heritable viability (review in Ahtiainen *et al.* 2001). Mobility is also positively associated with male mating success (Kotiaho *et al.* 1998b). Our results showed only a very weak negative relationship between pedipalp FA and mobility (Kendall's partial  $r_k = -0.086$ ; Fig. 3). Because pedipalp FA explains only 0.7% of the total variance in mobility, it is an unreliable measure of individual quality in *H. rubrofasciata*. Our estimate of  $r_k = -0.086$  is much smaller than the weighted mean effect size of  $r = -0.24$  between single trait's FA and male mating success (in nonhuman species excluding all sexually selected traits, experimental studies and composite scores; total estimates = 54) computed by Møller and Thornhill (1998) in

their meta-analysis (*see* also Palmer 1999), and the overall effect size of  $r = -0.26$  between FA and several fitness measures computed independently by Leung and Forbes (1996). However, because the deflating effect of measurement error has not been ruled out from those overall effect sizes, the unbiased estimate of the overall effect size has to be larger than the estimates above. On the other hand, the observed patterns of paradigm shift in studies concerning the relationship between FA and sexual selection (Simmons *et al.* 1999) suggest caution in correcting observed correlation coefficients, as published estimates might already be biased upward through the non-random sampling from the 'population' of all estimated correlations.

Our second goal was to find out the unbiased relationship between FA and sexual performance, which would be free of the large measurement error (54% of the mean FA). We used the effective reliability method to rule out the deflating effect of measurement error from FA-trait associations. Indeed, it is possible to calculate the unbiased Pearson's correlation coefficient, which is equivalent to the observed correlation coefficient divided by the square root of the product of the effective reliability estimates (Hedges & Olkin 1985, Rosenthal 1986). After the correction, pedipalp FA was still weakly related to male sexual performance ( $r = -0.183$  for FA and mobility). In literature, there are a few studies that have used corrected correlation coefficients to evaluate unbiased relationships between FA and several fitness measures of human men (reviewed by Gangestad & Thornhill 1999). Those estimates generally imply stronger relationship between FA and individual quality than this study.

Because of a variety of reasons (*see* Van Dongen 1998b, Whitlock 1998), heritability estimates of developmental instability, as well as correlations of developmental instability with other traits are expected to bias downward. The hypothetical repeatability can be used to correct for these biases in estimates of developmental instability, as it deals with the degree to which individuals with different levels of developmental instability differ in FA (Van Dongen 1998b, Whitlock 1998). The hypothetical repeatability refers to the proportion of variation in individual

FA that is due to the between-individual variation in developmental instability. To avoid confusion, we emphasize that our study used the effective reliability method that deals with the degree to how reliably two traits in a given correlation can be measured (Hedges & Olkin 1985, Rosenthal 1986). Effective reliability refers to the proportion of the total variation in a trait that is due to the true individual variation in the expression of a trait. This method yields unbiased correlation estimates between any two repeatable traits, both of which are corrected for measurement error (*see* Furlow *et al.* 1997, Gangestad & Thornhill 1999). Our results indicate that there can be considerable biases in observed correlation coefficients due to the large measurement error. Therefore, future studies should also provide unbiased estimates of FA-trait associations, provided that sample sizes are large enough for statistical significance.

Extreme outliers of FA were individuals with poor sexual performance. This result did not change qualitatively with the outlier criterion, i.e. when we varied the outlier criterion within the range of 0.2–0.4 mm, the outlier individuals were always significantly poorer in terms of sexual performance. Because the relative effect of measurement error on individual FA measurements decreases with an increase in trait FA, outliers were the specimens that could be measured most reliably in our study. These extreme outliers can be considered as biologically meaningful, suggesting that FA can be used as an indicator of phenodeviants (*see* Clarke 1993). Our results also showed that the kurtosis value of signed FA varied strongly depending on the inclusion of outliers. This indicates that true outliers are crucial in terms of increasing between-individual variability and leptokurtosis of FA. Because outliers, by definition, constitute only a small proportion of the overall data (2.1% in this case), large sample sizes are important in order to obtain enough between-individual variation in FA (*see* also Gangestad & Thornhill 1999).

Our power analysis showed that there is a non-linear sampling trade-off between required sample size and the number of within-subject repeats to achieve a given effect size (Fig. 5). Our results illustrated that increasing the number of within-subject repeats decrease required sample

sizes for the unbiased effect size at a decreasing rate. Given sample size and the number of within-subject repeats are large enough to separate true FA from measurement error, this indicates that it pays to measure more individuals than to increase the number of within-subject repeats to obtain accurate population-level estimates of FA-trait associations. Our power analysis also showed that there is a logarithmic increase in the required number of measurements when approaching the true effect size (Fig. 6). This result indicates that increasing sample size and within-subject repeats increase the possibility to estimate the true level of FA-trait associations (*see also Van Dongen 1999*), but this is the most effective in low number of within-subject repeats/large sample sizes. In fact, the deflating bias in observed correlation coefficients between FA and other repeatable traits can be easily removed by the effective reliability method.

## Acknowledgements

We want to thank the people collecting the spiders in the field and helping us in the laboratory, especially Ismo Eriksson, Simo Heikkilä, Eira Ihalainen, Antti Kanninen, Virpi Kortelainen, Helinä Nisu, Mari Saarinen, Jaana Suutari, Tapio Tikkanen, Jaana Tissari, Katja Tynkkynen and Saara Vertainen. We thank Dr. Janne S. Kotiaho, Dr. Sami Merilaita, Dr. Silja Parri, Dr. Joseph L. Tomkins and anonymous referees for valuable comments. The work was financed by the Academy of Finland to R.V.A. and J.M.

## References

- Ahtiainen, J. J., Alatalo, R. V., Kotiaho, J. S., Mappes, J., Parri, S. & Vertainen, L. 2001: Sexual selection in the drumming wolf spider *Hygrolycosa rubrofasciata*. — In: Toft, S. & Scharff, N. (eds.), *European Arachnology 2000*: 129–137. Aarhus University Press.
- Alatalo, R. V., Kotiaho, J., Mappes, J. & Parri, S. 1998: Mate choice for offspring performance: major benefits or minor costs? — *Proc. R. Soc. Lond. B* 265: 2297–2301.
- Badyaev, A. V. 1998: Environmental stress and developmental stability in dentition of the Yellowstone grizzly bears. — *Behav. Ecol.* 9: 339–344.
- Badyaev, A. V., Etges, W. J., Faust, J. D. & Martin, T. E. 1998: Fitness correlates of spur length and spur asymmetry in male wild turkeys. — *J. Anim. Ecol.* 67: 845–852.
- Bennett, A. T. D., Cuthill, I. C., Partridge, J. C. & Maier, E. 1996: Ultraviolet vision and mate choice in zebra finches. — *Nature* 380: 433–435.
- Brakefield, P. M. & Breuker, C. J. 1996: The genetical basis of fluctuating asymmetry for developmentally integrated traits in a butterfly eyespot pattern. — *Proc. R. Soc. Lond. B* 263: 1557–1563.
- Clarke, G. M. 1993: Patterns of developmental stability of *Chrysopa perla* L. (Neuroptera: Chrysopidae) in response to environmental pollution. — *Environ. Entomol.* 22: 1362–1366.
- Cohen, J. 1988: *Statistical power analysis for the behavioral sciences*. — Academic Press, New York.
- Cuthill, I. C., Swaddle, J. P. & Witter, M. S. 1993: Fluctuating asymmetry. — *Nature* 363: 217–218.
- Faul, F. & Erdfelder, E. 1992: GPOWER: *A priori, post-hoc, and compromise power analyses for MS-DOS* [Computer program]. — Dept. of Psychology, Bonn University.
- Furlow, B., Armijo-Prewitt, T., Gangestad, S. W. & Thornhill, R. 1997: Fluctuating asymmetry and psychometric intelligence. — *Proc. R. Soc. Lond. B* 264: 823–829.
- Gangestad, S. W. & Thornhill, R. 1999: Individual differences in developmental precision and fluctuating asymmetry: a model and its implications. — *J. Evol. Biol.* 12: 402–416.
- Graham, J. H., Freeman, D. C. & Emlen, J. M. 1993: Antisymmetry, directional asymmetry, and dynamic morphogenesis. — *Genetica* 89: 121–137.
- Hedges, L. V. & Olkin, I. 1985: *Statistical methods for meta-analysis*. — Academic press, Orlando.
- Hubert, W. A. & Alexander, C. B. 1995: Observer variation in counts of meristic traits affects fluctuating asymmetry. — *N. Am. J. Fish. Manage.* 15: 156–158.
- Jennions, M. D., Møller, A. P. & Petrie, M. 2001: Sexually selected traits and adult survival: A meta-analysis. — *Q. Rev. Biol.* 76: 3–36.
- Kokko, E. G., Floate, K. D., Colwell, D. D. & Lee, B. 1996: Measurement of fluctuating asymmetry in insect wings using image analysis. — *Ann. Entomol. Soc. Amer.* 89: 398–404.
- Kotiaho, J. S. 2000: Testing the assumptions of conditional handicap theory: costs and condition dependence of a sexually selected trait. — *Behav. Ecol. Sociobiol.* 48: 188–194.
- Kotiaho, J. S. 2001: Costs of sexual traits: a mismatch between theoretical considerations and empirical evidence. — *Biol. Rev.* 76: 365–376.
- Kotiaho, J., Alatalo, R. V., Mappes, J., Nielsen, M. G., Parri, S. & Rivero, A. 1998a: Energetic costs of size and sexual signaling in a wolf spider. — *Proc. R. Soc. Lond. B* 265: 2203–2209.
- Kotiaho, J., Alatalo, R. V., Mappes, J. & Parri, S. 1996: Sexual selection in a wolf spider: male drumming activity, body size and viability. — *Evolution* 50: 1977–1981.
- Kotiaho, J., Alatalo, R. V., Mappes, J. & Parri, S. 1999: Sexual signalling and viability in a wolf spider (*Hygrolycosa rubrofasciata*): measurements under laboratory and field conditions. — *Behav. Ecol. Sociobiol.* 46: 123–128.
- Kotiaho, J., Alatalo, R. V., Mappes, J., Parri, S. & Rivero, A. 1998b: Male mating success and risk of predation in a wolf spider: a balance between sexual and natural selection? — *J. Anim. Ecol.* 67: 287–291.

- Krebs, C. J. 1989: *Ecological methodology*. — Harper & Row, New York.
- Kronstedt, T. 1996: Vibratory communication in the wolf spider *Hygrolycosa rubrofasciata* (Araneae, Lycosidae). — *Revue Suisse De Zoologie hors série*: 341–354.
- Lessells, C. M. & Boag, P. T. 1987: Unrepeatable repeatabilities: a common mistake. — *The Auk* 104: 116–121.
- Leung, B. & Forbes, M. R. 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, M. R. & Houle, D. 2000: Fluctuating asymmetry as a bioindicator of stress: comparing efficacy of analyses involving multiple traits. — *Am. Nat.* 155: 101–115.
- Mappes, J., Alatalo, R. V., Kotiaho, J. & Parri, S. 1996: Viability costs of condition-dependent sexual male display in a drumming wolf spider. — *Proc. R. Soc. Lond. B* 263: 785–789.
- Moore, D. S. & McCabe, G. P. 1989: *Introduction to the practice of statistics*. — W. H. Freeman, New York.
- Møller, A. P. 1990: Fluctuating asymmetry in male sexual ornaments may reliably reveal male quality. — *Anim. Behav.* 40: 1185–1187.
- Møller, A. P. 1992: Female preference for symmetrical male sexual ornaments. — *Nature* 357: 238–240.
- Møller, A. P. 1993: Fluctuating asymmetry. — *Nature* 363: 217.
- Møller, A. P. 1999: Condition-dependent asymmetry is fluctuating asymmetry. — *J. Evol. Biol.* 12: 450–459.
- Møller, A. P., Cuervo, J. J., Soler, J. J. & Zamora-Munoz, C. 1996: Horn asymmetry and fitness in gemsbok, *Oryx g. gazella*. — *Behav. Ecol.* 7: 247–253.
- Møller, A. P. & Pomiankowski, A. 1993: Fluctuating asymmetry and sexual selection. — *Genetica* 89: 267–279.
- Møller, A. P. & Swaddle, J. P. 1997: *Asymmetry, developmental stability and evolution*. — Oxford University Press, Oxford.
- Møller, A. P. & Thornhill, R. 1998: Bilateral symmetry and sexual selection: a meta-analysis. — *Am. Nat.* 151: 174–192.
- Palmer, A. R. 1994: Fluctuating asymmetry analyses: A primer. — In: Markow, T. A. (ed.), *Developmental instability: its origins and evolutionary implications*: 335–364. Kluwer, Dordrecht.
- Palmer, A. R. 1999: Detecting publication bias in meta-analyses: A case study of fluctuating asymmetry and sexual selection. — *Am. Nat.* 154: 220–233.
- Palmer, A. R. & Strobeck, C. 1986: Fluctuating asymmetry: measurement, analysis, patterns. — *Annu. Rev. Ecol. Syst.* 17: 391–421.
- Palmer, A. R. & Strobeck, C. 1992: Fluctuating asymmetry as a measure of developmental stability: implications of non-normal distributions and power of tests. — *Acta Zool. Fennica* 191: 57–72.
- Rantala, M. J., Koskimäki, J., Taskinen, J., Tynkynen, K. & Suhonen, J. 2000: Immunocompetence, developmental stability and wingspot size in the damselfly *Calopteryx splendens* L. — *Proc. R. Soc. Lond. B* 267: 2453–2457.
- Rivero, A., Alatalo, R. V., Kotiaho, J. S., Mappes, J. & Parri, S. 2000: Acoustic signalling in a wolf spider: can signal characteristics predict male quality? — *Anim. Behav.* 60: 187–194.
- Roberts, J. R. 1996: *Spiders of Britain and Northern Europe*. — Oxford University Press, New York.
- Roldan, E. R. S., Cassinello, J., Abaigar, T. & Gomendio, M. 1998: Inbreeding, fluctuating asymmetry, and ejaculate quality in an endangered ungulate. — *Proc. R. Soc. Lond. B* 265: 243–248.
- Rosenthal, R. 1986: *Meta-analytic procedures for social research*. — SAGE Publications, California.
- Rowe, L., Repasky, R. R. & Palmer, A. R. 1997: Size-dependent asymmetry: fluctuating asymmetry versus antisymmetry and its relevance to condition-dependent signaling. — *Evolution* 51: 1401–1408.
- Simmons, L. W. 1995: Correlates of male quality in the field cricket, *Gryllus campestris* L.: Age, size, and symmetry determine pairing success in field populations. — *Behav. Ecol.* 6: 376–381.
- Simmons, L. W., Tomkins, J. L., Kotiaho, J. S. & Hunt, J. 1999: Fluctuating paradigm. — *Proc. R. Soc. Lond. B* 266: 593–595.
- Sokal, R. R. & Rohlf, F. J. 1981: *Biometry*. — W. H. Freeman and Co, San Francisco.
- Swaddle, J. P. & Cuthill, I. C. 1994: Preference for symmetric males by female zebra finches. — *Nature* 367: 165–166.
- Thornhill, R., Møller, A. P. & Gangestad, S. W. 1999: The biological significance of fluctuating asymmetry and sexual selection: A reply to Palmer. — *Am. Nat.* 154: 234–241.
- Uetz, G. W. & Smith, E. I. 1999: Asymmetry in a visual signaling character and sexual selection in a wolf spider. — *Behav. Ecol. Sociobiol.* 45: 87–93.
- Van Dongen, S. 1998a: The distribution of individual fluctuating asymmetry: why are the coefficients of variation of the unsigned FA so high? — *Ann. Zool. Fennici* 35: 79–85.
- Van Dongen, S. 1998b: How repeatable is the estimation of developmental stability by fluctuating asymmetry? — *Proc. R. Soc. Lond. B* 265: 1423–1427.
- Van Dongen, S. 1999: Accuracy and power in fluctuating asymmetry studies: effects of sample size and number of within-subject repeats. — *J. Evol. Biol.* 12: 547–550.
- Van Dongen, S., Molenberghs, G. & Matthysen, E. 1999: The statistical analysis of fluctuating asymmetry: REML estimation of a mixed regression model. — *J. Evol. Biol.* 12: 94–102.
- Van Valen, L. 1962: A study of fluctuating asymmetry. — *Evolution* 16: 125–142.
- Watson, P. J. & Thornhill, R. 1994: Fluctuating asymmetry and sexual selection. — *Trends Ecol. Evol.* 9: 21–25.
- Whitlock, M. 1998: The repeatability of fluctuating asymmetry: A revision and extension. — *Proc. R. Soc. Lond. B* 263: 1429–1431.
- Zar, J. H. 1996: *Biostatistical analysis*. — Prentice-Hall, New Jersey.