APPENDIX V
Fluctuating-asymmetry analysis: A step-by-step example
(To accompany Palmer & Strobeck 2003, CH 17. "Fluctuating asymmetry analyses revisited" in *Developmental Instability (DI): Causes and Consequences*. M. Polak ed., Oxford University Press)

NOTE: Figures and tables referred to as Fig.1, Table 1, etc. below are those figures and tables in the book chapter itself. Those referred to as Fig. V.1, Table V.1, etc., are contained in this appendix.

PREFACE

Fluctuating asymmetry (FA) analyses are not complicated. Nor do they require much statistical sophistication. But, as with any analysis that attempts to detect differences among groups where the signal is small relative to the noise, they do require considerable care to avoid being mislead by simplified statistical descriptors of bilateral variation. To help avoid some of the most common pitfalls of FA analyses, and to illustrate some of the ways in which FA analyses may be made more robust, and may be greatly clarified and simplified, we present a complete, step by step re-analysis of a published study of FA variation in thrips. Palmer (1994) provides a fuller explanation of the rationale for and the conduct of many of the analyses described below.

FA analyses depend heavily upon statistical inference (Palmer, 1994). This leads many inexperienced investigators to presume that statistical élan somehow substitutes for common sense. As we show below, many steps in a FA analysis are best preceded by visual inspection of the data, preferably via judiciously selected scatter plots. Many a false or misleading conclusion could be avoided if more efforts were made to examine the data visually, both to validate that they are well behaved (e.g., assumptions about the distribution of variation), and also (hopefully!) to obtain visual confirmation of statistical evidence for real biological differences. In many ways, differences between groups that are apparent visually in a scatter plot, are more likely to reflect biologically significant differences than those that depend primarily on low P values.

DATA AND GOALS OF THE CRESPI & VANDERKIST STUDY

Data from a previously published study (Crespi and Vanderkist, 1997) were generously provided by Bernie Crespi for this exercise, and may be downloaded from the internet: http://www.biology.ualberta.ca/palmer.hp/DataFiles.htm.

In the original Crespi/Vanderkist raw data, four individuals (of the original total of 140) were missing one or both measurements for one of the femurs. To ensure balanced and more straightforward analyses at all levels, all the data for these four individuals were excluded from the data files and analyses outlined below. For this reason, the descriptive data and statistical results differ slightly from those in the original study.

For FA analyses, it is often convenient to format the data file in two ways (Tables 2a and 7a of Palmer, 1994). The first facilitates error analysis and tests for departure from FA, and contains all the raw, replicate measurements for all samples, traits and individuals in a single column. The Crespi/Vanderkist data for these analyses are in the supplementary data file: CrespiData_forError. The second facilitates tests for FA outliers and comparisons of FA among groups of interest, and contains all the measurements for each individual thrips in a single row. The Crespi/Vanderkist data for these analyses are in the supplementary data file: CrespiData_forFA.

Crespi & Vanderkist (1997) asked a series of questions about bilateral variation of two traits (wings, foreleg femora) in a single species of haplo-diploid, gall-forming thrips (*Oncothrips tepperi*; Thysanoptera):

1) *Does FA vary with trait functionality?* Functionally significant traits (wings of dispersers) are predicted to be more developmentally stable, and therefore exhibit lower FA, than vestigial traits (wings of soldiers) (Palmer and Strobeck, 1986).

2) *Does FA vary with heterozygosity (sex)?* In these thrips, males are haploid and females are diploid. If heterozygosity affects DI (Vollestad et al. 1999), then males should exhibit higher FA than females for both traits.

3) *Does FA differ between castes?* Both sexes of *Oncothrips tepperi* exhibit a disperser (fully winged) and a non-disperser morph (soldier with reduced wings). However, members of both castes use their forelegs for fighting and walking. So if FA depends on functionality then wings of dispersers should exhibit lower FA than wings of soldiers, but forelegs of dispersers and soldiers should exhibit comparable FA. A test for differences in foreleg FA between castes therefore tests for some unexpected differences in DI between the castes. If foreleg FA does not differ between castes, this would strengthen conclusions about the effect of trait functionality on FA (question 1).

4) *Does the significance of any of the main effects (trait function, sex, caste) depend on interactions with other variables?* Larger sample sizes for specific biological questions may be achieved by pooling data from both traits, both sexes or both castes. However, such pooling may enhance or obscure differences between groups if the magnitude of one main effect depends on the state of a second. Tests for interactions between main effects may therefore reinforce the strength of conclusions about the statistical significance of a particular main effect.

MEASUREMENT PROCEDURE IN THE CRESPI & VANDERKIST STUDY

Because the measured differences between-sides depend on both true underlying FA and measurement error (ME), and because between-sides differences due to ME are indistinguishable from real differences between sides (Palmer and Strobeck, 1986), a reliable estimate of ME is essential for any FA analysis (Section V.A). Crespi and Vanderkist (1997) took three important precautions when measuring the lengths of wings and femora to ensure a reliable and unbiased estimate of ME.

First, all measurements of wings and femora were taken blind (i.e., without knowledge of the sex or caste of each individual). This avoided any conscious or unconscious bias that might influence the care with which measurements were taken.

Second, the first and second measurements of each trait were taken on separate days, without any reference to prior measurements. This order of replication ensures a) that effects of day or experience on measurement reliability are comparable between traits, between sexes, and between castes, and b) that replicate measurements were not influenced by the value for the initial measurement, or by being able to remember specific subjective decisions about the endpoints used when taking a measurement on a particular trait of a particular individual. In other words, this order of measurement ensured a reliable estimate of all the factors that contribute to ME.

Third, all measurements were taken by one individual. This ensured that differences in measurement among thrips were not confounded by virtually unavoidable differences in ME among measurers (Yezerinac, et al., 1992; Helm and Albrecht, 2000).

Because FA variation is such a small percentage of trait size (Palmer, 1996a), all three of these precautions are important to ensure the integrity of the data.

STEP BY STEP EXAMPLE OF AN FA ANALYSIS

STEP 1) Inspect data for bad raw measurements

Rationale. Very often, particularly in large data sets, errors unrelated to simple measurement imprecision may creep in. Examples include recording errors, transcription errors, data entry errors, calibration errors, sorting errors, spreadsheet cockups, acts of demonic intrusion, etc. Because these may greatly inflate estimates of ME, and therefore bias most descriptors of FA (Section V.A1), visual inspection of scatter plots of replicate measurements are an essential first step to any FA analysis.

Analysis: Visual inspection of scatter plots. Contrasting the difference between replicate measurements of pairs of traits (as in Fig. V.1) helps identify aberrant measurements as well as individuals that may have been unusually difficult to measure.

Results. Scatter plots of replicate measurements revealed four potentially anomalous ME values, all of which were for wings [points (i) - (iv); Fig. V.1].

Conclusion. Tests should be conducted to determine whether these apparent ME outliers are greater than expected due to simple sampling error.

STEP 2) Are apparent ME outliers more deviant than expected due to chance?

Rationale. If scatter plots reveal one or more points that appear to be outliers, such data may be eliminated from subsequent analyses, but only if they meet *a priori* statistical criteria (Sokal and Rohlf, 1995). Different outlier tests, however, apply different criteria. So, in the end, an element of arbitrariness persists.

Analysis: Outlier tests. For small sample sizes ($n < 25$), Dixon's test may be used to place a P value on suspected deviants (Sokal and Rohlf, 1995, p. 406-7). For larger sample sizes, the deviation of a potential outlier (X_i) is expressed as Grubb's test statistic (t_G), which is nothing more than the deviation of the observed value (X_i) from the sample mean ($\bar{X}$), expressed as a proportion of the sample standard deviation (SD):

$$t_G = (X_i - \bar{X}) / SD \qquad \text{(V.1)}$$

Values for Grubb's test statistic (t_G) are then compared against tabled critical values (e.g.,Table DD of Rohlf and Sokal, 1995).

Normally, the statistical question being asked in Grubb's test is: Does an observation deviate significantly from the sample mean? For ME, if all M_1, for example, were taken on one day and all M_2 were taken on another, a consistent difference between days is a possibility. Therefore comparison of outliers to the mean of $M_2 - M_1$ seems warranted. As it turns out, for the wings of dispersers M_2 was weakly (0.26% of wing length) but significantly larger than M_1 (t_s= 3.538, P< 0.001; see rows 3 and 4, Table V.1), so testing versus the mean is most appropriate. This slight, but consistent difference between M_1 and M_2 will increase the estimated ME for the wings of dispersers, which will affect the test of significance of FA relative to ME, but it will not inflate FA because this consistent difference cancels out when computing the difference between sides (R - L). No consistent differences were apparent between M_1 and M_2 in the other groups (wings of soldiers, femurs of dispersers and soldiers). However, because the expected value of the mean of

$M_1 - M_2$ is normally zero, it also seems wise to compare observations to zero. Furthermore, because the sign of the difference between a pair of measurements is arbitrary, Grubb's test should be done as a two-tailed test.

Finally, because more than one grouping of the data (trait or sample) is being examined for outliers, a sequential Bonferroni correction for multiple tests (Rice, 1989; Palmer, 1994) should be applied to the P values obtained from Grubb's test for individual outliers. The number of groups to use for this correction depends on the structure of the data. In tests for ME outliers, the number of groups would minimally be the number of traits (two, in the present example), because the ME would normally be the same for the same trait measured in different groups. However, if ease of measurement of the same trait differed between groups in the study (e.g., between sexes or between castes), and tests for ME outliers were done separately on each group, then the number of groups would have to be increased accordingly. In Oncothrips, wings are vestigial in soldiers, so ME might differ between the wings of soldiers and dispersers. Therefore it is wiser to test for ME outliers separately for soldiers and dispersers, and $N_{groups} = 4$ for the Bonferroni correction.

Results. Four wings in the Oncothrips data set appeared to exhibit higher than expected ME (points i - iv in Fig. V.1), however, only two (points i & iii) were compellingly larger than expected due to chance (Table V.1). Point (iv) was a marginally significant outlier when compared to zero, but not a significant outlier when compared to the sample mean.

Conclusion. To eliminate the potentially confounding effects that anomalous measurements can introduce into a FA analyses, such as spurious leptokurtosis (Section V.B2c) and reduced statistical power (Section V.A1g), observations (i) and (iii) should ideally be re-measured on the original material, if possible, or excluded from further FA analyses. Observation (iii) was also associated with a statistically aberrant FA estimate (observation (i) of Fig. V.3a; see Step 5 below), illustrating clearly how anomalous measurements can yield anomalous FA values.

Because of the small, but consistent difference between M_1 and M_2 for male disperser wings (see Step 2, Rationale), the test versus the mean is most appropriate for observation (iv) (Table V.1). Therefore, even though observation (iv) was a marginally significant outlier when compared to zero, observation (iv) can not be excluded reliably on statistical grounds. However, if re-measurement were possible, it would be wise to re-measure observation (iv).

STEP 3) Inspect data for aberrant individuals (trait size & asymmetry)

Rationale. In spite of care taken during measurement, differences between the sides of some individuals may be artificially inflated due to injury, wear, disturbances during development unrelated to DI, calibration errors between separate measurement sessions, or extreme effects of phenotypic plasticity etc. Such errors will yield artificially inflated values of FA and among-individual heterogeneity in DI (Sections III.B, III.C).

Analysis: Visual inspection of scatter plots. Scatter plots of left vs right measurements for each trait can help reveal outliers or errors.

Results & Conclusion. Scatter plots of left vs right femur and wing in Oncothrips suggest no grossly anomalous asymmetry measurements or extreme-sized individuals for either trait (Fig. V.2).

STEP 4) Inspect data for aberrant individuals (trait asymmetry)

Rationale. Although scatter plots of right versus left are informative for extreme trait sizes, and truly extreme asymmetry values, they may not reveal more subtle outliers where trait size varies considerably among individuals, as it does in Oncothrips (Fig. V.2). Scatter plots of asymmetry in one trait vs asymmetry in a second are more likely to reveal FA differences that have been artificially inflated due to injury, wear, disturbances during development unrelated to DI, or calibration errors between separate measurement sessions, etc. (Section V.A1). Identifying potentially anomalous individuals or traits is an important step in a FA analysis, since deviant individuals may seriously confound subsequent statistical analyses. For example, these outliers need to be detected, and removed if warranted, before conducting a test for the significance of FA relative to ME (Step 6 below) because they may yield a spuriously high estimate of FA relative to ME.

Such scatter plots may also yield the first hints of answers to two potentially interesting questions. Do some groups exhibit higher FA than others (wider scatter of points)? Are deviations from symmetry in one trait paralleled by deviations in the same direction of a second trait (i.e., is a slope to the scatter obvious), as might occur if traits were not developmentally independent (see Klingenberg 'integration' chapter in the original volume)?

Analysis: Visual inspection of scatter plots. Inspect scatter plots of (R-L) for one trait versus (R-L) for a second trait.

Results. Scatter plots of wing FA vs femur FA suggest two possible outliers for wing FA (i and ii) and one for femur FA (iii) of dispersers (Fig. V.3a). No outliers were apparent for soldiers (Fig.

V.3b).

Conclusion. Tests should be conducted to determine whether these three apparent FA outliers are more deviant than expected due to sampling error.

STEP 5) Are FA outliers more deviant than expected due to sampling error?

Rationale & Analysis. Same as for Step 2.

Results. All three possible outliers in Fig. V.3a met the statistical criteria for outlier status based on Grubb's test (see Step 2), regardless of whether they were compared to the sample mean or to zero (the expected value of the mean of a frequency distribution of R - L; Table V.2). However, the test for observation (iii) — female femur — became marginally non-significant after a sequential Bonferroni correction. If re-measurement were an option, it would be preferable to re-measure these individuals in case the anomalous FA values were due to recording or labelling errors.

Conclusion. Both observations for males, (i) and (ii), are legitimately excluded from subsequent analyses based on statistical criteria. As it turns out, observation (i) was also an outlier for wing ME (see Step 2 above), which illustrates nicely how anomalous measurements can yield anomalous asymmetry measures, and therefore how important such cautions are at the outset of an analysis.

The dubious value for the one female observation (observation iii; Table V.2) creates a problem. Compared to the sample mean, it meets Grubb's outlier criterion even after the sequential Bonferroni correction. But it does not meet Grubb's outlier criterion after the sequential Bonferroni correction when compared against zero. The safest procedure at this stage is to conduct subsequent analyses both with and without this suspicious value. Any significant comparisons that depend on the inclusion of this single observation are questionable. However, if the statistical significance of subsequent tests are not affected by inclusion of this observation, then it is probably safest to exclude it from the calculation and presentation of descriptive statistics, or compute the descriptive statistics both with and without this datum.

STEP 6) Are subtle asymmetries significantly greater than ME?

Rationale. Differences between the sides in studies of subtle asymmetries are often around 1% of trait size (Palmer, 1996a). In normal morphological or ecological studies measurements are rarely taken with much greater precision than this. Clearly, to have any hope of detecting meaningful differences in FA among groups of interest, the between-sides differences due to FA must be shown to be significantly greater than the between-sides differences due to ME (Section V.A).

Analysis: ANOVA of sides x individuals. A simple, two-way ANOVA (sides X individuals) performs this test nicely (Palmer and Strobeck, 1986; see Table 2 of Palmer, 1994, for the layout of the data and the correct computation of P values).

This ANOVA procedure is best done after both ME outliers (Steps 1 and 2) and FA outliers (Steps 3 - 5) have been removed, so that neither ME nor FA are inflated by spurious values.

Normally, the ANOVA procedure to test for FA relative to ME (Section V.A5) is conducted once for each trait, where all individuals for which replicate measurements were taken are included, since ME is assumed to be the same for the same trait measured in different individuals. However, as noted above (Step 2), this may not always be the case. For example, in Oncothrips, wings of soldiers are vestigial, so ME might be different for soldiers and dispersers. To guard against possible differences in ME among groups, particularly since ME differences among groups might give the mistaken impression that FA differs among groups (Section V.A1), it is wise to conduct separate analyses for each trait and each group separately, at least until ME is confirmed not to differ among groups.

Results. For both traits of both castes, the between-sides variation was highly significantly greater than that expected due to ME ($P< 0.0001$, Table V.3c). Note, however, how the error variance contributes from 32 to 45% of the total between-sides variance for three of the four groups (Table V.3f). Only for the wings of soldiers was the error variance a small fraction of the between-sides variance (0.23%; Table V.3f). These differences are also reflected in the repeatability of FA (ME5, Table V.3g).

Expressed in a more conventional way, ME1 (the average difference between replicate measurements, Table 3) makes up an even higher proportion of the between-sides variation: from 57 to 67% for femurs and disperser wings, and 5% for soldier wings (Table V.3j).

Note how FA10a (ME excluded) is always lower than FA4a (ME included). The difference between these two indices therefore represents the contribution of ME to FA.

FA clearly differs among the four groups. FA of soldier wings is nearly ten times higher than the FA for the remaining groups, regardless of how FA is calculated (Table V.3e, i, k). Notice how different ways of computing FA yield somewhat different estimates. Estimates derived from variances (FA10a and FA4a, Table V.3e,i) suggest greater FA variation among groups than the estimate derived from the mean absolute deviation (FA1, Table V.3k).

Conclusion. Because the between sides variation (MS_{SI}) was significantly greater than the error variation (MS_{err}), additional analyses of asymmetry variation are justified. For traits or samples where MS_{SI} is not significantly greater than MS_{err}, no further analysis is warranted.

The relatively high proportions of ME variation are also reflected in the dramatic drop in the approximate degrees of freedom for FA10 (Table V.3e) compared to that for the number of individuals in the analysis (Table V.3b) for these three groups: effective reductions in sample sizes of from 57 to 72%!

It should come as no surprise that when ME is a sizeable fraction of FA, the confidence in estimates of FA is lowered substantially, even where FA is significantly larger than ME statistically.

STEP 7) Is ME comparable among different traits and samples?

Rationale. Unless ME is factored out (e.g., via one of the forms of FA10, Table 1), the between-sides variation (FA) will always be inflated by ME (Fig. 2c) (see also Fig. 7 of Palmer, 1994). If ME is comparable among groups of interest, this allows multi-factor or multi-trait analyses to be conducted much more readily (see Step 10 below). However, if ME differs significantly among traits, or among groups of interest, differences in FA may arise that are an artifact of ME differences (Section V.A1f). Therefore, tests for differences in ME among traits or samples is an essential preliminary step in FA analyses.

Analysis: Levene's test for heterogeneity variance. See Section VI.A.

Results. The original test for significance of FA relative to ME revealed what appeared to be differences in ME among groups (Table V.3d): although the MS_{err} for wings of soldiers and dispersers were virtually identical, the MS_{err} of femurs of dispersers was only about 60% of that for soldiers ($P= 0.0039$; F-test of MS). Inspection of Fig. V.1, however, reveals no obvious differences in the error variation of femurs between soldiers and dispersers.

When the ME variation was scrutinized more closely by a 3-factor Levene's test for heterogeneity of variance, however, no significant overall differences were observed between sexes, castes, or traits, or among any of the interactions between these factors (Table V.4).

Conclusion. The absence of significant error variance heterogeneity (Table V.4) means subsequent analyses need not correct for possible ME differences between samples or traits. It also means differences among samples in FA1 may be used to infer differences in DI among samples, as long as the between-sides variation meets the criteria for ideal FA (see step 9 below).

The unusually low error variance for disperser femurs (Table V.3d), however, suggests caution when comparing the FA of disperser femurs to that of soldier femurs, unless such a comparison is done with FA10, where the ME has been partitioned out (see Step 10 below).

STEP 8) Does FA depend on trait size?

Rationale. Size-dependent effects can greatly complicate interpretation of FA variation among traits or samples (Section IV.A3).

However, in some cases between-sides differences may be small and the size range modest. In other cases, the size range may be large, but between-sides differences relatively constant. In both of these situations, dogmatic correction for body-size effects may yield apparent differences in DI where none exist (Fig. 2c). Therefore, tests for size-dependence should be conducted before applying any size correction. In addition, since the underlying model of size-dependent variability predicts a positive association (larger traits are more variable), corrections for size-dependence are only justified if the size-dependence is positive.

Analysis: Correlation tests. (see Section IV.A4)

Results. Scatter plots of trait asymmetry |R - L| vs trait size [(R+L)/2] for femurs and wings suggest no association between trait asymmetry and trait size except for the wings of male and female soldiers (Fig. V.4). The statistics support this observation. Femur FA did not depend on femur size for either caste or either sex, or when analyzed together (Table V.5a,b, c). Similarly, wing FA did not depend on wing size for either male or female dispersers or for both sexes of dispersers combined. However, wing FA did depend on wing size for both male and female soldiers regardless of whether the test was nonparametric (Table V.5a,b) or parametric (Table V.5c). However, after applying a sequential Bonferroni correction for multiple tests, only the association for male soldiers remained significant. Nonetheless, when data for male and female soldiers were combined, the dependence was highly significant. The highly significant dependence of FA on trait size for all wings pooled is not very meaningful because smaller (vestigial) wings are predicted to be more variable.

Rather unexpectedly, as noted by Crespi and Vanderkist (1997), the strong associations between trait asymmetry and trait size for wings of soldiers were negative (asymmetry was less in larger wings of smaller soldiers).

Conclusion. Only for wings of soldiers was a significant and biologically meaningful association found between trait asymmetry and trait size. However, this association was negative rather than the positive one that would be expected if variability increased with the mean size of a trait, therefore, no correction for scale effects is warranted.

One might argue that the negative size-dependence should be removed, but this depends on the cause. If a plausible developmental or methodological peculiarity could account for a decline in FA with increasing size, then some form of size correction might be warranted (see Leung, 1998). On the other hand, since FA sometimes declines with increasing body size because larger individuals are higher quality, this form of size-dependent FA presumably reflects real differences in underlying DI. In this case, no size correction is warranted.

STEP 9) Do traits exhibit ideal FA? Testing for antisymmetry and DA.

Rationale. To use deviations from symmetry as a measure of developmental precision, these deviations should exhibit ideal FA (mean 0, normal distribution of R-L variation; Section V.B). Departures from ideal FA include DA (Section V.B1) and departures from normality (Section V.B2).

Analysis: tests for departures from normality and for directional differences between sides. Tests for departures from normality are best conducted before tests for DA, simply because significant departures from normality compromise conventional parametric tests.

To test for departures from normality, first examine frequency distributions of (R-L) visually (Fig. V.5). Then compute the skew and kurtosis of the frequency distributions of (R-L) for all traits for the smallest subsamples of interest (e.g., see Tables V.6a,c). Finally, repeat these computations on data pooled at the next highest level (e.g., see Tables V.6b,d).

Tests for DA — one-sample t tests comparing the mean(R - L) to zero — may also be conducted for the smallest subsamples of interest (e.g., see Tables V.7a,c) and then repeated on data pooled at the next highest level (e.g., see Tables V.7b,d).

For Oncothrips, soldiers are known to have vestigial wings (Crespi and Vanderkist, 1997) therefore pooling of castes is not recommended for either of these analyses. As always, P-values for departures from normality and for DA must be corrected by sequential Bonferroni for multiple tests (see footnotes to Tables V.6 and V.7).

A useful alternative to conducting multiple separate tests for DA and then correcting the P-values afterwards by sequential Bonferroni, is a single two-way ANOVA (sex x caste) conducted using (R - L) of each trait (see Table V.8)

Results. The frequency distributions of R - L for femurs and wings appeared roughly normal (Fig.

V.5), except for the wings of soldiers, which appear both skewed and somewhat leptokurtic. Statistics more or less confirm this impression (Table V.6).

When sexes were pooled, both femurs and wings of soldiers were significantly leptokurtic, though only barely so after Bonferroni correction (Table V.6d). In addition, the wings of soldiers were also significantly skewed (Table V.6d).

When the sexes were analyzed separately, however, a tendency toward leptokurtosis was apparent among the femurs of female soldiers, and the wings of male soldiers and female dispersers, but none of these remained significant after Bonferroni correction (Table V.6c). In addition, neither trait exhibited significant skew for either sex or caste after Bonferroni correction (Table V.6c).

Fortunately, platykurtosis was not significant for either trait, regardless of sex or caste.

Note that if the ME and FA outliers (see Steps 2 and 5) were allowed to remain in the frequency distributions of (R - L), seven of eight were leptokurtic and negatively skewed (Table V.6a). This leptokurtosis and negative skew became more pronounced, and significant statistically for both traits and both castes, when data for the two sexes were pooled, even after Bonferroni correction for multiple tests (Table V.6b).

The frequency distributions of R - L for femurs and wings also showed no evidence of DA (Fig. V.5). Here too, statistics confirmed this impression (Table V.7). Only when the femurs of dispersers for both sexes were pooled, was any significant DA apparent (Table V.7d), but this significance was marginal and did not persist after a Bonferroni correction for multiple tests. The two-factor ANOVA (sex x caste) also indicated no significant DA for either trait (Table V.8).

<u>Conclusion.</u> Neither platykurtosis (Table V.6) nor deviations of (R-L) from zero (Tables V.7,8) were significant for either trait or caste, regardless of whether the data for the two sexes were pooled or not. Therefore both traits appear to exhibit ideal FA (Fig. 1) for both sexes and castes because these data show no evidence of either antisymmetry or DA.

STEP 10) Does FA differ significantly among traits or samples of interest?

<u>Rationale.</u> Tests for differences in FA among individuals, traits or groups are fundamentally tests for heterogeneity of variance, because indexes of FA all estimate the underlying DI variance (Section II.B).

<u>Analysis.</u> Assuming ME is not a sizeable fraction of between-sides variation and does not differ significantly among traits (Section V.A), a multi-way Levene's test provides a simple and robust test

for FA differences among groups for a wide variety of study designs (Sections VI.B - VI.D).

Results. When all data were analyzed together, FA differed significantly between traits, between castes, and between sexes, regardless of whether the analysis was conduced on unscaled (Fig. V.6a, Table V.9a) or size-scaled data (Fig. V.6b, Table V.10a). However, all two-way interactions, and the three-way interaction were also significant statistically (Tables V.9a, V.10a), therefore conclusions about differences in FA due to trait, caste or sex depend on which subsets of the data are included.

Among dispersers, FA of wings differed significantly from FA of femurs (Tables V.9b, Table V.10b), but FA did not differ between the sexes when analyzed together or separately (Tables V.9b-d, V.10b-d). However, when size effects were not removed wings exhibited higher FA (Fig. V.6a), but when size effects were removed wings exhibited lower proportional FA (Fig. V.6b).

Among soldiers, FA appeared to differ significantly, or nearly so, between traits and between sexes regardless of whether unscaled or size-scaled FA was used (Tables V.9e, Table V.10e). However, the interaction was also significant, or nearly so, and closer examination revealed that effect of sex was limited to wings (Tables V.9g, Table V.10g).

Finally, FA for femurs did not differ significantly between sexes or castes (Tables V.9h, Table V.10h).

Conclusion. Nearly all of the statistically significant variation in FA was limited to wings. As predicted by Crespi and Vanderkist (1997), the wings of soldiers had larger FA than the wings of dispersers (Fig. V.6a,b). However, contrary to predictions, the wings of diploid female soldiers exhibited higher FA than those of the haploid male soldiers, though this effect was not quite significant statistically for unscaled FA (Table V.9g). Wings exhibited lower proportional FA than femurs, perhaps because deviations from symmetry have a more negative impact on flight than on activities of the femurs.

Comparisons of the statistical results to the bar graphs are particularly informative.

STEP 11) Final presentation of results

Preface. The above analyses are all part of a proper FA analysis, and are needed to be convinced data and results are robust. Clearly, presentation of all these analyses in a paper would be excessive, and no journal would accept such a detailed exploration. However, some elements *are* essential to a final presentation of results.

Descriptive data, tests for departures from ideal FA

Text for publication	*Commentary*
Four individuals (3 female soldiers and 1 male disperser) were missing one or more femur measurements and were excluded to ensure analyses were balanced. Inspection of scatter plots revealed three extreme measurement errors that were found to be significant outliers, one of which yielded a significant FA outlier. One additional FA outlier was also rejected on statistical grounds. Therefore, data for 1 female disperser, 1 female soldier, and 2 male dispersers were excluded from all of the analyses.	Confirm that the data were inspected for robustness and that statistically significant outliers were rejected before conducting later analyses.
Two-way ANOVAs (caste x sex) on trait size, (R+L)/2 (Table V.11), for wings and femurs separately revealed that dispersers were larger than soldiers ($P= 0.036$ for femurs and $P= <0.001$ for wings) and that males had smaller femurs than females ($P< 0.001$). For wings, the effect of caste, sex, and the interaction were all highly significant ($P< 0.001$).	Basic description of trait sizes and the significance of trait-size differences for all groupings that seem important.
DA, as mean (R-L), varied among groups, but after sequential Bonferroni correction, none of the samples departed significantly from zero.	Description of DA variation and tests of significance.
After sequential Bonferroni correction, soldiers exhibited significant leptokurtosis of (R-L) for both femurs and wings, and significant skew wings (Table V.11). Cases of platykurtosis did not approach statistical significance so antisymmetry was not evident in these data.	Basic description of skew and kurtosis of (R-L) and significance of departures from zero. Note that the causes of significant leptokurtosis and skew are readily apparent in Figs. V.4 and V.5, Step 8 and 9.
The difference between sides (R-L) did not depend on trait size, (R+L)/2, for femurs ($P> 0.13$ for all groups, Spearman coefficient of rank correlation, Table V.11). Among soldiers, however, wing asymmetry decreased significantly with increasing trait size. Because decreases in trait asymmetry with increasing trait size are not expected due to allometric effects, no correction for trait size variation is warranted.	The dependence of trait asymmetry on trait size must be described. These results are from Table V.5.

Descriptors of FA, ME and tests for significance of FA relative to ME

Text for publication	*Commentary*
A sides by individuals ANOVA (Palmer, 1994) revealed only weakly significant DA for disperser femurs (Table V.12a), but this was not significant after a sequential Bonferroni correction. Trait size variation among individuals was highly significant (Table V.12b).	This analysis confirms the separate analyses for DA in Table V.11. Note that Table V.12 is a condensed version of Table V.3 to reduce redundancy and present the results in a more compact fashion.
Between-sides variation (FA) was significantly greater than ME for all four groups ($P < 0.001$; Table V.12c).	This means it is safe to interpret MS_{SI} as FA because no antisymmetry was present (Table V.11); if antisymmetry is present, it contributes to MS_{SI} (Palmer, 1994).
Measurement error appeared to differ among groups (Table V.12d), but these differences were not significant ($P > 0.1$ for all terms from a sex x caste x trait ANOVA on $\|M_1 - M_2\|$; results not shown).	This analysis is explained in Step 7. Because ME biases estimates of FA, confirmation that ME does not vary among groups is an important element of an FA analysis (Section V.A1). If ME did vary significantly, results like those in Table V.4 should be included in the final paper.
Although FA was significantly larger than ME for all traits, the repeatabilities of FA (Table V.12f) were not very high (38 - 51%) except for soldier wings (99.5%).	At least one measure of ME relative to FA should be included, either as repeatability or as ME1/ of Table V.3.

Differences in FA among groups

Text for publication	*Commentary*
FA varied significantly among groups of Oncothrips tepperi, but the pattern of variation was rather complex (Fig. V.7, Table V.13). All of the significant main effects (trait, caste, sex) and the significant interactions are byproducts of two principal differences. First, soldiers clearly exhibited greater wing FA than dispersers when wings were analysed separately (P< 0.001, analysis not shown). Second, the wings of female soldiers exhibited greater FA than that of males, but this effect was not quite significant when analysed separately because of the smaller sample sizes (P= 0.075, analysis not shown).	Only Fig. V.6a needs to be incorporated in the final paper, because the results are so similar to those in Fig. V.6b. Table V.13 is a highly condensed version of Table V.9. The P-values mentioned here are from Tables V.9e and V.9g, where only soldiers were analysed.
When other differences were tested using subsets of the data, no differences were observed. For example, when femurs were analysed separately, the effects of sex or caste were not significant (P > 0.34, analysis not shown).	The P-value mentioned here is from Table V.9h.
In addition, when wings of dispersers were analysed separately, the effect of sex was also not significant (P= 0.87).	The P-value mentioned here is from Table V.9d.
Similar results were obtained when the analysis was conducted using a size-adjusted index of FA (FA8a) with a couple of notable exceptions (analyses not shown). Among dispersers, wings exhibited proportionally lower FA than femurs (P< 0.001) and female soldiers exhibited significantly greater wing FA than males (P= 0.013).	This discussion refers to the pattern observed in Fig. V.6b and the statistics in Table V.10. The P-values mentioned here are from Tables V.10d,g. Even though no size-correction to FA was warranted for *within-trait* variation in FA (Step 8), wings of dispersers more than twice as large as femurs (Table V.11), so when comparing FA in the two traits of dispersers, proportional FA is more appropriate.

CONCLUSIONS FROM THE WORKED EXAMPLE

The above case study illustrates how FA analyses share much in common with house painting: the final step (painting) is the easiest part! The preparation is the hard work, and if not done right, the final results are shallow and worthless. One valuable outcome of this re-analysis of the Crespi/Vanderkist data set is a strong confirmation of the results and conclusions of the original study (Crespi and Vanderkist, 1997).

In FA analyses, the final tests for differences among samples or traits are relatively simple and straightforward (Step 10), but validating the data and underlying assumptions (Steps 1 - 9) take most of the time. We suspect as many budding FA analysts forget this simple point as frequently as do do-it-yourself painters. Statistical analyses will always yield patterns to which a clever biologist can fit an interesting story. Without Steps 1 - 9, the patterns revealed in Step 10 may bear no relation to reality.

LITERATURE CITED (APPENDIX V)

Crespi, B. J., and B. A. Vanderkist. 1997. Fluctuating asymmetry in vestigial and functional traits of a haplodiploid insect. Heredity **79**: 624-630.

Helm, B., and H. Albrecht. 2000. Human handedness causes directional asymmetry in avian wing length measurements. Anim. Behav. **60**: 899-902.

Leung, B. 1998. Correcting for allometry in studies of fluctuating asymmetry and quality within samples. Proc. Roy. Soc. Lond. B **265**: 1623-1629.

Palmer, A. R. 1994. Fluctuating asymmetry analyses: A primer. Pp. 335-364 in Developmental Instability: Its Origins and Evolutionary Implications, T. A. Markow, ed. Kluwer, Dordrecht, Netherlands.

Palmer, A. R. 1996a. Waltzing with asymmetry. BioScience **46**: 518-532.

Palmer, A. R., and C. Strobeck. 1986. Fluctuating asymmetry: measurement, analysis, patterns. Annu. Rev. Ecol. Syst. **17**: 391-421.

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution **43**: 223-225.

Rohlf, F. J., and R. R. Sokal. 1995. Statistical Tables. Freeman, San Francisco. 219 pp.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry. Freeman, New York. 887 pp.

Vøllestad, L. A., K. Hindar, and A. P. Møller. 1999. A meta-analysis of fluctuating asymmetry in relation to heterozygosity. Heredity 83: 206-218.

Yezerinac, S. M., S. C. Lougheed, and P. Handford. 1992. Measurement error and morphometric studies: Statistical power and observer experience. Syst. Zool. **41**: 471-482.

Table V.1. Results of Grubb's test for outliers applied to four suspicious wing-length measurements in Figs. V.1a,b. Two-tailed P values were obtained by comparing t_G (Equation V.1) against tabled critical values (e.g.,Table DD of Rohlf and Sokal, 1995). Note that the total N is 272 here because both the right and left wings of 136 individuals were measured.

	M_2 - M_1 for entire sample of wings			Possible outlier		Test vs. mean		Test vs zero	
Caste	N	Mean	SD	Label	(R-L)	t_G (mean)	$P_{2\text{-tail}}$	t_G (zero)	$P_{2\text{-tail}}$
soldier	90	0.00097	0.009780	(i)	0.055	4.986	< 0.001*	5.127	< 0.001*
soldier	90	0.00097	0.009780	(ii)	0.034	3.028	>0.05	3.169	>0.05
disperser	182	0.00203	0.007731	(iii)	0.039	4.066	< 0.001*	4.491	< 0.001*
disperser	182	0.00203	0.007731	(iv)	0.031	3.145	>0.05	3.569	< 0.05

* P values that remain significant at the 0.05 level after a sequential Bonferroni correction for multiple tests (Rice, 1989). N_{groups}= 4 for the Bonferroni correction because wings and femurs were examined separately for dispersers and soldiers. These data are from the supplementary data file: CrespiData_forError.

Table V.2. Results of Grubb's test for outliers applied to three suspicious points in Fig. V.3a. Two-tailed P values were obtained by comparing t_G (Eq. V.1) against tabled critical values (e.g., Table DD of Rohlf and Sokal, 1995).

		Entire sample			Possible outlier		Test vs. mean		Test vs zero	
Sex	Trait	N	Mean	SD	Label	(R-L)	t_G (mean)	$P_{2\text{-tail}}$	t_G (zero)	$P_{2\text{-tail}}$
Male	wing	32	-0.0048	0.01364	(i)	-0.0620	4.193	< 0.001*	4.546	< 0.001*
Male	femur	32	0.0005	0.00718	(ii)	-0.0245	3.478	< 0.01*	3.412	< 0.01*
Female	femur	59	0.0011	0.00725	(iii)	-0.0250	3.607	< 0.01*	3.450	< 0.02

* P values that remain significant at the 0.05 level after a sequential Bonferroni correction for multiple tests (Rice, 1989). N_{groups}= 4 for the Bonferroni correction because wings and femurs were examined separately for dispersers and soldiers. These data are from the supplementary data file: CrespiData_forFA.

Table V.3. Results from a 2-factor, mixed model ANOVA (sides= fixed factor, individuals= random factor, see Table 2 of Palmer, 1994, for a complete explanation of the analysis) on untransformed repeat measurements for two traits (femur, wing) and two castes (disperser, soldier) of Oncothrips tepperi.†

		Femur		Wing	
		Disperser	Soldier	Disperser	Soldier
Results from mixed model, 2-factor ANOVA					
Source of variation	Statistic				
a) Sides (S)	MS_S	0.000198	0.000000023	0.000086	0.000688
(tests for DA)	df	1	1	1	1
	F	5.211	0.000	1.004	0.055
	P	0.0249	0.9855	0.3192	0.8158
b) Individuals (I)	MS_I	0.012460	0.006432	0.006658	0.115967
(tests for trait size	df	87	43	87	43
differences among	F	327.882	95.003	77.684	9.259
individuals)	P	<0.0001	<0.0001	<0.0001	<0.0001
c) S x I interaction	MS_{SI}	0.000038	0.000068	0.000086	0.012525
(tests whether	df	87	43	87	43
asymmetries are	F	2.235	2.471	3.083	433.374
greater than ME)	P	<0.0001	0.0002	<0.0001	<0.0001
d) Error	MS_{err}	0.0000170	0.0000274	0.0000278	0.0000289
(variance of repeat measurements)	df	176	88	176	88
Descriptors of FA and ME derived from the above ANOVA results					
e) FA10a††		0.00366	0.00507	0.00607	0.08920
(in mm)	df	24.18	14.11	37.75	42.80
f) ME3= MS_{err} as % MS_{SI}		44.74	40.47	32.44	0.23
g) repeatability (ME5)		0.38	0.43	0.51	0.995
h) ME1= $0.798\sqrt{MS_{err}}$ § (in mm)		0.00329	0.00418	0.00421	0.00429
i) FA4a= $0.798\sqrt{MS_{SI}}$§§ (in mm)		0.00492	0.00658	0.00740	0.08931
j) ME1 as % FA4a		66.89	63.48	56.86	4.80
k) FA1 mean (mm)		0.00516	0.00584	0.00693	0.07589
SE		0.000389	0.000864	0.000658	0.012292

† Four ME and FA outliers (i and ii, Step 2, i and ii, Step 5) were excluded from this analysis. The data used in this analysis are from the supplementary data file: CrespiData_forError.

†† FA excluding ME. Because FA10a (Table 1) attempts to partition out the 'true' FA variance from the total between sides variance (asymmetry + ME). The df for the derived variance is a function of both the df of the initial variance (MS_{SI}) and the size of $MS_{err,}$ so they are only approximate (see Table 2d of Palmer, 1994).

§ Average difference between pairs of repeat measurements (Table 3b); obtained by converting the error variance to its equivalent in absolute deviations (Section V.A2; Eq. III.8, Appendix III).

§§ FA including ME. FA4a= $0.798\sqrt{MS_{SI}}$ because MS_{SI} = var(R-L) including ME.

Table V.4. Results from a Levene's test for heterogeneity of variance (3-factor ANOVA: sex, caste, trait) of ME (as $|M_2 - M_1|$) for Oncothrips tepperi.†

Source	df	Mean Square	F-test	P
Sex (A)	1	0.0000015	0.07	0.790
Caste (B)	1	0.0000112	0.55	0.460
AB	1	0.0000477	2.33	0.127
Trait (C)	1	0.0000423	2.07	0.151
AC	1	0.0000012	0.06	0.806
BC	1	0.0000409	2.00	0.158
ABC	1	0.0000174	0.85	0.357
Error	520	0.0000204		

† Both ME outliers (Step 2) and FA outliers (Step 5) were excluded from this analysis. The data used in this analysis are from the supplementary data file: CrespiData_forError.

Table V.5. Results of significance tests (Spearman coefficient of rank correlation, Kendall coefficient of rank correlation, least squares linear regression) of associations between trait size and trait asymmetry for both traits, sexes and castes of Oncothrips tepperi (data in Fig. V.4).†

			a) Spearman			b) Kendall			c) Linear regression		
Caste	Sex	N	ρ§	P		τ§	P		r	P	
Femur											
Disperser	Male	30	-0.2083	0.262		-0.1501	0.244		-0.214	0.255	
Disperser	Female	58	-0.1998	0.132		-0.1322	0.143		-0.149	0.265	
Soldier	Male	13	0.0303	0.916		0.052	0.805		0.003	0.992	
Soldier	Female	31	0.2509	0.169		0.1853	0.143		0.349	0.054	
Disperser	pooled	88	-0.0717	0.514		-0.0462	0.533		-0.063	0.558	
Soldier	pooled	44	-0.0374	0.818		-0.0325	0.756		0.057	0.712	
Pooled	pooled	132	-0.0777	0.374		-0.0532	0.366		-0.026	0.764	
Wing											
Disperser	Male	30	0.1216	0.513		0.0934	0.469		0.17	0.369	
Disperser	Female	58	-0.094	0.478		-0.0644	0.475		-0.066	0.622	
Soldier	Male	13	-0.7637	0.008	+	-0.6154	0.003	*	-0.855	<0.001	**
Soldier	Female	31	-0.4113	0.024		-0.2559	0.043		-0.339	0.062	
Disperser	pooled	88	0.0003	0.999		0.0008	0.991		0.014	0.897	
Soldier	pooled	44	-0.5314	<0.001	**	-0.3628	<0.001	**	-0.444	0.002	*
Pooled	pooled	132	-0.5453	<0.001	***	-0.3792	0.0001	***	-0.665	0.0001	***

† Significance of P values after sequential Bonferroni correction for multiple tests (N_{groups}= 8 when each caste and sex was analysed separately, N_{groups}= 4 when each caste was analysed separately but sexes were pooled, and N_{groups}= 2 when castes and sexes were pooled): + 0.1 > P > 0.05, * 0.05 > P > 0.01, ** 0.01 > P > 0.001. The data for these analyses are from the supplementary data file: CrespiData_forFA.

§ Corrected for ties.

Table V.6. Conventional tests for kurtosis and skew of (R - L) for different groupings of femur and wing measurements in both sexes and castes of Oncothrips tepperi. †

Trait/ sex	Caste	N	(R - L) Kurtosis	P		Skew	P	
			Data for all complete individuals included					
a) Sexes separate								
Femur								
Male	Disperser	32	2.902	< 0.001	**	-1.169	0.005	*
	Soldier	13	-0.680	ns		-0.361	0.559	
Female	Disperser	59	1.603	< 0.05		-0.629	0.043	
	Soldier	32	2.558	< 0.01		-0.414	0.318	
Wing								
Male	Disperser	32	7.759	< 0.001	**	-2.279	<0.001	***
	Soldier	13	1.864	< 0.05		-1.008	0.102	
Female	Disperser	59	1.488	< 0.05		0.048	0.878	
	Soldier	32	1.087	ns		-0.984	0.018	
b) Sexes pooled								
Femur	Disperser	91	2.076	< 0.001	**	-0.809	0.001	**
	Soldier	45	2.295	< 0.01	*	-0.550	0.120	
Wing	Disperser	91	7.358	< 0.001	**	-1.449	<0.001	***
	Soldier	45	2.061	< 0.01	*	-1.171	0.001	**
			All four ME and FA outliers excluded (Steps 2, 5)					
c) Sexes separate								
Femur								
Male	Disperser	30	-0.399	ns		-0.023	0.957	
	Soldier	13	-0.680	ns		-0.361	0.559	
Female	Disperser	58	-0.350	ns		0.061	0.847	
	Soldier	31	2.391	< 0.01	+	-0.388	0.357	
Wing								
Male	Disperser	30	-0.556	ns		-0.017	0.969	
	Soldier	13	1.864	< 0.05		-1.008	0.102	
Female	Disperser	58	1.777	< 0.05		0.005	0.986	
	Soldier	31	0.985	ns		-0.985	0.019	
d) Sexes pooled								
Femur	Disperser	88	-0.326	ns		0.045	0.860	
	Soldier	44	2.175	< 0.01	*	-0.540	0.131	
Wing	Disperser	88	1.311	< 0.05		0.042	0.870	
	Soldier	44	1.986	0.01	*	-1.176	0.001	**

† Kurtosis was computed using Eq. 6 and compared to separate critical values for platy- and lepto-kurtosis (Table 5). Skew was computed as $[\sum(X_i - \bar{X})^3 / (N*SD^3)]$, where N is the sample size, $\bar{X}$ is the sample mean, X_i is the value of X for individual i, and SD is the standard deviation of the sample computed using N rather than N-1, and its standard error computed following (Sokal and Rohlf, 1995, for g1, p. 138). N- sample size, SE- standard error, T_s- t statistic, P- probability, ns- not significant (P> 0.05). Asterisks indicate significance levels after a sequential Bonferroni correction for multiple tests [N_{groups}= 8 for (a) and (c), N_{groups}= 4 for (b) and (d); see footnote to Table V.5 for interpretation of symbols]. These analyses were conducted using data in the supplementary data file: CrespiData_forFA.

Table V.7. Conventional tests for DA (departures of mean (R-L) from zero) of femurs and wings for different combinations of sexes and castes of Oncothrips tepperi.†

Trait	Sex	Caste	N	(R-L) Mean	(R-L) SE	T_s	P
		Data for all individuals with complete measurements					
a) Sexes separate							
Femur	Male	Disperser	32	0.0005	0.00127	0.369	0.714
		Soldier	13	0.0034	0.00172	1.967	0.073
	Female	Disperser	59	0.0011	0.00094	1.204	0.234
		Soldier	32	-0.0014	0.00150	0.925	0.362
Wing	Male	Disperser	32	-0.0048	0.00241	1.996	0.055
		Soldier	13	0.0110	0.01739	0.632	0.539
	Female	Disperser	59	0.0001	0.00130	0.046	0.964
		Soldier	32	-0.0109	0.02217	0.490	0.628
b) Sexes pooled							
Femur		Disperser	91	0.0009	0.00075	1.196	0.235
		Soldier	45	0.0000	0.00121	0.009	0.993
Wing		Disperser	91	-0.0017	0.00121	1.366	0.175
		Soldier	45	-0.0045	0.01650	0.275	0.784
		All four ME and FA outliers excluded (Steps 2, 5)					
c) Sexes separate							
Femur	Male	Disperser	30	0.0013	0.00105	1.275	0.212
		Soldier	13	0.0034	0.00172	1.967	0.073
	Female	Disperser	58	0.0016	0.00084	1.881	0.065
		Soldier	31	-0.0015	0.00155	0.936	0.357
Wing	Male	Disperser	30	-0.0024	0.00153	1.536	0.135
		Soldier	13	0.0110	0.01739	0.632	0.539
	Female	Disperser	58	-0.0003	0.00127	0.224	0.824
		Soldier	31	-0.0102	0.02289	0.447	0.658
d) Sexes pooled							
Femur		Disperser	88	0.0015	0.00066	2.284	0.025
		Soldier	44	0.0000	0.00124	0.018	0.985
Wing		Disperser	88	-0.0010	0.00099	1.002	0.319
		Soldier	44	-0.0040	0.01687	0.234	0.816

† Abbreviations and symbols as in Table V.6. These analyses were conducted on data in the supplementary data file: CrespiData_forFA.

Table V.8. Results of tests for DA using a 2-factor ANOVA (sex x caste) of (R - L) for Oncothrips tepperi.†

Source	df	Mean Square	F	P
Femur				
Caste	1	0.000006	0.130	0.7191
Sex	1	0.000132	2.805	0.0964
Sex * Caste	1	0.000162	3.458	0.0652
Error	128	0.000047		
Wing				
Caste	1	0.000073	0.017	0.8959
Sex	1	0.002298	0.543	0.4626
Sex * Caste	1	0.003396	0.802	0.3721
Error	128	0.004233		

† All ME and FA outliers (Steps 2 and 5) were excluded from these analyses. These analyses were conducted on data in the supplementary data file: CrespiData_forFA.

Table V.9. Results from various ANOVA analyses of |R - L| variation for two traits (femur, wing), two castes (disperser, soldier) and two sexes (male, female) in Oncothrips tepperi.†

Source of variation	df	Mean Square	F-test	P
a) All traits, castes and sexes included				
Trait	1	0.0482	45.54	<0.001 ***
Caste	1	0.0451	42.69	<0.001 ***
Sex	1	0.0073	6.94	0.009 **
Trait * Caste	1	0.0426	40.30	<0.001 ***
Trait * Sex	1	0.0071	6.76	0.010 **
Caste * Sex	1	0.0068	6.45	0.012 *
Trait * Caste * Sex	1	0.0074	7.00	0.009 *
Residual	256	0.0011		
b) Dispersers only; both traits and sexes included				
Trait	1	0.0001338	5.16	0.024 *
Sex	1	0.0000075	0.29	0.591
Trait * Sex	1	0.0000017	0.07	0.795
Residual	172	0.0000259		
c) Femurs of dispersers only; both sexes included				
Sex	1	0.0000083	0.62	0.434
Residual	86	0.0000134		
d) Wings of dispersers only; both sexes included				
Sex	1	0.0000010	0.03	0.872
Residual	86	0.0000385		
e) Soldiers only; both traits and sexes included				
Trait	1	0.0663497	20.93	<0.001 ***
Sex	1	0.0103515	3.27	0.074 +
Trait * Sex	1	0.0106453	3.36	0.070 +
Residual	84	0.0031699		
f) Femurs of soldiers only; both sexes included				
Sex	1	0.0000010	0.03	0.862
Residual	42	0.0000336		
g) Wings of soldiers only; both sexes included				
Sex	1	0.0209958	3.33	0.075 +
Residual	42	0.0063062		
h) Femurs only, both castes and sexes included				
Caste	1	0.0000182	0.91	.342
Sex	1	0.0000006	0.03	.862
Caste * Sex	1	0.0000060	0.30	.584
Error	128	0.0000200		

† All four ME and FA outliers (Steps 2, 5) were excluded from these analyses. These analyses were conducted using the supplementary data file: CrespiData_forFA.

Table V.10. Results from various ANOVA analyses of |log(R/L)| variation for two traits (femur, wing), two castes (disperser, soldier) and two sexes (male, female) in Oncothrips tepperi.†

Source of variation	df	Mean Square	F-test	P
a) All traits, castes and sexes included				
Trait	1	0.081084	25.84	<0.001 ***
Caste	1	0.100617	32.06	<0.001 ***
Sex	1	0.041463	13.21	<0.001 ***
Trait * Caste	1	0.094507	30.12	<0.001 ***
Trait * Sex	1	0.046940	14.96	<0.001 ***
Caste * Sex	1	0.042646	13.59	<0.001 ***
Trait * Caste * Sex	1	0.045336	14.45	<0.001 ***
Residual	256	0.003138		
b) Dispersers only; both traits and sexes included				
Trait	1	0.0004057	21.38	<0.001 ***
Sex	1	0.0000066	0.35	0.557
Trait * Sex	1	0.0000110	0.58	0.447
Residual	172	0.0000190		
c) Femurs of dispersers only; both sexes included				
Sex	1	0.0000173	0.66	0.417
Residual	86	0.0000260		
d) Wings of dispersers only; both sexes included				
Sex	1	0.0000003	0.02	0.878
Residual	86	0.0000119		
e) Soldiers only; both traits and sexes included				
Trait	1	0.1282760	13.47	<0.001 ***
Sex	1	0.0615321	6.46	0.013 *
Trait * Sex	1	0.0675050	7.09	0.009 **
Residual	84	0.0095251		
f) Femurs of soldiers only; both sexes included				
Sex	1	0.0000692	1.33	0.255
Residual	42	0.0000521		
g) Wings of soldiers only; both sexes included				
Sex	1	0.1289679	6.79	0.013 *
Residual	42	0.0189982		
h) Femurs only, both castes and sexes included				
Caste	1	0.0000478	1.38	0.241
Sex	1	0.0000849	2.46	0.119
Caste * Sex	1	0.0000206	0.60	0.442
Error	128	0.0000345		

† All four ME and FA outliers (Steps 2, 5) were excluded from these analyses. These analyses were conducted using the supplementary data file: CrespiData_forFA.

Table V.11. Average trait size, DA, kurtosis and skew for different traits, castes and sexes of Oncothrips tepperi.†

					(R-L)		
Caste	Sex	N	(R+L)/2 Mean (SE)	Spearman ρ[a]	Mean (SE)[b]	Kurtosis	Skew
Femurs							
Disperser	M	30	0.28 (0.007)	-0.2083	0.0013 (0.00105)	-0.399	-0.023
Soldier	M	13	0.27 (0.007)	-0.1998	0.0034 (0.00172)	-0.68	-0.361
Disperser	F	58	0.37 (0.006)	0.0303	0.0016 (0.00084)	-0.35	0.061
Soldier	F	31	0.35 (0.002)	0.2509	-0.0015 (0.00155)	2.391 +	-0.388
Disperser	both	88	0.34 (0.006)	-0.0717	0.0015 (0.00066)	-0.326	0.045
Soldier	both	44	0.33 (0.006)	-0.0374	0.0000 (0.00124)	2.175 *	-0.54
Wings							
Disperser	M	30	0.78 (0.008)	0.1216	-0.0024 (0.00153)	-0.556	-0.017
Soldier	M	13	0.60 (0.013)	-0.094	0.0110 (0.01739)	1.864	-1.008
Disperser	F	58	0.79 (0.005)	-0.7637	-0.0003 (0.00127)	1.777	0.005
Soldier	F	31	0.35 (0.026)	-0.4113	-0.0102 (0.02289)	0.985	-0.985
Disperser	both	88	0.78 (0.004)	0.0003	-0.0010 (0.00099)	1.311	0.042
Soldier	both	44	0.42 (0.026)	-0.5314 **	-0.0040 (0.01687)	1.986 *	-1.176 **

† All four ME and FA outliers (Steps 2, 5) were excluded from these analyses. These analyses were conducted using the supplementary data file: CrespiData_forFA. Significance levels after sequential Bonferroni correction for multiple tests (N_{groups}= 8 when sexes were analysed separately, N_{groups}= 4 when sexes were pooled): + 0.1 > P > 0.05, * 0.05 > P > 0.01, ** P< 0.01.

[a] Spearman coefficient of rank correlation between |R-L| and (R+L)/2.

[b] No estimates of DA were significant after sequential Bonferroni correction.

Table V.12. Results from a 2-factor, mixed model ANOVA (sides= fixed factor, individuals= random factor, Palmer, 1994) on untransformed repeat measurements for two traits (femur, wing) and two castes (disperser, soldier) of Oncothrips tepperi.†

		Femur		Wing	
		Disperser	Soldier	Disperser	Soldier
Results from mixed model, 2-factor ANOVA					
Source of variation	Statistic				
a) Sides (S, df= 1)	MS_S	0.000198 *	<0.000001	0.000086	0.000688
b) Individuals (I)	MS_I	0.012460 ***	0.006432 ***	0.006658 ***	0.115967 ***
	df	87	43	87	43
c) S x I interaction	MS_{SI}	0.000038 ***	0.000068 ***	0.000086 ***	0.012525 ***
	df	87	43	87	43
d) Error	MS_{err}	0.0000170	0.0000274	0.0000278	0.0000289
	df	176	88	176	88
Descriptors of FA and ME derived from the above ANOVA results					
e) FA10a (mm)†		0.00366	0.00507	0.00607	0.08920
	df	24.18	14.11	37.75	42.80
f) repeatability (ME5)		0.38	0.43	0.51	0.995
g) FA4a (mm)§		0.00492	0.00658	0.00740	0.08931
h) FA1 mean (mm)§§		0.00516	0.00584	0.00693	0.07589
SE		0.000389	0.000864	0.000658	0.012292

† Computed as $0.798\sqrt{(MS_{SI} - MS_{err})}$ because the number of replicate measurements is two (Table 1). The df for FA10 are approximate (Palmer, 1994).

§ An estimate of FA including ME ($0.798\sqrt{MS_{SI}}$).

§§ Average |R-L| of untransformed measurements.

Table V.13. Results from ANOVA of |R - L| variation for two traits (femur, wing), two castes (disperser, soldier) and two sexes (male, female) in Oncothrips tepperi.†

Source of variation	df	Mean Square	F-test	P
Trait	1	0.0482	45.54	<0.001 ***
Caste	1	0.0451	42.69	<0.001 ***
Sex	1	0.0073	6.94	0.009 **
Trait * Caste	1	0.0426	40.30	<0.001 ***
Trait * Sex	1	0.0071	6.76	0.010 **
Caste * Sex	1	0.0068	6.45	0.012 *
Trait * Caste * Sex	1	0.0074	7.00	0.009 *
Residual	256	0.0011		

† All four ME and FA outliers (Steps 2, 5) were excluded from these analyses.

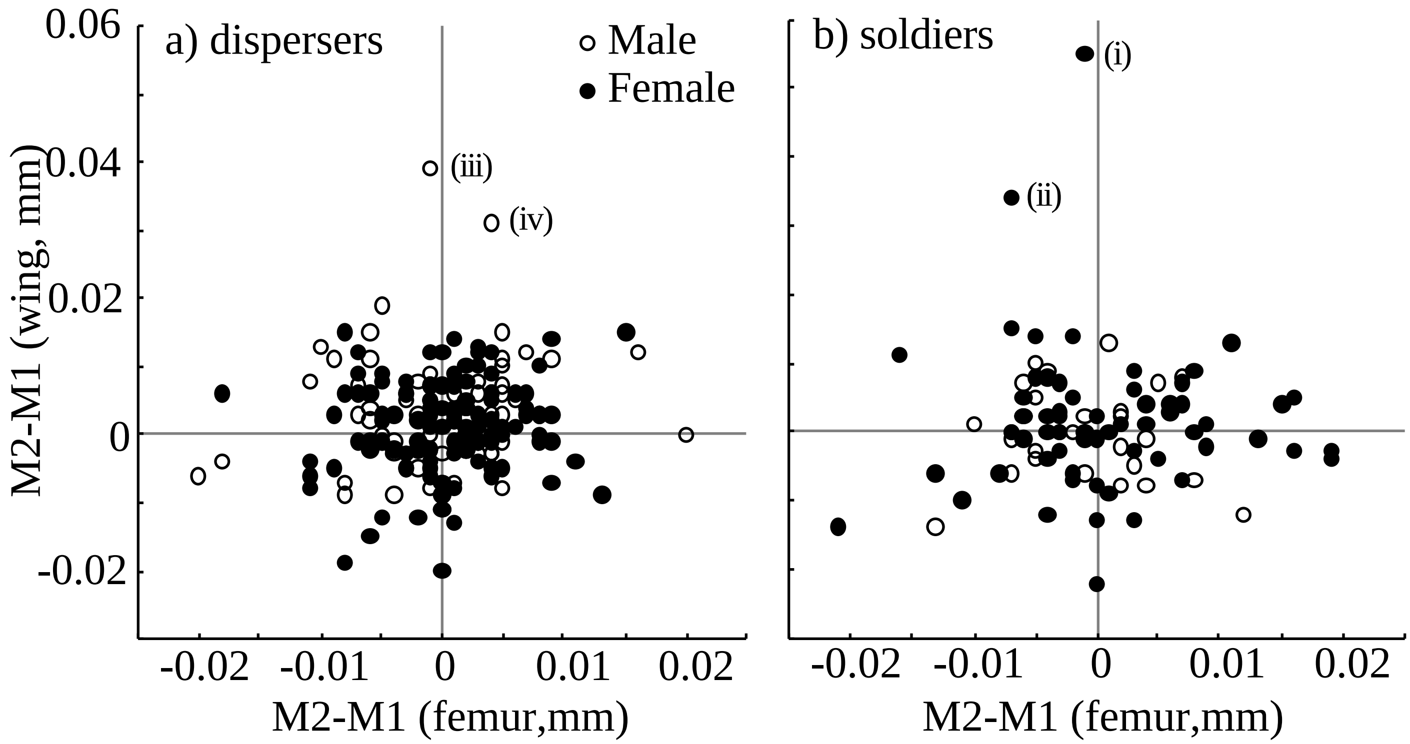

Figure V.1. Scatterplots of the difference between replicate measurements (M_2 - M_1 = measurement 2 - measurement 1) of wing vs femur measurements for each caste and each sex of Oncothrips tepperi. Points (i) to (iv) are possible outliers for wing measurement error (see Step 2 for how to handle such a problem). These data are from the supplementary data file: CrespiData_forError.

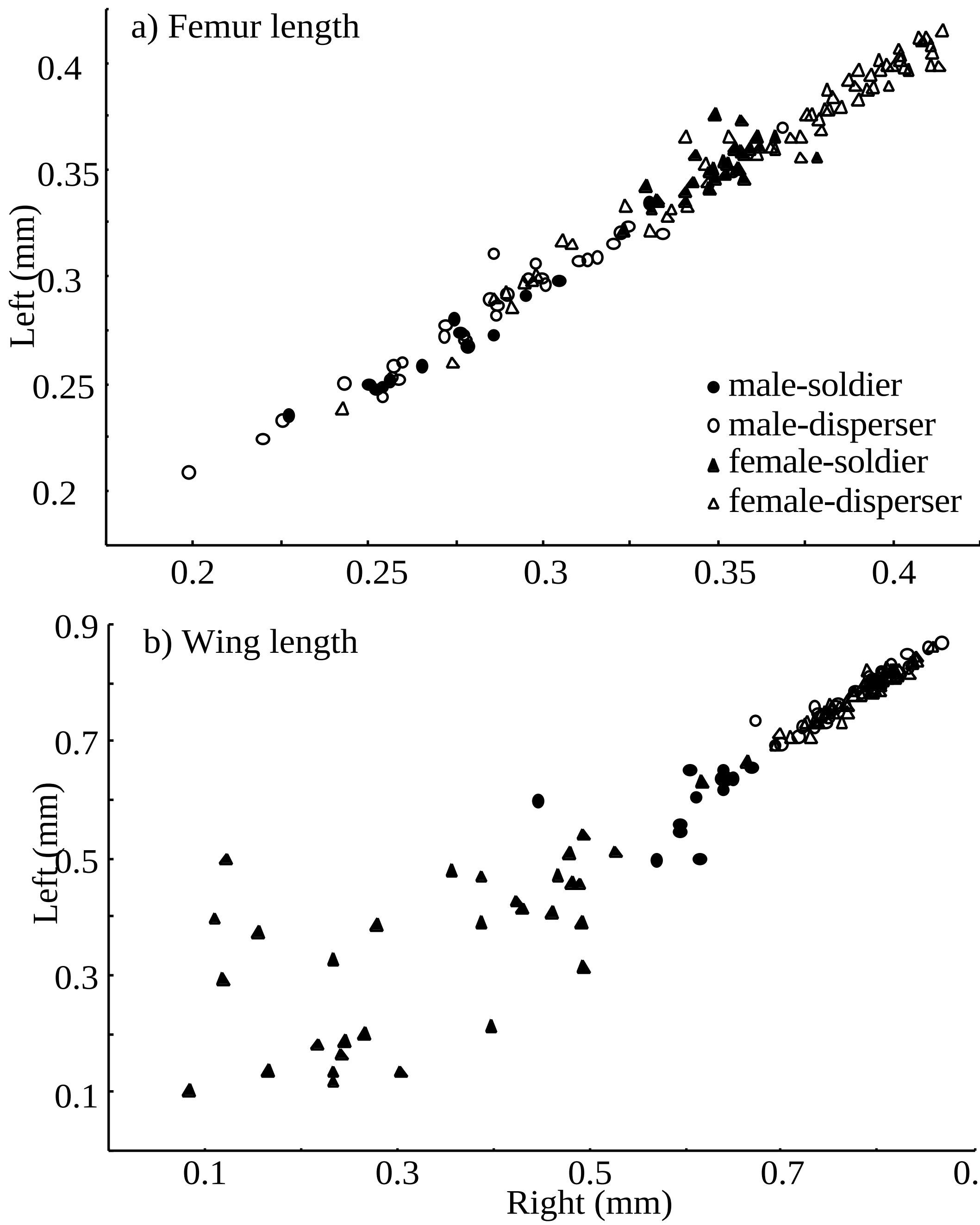

Figure V.2. Scatterplots of right side vs left side for both traits of Oncothrips tepperi (replicate measurements were averaged first). These data are from the supplementary data file: CrespiData_forFA.

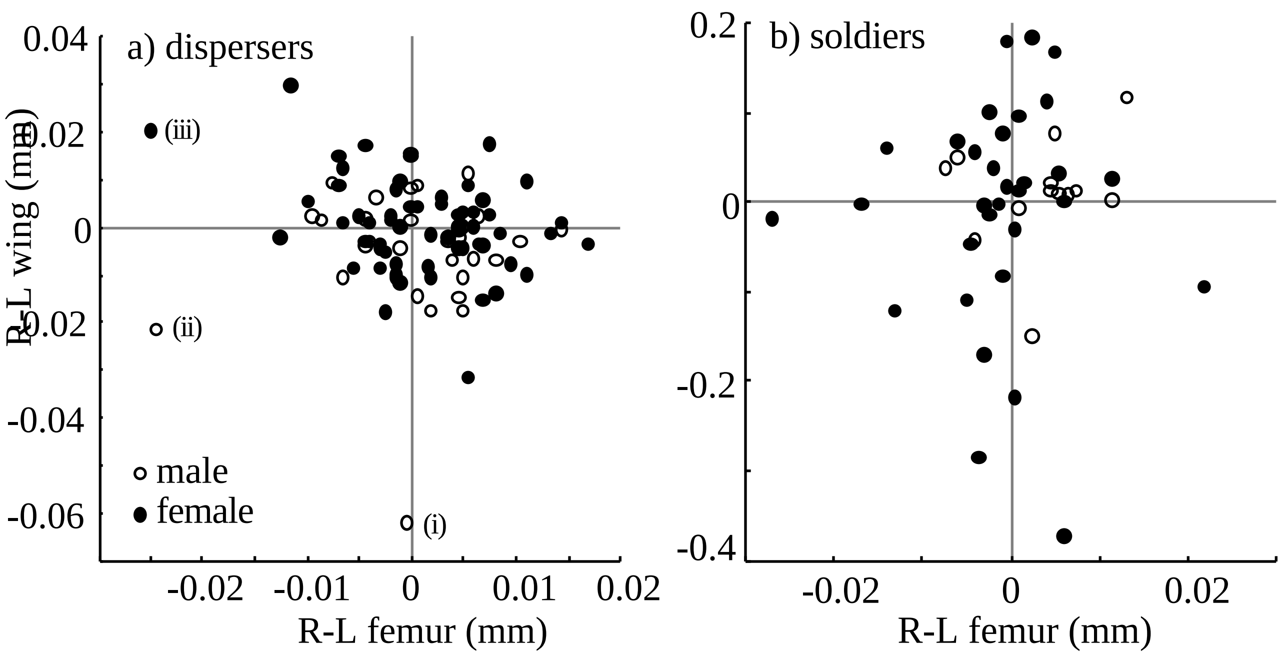

Figure V.3. Scatterplots of fluctuating asymmetries of wings vs fluctuating asymmetries of femurs for both sexes and castes (replicate measurements averaged first) in Oncothrips tepperi. These data are from the supplementary data file: CrespiData_forFA.

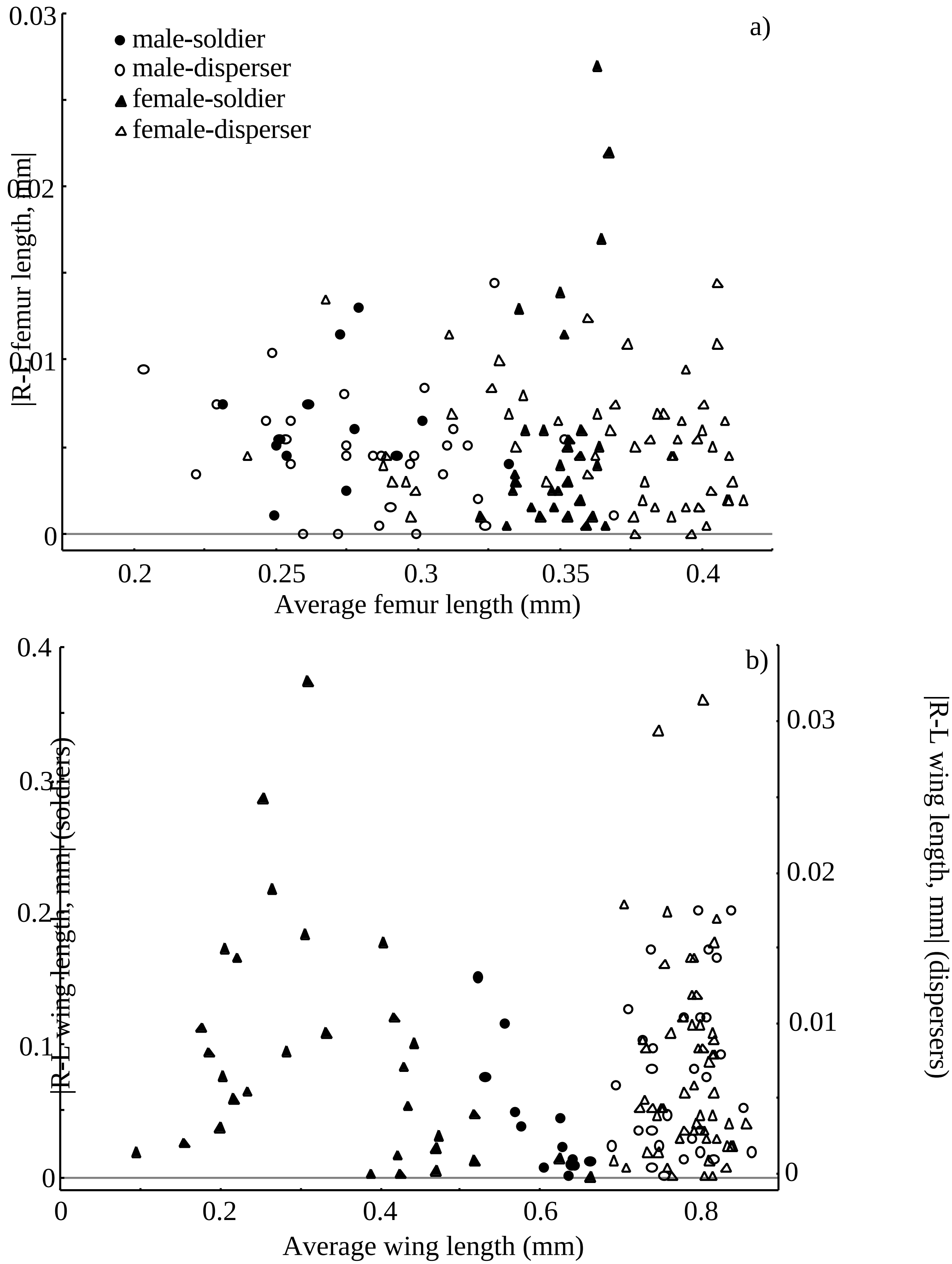

Figure V.4. Scatter plots of trait asymmetry |R - L| vs trait size [(R+L)/2] for femurs (a) and wings (b) of Oncothrips tepperi after all four outliers were moved (Steps 2, 5). These data are from the supplementary data file: CrespiData_forFA. Note the different axes for wings of soldiers (solid symbols) and dispersers (open symbols) in (b).

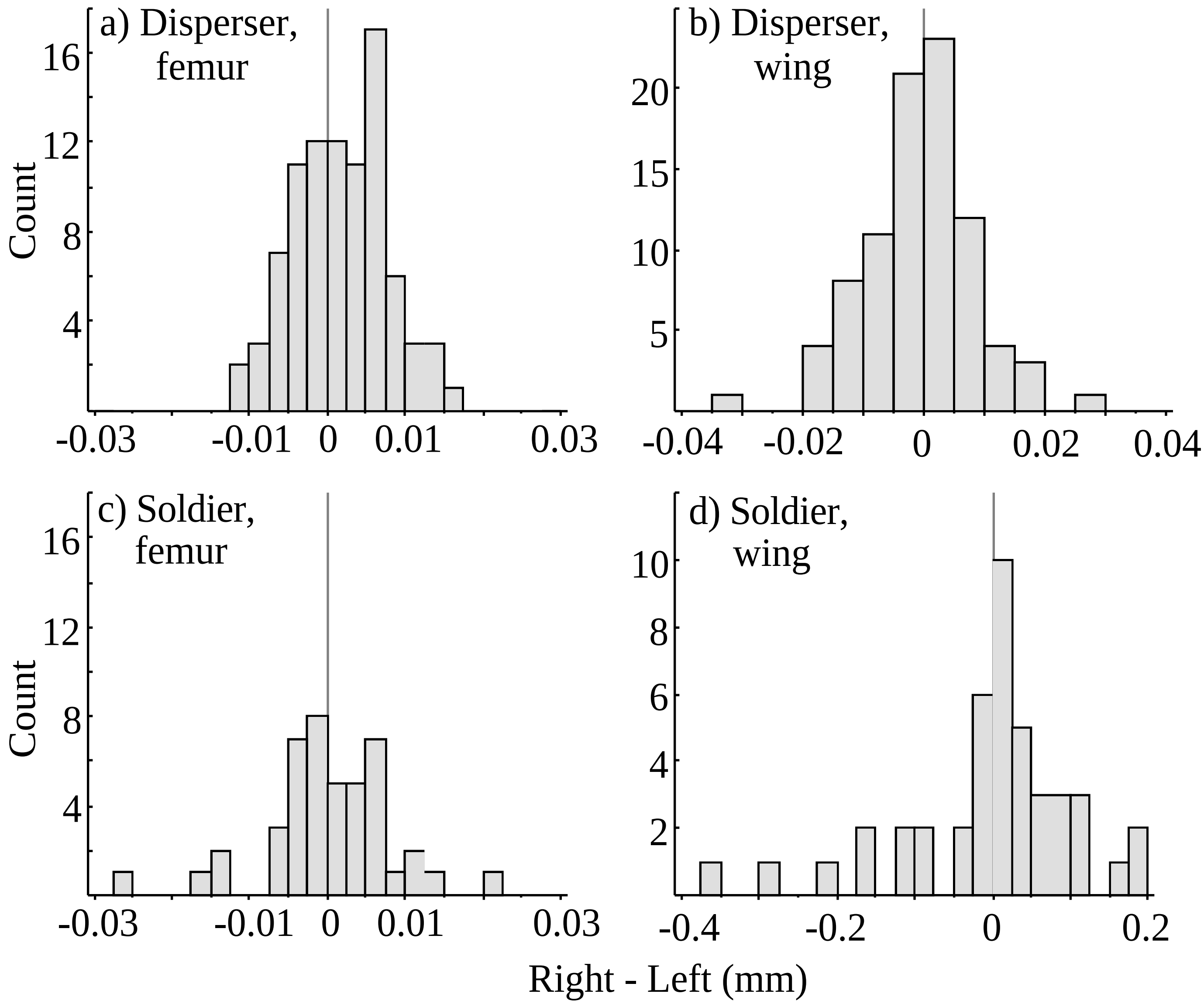

Figure V.5. Frequency distributions of (R - L) for femurs and wings of soldiers and dispersers of Oncothrips tepperi, after the four ME and FA outliers were eliminated (Steps 2, 5). These data are from the supplementary data file: CrespiData_forFA.

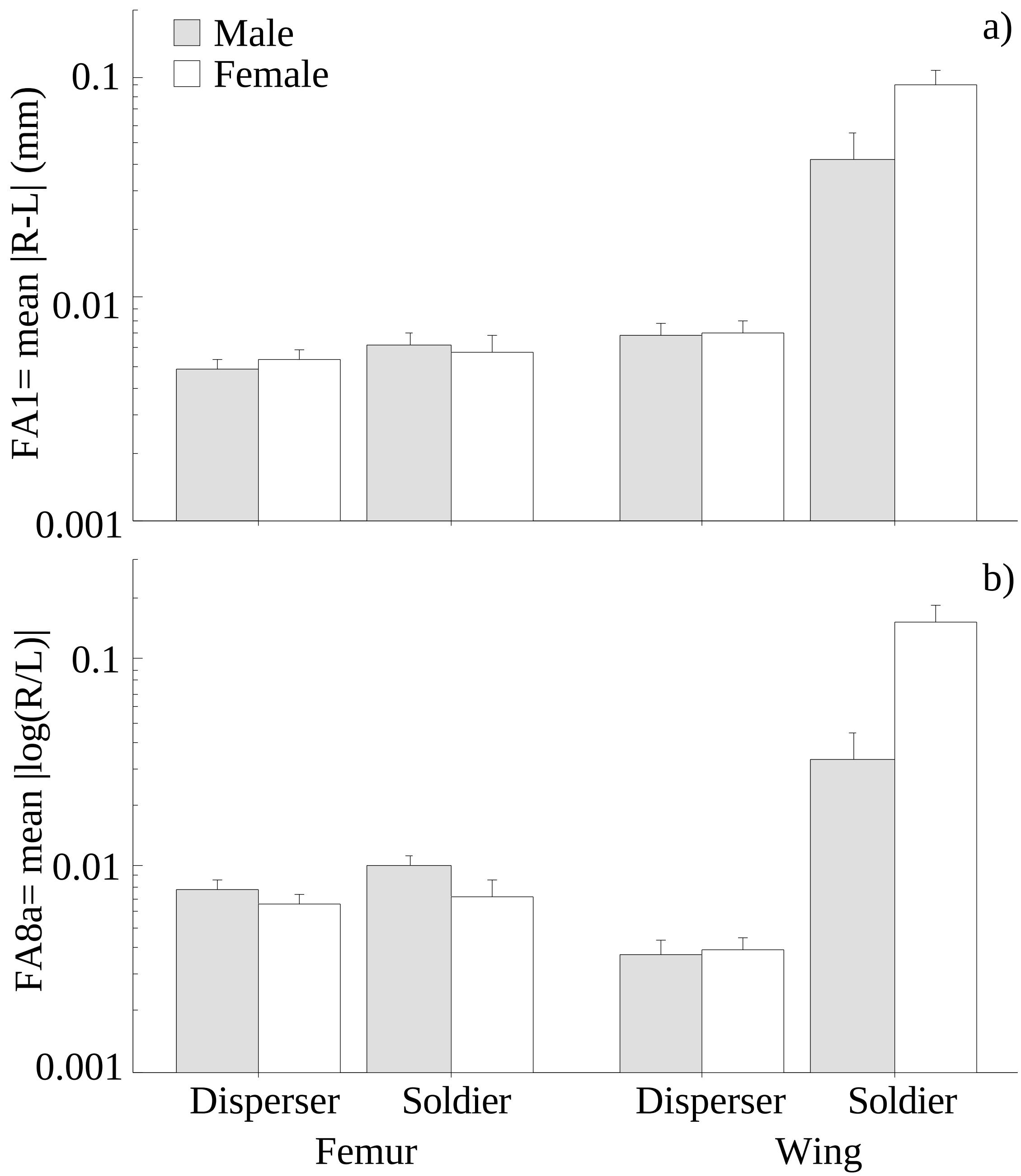

Figure V.6. (a) FA1 and (b) FA8a of femurs and wings, for both sexes and castes of Oncothrips tepperi. Note that these means include the effect of ME. All four ME and FA outliers (see Steps 2 and 5) were excluded from these analysis. These graphs were created from data in the supplementary data file: CrespiData_forFA.

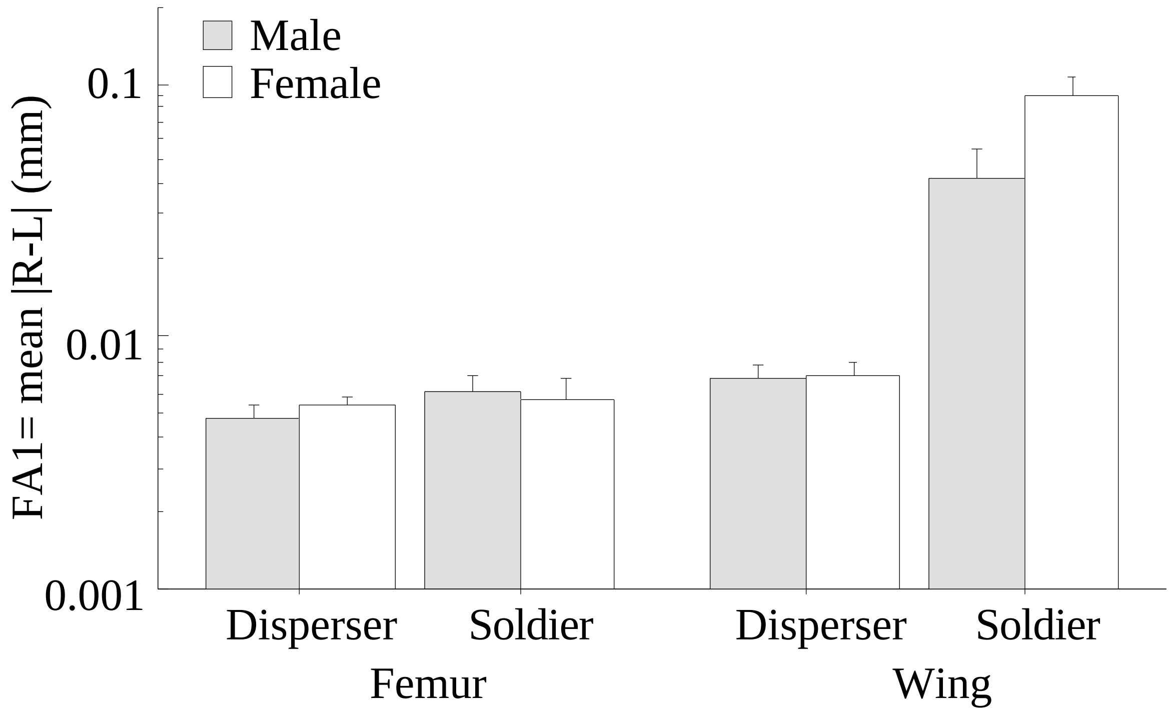

Figure V.7. Fluctuating asymmetry (FA1) of femurs and wings, for both sexes and castes of Oncothrips tepperi. Note that these means include the effect of ME. All four ME and FA outliers were excluded from this analysis.