INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI[®]
800-521-0600

NOTE TO USERS

Page(s) not included in the original manuscript are unavailable from the author or university. The manuscript was microfilmed as received.

82 AND 160

This reproduction is the best copy available.

UMI

**Assessing and Developing the Use of Fluctuating Asymmetry
as a Bioindicator of Environmental Stress
and Quality of Organisms**

by

Brian Leung

**A thesis submitted to
The Faculty of Graduate Studies and Research
In partial fulfilment of
The requirements for the degree of
Doctor of Philosophy**

Department of Biology

Carleton University

Ottawa, Ontario

May 1999

© 1999, Brian Leung

**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

395 Wellington Street
Ottawa ON K1A 0N4
Canada

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-42798-6

Canada

**The undersigned hereby recommend to
The Faculty of Graduate Studies and Research
Acceptance of the thesis:**

**Examining And Developing the Use of Fluctuating Asymmetry
as A Bioindicator Of Environmental Stress
and Quality of Organisms**

**Submitted by
Brian Leung**

**In partial fulfilment of the requirements
for the degree of Doctor of Philosophy**

Chair, Department of Biology

Thesis Supervisor

External Examiner

Abstract

Fluctuating asymmetry (FA) refers to subtle departures from perfect symmetry, and indicates an organism's inability to maintain precise development. FA has been related to putative stresses as well as to life history traits of individuals. Thus, FA has been advocated as a powerful, cost-effective general bioindicator of environmental stress and organismal quality. However, controversy surrounds the generality and strength of FA relations, whether FA should reflect organism-wide quality of individuals at all, and how FA should be analyzed.

In this thesis, I integrated statistical methods, mathematical modelling, and empirical studies to determine the usefulness of FA as a bioindicator of stress or predictor of organismal quality, and to develop more powerful methods to assay for FA-stress and FA-quality relations. I used meta-analytical methods and showed that FA-stress and FA-quality relations were likely real, but were typically weak. Thus, FA probably would not be reliable as a bioindicator of stress, unless methods could be found to increase the strength of FA relations. I constructed mathematical models to examine theoretical patterns of FA-stress and FA-quality relations. I presented mechanisms for why FA relations should be weak and I develop methodologies to improve the reliability of FA as a bio-indicator of stress. I used simulation models to examine methods of combining information across traits and to present methods that have the greatest power to detect non-spurious relations between FA and stress. I also used simulations to examine relations between FA and trait size (i.e., allometric relations). Allometry can obscure detection of relations between FA and stress and FA and quality. I showed that current methods to detect and control for allometric relations are inefficient, I presented methods to control optimally for allometry, and suggested problems that require further attention. I used methods derived above in

experimental and correlational studies. The methods provided increased power to the degree predicted. Still, I found only equivocal support for predicted relations between FA and quality. I concluded my thesis by discussing the limitations of FA as a bioindicator, and the circumstances under which FA may be used validly.

Acknowledgements

Where to start? I have interacted with many people during my degree, and it is largely this contact that has made my tenure as a Ph.D. student worth while and unique. Not only has it been intellectually stimulating to associate with the biology group at Carleton, but these interactions have been good for less definable reasons as well. I guess we are all human and social animals in the end.

I will begin by thanking my supervisor, Mark Forbes, with whom I have had a very interesting and productive degree. Whatever else can be said, no one can accuse us of not being productive, eh? I especially thank Linda Maddison, who has helped me find a little bit of peace and a place where I belong. I thank my parents John and Nora (same last name as mine), for their support. Dad, your advice and your ability to help me see more clearly has been invaluable. I guess we are not so different after all. I thank Lance Bishop, my friend and a fellow traveler on a similar path. Also, thanks for teaching me how to play the guitar. I would like to thank Rob Baker for his support and encouragement when I was working in Toronto. I thank my committee members, Lenore Fahrig, Scott Findlay, and Cam Wyndham, three very smart people who give me new insight into my thesis at each committee meeting. Lenore, I especially thank you for letting me to discuss some of my ideas with you when I first came to Carleton. I enjoyed talking to you. I thank my lab mates: Dean McCurdy, his wife Viv, Gina Schalk, her husband Mark Driedger (also thanks for the advice on math and statistics), Chris Yourth, Keiko Lui (thanks for allowing me to tease you...continuously), Kerry Hecker (whom I also thank for reviewing several of my manuscripts), and Sharleen Hawco. Dean and Chris, we've had more than a few good lab discussions, eh? I thank Pat Weatherhead's lab: Chris Parent (we've had some interesting philosophical discussions after beers), Kev Dufour (thanks

for the help on previous manuscripts. It was good to have someone with whom to discuss fluctuating asymmetry), Kent Prior, Kelley Kissner, Michelle Gunness, Gabriel Blouin-Demers, Sophie Sommerer. I thank Lenore Fahrig's lab: Darren Bender and Lutz Tischendorf (thanks for the discourse on computer programming), Michelle Lee (I guess I just like talking to you. I'm glad that you've been going to Friday beers regularly...these days), Tom Contraras (you are fairly soft spoken, but I'd probably end up listening to you over those more vocal in their opinions), Julie Brennan, Laurie Carr and Sheilagh Pope. I thank the group at U of T, specifically Che Elkin for all his help during my experiments. Finally, I'd like to thank Bruce Smith, John Jaenike (more than once), and Pierre Mineau for pre-reviewing various manuscripts. You all are the ones that I meant when I said that certain people had made my stay at Carleton worth while and unique.

Table Of Contents	
	Page
Abstract	iii
Acknowledgements	v
Table of Contents	vii
List of Tables	xiii
List of Figures	xiv
List of Appendices	xvi
General Introduction. Fluctuating Asymmetry: a Putative Bioindicator.	1
1. General Definitions and Concepts.....	1
2. Fluctuating Asymmetry: Background.....	4
3. Bioindicators and Advocated Advantages of Fluctuating	
Asymmetry	7
4. Controversies Surrounding Fluctuating Asymmetry.....	9
5. Directions and Objectives of Thesis.....	11
 Chapter 1 Fluctuating Asymmetry in Relation to Stress and Quality:	
Examination of Hypotheses Using Meta-Analysis.....	
1. Introduction.....	13
2. Methods.....	15
Meta-Analysis.....	15
Studies and Selection Criteria.....	17
Analytical Methods.....	18
Variables.....	20

Table Of Contents	
	Page
3. Results.....	21
General FA-Stress and FA-Fitness Relations.....	21
Effects of Traits and Thermomy on FA-Stress and FA-	
Quality Relations	22
4. Discussion.....	26
General FA-Stress and FA-Fitness Relations.....	26
Effects of Trait Type and Thermomy on FA-Stress and	
FA-Fitness Relations.....	27
 Chapter 2 Modelling Fluctuating Asymmetry in Relation to Stress and	
Quality	
1. Introduction.....	30
2. The Model.....	32
Equations and biological rationale.....	32
Simplifying Assumptions and Simulations.....	34
Setting Values of Each Term.....	35
Traits and FA Distributions.....	37
Random Nature of FA.....	37
3. Results.....	38
4. Discussion.....	47
Trait Types.....	47
FA Distributions.....	48
Random Nature of FA.....	50
5. Conclusions.....	52

Table Of Contents	
	Page
Chapter 3 Correcting for Allometry in Studies Relating Fluctuating Asymmetry to Quality Within Samples	
1. Introduction.....	53
2. Basic Model.....	54
3. Positive Size-Quality Relations: Current Rules.....	58
4. Alternative Rules to Control for Allometry.....	59
5. Inverse Size-Quality Relations.....	60
6. The Coefficient of Variation of Size.....	61
7. Discussion.....	67
8. Cautions and Other Theoretic Considerations.....	69
 Chapter 4 Fluctuating Asymmetry as a Bioindicator of Stress: Comparing Efficacy of Analyses Involving Multiple Traits.	
1. Introduction.....	72
2. Composite Fluctuating Asymmetry Analyses.....	73
3. Simulations.....	76
Baseline Simulations.....	76
Heterogeneous FA Distributions Between Traits.....	77
Leptokurtism.....	78
Correlations of FA Values Between Traits.....	80
Measurement Error.....	81
Number of Traits and Sample Size.....	81
4. Discussion.....	86

Table Of Contents

	Page
Recommendations.....	86
Mechanisms of Relative Benefits of CFA Indices	
Compared to Single Traits.....	89
Comparison of CFA Indices: Mechanisms of	
Differences in Power and Validity.....	90
FA-Quality Relations Within Samples.....	92
5. Summary.....	92
Chapter 5 Does Fluctuating Asymmetry Reflect Quality of Individuals	
or Samples: A Laboratory Study using a Damselfly (<u>Ishnura</u>	
<u>Verticalis</u>) Model System.	
1. Introduction.....	94
2. Methods.....	96
Field Collections.....	96
Experimental Setup.....	96
Data Collection.....	97
Analysis.....	98
Basic Statistics of Fluctuating Asymmetry.....	99
Fluctuating Asymmetry Versus Quality Among and	
Within Samples.....	100
Phenodeviants.....	101
3. Results.....	102
Basic Statistics of Fluctuating Asymmetry.....	102
Fluctuating Asymmetry Versus Quality.....	104

Table Of Contents

	Page
Phenodeviants.....	104
4. Discussion.....	113
Validity of Life History Traits.....	113
Relations between FA and Life History Traits.....	116
Implications of Results for FA as a Bioindicator.....	117
Chapter 6 Fluctuating Asymmetry in Relation to Indices of Quality in Damselflies (<u>Enallagma ebrium</u>) Captured in the Field.	
1. Introduction.....	120
2. Methods.....	121
Study Site.....	121
Field Collections.....	121
Variables.....	122
Basic Statistics of Fluctuating Asymmetry.....	123
Analyses.....	124
3. Results.....	125
Tests of Other Asymmetry Distributions and Assessment of Measurement Error.....	125
Relations Between FA, Wet Mass, Longevity and Mite Intensity.....	126
4. Discussion.....	130

Table Of Contents	
	Page
<hr/>	
Chapter 7 General Discussion: A Critical Assessment of Fluctuating Asymmetry as a Bioindicator.	
1. Introduction.....	134
2. Examining Fluctuating Asymmetry as a Bioindicator.....	135
3. Fluctuating Asymmetry in Perspective: A Summary of Limitations	151
4. Conclusion: The Utility of Fluctuating Asymmetry.....	153
Literature Cited.....	158
Appendices.....	174

List of Tables		
Table		Page
1-1	Effect size, standard errors, Z scores, number of species, and within group heterogeneity for FA-stress and FA-fitness relations.....	24
1-2	Results of tests for differences between average effect sizes of various groups.....	25
2-1	F-values for the correlation between FA of different traits within a sample of individuals.....	40
2-2	F-values for the correlation between FA and quality.....	41
3-1	Description of allometric rules used in this paper.....	62
5-1	Differences in life history traits (longevity and wet mass) between nutritional groups.....	106
5-2	Basic statistics associated with FA.....	107
5-3	Differences in CFA between nutritional groups.....	109
5-4	Relations between CFA and life history traits within samples and between FA of single traits and life history traits.....	110
5-5	Frequency of phenodeviants under different conditions of temperatures and feeding.....	111
6-1	Basic statistics associated with FA.....	127
6-2	Comparisons between life history traits (longevity and wet mass), intensity of mite parasitism, and CFA.....	129

List of Figures		
Figure		Page
2-1	The effects of one component of developmental stability, C_m , on strengths of FA-stress and FA-fitness relations.....	42
2-2	The effects of one component of developmental stability, C_a , on strengths of FA-stress and FA-fitness relations.....	43
2-3	The effects of one component of developmental stability, C_m , on FA variance and FA kurtosis.....	44
2-4	The effects of one component of developmental stability, C_a , on FA variance and FA kurtosis.....	45
2-5	FA versus quality showing a triangular bivariate distribution	46
3-1	Type I errors and power of Rules as underlying FA-size relations increased from no relation to isometry to $FA \propto size^2$, given positive relations between size and quality.....	64
3-2	Type I errors and power of Rules as underlying FA-size relations increased from no relation to isometry to $FA \propto size^2$, given positive relations between size and quality.....	65
3-3	Relations between coefficient of variation of size (CV), strength of FA-size relations, and Type I error rates of FA_u and FA_{rel}	66
4-1	Type I error rates and power for six CFA analyses and analyses using single traits under base line and heterogeneous conditions.....	83

List of Figures

Figure		Page
4-2	Type I error rates and power for six CFA analyses and analyses using single traits (ST) for three levels of kurtosis...	84
4-3	Type I error rates and power for six CFA analyses and analyses using single traits (ST) for three levels of correlations between FA values of traits.....	85
5-1	Differences in life history traits (longevity and wet mass) between phenodeviants and normal individuals.....	112
5-2	Flow diagram showing possible relations between the environment (stress), quality, FA, and life history traits.....	119
7-1	Idealized FA-quality response curve.....	157
H-1	Difference in power between CFA 2 and CFA 3.....	231
I-1	Asymmetry-quality plots within samples using a single trait and a composite measure combining 10 traits	232

List of Appendices

Appendix	Page
A Glossary of terms.....	174
B Data set showing source, correlate, species or organism, character used, trait types, therapy, statistic used, and sample size of each relation used in the meta- analysis.....	177
C General calculations used in the meta-analysis.....	205
D Parameters and terms used in Chapter 3.....	208
E Ranges of Type I error rates and power for Rules examined in Chapter 3.....	210
F Description of composite FA analyses (CFA).....	216
G Ranges of Type I error rates and power for analyses using single traits and CFA measures.....	219
H Estimated kurtosis values at which analyses using CFA 3 becomes stronger than CFA 2.....	225
I Asymmetry-quality plots comparing efficacy of FA of single traits to CFA 2.....	232

GENERAL INTRODUCTION

Fluctuating Asymmetry: a Putative Bioindicator

Environmental stress causes effects ranging from altered physiology, morphology, and biochemistry to species extinction. One area of research is aimed towards monitoring the presence and effects of stress on organismal quality and protecting populations of interest (e.g., \$133 million was spent on Marine monitoring programs in 1985 in the USA alone, Breckenridge and Olsen 1995). To this effect, researchers have been increasingly looking towards bioindicators as sensitive biologically relevant measures of stress or organismal quality (Huggett *et al.* 1992). Fluctuating asymmetry (FA) is a potential bioindicator that has been advocated to have many advantages over traditional monitoring techniques (e.g., Clarke 1993a, Sommer 1996). My thesis focuses on assessing and developing FA as a bioindicator. I divide my introduction into five parts:

1. I look at some general definitions and concepts (i.e., stress, quality, and fitness) that are integral to my thesis.
2. I define FA, define it and give background details.
3. I discuss the reasons why researchers have increasingly studied FA, and discuss the advocated advantages of FA as a bioindicator.
4. I discuss the controversies that surrounded FA at the time I began my thesis.
5. I present the direction and objectives of my thesis.

1. General Definitions and Concepts

Some of the terminology used in the FA literature is vague. In this section, I further discuss some of the more important and prevalent terms and concepts in

the FA literature to avoid subsequent misconceptions. Throughout the thesis, I also highlight each new term in bold and provide a glossary of these terms, for convenience of the reader. I define the term **stress** as any factor which negatively affects an organism such that physiology, morphology, or biochemistry is affected. If these changes relate to reduced current or future fitness or quality, or if they result in a change in population demographics (specifically population decline), then I consider the stress to be a biologically relevant stress. Stresses may either be natural in a system, or anthropogenic (e.g., nutritional deprivation and pollution, respectively).

I note that there are many possible biological definitions of stress. However, definitions typically infer that stress is some factor that has negative consequences for organisms. For example, Brett (1958) defined stress as a state produced by an environment or other factor which extends the adaptive responses of the animal beyond the normal range, or which disturbs the normal functioning to such an extent that the chances of survival are significantly reduced. Stress also has been variously defined as any condition that results in reduced growth, any condition that prevents an organism from realizing its 'genetic potential' (Begon et al. 1990), any environmental change which reduces the fitness of organisms (Koehn and Bayne 1989), or any factor that causes some response by the population of interest (Underwood 1989). Often, researchers use stressor to refer to a detrimental factor and stress to refer to the condition of an individual induced by that factor (e.g., Mayer et al. 1992). In my thesis, I use stress to refer to both the factor and condition induced by that factor to remain consistent with the FA literature.

A common view of stress is in terms of its energetic consequences for organisms (Koehn and Bayne 1989). Stress causes suborganismal reactions in organisms (e.g., changes in rate of protein synthesis, or allocation of resources).

This would be true regardless of whether the stress was general such as nutritional deprivation or specific such as a pollutant or a parasitic infection. For example, biochemical products would need to be produced to detoxify pollutants, immune response would need to be mounted to control the parasitic infection, resources would need to be allocated to repair damage caused by a stress that was not neutralized. These reactions require energy. As such, less energy would be available for other processes important for fitness such as growth or reproduction (Koehn and Bayne 1989). If a specific stress was intense enough, a systemic stress response may occur (i.e., the so-called general adaptation syndrome or GAS, Mayer et al. 1992, after Seyle 1976). After a certain level, exhaustion may occur, and decreased fitness or population declines may follow. An important point is that changes at the suborganismal level, in terms of biochemistry, physiology, histology and morphology, should precede drastic changes in life history traits of individuals or population demographics (Mayer et al. 1992). This is the logic underlying the use of bioindicators that will be discussed in the following section.

As mentioned, a biologically relevant stress should reduce the quality or fitness of organisms. Quality is a property of the individual(s) and is influenced by both genetic factors and the environment surrounding the individual(s). I illustrate the difference between quality and fitness using the following example: in a totally benign environment, differences in quality may have little relation to fitness. In contrast, in a stressful environment, only high quality individuals may be able to compensate for such stresses, while low quality individuals succumb. Thus, while quality may not relate to fitness under all circumstances, high quality individuals will on average have higher fitness given variable and often stressful environments. Similarly, a low quality population would be more likely to suffer population declines than a high quality one under a stressful environment. In

contrast, fitness is a relative measure within a population, and is in the strict sense not valid for comparisons between populations (Begon et al. 1990).

Although both of the terms quality and fitness are used in the FA literature, I use the term quality because FA is typically viewed as being affected by quality (see below). It is conceptually simpler to think of FA as being influenced by an inherent property of an individual (quality) rather than by some measured outcome (fitness). The term health also appears in the literature. Health has been defined as the residual capacity to withstand stress. The more stressed, the less capable is the organism of withstanding further stress (Mayer et al. 1992 after Bayne et al. 1985). The terms health and quality are synonymous.

I note that quality and fitness could not be measured directly. I was necessarily constrained to use life history traits as indirect measures. Typical life history traits include survival, growth, and reproduction (Mayer et al. 1992).

2. Fluctuating Asymmetry: Background

Fluctuating asymmetry (FA) refers to small directionally random deviations from perfect bilateral symmetry (Palmer and Strobeck 1986). These subtle departures from symmetry may either be viewed at the individual level, or at the population level. At the individual level, the absolute right minus left values are typical measures of FA. At the population level, FA refers to some measure of the variance of the asymmetry distribution. While some researchers have suggested different terminology to distinguish asymmetry at the individual and population level (e.g., Palmer (1994) suggested subtle asymmetry to refer to asymmetry at the individual level), I use the term FA for both levels to remain consistent with the literature.

FA is believed to result from two opposing processes termed developmental noise and developmental stability (Palmer 1994).

Developmental noise is random and non-genetic by definition and causes increases in FA magnitude by disrupting precise development (e.g., effects of thermal noise on enzyme activity, or small random differences in physiological processes between cells, Palmer 1994). Developmental stability, refers to an organism's ability to correct or reduce disruption during development and may have a genetic component (e.g., negative feedback system which regulates enzyme activity within and between cells, hormonal regulation of non-contiguous structures, Palmer 1994).

Stress may reduce the amount of energy available to be allocated to developmental stability, thereby compromising the ability of organisms to correct for subtle departures from symmetry (Mitton 1993, Campbell and Emlen 1996). Similarly, high quality individuals may be able to perform the processes underlying developmental stability better than low quality individuals (Møller 1993a, Polak and Trivers 1994). Thus, stressed or low quality individuals should have higher FA on average (e.g. Clarke 1993a, Møller 1993a). Given relations between FA and quality and FA and stress, FA has been advocated as a bioindicator (e.g., Leary and Allendorf 1989, Clarke 1995), and as a general measure of quality (Leary and Allendorf 1989, Polak and Trivers 1994). Note however, that FA is determined during development. Thus, there is an implicit assumption that stress and quality during development, reflect organismal quality after development is completed. Hypothesized relations between FA and post-developmental quality appear to be supported by inverse relations between FA and post-developmental life history traits such as survival (e.g. Naugler and Leech 1994) and mating success (e.g. Harvey and Walsh 1993). Hypothesized relations between FA and stress appear to be supported by positive relations

between FA and environmental stresses such as pollution (e.g. Clarke 1993a) or genetic stresses such as homozygosity (e.g., Leary et al. 1985).

We note that FA is not the only type of asymmetry. Other forms of asymmetry include directional asymmetry and antisymmetry (Van Valen 1962). **Directional asymmetry** refers to the situation in which there is a genetic predisposition towards one side being larger (e.g., in the human heart, the left side is typically larger than the right side). **Antisymmetry** refers to the situation in which there is a genetic predisposition towards one side being larger, but it does not matter which side is larger (e.g., male fiddler crabs have one large claw and one small one. However, either side can be larger, resulting in a bimodal asymmetry distribution). Directional asymmetry can be assayed for if the mean asymmetry differs from zero (Palmer and Strobeck 1986). In contrast, because FA is directionally random, the mean asymmetry should be zero. Antisymmetry is indicated by significant platykurtosis (Palmer and Strobeck 1986). Platykurtosis refers to a 'broad peaked' distribution with more data falling at intermediate values compared to a normal distribution (Zar 1984).

Researchers have questioned the validity of using these other forms of asymmetry, because genetic predispositions play a role in causing asymmetry (Palmer and Strobeck 1992). Therefore, with these other forms of asymmetry, we can not be certain what is the optimal phenotype. Thus, the consequences of stress or quality on the degree of asymmetry may be confounded by genetic effects. In contrast, for FA, the optimal phenotype is assumed to be perfect symmetry. Departures from perfect symmetry reflect an organisms inability to maintain homeostasis (Palmer and Strobeck 1986). We note however, that there is controversy regarding the validity of these other patterns of asymmetry for bioindicators of quality or stress (e.g., Graham et al. 1993).

3. Bioindicators and Advocated Advantages of Fluctuating Asymmetry

In certain situations, it may be advantageous to use **bioindicators** rather than directly measuring an environmental stress. For example, the degree of stress experienced by organisms may be difficult to measure in the field (e.g., the degree of food availability and nutritional stress). For anthropogenic stresses, it may be possible to measure the concentration of a given contaminant. However, such chemical analyses would not take into account exposure of biological organisms to this contaminant nor its effect on biota (Melancon 1995). Even when contaminants exist in the environment, they are not necessarily available to biological organisms. Even when they are assimilated by the biota, the contaminants are not necessarily toxic (Peakall 1992, Melancon 1995). Since we are primarily interested in biological consequences of stress, it seems reasonable to examine biological organisms.

Given the wide range of effects stress has on biological organisms, it should not be surprising that a wide range of bioindicators exist. Different bioindicators may indicate stress at different levels of biological organization and may serve different purposes. Most research has been conducted at the population, community, or ecosystem level (e.g., population decline or changes in species composition) (Mayer et al. 1992). However, detecting changes at this level may not be optimal, since damage would already have occurred, possibly irreversibly (Mayer et al. 1992, Bunn 1995). Ideally, we should detect stresses and take remedial actions before they have a major irreversible effects on populations of interest (Bunn 1995).

Bioindicators also include biochemical, physiological, or morphological responses to stress (also termed biomarkers, Melancon 1995). Suborganismal effects should occur before stress has consequences at higher levels of

organization. Thus, suborganismal responses should be much more sensitive to stress than changes in populations, and we should be able to detect stresses much earlier and have greater success in remedial actions by examination of the effects of stress on lower levels of organization. However, even at lower levels of organization, there are many possible bioindicators, each with different uses and limitations. To use these bioindicators effectively and appropriately, their purposes and limitations must be understood (Huggett et al. 1992).

One putative bioindicator, and the topic of this thesis is FA. FA has been argued to have advantages over traditional monitoring methods (Clarke 1993a). As with other suborganismal bioindicators (Mayer et al. 1992), FA may be more sensitive than population level indicators of stress that are most often used. Further, FA may have fewer fitness consequences than other life history traits such as size or fecundity. Thus, as energy becomes less available (i.e., due to the presence of stress), organisms may choose to allocate more energy to processes important for fitness or survival compared to developmental stability. As a result, developmental stability may be compromised before traditional life history measures. Thus, FA may be more sensitive to stress than other life history measures, and more effectively allow remedial action (Clarke 1993a).

In contrast to other methods of biomonitoring (Bunn 1995), the biological relevance of FA appears to be clear. FA has been found to relate to quality as measured by life history traits (e.g., Harvey and Walsh, 1994). Thus, differences in FA between populations should reflect differences in quality (Polak and Trivers 1994, Clarke 1995, Sommer 1996), and we might avoid the "so-what" question associated with other biomonitoring techniques (Mayer et al. 1992, Bunn 1995). Further, if FA reflects quality, then FA should reflect the synergistic interactions between stresses (Clarke 1993a). Given that multiple stresses likely exist in nature, such synergisms are necessary to consider (Adams et al. 1998). Other

bioindicators may have laboratory derived data that relates dose to effect, but this may not be representative of the potential synergistic effect which may occur in nature. Also, if FA reflects the general quality of individuals, then it may also be used as a convenient surrogate of quality (Polak and Trivers 1994). Thus, FA has been argued to be useful in conservation biology or breeding programs where it is important to identify and breed high quality individuals (Leary and Allendorf 1989).

FA may provide logistical advantages as well. First, FA most often consists of simple morphological measurements. As such, FA is relatively cheap and easy to measure (Clarke 1993a), and does not require animals to be killed. Further, FA can be measured on essentially any organism that has bilateral traits. Therefore, FA has the potential to be widely applicable and can be directly measured on populations of interest rather than on test organisms. Given the ease of measurement of FA, it is not surprising that relations between FA and putative stresses and quality measures have been extensively studied in the field (unlike many other bioindicators).

4. Controversies Surrounding Fluctuating Asymmetry

While some researchers have argued strongly for the utility of FA as a bioindicator, their reviews were qualitative (e.g., Leary and Allendorf 1989), and the papers cited may have been chosen because they supported an advocated view, rather than because they were typical of the literature. In fact, some researchers have found no relations between FA and fitness traits (e.g. Eggert and Sakaluk 1994), or FA and stress (e.g. Patterson and Patton 1990; Owen and McBee 1990). Others have found confounding positive relations (predictions are for negative relations) between FA and factors such as mate choice and survival

(e.g. Packer and Pusey 1995). Still other researchers have found that, within studies, significance of FA-stress and FA-fitness relations often differ between traits or samples (e.g., Blanco et al. 1990; Polak 1993). Thus, it is possible that the significant FA-stress and FA-fitness relations observed may be an artifact of the large number of tests that have been conducted. Alternatively, the non-significant relations may reflect low power (Palmer 1994). Clearly, a quantitative analysis of the literature would be useful to examine the generality of FA-quality and FA-stress relations.

The apparent heterogeneity in the literature has spurred controversy regarding which types of traits and which types of organisms should have the greatest sensitivity to detect FA-quality or FA-stress relations (e.g., Møller 1993a, Novak et al. 1993). Perhaps this is not surprising given that much of the theory surrounding FA also has been based on verbal arguments. For example, it has been argued that a FA distribution should be normally distributed, centered around zero (Palmer and Strobeck 1986). Deviations from normality or from a mean of zero should indicate genetic predispositions for asymmetry, and asymmetry of such traits may not validly measure developmental stability, quality, or stress (Palmer and Strobeck 1992). Mathematical models may allow the logic underlying these verbal arguments to be assessed. While mathematical models do not prove that a theory is correct, they can be used to determine logical end points given a set of assumptions. In contrast, it is sometimes difficult to determine the logical end point of a given set of assumptions, using verbal models.

Finally, there has been a wide range of methods used to analyze patterns of FA (see Palmer and Strobeck 1986, 1992 for reviews). Questions remain regarding which methods should be used. If FA is to be used as a sensitive bioindicator, it is critical to assess possible confounding factors associated with

FA, as well as to determine the most reliable means of detecting relations between FA and stress and FA and quality.

5. Direction and Objectives of Thesis

The objectives of this thesis are to 1) Assess support for underlying relations between FA and stress and FA and quality. 2) Examine the controversies, limitations, and problems associated with use of FA 3) Develop methodology to ameliorate some of the problems. 4) Finally, come to some conclusion concerning the applicability of FA as a biomonitor of stress or quality. In my thesis, I study FA in several ways. I assess the existing literature, use theoretical models, experimental laboratory studies, and correlational field studies.

In chapter 1, I conduct a quantitative literature review (meta-analysis) to examine the empirical evidence surrounding some of the above controversies.

In chapter 2, I develop a computer simulation model to critically examine the logic behind existing FA hypotheses and, using current assumptions underlying FA theory, explain and predict patterns that should be observed in nature.

In Chapter 3, I examine the problem of allometry. Larger traits may have more opportunity to become asymmetric. In such a case, FA would not necessarily reflect quality, but could simply reflect differences in size. I use simulation models to examine methodology to assess and control for allometry.

In Chapter 4, I examine methods to combine FA across traits. Researchers have argued that assessing FA across several traits yield better measures of developmental stability compared to analyses based on single traits (e.g., Leary and Allendorf 1989). Several methods to combine traits have been used or

presented (e.g., see Palmer 1994). However, quantitative comparison of their relative power and validity is lacking.

In Chapter 5, I conduct an experimental study, raising organisms (the damselfly Ishnura verticalis) under different environmental conditions. I examine the ability of FA to reflect life history traits, within and between samples.

In Chapter 6, I conduct a study on organisms collected from the field, and examine correlations between FA and life history traits.

Finally, In Chapter 7 (general discussion), I synthesize the work in this thesis, put it into the context of the general literature, and come to some conclusion concerning the utility of FA as a bioindicator.

CHAPTER 1

Fluctuating Asymmetry in Relation to Stress and Quality: Examination of Hypotheses Using Meta-Analysis.

1. Introduction

As mentioned in the general introduction, there has been controversy over whether fluctuating asymmetry is related to stress during development, or to quality, of organisms. Many researchers have found predicted relations between FA and stresses such as inbreeding and pollution (Clarke 1993a; also see Palmer and Strobeck 1986 for a review), or between FA and quality traits such as survival, competitive ability, and mating success (e.g., Ueno 1994). However, some researchers have found no relations between FA and either stress or quality measures (e.g., Owen and McBee 1990 and Eggert and Sakaluk 1994, respectively), whereas others have found relations opposite to those predicted (such as positive relations between FA and mate choice, survival, or parasitism: e.g., Packer and Pusey 1995). Furthermore, FA-stress and FA-quality relations can differ between samples, or between traits, of organisms (e.g., Blanco *et al.* 1990). Taken together, these results suggest that character FA may not consistently index stress or quality.

To account for these discrepancies, researchers have discussed which traits or which organisms should provide the strongest FA-stress or FA-quality relations. For example, researchers have argued that **sexually-selected traits**, which are under directional selection for increased size, should be quite susceptible to disruption by stress during development (Møller 1993a; Manning and Chamberlain 1994; Hill 1995). Also, researchers have reasoned that only

high quality individuals should develop large and symmetrical sexual characters (Møller 1993a; Watson and Thornhill 1994). In contrast, traits under stabilizing selection should not vary much in FA and may remain stable regardless of individual quality, even under stressful developmental conditions (Møller 1993a; Hill 1995).

Other researchers have suggested that FA of performance traits should relate to quality, due to direct impacts of asymmetry on the functioning of such traits. A corollary of this idea is that FA of non-performance traits may not relate to quality (Eggert and Sakaluk 1994; Palmer 1994). Note that this argument makes no stipulations about differences in FA-stress relations between performance and non-performance traits.

Finally, some researchers have hypothesized that FA-stress or FA-quality relations should be weaker in homeotherms than in poikilotherms, because homeostasis of homeotherms should lessen susceptibility of individuals to stresses (Wooten and Smith 1986; See Novak et al. 1993 for a review).

At the time that I conducted this study, most of the above arguments were based on qualitative examinations of the literature. In this study, I quantitatively assessed the literature. I used meta-analysis (Rosenthal 1991; Arnqvist and Wooster 1995) to test whether FA-stress or FA-quality relations were non-spurious, given the large number of relations examined. In fact, many papers have published both significant and non-significant relations. I next determined whether the tests in the literature supported specific hypotheses with respect to trait type and thermy, as factors influencing the strength of FA-stress or FA-quality relations.

2. Methods

Meta-Analysis

Meta-analysis provides a method for looking for general relations between variables (Rosenthal 1991; Arnqvist and Wooster 1995), provided hypotheses of interest are similar between studies (e.g., FA relates negatively to quality). Meta-analysis estimates the strength of relations and takes into account type II errors by synthesizing different statistics across many studies. For meta-analysis to be effective, it must be used properly and its limitations must be acknowledged. First, meta-analysis is limited by the quality of primary studies collated, as in any synthesis study. Thus, there should be an explicit non-subjective criteria by which studies are excluded. Second, the literature collated may produce bias if papers included are not representative of existing literature. This may occur if citation lists from other papers are used, as papers generally cite references that show phenomenon of interest (e.g., significant FA-stress relations). A more unbiased approach would be to use a CD-ROM search, which also would likely provide a more comprehensive search than relying on citations or randomly searching through journals. A third problem is dependence of data points, both statistical and biological. Statistical dependence may occur when multiple measures (e.g., quality traits) are taken from the same individual(s). Biological dependence may occur when similar organisms are used in several studies (i.e., different tests on the same species). Finally, publishing bias may occur, whereby significant results are published more often than non-significant results. Although we can not eliminate this problem, we can determine the number of unpublished non-significant relations that must exist for published significant results to be explained by type I errors ($\alpha = 0.05$).

Some other common criticisms of meta-analysis are that different statistics are combined across studies, study conditions often differ, and other unknown factors can strongly influence relations in different studies. There is an extensive literature concerning methods to combine different statistics. Typically, various statistics have algebraic relations to one another and can be transformed into a common currency. Parametric and non-parametric measures also are comparable (e.g., Pearson's and Spearman's r), once transformed into a common currency (Glass et al., 1981).

I note that use of different statistics, different study conditions, or the presence of unknown factors may increase heterogeneity (variation) of effect sizes. However, this does not imply that meta-analysis should be restricted to studies that use identical statistics and study conditions. To do so would greatly reduce the sample size, resulting in reduced power of the meta-analysis. Furthermore, such restrictions are unnecessary given that statistical methods exist that take into account unexplained heterogeneity (see below). However, if specific factors are suspected of influencing effect size, subdividing relations into appropriate categories may allow specific hypotheses to be tested (e.g., poikilotherms have stronger FA-stress relations than homeotherms). However, subdividing relations into all possible categories in the absence of a priori hypotheses would increase the probability of making a type I error. A final criticism is that collating relations across studies results in losing detail from individual studies. While this concern is understandable, any synthesis necessarily sacrifices detail of specific studies in order to find general relations [see Hedges (1994a) for an additional discussion of meta-analysis].

Studies and Selection Criteria

I found 59 papers from 1985-1995 that were appropriate for my analyses, largely by using a CD-ROM search of Biological Abstracts (BIOSIS). These studies reported a total of 397 FA-stress and FA-quality relations. A list of these studies, the species examined, the sample sizes, and the relations between character FAs and various stresses or quality measures, is presented in Appendix B.

Although many other papers were collated, these studies had to be omitted because they did not meet my selection criteria. I excluded relations (and entire studies) if traits showed either directional asymmetry or anti-symmetry, given the controversy that surrounds the validity of these forms of asymmetry as measures of developmental stability (Palmer 1994, Graham *et al.*, 1993). Papers were included only if researchers examined correlates of quality (e.g., body size, dominance, ornament size, reproductive success, survival). Studies reporting stresses such as inbreeding and pollution were included; these stresses were assumed to be biologically relevant. I excluded relations within studies that did not provide adequate information (i.e., strength and direction of relations). To avoid bias, I excluded entire studies if they only provided adequate information for significant results, but not for non-significant results.

I also excluded three other types of studies: studies on hybridization (because hybridization could result in either an increase or decrease in quality, reviewed by Palmer and Strobeck, 1986), studies on competitive interactions between species differing in their FA measures (because differences in competitive ability could be due to FA effects, species effects, or both factors) and two studies on plants (because specific hypotheses regarding trait type and thermy primarily dealt with animals). If studies presented several measures of

FA, I used the index most common between studies (i.e. [right-left] values). Otherwise, I used the measure of FA presented by the researcher.

Analytical Methods

I calculated the overall effect size and compared effect sizes between different groups (following methods of Shadish and Haddock 1994). I used a random effects model (as opposed to a fixed effects model) because there were reasons (see results) to believe that effect sizes were more variable than expected due to just sampling error [see Hedges (1994a) and Raudenbush (1994) for detailed discussions of differences between the two models].

To combine the wide range of studies and statistics, I converted all results into a common effect size following Shadish and Haddock (1994) and as detailed below. When means and standard errors were given, I used these values to perform a Levene's test to compare FA variances. This was done for consistency, in preference to using the statistic described in each study.

Levene's test is powerful for detecting relations and robust to deviations from normality (Palmer and Strobeck 1992). I only compared results between 2 groups being tested, so that I could easily define the direction of any relation. When more than 2 groups were present, the ANOVA test was collapsed into 2 groups (Glass et al., 1981). For example, control group(s) were compared with combined experimental group(s). If groups differed in putative stress levels, I compared the group with the lowest level of stress versus that with the highest level of stress.

Student t-values from Levene's test, F-values from ANOVAs, chi-square values, and Spearman correlation coefficients were all transformed into Pearson's r (equations given in Glass et al., 1981). When direct transformations were not available, but p-values were given, p-values were transformed into the

normal deviate Z score and then transformed into Pearson's r (Rosenthal, 1991). Binomial probabilities also were transformed first into Z scores (equation in Snedecor and Cochran 1980). From the 397 relations, 180 relations provided correlation coefficients directly, 124 provided means and standard errors (or standard deviations) for calculation of Levene's test, 65 provided t-values or F-values from ANOVAs, 20 were based on p-values, 5 provided chi-square values, and 3 provided binomial probabilities (Appendix B).

To take into account the confidence that I had in each test, effect sizes were weighted by variance (Appendix C). Sampling variance is a function of sample size; larger sample sizes result in better estimates of the true effect size (Shadish and Haddock 1994). To eliminate both statistical and biological dependence of data points, which may bias the meta-analysis (Hedges 1994a), I weighted by the inverse of the number of correlations per species (Appendix C). Thus, each species by trait type combination contributed a single data point in the meta-analysis. This procedure was similar to Steinkamp and Maehr (1983) who weighted by the number of correlations in each independent sample.

I estimated the 'fail safe' number following Hedges and Olkin (1985, pp305-306). The fail safe number is essentially the minimum number of unreported non-significant relations needed for the results of the meta-analysis to reach the type I error rate ($\alpha = 0.05$). In this paper, species were used as independent data points rather than studies. Thus, I calculated the minimum number of species needed to have been examined with respect to FA-stress or FA-quality relations and not reported, for any significant results in this paper to be due to chance. This fail safe number represents a conservative estimate, as many significant results were excluded following the criteria listed above. As well, the fail safe number assumes that additional relations have zero contribution to

the strength of the meta-analysis (relations in the predicted direction would improve the significance of the meta-analysis).

Heterogeneity of effect sizes was examined using the heterogeneity statistic, Q_w (Shadish and Haddock 1994; Appendix C). This test statistic was calculated to determine whether variation in effect sizes was more than expected by random sampling error. The presence of significant heterogeneity would indicate that a random effects model should be used for subsequent analysis (Hedges 1994a). The difference between fixed and random effects models is that fixed effects models assume that all studies estimate the same effect size, whereas random effects models assume that effect sizes may differ between studies due to unknown factors. I then used the Z test to determine whether there was a true relation between FA and stress or FA and quality (Appendix C). Differences between the average effect size of trait types were compared to one another using contrasts in an analogous fashion to parametric statistics (Appendix C). For ease of interpretation, I backtransformed Fisher's z to Pearson's r , after manipulations were completed.

Variables

I separated FA-stress and FA-quality relations, as researchers have suggested that these relations should be different (Palmer 1994). I first compared FA-stress and FA-quality relations using sexually-selected traits versus those using traits that were not sexually-selected. As mentioned, sexually-selected traits are believed to be less developmentally stable and include traits such as antlers, and ornamental tail feathers. Examples of traits that are not sexually-selected are gill raker numbers and most insect wings (Appendix B).

Next, I examined the effects of performance and non-performance traits on FA-stress and FA-quality relations. In the FA literature, performance traits typically refer to traits in which asymmetry may mechanically influence functioning of that trait. My categorizations of traits as performance included traits that directly influence movement. Other traits were classified as non-performance, as it would be difficult to perceive asymmetry mechanically influencing their functioning (Appendix B). For example, tail feathers may be considered a performance trait because size asymmetry may influence flight performance, thereby influencing male quality (Palmer 1994). In contrast, chest plumage coloration would not be a performance trait. Asymmetry of chest plumage may influence mate choice by females (Swaddle and Cuthill 1994a), but would do so via other mechanisms than mechanically influencing male quality. Although there is little empirical evidence concerning which traits experience reduced mechanical efficiency with increased asymmetry, I believed it timely to address this issue, given current hypotheses (see Palmer 1994).

3. Results

General FA-Stress and FA-Quality Relations

Of the 397 relations, 246 were FA-stress and 151 were FA-quality relations. When I weighted correlations by species, I produced 37 independent FA-stress relations (for 18 homeotherms and 19 poikilotherms) and 61 independent FA-quality relations (for 42 homeotherms and 19 poikilotherms). Species used in this study are listed in Appendix B. When I further separated relations by trait types within species, different sample sizes again were generated (Table 1-1). These

sample sizes were not necessarily lower as a single species could contribute two data points to an analysis, if two types of traits were considered.

Overall, I found that FA was significantly related to both stress and quality in the predicted directions (i.e., positive FA-stress and negative FA-quality relations; Table 1-1). Within group heterogeneity (Q_w) was significant for all groups, except for FA-stress relations based on sexually-selected traits (Table 1-1). Thus, there was typically variation in effect sizes beyond that expected due to just sampling error. As indicated, this required use of a random effects model for analyses.

Next, I examined the required number of additional non-significant results to exist and remain unpublished for FA-stress and FA-quality relations reported in Table 1-1 to reach the type I error rate ($\alpha = 0.05$). As a conservative estimate, I found that an additional 176 species and 435 species would have had to been examined and found to produce non-significant FA-stress and FA-quality relations, respectively, for the results reported herein to have resulted from chance.

Effects of Traits and Thermy on FA-Stress and FA-Quality Relations

FA of both sexually-selected and traits not under-sexual selection were significantly related to stress and quality in the predicted directions (Table 1-1). Furthermore, no differences in effect sizes between these trait types were apparent suggesting that, in general, FA of sexually-selected traits did not provide a better measure of stress or quality than non-sexually-selected traits (Table 1-2). FA-quality relations of both performance and non-performance traits also were significant (Table 1-1). The significance of FA-quality relations in non-performance traits suggested that FA related to quality even when FA magnitude

was unlikely to influence functioning of that trait. Finally, average effect sizes were significant and statistically indistinguishable for both homeotherms and poikilotherms (Table 1-1, 1-2). Thus, the literature did not support the hypothesis that FA should be more strongly related to stress or quality for poikilotherms.

Table 1-1. Effect size (Pearson's r backtransformed from Fisher's z), standard errors of average effect sizes (SE), Z scores, number of species (# spp), and within group heterogeneity (Qw) for FA-stress and FA-quality relations. Summary data are grouped for all relations combined (ALL), and separated by trait type and thermy. P=performance traits, NP=non-performance traits, S=sexually-selected traits all groups, NS=non-sexually-selected traits all groups, PO=poikilotherms, HO=Homeotherms. Normal deviate Z scores above 1.645 (or below -1.645) indicate that the average effect size is significantly different from zero for one-tailed tests ($\alpha = 0.05$). Predictions were for positive FA-stress and negative FA-quality relations. Qw follow a chi-square distribution (Degrees of freedom=# Spp-1). Significant results in the predicted direction are indicated by an asterisk.

	r	SE	Z	# Spp	Qw
FA versus Stress					
ALL	0.17	0.04	3.95*	37	180*
P	0.15	0.05	3.15*	26	81*
NP	0.18	0.06	3.22*	23	120*
S	0.18	0.04	4.09*	11	12
NS	0.16	0.05	3.79*	37	181*
PO	0.21	0.06	3.36*	18	124*
HO	0.13	0.06	2.01*	19	54*
FA versus Quality					
ALL	-0.26	0.06	-4.69*	61	369*
P	-0.23	0.07	-3.32*	40	241*
NP	-0.27	0.08	-3.30*	26	155*
S	-0.27	0.08	-3.37*	46	251*
NS	-0.26	0.07	-3.56*	18	135*
PO	-0.27	0.07	-3.94*	19	83*
HO	-0.25	0.08	-3.40*	42	262*

Table 1-2. Results of tests for differences between average effect sizes of various groups. Group labels and sample sizes are the same as Table 1-1. For example, S versus NS compares average effect sizes of sexually-selected versus traits not under sexual selection. Contrast values follow a chi-square distribution with one degree of freedom ($\chi^2_{0.05,1} = 3.84$).

	GROUPS COMPARED	χ^2	Result
FA vs Stress	S vs NS	0.099	NS
	PO vs HO	0.78	NS
FA vs Quality	S vs NS	0.0070	NS
	PO vs HO	0.45	NS

N.B. There was no a priori prediction for differences between performance and non-performance traits for FA-stress relations. The prediction for FA-quality relations was that relations should occur only in performance traits (see Table 1-1 for further information).

4. Discussion

General FA-Stress and FA-Quality Relations

The results of the meta-analysis show that FA-stress and FA-quality relations exist; however, relations were highly variable and weak (average effect size was $r=0.17$ for FA-stress relations and $r=0.26$ for FA-quality relations). In other words, on average, a researcher would find a relation that explained only 2.8% and 6.7% of the variation (i.e., r^2 values) for FA-stress and FA-quality relations, respectively. Hence, FA may be a poor general predictor of either stress or quality, at least as currently measured. The reasons for variable and low effect sizes are not clear. Suggested problems typically include inadequate sample size, low power of tests, and measurement error (Palmer 1994). However, meta analysis takes power into account and effect sizes are not dependent upon sample sizes (Armqvist and Wooster 1995). Measurement error, on the other hand, may be especially important for FA studies, given the small magnitude of FA measures (Palmer 1994). The effect of measurement error would be to weaken relations and reduce effective sample size (Palmer 1994). Reductions in effective sample size and power have much less effect in meta-analysis than in primary studies, given that meta-analysis combines sample sizes across many studies. Thus, perhaps it is not surprising that I found significant FA-stress and FA-quality relations even though the primary literature appears equivocal. Notwithstanding, measurement error would reduce the average strength of relations, and may account partially for the low average effect sizes observed. Clearly, measurement error is important to take into consideration when researchers conduct primary studies.

Other possible reasons for the variable and low effect sizes include: poor choice of stresses and quality traits; and/or differential sensitivities of FA of traits and (or) organisms to stresses and quality (e.g., Palmer 1994; Hill 1995). Most studies did not directly validate stress or quality measures for their study organisms (which is particularly important when no FA-stress or FA-quality relations are found, e.g., Owen and McBee 1990; Eggert and Sakaluk 1994). I suggest that future researchers explicitly explore the biological relevance of stresses or quality measures that they monitored.

Effects of Trait Type and Thermity on FA-Stress and FA-Quality Relations

As indicated, specific hypotheses exist that suggest that FA of sexually-selected traits should be more strongly related to stress and quality than other traits (Møller 1993; Hill 1995); that FA-quality relations are a result of direct consequences of asymmetry on performance of the chosen character and that such relations may not be found in non-performance traits (Eggert and Sakaluk 1994; Palmer 1994); and that FA relations should be stronger in poikilotherms than in homeotherms (Wooten and Smith 1986; see Novak et al., 1993 for a review). I deal with each of these hypotheses in turn.

In general, FAs of sexually-selected traits were not more strongly related to stress or quality than other traits. There is always the possibility that unknown factors confounded differences (e.g., better quality measures may have been used when FA was obtained from non-ornamental traits). However, at present, the literature does not support the contention that FA of sexually-selected traits is better for indexing stress or quality. I note that, with respect to FA-stress relations, 9 of the 11 data points (species) were found in a single study (Møller

1995a). Thus, there is still need for more studies on FA-stress relations, using ornamental traits.

There also was little evidence that FA-quality relations were due to direct mechanical impacts of FA on quality (cf. Eggert and Sakaluk 1994; Palmer 1994). However, I note that classifying performance and non-performance traits was difficult (Appendix B). Yet, I believe that non-performance traits reasonably reflect traits in which asymmetry is unlikely to have a mechanical impact, because I conservatively categorized traits that could possibly be influenced by FA as performance traits. In fact, given the typically small magnitude of FA, I would expect that FA rarely influences performance mechanically, even in locomotory traits. Given this consideration, my results suggest that FA can relate to quality beyond direct impacts of FA on performance. Therefore, FA of some traits could reflect some underlying measure of individual quality [see Møller (1993a) and Palmer (1994) for the two opposing views]. However, FA may still remain a poor predictor of individual quality, due to the low average effect sizes (this study) and (or) the random nature of FA (Palmer 1994). Finally, my study indicated that presence or absence of internal regulation (or homeostasis in homeotherms) did not relate to the strength of FA-stress or FA-quality relations.

In conclusion, although FA relations appear to be real, FA does not appear to be a sensitive measure of either stress or quality, at least given current methods used in studies. Further, there is currently little evidence that broad groupings of traits (e.g., sexually-selected and performance traits) or of organisms (e.g., poikilotherms) are useful for predicting the strength of FA-stress or FA-quality relations. However, it is clear that there was considerable heterogeneity of effect sizes. Thus, it is always possible that some other grouping variable may be predictive of FA-stress and FA-quality relations. If such grouping variables were identified, additional meta-analyses may be appropriate to

examine the empirical support for new hypotheses. Finally, it is still possible that FA of specific traits or of specific organisms may reliably relate to quality and stress, whereas other traits will always be poor predictors; such reliable traits may have to be determined individually via direct experimentation.

CHAPTER 2

Modelling Fluctuating Asymmetry in Relation to Stress and Quality.

1. Introduction

As mentioned in the general introduction, controversy abounds in the FA literature. For example, some asymmetry distributions may reflect genetic pre-dispositions for asymmetry. It has been argued that skewed distributions may arise due to directional asymmetry, platykurtic distributions may arise due to antisymmetry, and leptokurtic distributions may arise due to a combination of antisymmetry and true FA (Palmer and Strobeck 1992; Palmer 1994). Thus, for these distributions, larger asymmetry values may not be indicative of higher stress or lower quality (Palmer and Strobeck 1992, but see Graham *et al.* 1993). While most researchers exclude skewed and platykurtic asymmetry distributions from analyses, leptokurtic asymmetry distributions are observed and used commonly (e.g., Harvey and Walsh 1993; Polak 1993). One question is whether leptokurtic asymmetry distributions reflect genetic pre-dispositions for asymmetry and thus should be avoided, or whether the developmental processes underlying true FA distributions (i.e., developmental noise and developmental stability) also might result in leptokurtic asymmetry distributions.

Another source of controversy is whether FA should reflect quality at both the population and individual levels. Some researchers argue that while FA may accurately index stress to a population, FA may be a poor indicator of individual quality due to the random nature of FA (Palmer 1994). The inability of unsigned FA to index organism-wide developmental stability or quality is thought to be demonstrated by the typical lack of correlations between unsigned FAs of

different traits within samples of individuals (Palmer and Strobeck 1986, Watson and Thornhill 1994; Dufour and Weatherhead 1996).

Finally, as mentioned in Chapter 1, researchers have argued that FA of some traits should be more reliable indices of quality and stress than others. Specifically, sexually-selected traits should show stronger FA-stress and FA-quality relations than other morphological traits (e.g., Møller 1993). Sexually-selected traits should have lower developmental stability, and should therefore be more susceptible to perturbation by stress and should vary more with differences in quality.

In this paper, I present a simple and tractable mathematical framework for examining determinants of asymmetry distributions and their effect on relations between asymmetry values and stress to populations or quality of individuals. In chapter 1, I examined the empirical support for several hypotheses. However, theoretical models still are needed to examine the logic behind emerging FA theory and to provide insight into which areas of investigation should be prioritized. I modeled generation of asymmetry distributions with respect to current problems regarding the reliability of FA of different trait types as a measure of stress and fitness (i.e., condition dependent sexually-selected traits versus other morphological traits), the determinants and usefulness of leptokurtic asymmetry distributions as potential indices of stress and fitness, and the consequences of the random nature of FA generation with respect to FA-fitness relations.

2. The model

Equations and Biological Rationale

Developmental noise and developmental stability are believed to be opposing processes which increase and decrease FA, respectively (Palmer 1994).

Mathematically, these relations can be represented by either division

[$FA = Dn / (1 + Ds)$] or subtraction [$FA = |Dn| - Ds$], where Dn and Ds are the effects of developmental noise or perturbations and developmental stability, respectively.

In both equations, Dn increases FA and Ds decreases FA. However, in the subtraction equation, individuals can be perfectly symmetrical, assuming that when $Ds > Dn$, FA is set to zero (if FA is not set to zero, then individuals with high developmental stability would become increasingly asymmetrical which is not logical). In the division equation, Ds reduces the effects of Dn on asymmetry but individuals will not be perfectly symmetrical. Perfection likely is rare in nature and that small amount of asymmetry usually will be present for metric traits (Palmer and Strobeck 1992). Thus, I use the division equation:

$$FA_{ij} = Dn_{ij} / (1 + Ds_{ij})$$

2-1

Where the subscript i denotes an individual and the subscript j denotes a character. Note that in the absence of developmental stability (Ds), FA solely reflects developmental noise (Dn). While Dn is due to random processes, Ds has a genetic component (Palmer and Strobeck 1992), and therefore may relate positively to the quality of an individual (i.e., higher quality individuals have higher developmental stability, Møller 1993a).

Developmental stability may relate to other genetic factors as well. As indicated, processes such as directional selection for increased trait size (e.g., for condition dependent sexually-selected traits) may select against modifiers which buffer development and result in reduced developmental stability for these traits (Møller 1993a; Manning and Chamberlain 1994). In contrast, processes such as stabilizing selection (e.g., on performance traits) may favour high developmental stability (Møller 1993a). Given that different traits are under different selection pressures and developmental histories, it is reasonable to assume that there will be intrinsic differences in developmental stability between different traits. I use a new term, **character-specific buffering**, to represent components of developmental stability which are specific to a given trait. Currently, the exact nature of the relation between quality and the underlying processes of developmental stability is unknown. However, it is commonly believed that some positive relation exists (e.g., Møller 1993a). It is conceivable, though, that not all processes are related to individual quality. Mathematically, those processes of developmental stability dependent upon individual quality can be represented multiplicatively (C_m), whereas those processes not dependent upon individual quality can be represented additively (C_a) as follows:

$$D_{sij} = C_{aj} + C_{mj} \times Q_i \quad 2-2$$

combining equations 1 and 2 yields,

$$F_{Aij} = D_{nij} / (1 + C_{aj} + C_{mj} \times Q_i) \quad 2-3$$

Q is individual quality. The terms F_A , D_n , D_s and the subscripts i and j are as above. At present, there is no reason to believe that one component of

developmental stability (Cm or Ca) is more important or more common than the other in nature, nor is there any a priori hypothesis regarding whether selection should affect Cm or Ca most. For these reasons, I include both Cm and Ca.

Finally, stress is believed to decrease developmental stability (Palmer and Strobeck 1992). I assume that researchers are interested in monitoring biologically important stress. The reduction of quality due to stress can be represented mathematically using division [$Q'=Q/(1+S)$] or subtraction [$Q'=Q-S$]. The difference is that quality reaches zero linearly, using the subtraction equation, whereas quality approaches zero asymptotically, using the division equation. I chose the division equation because I felt that an asymptotic relation is biologically more realistic.

In notation,

$$Q^*_{ij}=Q_{ij}/(1+ S) \quad 2-4$$

$$Ds^*_{ij}=Ca_j+Cm_j \times Q^*_i \quad 2-5$$

$$FA^*_{ij}=Dn_{ij}/(1+Ds^*_{ij}) \quad 2-6$$

Where S represents environmental stress during development as a factor influencing all individuals in a population, * refers to values in the presence of stress (e.g., Q is individual quality in the absence of stress, and Q* is individual quality in the presence of stress). All other terms are as above.

Simplifying Assumptions and Simulations

I made the following simplifying assumptions: Ca and Cm were constants for a given character in a population (equation 2-2), (Ds-Q) relations were linear and all variation in Ds was accounted for by differences in Q between individuals

(equation 2-2), stress affected all individuals in a population and stress increased FA only through decreases in quality (equations 2-4 to 2-6). Variation not included in the model, such as variation around the (Ds-Q) relation, would add complexity to the model, behaving as additional random noise, and would decrease the absolute strength of resulting FA-stress and FA-fitness relations. However, I was not trying to predict the absolute strength of such relations. Also, there was no reason to expect, or not expect, a linear relation between Ds and Q, or an inversely proportional relation between S and Q. Thus, I chose the simplest equations. Finally, although stress could affect terms other than quality (e.g., Dn, Cm, or Ca), such effects need not relate to a decrease in quality, and therefore may be biologically meaningless. Thus, my model was consistent with the implicit assumption that stress, which increases FA, causes a decrease in the average quality of a population. I emphasize that the above simplifying assumptions make results from this model tractable, yet do not detract from the way in which FA is believed to be generated.

To conduct the simulation, I needed to assign parameters for each of the terms above. Although different ranges of values could yield different strengths of FA-quality or FA-stress relations, I emphasize that the magnitude of the values used were unimportant. It was the relative values of each term with respect to one another that allowed insight into various phenomena, such as the effects of level of developmental stability on the relative strength of FA-stress and FA-quality relations.

Setting Values of Each Term

I first generated FA values for each individual-by-trait combination separately and later examined how character asymmetry related to stress and fitness. For each

individual by trait combination, values of D_n and Q (equation 2-3) were chosen randomly from normal distributions such that D_n and Q were unrelated. I used 10 'perturbation events' to determine the final D_n value. Each perturbation event had an additive effect on D_n , increasing D_n by a value between -1 to 1 units (standard deviation=1/3 units). I allowed Q values to range from 0 to 1 units (standard deviation=1/6 units). Thus, D_n values were randomly generated from a normal distribution around a mean of zero, with no pre-disposition towards asymmetry and no relation to quality (sensu Palmer and Strobeck 1992). I emphasize that the processes which generate asymmetry values in my model reflect how FA is supposed to be generated, and that conclusions from this model should help refine emerging FA theory. My model generates predictions for patterns that we should observe in nature, if the underlying assumptions of FA theory are correct.

As indicated, I was interested in examining the FA distributions and indexing abilities at different values of C_m and C_a for given distributions of D_n and Q . Values of C_m and C_a (equation 2-2) ranged from .007 to 102 units. 15 different C_m and C_a values were used. Each different combination of C_m and C_a values was considered a different trait. When C_m was varied for each trait, C_a was held constant at a value of 0. When C_a was varied, C_m was held constant at a value of 50. Although, values below or above those used in my model were possible, I was not attempting to predict what ranges of C_m and C_a occur in nature (as this would not be possible without empirical studies). I felt, however, that the ranges of C_m and C_a values used were adequate to illustrate the potential effects of different levels of character specific buffering and average developmental stability on FA distributions and indexing abilities, as defined in Table 2-1.

S (equation 2-4) was set at a value of 0 and 2 for the unstressed and stressed populations respectively. In nature, many different levels of stress are possible. Again, I chose levels that were illustrative. Stress levels too low or too high would not have changed the trends generated, with respect to C_m and C_a values, but could have made interpretations less clear.

Traits and FA Distributions

I examined the effect of different traits (i.e., different character specific buffering and average developmental stability) on FA-stress and FA-quality relations and the variance and kurtosis of FA distributions. I used 100 individuals in each simulation. For FA-stress relations, this consisted of 50 individuals in each stressed population and 50 individuals in each unstressed population. To quantify the relative strengths of FA-stress and FA-quality relations with respect to trait types (i.e., indexing abilities), I used t-values from the Levene's test (Palmer and Strobeck 1992) and coefficients of determination (r^2), respectively. I ran 50 simulations for each character to estimate the amount of variation in indexing ability, variance, and kurtosis. All statistics were calculated following Zar (1996)

Random Nature of FA

As mentioned, FA is believed to be generated by developmental noise, which is random by definition. I examined the effects of the random nature of FA on the ability of unsigned FA to index individual quality. I examined correlations between unsigned FAs of 10 different traits, and then I examined unsigned FA versus quality for each trait. I examined whether FA-quality relations for two traits related

to the likelihood of finding FA-FA correlations between those traits. In this separate set of simulations, I used the following parameter values: C_m values ranged from 0.5 to 5, C_a values were set at 0. I only used an unstressed population ($S=0$) because the question was really one about FA-quality relations. Again, D_n values were randomly determined for each trait. Hence these traits would be regarded as developmentally unlinked. Finally, I plotted FA versus quality for one "representative" trait to examine the bivariate distribution of data points. Here, C_m , C_a and S values were set at 100, 0 and 0 respectively as these values resulted in a strong FA-quality relation compared to other values used in the simulation. The ranges of D_n and Q values were as above.

3. Results

The ability of FA to index stress and FA to index quality changed with measures of character specific buffering, depending on relative values of C_m and C_a . Indexing ability initially increased with increasing C_m values and then eventually declined slightly (Figures 2-1a,b). In contrast, indexing ability decreased with increasing values of C_a (Figures 2-2a,b). FA variance, on the other hand, declined with increasing values of both C_m and C_a (Figures 2-3a, 2-4a). These results indicate that indexing ability can not be predicted from the level of developmental stability of a given trait (e.g., low average developmental stabilities may result in high FA variances for sexually-selected traits, but these traits may not be good indices). In brief, developmental stability and asymmetry variance is increased by both C_m and C_a , yet indexing ability is not.

The type of asymmetry distribution also was influenced by character specific buffering. Average kurtosis values were consistently positive and increased with increasing C_m values and with decreasing C_a values (Figures 2-

3b, 2-4b). The proportion of significant kurtosis values ranged from 6%, for low C_m and high C_a values, to 68% for high C_m and low C_a values. All significant kurtosis values were leptokurtic. Platykurtosis and antisymmetry were not generated by my model. Yet, in the absence of genetic pre-dispositions towards asymmetry (as in this model), I found both normal and leptokurtic distributions.

As indicated, I compared FA-quality relations to FA-FA relations for specific traits. Significant FA-FA relations occurred infrequently suggesting that, in general, individual traits should be poor predictors of organism-wide FA (Table 2-1). However, unsigned FA was related to quality for 9 out of the 10 traits. These results suggest that unsigned FA may consistently relate to individual quality and developmental stability even though unsigned FAs of different traits do not correlate (i.e., lack of organism-wide asymmetry does not necessarily imply lack of organism-wide developmental stability).

Although unsigned FA of traits were often significantly related to quality in my model, I found that FA-quality relations were typically triangular in their bivariate distribution (Figure 2-5; $df=98$, $r^2=0.22$, $p<0.0001$). I note that the strength of the FA-quality relation was affected strongly by the individuals with high FA magnitudes. To illustrate, removal of the top 15% of the FA distribution yielded a much weaker relation ($df=83$, $r^2=0.066$), whereas consideration of only the top 15% of the distribution nearly doubled the strength of the original FA-quality relation ($df=13$, $r^2=0.43$). Thus, unsigned FA should be a very poor predictor of quality at low values of FA, but may reliably indicate low quality at high values of FA.

Table 2-2. F-values for the correlation between FA and quality, using the traits as in Table 2-1. In this table, F-values > 3.94 indicate a significant correlation between character FA and quality of the individual (df=98); these are indicated by an *. Comparisons of Tables 2-2 and 2-3 show that FA can be significantly related to quality even when the FAs of traits are not typically related within a sample of individuals. Note that characters 1 to 10 are merely labels representing different character specific buffering values.

Character	F-values
1	4.34*
2	8.06*
3	6.93*
4	6.68*
5	9.57*
6	5.51*
7	1.08
8	8.39*
9	15.90*
10	14.89*

Fig 2-1. The effects of one component of developmental stability, C_m , on strengths of FA-stress and FA-quality relations, using t-values and Pearson correlations (a and b respectively). Error bars represent ± 1 standard deviation. Note that as this aspect of character specific buffering increases, so does the ability of FA to index stress and quality.

Fig 2-2. The effects of one component of developmental stability, Ca, on strengths of FA-stress and FA-quality relations, using t-values and Pearson correlations (a and b respectively). Error bars represent ± 1 standard deviation. Note that as this aspect of character specific buffering increases, the ability of FA to index stress and quality decreases.

Fig 2-3. The effects of one component of developmental stability, C_m , on FA variance and FA kurtosis (a and b respectively). Error bars represent ± 1 standard deviation. Here, FA variance decreases with increasing C_m indicating that lower developmental stability (i.e., higher FA variance) need not be equivalent to higher indexing ability (compare with Fig 2-1). Also note that there is a tendency towards leptokurtism (positive kurtosis values) indicating that processes underlying FA may also result in leptokurtic asymmetry distributions.

Fig 2-4. The effects of one component of developmental stability, C_a , on FA variance and FA kurtosis (a and b respectively). Error bars represent ± 1 standard deviation. Here, FA variance decreases with increasing C_a as does indexing ability (compare with Fig 2-2). Thus, indexing ability relates to the relative values of C_m and C_a rather than the absolute level of developmental stability per se (higher values of both C_m and C_a result in higher developmental stability, compare with Fig 2-3a). As in Fig 2-3, there is a tendency towards leptokurtic asymmetry distributions.

Fig 2-5. FA versus quality showing a triangular bivariate distribution ($r^2=0.22$, $df=98$, $p<0.0001$). FA has very poor predictive power with respect to individual quality at low FA values. However, at high FA values, quality is lower than average.

4. Discussion

Trait Types

Researchers have argued that condition dependent sexually-selected traits are often under directional selection for increased magnitude which may select against modifiers that buffer development and which may result in higher susceptibility of such traits to disruption by stress (Møller 1993a; Manning and Chamberlain 1994; Hill 1995). The evidence typically cited for low developmental stability is a much higher FA magnitude (variance) of sexually-selected traits compared to other morphological traits (Manning and Chamberlain 1994). However, my results suggest that both components of character-specific buffering (C_m and C_a) increase developmental stability, but have opposite effects on indexing ability. Thus, we should not necessarily expect sexually-selected traits to be better indices of stress or fitness than other traits, due to their lower developmental stability. Consider, for example, the absence of any developmental stability ($D_s=0$). If $D_s=0$, then FA would be due to random processes with no genetic component (i.e., to D_n alone). Hence, character FA would have no relation to quality (Palmer and Strobeck 1992) beyond direct effects of FA on quality. Similarly, if $D_s=0$ and FA did not relate to quality, then there would be no reason to assume that stress which increases FA typically affects average quality (or fitness) of a population. In this case there would be no reason to prefer FA over traditional monitoring techniques, such as LD_{50} values (cf. Leary and Allendorf 1989; Clarke 1993a who advocated the use of FA as a biologically meaningful indicator). Empirical tests for differences in the strength of FA-stress and FA-fitness relations, between sexually-selected and other traits

also show no evidence of differences between trait types (Leung and Forbes meta-analysis, 1996, Chapter 1).

Differences in character-specific buffering and average developmental stability between different traits would mean that FA-stress and FA-fitness relations are inherently variable between different traits. Non-significant relations could result even under ideal circumstances. Factors such as low sample size, directional asymmetry, antisymmetry, measurement error, and FA-size relations have been suggested as reasons for variable or non-significant results (Palmer 1994). While these factors are important in empirical studies, I emphasize that these factors were not a problem in my model: large samples were generated, there were no genetic pre-dispositions toward asymmetry, and measurement error and allometric relations were absent.

Taken together, my results suggest that it would be erroneous to assume that no difference in FA indicates an absence of stress or that differences in FA indicates differences in quality, given that FA of some traits should not relate to stress or quality regardless of experimental rigor. Ideal FA of other traits may be associated with stress and quality consistently, but these traits may have to be validated empirically for each biological system. It is also conceivable that stress may affect terms other than quality in the FA equations. If changes in these terms are not correlated with changes in individual quality, then increases in FA variance will not necessarily relate to biologically important stress.

FA Distributions

Asymmetry distributions can be normal or leptokurtic when modelled according to verbal ideas about how FA is generated. Palmer and Strobeck (1992) suggested that leptokurtism may arise due to combining normal and platykurtic asymmetry

distributions in a population. Thus, there could be genetic pre-dispositions towards asymmetry causing leptokurtism, and researchers may wish to avoid these particular distributions. However, I emphasize that leptokurtic asymmetry distributions (LADs) could arise even when there were no genetic pre-dispositions towards asymmetry and given that the optimal phenotype was perfect symmetry (in my model), and even in the absence of extraneous factors (e.g., strong differential mortality).

I explain the tendency toward LADs as follows: even though developmental noise of individuals in a population may be normally distributed, developmental stability causes some individuals to be able to buffer these developmental accidents. The effect of differential developmental stability would be to cause more individuals to have low FA values than expected from a normal distribution since individuals can have low FA because of either chance or because of high developmental stability. Some "unlucky" individuals will have simultaneously high developmental noise and low developmental stability. Thus, there will be proportionally more individuals in the center and the tail of the asymmetry distribution which defines a LAD.

Given that we should expect to find LADs due to theoretical considerations, researchers are justified in assaying both normal FA distributions and LADs with respect to stress and fitness relations (cf. Palmer and Strobeck 1992; Palmer 1994). Furthermore, I expect very few individuals to have large asymmetry values (especially if such individuals are more susceptible to mortality, e.g., Naugler and Leech 1994). Unless there is strong evidence that such outliers do not reflect FA (e.g., deviations caused by injury, Palmer 1994), it is important to include them, given that these "outliers" are expected, and that they can account for a considerable portion of the variation of FA-stress and FA-fitness relations.

Random nature of FA

As indicated previously, some researchers believe that unsigned FA should be a poor index of individual quality due to the fact that developmental noise is generated by random processes and is independent for different traits that are developmentally unlinked (cf. Palmer 1994). The inability of FA to relate to quality of individuals is believed to be illustrated by the typical lack of correlations between unsigned FAs of different traits within samples of individuals (Dufour and Weatherhead, 1996; and references therein). I also found that correlations between unsigned FAs of different traits within individuals were uncommon in my model. However, I argue that the lack of organism-wide FA does not necessarily invalidate unsigned FA as a measure of organism-wide developmental stability or quality (cf. Palmer 1994, Watson and Thornhill 1994). It is clear from this model that researchers should expect to find fewer significant FA-FA correlations than FA-quality correlations, due to the presence of 2 independent terms (D_n) for developmental noise in FA-FA relations. Although developmental noise also will reduce the strength of FA-quality relation, the D_n term only appears once in the FA-quality relation. Thus, unsigned FA of certain traits may consistently relate to organism-wide developmental stability and quality or fitness, even though unsigned trait FA magnitudes often are not concordant within samples of individuals.

The question of whether FA can be used to index quality or whether it can be used in mate choice requires slightly different considerations (cf. Møller 1993a and Palmer 1994 for opposing views). On the one hand, FA may reflect developmental stability, which may relate to individual quality (Møller 1993a). On the other hand, FA is generated by developmental noise, hence even

developmentally unstable, low quality individuals may have low FA due to chance (Palmer 1994). Both arguments have merit. I found that significant FA-quality relations had typically triangular bivariate distributions. Thus, unsigned FA had little predictive power at low values of FA, which contained both high quality individuals and lucky individuals. High values of FA in some traits, on the other hand, could reliably indicate low quality. Thus, females may still choose on the basis of symmetry by selecting against high FA individuals. However, the importance of asymmetry to mate choice in nature is questionable, because few individuals should be in the tails of the FA distribution. Hence there may be few individuals to select against. I further note that FA-quality relations of performance traits need not be triangular, as FA may directly affect the quality of individuals. For performance traits, it is possible for FA to be predictive of quality at both high and low values of FA. However, I emphasize that unsigned FA may relate to quality even when there is no direct effect of FA on the functioning of a trait (cf. Palmer 1994).

The mechanism underlying FA-quality relations also may affect interpretations of FA-stress relations. Typically, it is assumed that stress that causes an increase in FA is biologically relevant (cf. Clarke 1993a), and that the increase in FA reflects a decrease in the average quality of individuals in a population. However, if unsigned FA relates to quality only through direct effects of FA on quality, our interpretation of FA-stress relations may be limited to direct costs of increased FA. Given that increases in the average magnitude of FA is often small and that most individuals will remain symmetric solely due to chance, the effects of direct costs of increased FA on the average quality of a population should be minor. In this case, the biological importance of FA-stress relations should be questioned. Thus, the mechanisms behind FA-quality relations are important to examine.

5. Conclusions

I believe that the terms used in my model are simple, yet biologically reasonable, and that this model reflects current views on factors influencing expression of trait FA. I conclude that there is currently little logical reason to expect that the strength of FA-stress and FA-fitness relations should be typically higher in sexually-selected traits or other developmentally unstable traits. Furthermore, FA may not relate to stress or fitness even in the absence of confounding factors, hence FA should be interpreted with caution. In addition, both normal asymmetry distributions and LADs can result from developmental process underlying true FA, and neither should be avoided. I further conclude that unsigned FA may relate to quality, even when FA does not have direct effects on the functioning of traits. Such relations may occur even though there is typically a lack of organism-wide unsigned FA and even though FA is caused by random developmental noise. However, relations between FA and stress and FA and quality should typically be weak, even if developmental stability were perfectly correlated with quality, due to the effects of developmental noise. Thus, FA may not reliably indicate either stress or quality, at least as currently studied. Finally, I suggest that the interpretation of FA-stress relations is dependent upon the mechanism behind FA-quality relations.

CHAPTER 3

Correcting for Allometry in Studies Relating Fluctuating Asymmetry to Quality Within Samples.

1. Introduction

Cumulative results suggests that FA-quality relations exist as predicted, but are weak on average and heterogeneous (Leung and Forbes 1996, Chapter 1). FA relations may be weak and heterogeneous because allometric relations (i.e., FA-size) confound FA-quality relations. Traits which grow larger may have more 'opportunity' to become asymmetric, resulting in positive FA-size relations; e.g., FA of legs may be greater than ears because legs are larger than ears rather than because Ds differs. Relative FA [FA scaled by trait size; $FA_{rel} = 2x | \text{right-left} | / (\text{right} + \text{left})$] is the method most commonly suggested and used to correct for allometry (Palmer 1994), although a few studies have used other methods (multiple regression, e.g., Solberg and Sether 1993; partial regression, e.g., Harvey and Walsh 1993). However, FA may not relate to size in all traits, and corrections should be done only when there is clear evidence for FA-size relations (Palmer 1994). Typically, allometric relations are detected using significance of FA-size correlations as a criterion.

Swaddle et al. (1994) noted that FA_{rel} is valid only when FA-size relations are isometric (i.e., FA proportional to size). However, isometry rarely has been tested (but see Møller 1992, McLachlan and Cant 1995). Further, researchers have not examined explicitly the consequences of using FA_{rel} when FA-size relations are anisometric (i.e., not isometric), nor examined efficacy of tests for

isometry, nor suggested alternative methods. Also, researchers have suggested that size corrections may obscure FA-quality relations if size relates to quality (Palmer 1994); again, no study has explicitly examined the consequences of size-quality relations.

In this paper, I used a simulation model to determine potential consequences of allometry and to suggest rules to correct for size (hereafter simply termed Rules). I used Type I error rates (finding significant FA-quality relations when no underlying relations exist) and power (probability of finding significance when true FA-quality relations exist) as criteria to judge the efficacy of different Rules. Mathematical models allow the presence or absence of FA-quality, FA-size, and size-quality relations to be specified. Thus, we can examine conditions where inflated Type I error rates occur and compare power of different Rules. In empirical studies, different Rules may yield different results, however, we would not know whether differences occurred because of increased Type I error rates or increased power.

2. Basic Model

FA_u and FA_{rel} referred to absolute FA uncorrected for size and absolute FA divided by average trait size, respectively. Equations simulated 3 relations: 1) component of size variation related to quality variation, 2) component of FA_u variation related to size variation, 3) component of FA_u variation related to quality variation (quality represented any measure that researchers could relate to trait size or FA; e.g., survival). FA_u was defined as,

$$FA_{ui} = |R_i - L_i|$$

where R and L are the right and left trait sizes of individual i , respectively. Trait sizes were determined by summing growth magnitudes (g) and asymmetry in growth (A) across j growth intervals for individual i . I assumed that A was directionally random; thus R could be less than or greater than L and $(R-L)$ would be distributed around a mean of zero. In notation,

$$R_i = \sum (g_{ij} + A_{ij}) \quad 3-2$$

$$L_i = \sum g_{ij} \quad 3-3$$

Growth (g) was modelled as,

$$g_{ij} = \mu + \alpha * f(q_i) + \beta * V_i \quad 3-4$$

where μ was mean growth, $f(q)$ was a function of quality, and V was variation in growth independent of quality. α and β were constants describing relative contributions of q and V to growth (e.g., if $\alpha:\beta$ was 1:0, 0:1, quality would account for 100% and 0% of size variation, respectively). $f(q)$ and V followed normal distributions (mean: $\bar{x}=0$, standard deviation: $\sigma=1/3$). μ was set at 10. I considered both positive ($\alpha>0$) and inverse ($\alpha<0$) size-quality relations. I allowed 2 growth intervals (j). I did not consider perfect size-quality correlations ($\alpha:\beta = 0:1$; perfect size-quality correlations are unlikely in nature although relations can be strong).

Developmental noise (Dn) increased A and was modelled as,

$$Dn_{ij} = a * f(g_{ij}) + b * c \quad 3-5$$

where $f(g_{ij})$ was a function of size, a and b were constants determining contribution of size on asymmetry, c denoted processes independent of size but which affected D_n , subscripts were as above. $f(g_{ij})$ and c followed normal distributions ($\bar{x}=0$; $\sigma = g_{ij}/300$ and $1/300$, respectively). Standard deviations were small such that FA comprised only a small proportion of trait size (Swaddle et al. 1994). As $a:b$ ratio increased, the contribution of size on A [$f(g_{ij})$] increased relative to other processes (c) (e.g., if $g_{ij}=10$ for all individuals, and $a:b = 0.1:1$, relative contribution of size on A would be 1:1 because $f(g_{ij})$ was scaled by size). When $a:b$ was 1:0, D_n was isometrically related to size. I also examined anisometric relations where $D_{n_{ij}}$ was proportional to growth squared ($f(g_{ij}^2)$ has $\bar{x}=0$ and $\sigma=g_{ij}^2/300$).

Developmental stability (D_s) reduced the effect of D_n on A ,

$$A_{ij}=D_{n_{ij}}/(1+D_{s_i}) \quad 3-6$$

and

$$D_{s_i}=\delta \cdot q_i \quad 3-7$$

where δ was a constant determining the magnitude of D_s . I modelled D_s as a linear function of quality (q) where q explained all the variation in D_s . Although other variation exists in nature, such variation only weakens relations between A and q and was not important for this model. Different strengths of A - q relations were generated by varying δ (the strength of A - q relations increased as δ increased).

The purpose of this model was to examine the validity of Rules, to determine the potential consequences of allometry, and to compare efficacy of different Rules. I examined validity of Rules using Type I error rates (i.e., when $\delta=0$). If error rates were reasonably low (compared to predicted alpha of 5%), I

examined power (i.e., when $\delta > 0$). I determined consequences of allometry using relative power (RP); i.e., power of Rules in the presence of allometric relations relative to power of FA_u in the absence of FA-size relations. FA_u in the absence of FA-size relations represented results that we should expect if allometry was not a confounding factor, and was a baseline for comparison. I examined efficiency of Rules for size correction under given allometric relations, using differences in absolute power (DP) between Rules and FA_u under those given allometric relations. Thus,

RP=power of Rule in presence of allometric relations

$$\div \text{power of } FA_u \text{ in absence of allometric relations} \times 100\% \quad 3-8$$

DP=power of Rule - power of FA_u

(both in presence of same allometric relations) 3-9

I ran 1000 simulations for each of 200 sets of parameter values [sample sizes: $N=40$ and 100 , 5 size-quality relations ($\alpha:\beta$), 5 FA-size relations: (a:b), and 4 FA-quality relations (δ); see Appendix D for values]. I primarily considered FA-size relations ranging from no relation (a:b = 0:1) to isometric (a:b = 1:0). However, I ran additional simulations for FA-size relations greater than isometry ($FA \propto \text{size}^2$) using only parameters which produced the greatest consequences of allometry (see below).

I examined 11 Rules based on FA_u , FA_{rel} , and multiple regression and different criteria to determine when size corrections should be conducted (Table 3-1). Rules included conducting size corrections when FA-size relations were significant or correcting using FA_{rel} only when FA-size relations were isometric (Swaddle et al. 1994). I identified departures from isometry following Niklas

(1994). Statistics generated by simulations were compared to Statistica and SAS, to ensure accuracy of the model. All tests were based on one-tailed hypotheses predicting inverse FA-quality relations.

3. Positive Size-Quality Relations: Current Rules

I considered Rules 1 to 7 to be the current rules for size correction (Table 3-1). In the absence of size-quality relations ($\alpha:\beta = 0:10$), Type I errors were approximately as predicted, regardless of which Rule was used (Fig 3-1a; Appendix E). Further, FA-size relations did not greatly affect power ($RP > 80\%$), and power differed only marginally between Rules ($DP < 15\%$, Fig 3-1b, Appendix E). In contrast, when size and quality related positively ($\alpha > 0$), indiscriminately using FA_{rel} (Rule 2) resulted in inflated Type I errors (as high as 84%, Appendix E) when underlying FA-size relations was less than isometry ($b > 0$; Fig 3-1c). Type I errors as high as 17% and 29% occurred if FA_{rel} was used when FA-size relations did not differ significantly from isometry (Rules 4 and 5, respectively). Thus, I did not consider Rules 2, 4, and 5 for examining power, because based on Type I error rates above, these allometric corrections should be avoided.

Rules 1, 3, 6, and 7 yielded predicted or conservative Type I error rates (Fig 3-1c). However, power of detecting FA-quality relations was severely reduced (Fig 3-1d). If allometric relations were not removed (Rule 1), RP could be as low as 2% (Appendix E). Although there were benefits if size corrections were conducted when FA-size relations were significant (Fig 3-1d), RP values could still be as low as 9% (Rule 3) and 15% (Rule 7) (Appendix E). In contrast, multiple regression (Rule 6) yielded greater power when FA-size relations were strong, but weaker power when FA-size relations were weak or absent (Fig 3-1d, Appendix E). Thus, when size related positively with quality, correcting for

allometry was important. However, Rules currently used to determine whether size correction should be conducted were inefficient.

4. Alternative Rules to Control for Allometry

Rules may be inefficient if underlying FA-size relations are not detected. If size relates positively to quality, FA-size relations can obscure FA-quality relations since size and quality theoretically have opposite effects on FA (Fig 3-1d). The reverse also could occur; FA-quality relations may obscure FA-size relations; e.g., average strength of FA-size relations generated decreased from $r=0.29$ to $r=-0.04$ (r =Pearson's correlation coefficient) as FA-quality relations increased from non-existent ($\delta=0$) to strong ($\delta=10$), even though the underlying FA-size relation was identical and isometric (a:b ratio = 1:0). One possible solution is to remove relations between size and quality (i.e., using residual size from a size-quality regression) before testing for FA allometry. The probability of detecting relations between FA and residual size was not affected by FA-quality relations. However, the strength of FA-residual size relations diminished as the strength of size-quality relations increased.

I used FA_{rel} and multiple regression when FA-residual size relations were significant (Rules 8 and 9, respectively). Type I errors were only slightly inflated (maximum of 10.3% and 8.9% for Rules 8 and 9, respectively) (Fig 3-1c; Appendix E). However, power was greatest using Rules 8 and 9. The minimum RP was 34% and 43% using Rule 8 and 9, respectively (compared to 9% and 15% using rule 3 and 7, respectively). The maximum DP was 48% and 42% using Rules 8 and 9, respectively (compared to 30% and 28% using Rules 3 and 7, respectively, Appendix E). I also considered partial regressions removing size from FA. However, this was consistently weaker than FA_{rel} or multiple regression

(although Type I error rates were comparable to multiple regression), and thus was not presented.

Next, I examined FA-size relations greater than isometry ($FA \propto size^2$). I used strong quality-size relations ($\alpha:\beta$ ratio = 6:4) with no component of Dn independent of size ($b=0$) and a sample size of 100. These parameters resulted in the greatest consequences of allometry in the above analysis. Rule 9 yielded considerably better power than Rule 8. RP ranged from 2% to 54% for Rule 8, and from 39% to 74% for Rule 9 (Appendix E). Both yielded predicted or conservative Type I error rates.

5. Inverse Size-Quality Relations.

When size related inversely to quality, Rule 1 could yield greatly inflated Type I errors when FA-size relations existed (Fig 3-2a). Thus, Rules 4 and 5 (using FA_u in the absence of isometry) made little sense, and were modified (see Rules 10 and 11, Table 3-1). Rules 2 and 3 were conservative if FA-size relations were less than isometric; they yielded appropriate Type I errors if FA-size relations were isometric, but were inflated when $FA \propto size^2$ (Fig 3-2a). Rules 8 and 9 also resulted in inflated Type I errors (correcting for allometry when FA-residual size relations were significant). Only Rules 6 and 7 yielded expected Type I error rates across all parameters examined (Fig 3-2b). Thus, when size related inversely with quality, raw values of size were needed and residual size was not appropriate.

Use of FA_{rel} (Rule 3) only failed when FA-size relations were greater than isometry. Thus, I considered rules which corrected for allometry when FA-size relations were significant; but used either FA_{rel} or multiple regression depending on deviations from isometry (Rules 10 and 11, Table 3-1). Type I errors were

improved but still inflated (maximum error rates were 18.4% and 39.2% for Rules 10 and 11, respectively).

Because FA_u had inflated Type I error rates when FA-size relations existed, DP was invalid. Thus, I only presented RP (FA_u had expected Type I error rates in the absence of FA-size relations). RP could be reduced by more than half by using multiple regression (Fig 3-2c). Correcting for size when FA-size relations were significant (Rule 7) yielded little difference compared to always using multiple regression (Rule 6). This was because when size related inversely with quality, inverse FA-quality relations indirectly inflated the strength of positive FA-size relations.

6. The Coefficient of Variation of Size.

I examined the coefficient of variation of size ($CV = \sigma/\bar{x} * 100\%$, where σ was standard deviation and $\bar{x}$ was sample mean). I incrementally increased underlying mean size (μ ; eq. 3-1) from 10 to 100. This resulted in a decrease in CV from 23% to 2% ($\alpha:\beta = \pm 6:4$, $a:b=1:0$, $N=100$). I examined RP for Rules with predicted Type I errors in the above analysis. I also examined Type I error rates for the most common rules [Rule 1: FA_u and Rule 2: FA_{rel}] for inverse and positive size-quality relations, respectively.

As CV decreased, the strength of FA-size relations decreased as did the effect of allometry on power and Type I error rates of FA_u (rule 1) and FA_{rel} (rule 2) (Fig 3-3a). Thus, at low CV, size corrections using FA_{rel} may not be important. In contrast, reductions in RP were dependent upon the strength of size-quality relations rather than CV, using rule 6 or 7 (if size related inversely with quality) and using rule 6 (if size relates positive with quality). For positive size-quality relations, using rule 9 was stronger than rule 1 regardless of CV (Fig 3-3).

Table 3-1: Description of allometric Rules used in this paper. Suggested Rules are indicated by an ***.

Rules	Description
1	Use FA_u regardless of strength of FA-size relations.
2	Use FA_{res} regardless of strength of FA-size relations
3	Use FA_{res} only when FA-size relations significant; otherwise use FA_u . This is the most common rule currently used.
	***Use rule 3 if the coefficient of variation of size (CV) is small.
4	Use FA_{res} only when FA-size relations isometric; otherwise use FA_u . Deviations from isometry was identified by intercept significantly different from zero in regression between $ FA $ and size (Niklas 1994).
5	Use FA_{res} only when FA-size relations isometric; otherwise use FA_u . Deviations from isometry was indicated by scaling component (κ) significantly different from one in the allometric equation $FA = \gamma \text{size}^\kappa$ (Niklas 1994).
6	Use multiple regression (multiple regression) regardless of strength of FA-size relations.
7	Use multiple regression only when FA-size relations significant.
	***Use rule 7 if inverse size-quality relations exist and CV is large.
8	First, take residual size after removing relation between size and quality. Use FA_{res} if FA-residual size relations significant; otherwise use FA_u .
9	First, take residual size after removing relation between size and quality. Use multiple regression if FA-residual size relations significant; otherwise use FA_u .
	***Use rule 9 if positive size-quality relations exist and CV is large.

Rules	Description
10	Correct for size when FA-size relations significant, otherwise use FA_u . If FA-size relations significant, use FA_{rel} only when FA-size relations isometric; otherwise use multiple regression. Deviations from isometry determined as for rule 4.
11	Correct for size when FA-size relations significant, otherwise use FA_u . If FA-size relations significant, use FA_{rel} only when FA-size relations isometric; otherwise use multiple regression. Deviations from isometry determined as for rule 5.

Fig 3-1. Type I errors ($\delta=0$; panels a,c) and power (% significant; panels b,d) of Rules (Table 3-1) were illustrated as underlying FA-size relations increased from no relation (a:b = 0:1) to isometry (a:b = 1:0) to $FA \propto size^2$. Relations were shown in absence of size-quality relations ($\alpha:\beta = 0:1$, panels a,b) and in presence of strong positive size-quality relations ($\alpha:\beta = 6:4$, avg size-quality relation; $r=0.83$, panels c,d). Rules 1 to 9 were symbolized by: 1) +, 2) Δ , 3) $\circ$, 4) $\bullet$, 5) $\blacklozenge$, 6) $\blacksquare$, 7) ∇ , 8) $\diamond$, 9) $\square$. Horizontal line represented predicted Type I error rate. Type I errors were similar for all Rules and were not inflated in (a) but were highly inflated for Rules 2 (Δ), 4 ($\bullet$), and 5 ($\blacklozenge$) in panel c. Thus, power was examined only for Rules 1, 3, and 6-9. Power was similar in the absence of size-quality relations (b). In contrast, when size-quality relations were strong (d), Rules 3 ($\circ$) and 7 (∇) were marginally more powerful than Rule 1 (+) when FA-size relations $\leq$ isometry. Rule 7 was much more powerful when $FA \propto size^2$. Rule 6 ($\blacksquare$) was more powerful when FA-size relations were stronger, but less powerful when FA-size relations were weaker. Rules 8 ($\diamond$) and 9 ($\square$) yielded similar power for FA-size relations $\leq$ isometry. Rule 9 was the most powerful overall when all parameter sets were considered. Power was presented only for $\delta=1$ (other values of δ yielded similar trends). Each data point represented 1000 simulations.

FA-Size Relations (a:b ratio)

Fig 3-2. Type I errors ($\delta=0$, Panels a, b) and relative power (RP, Panel c) of Rules (Table 3-1) were illustrated as underlying FA-size relations increased from no relation ($a:b = 0:1$) to isometry ($a:b = 1:0$) to $FA \propto size^2$. Because Rule 1 had inflated Type I errors, Rules 4 and 5 were excluded and absolute differences in power (DP) also were not shown (see text). RP was presented only for Rules 6 and 7 (other Rules had greatly inflated Type I errors). Horizontal line represented predicted Type I error rate. Rules 1 to 9 symbolized as in Fig 3-1; Rules 10) ▲ and 11) ▼. Relations were in the presence of strong inverse size-quality relations ($\alpha:\beta = -6:4$). RP was examined only for $\delta=1$ (other values of δ yielded similar trends). Each data point represented 1000 simulations.

Fig 3-3. Relations were shown between coefficient of variation of size (CV) and strength of FA-size relations (Pearson's r , Panel a). FA-size relations were unconfounded by FA-quality relations: $\delta=0$). Relations were also shown between CV and Type I error rates of FA_u (Rule 1) when size-quality relations were inverse ($\alpha:\beta = -6:4$) and FA_{rel} (Rule 2) when size-quality relations were positive ($\alpha:\beta = 6:4$) (Panel a). Relations were also shown between CV and relative power (RP; for $\delta=1$) of Rules for positive size-quality relations (Panel b) and inverse size-quality relations (Panel c). FA-size relations simulated were isometric ($a:b = 1:0$). RP was presented only for Rules with reasonable Type I errors (see Figs 3-1 and 3-2). Rules were symbolized as in Fig 3-1. Each data point represented 1000 simulations.

7. Discussion

Researchers need to separate relations between FA and quality from relations between FA and size. I examined 48 studies (not including studies using trait size as a measure of quality) and found 18 studies which corrected for size using FA_{rel} in part of their study, and 4 using other methods (i.e., partial regressions, multiple regression, ANCOVA). 20 studies did not assay for FA-size relations, 1 found relation greater than isometry ($FA \propto size^{2.5}$, McLachlan and Cant 1995), and the others used significance of FA-size relations to indicate presence (or absence) of allometry. Only 20 studies tested for size-quality (or stress) relations. 11 found that size related positively with quality (or size decreased with stress), 7 found no relation, and 2 found inverse relations.

Allometry clearly has not been addressed adequately, given methodology used in existing studies and given the results of this paper; e.g., if isometry exists and FA is not corrected, up to 98% of the power could be lost if positive size-quality relations exist, or Type I errors could be up to 77% if inverse size-quality relations exist. Inappropriately using FA_{rel} can yield Type I errors as high as 85% or may result in 98% loss of power depending on direction of size-quality relation. Researchers typically use significance of FA-size correlations to detect allometric relations. However, this may yield little benefit when inverse FA-quality and positive size-quality relations exist (as appears to be the most common scenario, see above). FA-quality and FA-size relations partially cancel one another out such that both could appear non-significant. Thus, allometry may remain undetected and uncorrected. Further, as noted by Møller (1992a) and Swaddle *et al.* (1994), corrections using FA_{rel} is theoretically only appropriate if FA-size relations are isometric. Moreover, one component of FA measures is

measurement error. FA_{rel} would be biased also if measurement error did not scale isometrically with size. If we could determine when isometry existed, FA_{rel} could be the most powerful index. Unfortunately, standard methods to detect deviations from isometry appear unreliable and using FA_{rel} can result in inflated Type I error rates (Rules 4,5,10, and 11).

The decision of which rule to use requires consideration of trade-offs between the possibility of inflated Type I error rates and power. Trade-offs depend on strength and direction of size-quality relations, strength of FA-size (or residual size) relations, and CV (see Table 3-1 for suggested Rules to use under different conditions). If size does not relate to quality, corrections for size will have only marginal effect on power or Type I error rates, regardless of underlying relations between FA and size. If size relates to quality, CV is important. As CV decreases, potential consequences of allometry decrease, underlying magnitudes of FA-size correlations decrease, and consequences of correcting for allometry using FA_{rel} decrease. In contrast, reduction in RP using multiple regression depends on the strength of size-quality relations and not on CV, because variation in quality is removed along with size (Palmer 1994, see also multicollinearity, Neter *et al.* 1983). Thus, for low CV, multiple regression may sometimes be disadvantageous. When size relates inversely to quality, inverse FA-quality relations may inflate the probability of finding positive FA-size relations and size corrections may appear necessary even if no underlying FA-size relation exist. One potential solution is to use multiple regression (Rule 7) when CV is high but to use FA_{rel} (Rule 3) when CV is low (note that low CV is subjective, but see Fig 3). Although FA_{rel} is biased when underlying FA-size relations differ from isometry, this bias decreases as CV decreases. For high CV, multiple regression may be necessary as the consequences of using FA_{rel} can be great (e.g., 78% Type I errors when $CV=23\%$ and $FA \propto size^2$). In contrast, for

positive size-quality relations multiple regression (Rule 9) should be more powerful than Rule 1 regardless of CV (but see cautions below). This is because when CV is low, FA-size relations are significant less often and multiple regression is used less often, thus canceling negative consequences of multiple regression.

8. Cautions and Other Theoretic Considerations

Conclusions of this paper necessarily are dependent upon underlying assumptions. Here, I discuss limitations of this model and problems that require further considerations. First, while I have suggested use of multiple regression to control for allometry, its use must be approached with caution. For example, deviations from linearity could invalidate simple multiple regression (Neter et al. 1983). Although I found that multiple regression was robust to the non-linear scenario where $FA \propto size^2$, many other possible curvilinear relations exist, which were not explored in this model. However, some curvilinear relations would not be expected due to size scaling (FA should be positively related to size because larger individuals have greater opportunity to be asymmetrical), and it would be questionable whether size corrections should be conducted where size does not causally influence FA (e.g., U-shaped FA-size curve). Nevertheless, it is possible to examine size relations and use transformations or more complex models if curvilinear relations exist (Neter et al. 1983). Also, I did not examine consequences of allometry on other forms of asymmetry (antisymmetry or directional asymmetry, Palmer 1994).

Another problem occurs when there is strong interdependence between independent variables (i.e., multicollinearity); FA-quality relations could not be interpreted except in the context of size (Neter et al. 1983). While multiple

regression had appropriate Type I error rates when FA was causally dependent on size, Type I error rates could be inflated when FA and size were correlated indirectly; e.g., if FA and size were related through some common unknown variable and if size related to quality through another independent unknown variable, then FA would relate spuriously to quality using multiple regression. Unfortunately, separation of correlation from causation may not be possible. Nevertheless, multiple regression still offers the best option when CV is high. However, it may be better to use FA_u or FA_{rel} when CV is low.

Finally, I did not examine tests of FA differences between samples in this paper (ca. half of current studies test whether stressed samples have higher FA than unstressed samples). For FA-quality relations within samples, we can measure directly the component of quality of interest (e.g., longevity). Other unknown components of quality could relate to both size and FA, but it would be the component of interest that would determine the consequences of allometry on detection of FA relations (e.g., in the absence of longevity-size relations, underlying FA-size relations should not greatly confound FA-longevity relations). In contrast, for tests of FA differences between samples, quality would be an unmeasured quantity, and could not be removed. Hence, FA-quality relations may obscure detection of FA-size relations when positive size-quality relations exist. Solutions to this problem have yet to be determined.

This paper highlights potential consequences of allometry to FA studies, suggests rules to correct for size, and suggests when allometry should not be a problem. While these Rules provide advantages over current methods, they have flaws and more efficient and robust Rules could be possible. Given that allometry can greatly affect power even when size corrections are conducted, I suggest researchers report CV and the strength and direction of size-quality and FA-size (residual size) relations to determine the potential effects of allometry. This is

similar to arguments that measurement error should be reported because it reduces reliability of FA (Palmer 1994).

CHAPTER 4

Fluctuating Asymmetry as a Bioindicator of Stress: Comparing Efficacy of Analyses Involving Multiple Traits.

1. Introduction

Applied biologists are interested in monitoring environmental stress, preferably before stress irreversibly damages populations (Clarke 1993a, Bunn 1995). Researchers have used changes in the biota to indicate stress (i.e., bio-indicators). While a variety of bio-indicators have existed, fluctuating asymmetry (FA) has received increasing attention in the last decade. FA variance has been found to increase with stress (reviewed in Leung and Forbes 1996, Chapter 1), and may therefore provide a cost effective simple indicator of environmental stress (Clarke 1993). Also, FA has been related to quality of organisms; therefore a change in FA should be biologically relevant (Sommer 1996). However, empirical relations between FA based on single traits and stress are weak on average and heterogeneous (Leung and Forbes 1996, Chapter 1). Further, we would predict relations between FA and stress should be weak based on mathematical algorithms generating FA (Leung and Forbes 1997a, Chapter 2). Thus, if researchers are serious about using FA as a bio-indicator of stress, methods must be found to increase the reliability of FA to detect stress.

Previous analyses and mathematical models usually have considered FA based on single paired traits. Analyses that combine information across traits should prove to be more reliable detectors of stress (Leary and Allendorf 1989, Watson and Thornhill 1994). No study has quantitatively compared the efficacy or the validity of such analyses, although some composite indices of FA (CFAs)

have been used in empirical studies (e.g., Whitlock 1993, Dufour and Weatherhead 1996, Manning and Ockenden 1994) and other analyses have been suggested (Palmer 1994, Zhivotovsky 1992). Of those studies that have used a CFA measure, the strengths and directions of FA relations for single traits typically are not reported (with the exception of Manning and Ockenden 1994), and different CFA analyses typically are not used on the same data sets. For convenience, I will refer to analyses combining information on FA distributions across traits as CFA analyses.

In this paper, I examine properties of CFA analyses using simulations. In a simulation study, we can work with a model for FA whose properties are known, and ask how much information about the model can be recovered from various measures of FA. Furthermore, we can isolate and study a number of factors which could affect the relative efficacy of CFA analyses, such as the strength of FA-stress relations, the magnitude of measurement error, the variance in developmental stability of different traits, and deviations from normality. The objective of this study is to help researchers decide which CFA analysis should be used, given different patterns of FA distributions. I determine which analyses are valid, in terms of the probability of concluding that FA of two samples differ when samples come from identical FA distributions (i.e., Type I error rates) and which analyses provide the greatest probability of detecting relations when FA differs between two populations (i.e., power).

2. Composite Fluctuating Asymmetry Analyses

I consider six analyses of fluctuating asymmetry across multiple traits, which I refer to as CFA analyses. These are defined below, and summarized in Appendix F. The first three methods all involve calculation of a single index of FA across

traits, which are then analyzed by a simple t-test. The last three methods are multivariate analyses, and do not involve calculation of a simple index of FA.

The first composite statistic (CFA 1) is the sum (or mean) of absolute FA values ($|FA|$) for all traits for each individual (e.g., Whitlock 1993, Clarke and McKenzie 1992). This statistic is simple to calculate, but traits with higher FA magnitudes would be weighted more heavily than more developmentally stable traits. Some researchers have standardized FA by the mean size of the trait in each individual (e.g., Thornhill et al. 1995) or the average trait size of the population (Palmer and Strobeck 1986) before calculating CFA1. This would be appropriate to control for size differences when FA magnitude across traits scales positively and isometrically to trait size. However, this would not necessarily control for differences in developmental stability across traits. I do not model nor consider allometric relations in this chapter (but see Leung 1998, Chapter 3).

Another way to standardize FA magnitudes of traits is to divide each FA value by the average $|FA|$ of that trait in the populations of interest such that all traits contribute equally to CFA measure. This is my CFA 2. To my knowledge CFA2 has not been applied before. I consider CFA 2 because traits are likely to differ in their developmental stability and because the relationship between the magnitude of developmental stability and indexing ability is unclear (Leung and Forbes 1997a, Chapter 2). Another method of standardizing FA magnitudes across traits is to rank $|FA|$ values for each trait and then sum the ranks (e.g., Zakharov et al. 1991, Dufour and Weatherhead 1996). I call this measure CFA 3. Each of these three CFA indices can then be analyzed by a t-test, as I do in my simulations, or by an ANOVA if there are more than two samples of individuals to compare.

The fourth analysis (CFA 4), uses standardized generalized variance to quantify asymmetry (Zhivotovsky, 1992). This index, called GFA by Zhivotovsky (1992), has the advantage of statistically pooling information across characters, taking into account correlations between signed asymmetries. GFA also has the advantage of increasing degrees of freedom, which are a function of both sample size and number of traits examined. However, GFA indices are not available on many statistical packages, and tables of values are not comprehensive (see tables by Zhivotovsky 1992). Palmer (1994), noting the complexity of GFA suggested Two-Way ANOVAs as an alternative, with population (stressed versus unstressed) and traits as independent variables, and $|FA|$ as the dependent variable. This is CFA 5. I note that Two-Way ANOVAs are typically analyzed using 2-tailed testing. However, I had a priori predictions with respect to the direction of relations. Thus, I adjusted the critical value and only considered significance if the relation was in the predicted direction.

MANOVA has also been used to integrate information across traits and to determine differences in $|FA|$ between samples (e.g., Alados et al. 1993, Clarke 1993b). This is CFA 6. In this study, I used Wilks' Lambda (Λ) as my test statistic, and converted Λ into the F-statistic following Lindeman et al. (1980).

All statistics generated by my computer programs were verified by comparing output from several simulations with output generated by STATISTICA statistical package (Statsoft inc. 1993) or QUATTRO PRO spread sheet (Corel Co. Ltd. 1996). FA differences between samples for single traits were determined by comparing variances using Levene's test (i.e., t-test on absolute FA values as suggested by Palmer and Strobeck 1992).

3. Simulations

Baseline Simulations

I began by generating FA values using a baseline model. The base line model was as follows: the underlying distribution of FA values was normal, centered at zero, with variance ($\sigma^2_{p,t}$), where the subscripts p and t denote population and trait, respectively. Populations p=1 and p=2 refer to the unstressed and stressed populations, respectively. This distribution conforms to typical views of FA (Palmer and Strobeck 1986). Furthermore, FA values were independent between traits, measurement error (ME) was not considered at this point, and the underlying distributions of FA values were identical between traits (i.e., FA distributions were homogeneous between traits). For my baseline simulations, I used 10 traits (k) and a sample size (N) of 40. Subsequently, I modified the baseline model, altering parameters. I used 'conditions' as a convenient term to refer to these patterns and parameter sets that I modelled. For each condition, I examined a number of levels; e.g., I examined three levels of measurement error.

I analyzed FA difference between populations using the six CFA analyses described above. I used the ratio of variances between populations as a convenient measure of the effect size ($E = \sigma^2_{2,t} / \sigma^2_{1,t}$), where E, $\sigma^2_{2,t}$, and $\sigma^2_{1,t}$ refer to the effect size and the population variances of the putatively stressed and unstressed populations, respectively. I considered E=1, 1.2, and 1.5. When E=1, the two FA distributions are identical, and there is no stress. Here, significant differences between "stressed" and "unstressed" samples are due to Type I errors. For the other two E values, differences in the percentage of significant analyses reflect differences in power. These E values offer a range of strengths

of FA-stress relations. To determine Type I error rates and power, I ran 1000 simulations for each set of parameter values.

I found that Type I error rates were as predicted for all CFA analyses when no differences in FA variance were modelled between populations ($E=1$) (panel a of Fig 4-1, Appendix G). Thus, all CFA indices could be used validly under these baseline conditions. Power was dramatically improved using CFA analyses compared to single traits, being as much as four times as great (panel c of Fig 4-1, Appendix G). CFA 1, 2 and 5 yielded very similar power. These analyses could provide marginally greater power than CFA 3 and 4 analyses (10%-15%). CFA 6 could be as much as 50% less powerful than the other analyses (panel c of Fig 4-1, Appendix G).

Heterogeneous FA Distributions Between Traits

In nature, we might expect the magnitude of FA values to differ between traits. Traits under strong directional selection may have a high FA variance compared to other morphological traits, whereas traits whose asymmetry affects their performance may have lower FA variances. For example, Møller and Höglund (1990) found that sexually-selected traits had average FA values an order of magnitude greater than other morphological traits. Gummer and Brigham (1995) found that FA variance in bat leg tibias were about twice as large as wings. Clearly, FA can differ appreciably between traits. I considered the following five levels of heterogeneity across my 10 traits:

1) Variances of traits 1-5 were three times the variance of traits 6-10.

Traits 1-5 were homogeneous. Traits 6-10 were homogeneous [i.e., $\sigma^2_{1,1-5}=3 \times \sigma^2_{1,6-10}$].

2) Variance of one trait was three times the variance of the other traits.

The other traits were homogeneous.

3) Variances of traits 4-6 were three times the variance of traits 1-3.

Variances of traits 7-10 were ten times the variance of traits 1-3. Traits 1-3, 4-6, and 7-10 were homogeneous.

4) Variances of traits 1-5 were ten times the variance of traits 6-10. Traits 1-5 were homogeneous. Traits 6-10 were homogeneous.

5) Variance of one trait was ten times the variance of the other traits. The other traits were homogeneous.

Other parameters were as for the baseline simulations. E values were homogeneous between traits.

As with the baseline simulations, I found that Type I error rates were as predicted for all analyses (panel b of Fig 4-1, Appendix G) and that CFA analyses provided much greater power than analyses using single traits. Again, CFA 1, 2, and 5 were the most powerful analyses and CFA 6 was the weakest of the CFA analyses (panel d of Fig 4-1, Appendix G). In contrast, however, I found that CFA 2 could be marginally more powerful than CFA 1 and 5, yielding as much as 10% more power depending on the magnitude and nature of the heterogeneity (panel d of Fig 4-1, Appendix G).

Leptokurtism

Leptokurtic distributions occur when more data points fall in the center and at the tails of the distribution compared to a normal distribution (Zar 1984). Leptokurtic FA distributions may reflect a composite of many heterogeneous FA distributions (Houle 1997), which can arise due to differences in developmental stability

between individuals (Leung and Forbes 1997a, Chapter 2). Specifically, the underlying FA distribution from which an individual FA value is chosen would have a lower variance for an individual with high developmental stability compared to an individual with low developmental stability. Thus, I modelled FA variances that were heterogeneous between individuals within traits (note that $\sigma^2_{p,t}$ that was homogeneous between traits). Again, E values were as for baseline conditions.

I considered three levels of kurtosis:

- 1) $\sigma^2_{1,t}$ ranged from one to three
- 2) $\sigma^2_{1,t}$ ranged from one to ten
- 3) $\sigma^2_{1,t}$ ranged from one to one hundred.

Levels 1 to 3 produced average kurtosis values (g) of ca. $g = 0.7$, 2.9 , and 7.6 , respectively. Ranges of $\sigma^2_{2,t}$ could be calculated from E times $\sigma^2_{1,t}$.

I found that CFA 4 had inflated Type I errors that could be as high as 34% when $g = 7.6$ (top panel Fig 4-2, Appendix G). Thus, I did not consider power of CFA 4 for leptokurtic FA distributions. The other analyses had predicted levels of Type I errors. It appeared as though power of both analyses using single traits and CFAs decreased as kurtosis increased. However, it is unknown whether this was a function of kurtosis per se, or whether it was due to the additional component of variation. The most powerful CFA analysis depended upon the level of kurtosis. When kurtosis was low, CFA 1,2 and 5 provided the most powerful analyses. When kurtosis was high, CFA 3 provided the most powerful analysis and could yield up to 25% more power than the other CFA indices (bottom panel Fig 4-2, Appendix G). When kurtosis was high, CFA 2 provided marginally greater power than CFA 1.

Correlations of FA Values Between Traits

If there is a component of developmental stability that is organism-wide, then we would expect weak concordance among traits, albeit typically not significant (Leung and Forbes 1997a, Chapter 2). Further, we would expect $|FA|$ values but not necessarily signed FA values to be concordant. I modelled this by defining $\sigma^2_{p,t}$ as being composed of two variances ($\sigma^2_{p,t} = V1^2_{p,t} + V2^2_{p,t}$), where $V1^2_{p,t}$ represents the portion of $\sigma^2_{p,t}$ that is uncorrelated between traits within individuals and $V2^2_{p,t}$ represents a component of $\sigma^2_{p,t}$ that is common between traits. Absolute values comprising $V2^2_{p,t}$ were correlated but were directionally random such that signed FA values would not be correlated, but $|FA|$ would be correlated. The common component (C) of $\sigma^2_{p,t}$ could be expressed as a proportion of the total variance (i.e., $C = V2^2_{p,t} / \sigma^2_{p,t}$). The underlying distribution from which FA values were chosen was homogeneous between individuals and traits, within populations.

I examined three levels: 1) $C=0.091$, 2) $C=0.23$, and 3) $C=0.33$. Levels 1-3 corresponded to average FA-FA correlations of $r=0.0052$, $r=0.039$, and $r=0.085$, respectively. Thus, I modelled very weak concordance between traits.

I found that CFA 4 and CFA 5 yield inflated Type I error rates when $|FAs|$ were concordant across traits (top panel Fig 4-3, Appendix G). Thus, power of these CFA analyses was not considered when FA values were concordant between traits. I found that power of the other CFA analyses declined as the magnitude of FA-FA concordance increased. However, the relative positioning of these CFA analyses did not change from baseline conditions, with CFA 1 and CFA 2 yielding the greatest power (bottom panel Fig 4-3, Appendix G).

Measurement Error

Measurement error reduces the ability of FA to detect stress (Palmer 1994). It is unknown what effect measurement error should have on the relative efficacy of different CFA analyses. Similar to above, I modelled $\sigma^2_{p,t}$ as being composed of two variances ($\sigma^2_{p,t} = V1^2_{p,t} + V^2_{ME}$) where V^2_{ME} is the variation due to measurement error. The degree of measurement error (ME) could be expressed as a proportion of $\sigma^2_{1,t}$ (i.e., $ME = V^2_{ME}/\sigma^2_{1,t}$).

I examined three levels: 1) $ME=0.091$, 2) $ME=0.23$, and 3) $ME=0.33$.

Stressed and unstressed populations experienced the same value of V^2_{ME} .

Not surprisingly, the power to detect differences between traits decreased with increasing ME. However, Type I errors and relative positioning of CFA analyses in terms of power remained unchanged compared to the baseline simulations. These results are presented only in Appendix G.

Number of Traits and Sample Size

Both the number of traits (k) combined into a CFA measure and the sample size (N) would affect the power of a CFA analysis to detect stress. However, it is unknown whether these factors influence the relative efficacy of different CFA analyses. Thus, I examined different values of k and N. I used $k=5$ ($N=40$), and $N=80$ ($k=10$) and $N=20$ ($k=5$), as three separate sets of simulations. When $N=20$, I set $k=5$ because it was not possible to use CFA 4 with $N=20$ and $k=10$ (with 10 individuals in each of the stressed and unstressed populations, and 10 traits, there would be zero degrees of freedom).

NOTE TO USERS

Page(s) not included in the original manuscript are unavailable from the author or university. The manuscript was microfilmed as received.

This reproduction is the best copy available.

UMI

Fig 4-1. Type I error rates ($E=1$, panels a and b) and power ($E=1.5$, panels c and d) for six CFA analyses and analyses using single traits (ST) were shown under different conditions. The conditions were baseline (panels a and c) and FA distributions heterogeneous between traits (panels b and d). Baseline conditions were that the underlying FA distributions of all traits were normally distributed, centered at zero, uncorrelated, without measurement error, and homogeneous between traits. We used 40 individuals and 10 traits for the baseline simulations. For the other conditions, we modified one property of the model and maintained baseline properties for the rest of the model. We presented one level of heterogeneity between traits (level 5, see Appendix F and text). CFA analyses were described in Appendix 4-E.

Type I error rate

a

b

c

d

Analysis

Fig 4-2. Type I error rates ($E=1$) and power ($E=1.5$) for six CFA analyses and analyses using single traits (ST) were shown for three levels of kurtosis (g). Type I error rates were inflated for CFA 4, thus power of CFA 4 was not considered (NA=not applicable). These levels corresponded to levels 1-3 in the text and in Appendix F. CFA analyses are described in Appendix E.

Fig 4-3. Type I error rates ($E=1$) and power ($E=1.5$) for six CFA analyses and analyses using single traits (ST) were shown for three levels of correlations between FA values of traits (r). Type I error rates were inflated for CFA 4 and 5, thus power of CFA 4 and 5 were not considered (NA=not applicable). These levels corresponded to levels 1-3 in the text and in Appendix F. CFA analyses are described in Appendix E.

Analysis

4. Discussion

Fluctuating asymmetry potentially could offer several advantages as a bio-indicator of environmental stress (Clarke 1993a), provided FA is used and interpreted correctly. Researchers argue that FA should be higher in a stressed population compared to a relatively unstressed one, and that elevated FA indicates the presence of stress. Therefore, it is critical that methods of analyzing FA do not falsely indicate differences in FA between populations beyond predicted Type I error rates, but simultaneously maximize power or the probability of detecting differences in FA when they exist. I first present recommendations concerning how FA should be analyzed. Later, I present reasons for advantages/disadvantages of different analyses.

Recommendations

CFA analyses typically provided much greater power than analyses based on single traits. For example, CFA 2 could detect differences in FA among populations 85% of the time compared to only 20% using FA of single traits. In fact, to obtain the same power of CFA 2, we would require a sample size of $N \cdot k$ with analyses using single traits, where N is the sample size using CFA 2, and k is the number of traits combined (reason detailed below). Thus, CFA analyses should be preferred over analyses using single traits (see Leary and Allendorf 1989, Watson and Thornhill 1994, for qualitative arguments).

However, not all CFA analyses were equally useful. I suggest that researchers analyze data using standardized summed FA values across traits (i.e., CFA 2 and CFA 3, depending on the level of kurtosis). Both parametric

analyses (CFA 2) and non-parametric ranked analyses (CFA 3) have predicted Type I error rates and yield the greatest power. For normal FA distributions that were heterogeneous between traits and for slightly leptokurtic FA distributions, CFA 2 provided the greatest power (although the difference in power between CFA 1, 2 and 5 was less than 10%). Power of CFA 2 was essentially identical to CFA 1 (unstandardized summed FA values across traits) and CFA 5 (Two-Way ANOVAs) when FA distributions were homogeneous between traits. We should expect, however, that FA distributions typically would not be identical between traits in nature (see rationale in Leung and Forbes 1997a, Chapter 2).

If FA distributions are leptokurtic, I suggest either CFA 2 or CFA 3. For low kurtosis values, CFA 2 was up to about 15% more powerful than CFA 3, whereas for high kurtosis values, CFA 3 was up to about 25% more powerful than CFA 2. My results are consistent with those of Gangestad and Thornhill (1998) in that CFA 2 yielded predicted Type I error rates regardless of the degree of kurtosis. Further, CFA 2 was more powerful than CFA 3 when the signed FA distribution was normal, even though the $|FA|$ distribution on which analysis was conducted theoretically would be non-normal. However, my results differ from Gangestad and Thornhill (1998) in that CFA 3 (a non-parametric measure of FA) provided much greater power as the degree of non-normality of the signed FA distribution increased.

The obvious question, of course, is what constitutes high versus low kurtosis? Unfortunately, the answer is not as simple as whether the FA distributions are significantly leptokurtic. The point at which CFA 3 becomes more powerful than CFA 2 is a function of sample size and of effect size; thus, it is impossible to give a precise kurtosis value at which CFA 3 should be used. Nevertheless, I present tables of kurtosis values at which CFA 3 becomes more powerful, examining a range of sample sizes (N), effect sizes (E), and number of

traits combined (k) (Appendix H). I note that there is a lot of overlap in terms of which index should be most powerful at intermediate levels of kurtosis (figure H-1).

The other CFA analyses (4-6) were flawed. Specifically, CFA 4 (GFA) yielded inflated Type I errors when signed FA distributions were leptokurtic (cf. Monte Carlo simulations by Zhivotovsky 1988), and when $|FA|$ values were related between traits. Further, CFA 4 was much more difficult to calculate and was not amenable to most statistical packages. Similarly, CFA 5 (Two-Way ANOVA) yielded inflated Type I errors when $|FA|$ values were related between traits. For example, there was a 12% Type I error rate using CFA 5 when the average correlation was only $r=0.085$ and typically non-significant (corresponding to $r^2=0.007$ or only 0.7% of the variation explained). Thus, researchers could not use significance as an indication of when relations between $|FA|$ values of different traits would be a problem. While CFA 6 (MANOVA) had predicted Type I error rates, its power was considerably less than the other CFA indices (I could detect differences in FA between populations 50% more often using CFA 2 compared to CFA 6).

Mechanisms of Relative Benefits of CFA Indices Compared to Single Traits

The mechanisms underlying relative benefits of GFA, Two-Way ANOVAs, and MANOVA are clear. GFA and Two-Way ANOVAs reduce the standard error by increasing the effective sample size and degrees of freedom (Zhivotovsky 1992 and Palmer 1994, respectively). MANOVA estimates coefficients associated with each trait FA such that the maximal amount of variation is explained (Lindeman *et al.* 1980). The reasons why analyses using CFA 1-3 are more powerful is perhaps less clear.

As above, CFA 1-3 also act to reduce the standard error. $|FA|$ of each trait of each individual comes from a distribution of $|FA|$ values from which one value is chosen. As more traits are measured, the standard error around the mean $|FA|$ across all traits for a given individual would decrease relative to mean $|FA|$, following the central limit theorem (Zar 1984). This resulted in reduced standard error in a population. Thus, we would have an increased probability of detecting differences between populations, if $|FA|$ values of individuals in one population were generally greater than those in another population (as we would predict to occur due to stress).

Specifically, in the absence of FA-FA correlations, the standard error would be approximately reduced by a factor of $(k)^{0.5}$, where k is the number of traits combined. Thus, on average, we would require a sample size of $k*N$ using single traits to obtain the same power as analyses using CFA 2. Of course, the advantage of combining traits decreases as the degree of correlation between traits increase; for perfect correlation between traits, analyses using CFA 2 would provide no advantage over single traits.

Comparison of CFA Indices: Mechanisms of Differences in Power and Validity

CFAs 1-5 had similar power under baseline conditions because all of them acted to reduce the standard error (although not with the same efficiency). The difference in power between CFA 2 and CFA 3 reflected the difference between parametric versus non-parametric CFA measures; these differences in power depended upon the degree of non-normality (Appendix H). When FAs were heterogeneous between traits, higher power of CFA 2 versus CFAs 1 and 5 occurred because CFAs 1 and 5 would implicitly weight asymmetric traits more heavily than less asymmetric traits. To illustrate the potential consequences of heterogeneity consider a trait whose FA variance was an order of magnitude larger than other traits (e.g., Møller and Höglund 1990). CFA 1 and 5 would primarily reflect this one trait. Thus, the standard error would primarily reflect a single trait rather than being optimally reduced by taking the standardized average of many traits. Not surprisingly, CFAs 1, 2 and 5 were essentially identical when FA distributions were homogeneous between traits. In such cases, traits would be weighted approximately equally using these analyses. However, CFA 1 and 5 could potentially be equal or stronger than CFA 2 if highly asymmetric traits were consistently more strongly affected by stress.

As noted above, CFA 5 treated traits as independent data points and therefore had inflated Type I errors when $|FA|$ values of traits were correlated. GFA was also sensitive to correlations of unsigned asymmetries. Although GFA takes into account signed asymmetries, and hence may not be biased by developmentally linked traits, unsigned asymmetries may be correlated without correlations between signed asymmetries. Hence, GFA could be biased by slight concordance among traits (albeit typically non-significant) due to some organism

wide component of developmental stability or quality (Leung and Forbes 1997a, Chapter 2).

MANOVA was typically the weakest index. This was because a separate parameter needed to be estimated for each trait. Thus, more variability needed to be explained to find significance. In contrast, CFA 2 increased in power following the central limit theorem, but did not concurrently require additional variation to be explained to find significance. MANOVA also had to be two-tailed, since each trait had a parameter estimate that could be in either direction, whereas the other analyses could be analyzed using one-tailed testing. Although there are situations in which two-tailed testing would be appropriate, MANOVA would still be the weakest index.

Researchers also may ask whether single traits sometimes can be more reliable than CFA indices. If the effect sizes of FA-stress relations are heterogeneous, this is certainly possible. CFA 2 typically would be more powerful than the average of single traits. However, CFA 2 may be a combination of strong FA relations and weak FA relations. Weak FA relations would act like random noise and could result in a weaker CFA index compared to reliable single traits. However, two problems would be associated with using reliable single traits. First, consideration of multiple individual traits would require some correction for multiple tests, and would therefore result in reduced power (e.g., sequential Bonferonni, Palmer 1994). Second, it is unknown which trait would be most reliable. Thus, a priori testing would be required to find such traits without causing inflated Type I error rates. Further, if there were a priori reasons for expecting one trait to be more reliable than another, it could be possible to weight CFA 2 values by the estimated reliability of individual traits (e.g., taking into account measurement error). However, one cautionary note regarding such weighting procedures is that inappropriate weighting would weaken CFA 2

indices. Here, CFA 2 would reflect some traits more than others; hence reduction in standard error due to taking averages would not be as great. Thus, while weighting may provide advantages even over CFA 2, optimal methods of weighting have yet to be determined.

FA-Quality Relations Within Samples

This study dealt exclusively with detection of FA differences between samples. However, many studies examine relations between FA and some measure of quality (e.g., mating success) within studies. The logic underlying the findings of this paper extends to tests of FA-quality relations. Further, although single traits sometimes relate significantly to quality, FA-quality distributions within samples should typically be triangular in distribution, and FA of single traits may only be predictive of quality at high FA values (Leung and Forbes, 1997a, Chapter 2; but cf. Rowe *et al.* 1997 for antisymmetry-quality distributions). However, with composite measures, symmetry need not be the common phenotype due to the central limit theorem, and FA can be potentially predictive of quality over the entire range of FA values (Appendix I).

5. Summary

It may be possible to use FA to monitor stress. However, current methods of analyses usually employ FA of single traits and appear to be weak and unreliable. If FA is to be used, it is essential to maximize the reliability of FA as a monitor of stress. Composite measures of FA can provide vast advantages over single traits in terms of the probability of detecting FA differences between populations. However, different methods of CFA analyses differ both in their

validity (Type I error rates) and power, and it is critical to determine which CFA analyses should be used. In this study, I demonstrate that the most powerful and robust CFA analyses are simply t-tests using standardized summed FA values. When kurtosis is low, parametric summed FA values provide the greatest power (CFA 2). When kurtosis is high, non-parametric summed ranked FA values provide the greatest power (CFA 3).

CHAPTER 5

Does Fluctuating Asymmetry Reflect Quality?

A Laboratory Study using a Damselfly (Ishnura verticalis) Model System.

1. Introduction

Changes in quality or fitness of organisms or demographics of populations may arise due to differences in environmental conditions (i.e., the presence of environmental stress). The role of a bioindicator is to monitor or predict reduction in quality or fitness of individual organisms or decline of populations. Thus, for a bioindicator to be valid, changes in the bioindicator should reflect changes in either current or future quality, an assumption which is often untested (Bunn 1995).

Fluctuating asymmetry (FA) has been receiving increasing attention as a putative bioindicator (Leary and Allendorf 1989). FA has received attention because FA has been found to increase in the presence of putative stresses and because there is substantial evidence that FA is correlated with life history traits, such as mating success or longevity, which are believed to relate to quality (reviewed in Leung and Forbes 1996, Chapter 1). Thus, FA has been suggested as a biologically meaningful bioindicator because differences in quality between populations should be reflected by differences in FA (Clarke 1995, Sommer 1996).

In this study, I tested the theoretical applicability of FA as a bioindicator using a damselfly (Ishnura verticalis) model system. I tested whether differences in FA reflected quality within samples and related to differences in the quality between samples subjected to different environmental conditions (stresses). I

used the life history traits wet mass and longevity in the laboratory as surrogates of quality. I used differences in food availability and the presence/absence of parasitic mite to generate differences in quality between groups.

My reasons for my choices of life history traits and environmental conditions were as follows: reduced wet mass in adult damselflies is correlated with reduced longevity (e.g., Leung and Forbes 1997b) and size can relate to lifetime reproductive success (Harvey and Walsh 1993). In males, lifetime reproductive success also is strongly linked to longevity (Banks and Thompson 1985). Although longevity was measured in the laboratory in the absence of feeding, this measure of quality likely has importance in nature as well. Evidence for I. verticalis suggest that starvation conditions may occur frequently for adult damselflies in the field (Richardson and Baker 1997). The ability to survive, given nutritional deprivation, should be important. Further, in a related damselfly, Enallagma ebrium, 24% of damselflies died within 2 days in the absence of feeding (Leung and Forbes 1997b). E. ebrium does not feed in inclement weather (Forbes pers. obs). In our study area, periods of 2 consecutive days of rainy weather was common. Other studies with inclement weather also showed that the largest male damselflies had the greatest lifetime reproductive success (Harvey and Walsh 1993, cf. Banks and Thompson 1985 where inclement weather was rare). Thus, the probability of surviving such stressful periods, and reaching sexual maturity was dependent upon longevity in the absence of feeding, which in turn related to wet mass.

Nutritional deprivation during larval stages was used in this study as it has been shown to decrease the size and weight of adult damselflies (Richardson and Baker 1997). I also used mite parasites as a potential stress because a previous study found that mite intensity was inversely associated with FA of damselflies (Bonn et al. 1996). Unfortunately, the larval damselflies removed all

mites before damselfly emergence. Prior to emergence, Arrenurus pseudosuperior mites are phoretic on their hosts and do not directly feed (Smith 1988). Thus, I could only examine the consequences of mite colonization on life history traits and FA. Consequently, I de-emphasize the parasitism portion of my study and focus on the nutritional treatment portion of my experiment.

Methods

Field Collections

I collected 74 males and 158 females F-1 larval Ischnura verticalis (F indicates final instar, F-1 penultimate instar, etc.) during June 1997 at Burns Reservoir (43° 29'N; 80° 02'W), South Western Ontario. I collected A. pseudosuperior mite eggs at Charleston Lake (44°32'N, 76°01'W).

Experimental Setup

I kept F-1 damselfly larvae at 10° C and fed them sparingly to prevent damselfly larvae from molting to the final instar before mites hatched. When mites hatched, I transferred the F-1 larvae housed in 75 mL vials to an environmental chamber at 25°C (16:8 L:D) and fed them 1 white worm (Enchytraeidae) per day. On the day larvae molted to the final instar, it was placed in either the high food availability (1 white worm/day) or low food availability (1 white worm/3 days) environment. Animals in the high food environment were not given worms ad libitum because uneaten worms died and fouled the water.

I tested the effect of mite treatment by placing damselflies into mite absent and mite present groups. Damselflies in the mite present group, were exposed to

a one time treatment of >50 mites during July 2 to July 10. Mite larvae were pipetted from their aquaria and placed in vials containing damselflies. After being subjected to mites for 1-2 hours, I placed damselflies in vials with water free of mites where they were raised to emergence, under their assigned nutritional group. As mentioned, mites were removed from damselfly larvae before emergence. Thus, I could not examine direct effects of parasitism by mites on damselflies, since Arrenurus mites are phoretic on larvae and do not directly parasitize their hosts until after emergence (Smith 1988). Further, the effect of mite colonization on life history traits (longevity and wet mass) was negligible (Two-Way MANOVA: Mite treatment: Wilk's $\Lambda=0.99$, $p>0.75$, Sex: Wilk's $\Lambda=0.89$, $p<0.005$, $df=2$, 105). For these reasons I do not consider mite colonization further in this study.

Data Collection

I recorded wet mass of newly-emerged adults and whether damselflies were phenodeviants. Phenodeviants had curled wings or wings that failed to extend; they could not fly and would quickly die in nature. I excluded phenodeviants from my main analysis as FA could not be measured on these individuals and they were clearly abnormal compared to individuals found in nature; I analyzed frequency of phenodeviants later. I measured wet mass of individuals using an analytical balance (accurate to 0.0001 g), 24 hours after emergence. I waited 24 hours to allow damselfly wings to dry and harden before handling; this reduced damage and variability in mass due to moisture. After weighing, animals were housed in 100 mL plastic cups supplied with screen mesh over 1 cm deep water; water helped prevent desiccation. Each cup contained a wooden dowel as a

perch and was covered with perforated aluminum foil to prevent escape. I checked for moulted, emerged, and dead damselflies once daily.

After damselflies died, I measured both forewings (FW) and hindwings (HW) from the nodus to the 2nd radial vein (R2, Askew 1988), and the metathoracic tibiae (MT) and mesothoracic tibiae (MS). I used a projector to enlarge images onto a screen (17 x magnification for tibiae and 13 x magnification for wings). Each enlarged image was drawn twice onto paper and measured using digital calipers to estimate measurement error. Means were used to reduce measurement error (Yezerinac et al. 1992).

Analysis

I first examined whether life history traits differed between samples, without which I may not expect to find differences in FA between samples. I used Two-Way ANOVAs with sex and food availability as independent variables and life history traits as dependent variables. I predicted a priori that individuals with lower food availability should have lower wet mass and shorter longevity. I conducted one-tailed tests. Typically, ANOVAs are viewed as two-tailed. However, I treated ANOVAs as one-tailed tests by first considering the direction of relations and then adjusting the significance level to reflect one-tailed predictions. I also examined relations between longevity and wet mass to examine the internal consistency of our life history traits (using an ANCOVA to control for sex and nutritional group). I predicted positive relations between longevity and wet mass. I only present analyses based on individuals included in our FA analyses (see below for reasons why some individuals had to be excluded). I note however, that analyses based on all normal individuals yielded

normal individuals yielded the same conclusions. I conducted sequential Bonferonni corrections (Rice 1989).

Basic Statistics of Fluctuating Asymmetry

I began my analysis of FA by testing for potential confounding factors. I assayed for directional asymmetry in each trait by subtracting left from right values and determining if 95% confidence intervals overlapped zero (Palmer and Strobeck 1986). Anti-symmetry was indicated by significant platykurtosis (platykurtosis has negative kurtosis values) (Palmer and Strobeck 1986). I note that there is controversy over whether these other forms of asymmetry may be used to index stress or quality, and whether these forms of asymmetry actually relate more reliably to stress or quality (Graham *et al.* 1993, Rowe *et al.* 1997).

I tested for allometric relations between FA and trait size following methods presented by Leung (1998, Chapter 3). I only controlled for positive relations between FA and trait size, as this is when allometry may cause spurious FA-quality relations (Leung 1998, Chapter 3). I also presented the coefficient of variation of size (CV_{size}). Consequences of allometry should decrease as CV_{size} decreases. I corrected for allometry using relative FA ($|right-left|/avg(right+left)$) if (CV_{size}) was low. Use of relative FA had little bias when CV_{size} was low, even when FA-size relations were not isometric (Leung 1998, Chapter 3).

I determined measurement error using a Two-Way mixed model ANOVA with sides as the fixed factor, individual as the random factor and trait length as the response variable (Palmer and Strobeck 1986). Following Yezerinac (1992), percent measurement error (%ME) was defined as:

$$\%ME = s^2_{within} / (s^2_{within} + s^2_{among}) \times 100$$

Where s^2_{within} is the variance of measures within sides of individuals (i.e. due to measurement error) and s^2_{among} is the true FA variation. However, the contribution of measurement error to the total variance is considerably less, because I used the mean of two measurements for each individual. Following logic from the central limit theorem (Zar 1984, pp86-87),

$$s^2_{\text{within}'} = (s^2_{\text{within}})/m$$

5-2

where $s^2_{\text{within}'}$ is the variance of mean measurement values due to measurement error and m is the number of measurements. Replacing s^2_{within} in equation 5-1 with $s^2_{\text{within}'}$ of equation 5-2 estimates the contribution of measurement error using mean rather single measurements of each individual. I identify this contribution as %ME_{mean}.

Fluctuating Asymmetry Versus Quality Among and Within Samples

I produced a composite measure of asymmetry (CFA) based on the 4 traits measured. I standardized each FA of each trait by dividing by the mean FA magnitude for that trait and then combined FAs of traits so all traits contributed equally to the CFA measure (CFA 2, Chapter 4, Appendix F). This CFA index requires all traits to be intact; thus, I excluded individuals with injured or missing traits from the analysis.

To examine whether FA reflected quality between groups, I used a Two-Way ANOVA with sex and food availability as independent factors and CFA as the dependent variable. To examine whether FA related to quality within groups, I

used an ANCOVA model to control for sex and food availability, and examined correlations between CFA and life history traits. Predictions were one-tailed for CFA to be higher in the group exposed to low food availability and for inverse relations between CFA and wet mass or longevity.

I used parametric statistics even though FA values are not normally distributed following recommendations by Gangestad and Thornhill (1998). Simulations by Leung (1998, Chapter 3 and Chapter 4) also indicated that parametric statistics showed appropriate Type I errors when using testing for FA relations using linear correlations.

Phenodeviants

I compared the frequency of phenodeviants between groups exposed to high and low food availability using a chi-square test. Following Zar (1984), I pooled the frequencies for males and females as relations appeared homogeneous ($\chi^2=0.29$, $df=1$, $p>0.5$). Next, I compared wet mass and longevity of phenodeviants versus normal individuals within each sex-stress combination. Finally, I attempted to isolate the factor causing phenodeviation.

Two aspects of my experimental design may have caused the high incidence of phenodeviants. First, because I had intended to examine the consequences of mite parasitism, I kept F-1 damselflies under cool conditions to slow development, to coincide the timing of emergence of mites with moulting into the final instars of damselflies (see above). Second, I fed damselflies white worms, because white worms were easier than Daphnia (a traditional food for larval damselflies) to raise in large numbers.

I separated these two potential causes of phenodeviation by conducting a 2x2 design during late October 1997 to mid January 1998. Individuals in F-1

were collected from the field and half were housed in a refrigerator at ca. 6° C and half were kept at room temperature (ca 20° C). All were fed daphnia during this week. Following the first week, all larvae were kept at room temperature and half from each temperature treatment were fed ad libitum with worms and half were fed ad libitum daphnia every two days. Larvae were reared until they emerged or died. We used a Chi-Square test to analyze results.

Results

56 animals died before emergence, 66 were phenodeviants, and 110 emerged properly. Of the 110 animals that emerged properly, only 50 animals had all traits intact, a requirement of my FA analyses. I found that food availability was significantly related to both longevity and wet mass using a one-tailed Two-Way ANOVA to control for sex, and conducting sequential Bonferonni corrections. Specifically, individuals in the low food treatment had lower wet mass and survived in the lab for a shorter period of time (Table 5-1). Further, using a Two-Way ANCOVA to control for sex and food availability, wet mass and longevity were positively correlated as predicted ($r=0.37$, $t=2.69$, $df=45$, $p<0.005$). Thus, it was reasonable to assay for differences in FA between the two feeding environments.

Basic Statistics of Fluctuating Asymmetry

I found a general trend towards leptokurtosis (positive kurtosis values) with 14 of 16 tests (4 traits in each of 4 groups) having positive kurtosis values and 5 showing significant leptokurtosis ($p<0.05$, Table 5-2). There was no obvious tendency for directional asymmetry (Table 5-2).

Next, I examined relations between size and life history traits and size and FA to determine whether I should correct for allometry. I found that CV_{size} was generally small (2%-7%) and that size was typically uncorrelated with life history traits (Table 5-2). Thus, allometric relations between size and FA should not greatly influence detection of relations between FA and life history traits. Nevertheless, I tested for FA versus residual size relations and found such relations for the trait MS in 2 groups (FA-size relations were not significant for the other traits; Table 5-2). For my FA-quality relations within groups, I corrected using relative FA values for these groups. I note, however, that results were the same whether or not I corrected for allometric relations.

I used a Two-Way ANOVA to control for sex and found that size of wings differed between feeding treatments (FW: $F_{1,46}=21.66$, $p<0.001$; HW: $F_{1,46}=16.84$, $p<0.001$) but tibias did not (MT: $F_{1,46}=0.72$, $p>0.25$, MS: $F_{1,46}=1.28$, $p>0.25$). Specifically, size was larger with high food availability. While it was possible that FA-size allometry could confound detection of FA differences between groups, the effect would likely be small given the small CV_{size} and the general lack of FA-size correlations. I tested for FA differences between groups using both FA uncorrected for size and relative FA, and found that both yielded the same conclusions. Thus, allometry likely was not a problem and I presented results for only FA uncorrected for size for my test of FA differences between groups.

I tested for measurement error. I found that the interaction term between sides and individuals was significant for all traits, indicating that FA was present in all traits (MS: $F_{108,218}=6.82$, $p<0.0001$, MT: $F_{118,238}=31.72$, $p<0.0001$, FW: $F_{108,218}=37.44$, $p<0.0001$, HW: $F_{99,200}=7.78$, $p<0.0001$). Percent measurement error was calculated as 2.67%, 12.85%, 14.64%, and 3.15% for FW, HW, ML, and HL, respectively.

Fluctuating Asymmetry Versus Quality

CFA was not significantly different between feeding groups (Table 5-3) even though food availability affected quality as measured by wet mass and longevity. However, the trend was in the correct direction. In contrast, an ANCOVA model, controlling for sex and food availability, indicated that CFA was inversely related to both longevity and wet mass in the predicted direction (Table 5-4). Further the relation between CFA and longevity or wet mass was consistently inverse as predicted, within each group. Thus, there was evidence that FA related to quality within samples, but did not reflect differences in quality between samples.

As expected, the strength of relations was strongest using CFA compared to using FA of single traits; r-values were approximately twice as great using CFA compared to the average of single traits. Although FA-quality relations using single traits were consistently inverse, as predicted, we likely would have concluded that FA was not generally related to quality using single traits, as most relations were non-significant, even without sequential bonferonni adjustments. Thus, the use of CFA measures appeared to offer large advantages over the use of FA of single traits.

Phenodeviants

I found that there was a significantly greater frequency of phenodeviants in the well fed group compared to the poorly fed group ($\chi^2=6.4$, $N=176$, $df=1$, $p<0.02$). The phenodeviants had lower wet mass and longevity than normal individuals in all groups (Fig 5-1). Finally, phenodeviation appeared to be caused by the white worm diet. All individuals fed *Daphnia* emerged as normal adults, whereas 13 of

26 individuals fed white worms emerged as phenodeviants (Table 5-5). There was no indication that temperature influenced the incidence of phenodeviation.

Table 5-1. Differences in life history traits (longevity and wet mass) between nutritional groups (Food) are shown using a 2-Way ANOVA. We had one-tailed predictions that longevity and wet mass would be lower in the group exposed to lower food availability. The abbreviations were as follows: Males and females exposed to high food availability (MH and FH, respectively) and low food availability (ML and FL, respectively). Sequential Bonferonni corrections were conducted for the two life history traits. Degrees of freedom for each level of the Two-Way ANOVA was $df=1, 45$.

		Wet mass	Longevity
Mean	FL	0.015	3.40
	FH	0.017	3.71
	ML	0.012	2.56
	MH	0.016	4.33
Two-Way ANOVA			
Food	F-value	16.48	3.39
	one-tailed p-value	<0.001	<0.05
Sex	F-value	10.26	0.04
	p-value	<0.001	>0.75
Interaction	F-value	0.51	1.69
	p-value	>0.20	>0.20

Table 5-2: Basic statistics associated with FA. Kurtosis (Kurt), mean signed FA values (Mean), standard errors (se), FA-size relations (Pearson's r), relations between size and life history traits (Pearson's r), and the coefficient of variation of size (CV_{size}) are shown for each group. To test for allometry, FA was regressed against size using residuals after removing life history trait variation from size variation following Leung (1998), Chapter 3. Size corrections were only conducted for positive FA-residual size relations. The abbreviations were as follows: Males and females exposed to high food availability (MH and FH, respectively) and low food availability (ML and FL, respectively); for each morphological trait: forewing (FW), hindwing (HW), mesothoracic tibia (MS), and metathoracic tibia (MT); for life history traits: longevity (Long) and wet mass (Mass). Tests of kurtosis, means, and size-quality relations were two-tailed and FA-size relations were one-tailed (predicting positive relations). Sample sizes were 9, 17, 9, 15 for MH, FH, ML, and FL, respectively. Significance ($\alpha < 0.05$) in the predicted direction was indicated by *.

Table 5-2.

		Mean(se)	Kurt(se)	FA vs residual		Size vs		CV _{size}
				size		quality		%
				Mass	Long	Mass	Long	
FL	FW	0.10(0.17)	2.21(1.12)	0.33	0.41	0.22	0.23	5.36
	HW	-0.42(0.59)	10.02(1.12)*	0.10	0.05	0.23	0.26	5.77
	MS	-0.41(0.32)	0.88(1.12)	0.36	0.44*	-0.30	-0.27	4.67
	MT	0.10(0.37)	1.27(1.12)	-0.04	-0.22	-0.41	-0.09	3.48
FH	FW	0.03(0.53)	14.53(1.06)*	0.07	-0.01	0.33	0.07	3.34
	HW	0.08(0.12)	-0.83(1.06)	-0.50	-0.50	0.35	0.14	3.61
	MS	0.37(0.34)	1.91(1.06)	0.33	0.34	0.06	0.20	6.43
	MT	0.01(0.18)	0.31(1.06)	-0.03	-0.07	0.24	0.16	4.64
ML	FW	0.00(0.23)	1.45(1.40)	0.16	0.01	0.08	-0.71*	5.99
	HW	0.15(0.58)	1.60(1.40)	0.56	0.37	0.04	-0.72*	6.42
	MS	-0.42(0.43)	3.89(1.40)*	-0.17	-0.37	-0.11	-0.61	4.59
	MT	0.05(0.26)	-2.25(1.40)	-0.03	0.22	-0.14	-0.66*	5.75
MH	FW	-0.13(0.35)	3.20(1.40)*	-0.10	-0.11	0.26	0.28	3.45
	HW	-0.26(0.35)	5.50(1.40)*	-0.12	-0.09	0.15	0.19	3.06
	MS	-0.05(0.17)	0.94(1.40)	0.85*	0.75*	-0.06	-0.42	4.00
	MT	-0.12(0.29)	0.34(1.40)	0.18	0.02	0.33	-0.06	2.13

Table 5-3. Differences in CFA between nutritional groups (Food) are shown using a Two-Way ANOVA. I had a one-tailed prediction that CFA would be lower in the nutritionally unstressed groups. Groups are as in Table 1. Degrees of freedom for each level of the Two-Way ANOVA was $df=1, 45$.

Group	Mean CFA values	
FL	1.05	
FH	0.95	
ML	0.98	
MH	0.74	

Two-Way ANOVA		
Food	F	0.56
	One-tailed p	>0.2
Sex	F	0.39
	P	>0.5
Interaction	F	0.10
	P	0.75

Table 5-4. Relations between CFA and quality within samples and between FA of single traits and quality are shown using an ANCOVA model to control for sex and nutritional group. Single traits were analyzed only to allow comparison of the average strengths of FA relations using single traits compared to CFA measures (Chapter 4). Traits are as in Table 5-2. Sequential Bonferonni corrections were conducted separately for CFA analyses. Relations between CFA and quality were one-tailed predicting inverse asymmetry-quality relations.

Life history trait	Trait Asymmetry	Degrees of Freedom	Pearson's r-value	t-value	One-tailed p-value
Wet mass	CFA	45	-0.34	-2.39	0.02
Longevity	CFA	45	-0.27	-1.86	0.04
	Single Traits				
Wet mass	FW		-0.04	-0.26	0.40
	HW		-0.12	-0.82	0.42
	MS		-0.17	-1.12	0.40
	MT		-0.34	-2.44	0.04
	Average		-0.17	-1.16	
Longevity	FW		-0.15	-0.99	0.49
	HW		-0.08	-0.51	0.31
	MS		-0.12	-0.84	0.40
	MT		-0.21	-1.42	0.32
	Average		-0.14	-0.94	

Table 5-5. Frequency of phenodeviants under different conditions of temperatures and feeding. Damselfly larvae were either exposed to either 20° C or 6° C temperatures during their F-1 larval stage. Larvae were fed either Daphnia or white worms during their final larval instar.

		Frequency of phenodeviants	Frequency of normal individuals
Daphnia	20° C	0	12
	6° C	0	12
White worms	20° C	7	6
	6° C	6	7
$\chi^2=16.42$		df=4	p<0.005

Fig 5-1. Differences in life history traits (top panel: longevity and bottom panel: wet mass) between phenodeviants and normal individuals are shown. The abbreviations were as follows: Males and females exposed to high food availability (MH and FH, respectively) and low food availability (ML and FL, respectively). Sample sizes for the above groups were 7, 21, 3, 16, respectively for phenodeviants, and 17, 33, 20, 40, respectively for normal individuals. All differences between phenodeviants and normal individuals were significant. The probability of 4 of 4 relations being significant at $\alpha=0.05$ was $p<0.00001$ (using binomial expansion).

4. Discussion

Fluctuating asymmetry has been suggested as a potential bioindicator (Leary and Allendorf 1989). Moreover, FA has been tied to both the presence of putative stresses as well as differences in individual quality (Leung and Forbes 1996, Chapter 1). Thus, FA may be particularly relevant because differences in FA should reflect differences in the quality of samples due to different environmental conditions (stress levels) (Clarke 1995).

In this study, I found equivocal support for FA as a bioindicator. On the one hand, I found that FA related to life history traits of individuals within samples, but did not reflect differences in life history traits between samples. This would suggest that FA might not be a good bioindicator of the average quality of a sample, despite relations that may be found between FA and quality within samples (see Leung and Forbes 1996 for a review, Chapter 1). This interpretation, of course, is dependent upon the validity of my life history traits as measures of quality. First, I will discuss reasons for why the life history traits used in my study may be (or may not be) good measures of quality. Next, I will discuss potential reasons for why FA related to life history traits within samples, but did not reflect differences between samples. Finally, I will discuss implications of this study for the use of FA as a bioindicator of stress and of differences in quality due to stress.

Validity of Life History Traits

There was clearly stress present in both high and low food groups. This highlights the importance of testing for the validity of stresses (i.e., whether the 'stressed' group was in fact of lower quality), a factor not addressed in most

studies (as argued by Leung and Forbes 1996, Chapter 1). I had generated two different 'environmental conditions' by providing one group with high food availability and the other with low food availability, with the intention that the group with low food availability would represent the 'stressed' group and would have lower quality. Unfortunately, my use of a white worm diet had adverse effects. It turned out that a white worm diet resulted in a high frequency of phenodeviation, with the group exposed to high food availability having a higher frequency of phenodeviants. Phenodeviants could be regarded both as very low quality individuals (as they would likely die quickly in nature) and as developmentally unstable individuals (by definition, phenodeviants were not able to develop properly).

Clearly, phenodeviants were of low quality. The question that remained was whether normal individuals differed in quality between groups, and thus whether I should assess differences in FA between groups of normal individuals. Normal animals exposed to high food availability had significantly higher wet mass and longevity than poorly fed animals. The dilemma was whether the greater longevity and wet mass of well fed normal animals reflected higher quality, or whether the higher frequency of phenodeviants of the well fed animals suggested that normal well fed animals were of lower quality and that the life history traits measured were invalid under my experimental conditions. For example, it could be that individuals with higher wet mass were actually of lower quality when fed a white worm diet, given the high fat and low chitin content of white worms.

I argue that the life history traits measured remained valid despite the higher frequency of phenodeviants in the well fed group, and that the well fed normal individuals were of higher quality. First, phenodeviants and normal individuals were clearly different, and a higher frequency of phenodeviants in the

well fed group did not necessarily indicate that normal well fed individuals had compromised quality compared to poorly fed individuals. Second, phenodeviants, which should have severely compromised fitness, had lower wet mass and shorter longevity. This was consistent with the interpretation that my life history traits remained valid measures of quality. Third, despite the higher frequency of phenodeviants in the well fed group, well fed normal individuals had higher wet mass and longevity in contrast to the lower wet mass and longevity of phenodeviants, suggesting that the effect of diet was different for normal individuals and phenodeviants. Fourth, my rationale for using longevity measured in the laboratory would not change, regardless of diet. The ability to survive longer under nutritional deprivation would be important given inclement weather that occurred in my study area. Damselflies do not feed in inclement weather (Forbes, pers obs.), and an animal that can survive periods of rainy weather would be more likely to survive the maturation period and have the opportunity to mate. Further, I. verticalis appears to often experience nutritional deprivation in the field (Richardson and Baker 1997). After maturation, longevity remains an important factor for lifetime mating success (Banks and Thompson 1985). Finally, wet mass and longevity were positively related to one another, as predicted. If being exposed to high food availability and having higher wet mass was detrimental to normal individuals, then there would be no reason to expect wet mass and longevity to remain positively correlated. Thus, there was no reason to expect that the inflated frequency of phenodeviants invalidated my life history traits as measures of quality. Given that life history traits appeared to be behaving as expected and differed significantly between well fed and poorly fed groups, it was reasonable to assess differences in FA between groups of normal individuals to determine whether differences in FA reflected differences in quality.

Relations between FA and Life History Traits

As mentioned, FA did not differ significantly between samples, despite significant relations between FA and life history traits within samples and significant differences in life history traits between samples. The discrepancy of FA relations within and between samples was not predicted. One possibility is that a Type II error occurred for the test of FA differences between samples but did not occur within samples. The alternative explanation would be that environmental factors could affect quality and developmental stability independently, such that in the presence of heterogeneous stresses, FA and quality may no longer be correlated.

The ideal relation between FA, stress, and quality seems to be that stress reduces quality, and increases FA via that decrease in quality (Fig 5-2a). Due to this causal relation, a change in FA would reflect a change in quality and therefore would be biologically relevant. This might occur because stress is energetically costly (Koehn and Bayne 1989, Mayer *et al.* 1992), and stress may therefore reduce the amount of energy available for processes related to quality and developmental stability (Mitton 1993). Despite this general theory, I note that little is known about the actual mechanisms whereby stress can compromise developmental stability. Thus, it seems possible that some environmental factors may directly affect developmental stability without having a proportional effect on quality (Fig 2b).

Direct effects of white worms on developmental stability may account for the discrepancy in my results. White worm diet resulted in an increased frequency of phenodeviants, another measure of developmental stability. It is possible that white worm diet directly compromised pathways responsible for development, which may have affected FA. If this were the case, FA could be

affected without a proportional change in quality. Well fed individuals would not have experienced nutritional deprivation, but may have encountered greater negative effects of the white worm diet because they were fed more. The combination of a greater negative effect of white worm diet and benefits of increased food intake could have cancelled out differences in FA between groups. In contrast, FA may reflect quality within samples as all individuals within samples would have been fed identically and the direct effect of white worm diet on developmental stability would have been approximately homogeneous between individuals within samples. Thus, the presence of heterogeneous stresses may decouple relations between FA and quality.

Implications of Results for FA as a Bioindicator

My results indicate that we cannot necessarily extrapolate from finding FA-quality relation within samples to the ability of FA to reflect quality between samples and to the usefulness of FA as a bioindicator (cf. Clarke 1995). Further, if the absence of relation between FA and life history traits between samples occurred due to the presence of heterogeneous stresses, each affecting FA and quality differently, then the interpretation of patterns of FA becomes more difficult. One of the most attractive features of FA was that it has been linked to both putative stresses as well as to fitness traits (Clarke 1995). Thus, in contrast to other biomonitoring methods, the biological relevance of FA appeared to be clear. However, if factors can affect FA without affecting quality, then the biological relevance of FA becomes less clear. Direct effects of environmental factors on FA could be potentially a very serious problem. While it has already been argued that FA may not be applicable in all traits for all biological systems (e.g., Leung and Forbes 1996, 1997a, Chapters 1 and 2), there may be some traits of some

organisms where FA typically relates to quality. However, if unusual environmental conditions influence FA without affecting quality, then relations between FA and quality could be confounded, even when previous evidence suggests that a system is reliable. For example, in my study, the absence of differences in FA did not indicate the absence of differences in quality and could have been due to direct effects of white worm diet on FA. I note however that conditions in the laboratory may have differed from natural conditions (i.e., the effects of the white worm diet). Thus, it is unknown how often stress directly affects FA and confounds FA-quality relations in nature.

While it may appear that I am suggesting that FA would not be a useful bioindicator, this is not necessarily the case. I believe that it is important to identify the potential problems associated with using FA as a bioindicator, such that FA can be interpreted properly. Potential problems include: 1) direct effects of stress on FA (this chapter). 2) relations between FA and quality or stress should not occur in all traits (Leung and Forbes 1997a, Chapter 2). 3) relations between FA and quality or stress may be obscured by pre-sampling mortality (Campbell and Emlen 1996), allometry (Leung 1998, Chapter 3), or measurement error and power (Palmer 1994). These problems indicate that FA is not universal nor as easy to use as previously suggested (e.g., Clarke 1993a). The applicability of FA would be much narrower and would require testing of the reliability of FA for each biological system of interest, before we could have confidence in our interpretations of FA patterns. However, other bioindicators each have their own problems as well (e.g., Bunn 1995, Sommer 1996), and typically require a priori knowledge of natural history and a priori testing of the conditions which may influence the validity and interpretation of each bioindicator. Thus, it may be timely to compare explicitly the costs and benefits of FA versus other bioindicators, using a more balanced view of FA.

Fig 5-2. Flow diagram showing possible relations between the environment (stress), quality, FA, and life history traits. In an ideal scenario (a), there is an underlying inverse relation between FA and quality. Stress decreases quality and increases FA via this decrease in quality. In this scenario, an increase in FA indicates a decrease in quality due to stress; the absence of differences in FA indicates the absence of differences in quality. Alternatively, factors in the environment (i.e., white worm diet) may influence FA beyond their effect on quality (b). In this case, heterogeneous stresses may confound relations between FA and quality (i.e., high white worm availability simultaneously increases quality, thereby decreasing FA, but also directly affects developmental stability and increases FA). FA may relate to quality under homogeneous environments (i.e., FA-quality relations within samples in this study), but FA may no longer reflect quality given heterogeneous environments (i.e., absence of FA-quality relations between samples in this study). Since quality can not be directly measured we would be constrained to use life history traits serve as surrogate measures. Thus, the strength of correlations between stress and life history traits or FA and life history traits depends upon the relation between the life history trait and underlying quality. Note that life history traits do not directly affect FA, or vice versa, but are correlated via underlying quality.

a

b

CHAPTER 6

Fluctuating Asymmetry in Relation to Indices of Quality in Damselflies (Enallagma ebrium) Captured in the Field.

1. Introduction

In this study, I tested the assumption that FA is related quality for the damselfly Enallagma ebrium (Hagen) caught in the field. I also tested whether FA was positively related to mite intensity as found in another study (Bonn et al. 1996). I used the life history traits wet mass and longevity as measures of individual quality. The choice of these life history traits were discussed in Chapter 5. I first examined the relation between longevity and wet mass to assess whether survivorship was non-random with respect to my measure of quality.

I also examined the relation between FA and the intensity of parasitism by mites. Mites are important as they negatively influence survival and mating success (Forbes 1991a). Although Arrenurus mites are phoretic on developing larvae and only begin to feed on damselflies after they emerge (Smith 1988), damselflies of higher quality may be less susceptible to mite parasites. Thus, while mites should not directly influence quality during larval development, mite intensity may still correlate with FA. Alternatively, mite parasitism immediately after damselfly emergence may cause a hydrostatic stress and may directly influence wing development and increase FA, as argued by Bonn et al. (1996). Regardless of the mechanism, the hypothesis is clear: I predict a positive relation between FA and the intensity of mite parasitism (see also Bonn et al. 1996).

2. Methods

Study Site

My study site was at Two Island Lake (TIL), previously described in Forbes (1991a). TIL is located within 10 km of Chaffeys Locks, Ontario, Canada (44° 34'N; 79° 15'W).

Field Collections

I collected 4 groups of E. ebrium on June 28-30, 1995 consisting of 32 males and 33 females with mature colouration (henceforth termed adults), and 48 males and 44 females lacking mature colouration (mauve body colouration with vitreous sheen to wings). Damselflies lacking mature colouration were newly-emerged, pre-reproductive adults, and were probably 1-2 d old (henceforth termed tenerals). For comparisons involving mature adults, I controlled for approximate age by collecting only damselflies with engorged Arrenurus mites (Forbes 1991b). Arrenurus mites parasitizing E. ebrium engorge over the pre-reproductive period and drop off after the host reaches sexual maturity (Forbes 1991a). By including damselflies that were parasitized by one or more engorged mites, I ensured that damselflies were past the pre-reproductive period, but not old enough for mites to have detached. I estimated that my adult damselflies had emerged 14-17 days earlier (following work on damselflies in outdoor insectaries, Forbes 1991a).

All adult females were captured in tandem with males away from the pond edge to ensure that females had not already oviposited (following Forbes and Baker 1991). Thus, I controlled for reproductive condition as females that oviposited might be lighter than females that had not oviposited that day.

During transport, damselflies were placed in rectangular cages with nylon meshing on the top and sides (42 cm X 30 cm X 30 cm). To prevent desiccation of damselflies, wet paper towels were placed at the bottom of cages. A wire screen on top of the paper towel precluded adhesion of damselflies to the wet paper towel. Teneral and adults were placed in separate cages to avoid injuries to the weakly flying teneral, especially given that the exocuticles of teneral were not fully hardened and their wings often were not fully dried.

Variables

I recorded wet mass (measured to the 0.0001 g using an analytical balance) and the number of mites for each damselfly. For mature adults, I corrected for the collective mass of engorged mites by taking the average weight of 110 mites (3.086×10^{-5} g) multiplied by the number of engorged mites on any adult. I did not correct for the negligible wet mass of unengorged mites on teneral damselflies. Unengorged mites were so small that their removal was impossible without destroying them.

I used longevity of individuals housed in the laboratory as a second measure of quality. Damselflies were not fed during the course of these experiments. Such nutritional deprivation often occurs during a series of rainy days as damselflies do not forage at that time (Forbes unpublished data). All damselflies were housed individually in numbered 285 mL plastic cups provided with perforated, 10 cm lids of aluminum foil. Each cup contained a 7.5 x 0.3 cm perch of doweling, a 5 cm diameter screen at the bottom of the cup, and ca. 25 mL of water underneath the screen to prevent desiccation. Individuals were kept in darkness at ca. 100% humidity (to simulate dark rainy days). Individuals were checked twice daily (at 0800 h and 2000 h) and deaths were recorded.

I measured asymmetries of five traits: the tibia of the metathoracic tibia (LEG), the forewing from the nodus to the pterostigma (FWNP), the forewing from the pterostigma to the second radial vein (FWP2), the hindwing from the nodus to the pterostigma (HWNP), and the hindwing from the pterostigma to the second radial vein (HWP2). To reduce measurement error, I used an Image-Analysis Digitizer which greatly enlarged the size of each character (10-20x), and took the average of three measurements of each character for each individual (following Yezerinac et al. 1992). FA was determined for each character pair as $| \text{right-left} |$.

I produced a composite measure of asymmetry using the five traits described above. I standardized each FA of each trait by dividing by the mean FA magnitude for that trait and then combined FAs of traits so all traits contributed equally to the CFA measure (i.e., CFA2, Chapter 4, Appendix F). I excluded individuals with injured or missing traits.

Basic Statistics of Fluctuating Asymmetry

I tested for directional asymmetry and antisymmetry as described in Chapter 5. I tested and corrected for allometric relations between FA and trait size as described in Chapter 5 (also see Leung 1998, Chapter 3). I determined measurement error using a Two-Way mixed model ANOVA as described in Chapter 5, except here analyses were based on three replicates of each of 31 individuals. Measurement error for the average of three measurements per trait was calculated as in Chapter 5.

Analyses

First, I examined relations between life history traits. I predicted a positive relation between wet mass and longevity following reasoning discussed in Chapter 5. I also examined relations between mite intensity and wet mass and longevity. Mite parasitism may directly reduce wet mass and longevity of adults. In such a case, it may be necessary to control for mite parasitism in examining relations between FA and life history traits. Mite parasitism may also relate to wet mass of newly emerged teneral. Although this likely would not occur due to direct parasitism (since there would have been little time since emergence for mites to engorge), it may occur if wet mass was related to the susceptibility to attachment by mites. I examined relations using ANCOVAs to control for sex (male/female) and damselfly age (teneral/adult).

Following these preliminary analyses, I tested for relations between CFA and life history traits and CFA and mite intensity using ANCOVAs to control for sex and age. I also examined relations between CFA and residual longevity and residual wet mass after controlling for mite intensity (using partial correlations). As mentioned, mites can affect longevity of aquatic insects (Smith 1988) including damselflies (Forbes and Baker 1991). Parasitism by larval water mites can depend on factors such as microhabitat or timing of emergence of insects (Smith 1988) rather than quality of individuals per se (but see Forbes and Baker 1990, Leung and Forbes 1997b).

Results

Tests of Other Asymmetry Distributions and Assessment of Measurement Error

Only one of sixteen tests showed a significant departure from a mean signed FA value of zero (Table 6-1). This was probably due to chance rather than due to directional asymmetry, given the number of tests conducted. I found no significant platykurtism, thus no evidence for antisymmetry. The five significant departures from normality showed positive kurtosis, i.e., leptokurtosis (Table 6-1).

Only one trait of one group showed significant positive relations with size (Table 6-1). This likely was due to chance rather than a true allometric relation given the large number of tests conducted. Thus, I did not correct for allometry. Further, the coefficient of variation of size (CV_{size}) was low (2%-7%), and thus, allometric relations would not greatly affect relations between FA and life history traits.

I tested for measurement error using a Two-Way mixed model ANOVA. The interaction term between sides and individuals was significant for all traits (LEG: $F_{30,124}=6.32$, $p<0.0001$, FWNP: $F_{30,124}=13.21$, $p<0.0001$, FWP2: $F_{30,124}=6.78$, $p<0.0001$, HWNP: $F_{30,124}=15.27$, $p<0.0001$, HWP2: $F_{30,124}=7.81$, $p<0.0001$). Thus, FA was present in all traits. The measurement error or %ME was 17% to 36% of the asymmetry variation in characters (36.0% of the variation in LEG FA, 19.7% of FWNP FA, 34.1% of FWP3 FA, 17.4% of HWNP FA, and 30.6% of HWP3 FA). By using the mean of three measurements for each character of each individual, the proportion of FA variation due to measurement error (%ME_{mean}) was reduced to 15.8% in LEG FA, 7.6% of FWNP FA, 14.7% of FWP3 FA, 6.5% of HWNP FA, and 12.8% of HWP3 FA.

Relations Between CFA, Wet Mass, Longevity and Mite Intensity

Wet mass was positively related to longevity and mite intensity was inversely related to both wet mass and longevity (Table 6-2). Thus, life history traits and mite intensity related to one another as expected and I felt it reasonable to assay for relations between CFA and life history traits.

I found no evidence of the predicted relations between FA and longevity, wet mass, or mite intensity (Table 6-2). Because mite intensity was related to longevity and wet mass, I factored out mite intensity using partial correlations and re-examined the relations between FA and corrected longevity and corrected wet mass. Again, there was no evidence that FA related to life history traits after controlling for mite intensity (Table 6-2).

Kurtosis(se)		Mean(se)	FA vs		Size vs		CV _{size} %			
			residual size		quality					
			Mass	Long	Mites	Mass	Long	Mites	CV _{size}	
AF	LEG	1.59 (0.80)	0.02 (0.02)	0.00	-0.01	0.01	0.25	0.07	-0.18	3.83
	FWNP	-0.19 (0.8)	0.02 (0.01)	-0.15	-0.17	-0.21	0.66*	0.40*	-0.22	3.55
	FWP2	1.26 (0.8)	0.01 (0.01)	0.18	0.19	0.17	0.08	-0.14	-0.31	4.72
	HWNP	-0.79 (0.82)	0.00 (0.01)	0.15	0.23	0.20	0.62*	0.51*	-0.07	4.11
	HWP2	-0.95 (0.82)	0.00 (0.01)	-0.28	-0.25	-0.19	0.20	-0.06	-0.43	5.24
TM	LEG	24.36 (0.7)*	0.05 (0.03)	-0.75	-0.75	-0.74	0.23	0.06	-0.13	5.28
	FWNP	1.45 (0.69)*	0.01 (0.01)	-0.07	-0.14	-0.04	0.53*	0.37*	-0.09	3.59
	FWP2	-0.25 (0.7)	0.00 (0.01)	-0.15	-0.09	-0.07	0.20	0.05	-0.01	6.28
	HWNP	30.5 (0.71)*	-0.02 (0.03)	0.38*	0.47*	0.37*	0.51*	0.36*	-0.14	3.90
	HWP2	2.28 (0.73)	0.00 (0.01)	-0.05	-0.06	-0.08	0.37*	0.20	-0.14	5.65
TF	LEG	0.72 (0.72)	0.00 (0.01)	-0.47	-0.48	-0.48	0.38*	0.20	0.04	4.21
	FWNP	0.79 (0.7)	0.01 (0.01)	0.12	0.18	0.13	0.35*	0.34*	-0.01	2.67
	FWP2	0.2 (0.73)	-0.01 (0.01)	0.16	0.16	0.18	0.39*	0.13	-0.11	5.14
	HWNP	-0.54 (0.72)	0.03 (0.01)*	0.02	0.09	0.06	0.30*	0.37*	0.01	2.72
	HWP2	-0.03 (0.72)	-0.02 (0.01)	0.04	0.09	0.11	0.35*	-0.01	-0.15	4.23

Table 6-2: Comparisons between longevity (LONG), wet mass (MASS), intensity of mite parasitism (MITE), and CFA. ANCOVAs were used control for sex (male/female) and age (adult/teneral) and examine relations between continuous variables. Tests were one-tailed. I predicted positive relations between longevity and wet mass, inverse relations between mite intensity and life history traits, inverse relations between CFA and life history traits, and positive relations between CFA and mite intensity. All relations were in the predicted direction and one-tailed p-values were reported, with the exception of CFA versus wet mass (which was therefore not considered). Sequential Bonferonni corrections were conducted for each set of hypotheses [i.e., 1) correlaton between life history traits, 2) correlation between CFA and life history traits]. df=degrees of freedom, r= Pearson's correlation coefficient, t=t value, and p=probability of occurring due to chance.

Comparison	df	R	t	p
LONG vs MASS	149	0.34	4.39	<0.001
MITE vs LONG	151	-0.21	-2.61	<0.01
MITE vs MASS	149	-0.18	-2.25	<0.02
<u>CFA analysis</u>				
CFA vs LONG	125	-0.05	-0.57	>0.2
CFA vs MASS	124	0.14	1.61	N/A
CFA vs MITE	125	0.01	0.08	>0.2
residual life history, controlling for mites				
CFA vs LONG	125	-0.05	-0.61	>0.2
CFA vs MASS	124	0.15	1.63	N/A

4. Discussion

Researchers have argued that higher quality individuals should have greater developmental stability and lower FA on average (e.g., Møller 1993a, Polak and Trivers 1994). Thus, FA should be negatively related to life history traits such as longevity or mating success as has been found (e.g., Møller 1992b, Naugler and Leech 1994, Ueno 1994). Two previous studies demonstrated significant negative relations between FA of wings and lifetime or short-term mating success in two species of damselflies (Harvey and Walsh 1993, Cordoba-Aguilar 1995, respectively). In contrast, I found that FA did not relate to longevity even though longevity has been highly correlated with lifetime mating success in other species of coenagrionid damselflies (Fincke 1988, Banks and Thompson 1985). Moreover, FA was not related to an index of quality: wet mass of newly-emerged or mature adult damselflies.

One explanation for my non-significant results, is that FA is not consistently related to quality. Such neutral results are important, but only if other factors that may confound predicted relations have been controlled for or assessed. Important other factors include the following: inappropriate choice of traits to measure FA (Watson and Thornhill 1994, Møller 1993a, Palmer 1994), measurement error (Palmer 1994), and inappropriate measures of quality.

My choice of characters was appropriate because all five characters showed true FA. Moreover, FA-fitness relations have been shown in other species of damselflies using similar characters (i.e. damselfly wing length; Harvey and Walsh 1993, Córdoba-Aguilar 1995). However, I note that these other studies measured traits on live animals with calipers and did not assess

measurement error. I assessed measurement error in my study, and found that measurement error was low and that FA was detectable.

In order to assess whether my measures of quality were appropriate, it was useful to ask whether these measures behaved as expected. I expected longevity and wet mass to be positively related. If longevity was not related to wet mass, then it could suggest that study conditions invalidated these life history traits as measures of quality, for some unknown reason. If longevity and wet mass related positively, then I would be more confident that these life history traits were good measures of quality. I found that my life history traits behaved as predicted.

My measures of quality likely were appropriate in nature as well as in my laboratory setting. 15% percent of my damselflies died within the first day and a half of being nutritionally stressed in the laboratory, 24% within 2 days, and over half within 3 days. Consecutively rainy days ranged from 1 day to 7 days with an average of 1.8 days (data from QUBS weather station). Given that E. ebrium do not feed during rainy periods, it is not surprising that numbers of mature damselflies often appear to decline after a series of rainy days. Thus, nutritional stress likely occurs in nature and longevity measured in the lab likely reflects the probability of surviving to maturation (and to potential matings) in nature under these stressful conditions. Also, wet mass was an appropriate measure of quality (the probability of high fitness) given that it related to longevity under stressed conditions, and given that nature is often variable and stressful. Thus, if FA reflects quality as hypothesized, I would predict inverse FA-longevity (and FA-wet mass) relations under the conditions of my experiment.

I also had expected relations between FA and intensity of mite parasitism (as found by Bonn et al. 1996). Mites may be stressful during development, given that number of mites was inversely related to wet mass in teneral (Forbes and

Baker 1990, Leung and Forbes 1997b). Alternatively, mites may cause a short term hydrostatic stress during wing expansion, as argued by Bonn et al. (1996). However, I found no relation between FA and mite intensity.

While it is impossible to definitively exclude all possible confounding factors, the simplest and most likely explanation is that FA is not consistently related to quality, at least across damselfly species. Interestingly, these results differ from results of Chapter 5 of this thesis, in which I had found predicted relations between FA and life history traits within samples for another species of damselfly. One possibility is that the relation between FA and quality is trait- and organism-specific (as argued in Chapter 2, Leung and Forbes 1997a). While I had used similar traits between studies, FA of these traits may relate differently to quality between species. Alternatively, the discrepancy between results may reflect differences between experimental studies and field studies. In Chapter 5, all animals were raised under identical conditions within samples. Here, animals were collected from the field and may have encountered heterogeneous stresses. If stresses differ in their effect on quality and FA, then the presence of heterogeneous stresses could obscure relations between quality and FA (see arguments in Chapter 5).

My study may have differed from other field studies on closely related organisms which found predicted relations, such as Harvey and Walsh's (1994) and Córdoba-Aguilar's (1995) studies, because FA may sometimes influence flight performance in the field [e.g., aerodynamic efficiency (Palmer 1994)]. This was a laboratory study, which simulated specific conditions of rainy weather in nature during which damselflies do not feed. However, the possibility that FA-fitness relations are due to direct consequences of FA on performance needs further testing.

Notwithstanding, researchers should be cautious when using FA to index quality. It would be erroneous to conclude that differences in FA for E. ebrium (and possibly many other organisms) indicate differences in quality. This admonition is also supported by other work. Namely, a quantitative synthesis of the current FA literature (i.e., a meta-analysis), which suggests that FA relations are typically weak and highly heterogeneous (Chapter 1, Leung and Forbes 1996), and a mathematical model, which suggests that relations between FA and quality (and FA and stress) should be of variable strength due to theoretic considerations (Chapter 2, Leung and Forbes 1997a). If FA is to be used as a bio-indicator of differences in quality, its validity must be assessed for each study system.

CHAPTER 7

GENERAL DISCUSSION:

A Critical Assessment of Fluctuating Asymmetry as a Bioindicator.

1. Introduction

Fluctuating asymmetry (FA) is a research topic that has generated both staunch advocates, as well as harsh critics. Advocates have variously stated that "morphological variability [FA] may provide a valuable early indicator of environmental and genetic stress" (Leary and Allendorf 1989); "No existing techniques designed to detect possible adverse impact on native animal populations can match the FA approach for sensitivity" (Clarke 1993a); "FA is now known to be diagnostic for a wide range of environmental and genetic stressors..."; "FA is a good measure of phenotypic quality. Indeed, it may be the best measure we have" (Polak and Trivers 1994) ; "Developmental stability [FA] may be used as a surrogate for more direct fitness estimation for use in conservation biology and biomonitoring" (Clarke 1995); "Indications are that developmental stability [FA] of individuals could provide a sensitive in situ marker of population fitness" (Sommer 1996); "...it [FA] could be viewed as a general health certificate of an individual" (Møller and Swaddle 1997). Other researchers have been more critical of the potential of FA as a bioindicator: "Hence, FA may be a poor general predictor of either stress or fitness [due to the low average effect size observed]" (Leung and Forbes 1996, Chapter 1); "Just how powerful is this tool [FA], and how reliable? Perhaps very, but those wishing to apply it must recognize that the biological signal is exceedingly small and that not all deviations from symmetry provide a useful signal" (Palmer 1996); "The problem

in determining the importance of asymmetry is that the effects of many of the phenomena are, if present, so small that observational and experimental results may easily be influenced one way or the other by seemingly trivial details of experiment or observation" (Houle 1998).

Given the increasing interest and controversy surrounding FA, it would be timely to review what we now know about FA and develop a balanced view on the usefulness of FA as a bioindicator of environmental stress or of individual quality. In brief, the appropriate use of any tool requires an understanding of its limitations as well as its benefits. In the following sections, I critically examine FAs potential as a bioindicator, using nine criteria to judge the efficacy and purpose of FA. I briefly examine several other methods of biomonitoring and their limitations to put FA into context. Finally, I come to a conclusion about the applicability of FA as a bioindicator.

2. Examining Fluctuating Asymmetry as a Bioindicator

Fluctuating asymmetry is only one of many possible biomonitoring techniques. Other techniques can range from toxicity testing using LC_{50} values to looking at possible bioaccumulation based on chemical properties of pollutants (Bioconcentration factors, BCF). Biomonitoring methods also include specific biomarkers that indicate exposure to specific stresses, as well as more general bioindicators that may be influenced by a number of stresses and may give some indication of overall organismal quality (reviewed in Kovacs 1992, Huggett *et al.* 1992, Kramer 1994). Some methods are well established and required under (US) law. Others are like FA and are highly speculative and controversial. I compare FA to four methods that are commonly used and well studied. Although previous authors have briefly compared FA versus LC_{50} levels (see general

introduction, Clarke 1993a), few comparisons have been made between FA and other bioindicators.

I will discuss toxicity testing (LC_{50} levels), estimations of bioaccumulation, Cholinesterase inhibition (as the representative stress-specific bioindicator), and Scope for Growth (SFG) (as the representative general stress bioindicator). I compare FA to these methods because these methods appear to be the most widely accepted and represent the breadth of possible methods. Traditional environmental monitoring consists primarily of measuring LC_{50} levels (Heath 1998) and is required under law (e.g., pulp and paper effluent regulation in Fisheries Act SOR/92/269). The LC_{50} is the concentration of chemical causing lethality to 50% of test organisms. Examination of the bioconcentration factor (BCF) is required under the EPA regulation. BCFs are predicted on the basis of quantitative structure-activity relations (QSARs); QSARs essentially measure the properties of chemicals (specifically the hydrophobicity) that relate to the potential to accumulate in biological tissue (Barron 1995). Cholinesterase (AChE) inhibition arguably represent the best success story of bioindicators (Mineau 1998), and represent characteristics of toxicant-specific bioindicators (other specific bioindicators include aminolevulinic acid dehydratase, which responds to lead, or metallothionein, which responds to certain heavy metals, see Stegeman et al. 1992). AChE inhibition is one of the few bioindicators formally accepted as a legal argument to indicate environmental damage (Mineau 1998). Scope for Growth (SFG) is probably the most developed and rigorously field tested of the general stress indicators (Mayer et al. 1992). SFG has been widely used to monitor aquatic environments (Smaal and Widdows 1994). SFG measures the energy potentially available for growth and reproduction. Other general bioindicators exist and show promise but have not been as well studied, have not achieved common use in biomonitoring, and will not be discussed here [e.g.,

heat stress proteins (Stegeman et al. 1992), or biological early warning systems (BEWS) examining physiological or behavioural changes (Baldwin and Kramer 1994)].

Criteria that we can use to evaluate bioindicators include:

1. **Biological relevance.** In other words, a change of the bioindicator should reflect a current or future change in the quality of individuals or populations (Huggett et al. 1992, Mayer et al. 1992, Bunn et al. 1995).
2. **Sensitivity.** Ideally, the bioindicator should be more sensitive than traditional endpoints such as survival, reproduction, and growth. The relative sensitivity compared to other bioindicators would also be important (Huggett et al. 1992).
3. **General versus specific indicators** (Mayer et al. 1992, Mineau 1998). General indicators may reflect decreased quality due to a variety of stresses. General indicators may be important to indicate the presence of some potentially unknown stress and potentially reflect the general quality of individuals or a population (FA would be a general indicator). General indicators would not identify the causal agent. Specific indicators may detect the presence of specific stresses (e.g., cholinesterase inhibition is a bioindicator of pesticides that interfere with cholinesterase activity, Melancon 1995).
4. **Relative ease and cost of use** (Huggett et al. 1992, Mineau 1998).
5. **Biological specificity.** Can the bioindicator be used in many biological systems (Huggett et al. 1992)? Can the bioindicator be measured in endemic organisms, or does it require test organisms (Mayer et al. 1992)?
6. **Field validation.** Has the biomarker been tested in the field, and is it amenable to use under field conditions (Huggett et al. 1992)?

7. Quantifiable change in response. Bioindicators should respond in a dose or time dependent manner to the stress so the magnitude of exposure or effect can be determined. (Mayer et al 1992, Mineau 1998).
8. Ecological Factors. Variability due to other factors such as season, temperature, sex, weight, etc. should be understood and acceptable (Huggett et al. 1992, Mayer et al. 1992).
9. Permanence of response. Is the bioindicator transient, or a permanent consequence of exposure or effect of stress (Huggett et al .1992).

These criteria provide guidelines to assess the purpose and utility of bioindicators, although not all criteria need (or can) be fulfilled for a bioindicator to be useful (Mayer et al. 1992). I use these criteria as a framework to critically examine the usefulness of FA.

1) Biological relevance

One of the most attractive features of FA is that it is believed to be biologically relevant and to relate strongly to quality (e.g., Polak and Trivers 1994). Thus, differences in FA would imply that either a change in quality had occurred or would occur in the future. This assertion appears to be valid in some biological systems, given that FA is related to quality (Clarke 1995, Leung and Forbes 1996, Chapter 1). However, the ability of FA to reflect organismal quality likely is not a property of all biological systems, given the heterogeneity of empirical relations observed (Leung and Forbes 1996, Chapter 1) and given theoretic considerations (Leung and Forbes 1997a, Chapter 2).

For systems in which FA relates to quality, FA may provide advantages over traditional monitoring techniques such as LC₅₀s or BCFs (see also Clarke 1993, Sommer 1996). Toxicity for one species typically is not indicative of toxicity

for another species (Peakall 1992). Similarly, BCF values obtained for a test species in the laboratory can not be used to predict BCFs for other species or the same species in the field (Sijm 1991). Thus, use of test organisms (which is typically done) may have little relevance for natural populations of interest. Further, while BCFs may indicate the accumulation potential of a substance, it would not indicate toxicity of that substance (McKim and Schmieder 1991).

The biological relevance of bioindicators such as AChE inhibition appear to be clear. AChE activity relates to organism function; high levels of inhibition can directly cause death (50% inhibition is used to indicate that death occurred due to pesticides, Melancon 1995). Lower levels of inhibition can influence behaviour and physiology, which may in turn affect fitness (Mayer *et al.* 1992, Mineau 1998). Thus, AChE inhibition is obviously biologically relevant. SFG would also likely have relevance given that SFG measures the potential amount of energy available for growth and reproduction, two traits that should be important for fitness (Smaal and Widdows 1994).

2) Sensitivity

There are two aspects of sensitivity that are relevant in FA analysis. One is power of analyses using FA to detect stress or changes in quality. The other is the timing at which relations appear (i.e., does stress affect FA before it affects other life history traits). Due to developmental noise, we should expect FA of single traits to have low power and to be only a weak indicator of environmental stress or of individual quality (Leung and Forbes 1997a, Chapter 2). This problem may be ameliorated by using composite measures (Chapter 4), but even composite measures can do little if the relation between developmental stability and quality is weak or absent (Chapter 6). Thus, even with composite measures,

the sensitivity of FA would likely be case-specific, given that there is no reason to predict that the relation between developmental stability and quality should be homogeneous across species.

It is possible that sensitivity of FA may be greater than traditional endpoints (e.g., survival, Clarke and McKenzie 1992). The arguments and mathematical models (e.g., Leung and Forbes 1997a, Chapter 2), that suggest FA (of single traits) should not be sensitive, are with respect to statistical power; they do not take into account life history trade-offs which may occur. FA may have minimal consequences to organism function compared to other measures such as growth, reproduction, or survival. Under stressed conditions, organisms may allocate resources to more important life history traits and compromise developmental stability. Thus, FA may be increased before traditional life history traits are affected, and may predict potential future effects on fitness (Clarke and McKenzie 1992). Some evidence suggests that FA changes in the presence of stress when other traditional monitoring techniques fail to indicate a stress (Clarke 1993a), and that FA changes before other endpoints such as survival (Clarke and McKenzie 1992). However, other evidence suggests that changes in life history traits due to stress can occur without changes in FA (Premchand *et al.* 1998). Thus, this aspect of sensitivity would also likely be system-specific.

While there is some evidence that FA can sometimes be more sensitive to detect a potential problem than monitoring techniques such as BCFs or LC₅₀s (Clarke 1993a), to my knowledge there has not been comparison of sensitivities of FA versus techniques such as AChE inhibition or SFG. Like FA, AChE inhibition and SFG are also suborganismal; thus, responses of these bioindicators to stress may occur before changes in traditional life history traits (Mayer *et al.* 1992).

3) General versus specific indicators

It is debatable what types of characteristics are useful in bioindicators. Some argue that bioindicators should be specific and only responds to one or a few toxicants (e.g., Mineau 1998). Others argue that general indicators are useful too, both to detect a wide range of stresses (Mayer et al. 1992), as well as to take into account synergistic interactions between stresses and predict organism quality (Mayer et al. 1992; Clarke 1993). In fact, some researchers have noted that interactions between multiple stresses, both anthropogenic and natural, must be taken into account for remedial action to be effective (Adams et al. 1998).

FA would be classified as a general indicator. FA has been shown to respond to a wide range of stresses and has been correlated to many life history traits (see Møller and Swaddle 1997 for a review). Thus, FA's use as a bioindicator (if any) would be to indicate the overall quality of a population (or individuals) and to identify which populations could be at risk, rather than determining the exact cause of the reduction in quality (Clarke 1995). As mentioned, SFG would also be a general indicator and would be affected by many stresses. Like FA, SFG may indicate the overall quality of a population (or individuals) and may be able to respond to synergistic interactions between stresses. In contrast, AChE inhibition is relatively specific to pesticides. Thus, causality is not as much of a problem compared to FA analysis. However, due to the specificity of AChE, it would not be useful to indicate most other stresses that may influence quality (Heath 1998).

Techniques such as LC₅₀ tests and BCFs could be applied to any toxicant. In fact, a huge data base exists of LC₅₀ values for variety of chemicals and test organisms; LC₅₀ may be useful to compare relative toxicities and species sensitivities of different chemicals (Heath 1998). A huge data base also exists for

BCFs and could be used to indicate the potential for accumulation of a given substance (Barron 1995). However, these methods do not indicate whether a population is actually under stress in nature.

4 and 5) Relative ease and cost of use and biological specificity

Other advocated advantages of FA are that FA is very easy and cheap to measure (Clarke 1993a), FA does not require sacrifice of animals, and FA is relatively non-intrusive. Use of FA can be divided into three sections: data collection, data analysis, and interpretation. FA is typically measured as the difference between the right and left sides of a bilateral trait (Palmer and Strobeck 1986). Thus, the technical difficulty in measuring FA is minimal, and would require minimal experience and minimal equipment. Cost and ease of data collection is perhaps the clearest advantage of FA. The analysis of FA has several complications, including allometry and measurement error (see Leung 1998, Chapter 3, for a discussion of allometry, Palmer 1994 for a discussion of measurement error; and Leung and Forbes 1997b, Chapter 5, to determine percent reduction of measurement error by using multiple measurements). Despite these complications, analysis of FA is still relatively simple, requiring only a basic knowledge of statistics that most ecologists should have (linear regression and ANOVA models). The interpretation of differences in FA between samples, or the lack thereof, requires consideration of many factors, and will be discussed below. Thus, the major limitation of FA is interpreting changes in a quantitative and biologically meaningful way (e.g., how much has quality changed?).

FA may also be advantageous because it can be measured on almost any organisms. In fact, FA has been studied on traits ranging from teeth to feathers,

and from organisms ranging from plants to humans (see Leung and Forbes 1996, Chapter 1, Møller and Swaddle 1997, for reviews). However, as noted above, FA may not relate to quality or stress in all biological systems.

For comparison to FA, AChE is also inexpensive to analyze (although likely more expensive than FA analysis, since FA requires nothing more than morphometric measurements). Effect of percent inhibition of AChE activity appear to be fairly consistent for mammals and birds (Mineau pers. com.), however relations may be more variable for other vertebrates and little is known for invertebrates (Mineau 1998). Thus, the reliability of AChE inhibition may also be specific to the biological system. One disadvantage of examining AChE is that AChE inhibition is typically measured in the brain (note that AChE activity differs in different parts of the brain, Peakall 1992), and would require sacrifice of animals; in contrast, FA can be non-intrusive. While AChE can also be measured non-destructively in blood samples, activity is much more variable and inhibition is short lived (Melancon 1995). Thus, blood AChE inhibition could indicate exposure to pesticides, but likely would not reflect effect.

For SFG, one limitation is that measurement of individual physiology requires more training than other bioindicators (e.g., FA); further initial costs of equipment is high (Smaal and Widdow 1994). Finally, most work is currently done using mussels as a test organism, due to ease of handling and other reasons (discussed in Smaal and Widdows 1994). Effects of stresses on mussels may not reflect effects on other species of interest. Unfortunately, SFG may be more difficult to measure on other organisms. Unlike mussels, handling and laboratory manipulations may in and of itself cause stress to some organisms. Thus, SFG may not be amenable to many biological systems.

6 and 7) Field validation and quantifiable changes in response.

Here, I discuss the areas that perhaps is most lacking in FA studies. FA has been extensively measured in the field showing that it is amenable to field analysis and that FA relates (sometimes) to stress and quality under field conditions. In this respect FA has been field validated in many biological systems. However, FA has been studied in most species only once, under one set of environmental conditions, and for one specific stress (Møller's work on barn swallows is a notable exception; see Møller and Swaddle 1997 for a review; see Houle 1998 for criticism of their work). Repeated tests should be done to determine if FA of a species of interest would consistently reflect changes in quality or stress. Ideally, we would want to quantify changes in FA with respect to changes in quality or stress (similar to a dose-response curve, Fig 7-1) such that we could interpret a change in FA in terms of quality or stress (Mayer et al. 1992, Mineau 1998). Unfortunately, such response curves are rare in the FA literature (but see Zakharov et al. 1991).

Predictive FA-quality curves would need to be determined under a range of environmental conditions. Heterogeneous stresses, both natural and anthropogenic, are likely found in the environment. In fact, one of the potential benefits of use of FA is that FA should reflect synergistic interactions between stresses and that FA should reflect overall quality of individuals or populations (Clarke 1995). In the ideal scenario, FA would correlate with quality, stress would reduce quality and would increase FA via that decrease in quality. Here, FA would be consistently linked to quality, and FA could potentially be used to predict quality (Fig 5-2). However, if some environmental factors directly affect developmental stability (and FA) then it may be very difficult to predict quality using FA (see Chapter 5). Disproportionate effects of environment on FA and quality can be potentially a serious problem, and may limit the utility of FA as an

predictor of quality for some biological systems. Thus, it would be important to test the sensitivity of FA-quality relations under different environmental conditions, given the variety and combinations of stresses that may be present in the environment.

Finally, we do not have the necessary evidence to conclusively state why FA relates to quality or why stress increases FA, despite mathematical models (Emlen et al. 1993, Graham et al. 1993) and verbal arguments (see introduction, Mitton 1993). One unknown is whether there is some maximum asymmetry, beyond which FA no longer reflects changes in stress. If stress increases FA through a reduction of developmental stability (as typically argued, e.g., Palmer and Strobeck 1992), then we may expect that as stress increases and developmental stability decreases, FA may reach an asymptote (where FA reflects only developmental noise, Leung and Forbes 1997a, Chapter 2). In contrast, if stress increases FA both through increasing developmental noise and decreasing developmental stability (Emlen et al. 1993), this asymptote would not necessarily be reached. We do not currently have the empirical data to determine whether such an asymptote exists.

In comparison to FA, field validation for LC₅₀ levels is rare and results obtained under laboratory conditions do not necessarily reflect what would occur under field conditions. In the lab, organisms are usually exposed to a given concentration of a toxicant for a relatively short duration of time. In nature, exposure could have occurred for years (Peakall 1992). Finally, there usually should be numerous stresses in nature that are often unmeasured. Synergisms between stresses (natural and anthropogenic) are essential to address, given that stresses do not exist nor act in isolation in nature (Adams et al. 1998). Thus, while dose response curves can be generated under very specific conditions for traditional toxicity tests, these tests would not take into account synergism

between stresses, and therefore may not be predictive of toxicity in nature (Clarke 1993a). Similarly, BCFs are typically determined in the laboratory. Like LC₅₀ tests, BCFs would not take into account synergism between stresses or rates of accumulation in nature (see below).

For AChE inhibition, dose response curves can be obtained under a given set of conditions. However, AChE inhibition would not take into account the synergistic interactions of multiple stresses. Thus, while AChE would clearly indicate exposure, its relation with death is more controversial (Meyer et al. 1992). For example in fish, death after exposure has occurred anywhere between 5.4% to 92% of normal brain AChE activity, but survival may occur even with activities as low as 10% to 20% of normal (Mayer et al. 1992). It should be noted that the confidence that we do have with AChE inhibition was derived through extensive testing and validation (Mineau 1998). For previously untested species (e.g., most invertebrates, Mineau 1998), extensive testing would again be required to generate base line data (Melancon 1995).

As mentioned SFG has been rigorously field tested (Mayer et al. 1992). Further, SFG, which is simply a measure of the amount of available energy, may range from positive values under benign conditions to negative values near lethality. Thus, SFG should provide a quantifiable measure of stress, and should relate to overall organism quality. However, while animals are obtained from the field, actual SFG measurements are conducted in the laboratory (Smaal and Widdow 1994). Thus, observed SFGs do not necessarily reflect natural situations. Further, as mentioned, SFG measures the potential rather than the actual amount of energy used for growth and reproduction; accurate prediction of actual growth and reproduction using SFG is more difficult (Smaal and Widdow 1994).

8) Ecological factors.

A good understanding of the natural history of organisms would be required to interpret FA correctly. Factors such as age have been correlated with FA, possibly due to differential mortality of high FA individuals (e.g., Forbes et al. 1997). Sexual differences may exist (e.g., Møller 1993b). Seasonal effects may play a role for short-lived organisms (Forbes et al. 1997). Other factors such as temperature or weight would not necessarily be a problem, because FA would be viewed as a general indicator of overall quality, and these factors arguably affect quality. Finally, lethal stresses and pre-sampling mortality may seriously compromise detection of FA differences between samples (Campbell and Emlen 1996). If FA is related to quality, as is often assumed, then it seems reasonable that high FA individuals should have a higher probability of mortality. Stress may increase FA of a sample, but if more individuals die in the stressed sample, and if individuals with high FA disproportionately suffer, then FA may not differ between samples, despite the presence of stress.

Pre-sampling mortality does not necessarily invalidate FA as a bioindicator, but it forces us to reinterpret differences (or the lack of differences) in FA between samples. If differential pre-sampling mortality has occurred, we may not be able to use FA to determine whether a stress is present. However, FA may still be useful to reflect the general quality of a sample; i.e., pre-sampling mortality that removes the high FA individuals may also remove the low quality individuals, such that FA and quality remain correlated. As such, FA's role (if any) should be as a potential general quality indicator of a sample rather than as an indicator of any specific stress. Further, if data on population abundance is also collected, then the change in population abundance due to higher mortality in the stressed population may be detected. Collecting data on both population

abundance and FA would therefore allow FA to serve as a sub-lethal bioindicator to allow (putatively) FA to detect problems before differences in mortality occur.

As mentioned, LC_{50} tests are determined in the lab using test species and do not necessarily reflect what occurs in nature. The heterogeneous mix of stresses which exists in nature may make predictions based on LC_{50} tests difficult (Clarke 1993a). Similarly, BCFs are determined under standardized conditions. The theory underlying the use of QSARs for biomonitoring assume that chemicals are under equilibrium conditions between the sediment and the aqueous phase of the sediment, that lipid content of an animal is the primary determinant of bioaccumulation, and that bioaccumulation is simply driven by thermodynamic considerations partitioning the chemical between lipid and water phases of the animal (Barron 1995). The theory does not take into account metabolic degradation of the chemical (Barron 1995). This is likely a poor assumption given that some metabolic pathways are specifically designed to detoxify chemicals. Further, factors such as pH, sediment type, position in the water column may affect bioavailability (Barron 1995). Uptake may be affected by food intake (type and rate), rate of organism growth, pH, temperature, and interactions with other chemicals (Sijm 1991). Environmental effects (e.g., temperature) may also affect biochemistry and physiology of organisms, which may in turn affect metabolism of chemicals and eventual bioaccumulation (Barron 1995). Not surprisingly, BCF values obtained for a test species in the laboratory can not be used to predict BCFs for other species or the same species in the field (Sijm 1991).

Although AChE activity varies due to ecological factors (Peakall 1992), AChE inhibition appears to be fairly robust to most ecological factors, at least in birds and mammals (Mineau, pers com.). In blood (the non-destructive method of testing for inhibition), AChE activity (and inhibition) is influenced by factors such

as age, season, diurnal effects (Peakall 1992), mercury contamination, severe infection, anemia, malnutrition, liver disease (Mayer et al. 1992), type of pesticide, and species (Melancon 1995).

Unlike AChE inhibition in the brains of some vertebrates, SFG is affected by many different conditions including body size, reproductive status and genetic characteristics (Smaal and Widdow 1994); these factors may decrease SFG but do not necessarily imply reduced quality. Thus, it is important to obtain base line data and to know the natural history of organisms (Mayer et al. 1992, Smaal and Widdows 1994). This would need to be done for each new species examined.

9) Permanence of response

As mentioned, FA is set during development. This has both advantages and disadvantages. The advantage is that FA would not be affected by factors such as environmental fluctuations after development (e.g., due to season for long lived organisms), effects of handling, or hormonal changes (e.g., due to reproductive status) which may create error in bioindicators (Mayer et al. 1992). Further, increases in stress may cause exhaustion. Other bioindicators may reduce activity due to energetic costs of responses and actually experience a reduction in response under some levels of stress (Mayer et al. 1992). This may not be a problem with FA; FA arises due to reduction in response (i.e., reduced developmental stability has less energy required).

The disadvantage is that FA can only measure stress during development and can not measure post-developmental stress. In fact, there is the implicit assumption that quality derived during development is related to quality after development. This assumption appears to be supported by relations observed between FA and post-developmental quality (e.g., longevity and wet mass,

Chapter 5). Further, robust predictive response curves may also be difficult to develop because the timing of stresses may be important (e.g., Parsons 1990, Clarke 1995). While researchers have argued that stress during critical periods of development should increase FA (Parsons 1990, Clarke 1995), it is unknown whether stress which occurs before trait development would also increase FA. We should expect that if FA relates to post-developmental quality, as the evidence suggests, then a stress that occurred prior to trait development but reduced quality should still result in inflated asymmetry. Further, if a stress occurred during trait development, but quality was not irreversibly reduced, then the organism may have the ability to correct the asymmetry later during development (i.e., processes associated with developmental stability). Despite these possibilities, we simply do not know how timing of stresses should influence a FA-quality response curve.

Due to the fact that FA only measures stress during development, FA would probably serve best as a measure of chronic stress rather than as a measure of acute stress (relative to the life span of the organism). In contrast, for SFG, the response is short term, and may be used to measure the amount of energy available at a given point in time (Smaal and Widdows 1994). For AChE inhibition in brains, inhibition can be time dependent and recovery can begin a few hours after exposure ceases, or inhibition can be measurable weeks after exposure ceases (Peakall 1992). Again, it is important to obtain base line data for each species.

3. Fluctuating Asymmetry in Perspective: A Summary of Limitations

Arguably, the purpose of biomonitoring is to assess whether a population is at risk. Specifically, we are interested in determining whether environmental conditions (stresses) has already or may eventually reduce population quality and cause populations to decline. To this effect, environmental monitoring can be viewed at 3 levels. We may wish 1) to determine whether stresses are present, and to identify these stresses; 2) to determine whether organismal quality has changed; and 3) at the highest level, to quantitatively predict the degree to which quality has been compromised. Of course, it is important to understand the limitations of any method we use to monitor the environment.

FA has the potential to play a role at all three of these levels. FA has been found to increase in the presence of many stresses. Therefore, a change in FA may indicate that a population is being subjected to stress, although FA would not be able to indicate the exact source of that stress. At this level, we could not interpret changes in FA in terms of changes in quality with any certainty without extensive a priori testing. Nor could we be certain that changes in FA would occur before changes in other life history parameters (e.g., Premchand et al. 1998). Also, there could be many reasons that FA relations are not found even when stress has had an effect on quality of individuals or populations. Thus, we could not infer anything from an absence of changes in FA. The use of FA without extensive a priori testing may be sufficient to indicate situations that warrant further attention, but would not allow us to state with certainty whether harm to individuals or populations had occurred or would occur in the future.

Extensive a priori testing would be necessary for FA to play a role in predicting changes in quality and identifying populations at risk, and potentially

characterizing the degree of that risk. First, for each biological system, we would require a good knowledge of the natural history of the study organism. Relations between FA and quality would have to be assayed under a range of environmental conditions and different levels of stress, to test the robustness of FA's predictive power. The reliability of FA would have to be validated for each new system, given the heterogeneity between biological systems. The utility of FA may range from stating with confidence that quality has been affected, if changes in FA consistently relates to a change in quality, to actually be able to predict quantitatively the degree of change in quality, if the FA-quality response curve is sufficiently robust. However, researchers should note that the hypothesized potential of FA might not ever be achieved, despite all effort expended to test the validity of FA. FA may simply be an unreliable measure of quality in some biological systems.

Given the limitations associated with FA, the effort required for validation, and the uncertainty in terms of interpretations, it might be tempting for researchers to simply dismiss FA as a bioindicator. However, as illustrated above, all biomonitoring techniques have their limitations, many of which are common with those associated with FA. For example, a common limitation of bioindicators is that it is difficult to actually predict effect or changes in quality. One might be able to infer that some change in quality has occurred. But predicting the magnitude of change is more difficult. Further, the absence of a response would not necessarily indicate that quality has not been affected. Also, typically it is unclear how synergistic interactions between multiple stresses affect the relation between the bioindicator response and quality (especially for specific bioindicators), despite the obvious importance of multiple stresses. Finally, heterogeneity between species appears to be the rule. For all bioindicators,

extensive validation and collection of background and natural history data is likely required for each species of interest.

The purpose of the above discussion on currently used biomonitoring techniques is not to suggest that current methods are not useful. For example, traditional techniques such as toxicity testing and BCF are useful to get a base line data on the potential of chemicals to cause harm, either due to toxicity or accumulation for a given species, although we can not necessarily extend predictions to effects under field conditions. Specific bioindicators are useful to indicate exposure, but robust quantitative prediction of effect for a given toxic substance is difficult given the presence of multiple stresses, each of which may act synergistically to affect quality. Commonly used general indicators are useful to detect the presence of unknown stresses, although actual quantitative estimation of quality is more difficult. Further, identifying the causal stress requires other techniques besides stress-general bioindicators.

Clearly, all of these biomonitoring methods provide information that we want to know. They are useful despite their limitations, provided inferences are drawn appropriately. Thus, despite the fact that FA is limited, it may still serve a useful purpose so long as it is interpreted correctly. Thus the question is what inferences can be drawn using FA, given its limitations, and what role can FA play in biomonitoring?

4. Conclusions: The Utility of Fluctuating Asymmetry

If FA were to be used, it would probably be as a general indicator of quality (potentially indexing the synergistic effects of all stresses in the environment). FA would not be useful to isolate causes of harm (also see Clarke 1995). Further, given that FA forms during development and would be increased by stress after

development, FA probably would be most useful to measure chronic rather than acute stress (relative to an organism's life span).

I note that FA falls short of the apparent expectations advocated in the literature. Although FA is one the simplest bioindicators to measure, and is relatively simple to analyze, precise interpretation of FA patterns is more difficult. Without extensive validation, we would not be able to assume that FA would reliably reflect quality of individuals or populations, or the presence of stress. To determine the robustness of FA-quality relations, validation would be required for each new biological system, under a range of environmental conditions that the organisms would reasonably be expected to encounter. A good understanding of the natural history of each species would also be necessary. Finally, given that differential mortality should reduce FA differences, we likely would not use FA in isolation to determine whether stress is present. We would need to measure traditional life history estimates or population size estimates as well.

Given the extensive validation required, and the uncertainty that FA will relate reliably to quality in a given biological system, researchers may decide that it is not worth the effort to obtain this base line data. However, one should note that 1) validation would be required for other bioindicators as well. 2) The generality and reliability of all bioindicators as predictors of quality appears to be uncertain. 3) Significant FA relations have been found in a wide range of organisms and traits. 4) If validation suggests that FA can reliably be used either to indicate a change in quality has occurred or to quantitatively predict the change in quality, then FA would be likely the simplest, most cost-effective, non-intrusive measure of quality available.

Of course, researchers could measure life history traits directly, rather than relying on indirect measures of quality. However, FA may offer several advantages over life history traits: 1) changes in FA may be more sensitive to

stress than other life history traits (e.g., Clarke and McKenzie 1992). 2) Some life history traits be more difficult or time consuming to measure (e.g., reproductive success). 3) Although some life history traits are easy to measure (e.g., body size), FA may provide additional information (e.g., both body size and FA independently related to mating success in damselflies, Harvey and Walsh 1993).

To determine the validity of FA, a pilot study could be conducted to determine if there are reasons to expect FA to relate to either stress or quality. If such relations exist, researchers could proceed to do a more thorough test of the robustness of FA's predictive power. If researchers decide that it is too much effort to validate FA, FA may still play a role in biomonitoring (albeit a reduced one). Arguably, regardless of whether FA is validated, gross traits should be assessed (e.g., body size or condition, Mayer *et al.* 1992), as such traits are often determinants of fitness. Any time traits such as body condition or body size is assessed, FA could simultaneously be assessed with little additional work (FA is nothing more than morphometric measurements). Without validation, we would not be able to definitively state that FA reflected quality or stress, or that FA was more sensitive than other life history traits. However, we know that FA is often related to stress and quality. Therefore, even without full knowledge, we could use a change in FA to reasonably suggest that a population may be at risk and that additional detailed and costly examination should be conducted. As long as limitations concerning interpretations of FA patterns is considered (e.g., the absence of FA differences does not necessarily indicate the absence of effects), FA may still provide a useful tool even without validation. The actual decision to use FA would need to be made by the individual researcher and would likely depend upon the question of interest, the importance of the biological system under study, and which other bioindicators were available and their limitations

and benefits. Likely, if FA were to be used, it should be treated as one weapon in an arsenal of bioindicators, especially given FA's non-specificity to stresses and FA's inability to isolate causes of harm.

Fig 7-1. Idealized FA-quality response curve, where FA is the average FA in a sample, and quality is the average quality in that sample. If this curve can be generated, and is robust to varying environmental conditions, then FA could be used to predict quality. In this idealized figure, FA increases measurably before quality begins to sharply decrease. Thus, we could theoretically identify FA values for which we would not need to perform remedial action (the "Safe" region), where populations may be "At Risk", and where populations are likely being affected. In reality, little work has been done to generate such predictive curves; it is unknown how common or robust such curves could be. It is also unknown how often FA would actually be more sensitive than traditional life history traits.

Fluctuating Asymmetry

Literature cited

- Adams, S.M., Ham, K.D., and LeHew, R.F. 1998. A framework for evaluating organism responses to multiple stressors: mechanisms of effect and importance of modifying ecological factors. pp13-22. In Cech, J.J. Jr., Wilson B.W., and Crosby D.G. (eds). Multiple stresses in Ecosystems. Lewis Publishers, Baco Raton.
- Alados, C.L., Escos, J. and Emlen, J.M. 1993. Developmental instability as an indicator of environmental stress in the pacific hake (Merluccius productus). US National Marine Fisheries Service: Fishery Bulletin 91:587-593.
- Anderson, D.W. 1998. Evaluating and impact of multiple stressors on ecosystems: three classic case histories. pp3-8. In Cech, J.J. Jr., Wilson B.W., and Crosby D.G. (eds). Multiple stresses in Ecosystems. Lewis Publishers, Baco Raton.
- Arnqvist, G., and Wooster, D. 1995. Meta-analysis: synthesizing research findings in ecology and evolution. Trends in Ecology and Evolution, 10: 236-240
- Askew, R.R. 1988. The dragonflies of Europe. Harley Books. Colchester. UK.
- Baldwin, I.G., and Kramer K.G.M. Biological early warning systems (BEWS). pp 1-28. in Kramer, K.J.M. (ed). Biomonitoring of coastal waters and estuaries. CRC Press. Baco Raton.
- Banks M.J. and Thompson, D.J. 1985. Lifetime mating success in the damselfly Coenagrion puella. Animal Behavior 33:1175-1183.
- Barron, M.G. 1995. Bioaccumulation and bioconcentration in aquatic organisms. pp 652-666. in Hoffman, D.J., Rattner, B.A., Burton, G.A. Jr., Cairns, J. Jr. (eds.) Handbook of Ecotoxicology. Lewis Publishers, Boca Raton, USA.

- Beacham, T. D. and Witler, R. E. 1987. Developmental stability and heterozygosity in chum (Oncorhynchus keta) and pink (Oncorhynchus gorbuscha) salmon. *Canadian Journal of Zoology*, 65: 1823-1826
- Beacham, T. D., 1990. A genetic analysis of meristic and morphometric variation in chum salmon (Oncorhynchus keta) at three different temperatures. *Canadian Journal of Zoology*, 68: 225-229
- Begon, M., Harper, J.L., and Townsend, C.R. 1990. *Ecology: Individuals, populations, and communities* (2nd ed). Blackwell scientific publications. Boston.
- Blanco, G., Sanchez, J. A., Vazquez, E., Garcia E., and Rubio, J. 1990. Superior developmental stability of heterozygotes at enzyme loci in Salmo salar L. *Aquaculture*, 84: 199-209
- Bonn, A., Gasse, M., Rolff, J. and Martens, A. 1996. Increased fluctuating asymmetry (FA) in the damselfly Coenagrion puella associated with ectoparasitism by water mites: implications for FA theory. *Oecologia* 108: 596-598.
- Breckenridge, R.P., Olson, G.L. 1995. Identification and use of biomonitoring data. pp198-219. in Hoffman, D.J., Rattner, B.A., Burton, G.A. Jr., Cairns, J. Jr. (eds.) *Handbook of Ecotoxicology*. Lewis Publishers, Boca Raton, USA.
- Brett, J.R. 1958. Implications and assessment of environmental stress. in Larkin, P.A., (ed). *Investigations of fish power problems*. University of British Columbia Press. Vancouver.
- Bunn, S.E. 1995. Biological monitoring of water quality in Australia: workshop summary and future directions. *Australian Journal of Ecology* 20: 220-227.

NOTE TO USERS

Page(s) not included in the original manuscript are unavailable from the author or university. The manuscript was microfilmed as received.

This reproduction is the best copy available.

UMI

- Dufour, K.J. and Weatherhead, P.J. 1996. Estimation of organism-wide asymmetry in red-winged blackbirds and its relation to studies of mate selection. *Proceedings of the Royal Society of London Series B* 263:769-775.
- Dufour, K.J. and Weatherhead, P.J. 1998. Bilateral symmetry and social dominance in captive male red-winged blackbirds. *Behavioral Ecology and Sociobiology* 42: 71-76.
- Eggert A.K. and Sakaluk, S.K. 1994. Fluctuating asymmetry and variation in the size of courtship food gifts in decorated crickets. *American Naturalist* 144:708-716.
- Emlen, J.M., Freeman, D.C., and Graham, J.H. 1993. Nonlinear growth dynamics and the origin of fluctuating asymmetry. *Genetica* 89: 77-96.
- Evans, M. R., Martins, T. L. F. and Haley, M. P. 1995. Inter- and intra-sexual patterns of fluctuating asymmetry in red-billed streamertail: should symmetry always increase with ornament size. *Behavioural Ecology and Sociobiology* 37: 15-23
- Fincke, O.M. 1988. Sources of variation in lifetime reproductive success in a nonterritorial damselfly (Odonata: Coenagrionidae). In T.H. Clutton-Brock (ed.). *Reproductive success: Studies of individual variation in contrasting breeding systems*. Univ Chicago Press pp24-43.
- Forbes, M.R.L. 1991a. Parasites and reproductive success of male hosts: competing hypotheses and tests with Enallagma damselflies and parasitic mites. PhD thesis. Univ Toronto, Canada 196pp.
- Forbes, M.R.L. 1991b. Ectoparasites and mating success of male Enallagma ebrium damselflies (Odonata:Coenagrionidae). *Oikos* 60:336-342.

- Forbes M.R.L. and Baker, R.L. 1990. Susceptibility to parasitism: experiments with the damselfly *Enallagma ebrium* (Odonata: Coenagrionidae) and larval water mites, *Arrenurus* spp. (Acari: Arrenuridae). *Oikos* 58:61-66.
- Forbes M.R.L. and Baker, R.L. 1991. Condition and fecundity of the damselfly, *Enallagma ebrium* (Hagen): the importance of ectoparasites. *Oecologia* 86:335-341.
- Forbes, M.R., Leung, B., and Schalk, G. 1997. Fluctuating asymmetry in *Coenagrion resolutum* (Hagen) in relation to age and male pairing success (Zygoptera:Coenagrionidae). *Odonatologica* 26: 9-16.
- Fowler, K. and Whitlock, M.C. 1994. Fluctuating asymmetry does not increase with moderate inbreeding in *Drosophila melanogaster*. *Heredity* 73:373-376.
- Gangestad, S.W. and Thornhill, R. 1998. The analysis of fluctuating asymmetry redux - the robustness of parametric statistics. *Animal Behaviour* 55:497-501.
- Gest, T. R., Siegel, M. T. and Anistranski, J. 1987. The long bones of neonatal rats stressed by cold, heat, and noise exhibit increased fluctuating asymmetry. *Growth* 50: 385-389.
- Glass, G. V., McGaw, B. and Smith, M. L. 1981. Meta-analysis in social research. Sage Publications, Beverly Hills, California
- Graham, J.H., Freeman D.C. and Emlen, L.M. 1993. Antisymmetry, directional asymmetry and dynamic morphogenesis. *Genetica* 89: 121-137.
- Gummer, D.L. and Bringham, R.M. 1995. Does fluctuating asymmetry reflect the importance of traits in little brown bats (*Myotis lucifugus*). *Canadian Journal of Zoology* 73: 990-992.

- Harvey, I. F. and Walsh, K. J. 1993. Fluctuating asymmetry and lifetime mating success are correlated in males of the damselfly Coenagrion puella. *Ecological Entomology* 18: 198-202.
- Heath, A.G. 1998. Physiology and Ecological Health. pp. 59-90. In Cech, J.J. Jr., Wilson B.W., and Crosby D.G. (eds). *Multiple stresses in Ecosystems*. Lewis Publishers, Baco Raton.
- Hedges, L. V., 1994a. Statistical considerations. Pages 30-37 in Cooper, H. and L. V. Hedges (ed.). *Handbook of Research Synthesis*. Russel Sage Foundation, New York.
- Hedges, L. V. 1994b. Fixed effects models. Pages 286-299 in Cooper, H. and L.V. Hedges (ed.). *Handbook of Research Synthesis*. Russel Sage Foundation, New York.
- Hedges, L. V. and Olkin, I. 1985. *Statistical methods for meta-analysis*. Academic Press.Orlando, Florida
- Hill, G. E., 1995. Ornamental traits as indicators of environmental health. *Bioscience* 45: 25-31
- Houle, D. 1997. Comment on "A meta-analysis of the heritability of developmental stability" by Moller and Thornhill. *Journal of Evolutionary Biolology* 10: 17-20.
- Houle, D. 1998. High enthusiasm and low r-squared. *Evolution* 52: 1872-1876.
- Huggett, R.J., Kimerle, R.A., Mehrle, P.M. Jr., and Bergman H.L. 1992. Biomarkers: biochemical, physiological, and histological markers of anthropogenic stress. Lewis publishers. Boca Raton.
- Kieser, J.A. and Groeneveld, H.T. 1991. Fluctuating odontometric asymmetry, morphological variability, and genetic monomorphism in the cheetah Acinonyx jubatus. *Evolution* 45:1175-1183.

- Koehn, R.K. and Bayne, B.L. 1989. Towards a physiological and genetical understanding of the energetics of the stress response. *Biological Journal of the Linnean Society* 37: 157-171.
- Kovács, M. 1992. *Biological indicators in environmental protection*. Ellis Horwood Ltd. Chinchester, UK
- Kramer, K.J.M. 1994. *Biomonitoring of coastal waters and estuaries*. CRC Press. Baco Raton.
- Leary, R.F. and Allendorf, F.W. 1989. Fluctuating asymmetry as an indicator of stress: implications for conservation biology. *Trends in Ecology and Evolution* 4: 214-216.
- Leary, R. F., Allendorf, F. W., and Knudsen, K. L. 1984. Superior developmental stability of heterozygotes at enzyme loci in salmonid fishes. *American Naturalist* 124: 540-551
- Leary, R. F., Allendorf, F. W. and Knudsen, K. L. 1985. Developmental instability as an indicator of reduced genetic variation in hatchery trout. *Transactions of the American Fisheries Society* 114: 230-235.
- Leary, R. F., Allendorf, F. W. and Knudsen, K. L. 1991. Effects of rearing density on meristics and developmental stability of rainbow trout. *Copeia* 1991: 44-49.
- Leung, B. 1998. Controlling for allometry in studies of fluctuating asymmetry and quality within samples. *Proceedings of the Royal Society of London Series B*. 265:1623-1629.
- Leung, B. and 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. and Forbes, M.R. 1997a. Modeling fluctuating asymmetry in relation to stress and fitness. *Oikos* 78: 397-405.

- Leung, B. and Forbes, M.R. 1997b. Fluctuating asymmetry in relation to indices of quality and fitness in the damselfly, Enallagma ebrium (Hagen). *Oecologia* 110: 472-477.
- Liggett, A. C., Harvey, I. F. and Manning, J. T. 1993. Fluctuating asymmetry in Scatophaga stercoraria L.: successful males are more symmetrical. *Animal Behaviour* 45: 1041-1043.
- Lindeman, R.H., Merenda, P.F. and Gold, R.Z. 1980. Introduction to bivariate and multivariate analysis. Scott Foreman and Co. USA.
- Malyon, C. and Healy, S. 1994. Fluctuating asymmetry in antlers of fallow deer, Dama dama, indicates dominance. *Animal Behaviour* 48: 248-250.
- Manning, J. T. and Chamberlain, A. T. 1994. Fluctuating asymmetry in gorilla canines: a sensitive indicator of environmental stress. *Proceedings of the Royal Society of London Series B*, 255: 189-193.
- Manning, J. T. and Hartley, M. A. 1991. Symmetry and ornamentation are correlated in the peacock's train, *Animal Behaviour*, 42: 1020-1021.
- Manning, J.T. and Ockenden, L. 1994. Fluctuating asymmetry in racehorses. *Nature* 370: 185-186.
- Markow, T. A. and Gottesman, I. I. 1989. Dermatoglyphic fluctuating asymmetry in twins and singletons. *Hereditas*, 110: 211-215.
- Markow, T. A. and Martin, J. F. 1993. Inbreeding and developmental stability in a small human population. *Annals of Human Biology*, 20: 389-394.
- Markow, T. A. and Ricker, J. P. 1992. Male size, developmental stability, and mating success in natural populations of three drosophila species. *Heredity*, 69: 122-127.
- Mayer, F.L., Versteeg, D.J., Mckee, M.J., Folmar, L.C., Graney, R.L., McCume, D.C., Rattner, B.A. 1992. Physiological and nonspecific biomarkers. pp 5-86. in Huggett, R.J., Kimerle, R.A., Mehrle, P.M. Jr., and Bergman H.L.

- (eds). Biomarkers: biochemical, physiological, and histological markers of anthropogenic stress. Lewis publishers. Boca Raton.
- McKim, J.M. and Schmieder P.K. 1991. Bioaccumulation: does it reflect toxicity? pp161-188. in Nagel, R and Loskill R. (eds) Bioaccumulation in aquatic systems: Contributions to the assessment. VCH. Weinheim.
- McLachlan, A. and Cant, M. 1995. Small males are more symmetrical: mating success in the midge Chironomus plumosus L. (Diptera: Chironomidae). *Animal Behaviour* 50: 841-846.
- Melancon, M.J. 1995. Bioindicators used in aquatic and terrestrial monitoring. p220-240. in Hoffman, D.J., Rattner, B.A., Burton, G.A. Jr., Cairns, J. Jr. (eds.) *Handbook of Ecotoxicology*. Lewis Publishers, Boca Raton.
- Mineau, P. 1998. Biomarkers: are there linkages to ecological effects. p91-107. in Cech, J.J. Jr., Wilson B.W., and Crosby D.G. (eds). *Multiple stresses in Ecosystems*. Lewis Publishers, Baco Raton.
- Mitton, J.B. 1993. Enzyme heterozygosity, metabolism, and developmental stability. *Genetica* 89: 47-65.
- Møller, A. P., 1990. Fluctuating asymmetry in male sexual ornaments may reliably reveal male quality. *Animal behaviour* 40: 1185-1187
- Møller, A. P., 1992a. Patterns of fluctuating asymmetry in weapons: evidence for reliable signalling of quality in beetle horns and bird spurs. *Proceedings of the Royal Society of London Series B* 248: 199-206.
- Møller, A.P. 1992b. Female swallow preference for symmetrical male sexual ornaments. *Nature* 357:238-240.
- Møller, A. P., 1993a. Developmental stability, sexual selection, and speciation. *Journal of Evolution Biology* 6: 493-509.

- Møller, A. P., 1993b. Morphology and sexual selection in the barn swallow Hirundo rustica in Chernobyl, Ukraine. *Proceedings of the Royal Society of London Series B* 252: 51-57.
- Møller, A. P., 1993c. Patterns of fluctuating asymmetry in sexual ornaments predict female choice. *Journal of Evolutionary Biology* 6: 481-491.
- Møller, A. P., 1993d. Female preference for apparently symmetrical male sexual ornaments in the barn swallow Hirundo rustica. *Behavioural Ecology and Sociobiology* 32: 371-376.
- Møller, A. P., 1995a. Patterns of fluctuating asymmetry in sexual ornaments of birds from marginal and central populations. *American Naturalist* 145: 316-327.
- Møller, A. P., 1995b. Developmental stability and ideal despotic distribution of blackbirds in a patchy environment. *Oikos* 72: 228-234.
- Møller, A. P. and Höglund, J. 1990. Patterns of fluctuating asymmetry in avian feather ornaments: implications for models of sexual selection. *Proceedings of the Royal Society of London Series B*, 245: 1-6.
- Møller, A. P., Magnhagen, C., Ulfstrand, A. and Ulfstrand, S. 1995. Phenotypic quality and molt in the barn swallow, Hirundo rustica. *Behavioural Ecology* 6: 242-249.
- Møller, A.P. and Pomiankowski, A. 1993. Fluctuating asymmetry and sexual selection. *Genetica* 89:267-279.
- Møller, A.P. and Swaddle, J.P. 1997. *Asymmetry, developmental stability, and evolution*. Oxford University Press. Oxford.
- Mooney, M. P., Siegel, M. I. and Gest, T. R. 1985. Prenatal stress and increased fluctuating asymmetry in the parietal bones of neonatal rats. *American Journal of Physical Anthropology* 68: 131-134.

- Naugler, C. T. and Leech, S. M. 1994. Fluctuating asymmetry and survival ability in the forest tent caterpillar moth Malacosoma disstria: implications for pest management. *Entomologia Experimentalis et Applicata*, 70: 295-298.
- Neter, J., Wasserman, W., and Kutner, M.H. 1983. Applied linear regression models. Richard D. Irwin Inc, Illinois.
- Niklas, K.J. 1994. Plant allometry: the scaling of form and process. University of Chicago Press, Chicago.
- Nilsson, J., 1994. Energetic stress and the degree of fluctuating asymmetry: implications for a long-lasting, honest signal. *Evolutionary Ecology* 8: 248-255.
- Novak, J. M., Rhodes, O. E., Smith., M. H. Jr. and Chessser, R. K. 1993. Morphological asymmetry in mammals: genetics and homeostasis reconsidered. *Acta Theriologica* 38: 7-18.
- Oakes, E. J. and Barnard, P. 1994. Fluctuating asymmetry and mate choice in paradise whydahs, Vidua paradiaea: an experimental manipulation. *Animal Behaviour* 48: 937-943.
- Owen, R. D. and McBee, K. 1990. Analysis of asymmetry and morphometric variation in natural populations of chromosome-damaged mice. *Texas Journal of Science* 42: 319-332.
- Packer, C. and Pusey, A. E. 1993. Should a lion change its spots? *Nature* 362: 595.
- Palmer, A. R. 1994. Fluctuating asymmetry analysis: a primer. Pages 335-364 in Markow, T.A. (ed.). *Developmental Instability: Its Origins and Evolutionary Implications*. Kluwer, Dordrecht.
- Palmer, A.R. 1996. Waltzing with asymmetry. *BioScience* 46: 518-532.

- Palmer, A. R. and Strobeck, C. 1986. Fluctuating asymmetry: Measurement, analysis and patterns. *Annual Review of Ecology and Systematics* 17: 391-421.
- Palmer, A. R. and Strobeck, C. 1992. Fluctuating asymmetry: implications of non-normality. *Acta Zoologica Fennica*, 191: 57-72.
- Pankakoski, E. 1985. Epigenetic asymmetry as an ecological indicator in muskrats. *Journal of Mammalogy*, 66: 52-57.
- Parsons, P.A. 1990. Fluctuating asymmetry: an epigenetic measure of stress. *Biological Reviews (Cambridge)* 65:131-145.
- Patterson, B. D. and Patton, J. L. 1990. Fluctuating asymmetry and allozymic heterozygosity among natural populations of pocket gophers (Thomomys bottae). *Biological Journal of the Linnean Society* 40: 21-36
- Peakall, D. 1992 *Animal biomarkers as pollution indicators*. Chapman and Hall, London.
- Polak, M. 1993. Parasites increase fluctuating asymmetry of male Drosophila nigrospiracula: Implications for sexual selection. *Genetica* 89: 255-265.
- Polak, M. and Trivers, R. 1994. The science of symmetry in biology. *Trends in Ecology and Evolution* 9:122-124.
- Premchand, A., Mawri, F., Gladstone, S., Freeman, D.C. 1998. Is fluctuating asymmetry a reliable biomonitor of stress? A test using life history parameters in soybean. *International Journal of Plant Sciences* 159: 559-565.
- Radesäter, T. and Halldórsdóttir, H. 1993. Fluctuating asymmetry and forcep size in earwigs, Forficula auricularia. *Animal Behaviour* 45: 626-628.
- Raudenbush, S. W. 1994. Random effects models. Pages 301-320 in Cooper, H. and Hedges, L.V. (ed.). *Handbook of Research Synthesis*. Russel Sage Foundation, New York.

- Richardson, J.M.L. and Baker, R.L. 1997. Effect of body size and feeding on fecundity in the damselfly Ischnura verticalis (Odonata: Coenagrionidae). *Oikos* 79: 477-483.
- Rosenthal, R. 1991. Meta-analytic procedures for social research (Rev. ed). Sage, Beverly Hills, California.
- Rowe, L., Repasky, R.R. and Palmer, A.R. 1997. Size-dependent asymmetry: fluctuating asymmetry versus antisymmetry and its relevance to condition dependent signaling. *Evolution* 51: 1401-1408.
- Ryan, M. J., Warkentin, K. M., McClelland, B. E. and Wilczynski, W. 1995. Fluctuating asymmetry and advertisement call variation in the cricket frog, Acris crepitans. *Behavioural Ecology* 6: 124-131
- Shadish, W. R. and Haddock, C. K. 1994. Statistically describing and combining studies. Pages 262-281 in Cooper, H. and L. V. Hedges (ed.). *Handbook of Research Synthesis*. Russel Sage Foundation, New York.
- Siegel, M. I. and Mooney, M. P. 1987. Perinatal stress and increased fluctuating asymmetry of dental calcium in the laboratory rat. *American Journal of Physical Anthropology* 73: 267-270.
- Sijm, D.T.H.M. 1991. Extrapolating the laboratory results to environmental conditions. pp151-160. in Nagel, R and Loskill R. (eds) *Bioaccumulation in aquatic systems: Contributions to the assessment*. VCH. Weinheim.
- Smaal, A.C. and Widdows, J. 1994. The Scope For Growth of bivalves as an integrated response parameter in biological monitoring. pp247-262. in Kramer, K.J.M. (ed). *Biomonitoring of coastal waters and estuaries*. CRC Press. Boca Raton.

- Smith, B.P. 1988. Host-parasite interaction and impact of larval water mites on insects. *Annual Review of Entomology* 33:487-507.
- Snedecor G. W. and Cochran, W. G. 1980. *Statistical methods*. Iowa State University Press. Ames, Iowa.
- Solberg, E. J. and Sether, B. 1993. Fluctuating asymmetry in the antlers of moose (Alces alces): does it signal male quality? *Proceedings of the Royal Society of London Series B*, 254: 251-255.
- Sommer, C. 1996. Ecotoxicology and developmental stability as an in situ monitor of adaptation. *Ambio* 25: 374-376.
- Stegeman, J.J., Brouwer, M., Di Giulio, R.T., Förlin, L., Fowler, B.A., Sanders, B.M., and Wan Weld, P.A. 1992. Enzyme and Protein Synthesis as Indicators of Contaminant Exposure and Effect. pp 235-336. in Huggett, R.J., Kimerle, R.A., Mehrle, P.M. Jr., and Bergman H.L. (eds). *Biomarkers: biochemical, physiological, and histological markers of anthropogenic stress*. Lewis publishers. Boca Raton.
- Steinkamp, M. J. and Maehr, M. L. 1983. Affect, ability, and science achievement: a quantitative synthesis of correlational research. *Review of Educational Research* 53: 369-396.
- Strong, L. and James, S. 1992. Some effects of rearing the yellow dung fly Scatophaga stercoraria in cattle dung containing ivermectin. *Entomologia Experimentalis et Applicata* 63: 39-45.
- Swaddle, J. P. and Cuthill, I. C. 1994a. Female zebra finches prefer males with symmetric chest plumage. *Proceedings of the Royal Society of London Series B* 258: 267-271.
- Swaddle, J. P. and Cuthill, I. C. 1994b. Preference for symmetric males by zebra finches. *Nature* 367: 165-166.

- Swaddle, J. P. and M. S. Witter, 1994. Food, feathers, and fluctuating asymmetries. *Proceedings of the Royal Society of London Series B* 255: 147-152.
- Swaddle, J.P., Witter, M.S., and Cuthill, I.C. 1994. The analysis of fluctuating asymmetry. *Animal Behaviour* 48: 986-989.
- Thornhill, R. and Sauer, P. 1992. Genetic sire effects on the fighting ability of sons and daughters and mating success of sons in a scorpionfly. *Animal Behaviour* 43: 255-264.
- Thornhill, R., 1992a. Fluctuating asymmetry and the mating system of the Japanese scorpionfly, Panorpa japonica. *Animal Behaviour* 44: 867-879.
- Thornhill, R., 1992b. Female preference for the pheromone of males with low fluctuating asymmetry in the Japanese scorpionfly (Panorpa japonica). *Behavioural Ecology* 3: 277-283.
- Thornhill, R., Gangestad, S.W., and Comer, R. 1995. Human female orgasm and mate fluctuating asymmetry. *Animal Behavior* 50:1601-1615.
- Ueno, H., 1994. Fluctuating asymmetry in relation to two fitness components, adult longevity and male mating success in the ladybird beetle, Harmonia axyridis. *Ecological Entomology* 19: 87-88.
- Underwood, A.J. 1989. The analysis of stress in natural populations. *Biological Journal of the Linnean Society* 37: 51-78.
- Van Valen, L. 1962. A study of fluctuating asymmetry. *Evolution* 16: 125-142.
- Watson, P. J. and Thornhill, R. 1994. Fluctuating asymmetry and sexual selection. *Trends in Ecology and Evolution* 9: 21-25.
- Whitlock, M. 1993. Lack of correlation between heterozygosity and fitness in forked fungus beetles. *Heredity* 70:574-581.

- Wiener, J. G. and Rago, P. J. 1989. A test of fluctuating asymmetry in Bluegills (Lepomis macrochirus Rafinesque) as a measure of pH-related stress. *Environmental Pollution* 44: 27-36.
- Witter, M. S. and Swaddle, J. P. 1994. Fluctuating asymmetries, competition and dominance. *Proceedings of the Royal Society of London Series B* 256: 299-303.
- Wooten, M. C. and Smith, M. H. 1986. Fluctuating asymmetry and genetic variability in a natural population of Mus musculus. *Journal of Mammalogy* 67: 725-732.
- Yezerinac, S.M., Loughheed, S.C. and Handford P. 1992. Measurement error and morphometric studies: statistical power and observer experience. *Systematic Biology* 41:471-482.
- Zakharov, V. M., Pankakoski, E., Sheftel, B. I., Peltonen, A. and Hanski, I. 1991. Developmental stability and population dynamics in the common shrew, Sorex araneus. *American Naturalist* 138: 797-810.
- Zar, J.H. 1984. *Biostatistical Analysis* (2nd ed). Prentice Hall. Upper Saddle River, NJ.
- Zar, J.H. 1996. *Biostatistical Analysis* (3rd ed). Prentice Hall. Upper Saddle River, NJ.
- Zhivotovsky, L.A. 1988. Some methods of analysis of correlated characters. In: Weir, B.S., Eisen, E.J., Godman, M.M. and Namkoong, G. (eds). *Proc. Second Int. Conf. Quant. Genetics*: 423-432. Sinauer Associates, Sunderland, Massachusetts, USA.
- Zhivotovsky, L.A. 1992. A measure of fluctuating asymmetry for a set of characters. *Acta Zoologica Fennica* 191:73-77.

Appendix A.

Glossary of Terms

Allometry-Relation between size and some other variable (Niklas 1994).

Bioindicator-responses of living organisms that may indicate exposure to stresses, predict future harm, or may cause harm (Melancon 1995). A subset of bioindicators are biochemical, physiological, or morphological responses to stress (termed biomarkers, Melancon 1995).

Character specific buffering-A suite of processes that tend to resist disruption of symmetry during development of specific characters. The expression of some of these processes may be dependent upon quality of an individual or upon factors associated with individual quality (Cm in my model). The expression of other processes may be independent of quality (Ca in my model). Character specific buffering likely differs between traits depending on selective forces on those traits (e.g., directional selection for trait size or stabilizing selection). Thus, by definition, character specific buffering is a component of developmental stability.

Composite Fluctuating Asymmetry (CFA)-A convenient term to refer to analyses based on information on FA distributions combined across traits.

Developmental Noise- A suite of processes that tend to disrupt precise development, such as: a) small, random differences in rates of cell division, cell growth, and cell shape change, b) effects of thermal noise on enzymatic activity, c) small, random differences in rates of physiological processes among cells (*sensu stricto* Palmer 1994)

Developmental stability- A suite of processes that tend to resist or buffer the disruption of precise development such as a)negative feedback systems that regulate enzyme activity (both concentration and catalytic rate) within

and between cells, b) central nervous regulation of non-contiguous structures, or c) hormonal regulation of non-contiguous structures (*sensu stricto* Palmer 1994).

Effect size-A measure of the strength of a relation between two variables (e.g., FA and stress)

Fitness- has been defined as the relative contribution by an individual (relative to other individuals in the present population) to the gene pool of the next generation (Begon *et al.* 1990).

Fluctuating asymmetry (FA)-Subtle deviations from perfect bilateral symmetry.

FA is a function of both developmental noise and developmental stability.

FA is most often measured as the absolute left minus right values of a bilateral character pair (e.g., arm lengths).

Indexing ability-A convenient term used in this model to describe the *relative* strengths of FA-stress and/or FA-quality relations of different traits.

Presumably, the relative strength of these relations indicate the ability of FA to index either stress or quality of individuals.

Isometry- Special case of allometry when some variable is proportional to some body size measurement (assuming that the variable is in the same units as size, Niklas 1994)

Leptokurtic asymmetry distribution (LAD)-Identical to the definition of FA except that variation is leptokurtically distributed rather than normally distributed. As with FA, LAD is also determined by developmental noise and developmental stability.

Leptokurtosis-A distribution with more values around the center and tails of the distribution compared to expected values of a normal distribution (Zar 1984). Leptokurtosis is identified by positive kurtosis values.

Performance trait- A trait in which asymmetry has a negative mechanical consequence on the optimal functioning of a character, and may therefore directly affect quality or fitness.

Platykurtosis-A distribution with more values between the center and tails of the distribution compared to expected values of a normal distribution (Zar 1984). Platykurtosis is identified by negative kurtosis values.

Quality-At the individual level, I define quality as the probability of high fitness. At the population level, I define quality as a characteristic of that population that influences the probability of population decline or extinction. A low quality population would be comprised of comparatively low quality individuals.

Sexually-selected trait-A trait which affects intra-specific mate choice or mate competition. These traits often are under directional selection for increased magnitude and may, therefore, have low developmental stability (Møller 1993).

Stress-Any factor which negatively affects an organism such that physiology, morphology, or biochemistry is affected. A biologically relevant stress is a factor that reduces current or future fitness or quality, or results in a change in population demographics (specifically population decline).

Appendix B

Data set showing source, correlate, species or organism, character used, trait types, thermy, statistic used, and sample size of each relation used in the meta-analysis. More detailed information may be obtained from source numbers and corresponding references cited at the end of Appendix B. Correlate refers to the type of stress or the fitness measure examined with respect to FA. For tractability, correlates and characters were grouped into general types (e.g., pollution included toxic wastes, fertilizer effluent, etc., wing/tail included feathers, actual wing length, etc.). We categorized each relation as sexually-selected (S), non-sexually-selected (NS), performance (P), and non-performance traits (NP), and poikilotherms (PO), and homeotherms (HO). The statistic used refers to the type of data extracted from primary studies and transformed into a common currency. MS=means and standard errors (or standard deviations), PC=Pearson's correlation coefficient, LT=Levene's test, PV=p-values, CS=Chi-square values, SC=Spearman's correlation coefficient, and BP=Binomial probabilities.

FA-STRESS RELATIONS:

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
1	Temperature	<u>Oncorhynchus keta</u>	fin	NS	P PO	MS	40
				NS	P PO	MS	40
			pattern	NS	NP PO	MS	40
2	homozygosity	<u>Oncorhynchus keta</u>	gill	NS	NP PO	PC	94
				NS	NP PO	PC	94
			fin	NS	P PO	PC	94
				NS	P PO	PC	94
			gill	NS	NP PO	PC	100
				NS	NP PO	PC	100
2	homozygosity	<u>Oncorhynchus keta</u>	fin	NS	P PO	PC	100
				NS	P PO	PC	100
			gill	NS	NP PO	PC	1003
				NS	NP PO	PC	100
			fin	NS	P PO	PC	100

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
3	homozygosity	<u>Oncorhynchus gorbusha</u>	gill	NS	P	PO	PC
				NS	NP	PO	PC
				NS	NP	PO	PC
			fin	NS	P	PO	PC
				NS	P	PO	PC
				NS	P	PO	PC
	homozygosity	<u>Salmo salar</u>	fin	NS	P	PO	LT
				NS	P	PO	LT
				NS	NP	PO	LT
			gill	NS	P	PO	LT
				NS	P	PO	LT
			gill	NS	NP	PO	LT
				NS	P	PO	LT
			fin	NS	P	PO	LT
			gill	NS	NP	PO	LT
			fin	NS	NP	PO	LT
			fin	NS	NP	PO	LT

Source	Correlate	Species/Organism	Character	Trait type	Thermy	Stat	Sample size
4	pollution	<u>Chrysopa perla</u>	gill	NS	NP	LT	71
				NS	P	MS	86
			wing\tail	NS	P	MS	86
				NS	P	MS	86
				NS	P	MS	86
5	pollution	<u>Palaemon elegans</u>	sensory	NS	NP	PV	88
				NS	NP	PV	80
6	homozygosity	<u>Tisbe holothuriae</u>	leg	NS	P	LT	60
				NS	NP	LT	60
			bristle\spine	NS	NP	LT	60
				NS	NP	LT	60
				NS	NP	LT	60
7	temperature	<u>Lucilia cuprina</u>	combination	NS	NP	MS	75
8	homozygosity	<u>Apis mellifera</u>	wing\tail	NS	P	PV	11
				NS	P	PV	11
				NS	P	PV	11

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
8	homozygosity	<u>Apis mellifera</u>		NS	P PO	PV	11
				NS	NP PO	PV	11
			hamuli#	NS	P PO	PV	13
			wing\tail	NS	P PO	PV	13
				NS	P PO	PV	13
				NS	P PO	PV	13
				NS	P PO	PV	13
			hamuli#	NS	NP PO	PV	13
			wing\tail	NS	P PO	PV	16
				NS	P PO	PV	16
9	insecticide	<u>Musca vetustissima</u>		NS	P PO	PV	16
				NS	P PO	PV	16
			hamuli#	NS	NP PO	PV	16
			wing\tail	NS	P PO	LT	284
				NS	P PO	LT	300
				NS	P PO	LT	300
				NS	P PO	LT	300
				NS	P PO	LT	284
				NS	P PO	LT	284
				NS	P PO	LT	284

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
12	homozygosity	<u>Drosophila melanogaster</u>	bristle/spine	NS	P	PO	LT 300
				NS	P	PO	LT 300
				NS	P	PO	LT 254
				NS	P	PO	LT 270
				NS	P	PO	LT 270
				NS	P	PO	LT 254
				NS	P	PO	LT 270
				NS	P	PO	LT 270
13	temperature	<u>Rattus norvegicus</u>	leg	NS	NP	PO	PV 2035
				NS	NP	PO	PV 2027
				NS	P	HO	LT 94
				NS	P	HO	LT 85
15	homozygosity	<u>Salmo clarki</u>	combination	NS	P	HO	LT 70
				NS	NP	PO	MS 51
16	homozygosity	<u>Salvelinus fontinalis</u>	combination	NS	NP	PO	PC 28
				NS	NP	PO	PC 59

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
17	density	<u>Salmo gairdneri</u>		NS	NP PO	PC	50
				NS	NP PO	PC	300
				NS	NP PO	PC	57
				NS	NP PO	PC	60
				NS	NP PO	PC	60
				NS	NP PO	PC	25
				NS	NP PO	PC	65
				NS	NP PO	PC	66
				NS	NP PO	PC	51
				NS	NP PO	PC	49
				NS	NP PO	PC	27
				NS	NP PO	PC	51
				NS	P PO	CS	100
				NS	P PO	CS	100
17	density	<u>Oncorhynchus mykiss</u>	fin	NS	NP PO	PC	51
				NS	P PO	CS	100
				NS	NP PO	CS	100
17	density	<u>Salmo Clarki</u>	gill	NS	NP PO	PC	51
				NS	P PO	CS	100
				NS	NP PO	CS	100

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
20	env degradation	<u>Gorilla gorilla</u>	sensory	NS	NP PO	CS	100
			tooth\skull	S	NP HO	PC	38
20	env degradation	<u>Gorilla gorilla</u>	tooth\skull	NS	NP HO	PC	45
				NS	NP HO	PC	45
				S	NP HO	PC	56
				NS	NP HO	PC	85
				NS	NP HO	PC	90
				NS	NP HO	PC	30
				NS	NP HO	PC	32
				NS	NP HO	PC	32
				NS	NP HO	PC	52
				NS	NP HO	PC	72
				NS	NP HO	PC	77
22	twinning	<u>Homo sapiens</u>	finger\hand	NS	NP HO	MS	52
				NS	NP HO	MS	67
				NS	NP HO	MS	52

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
23	homozygosity	<u>Homo sapiens</u>	finger\hand	NS	NP HO	MS	67
				NS	NP HO	MS	52
				NS	NP HO	MS	67
				NS	NP HO	MS	52
				NS	NP HO	MS	67
31	radiation	<u>Hirundo rustica</u>	wing\tail	NS	NP HO	MS	85
				NS	NP HO	MS	85
				S	P HO	MS	121
				NS	P HO	MS	121
				NS	P HO	MS	119
32	marginal habitat	<u>Macrodipteryx longipennis</u>	wing\tail	NS	P HO	MS	119
				NS	P HO	MS	119
				NS	P HO	MS	62
				NS	P HO	MS	62
				S	P HO	MS	62

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
				NS	P	HO	MS 62
			wing\tail	NS	P	HO	MS 36
			leg	NS	P	HO	MS 36
			wing\tail	NS	P	HO	MS 36
				NS	P	HO	MS 56
			leg	NS	P	HO	MS 56
			wing\tail	S	P	HO	MS 56
				NS	P	HO	MS 56
				NS	P	HO	MS 30
			leg	NS	P	HO	MS 30
			wing\tail	NS	P	HO	MS 30
				NS	P	HO	MS 30
			wing\tail	NS	P	HO	MS 24
			leg	NS	P	HO	MS 24
			wing\tail	S	P	HO	MS 24
				NS	P	HO	MS 24
32	marginal habitat	<u>Psalidoprocne obscura</u>	wing\tail	NS	P	HO	MS 24
			leg	NS	P	HO	MS 24
			wing\tail	S	P	HO	MS 24
				NS	P	HO	MS 24

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
32	marginal habitat	<u>Discurus paradisaeus</u>	leg	NS	P HO	MS	40
				NS	P HO	MS	40
				NS	P HO	MS	40
				NS	P HO	MS	40
				NS	P HO	MS	88
				NS	P HO	MS	88
		<u>Terpsiphone viridis</u>	leg	S	P HO	MS	88
				NS	P HO	MS	88
				NS	P HO	MS	44
				NS	P HO	MS	44
				NS	P HO	MS	44
				NS	P HO	MS	38
		<u>Vidua paradisaea</u>	leg	NS	P HO	MS	38
				NS	P HO	MS	38
				S	P HO	MS	38
				NS	P HO	MS	38
			wing\tail	NS	P HO	MS	16
				NS	P HO	MS	16

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
32	marginal habitat	<u>Euplectes jacksoni</u>	leg	NS	P	HO	MS 16
			wing\tail	NS	P	HO	MS 16
				NS	P	HO	MS 40
			leg	NS	P	HO	MS 40
			wing\tail	S	P	HO	MS 40
				NS	P	HO	MS 24
		<u>Quiscalus mexicanus</u>	leg	NS	P	HO	MS 24
			wing\tail	NS	P	HO	MS 24
				NS	P	HO	MS 90
			leg	NS	P	HO	MS 90
			wing\tail	S	P	HO	MS 90
				NS	P	HO	MS 42
33	patch size	<u>Turdus merula</u>	leg	NS	P	HO	MS 42
			wing\tail	NS	P	HO	MS 42
			wing\tail	NS	P	HO	MS 36
				NS	P	HO	MS 36

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
35	temperature	<u>Rattus norvegicus</u>	leg	NS	P HO	MS	15
				NS	P HO	MS	15
				NS	P HO	MS	29
				NS	P HO	MS	29
37	noise	<u>Rattus norvegicus</u>	leg	NS	P HO	LT	100
				NS	P HO	LT	91
				NS	P HO	LT	76
37	nutrition	<u>Sitta europaea</u>	wing\tail	NS	P HO	MS	68
39	pollution	<u>Peromyscus leucopus</u>	tooth\head	NS	NP HO	LT	50
				NS	NP HO	LT	50
				NS	NP HO	LT	50
				NS	NP HO	LT	50
39	pollution	<u>Peromyscus leucopus</u>	finger\hand	NS	NP HO	LT	50
				NS	P HO	LT	50
				NS	P HO	LT	50
39	pollution	<u>Sigmodon hispidus</u>	tooth\head	NS	NP HO	LT	24

Source	Correlate	Species/Organism	Character	Trait type	Thermy	Stat	Sample size	
42	inbreeding	<u>Thomomys bottae</u>	tooth\head	NS	NP	HO	SC	18
				NS	NP	HO	SC	18
				NS	NP	HO	SC	18
				NS	NP	HO	SC	18
				NS	NP	HO	SC	18
43	parasitism	<u>Drosophila nigrospiracula</u>	bristle\spine	NS	NP	PO	MS	151
				NS	P	PO	MS	133
				NS	NP	PO	PC	22
				NS	P	PO	PC	22
				NS	NP	HO	LT	72
46	noise	<u>Rattus norvegicus</u>	tooth\head	NS	P	PO	MS	67
				NS	P	PO	MS	67
48	drug	<u>Scatophaga stercoraria</u>	wing\tail	NS	P	PO	MS	67
				NS	P	PO	MS	67
51	nutrition	<u>Sturnus vulgaris</u>	wing\tail	NS	P	HO	MS	35
			pattern	S	NP	HO	MS	35
56	homozygosity	<u>Bolitotherus cornutus</u>	combination	NS	NP	PO	PC	65
57	pH	<u>Lepomis macrochirus</u>	gill	NS	NP	PO	SC	11

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
			scale	NS	NP	PO	SC 11
			fin	NS	P	PO	SC 11
	density		gill	NS	NP	PO	SC 11
			scale	NS	NP	PO	SC 11
			fin	NS	P	PO	SC 11
58	competition	<u>Sturnus vulgaris</u>	wing\tail	NS	P	HO	LT 36
59	density	<u>Sorex araneus</u>	combination	NS	NP	HO	PC 6
				NS	NP	HO	PC 6

Source	Correlate	Species/Organism	Character	Trait type	Thermy	Stat	Sample size	
				S	NP	HO	PC	8
				S	NP	HO	PC	8
				S	NP	HO	PC	8
20	character magnitude	<u>Gorilla gorilla</u>	tooth\skull	S	NP	HO	PC	56
21	character magnitude	<u>Pavo cristatus</u>	wing\tail	S	NP	HO	SC	17
24	mating success	<u>Drosophila simulans</u>	wing\tail	NS	P	PO	MS	100
				NS	P	PO	MS	100
			bristle\spine	NS	NP	PO	MS	100
				NS	NP	PO	MS	100
				NS	NP	PO	MS	100
		<u>Drosophila mojavensis</u>	wing\tail	NS	P	PO	MS	44
			bristle\spine	NS	NP	PO	MS	44
25	character magnitude	<u>Hirundo rustica</u>	wing\tail	NS	P	HO	PC	96
26	character magnitude	<u>Phasianus colchicus</u>	wing\tail	S	P	HO	SC	10
		<u>Tetrao tetrix</u>		S	P	HO	SC	10
		<u>Pharomachrus mocinno</u>		S	P	HO	SC	10

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
26	character magnitude	<u>Tanyptera galatea</u>		S	P HO	SC	10
		<u>Mevops nubicus</u>		S	P HO	SC	10
		<u>Cyonolechia mocoa</u>		S	P HO	SC	10
		<u>Poplairaia conversi</u>		S	P HO	SC	10
		<u>Macrodipteryx longipennis</u>		S	P HO	SC	10
		<u>Muscivora tyvannus</u>		S	P HO	SC	10
		<u>Teleoneara filicouda</u>		S	P HO	SC	10
		<u>Paradisaea rubra</u>	wing\tail	S	P HO	SC	10
		<u>Terpsiphone veridis</u>		S	P HO	SC	10
		<u>Hirundo rustica</u>		S	P HO	SC	10
27	character magnitude	<u>Nectarinia famosa</u>		S	P HO	SC	10
		<u>Euplectes jarkvoni</u>		S	P HO	SC	10
		<u>Vidua macroura</u>		S	P HO	SC	10
		<u>Lucanus dorcus</u>	horns	S	NP PO	PC	20
		<u>Heliocapris gigas</u>		S	NP PO	PC	20
		<u>Cupris ochus</u>		S	NP PO	PC	20

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
		<u>Drepanocerus sinicus</u>		S	NP	PO	PC 20
		<u>Liatongus fairmairei</u>		S	NP	PO	PC 20
		<u>Onthophagus mnischei</u>		S	NP	PO	PC 20
		<u>Chalcosoma atlas</u>		S	NP	PO	PC 20
		<u>Bolitotherus cornutus</u>		S	NP	PO	PC 20
		<u>Shasianus colchicus</u>	spurs	S	NP	HO	PC 10
		<u>Plectropterus gambensis</u>		S	NP	HO	PC 10
		<u>Chauna torquata</u>		S	NP	HO	PC 10
		<u>Jacana spinosa</u>		S	NP	HO	PC 10
		<u>Hoplopterus spinosus</u>		S	NP	HO	PC 10
27	character magnitude	<u>Chionis alba</u>		S	NP	HO	PC 10
28	premating time	<u>Hirundo rustica</u>	wing\tail	S	P	HO	MS 23
	egg-laying date			S	P	HO	MS 24
				S	P	HO	MS 23
				S	P	HO	MS 24
29	character magnitude	<u>Pavo cristatus</u>	wing\tail	S	P	HO	PC 20

Source	Correlate	Species/Organism	Character	Trait type	Thermy	Stat	Sample size	
		<u>Hirundo rustica</u>		S	P	HO	PC	20
		<u>Nectarinia johnstoni</u>		S	P	HO	PC	20
		<u>Vidua regia</u>		S	P	HO	PC	20
		<u>Euplectes progne</u>		S	P	HO	PC	20
		<u>Euplectes jacksoni</u>		S	P	HO	PC	20
		<u>Gallus gallus</u>		S	P	HO	PC	20
		<u>Phasianus colchicus</u>		S	P	HO	PC	20
		<u>Tetrao tetrix</u>		S	P	HO	PC	20
		<u>Centrocercus urophasianus</u>		S	P	HO	PC	20
		<u>Chiroxiphia linearis</u>		S	P	HO	PC	20
		<u>Parotia Lawesii</u>		S	P	HO	PC	20
30	premating period	<u>Hirundo rustica</u>	wing\tail	S	P	HO	MS	19
	prelaying period			S	P	HO	MS	19
	second brood			S	P	HO	MS	19
	survival	<u>Turdus merula</u>		NS	P	HO	MS	51
				NS	P	HO	MS	51

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
34	condition	<u>Hirundo rustica</u>	wing\tail	S	P	HO	PC 40
				S	P	HO	PC 40
				S	P	HO	PC 40
				NS	P	HO	PC 48
				NS	P	HO	PC 48
				NS	P	HO	PC 48
				NS	P	HO	PC 691
				NS	P	HO	PC 691
				NS	P	HO	PC 691
36	longevity	<u>Malacosoma dissdtria</u>	leg	NS	P	PO	PC 25
				NS	P	PO	PC 15
38	mating success	<u>Vidua paradisaea</u>	wing\tail	S	P	HO	LT 12
	character magnitude			S	P	HO	PC 73
40	longevity	<u>Panthera Leo</u>	face spots	NS	NP	HO	PC 176
			face spots	NS	NP	HO	PC 87
41	growth	<u>Ondatra zibethicus</u>	combination	NS	NP	HO	PC 9

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
	weight			NS	NP	HO	PC 9
	growth			NS	NP	HO	PC 316
				NS	NP	HO	PC 239
44	character magnitude	<u>Forficula auricularia</u>	forcep	S	P	PO	PC 116
	mating success		cerci	S	P	PO	MS 13
	condition			S	P	PO	SC 75
				S	P	PO	SC 75
			wing/tail	NS	P	PO	SC 75
				NS	P	PO	SC 75
45	mating call	<u>Acris crepitans</u>	leg	NS	P	PO	MC 85
				NS	P	PO	MC 85
				NS	P	PO	MC 85
			sensory	NS	NP	PO	MC 50
				NS	NP	PO	MC 50
				NS	NP	PO	MC 50
				NS	NP	PO	MC 55

Source	Correlate	Species/Organism	Character	Trait type	Thermy	Stat	Sample size	
47	character magnitude	<u>Alces alces</u>	antler	NS	NP	PO	MC	55
				NS	NP	PO	MC	55
				NS	NP	PO	MC	42
				NS	NP	PO	MC	42
				NS	NP	PO	MC	42
				NS	NP	PO	MC	55
				NS	NP	PO	MC	55
				NS	NP	PO	MC	55
				S	NP	HO	PC	644
				S	NP	HO	PC	665
				S	NP	HO	PC	429
				S	NP	HO	PC	388
				S	NP	HO	MC	574
	S	NP	HO	MC	585			
body mass				NP	HO	MC	387	
				NP	HO	MC	354	

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
49	mating success	<u>Taeniopygia guttata</u>	pattern	S	NP	HO	LT 24
				S	NP	HO	LT 24
50	mating success	<u>Taeniopygia guttata</u>	pattern	NS	NP	HO	LT 120
51	character magnitude	<u>Sturnus vulgaris</u>	wing/tail	NS	P	HO	MC 34
52	fighting ability	<u>Panorpa japonica</u>	wing/tail	NS	P	PO	MS 80
	mating success			NS	P	PO	MS 81
	survival			NS	P	PO	SC 21
				NS	P	PO	SC 22
	mating success			NS	P	PO	SC 21
				NS	P	PO	SC 22
53	mating success	<u>Panorpa japonica</u>	wing/tail	NS	P	PO	BP 25
				NS	P	PO	BP 25
				NS	P	PO	BP 15
54	nutritional status	<u>Panorpa vulgaris</u>	wing/tail	NS	P	PO	LT 63
			leg	NS	P	PO	LT 63
			wing/tail	NS	P	PO	LT 81

<u>Source</u>	<u>Correlate</u>	<u>Species/Organism</u>	<u>Character</u>	<u>Trait type</u>	<u>Thermy</u>	<u>Stat</u>	<u>Sample size</u>
55	longevity	<u>Harmonia axyridis</u>	wing\tail	NS	P PO	LT	42
				NS	P PO	SC	35
				NS	P PO	SC	37
58	dominance	<u>Sturnus vulgaris</u>	wing\tail	NS	P HO	LT	24
				NS	P HO	LT	24
	body mass			NS	P HO	LT	48
	fat stores			NS	P HO	PV	48
59	breeding success	<u>Sorex araneus</u>	combination	NS	NP HO	PC	6
				NS	NP HO	PC	6

Source list: 1) Beacham, 1990. 2) Beacham and Witter, 1987. 3) Blanco et al., 1990. 4) Clarke, 1993a. 5) Clarke, 1993b. 6) Clarke, Brand and Whitten, 1986. 7) Clarke and McKenzie, 1992. 8) Clarke, Oldroyd and Hunt, 1992. 9) Clarke and Ridsdill-Smith, 1990. 10) Eggert and Sakaluk, 1994. 11) Evans, Martins and Haley 1995. 12) Fowler and Whitlock, 1994. 13) Gest, Siegel and Anistranski, 1987 14) Harvey and Walsh, 1993. 15) Leary, Allendorf and Knudsen, 1985 16) Leary, Allendorf and Knudsen, 1984. 17) Leary, Allendorf and Knudsen, 1991 18) Liggett, Harvey and Manning, 1993. 19) Malyon and Healy, 1994. 20) Manning and Chamberlain, 1994. 21) Manning and Hartley, 1991. 22) Markow and

- Gottesman, 1989. 23) Markow and Martin, 1993. 24) Markow and Ricker, 1992. 25) Møller, 1990. 26) Møller and Höglund, 1990. 27) Møller, 1992a. 28) Møller, 1992b. 29) Møller, 1993c. 30) Møller, 1993d. 31) Møller, 1993b. 32) Møller, 1995a. 33) Møller, 1995b. 34) Møller et al., 1995. 35) Mooney, Siegel and Gest, 1985. 36) Naugler and Leech, 1994. 37) Nilsson, 1994. 38) Oakes and Barnard, 1994. 39) Owen and McBee, 1990. 40) Packer and Pusey, 1993. 41) Pankakoski, 1985. 42) Patterson and Patton, 1990. 43) Polak, 1993. 44) Radesäter and Halldórsdóttir, 1993. 45) Ryan et al., 1995. 46) Siegel and Mooney, 1987. 47) Solberg and Sether, 1993. 48) Strong and James, 1992. 49) Swaddle and Cuthill, 1994a. 50) Swaddle and Cuthill, 1994b. 51) Swaddle and Witter, 1994. 52) Thornhill, 1992a. 53) Thornhill, 1992b. 54) Thornhill and Sauer, 1992. 55) Ueno, 1994. 56) Whitlock, 1993. 57) Wiener and Rago, 1989. 58) Witter and Swaddle, 1994. 59) Zakharov et al., 1991.

Appendix C

To determine weighted effect sizes for a random model, we began with a fixed effect model and determined the variation of true effect sizes (σ^2_{θ}). We then used a weighted average to estimate the effect size in a fixed effect model (modified from Shadish and Haddock 1994; and Steinkamp and Maehr, 1983).

$$z = \sum W_{ij} z_{ij} / \sum W_{ij} \quad (1)$$

$$v_{ij} = 1 / (N_{ij} - 3) \quad (2)$$

$$W_{ij} = 1 / (v_{ij} * C_i) \quad (3)$$

$$v = 1 / \sum W_{ij} \quad (4)$$

Where z is the effect size, W is the weighting factor, N is the sample size, C is the number of correlations for a given species by trait type combination, z is the weighted average effect size, v is the sampling variance for a correlation, and v is the variance of z corrected for the number of tests per combination. The subscript i and j refer to species and correlations within species respectively. Note that $1 / (N_j - 3)$ is the variance of Fisher's z . Note that the function of C is to control for any dependence of data points such that each species by trait type combination contributes only one data point to the meta analysis.

Next, we computed the within group heterogeneity (Q_w) which indicates whether the variation between effect sizes is greater than would be expected due to sampling error. Q_w also estimates the unconditional variance of effect sizes (unconditional indicates a random effects model where there is variation in true effect sizes and variation is not due solely to sampling error; see Shadish and Haddock, 1994).

$$Q_w = \sum W_{ij}(z_{ij} - \bar{z})^2 \quad (5)$$

Q_w follows a chi-square distribution with $k-1$ degrees of freedom where k =number of species by trait type combinations. Other notation is as above.

Following Shadish and Haddock (1994):

$$E(Q) = c(\sigma^2_{\theta}) + (k-1) \quad (6)$$

where $E(Q)$ is the expected value of Q , σ^2_{θ} is the variance of true effect sizes and k is as above. c is given by:

$$c = \sum 1/(v_{ij}^* C_i) - [\sum 1/(v_{ij}^2 C_i) / \sum 1/(v_{ij}^* C_i)] \quad (7)$$

where notation is as above. Rewriting equation 6:

$$\sigma^2_{\theta} = [Q_w - (k-1)]/c \quad (8)$$

Finally, we obtained an unconditional variance estimate for each type of test that includes the sampling variance and the variation of true effect sizes:

$$v_{ij}^* = \sigma^2_{\theta} + v_{ij} \quad (9)$$

This was the basis for the random effects model. Hence forth, an asterisk indicates the random effects model; otherwise all notation is the same as above. v_{ij}^* of equation 9 was used in place of v_{ij} of equation 2 and calculations proceeded as it would for a fixed effect model in equations 1 to 4 (Shadish and Haddock, 1994).

Using the recalculated random effects values, significant average effect sizes ($z^* > 0$) was determined by:

$$Z = z^* / (v^*)^{.5} \quad (10)$$

Where Z is the normal standard deviate and one-tailed $\alpha = 0.05$ occurs when $Z = 1.645$. z^* is the weighted average effect size, and v^* is the variance of z^* .

Group comparisons were conducted using the following equation (Hedges, 1994b), again with the variance from the random effects model. Thus,

$$\chi^2 = (z_{1.} - z_{2.})^2 / (v_{1.} + v_{2.}) \quad (11)$$

Where $z_{i.}$ is the weighted average effect size for groups, $v_{i.}$ is the variance of each $z_{i.}$ ($i = \text{groups 1 and 2}$). This statistic follows a chi-square distribution with $df = 1$.

Finally, z values were backtransformed into Pearson's r using:

$$r = (e^{2Z} - 1) / (e^{2Z} + 1) \quad (12)$$

Appendix D

Parameters and terms used in Chapter 3.

Notation	Definition	Equation or values used
μ	Mean growth	(see equation 3-4)
δ	Parameter determining the strength of FA-quality relations.	δ values used were 0, 1, 5, and 10. In the absence of confounds, these corresponded to FA _u - quality relations of ca: $r=0$, $r=-0.21$, $r=-0.35$, and $r=-0.40$, respectively (see equation 3-7).
$\alpha:\beta$	Ratio determining the contribution of quality on size. The greater α is relative to β , the greater the strength of size-quality relations.	$\alpha:\beta$ ratios used were: -6:4, -4:6, 0:10, 4:6, and 6:4. These corresponded to size-quality relations of ca. $r=-0.83$, $r=-0.55$, $r=0$, $r=0.55$, and $r=0.83$, respectively (see equation 3-4).
a:b	Ratio determining the contribution of size on asymmetry. The greater a is relative to b, the greater the strength of size-FA _u relations.	a:b ratios used were: 0:1, 0.05:1, 0.1:1, 0.2:1, and 1:0. In the absence of confounds, these corresponded to size-FA _u relations of ca. $r=0$, $r=0.08$, $r=0.19$, $r=0.30$, and $r=0.37$, respectively (see equation 3-5).
CV	The Coefficient of variation of size	standard deviation $\div$ mean size

Notation	Definition	Equation or values used
DP	Absolute differences in power between an FA index and FA _u , under a given set of parameters (i.e., allometric relations).	DP=power of Rule - power of FA _u both in presence of same allometric relations
FA _{rel}	Relative asymmetry, i.e., absolute asymmetry scaled by average trait size	$FA_{rel} = .5x R-L / (R+L)$
FA _u	Absolute values of asymmetry not corrected for allometric trait size	$FA_u = R-L $
RP	Relative power of an FA index compared to FA _u in the absence of allometric relations.	RP=power of Rule in presence of allometric relations ÷ power of FA _u in absence of allometric relations x100%

Appendix E

Ranges of Type I error rates ($\delta=0$) for Rules examined (see Table 3-1) are presented for positive and inverse size-quality relations ($\alpha:\beta$ ratios) and at different size-FA relations ranging from no relation ($a:b = 0:1$) to isometry ($a:b = 0:1$) to FA $\propto$ size². Ranges of power was examined for Rules which did not show greatly inflated Type I errors in any parameter set. Both absolute differences in power (DP) and relative power (RP) was examined (defined in Appendix D) for positive size-quality relations. Only RP was examined for inverse size-quality relations because FA_u (Rule 1) had inflated Type I errors. Ranges of power were generated from 3 sets of FA-quality relations ($\delta=1, 5, \text{ and } 10$) and 2 sample sizes $N=40$ and 100 .

Type I error rates ($\delta=0$) for tests of FA-quality relations within samples.

pos quality	size-FA	Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6	Rule 7	Rule 8	Rule 9
-size ($\alpha:\beta$)	(a:b)									
0:10	0:1	4.3-4.8	4.3-4.6	4.0-4.5	4.2-4.7	4.2-4.6	4.2-4.7	4.2-4.6	4.2-4.6	4.2-4.6
	0.05:1	4.6-4.8	4.9-5.4	5.0-5.2	4.7-5.2	5.2-5.4	4.9-5.3	4.9-5.1	5.1-5.4	4.9-5.3
	0.1:1	5.1-5.7	4.0-5.9	4.5-5.8	3.7-5.5	3.9-5.5	5.1-6.5	5.3-6.2	5.0-6.0	5.6-6.4
	0.2:1	5.9-6.5	5.6-5.8	5.8-6.3	5.9-6.0	5.7-5.8	5.5-7.0	5.6-6.8	5.8-6.7	5.6-7.1
	1:0	5.2-5.5	5.1-5.6	5.2-5.5	5.3-5.6	5.3-5.9	5.1-5.6	5.2-5.5	5.5-5.7	5.5-5.6
4:6	0:1	4.6-5.4	27.2-55.2	4.6-6.0	7.9-11.4	13.2-14.5	4.5-5.0	4.6-5.4	5.1-7.5	5.5-6.3
	0.05:1	3.5-3.5	22.3-43.0	3.6-5.8	0.8-12.7	13.8-17.9	5.3-6.5	4.0-4.7	5.7-9.2	6.0-6.3
	0.1:1	0.8-2.0	15.4-25.5	2.4-6.7	0.9-12.8	12.2-15.4	4.2-4.9	2.7-3.3	4.0-10.2	3.7-4.3
	0.2:1	0.2-0.8	7.8-11.0	2.2-5.1	6.8-8.1	6.8-8.1	4.6-5.9	3.5-4.1	3.4-7.7	4.4-5.7
	1:0	0.0-0.7	5.2-6.2	1.3-4.3	4.6-5.3	4.4-5.3	5.0-7.3	3.8-7.2	2.4-4.8	4.6-7.3
6:4	0:1	4.8-6.1	52.0-84.5	5.0-7.0	11.5-17.5	25.4-26.5	5.1-6.0	5.2-6.4	7.1-10.3	7.2-8.9
	0.05:1	2.1-2.4	33.1-70.9	2.5-5.2	14.7-15.0	20.1-29.0	4.2-5.8	2.8-3.0	4.4-9.2	4.5-6.6
	0.1:1	0.1-1.0	23.9-41.7	1.1-7.1	12.9-17.5	14.9-22.9	5.5-6.1	2.3-3.0	3.5-9.9	5.4-5.6
	0.2:1	0.0-0.2	12.0-16.5	0.3-4.8	8.1-9.6	8.4-11.2	4.4-6.3	2.9-3.6	3.1-5.5	4.4-6.3

1:0	0.0-0.0	4.5-4.8	0.3-1.1	3.1-3.4	3.7-4.0	5.3-5.9	3.1-4.5	1.7-2.1	5.3-5.8
size ²	0.0	0.0	0.0	0.0	0.0	5.8	5.8	0.0	5.8

inv. Quality	size-FA	Rule 1	Rule 2	Rule 3	Rule 6	Rule 7	Rule 8	Rule 9	Rule 10	Rule 11
-size ($\alpha:\beta$)	(a:b)									
-4:6	0:1	4.9-4.9	0.0-0.2	3.6-3.7	4.1-4.9	3.8-3.8	4.8-4.8	4.7-4.8	3.6-3.6	3.6-3.6
	0.05:1	7.0-7.2	0.1-0.4	4.0-4.8	4.7-4.7	4.7-5.1	6.3-7.0	6.3-7.0	4.0-4.8	4.0-4.8
	0.1:1	13.4-18.8	0.8-2.1	6.3-7.3	5.0-6.0	6.6-7.3	12.0-13.9	11.8-13.9	6.3-7.1	6.3-7.0
	0.2:1	18.7-34.4	2.3-2.3	4.9-6.8	5.5-5.5	6.6-7.3	14.1-15.3	13.7-15.0	4.8-6.7	5.0-6.7
	1:0	24.9-49.5	4.5-5.2	6.3-8.3	4.5-5.2	5.8-6.7	14.7-17.0	12.0-15.9	5.4-8.1	5.6-7.8
-6:4	0:1	4.6-5.7	0.0-0.1	1.9-2.8	4.1-5.7	2.5-2.9	4.4-5.4	4.4-5.4	1.9-2.8	1.9-2.8
	0.05:1	9.1-15.0	0.0-0.6	3.3-4.7	4.8-6.0	3.9-5.7	8.4-13.7	8.4-13.7	3.3-4.7	3.2-4.7
	0.1:1	19.0-33.0	0.0-0.8	5.1-6.0	4.8-5.1	6.0-7.0	16.5-26.9	16.3-26.9	5.1-5.9	5.3-5.7
	0.2:1	31.2-59.8	1.3-2.0	3.9-6.3	5.2-5.7	6.9-6.9	26.9-41.2	26.8-40.7	3.7-6.2	3.8-6.1
	1:0	43.7-77.2	3.7-5.9	4.6-10.7	4.1-5.9	6.4-7.3	35.9-43.4	34.8-41.9	3.9-9.5	3.8-9.5
	size ²	99.1	78.5	78.5	5.3	5.3	80.9	16.7	18.5	39.2

power of tests of FA-quality relations within samples for positive size-quality relations.

pos quality -size ($\alpha:\beta$)	Size-FA (a:b)	Rule 1	Rule 3	Rule 6	Rule 7	Rule 8	Rule 9
<u>Relative Power (RP)</u>							
0:10	0:1	100-100	96-100	99-100	99-100	98-100	100-100
	0.05:1	95-101	92-99	95-100	95-100	93-100	96-101
	0.1:1	94-100	89-98	95-100	94-100	90-98	95-100
	0.2:1	88-99	90-98	92-99	91-99	91-98	92-99
	1:0	81-98	93-100	91-99	90-99	95-101	91-99
4:6	0:1	100-100	100-100	81-97	100-100	100-106	100-106
	0.05:1	79-99	80-99	76-96	80-99	86-99	83-99
	0.1:1	56-96	61-96	81-97	63-96	81-98	77-97
	0.2:1	28-90	40-92	74-98	45-92	59-97	65-96
	1:0	17-82	37-89	78-98	49-89	63-96	73-96
6:4	0:1	100-100	100-100	49-79	100-100	100-108	100-108
	0.05:1	72-98	72-98	49-78	73-98	81-98	82-98
	0.1:1	33-91	35-91	47-74	35-91	54-93	52-93
	0.2:1	11-78	16-78	44-77	18-78	37-86	42-85
	1:0	2-52	9-54	47-77	15-54	34-77	43-76
	Size ²	0-1	2-35	39-74	38-56	2-54	39-74

pos	Size-FA	Rule 1	Rule 3	Rule 6	Rule 7	Rule 8	Rule 9
quality	(a:b)						
-size							
(α : β)							
<u>Difference Power (DP)</u>							
0:10	0:1	NA	-1.6-0.0	-1.0-0.0	-0.5-0.0	-0.8-0.0	-0.1-0.3
	0.05:1	NA	-4.1-0.0	-0.9-0.0	-0.8-0.0	-3.1-0.0	-0.5-0.4
	0.1:1	NA	-4.1-0.0	0.1-1.3	-0.3-0.9	-3.4-0.0	0.1-1.3
	0.2:1	NA	-0.9-1.5	0.1-3.3	0.1-2.8	-0.8-2.4	0.1-3.1
	1:0	NA	1.2-9.1	0.7-5.9	0.7-6.2	1.2-9.2	0.7-6.3
4:6	0:1	NA	0.0-0.1	-14.4-0.0	0.0-0.2	0.0-2.3	0.0-2.3
	0.05:1	NA	0.0-1.0	-7.5-1.5	0.0-0.9	0.3-4.8	0.3-4.3
	0.1:1	NA	0.3-5.4	0.2-18.0	0.3-4.7	1.6-18.4	1.4-15.0
	0.2:1	NA	1.3-19.1	7.7-36.0	1.6-16.0	6.5-40.0	6.1-32.2
	1:0	NA	2.6-29.8	15.8-43.9	4.1-27.8	13.8-48.4	13.9-42.5
6:4	0:1	NA	0.0-0.0	-39.3-0.0	0.0-0.0	0.1-3.1	0.1-3.2
	0.05:1	NA	0.0-0.1	-29.7-0.0	0.0-0.2	0.1-5.4	0.1-4.9
	0.1:1	NA	0.0-1.6	-19.3-10.2	0.0-1.2	1.7-15.9	1.5-13.6
	0.2:1	NA	0.1-6.8	-0.8-27.0	0.2-4.5	7.7-31.3	7.2-24.7
	1:0	NA	0.3-13.1	15.3-32.4	0.7-12.3	10.4-36.7	13.9-30.4
	Size ²	NA	1.3-33.1	27.9-72.6	27.6-54.8	1.7-52.6	27.9-72.5

Relative power of tests of FA-quality relations within samples for inverse size-quality relations.

Within Samples.			
inv. quality -size ($\alpha:\beta$)	size-FA (a:b)	Rule 7	Rule 8
<u>Relative Power (RP)</u>			
-4:6	0:1	78-96	79-97
	0.05:1	79-97	87-97
	0.1:1	73-97	84-97
	0.2:1	77-96	82-96
	1:0	76-96	77-96
-6:4	0:1	44-79	49-79
	0.05:1	47-77	49-77
	0.1:1	46-74	47-74
	0.2:1	45-72	47-72
	1:0	44-73	44-73
	size ²	45-63	45-63

Appendix F

Composite FA analyses (CFA). The CFA numbers correspond to numbers used in the text. Subscripts i=individual, j= trait, and k=number of traits per individual.

Index	Equation	Description
CFA 1	$CFA_i = \sum_{j=1 \text{ to } k} FA_{ij} $	Summation of FA values. CFA and FA denote composite asymmetry values for an individual and the absolute asymmetry for each trait, respectively.
CFA 2	$CFA_i = \sum_{j=1 \text{ to } k} FA_{ij} / \text{avg } FA_j $	Summation of standardized absolute FA values. FA values of a given trait are divided by the average absolute FA magnitude for that trait. Thus, FA values of all traits are given equal weight.
CFA 3	$CFA_i = \sum_{j=1 \text{ to } k} \text{ranked } FA_{ij} $	FA values of individuals are ranked for each trait. Thus, FA values are standardized across traits. The ranked values of each trait are summed to give the CFA value of an individual.

Index	Equation	Description
CFA 4	$GFA = G \cdot (\det S)^{1/k}$	Generalized FA variance. where G eliminates statistical bias, det S is the determinant of the variance-covariance matrix of FAs of different traits. GFAs of stressed versus unstressed populations are compared using F-ratio test. Zhivotovsky (1992) provides necessary matrix notation and formulae, and a brief table of G values and degrees of freedom.
CFA 5	$FA_{ijg} = \mu + \alpha_j + \beta_g + (\alpha\beta)_{jg} + \epsilon_{ijg}$	Two-Way ANOVA model (Type I) using trait and sample (stressed versus unstressed) as independent factors and FA as the dependent factors. μ , α_j , β_g , $(\alpha\beta)_{jg}$, ϵ_{ijg} , and the subscript g denote the grand mean, the effect of trait, the effect of stress, the interaction term, the error term, and population (grouping variable), respectively (Palmer 1994). Here, we are interested in whether the effect of stress (β_g) is significant. Equation following Lindeman <u>et al.</u> (1980 p137).

Index	Equation	Description
CFA 6	$D_{ig} = \sum u_j (FA_{ijg})$	MANOVA examining FA differences between samples. Coefficients (u) are calculated to maximize the differences between D values of stressed and unstressed populations. Equation following Lindeman <u>et al.</u> (1980, p171).

Appendix G

Ranges of Type I errors and Power were presented for analyses using single traits and CFA measures (see Appendix F for CFA measures). Baseline conditions were that the underlying FA distributions of all traits were normally distributed, centered at zero, uncorrelated, without measurement error, and homogeneous between traits. We used 40 individuals and 10 traits for the baseline simulations. We then modified parameters to examine five levels of heterogeneity between traits, three levels of leptokurtosis, three levels of correlations of FA values between traits, three levels of measurement error, two other sample sizes, and a different number of traits. Levels 1-3 for kurtosis corresponded with average kurtosis values (g) of ca. $g = 0.7$, 2.9 , and 7.6 , respectively. Levels 1-3 for correlations of FA values corresponded with average FA-FA correlations of $r=0.0052$, $r=0.039$, and $r=0.085$, respectively. Each level of each condition was examined at three different effect sizes ($E=1$, $E=1.2$, and $E=1.5$). Effect size (E) was measured as the ratio of FA variances between stressed ($\sigma^2_{2,t}$) and unstressed populations ($\sigma^2_{1,t}$) (i.e., $E=\sigma^2_{2,t} / \sigma^2_{1,t}$), where t denoted a trait. When $E=1$, there was no difference between populations and significance was due to Type I errors. Conditions are discussed in more depth in the text.

Type I error rates ($E=1$, see text) of CFA analyses at different parameter values.

Conditions/levels	% significant						
	Single Trait	CFA 1	CFA 2	CFA 3	CFA 4	CFA 5	CFA 6
<u>Baseline</u>	5.3	4.4	4.5	3.2	5.1	4.4	5.6
<u>Heterogeneous FA distributions</u>							
1) $\sigma^2_{1,1-5}=3 \times \sigma^2_{1,6-10}$	4.6	5.3	5.2	3.7	4.1	4.9	5.2
2) $\sigma^2_{1,1}=3 \times \sigma^2_{1,2}$ $\sigma^2_{1,2}=\sigma^2_{1,3}=\dots=\sigma^2_{1,10}$	4.7	3.9	4.4	3.0	4.0	4.1	5.3
3) $\sigma^2_{1,4-6}=3 \times \sigma^2_{1,1-3}$ and $\sigma^2_{1,7-10}=10 \times \sigma^2_{1,3}$	4.2	5.2	4.5	2.9	4.0	5.2	5.2
4) $\sigma^2_{1,1-5}=10 \times \sigma^2_{1,6-10}$	4.8	5.1	7.1	4.2	5.8	5.8	5.8
5) $\sigma^2_{1,1}=10 \times \sigma^2_{1,2}$ $\sigma^2_{1,2}=\sigma^2_{1,3}=\dots=\sigma^2_{1,10}$	4.8	4.2	5.0	3.3	4.1	4.3	4.8
<u>Leptokurtism</u>							
1) $\sigma^2_{p,t}$ ranged from 1-3	5.0	6.0	5.8	6.1	6.6	5.9	4.3
2) $\sigma^2_{p,t}$ ranged from 1-10	6.6	4.7	4.8	5.9	17.6	5.0	3.8
3) $\sigma^2_{p,t}$ ranged from 1-100	4.1	5.8	6.7	6.2	33.7	6.4	2.2
<u>Correlations of FA values</u>							
1) $C=0.091$	4.8	4.7	4.6	4.7	4.5	5.0	4.1
2) $C=0.23$	5.3	5.3	5.3	6.3	6.6	9.0	3.8

Conditions/levels	% significant						
	Single Trait	CFA 1	CFA 2	CFA 3	CFA 4	CFA 5	CFA 6
3) C=0.33	4.9	6.2	6.3	6.4	9.6	12.2	5.0
<u>Measurement error</u>							
1) ME=0.091	4.6	5.3	5.5	5.6	3.7	5.1	4.2
2) ME=0.23	4.3	5.1	5.2	4.8	4.2	4.9	4.4
3) ME=0.33	5.2	5.2	5.2	5.2	4.5	5.2	4.8
<u>Number of traits</u>							
1) k=5	6.3	6.0	5.8	4.7	5.1	5.4	6.1
<u>Sample Size</u>							
1) N=80	5.5	4.5	4.6	3.1	3.9	4.4	4.4
2) N=20 (k=5)	5.4	4.2	4.0	4.3	4.1	3.3	4.6

Power (% significant) of CFA analyses at different parameter values (condition numbers correspond with condition numbers and descriptions above).

		%significant						
E	Conditions/levels	Single Trait	CFA 1	CFA 2	CFA 3	CFA 4	CFA 5	CFA 6
<u>Baseline</u>								
1.2		10.2	34.9	35.3	27.4	25.8	35.1	9.4
1.5		20.6	85.2	84.8	70.3	74.1	85.1	28.3
<u>Heterogeneous FA distributions</u>								
1.2	1	8.6	31.4	31.5	26.8	26.8	32.6	8.5
	2	9.7	33.0	33.3	26.7	27.5	33.4	8.7
	3	8.9	29.4	31.6	26.4	25.8	30.2	8.5
	4	10.8	28.0	31.6	25.8	25.1	27.8	9.0
	5	10.7	29.2	34.7	28.3	26.7	29.4	8.3
1.5	1	19.1	83.5	83.7	72.9	74.6	84.3	29.5
	2	19.6	82.8	83.5	69.3	75.0	83.8	27.0
	3	19.2	79.0	84.3	71.5	77.4	80.1	29.0
	4	21.7	77.4	84.1	72.5	75.3	78.9	28.8
	5	20.9	75.4	84.0	70.2	74.7	76.3	27.8
<u>Leptokurtism</u>								
1.2	1	8.5	28.0	26.5	21.1	30.5	28.1	7.3

%significant								
E	Conditions/levels	Single Trait	CFA 1	CFA 2	CFA 3	CFA 4	CFA 5	CFA 6
	2	5.9	18.6	17.6	18.3	34.2	18.8	5.0
	3	4.5	10.2	10.7	16.0	43.3	10.0	2.3
1.5	1	18.8	75.1	74.4	62.5	73.0	76.4	23.3
	2	14.8	49.9	50.5	49.6	60.6	50.9	10.1
	3	8.5	20.9	24.4	46.0	59.4	22.6	4.3
<u>Correlations of FA</u>								
<u>values</u>								
1.2	1	10.3	32.6	31.3	26.1	26.4	35.2	8.3
	2	11.1	26.8	26.3	21.5	29.0	36.0	7.4
	3	8.2	21.0	20.9	18.1	29.4	35.2	5.1
1.5	1	21.1	85.2	84.0	72.6	74.4	86.9	28.4
	2	17.7	70.7	70.1	56.6	70.7	81.2	16.4
	3	20.5	61.5	60.7	52.4	70.5	78.5	14.4
<u>Measurement error</u>								
1.2	1	9.8	28.3	28.0	22.4	22.6	29.0	6.2
	2	9.7	22.1	22.8	17.9	19.8	24.1	6.2
	3	8.1	18.5	18.1	14.8	14.8	18.9	6.6
1.5	1	18.4	78.3	78.0	63.8	69.3	80.4	22.0

%significant								
E	Conditions/levels	Single Trait	CFA 1	CFA 2	CFA 3	CFA 4	CFA 5	CFA 6
	2	16.4	69.9	68.2	53.8	58.4	70.6	17.3
	3	13.9	58.6	57.3	44.2	47.5	60.0	12.6
	<u>Number of traits</u>							
1.2	1	8.1	20.9	20.9	16.7	17.3	21.3	7.4
1.5	1	19.9	59.2	59.2	46.9	55.8	59.7	21.4
	<u>Sample size</u>							
1.2	1	13.1	51.4	51.5	40.9	49.9	51.3	13.7
	2	8.8	14.4	14.1	12.2	11.8	15.4	6.2
1.5	1	32.7	98.6	98.1	90.7	98.3	98.7	64.6
	2	13.2	33.7	32.6	26.8	27.1	36.1	9.2

Appendix H

Estimated kurtosis values at which analyses using CFA 3 becomes stronger than CFA 2. I calculated t-values using CFA 3 and CFA 2 for 100 simulations. Between simulations, I increased the degree of kurtosis of the FA distribution, such that a range of kurtosis values was produced. I subtracted the t-values of CFA 2 from t-values of CFA 3 as a measure of the differences in power of these 2 analyses ($D = t_{CFA3} - t_{CFA2}$). When D was positive, analyses using CFA3 was stronger than CFA2. I regressed D against kurtosis and determined a best fit line. The point at which the best fit line intercepted the x-axis (i.e., $D=0$), indicated the point at which CFA 3 became stronger than CFA 2. However, this was only a rough estimate as there was a lot of overlap at intermediate levels of kurtosis. This is illustrated in the figure below, using a sample size $N=40$, number of traits $k=10$, and effect size $E=1.5$. The point at which CFA 3 became stronger than CFA 2 was at a kurtosis value of ca. $g=3.5$. Below, I present a table of kurtosis values at which CFA 3 became stronger than CFA 2. Each value presented is based on 10 x-intercepts (i.e., 1000 simulations), such that I could determine the variation in my estimates (standard deviation of 10 estimates shown in parentheses). The percent of significant relations (using CFA 2) is also indicated (square brackets) to show the range of powers that we examined. I examined a range of sample sizes (N), effect sizes (E) and number of traits (k) (500,000 simulations in total). Note that kurtosis values are based on signed FA values of single traits, whereas analyses using CFA 2 and CFA 3 are based on $|FA|$.

k=2

N	E=1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
10	0.85(0.31)[5]	0.70(0.26)[3]	1.21(0.41)[7]	0.95(0.34)[6]	1.00(0.33)[7]	0.72(0.24)[9]	0.78(0.26)[8]	0.94(0.31)[10]	0.69(0.34)[10]	0.39(0.22)[11]
20	2.62(0.80)[5]	2.46(0.76)[6]	2.58(0.79)[9]	2.37(0.73)[9]	1.92(0.61)[10]	2.01(0.63)[9]	1.66(0.52)[13]	1.97(0.60)[13]	1.88(0.59)[16]	1.74(0.54)[14]
30	4.11(1.26)[7]	3.42(1.04)[7]	3.58(1.09)[9]	3.48(1.06)[10]	2.68(0.87)[12]	2.94(0.91)[14]	3.02(0.93)[15]	2.61(0.79)[19]	2.54(0.78)[17]	2.51(0.77)[21]
40	5.46(1.69)[4]	4.40(1.34)[7]	4.47(1.37)[9]	4.22(1.29)[11]	3.59(1.13)[14]	3.37(1.03)[15]	3.70(1.14)[17]	3.33(1.04)[20]	2.91(0.90)[21]	2.53(0.80)[25]
50	6.18(2.10)[7]	5.45(1.69)[7]	4.97(1.58)[10]	3.74(1.19)[14]	3.95(1.31)[15]	4.20(1.27)[16]	4.00(1.23)[20]	3.40(1.10)[24]	3.35(1.05)[25]	3.39(1.04)[26]
60	7.48(2.28)[7]	5.28(1.66)[7]	4.36(1.45)[10]	5.40(1.67)[12]	4.89(1.49)[17]	4.48(1.37)[20]	3.82(1.26)[20]	3.63(1.13)[28]	3.63(1.12)[29]	3.53(1.13)[31]
70	7.28(2.27)[7]	7.39(2.25)[8]	5.65(1.85)[11]	5.56(1.72)[13]	5.57(1.70)[18]	5.16(1.60)[22]	4.38(1.37)[26]	4.15(1.29)[27]	3.07(0.96)[31]	2.44(0.83)[34]
80	9.03(2.78)[5]	7.86(2.39)[7]	6.61(2.05)[12]	4.85(1.56)[15]	5.44(1.68)[19]	4.58(1.49)[20]	4.14(1.30)[29]	3.55(1.13)[28]	3.62(1.19)[32]	2.87(1.10)[36]
90	8.45(2.68)[6]	8.90(2.73)[10]	5.86(1.98)[13]	5.51(1.78)[16]	3.85(1.57)[22]	4.64(1.50)[26]	3.31(1.05)[27]	2.99(0.99)[30]	2.66(1.23)[38]	3.26(1.02)[39]
100	11.84(3.71)[6]	8.54(2.65)[10]	6.70(2.08)[13]	5.37(1.88)[18]	5.20(1.65)[24]	4.11(1.34)[25]	4.34(1.53)[31]	3.42(1.26)[37]	2.62(1.11)[37]	1.82(1.17)[41]

k=3

N	E=1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
10	1.00(0.36)[5]	1.26(0.41)[6]	1.11(0.34)[6]	0.80(0.28)[8]	0.76(0.26)[8]	0.77(0.24)[10]	0.66(0.23)[12]	0.60(0.24)[12]	0.57(0.22)[13]	0.76(0.28)[15]
20	2.76(0.85)[6]	2.84(0.86)[7]	2.46(0.75)[10]	2.08(0.66)[9]	2.04(0.63)[10]	2.17(0.67)[14]	1.84(0.56)[16]	2.01(0.62)[18]	1.67(0.53)[21]	1.68(0.51)[20]
30	4.15(1.26)[7]	3.62(1.09)[8]	3.75(1.14)[9]	3.13(0.96)[11]	3.09(0.95)[15]	2.96(0.91)[14]	2.85(0.88)[18]	2.41(0.76)[21]	2.65(0.81)[25]	2.24(0.69)[25]
40	5.26(1.60)[5]	4.58(1.40)[9]	4.02(1.24)[10]	3.86(1.18)[14]	3.64(1.11)[18]	3.48(1.08)[17]	3.39(1.05)[21]	3.43(1.05)[23]	3.01(0.94)[31]	3.12(0.95)[30]
50	5.84(1.78)[6]	5.76(1.75)[9]	5.30(1.64)[12]	5.08(1.56)[16]	4.10(1.28)[17]	4.15(1.31)[21]	3.45(1.08)[25]	3.15(0.99)[30]	3.22(1.00)[33]	2.71(0.85)[35]
60	6.59(2.02)[7]	6.95(2.12)[10]	4.74(1.50)[13]	5.18(1.58)[16]	4.47(1.41)[20]	4.09(1.30)[23]	4.35(1.36)[31]	3.45(1.06)[31]	3.46(1.06)[36]	2.91(0.94)[37]
70	7.84(2.38)[7]	6.80(2.18)[10]	6.16(1.88)[18]	4.72(1.47)[18]	5.18(1.59)[24]	3.71(1.33)[24]	3.60(1.13)[31]	3.47(1.12)[37]	4.16(1.28)[38]	3.79(1.17)[43]
80	7.83(2.44)[8]	6.79(2.11)[11]	6.80(2.08)[16]	5.61(1.76)[21]	4.37(1.53)[24]	3.66(1.23)[28]	3.92(1.25)[35]	2.90(0.93)[40]	3.35(1.27)[44]	3.16(1.02)[46]
90	8.48(2.66)[6]	7.13(2.26)[11]	6.16(1.91)[15]	6.24(1.91)[21]	5.08(1.56)[25]	5.43(1.70)[33]	4.76(1.53)[34]	3.91(1.37)[43]	3.34(1.15)[47]	2.42(0.88)[49]
100	10.07(3.07)[9]	7.88(2.43)[12]	6.35(2.11)[16]	6.44(1.98)[23]	4.09(1.41)[26]	3.15(1.19)[31]	3.56(1.18)[36]	2.28(0.89)[43]	3.66(1.16)[50]	2.15(0.96)[52]

k=4

N	E=1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
10	1.35(0.45)[6]	1.32(0.41)[5]	0.92(0.29)[7]	0.62(0.31)[7]	0.95(0.32)[10]	0.73(0.23)[10]	0.75(0.23)[11]	0.69(0.23)[12]	0.59(0.21)[14]	0.67(0.28)[17]
20	2.79(0.85)[5]	2.61(0.79)[7]	2.53(0.77)[10]	2.35(0.71)[10]	2.26(0.69)[15]	2.16(0.66)[15]	1.90(0.58)[15]	1.75(0.55)[19]	1.60(0.49)[23]	1.64(0.51)[25]
30	4.28(1.31)[7]	3.94(1.21)[10]	3.23(1.00)[12]	3.48(1.06)[13]	2.89(0.89)[17]	3.29(1.00)[17]	2.87(0.87)[19]	2.82(0.85)[23]	2.33(0.72)[28]	2.27(0.70)[30]
40	5.30(1.61)[5]	4.63(1.41)[10]	4.29(1.30)[12]	4.08(1.24)[17]	3.83(1.16)[19]	3.67(1.13)[24]	3.52(1.08)[24]	3.25(0.98)[26]	2.91(0.90)[34]	2.46(0.79)[38]
50	6.54(2.00)[7]	5.64(1.72)[10]	5.10(1.59)[12]	4.42(1.37)[20]	4.25(1.33)[20]	4.35(1.34)[25]	4.13(1.26)[25]	3.94(1.19)[34]	3.27(1.01)[38]	3.02(0.92)[42]
60	6.77(2.07)[7]	6.09(1.87)[11]	5.62(1.71)[15]	5.37(1.66)[19]	4.57(1.41)[24]	4.22(1.31)[24]	4.49(1.37)[29]	3.84(1.19)[35]	3.87(1.18)[44]	3.34(1.04)[48]
70	7.98(2.46)[9]	6.16(1.91)[10]	6.41(1.99)[18]	4.84(1.49)[18]	5.11(1.57)[20]	4.25(1.36)[27]	3.93(1.24)[31]	4.17(1.29)[36]	3.22(1.06)[43]	3.69(1.16)[50]
80	8.79(2.68)[9]	6.26(1.98)[12]	5.57(1.89)[18]	5.98(1.86)[22]	5.52(1.70)[29]	4.54(1.47)[29]	4.29(1.35)[36]	3.92(1.26)[41]	2.73(0.87)[51]	3.10(1.00)[55]
90	9.30(2.86)[9]	6.92(2.16)[14]	5.11(1.82)[19]	6.44(1.99)[24]	4.91(1.59)[24]	4.60(1.44)[32]	4.51(1.44)[39]	4.14(1.32)[44]	3.09(1.00)[55]	2.53(0.86)[59]
100	8.62(2.64)[8]	6.67(2.13)[15]	6.94(2.17)[22]	5.03(1.55)[26]	4.85(1.63)[26]	4.02(1.29)[32]	4.26(1.34)[38]	3.38(1.05)[46]	3.27(1.05)[52]	2.13(0.81)[61]

k=5

<i>N</i>	<i>E=1.1</i>	<i>1.2</i>	<i>1.3</i>	<i>1.4</i>	<i>1.5</i>	<i>1.6</i>	<i>1.7</i>	<i>1.8</i>	<i>1.9</i>	<i>2.0</i>
10	1.14(0.35)[4]	1.07(0.39)[7]	0.95(0.33)[10]	0.98(0.30)[10]	0.80(0.30)[8]	0.79(0.27)[13]	0.64(0.22)[14]	0.80(0.26)[16]	0.67(0.23)[15]	0.58(0.19)[20]
20	2.51(0.77)[6]	2.39(0.74)[7]	2.38(0.72)[10]	2.36(0.72)[13]	1.97(0.60)[13]	2.05(0.62)[16]	1.92(0.58)[23]	1.56(0.49)[23]	1.87(0.58)[25]	1.74(0.53)[30]
30	4.36(1.33)[7]	3.71(1.12)[8]	3.71(1.12)[14]	3.12(0.96)[15]	3.15(0.97)[18]	3.01(0.93)[23]	3.03(0.92)[25]	2.66(0.81)[30]	2.45(0.75)[32]	2.25(0.70)[37]
40	5.45(1.66)[6]	4.88(1.49)[13]	4.38(1.32)[14]	4.25(1.29)[16]	3.95(1.21)[22]	3.48(1.06)[26]	3.34(1.02)[32]	2.76(0.85)[34]	2.77(0.86)[38]	2.74(0.86)[45]
50	6.01(1.83)[7]	5.59(1.70)[11]	5.10(1.56)[17]	4.60(1.40)[20]	4.07(1.26)[24]	3.68(1.14)[32]	3.64(1.11)[34]	3.63(1.11)[39]	3.09(0.95)[43]	2.95(0.90)[48]
60	6.70(2.07)[7]	6.51(2.00)[11]	6.40(1.95)[19]	4.61(1.43)[22]	4.17(1.31)[26]	4.11(1.25)[34]	3.93(1.20)[38]	3.62(1.12)[44]	3.19(0.98)[47]	3.52(1.08)[54]
70	7.71(2.55)[8]	6.45(1.96)[11]	6.26(1.92)[20]	5.31(1.64)[22]	4.52(1.43)[31]	4.35(1.37)[35]	4.03(1.25)[41]	3.28(1.02)[49]	3.50(1.13)[49]	2.84(0.99)[57]
80	6.67(2.11)[8]	6.78(2.09)[13]	5.86(1.80)[19]	5.51(1.72)[27]	5.03(1.53)[32]	4.60(1.42)[41]	3.58(1.26)[47]	3.56(1.12)[52]	3.69(1.15)[60]	3.50(1.10)[61]
90	7.74(2.45)[8]	7.76(2.40)[15]	6.18(1.92)[21]	5.05(1.60)[28]	5.39(1.67)[35]	5.59(1.73)[44]	4.78(1.46)[50]	3.46(1.12)[55]	3.44(1.09)[64]	2.44(0.86)[63]
100	8.38(2.58)[9]	6.42(2.11)[14]	5.99(1.84)[22]	5.55(1.72)[30]	5.51(1.70)[36]	4.61(1.41)[43]	4.11(1.37)[51]	2.90(0.98)[56]	2.57(0.98)[64]	3.21(1.02)[70]

k=10

N	E=1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
10	1.21(0.38)[7]	1.06(0.33)[8]	0.94(0.29)[11]	0.93(0.29)[14]	0.89(0.28)[16]	0.88(0.28)[18]	0.46(0.38)[20]	0.85(0.26)[20]	0.67(0.21)[26]	0.54(0.19)[28]
20	2.66(0.81)[6]	2.51(0.76)[10]	2.25(0.69)[14]	2.31(0.70)[18]	2.09(0.64)[19]	1.79(0.55)[26]	2.04(0.62)[30]	1.84(0.56)[34]	1.84(0.56)[41]	1.80(0.55)[42]
30	3.75(1.14)[7]	3.57(1.10)[12]	3.29(1.00)[16]	2.71(0.84)[21]	3.34(1.01)[29]	2.84(0.86)[32]	2.60(0.79)[38]	2.59(0.79)[38]	2.40(0.74)[49]	2.06(0.63)[54]
40	4.89(1.49)[8]	4.49(1.36)[15]	4.12(1.25)[20]	4.01(1.21)[25]	3.43(1.06)[34]	3.35(1.02)[41]	3.12(0.98)[47]	2.98(0.91)[52]	3.00(0.91)[60]	2.82(0.87)[63]
50	5.90(1.79)[8]	5.79(1.77)[13]	4.42(1.35)[21]	3.95(1.20)[30]	4.20(1.27)[37]	3.91(1.19)[47]	3.52(1.07)[52]	3.62(1.10)[61]	3.11(0.96)[65]	3.05(0.94)[69]
60	6.47(1.98)[10]	5.72(1.74)[17]	5.04(1.55)[22]	4.90(1.49)[34]	4.33(1.33)[41]	4.07(1.25)[50]	3.89(1.19)[60]	3.59(1.12)[62]	3.37(1.03)[71]	3.60(1.09)[77]
70	7.14(2.17)[9]	6.36(1.95)[18]	5.35(1.64)[26]	4.98(1.52)[40]	4.39(1.37)[48]	3.93(1.22)[56]	3.53(1.10)[63]	3.58(1.09)[69]	3.16(0.98)[76]	3.64(1.13)[80]
80	6.55(2.14)[10]	5.77(1.78)[20]	5.09(1.56)[28]	4.70(1.44)[40]	4.45(1.36)[49]	4.04(1.24)[56]	4.05(1.25)[67]	3.75(1.17)[74]	3.13(0.97)[78]	2.81(0.92)[83]
90	8.78(2.68)[12]	6.16(1.96)[20]	5.71(1.74)[32]	4.70(1.54)[43]	4.88(1.50)[53]	4.26(1.33)[63]	3.85(1.19)[69]	3.37(1.05)[77]	3.21(0.99)[82]	3.41(1.08)[86]
100	8.67(2.67)[12]	5.55(1.94)[22]	4.89(1.58)[35]	5.31(1.62)[46]	4.67(1.44)[56]	4.42(1.36)[67]	4.05(1.25)[73]	3.94(1.22)[77]	3.59(1.13)[85]	3.20(1.01)[89]

Fig H-1. The difference in power of CFA 2 and CFA 3. I subtracted the t-values of CFA 2 from t-values of CFA 3 as a measure of the differences in power of these 2 analyses ($D=t_{\text{CFA3}}-t_{\text{CFA2}}$). I regressed D against kurtosis and determined a best fit line. The point at which the best fit line intercepted the x-axis (i.e., $D=0$), indicated the point at which CFA 3 became stronger than CFA 2. However, note the high degree of overlap at intermediate levels of kurtosis. I used sample size $N=40$, number of traits $k=10$, and effect size $E=1.5$.

difference in t-values (CFA 3- CFA 2)

Appendix I

Fig I-1. Asymmetry-quality plots within samples using a single trait (top panel) and a composite measure combining 10 traits (bottom panel). FA-quality relations were generated using an algorithm following Leung and Forbes (1997, Chapter 2). Composite measure was calculated using CFA 2. Each plot contains 40 individuals. r value for single traits and for CFA 2 was $r=-0.19$ and $r=-0.63$, respectively. The r value of the single trait was representative of the average r value for the 10 single traits.

232a