

Individual Cryptic Scaling Relationships and the Evolution of Animal Form

W. Anthony Frankino^{1*}, Eric Bakota¹, Ian Dworkin², Gerald S. Wilkinson³, Jason B. Wolf⁴, and Alexander W. Shingleton⁵

¹ Department of Biology and Biochemistry, University of Houston, Houston Texas, 77204-5001, USA

² Department of Biology, McMaster University, Hamilton, Ontario, L9H 6X9, Canada

³ Department of Biology, University of Maryland, College Park, Maryland 20742

⁴ Milner Centre for Evolution and Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, UK

⁵ Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607

* *corresponding author*

Abstract

Artificial selection offers a powerful tool for the exploration of how selection and development shape the evolution of morphological scaling relationships. An emerging approach models the expression and evolution of morphological scaling relationships as a function of variation among individuals in the developmental mechanisms that regulate trait growth. These models posit the existence of genotype-specific morphological scaling relationships that are unseen or 'cryptic'. Within-population allelic variation at growth-regulating loci determines how these individual cryptic scaling relationships are distributed, and exposure to environmental factors that affect growth determines the size phenotype expressed by each individual on their cryptic, genotype-specific scaling relationship. These models reveal that evolution of the intercept and slope of the *population*-level static allometry is determined, often in counterintuitive ways, largely by the shape of the distribution of these underlying *individual*-level scaling relationships. Here we review this modeling framework and present the wing-body size individual cryptic scaling relationships from a population of *Drosophila melanogaster*. To determine how these models might inform interpretation of published work on scaling relationship evolution, we review studies where artificial selection was applied to alter the parameters of population-level static allometries. Finally, motivated by our review, we outline areas in need of empirical work and describe a research program to address these topics; the approach includes describing the distribution of individual cryptic scaling relationships across populations and environments, empirical testing of the model's predictions, and determining the effects of environmental heterogeneity on realized trait distributions and how this affects allometry evolution.

“It is suggested that selection provides a valuable tool for studying the constancy or lability of the growth patterns which determine morphology” – (Robertson 1962)

Introduction. The evolution of morphological scaling relationships is a primary mechanism by which morphological diversification occurs. Traits that evolve steep scaling relationships represent striking examples of exaggerated ornaments, weapons, and specialized tools (e.g., Gould 1974; Huxley 1932; Larabee and others 2017; Wilkinson and Dodson 1997). Traits that evolve shallow scaling relationships are perhaps less conspicuous (e.g., Eberhard 2009; Koh and others 2005) but are equally interesting developmentally, ecologically, and evolutionarily. Despite intense interest in how scaling relationships are maintained within biological groups (e.g., sexes, populations, species, etc.), and how they are modified to generate morphological diversity among groups, surprisingly little is known conclusively regarding the role of selection in these processes.

Humans have used artificial selection to create domestic breeds with desirable proportions for millennia (Clutton-Brock 1992; Larson and Fuller 2014). However, it is only 70 years since Robertson and Reeve’s pioneering work (Robertson 1962; Robertson and Reeve 1952) established artificial selection as a way to investigate the evolutionary malleability of relative trait size, body proportion, and morphological scaling. Since then, several studies have used artificial selection to determine the evolutionary independence of the intercept and slope of the allometric equation (e.g., Bolstad and others 2015; Egset and others 2012; Stillwell and others 2016); the relative roles of natural selection and developmental constraint in scaling relationship expression and evolution (e.g., Bolstad and others 2015; Booksmythe and others 2015; Frankino and others 2007; Frankino and others 2005; Kotrschal and others 2013); the heritability of individual traits or the scaling relationship between them (e.g., Emlen 1996; Weber 1990; Wilkinson 1993); the role of sexual selection in the production of relative trait size (e.g., Booksmythe and others 2015; Cayetano and others 2011; Ewing 1961; Menezes and others 2013; Wilkinson and Reillo 1994), etc. (see reviews in Baker and Wilkinson 2001; Frankino and others 2009; Houle and others, this volume; Pélabon and others 2014; Shingleton and Tang 2012; Voje and others 2013).

In tandem with these empirical studies, a variety of mathematical treatments have sought to identify how natural selection might transform one pattern of scaling into another (e.g., Bonduriansky and Day 2003; Fromhage and Kokko 2014; Kodric-Brown and others 2006;

Lande 1979). New information regarding the developmental regulation and integration of trait growth has been used increasingly to study the expression and evolution of morphological scaling relationships; space limitations preclude a discussion here of the mechanisms underlying growth regulation, however, interested readers are directed to several recent reviews for details (e.g., Casasa and Moczek, this volume; Cooper, this volume; Emlen and Allen 2004; Mirth and others 2016; Nijhout and others 2014; Shingleton 2011; Shingleton and Frankino 2018; Shingleton and Frankino 2013; Shingleton and others 2007; Shingleton and others 2008; Warren and others 2013). Some studies model formally the role of development in scaling expression (e.g., Nijhout and Wheeler 1996; Shingleton and others 2008) and an emerging approach distinguishes between two types of scaling relationships: *individual* cryptic scaling relationships, and the well-established concept of the *population-level* static allometry (Figure 1; Dreyer and others 2016; Houle and others, this volume; Shingleton, this volume). Distinguishing between these classes of scaling relationship generates potentially revealing ways of thinking about the mathematical representation of morphological scaling relationships, the developmental mechanisms that selection targets to alter scaling, how genetic variation among individuals in these mechanisms influences the response to selection, and the importance of environmentally-induced size variation in the evolution of scaling (Dreyer and others 2016; Shingleton and others 2007; Shingleton and others 2008, Shingleton, this volume). Below, we sketch the general aspects of this emerging approach and the new insights they offer. We then provide data describing the individual cryptic morphological scaling relationships central to this approach, and investigate what these tell us about the study population. Finally, we use existing data from the literature on artificial selection applied to morphological scaling relationships to explore static allometry evolution in the context of the model, before turning our attention to areas for future work.

The Model. Briefly, the model (Dreyer and others 2016; Shingleton, this volume) employs known developmental mechanisms that regulate trait and body growth to generate a latent morphological scaling relationship for each genotype (Figure 1A); this linear, *individual*-level morphological scaling relationship represents the continuum of size combinations for any two traits (e.g., appendages, eyes, or other organs, some proxy of body size, etc.) potentially expressed by a single genotype across the full range of possible size induced by variation in a single environmental factor (Shingleton and others 2007). Thus, each of these genotype-specific individual scaling relationships constitutes a linear function-valued trait (Stinchcombe and others 2012) described by a potentially unique intercept and slope. Because individuals realize just one

joint size phenotype on this genotype-specific continuum, the relationship itself is not observed; the model therefore refers to these as ‘individual cryptic scaling relationships’.

Allelic identity at growth-regulating loci determines the intercept and slope of the individual cryptic scaling relationship, and genetic variation among individuals at these loci generates a distribution of the individual cryptic scaling relationships in the population; when populations differ in the genetic variation they harbor, the pattern of individual cryptic scaling relationships will vary among populations (Figure 1B-D; see the original model for a formal, detailed presentation). When there is genetic variation in slopes, these distribution patterns can be divided into three categories based on the average point at which the individual cryptic scaling relationships converge (i.e., the mean point of intersection, or MPI) relative to the range of trait sizes potentially expressed by each individual: the ‘speedometer’, where the MPI is near or below the smallest trait sizes (Figure 1B); the ‘broomstick’, where the MPI is near or above largest trait sizes (Figure 1C); and the ‘seesaw’, where the MPI is toward center of the size distribution, near the bivariate mean (Figure 1D). While these categories are discrete in the model for the sake of convenience, in reality they exist along a continuum of possibilities that includes not only the location of the MPI but also how tightly the intersection points lie around it. Regardless of the distribution of the individual cryptic scaling relationships in a population, the individual phenotypes expressed on them are used collectively to fit the static allometry to the group (Figure 1A). As discussed below, the pattern of the individual cryptic scaling relationships in a population can have wide-ranging, important effects on how the group-level static allometry responds to selection.

The Sources of Size Variation. The individual cryptic scaling relationship captures size covariation between traits across the full range of possible size for a single genotype. Each genotype expresses a single phenotype that corresponds to a point on their individual scaling relationship; the location of this point is determined by the level of the environmental factor experienced during ontogeny. In practice, however, measurement error and developmental instability may place the measured phenotype slightly off the genotype-specific individual scaling relationship (Shingleton, this volume).

The original model (Dreyer and others 2016) was concerned with the impact of variation in nutrition on trait and body size. This is reasonable, as nutritional variation during growth affects final size in nearly all metazoans via conserved developmental mechanisms (Gokhale and

Shingleton 2015; Nijhout and others 2014; Oldham and Hafen 2003). However, other environmental factors can also generate variation in size (e.g., temperature or oxygen level Atkinson 1994; Frazier and others 2001; Harrison and others 2010; Harrison and others 2015), and may do so in a trait-specific manner. For example, in *Drosophila*, wing size is more sensitive to temperature than it is to nutrition, relative to the rest of the body, whereas thorax size is similarly sensitive to both nutrition and temperature. Consequently, the wing-thorax individual scaling relationships are steeper when size varies in response to temperature than when size varies in response to nutrition (Shingleton and others 2009). Further, size variation need not be due solely to environmental factors. Genetic variation in the level of systemic size regulators (e.g., growth hormones) also generates size covariation among traits (Shingleton and others 2007), and may produce a genetic scaling relationship that is quite different from an environmental scaling relationship for the same trait pairs (Dreyer and Shingleton 2011). Thus, environmental effects on trait size, body size, and size covariation can be complex and may vary in trait- and environment-specific ways.

Trait size is not simply a developmental consequence of body size (see Gokhale and Shingleton 2015 for review). This means that scaling relationships are not equivalent to a single reaction norm where the environment is body size that determines trait size. Rather, bivariate morphological scaling relationships are the combination of *two* reaction norms, one for trait size and one for body size (Shingleton and others 2007). Describing a scaling relationship as a single reaction norm obfuscates the scaling relationship itself with the developmental mechanisms that regulate or integrate trait and body growth to generate their individual reaction norms, confounds different types of scaling relationships (e.g., environmental versus genetic; Shingleton and others 2007), and risks mischaracterizing the developmental and evolutionary relationship between trait and body size. To avoid implying a hierarchy of growth regulation, where body size determines trait size, below we refer to morphological allometries where the size of one trait (x) is plotted against the size of a second trait (y); x usually refers to some index of body size, however, it could be the size of any morphological trait.

Whether the focus is environmentally or genetically determined scaling relationships, the source of size variation matters (Shingleton and others 2009; Shingleton and others 2007). Ultimately, the impact of different sources of size variation on the expression and evolution of scaling will depend on the genotypic variation that resides at growth-regulating loci, the level of variation

exhibited by relevant environmental factors in nature, and the degree to which these environmental factors influence phenotype expression.

Evolutionary Insight from the Model. The individual cryptic scaling relationships formalize how genotype by environment interactions influence the size of traits, and how size covaries between traits across the range of possible trait sizes. Under the Dreyer model (2016), these interactions reflect specific functional/mechanistic relationships, and genetic variation underlying these relationships structures the distribution of individual scaling relationships within populations. *More than just a static picture of observed variation, however, these distributions may account for the potentially predictable impact that selection can have on patterns of this variation, and how this affects expression and evolution of the population-level static allometry.* Evolutionary simulations using this modeling framework reveal three broad insights regarding how population-level static allometries are expected to evolve (Dreyer and others 2016). First, and perhaps most importantly, they show that the pattern of the individual cryptic scaling relationships in a population (Figure 1) can affect greatly the response of the population-level static allometry parameters to selection. This effect can be so strong that phenotypically indistinguishable populations - i.e., those that have the same trait size distributions, trait size covariances, and population-level static allometries - that differ in their underlying distributions of individual cryptic scaling relationships can evolve differently in response to the same pattern of selection; a single pattern of selection can produce steep or shallow slopes (i.e., hyper- or hypoallometry) or generate no response at all, depending on the distribution of the individual scaling relationships in the population. Collectively, such effects mean that predicting or interpreting the response of a population-level static allometry to selection may require knowledge of the distribution of the underlying individual cryptic scaling relationships.

Second, the model reveals that evolution of the static allometry slope depends on the mode and target of selection in somewhat surprising ways. Most notably, selection on the ratio of trait:trait size is thought to shift the intercept but not the slope (e.g., Wilkinson 1993). Yet this pattern of selection (called proportional selection in the model) can result in rapid evolution of hyper- or hypoallometry under some distributions of individual cryptic scaling relationships. In contrast, most models of scaling relationship evolution posit that the static allometry slope evolves in response to complex patterns of selection (e.g., Biernaskie and others 2014; Bonduriansky and Day 2003; Emlen and others 2012; Fromhage and Kokko 2014; Green 1992; Kodric-Brown and others 2006; Petrie 1992; Zeh and Zeh 1988). The Dreyer and others (2016) model

demonstrates that such correlational selection will alter the slope under only a subset of conditions. Interestingly, the model indicates that directional univariate selection on absolute trait size can indirectly cause swift evolution of the scaling relationship slope, indicating that complex patterns of multivariate selection or combinations of different selection patterns may not be the only – or even dominant – mechanism underlying slope evolution. These findings may inform studies that seek to determine the origin of trait exaggeration and sometimes complicated fitness functions posited to produce steep scaling relationships under sexual selection (e.g., Biernaskie and others 2014; Emlen and others 2012; Green 1992; Kodric-Brown and others 2006; Petrie 1992; Zeh and Zeh 1988).

Third and finally, the efficacy of, and response to, selection may be affected by the statistical method used to fit the population-level static allometry. Where such variation in the response to selection exists, using an ordinary least squares (OLS) Model I regression line as a reference to identify individuals for selection tends to result in a weaker evolutionary response than does using major axis (MA) Model II regression techniques (e.g., see Figure 4 in Dreyer and others 2016). This may be because, as compared to other line-fitting approaches, MA regression better captures the relationship between observed trait values and the underlying distribution of possible phenotypes from the individual cryptic scaling relationship in some populations (see Shingleton, this volume). Regardless of the cause, this finding may add a new dimension to ongoing debates regarding appropriate regression methodologies for fitting static allometries (e.g., Bolstad and others 2015; Egset and others 2012; Frankino and others 2009; Hansen and Bartoszek 2012; Houle and others 2011; Stillwell and others 2016; Shingleton, this volume).

In sum, by taking an approach that focuses on individual-level variation in the mechanisms regulating trait growth, this modeling framework (Dreyer and others 2016; Shingleton, this volume) offers new insights into the evolution of morphological scaling. It reveals that the response to selection of the population-level static allometry slope depends on the frequency of loci that generate pattern in the distribution of individual cryptic scaling relationships in a population. Moreover, the evolutionary trajectory and speed of any response is not only determined by how the individual relationships are distributed, but also by the target of selection, and the method used to fit the population-level static allometry. Below we use isogenic *Drosophila melanogaster* lineages to describe the pattern of individual cryptic wing-body size scaling relationships for a population of flies.

Individual Cryptic Scaling Relationships in *Drosophila*.

To quantify individual cryptic scaling relationships for several dozen *Drosophila melanogaster* genotypes, we manipulated access to nutrition for larvae from isogenic lineages and used the resulting individuals to estimate the sex-specific cryptic scaling relationship across the full range of size for each genotype.

System: Focal Traits & Isogenic Drosophila Lineages. The wing size-body size scaling relationship in *D. melanogaster* is a good model for study because it is relatively straightforward to estimate (Stillwell and others 2011) and is ecologically relevant. The likely functional importance of wing size, body size, and relative wing size (Bastock and Manning 1955; Bennet-Clark and Ewing 1968; De Jong and Bochdanovits 2003; Dudley 2002; Ewing 1961; 1964; Frankino and others 2009; Lane and others 2014; Menezes and others 2013), and geographic variation, thermally-induced plasticity, and long-term evolutionary trends in how these traits scale with one another (e.g., Azevedo and others 1998; Bolstad and others 2015; Frazier and others 2008; Gilchrist and Huey 2004; Gilchrist and others 2004; Gilchrist and others 2001; Lane and others 2014; Liefting and others 2009) suggest that the intercept and slope of this scaling relationship may be subject to strong and perhaps predictably variable selection.

We used flies from the *Drosophila* Genetic Reference Panel (DGRP), a collection of *D. melanogaster* lineages that were established from flies collected at a farmer's market in Raleigh, North Carolina, USA, and propagated through over 20 generations of full-sib mating to create ca. 200 isogenic lineages (Mackay and others 2012). Below, we use 'lineage' to refer to a specific isogenic line (i.e., a genotype) from the DGRP. We use 'population' in reference to a cohort of genetically identical individuals from a DGRP lineage. Thus, multiple populations may be used to estimate the individual cryptic scaling relationship parameters for a given genotype.

General Methods. Egg collection, larval rearing, larval diet, application of the starvation treatment, and phenotyping followed our standard lab protocols (Myers and Frankino 2012; Stillwell and others 2011; Stillwell and others 2016). To generate populations for estimation of the individual cryptic scaling relationship parameters for genotypes in the DGRP, we collected eggs over 12-20h from each lineage, transferred them in lots of 50 to 7ml of fly food, and reared larvae by genotype at 22°C until application of the starvation treatment. Larvae were removed from food at precisely timed developmental stages and starved for either 0-24h (referred as Day 0 flies) or 48-62h (Day 2 flies) before pupation to ensure expression of the full range of wing and

body and wing size for each lineage. Pupae were isolated into individual 2.5ml epitubes that had been punctured with a 20-gauge needle for gas exchange. The resultant adults fell into two size classes where the minimally starved individuals were ~75% larger than the individuals subject to longer starvation. The right wing and pupal case of each fly were imaged and used to estimate wing and body size, and sex-specific body (pupal)-wing size scaling relationships were fit to each lineage using Model I and Model II linear regression.

Experimental Design & Statistical Analysis. We used an incomplete blocking scheme; flies from each lineage were reared in populations divided among six experimental blocks, and populations from some lineages were replicated among blocks. At a minimum, each block contained populations from 12 lineages; populations from these lineages are included in all blocks.

All analyses were conducted in *R*. The scripts for the analyses, as well as the data analyzed, are available from Dryad. All size data used to fit the individual cryptic scaling relationships for each lineage were log-log transformed. Some theoretical studies suggest that Model II linear regression best captures the developmental mechanisms that generate morphological scaling between body parts in *Drosophila* (Shingleton, this volume) and so, where possible, the scaling relationship between wing and pupal size was fit using this approach (*R* package: *smatr*). However, for completeness, and when testing more complex statistical models (for example, where lineage was treated as a random factor), we fit the relationship using Model I linear regression, using maximum likelihood (*R* package: *lme4*), and Bayesian methods (*R* package: *MCMCglmm*). We excluded lineages if six or fewer individuals of each sex were measured; other cutoffs for the number of observations within lineages (e.g., 12 or 18 per sex) had no qualitative effect on our results.

Analysis & Results

There is Genetic Variation in Individual Scaling Relationships. We analyzed wing- and body (pupal)-size data from 87 of the DGRP lineages. The scaling relationship between wing and pupal size differs between sexes (Table 1), and consequently we analyzed males and females separately. In both sexes, the individual cryptic scaling relationship between wing and body size varies significantly among lineages when fit using a MA Model II regression (treating lineage as a fixed factor; sma test: *Wing = Pupa·Lineage*, Male: likelihood ratio = 257.0, DF = 80, $P < 0.0001$; Female: likelihood ratio = 257.4, DF = 82, $P < 0.0001$; Figure 2) or linear mixed-model

regression (treating lineage and block as a random factor, analogous to a Model I regression but with shrinkage towards the global mean slope and intercept across lineages; Table 2; images for this analysis are presented in material on Dryad). Male and female individual cryptic scaling relationship MA Model II slopes exhibited a positive relationship among lineages (MA regression: female MA slope = male MA slope, slope = 0.95, $R^2 = 0.1322$, $P < 0.001$), and this trend held when OLS Model I slopes were analyzed similarly (OLS regression: female OLS slope = male OLS slope, $R^2 = 0.2263$, $F_{(1,65)} = 22.82$, $P < 0.0001$; Figures 3, S1, S2).

The Wing-Body Individual Cryptic Scaling Relationships Differ by Sex. Conceptually, there are three categories of individual scaling relationship patterns or distributions: speedometer, seesaw and broomstick (Figure 1 and Dreyer and others 2016). While there is no clear way to capture the differences among these three patterns in a single statistic, one approach is to compare the location of the MPI with the bivariate mean of all individuals. We calculated the MPI for our DGRP lineages using the intersections of all pairwise combinations of the individual cryptic scaling relationships by sex using MA Model II regression (Figure 2). We then used Hotelling's two-sample T^2 to test whether the MPI differed significantly from the bivariate size mean, comparing the distribution of all pairwise intersection points of the individual scaling relationships with the location of the bivariate mean of wing and pupal size. Our protocols were designed to ensure that phenotypes were represented well from across the full range of possible sizes for each genotype. This may, however, produce size distributions in our experimental population that are unrepresentative of those in nature or typical artificial selection experiments, both of which may be deficient in small individuals. To guard against any potential bias this could introduce, we conducted a conservative analysis by using two estimates of the bivariate mean: one based on all individuals and one based only on the minimally starved (Day 0) treatment individuals. For males, the MPI does not differ significantly from the bivariate mean of all individuals nor from the bivariate mean of minimally starved individuals (Table 3). In other words, the MPI is close to the center of the range of male trait sizes, suggesting their pattern of individual scaling relationships is seesaw-like (Figure 2). In contrast, for females the MPI is significantly different from the bivariate mean trait size of all individuals, and closer to the origin, suggesting that their pattern of individual scaling relationships is speedometer-like (Figure 2; Table 3). The female MPI is also closer to the origin than the bivariate mean of minimally-starved individuals, but this difference is not significant (Table 3). The results were qualitatively the same when the individual scaling relationships were fit by OLS Model I regression, although

the MPIs of either sex were not significantly different than the bivariate mean trait size, for all individuals or for minimally starved individuals (Table 2).

Under the speedometer distribution of individual scaling relationships, lineages with larger mean relative wing size in the largest size classes should have steeper individual cryptic scaling relationships (i.e., the mean relative wing size of the larger size classed individuals from a lineage and the lineage's individual scaling relationship slope should be correlated positively), whereas there should be only a weak (if any) relationship between mean relative wing size and the slope of the individual cryptic scaling relationships in the smaller size classes (Figure 4A). In contrast, for the seesaw distribution, there should be a positive relationship between the slope of the individual scaling relationship and mean relative wing size among the largest individuals, but this relationship should be negative among the smallest individuals (Figure 4B). We found a significant negative correlation between the mean relative wing size of a lineage – using data from the smallest 25% of all individuals, based on pupal size – and the slope of the lineage's individual scaling relationship in males but not in females (Figure 4C, D). In contrast, we found a significant positive correlation between the mean relative wing size of a lineage – using data from the largest 25% of all individuals based on pupal size – and the slope of that lineage's individual scaling relationship for both females and males (Figure 4E, F). These findings are consistent with the analyses of the sex-specific individual cryptic scaling relationship distributions described above.

With the caveat that these statistical analyses may imperfectly capture differences between groups in the pattern of individual scaling relationships, collectively our data suggest that the pattern of individual scaling relationships is speedometer-like in females and seesaw-like in males. However, closer examination reveals an interesting relationship between these sex-specific distributions. The MPI does not differ significantly between males and females (Hotelling's two-sample T^2 test, $T^2 = 0.496$, $F_{2,13941}=0.248$, $P = 0.780$). This, combined with the observation that the slopes of the individual cryptic scaling relationships are correlated between sexes among lineages (Figures 3, S1, S2), suggests that males and females of the same lineage occupy different size ranges on more-or-less the same individual cryptic scaling relationships that are distributed as a seesaw (compare the size distributions, individual scaling relationships, and locations of the MPI by sex and see caption for Figure 2). In other words, males appear on these seesaw-distributed cryptic individual scaling relationships from the smaller wing and pupal sizes to just beyond the MPI. Females, however, essentially occupy the

larger pupal and trait sizes on this distribution of cryptic scaling relationships – the area where the individual cryptic scaling relationships fan outward and to the right from the MPI. Their location in the region of larger trait and body sizes on the overall seesaw distribution means that females effectively experience a speedometer distribution.

Discussion. Many empirical and theoretical studies, some discussed above, have sought to determine how different patterns of selection might alter morphological scaling relationships to produce morphological diversity. Perhaps because the parameters of the allometric equation describe how traits scale within a biological group, and because individuals do not express a scaling relationship slope or intercept, historically this work has focused more on the population-level static allometries than on the individuals used to fit these scaling relationships. However, recent empirical studies assume particular models of relative trait growth that produce individual scaling relationships, and they assume that (genetically determined) variation underlying growth parameters may produce variation among individual scaling relationships (Bolstad and others 2015; Egset and others 2012). The emerging framework described here formalizes this approach to explicitly model how genetic variation among individuals in the developmental mechanisms that regulate trait growth affects the expression and evolution of morphological scaling. Under this approach, genotype determines the continuum of trait and body size covariation potentially expressed by an individual across a gradient of some environmental factor. This continuum constitutes an individual cryptic scaling relationship for that genotype, whereas the realized phenotype of the individual is a single point (set of trait values) on that continuum, determined by the environment the individual experienced (Figure 1). Genetic variation among individuals at growth-regulating loci generates the distribution of the individual cryptic scaling relationships in the population, the pattern of which can affect strongly the response to uni- and multivariate selection.

Our analyses of the of individual cryptic wing-body size scaling relationship distribution for a population of *D. melanogaster* reveal variation among genotypes in the slopes of these relationships, and a positive correlation between the slopes for males and females within lineages. For our study population, the individual cryptic scaling relationships in males seem to be distributed in a seesaw-like pattern, where their mean point of convergence is near the bivariate mean for body and wing size. Females appear to exhibit a speedometer-like distribution, where the individual cryptic scaling relationships fan outward and upward from the mean point of convergence. However, sexual size dimorphism and the correlation of slope

between sexes among lineages means that females essentially occupy the larger body size-wing size portions of the same seesaw distribution occupied by the smaller sized males. This unexpected finding has important implications, which we explore below under Future Directions.

While the individual cryptic scaling relationships have not been estimated for any population subject to selection, it still may be worthwhile to review, in the context of the Dreyer and others (2016) model, experiments that have selected to change population-level scaling relationship parameters. Below, we review in this framework some experiments that have used artificial selection to alter the parameters of morphological scaling relationships. We then turn attention toward describing the kinds of data that are needed to test the model directly.

Robertson (1962) was the first to select on body proportion directly in this context. Using *D. melanogaster*, he selected on the wing:thorax ratio – that is wing size relative to body size – in both directions for 10 generations. Although the pooling of data between sexes and the omission of scaling relationship visualizations from the publication makes the details of the response to selection difficult to decipher, he generated populations of flies with divergent wing:thorax ratios and scaling relationship slopes. Under the Dreyer and others (2016) model, such a response is expected when size ratios are selected from a speedometer pattern of cryptic individual scaling relationships – a distribution consistent with our observations for females from the DGRP lineages (with the caveats that we are using different measures of wing and body size than Robertson 1962, and that our population is unusual in that it was created from flies collected from a farmer's market and then inbred). Intriguingly, Wilkinson (1993) selected on the eyestalk span:body length ratio on a natural scale to change the scaling relationship intercept in male stalk eyed flies. In addition to changing of the intercept as intended, the slopes of the selected groups diverged. However, there is no observable difference in slope between selected directions when these data are log-log transformed under MA Model II regression, and the results are equivocal under Model I regression (see analysis on Dryad). Thus, selecting on trait size ratios can clearly alter the intercept of morphological scaling relationships, although the effect this might have on the slope is unclear, as is the role of the distribution of individual cryptic scaling relationships in determining these responses.

Most studies that have employed artificial selection to alter body proportion, however, have used an individual's distance in morphospace from the population-level static allometry as the basis of a selection index; individuals with extreme residuals from the group-level static

allometry, sometimes weighed by their location in the size distribution along x or y, are retained to breed the next generation. This general approach has been used to change intercepts (Bolstad and others 2015; Bookmythe and others 2015; Egset and others 2012; Frankino and others 2007; Frankino and others 2005), slopes (Egset and others 2012; Stillwell and others 2016), and even the inflection point of curvilinear scaling relationships (Emlen 1996).

Three studies reported expected responses of the intercept when selection was focused on the residual distance of each individual perpendicular to the replicate-specific static allometry fit to each sex via reduced MA regression (Bookmythe and others 2015; Frankino and others 2007; Frankino and others 2005). Additionally, Kotrschal and others (2013) observed a 9% divergence in the intercept of brain-body size allometry in guppies after just two generations. In these studies, the slopes either did not change or did so to varying degrees and direction among treatments and replicates within studies (see reanalysis of log-log transformed data on Dryad from both Frankino and others 2007, and Frankino and others 2005). Such inconsistent responses of the slope suggest that these shifts are not the result of an indirect response of this parameter to selection on the intercept. Two additional studies – Egset and others (2012), and Bolstad and others (2015) – effectively targeted the intercept of the static allometries using a more complicated index for selection that included residuals relative to the static allometry. In these two studies, the intercepts responded rapidly to selection and slope remained unchanged, demonstrating unequivocally that the intercept can evolve independently of the slope. Under the Dreyer and others (2016) model, directional selection on the relative trait size as described here (proportional selection in the model) will not affect the slope of the scaling relationship when the distribution of individual cryptic scaling relationships is a seesaw and/or when selection is focused on individuals that are close to the MPI for the individual scaling relationships (which is close to the bivariate mean under the seesaw distribution). Interestingly, Bolstad and others (2015) intentionally selected individuals with extreme residuals that were close to the bivariate mean. In all these studies, the consistent response of the intercept paired with the lack of any, or any consistent, response in slope suggests that the trait pairs under consideration in each may possess a seesaw distribution.

Selecting to change the population-level scaling relationship slope, and generating a response, has proven more challenging than using artificial selection to change the intercept. Two studies – Bolstad and others (2015) and Stillwell and others (2016) – sought to change the scaling relationship slope directly. A third study, (Egset and others 2012) also attempted to alter the

scaling relationship slope, but methodological issues make interpretation of their findings less clear and so this study is not considered here in detail. Although Bolstad and others (2015) and Stillwell and others (2016) employed different selection procedures, both had the explicit aim of evolving the slope while keeping the mean size of both traits constant; this was accomplished by rotating the slope around the bivariate mean. In both studies, the response to selection was erratic but detectable, changing by ca. 0.5 - 1.0% per generation-- weaker (Stillwell and others 2016) and less consistent (Bolstad and others 2015; Stillwell and others 2016) on average than the response in studies that targeted the intercept. One possible explanation for this difference is that the slope of morphological scaling relationships may experience greater relative developmental or genetic constraint than does the intercept (Pélabon and others 2014; Schwenk and Wagner 2003; Voje and others 2013; Houle and others, this volume). This may be an unlikely explanation for at least three reasons. First, selection on relative trait size can sometimes change the slope indirectly (e.g., Robertson 1962; Weber 1990, and see Voje and others 2013), suggesting an absence of strong constraints. Second, developmental genetic manipulations alter the slope of trait-body scaling in *D. melanogaster* (e.g., Shingleton and Tang 2012; Tang and others 2011), suggesting slope evolution could occur among taxa through these evolutionarily conserved genetic pathways. Finally, the slopes of scaling relationships evolve among species over the longer term (e.g., Baker and Wilkinson 2001; Burkhardt and de la Motte 1985; Emlen and others 2005; McCullough and others 2015; Swallow and others 2005; see reviews in Emlen and Nijhout 2000; Frazier and others 2008; Voje and Hansen 2014; Voje and others 2013), providing evidence for the absence of absolute constraint. On the other hand, the mechanisms that regulate scaling relationship slope appear to affect absolute trait size in tandem (Emlen and others 2012; Shingleton and Tang 2012; Tang and others 2011; Warren and others 2013), and therefore may be less likely or unable to contribute to slope evolution without also affecting the bivariate mean. Consequently, a response to the selection regime used in these studies may require evolution of coordinated, condition (size)-dependent expression of different growth regulators; this level of developmental integration may constitute a constraint that would likely be difficult to overcome, and may help explain the low response to selection in these studies. However, it is important to note that the patterning mechanisms that determine intra-organ trait size (e.g., the size or location of wing veins relative to the size of the wing itself, the traits subject to selection in Bolstad and others 2015), may be distinct from those that determine inter-organ size (see Houle and others, this volume); thus, it is difficult to draw generalities about constraint from just two studies that are focused on such developmentally distinct morphological scaling relationships.

An alternative to the constraints hypothesis is that the relative lack of response resulted from the use of a selection regime that is ineffective when applied to the distribution of individual cryptic scaling relationships present in these populations. The model predicts that correlational selection will have only moderate efficacy when applied to a seesaw distribution of individual cryptic scaling relationships, and will generate an inconsistent response when applied to populations possessing a speedometer or broomstick distribution, depending on the direction of selection and method used to fit the scaling relationship (Dreyer and others 2016). Our data suggest that the distribution of wing-body size individual cryptic scaling relationships in the DGRP population is seesaw-like for males and speedometer-like and females (Figures 2A, B); if this distribution holds across populations, it may explain the observed moderate response to selection on the slope of this relationship (Stillwell and others 2016). While we do not know the pattern of individual scaling relationships for the traits from the population subject to selection in the Bolstad and others (2015) study (but see Houle and others, this volume), similar distribution effects could also explain their results.

More generally, the pattern of underlying individual scaling relationships, and the range of trait sizes expressed along those relationships, may strongly influence the response to selection. In nearly all the artificial selection experiments reviewed above, individuals were reared under conditions designed to minimize stress and thereby maximize survival. This makes sense in the context of artificial selection experiments where every individual is valuable, where inbreeding can reduce lineage fitness during the course of the experiment (e.g., Emlen 1996), and where the value of the experimental lineages increases with each generation. But rearing individuals under uniform or low-stress conditions is unlikely to produce the full range of possible trait and body sizes; this has at least four implications for interpreting the results of artificial selection experiments. First, rearing individuals under benign conditions means that most will grow to be near their genetically-determined maximal size; a scaling relationship fit only to individuals at the largest portion of the potential size distribution may differ from, or even misrepresent, the static allometry fit across the full range of body size (Figure 5A). It follows that fitting a scaling relationship to only the largest individuals could also adversely affect estimations of the pattern of individual cryptic scaling relationships, the location of the MPI, etc. (Figure 5B). Second and related, without environmental manipulation, related individuals will experience the same environmental conditions, and may therefore tend to cluster together in morphospace. Consequently, some sibships may be over- or under-represented in the selected group – not

only because of their genotype, but also because the common developmental environment produces similar phenotypes among related individuals (Figure 5A, 5B). In the context of artificial selection experiments, this could lead to inbreeding or otherwise alter the response to selection. Third, under conditions where there is little environmentally-induced variation in body size, selection is likely to focus more on the genetic static allometry than on the environmental static allometry (Shingleton and others 2007) of the group. Selection on genetic static allometries may target different developmental mechanisms than selection applied to environmental static allometries, and this in turn could affect the response to selection. Finally, the range of expressed trait sizes relative to the MPI can determine the pattern of individual scaling relationships that are exposed to selection. This means that artificial selection experiments that employ only a single developmental environment may draw individuals from different subsets of the individual cryptic scaling relationships as compared to designs that include environmental manipulations that affect size (Dreyer and others 2016). For example, the pattern of individual scaling relationships across the full range of potential trait sizes may be a seesaw, but a population of well-fed (i.e., large) individuals may occupy only the speedometer-like fan of this distribution (Figure 5C). Such environmentally-induced variation in size distributions would expose different portions of the pattern of individual scaling relationships among populations, which in turn could produce different responses to the same pattern of selection across experiments.

The potential impact of the rearing environment on the response to selection merits further consideration. In their experiment designed to rotate the slope about the bivariate mean, Stillwell and others (2016) applied a developmentally-timed starvation treatment to split cohorts to ensure that each selected lineage expressed the full range of wing and body size each generation, and to make certain that progeny from each cross were represented throughout the size distribution. In contrast, Bolstad and others (2015) observed a response to selection on the slope only for populations reared on full diets; no response to selection was detected when they applied starvation treatments to their populations. This difference in response between the Bolstad and others (2015) starved- and fully-fed companion experiments is challenging to reconcile, but it supports the hypothesis that changing the range of trait sizes expressed in the population may alter the portion of the individual cryptic scaling relationships exposed to selection. For example, if the distribution of individual cryptic scaling relationships in Bolstad and others (2015) follows the broomstick pattern, then starved individuals will exhibit smaller trait sizes and higher variance in relative trait size, relative to fully-fed individuals, which will tend to

be larger and express lower variation in relative trait size (Figure 1C); the pattern of observed trait means, variances and covariances will be affected differently across developmental environments under other distributions of individual scaling relationships. Regardless of the distribution of individual scaling relationships in Bolstad and others (2015) study population, their observations emphasize the importance of considering the size range of traits in the population subject to selection, the effect that these distributions have on which genotypes – that is which individual cryptic scaling relationships – are being selected, and how these might influence evolutionary outcomes and the interpretation of results (Figure 5C).

By definition, genotype by environment interactions mean that the phenotype expressed by each individual represents a single realization of all the possible trait values for that genotype across some environmental gradient. In most of the artificial selection studies reviewed above, populations were subject to mass selection to change the population-level static allometry based on the observed phenotype of a single individual for each genotype. Investigators selected these individuals based on their location in morphospace and the patterns of (co)variance exhibited by the traits in the population. This means that implicitly or explicitly, investigators made assumptions regarding how individuals might contribute to the response to selection. In effect, assumptions were made regarding the selected individual's cryptic scaling relationship, and by extension, how the individual cryptic scaling relationships are distributed in the population. Our data suggest that the pattern of individual scaling relationships may vary among groups in ways that the Dreyer and others (2016) model predicts will affect the response of the population-level static allometry to selection (Figure 5).

In some sense, these issues surrounding the response to selection are reflected in family-based versus mass-selection artificial selection approaches. Recently, Stewart and Rice (2018) described a large, mass-selection design to alter sexual size dimorphism. Sexual size dimorphism, like morphological scaling relationships, cannot be estimated for individuals. In their experiment, Stewart and Rice (2018) did not observe any evolutionary change in the pattern of sexual size dimorphism until after 100 generations of individual-based selection. In contrast, a family-based artificial selection experiment produced rapid divergence in sexual size dimorphism in just 15 generations (Bird and Schaffer 1972). In the latter case, the “cryptic” phenotype was the sexual size dimorphism function for each genotype (family), and using this as the basis of selection enabled much more rapid evolution of the trait relative to selection based on individual-level phenotypes. In the context of scaling relationship evolution, this

suggests that making predictions regarding the response to selection based on the distribution of individual cryptic scaling relationships may be more accurate than using the patterns of trait size (co)variance. Consequently, selecting individuals based on the underlying function of their individual cryptic scaling relationships may prove more profitable than selecting based their location in morphospace.

Areas for Future Work

Artificial selection has proven a useful tool for exploring how morphological scaling relationships evolve. An emerging, formal focus on variation among individuals in the mechanisms that regulate trait and body growth have offered new insights regarding the expression and evolution of morphological scaling. In this context, below we outline pressing areas that need to be addressed to move forward our understanding of how morphological scaling relationships evolve.

I: Describe the distributions of individual cryptic scaling relationships. In developing this study, we had hoped to use existing data to describe the distribution of individual cryptic scaling relationships for a few systems where selection had been used to alter morphological scaling relationship parameters, and to employ these to determine if the distributions explained aspects of the response to selection. We were unable to do so, because those data simply do not exist. For many study systems, full sib groups can be used as proxy for genotype to estimate individual (sibship) cryptic scaling relationships in a manner similar to how split sibships can be used to estimate plastic responses (e.g. Blanckenhorn 1998; Newman 1988). However, the requirements for such datasets to be useful are difficult to meet. To be most effective, such studies require enough observations within sib groups to ensure a robust estimation of the individual (sibship) cryptic scaling relationships. Moreover, sib groups should be split among diets or other relevant developmental environments to produce the full range of size variation for each trait, as this will enable categorization of the individual scaling relationship distribution (i.e., seesaw, broomstick, or speedometer). Finally, a large enough number of sib groups is needed to provide sufficient power to detect variation among cryptic scaling relationship intercepts and slopes, and to allow estimation of the mean intersection point among individual scaling relationships.

Existing data that might have been used to complement studies of selection on scaling relationship parameters lacked sufficient within-sibship sample sizes or did not include

environmental treatments to produce a range of trait sizes. For example, Vega-Trejo and others (2018) generated the kinds of data that could be used to generate individual cryptic scaling relationships for the population subject to selection on relative gonopodium length in mosquitofish (Booksmythe and others 2015); while the numbers of sibships in Vega-Trejo and others (2018) study is impressive, and sib groups were split between diets that generated variation in adult size, there are too few observations within sibships to allow robust fitting of their individual (sibship) cryptic scaling relationships. Unfortunately, such limitations exist for every published and unpublished dataset we examined for the current study, primarily because the experiments that generated these data were designed to address other topics. Thus, experiments must be designed explicitly to address fundamental questions regarding the distribution of individual cryptic scaling relationships within populations, the degree to which different traits share distributions within populations, and how these patterns vary among populations, species or other biological groups.

II. Formally test the model predictions. Once distributions of individual cryptic scaling relationships are known for a population, artificial selection can be applied to test specific predictions of the Dreyer and others (2016) model. Such tests include determining how the distribution of individual cryptic scaling relationships affects scaling relationship evolution; to do this, populations that differ in their distributions for the same trait pairs can be subjected to the same pattern of selection (Figure 5 D-F). Tests also include determining how easily scaling relationship slopes evolve under patterns of univariate versus multivariate selection. For the reasons outlined above, a family-based approach (e.g., Bird and Schaffer 1972) could enable estimation of the distribution of individual (sibship) cryptic scaling relationships and a family-based selection approach may prove profitable. It may seem critical to use a split-cohort design in any such experiment, where individuals from each cross are reared under different developmental environments to ensure that progeny from all crosses are represented across the full range of trait sizes, and can therefore potentially contribute to the response selection (Figure 5). However, consideration of environmental effects on size and how this affects the response to selection may benefit from a more nuanced approach, as outlined below.

III. Determine the role of environmental variation in shaping the response to selection. The distribution of size phenotypes expressed in natural populations may not span the full range of possible expression or could be skewed, and this could have strong effects on the response of the static allometry parameters to selection. Such possibilities can be explored empirically in a

three-step program. First, the trait size distributions should be assessed in natural, focal populations; this will provide context for the remaining work. Second, using the methodology described above, the distribution of individual cryptic scaling relationships should be determined for these populations. Finally, artificial selection can be applied to determine if the response to the same pattern of selection holds when applied to different subsets of the full individual cryptic scaling relationships distributions. The results of Bolstad and others (2015) suggest that the response may not be the same for populations that differ in the pattern or range of phenotypic variation they express. It is important to elucidate why this is so. The expected response of the population-level static allometry to selection will likely differ if populations express only a partial range of body and trait size, or if selection sees only a limited or biased portion of the size distribution - thereby effectively converting one distribution of the individual cryptic scaling relationships into another (e.g., turning the seesaw distribution into the speedometer; Figure 5C).

In conclusion, by focusing attention on the patterns of cryptic individual-level variation that underlies observed population-level scaling relationships, the framework we present provides a different approach to study the expression and evolution of morphological scaling. The emerging models that use this framework suggest wide ranging, counterintuitive, and important insights. Evaluating the validity of these insights is an empirical task, and applying selection to carefully constructed artificial populations where the distribution of individual scaling relationships has been defined is likely to be a key tool in this undertaking.

Acknowledgements

We would like to thank the organizers of the symposium for inviting our participation, and D. Houle and an anonymous reviewer for comments on an earlier version of this manuscript. Bas Zwaan, Doug Emlen, and Michael Jennions all kindly provided data that ultimately could not be used in our review. This work was supported by grants from the National Science Foundation to WAF (IOS-1558098), AWS (IOS-1557638), and GSW (DEB-9409369, DEB-9807937). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Atkinson D. 1994. Temperature and organism size: A biological law for ectotherms? *Advances in Ecology Research* 25:1–58.

- Azevedo RBR, James AC, McCabe J, Partridge L. 1998. Latitudinal variation of wing: Thorax size ratio and wing-aspect ratio in *Drosophila melanogaster*. *Evolution* 52:1353-1362.
- Baker, R.H. and Wilkinson, G.S. 2001 Phylogenetic analysis of eye stalk allometry and sexual dimorphism in stalk-eyed flies (Diopsidae). *Evolution* 55: 1373-1385
- Bastock M, Manning A. 1955. The courtship of *Drosophila melanogaster*. *Behaviour* 8:85-111.
- Bennet-Clark HC, Ewing AW. 1968. The wing mechanism involved in the courtship of *Drosophila*. *Journal of Experimental Biology* 49:117-128.
- Biernaskie JM, Grafen A, Perry JC. 2014. The evolution of index signals to avoid the cost of dishonesty. *Proceedings of Royal Society B: Biological Sciences* 281:1-6.
- Bird MA, Schaffer HE. 1972. A study of the genetic basis of the sexual dimorphism for wing length in *Drosophila melanogaster*. *Genetics* 72:475-487.
- Blanckenhorn, WU. 1998. Adaptive phenotypic plasticity in growth, development, and body size in the yellow dung fly. *Evolution* 52: 1394-1407
- Bolstad GH, Cassara J, Marquez E, Hansen TF, van der Linde K, Houle D, Pelabon C. 2015. Complex constraints on allometry revealed by artificial selection on the wing of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America* 112: 13284-13289.
- Bonduriansky R, Day T. 2003. The evolution of static allometry in sexually selected traits. *Evolution* 57:2450-2458.
- Booksmythe I, Head ML, Keogh JS, Jennions MD. 2015. Fitness consequences of artificial selection on relative male genital size. *Nature Communications* 7:11597.
- Burkhardt D, de la Motte I. 1985. Selective pressures, variability, and sexual dimorphism in stalk-eyed flies (*Diopsidae*). *Naturwissenschaften* 72:204-206.
- Casasa, S, Moczek, AP. 2019. Evolution of, and via, developmental plasticity: Insights through the study of scaling relationships. *Integrative and Comparative Biology*.
- Cayetano L, Makalkov AA, Brooks RC, Bonduriansky R. 2011. Evolution of male and female genitalia following release from sexual selection. *Evolution* 65:2171-2183.
- Clutton-Brock J. 1992. The process of domestication. *Mammal Review* 22:79-85.
- Cooper, KL. 2019. Developmental and evolutionary allometry of the mammalian limb skeleton. *Integrative and Comparative Biology*.
- De Jong G, Bochdanovits Z. 2003. Latitudinal clines in *Drosophila melanogaster*: Body size, allozyme frequencies, inversion frequencies, and the insulin-signalling pathway. *Journal of Genetics* 82:207-223.
- Dreyer AP, Saleh Ziabari O, Swanson EM, Chawla A, Frankino WA, Shingleton AW. 2016. Cryptic individual scaling relationships and the evolution of morphological scaling. *Evolution* 70:1703-1716
- Dreyer AP, Shingleton AW. 2011. The effect of genetic and environmental variation on genital size in male *Drosophila*: Canalized but developmentally unstable. *Public Library of Science ONE* 6:e28278.
- Dudley R. 2002. The biomechanics of insect flight: Form, function, evolution. Princeton NJ: Princeton University Press.
- Eberhard WG. 2009. Static allometry and animal genitalia. *Evolution* 63:48-66.
- Egset CK, Hansen TF, Le Rouzic A, Bolstad GH, Rosenqvist G, Pelabon C. 2012. Artificial selection on allometry: Change in elevation but not slope. *Journal of Evolutionary Biology* 25:938-948.

- Emlen DJ. 1996. Artificial selection on horn length-body size allometry in the horned beetle *Onthophagus acuminatus* (Coleoptera: Scarabaeidae). *Evolution* 50:1219-1230.
- Emlen DJ, Allen CE. 2004. Genotype to phenotype: Physiological control of trait size and scaling in insects. *Integrative and Comparative Biology* 43:617-634.
- Emlen DJ, Marangelo J, Ball B, Cunningham CW. 2005. Diversity in the weapons of sexual selection: Horn evolution in the beetle genus *Onthophagus* (Coleoptera: Scarabaeidae). *Evolution* 59:1060-1084.
- Emlen DJ, Nijhout HF. 2000. The development and evolution of exaggerated morphologies in insects. *Annual Review of Entomology* 45:661-708.
- Emlen DJ, Warren IA, Johns A, Dworkin I, Lavine LC. 2012. A mechanism of extreme growth and reliable signaling in sexually selected ornaments and weapons. *Science* 337:860-864.
- Ewing AW. 1961. Body size and courtship behaviour in *Drosophila melanogaster*. *Animal Behaviour* 9:93-99.
- Ewing AW. 1964. The influence of wing area on the courtship behaviour of *Drosophila melanogaster*. *Animal Behaviour* 12:316-320.
- Frankino WA, Shingleton A, Emlen D. 2009. Experimental approaches to studying the evolution of morphological allometries: The shape of things to come. In: Garland T, Rose M, editors. *Experimental Evolution: Concepts, Methods, and Applications*. University of California Press. p. 419-478.
- Frankino WA, Zwaan BJ, Stern DL, Brakefield PM. 2007. Internal and external constraints in the evolution of morphological allometries in a butterfly. *Evolution* 61:2958-2970.
- Frankino WA, Zwaan BJ, Stern DL, Brakefield PM. 2005. Natural selection and developmental constraints in the evolution of allometries. *Science* 307:718-720.
- Frazier MR, Harrison JF, Kirkton SD, Roberts SP. 2008. Cold rearing improves cold-flight performance in *Drosophila* via changes in wing morphology. *Journal of Experimental Biology* 211:2116-2122.
- Frazier MR, Woods HA, Harrison JF. 2001. Interactive effects of rearing temperature and oxygen on the development of *Drosophila melanogaster*. *Physiological and Biochemical Zoology* 74:641-650.
- Fromhage L, Kokko H. 2014. Sexually selected traits evolve positive allometry when some matings occur irrespective of the trait. *Evolution* 68:1332-1338.
- Gilchrist GW, Huey RB. 2004. Plastic and genetic variation in wing loading as a function of temperature within and among parallel clines in *Drosophila subobscura*. *Integrative and Comparative Biology* 44:461-470.
- Gilchrist GW, Huey RB, Balanya J, Pascual M, Serra L. 2004. A time series of evolution in action: A latitudinal cline in wing size in South American *Drosophila subobscura*. *Evolution* 58:768-780.
- Gilchrist GW, Huey RB, Serra L. 2001. Rapid evolution of wing size clines in *Drosophila subobscura*. *Genetica* 112-113:273-286.
- Gokhale RH, Shingleton AW. 2015. Size control: The developmental physiology of body and organ size regulation. *Wiley Interdisciplinary Reviews: Developmental Biology* 4:335-356.
- Gould SJ. 1974. The origin and function of "bizarre" structures: Antler size and skull size in the "Irish Elk", *Megaloceros giganteus*. *Evolution* 28:191-220.
- Green A. 1992. Positive allometry is likely with mate choice, competitive display and other function. *Animal Behavior* 43:170-172.

- Hansen TF, Bartoszek K. 2012. Interpreting the evolutionary regression: The interplay between observational and biological errors in phylogenetic comparative studies. *Systematic Biology* 61:413-425.
- Harrison JF, Kaiser A, VandenBrooks JM. 2010. Atmospheric oxygen level and the evolution of insect body size. *Proceedings of the Royal Society B: Biological Sciences* 277:1937-1946.
- Harrison JF, Shingleton AW, Callier V. 2015. Stunted by developing in hypoxia: Linking comparative and model organism studies. *Physiological and Biochemical Zoology* 88:455-470.
- Houle D, Jones LT, Fortune R, Sztepanacz JL. 2019 Allometry, Scaling, and Ontogeny of Form. *Comparative and Integrative Biology*.
- Houle D, Pelabon C, Wagner GP, Hansen TF. 2011. Measurement and Meaning in Biology. *The Quarterly Review of Biology* 86:3-34.
- Huxley JS. 1932. *Problems of Relative Growth*. London: Methuen & Co. Ltd.
- Kodric-Brown A, Sibly RM, Brown JH. 2006. The allometry of ornaments and weapons. *Proceedings of the National Academy of Sciences of the United States of America* 103:8733-8738.
- Koh I, Lee MS, Lee NJ, Park KW, Kim KH, Kim H, Rhyu IJ. 2005. Body size effect on brain volume in Korean youth. *Neuroreport* 16:2029-32.
- Kotrschal A, Rogell B, Bundsen A, Immier S, Maklakov AA, Kolm N. 2013. Artificial selection on relative brain size in the guppy reveals costs and benefits of evolving a larger brain. *Current Biology* 23:168-171.
- Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* 33:402-416.
- Lane SJ, Frankino WA, Elekonich MM, Roberts SP. 2014. The effects of age and lifetime flight behavior on flight capacity in *Drosophila melanogaster*. *The Journal of Experimental Biology* 217:1437-1443.
- Larabee FJ, Gronenberg W, Suarez AV. 2017. Performance, morphology and control of power-amplified mandibles in the trap-jaw ant *Myrmoterias* (Hymenoptera: Formicidae). *Journal of Experimental Biology* 220:3062-3071.
- Larson G, Fuller DQ. 2014. The evolution of animal domestication. *Annual Review of Ecology, Evolution, and Systematics* 45:115-136.
- Liefting M, Hoffmann AA, Ellers J. 2009. Plasticity versus environmental canalization: Population differences in thermal responses along a latitudinal gradient in *Drosophila serrata*. *Evolution* 63:1954-1963.
- Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM et al. 2012. The *Drosophila melanogaster* genetic reference panel. *Nature* 482:173-178.
- McCullough EL, Ledger KJ, O'Brien DM, Emlen DJ. 2015. Variation in the allometry of exaggerated rhinoceros beetle horns. *Animal Behaviour* 109:133-140.
- Menezes BF, Vigoder FM, Peixoto AA, Varaldi J, Bitner-Mathéa BC. 2013. The influence of male wing shape on mating success in *Drosophila melanogaster*. *Animal Behaviour* 85:1217-1223.
- Mirth, CK, Frankino, WA, and Shingleton, AW. 2016. Allometry and Size Control: What can studies of body size regulation tell us about the evolution of morphological scaling relationships? *Current Opinion in Insect Science*. 13: 93–98

- Myers EM, Frankino WA. 2012. Time in a bottle: The evolutionary fate of species discrimination in sibling *Drosophila* species. *Public Library of Science ONE* 7:e31759.
- Newman, RA, 1988. Adaptive plasticity in development of *Scaphiopus couchii* tadpoles in desert ponds. *Evolution* 42:774-783.
- Nijhout HF, Riddiford LM, Mirth C, Shingleton AW, Suzuki Y, Callier V. 2014. The developmental control of size in insects. *Wiley Interdisciplinary Reviews: Developmental Biology* 3:113-134.
- Nijhout HF, Wheeler DE. 1996. Growth Models of Complex Allometries in Holometabolous Insects. *The American Naturalist* 148:40-56.
- Oldham S, Hafen E. 2003. Insulin/IGF and target of rapamycin signaling: A TOR de force in growth control. *Trends in Cell Biology* 13:79-85.
- Pélabon C, Firmat C, Bolstad GH, Voje KL, Houle D, Cassara J, Rouzic AL, Hansen TF. 2014. Evolution of morphological allometry. *Annals of the New York Academy of Sciences* 1320:58-75.
- Petrie M. 1992. Are all secondary sexual display structures positively allometric and, if so, why? *Animal Behavior* 43:173-175.
- Robertson FW. 1962. Changing the relative size of the body parts of *Drosophila* by selection. *Genetics Research* 3:169-180.
- Robertson FW, Reeve E. 1952. Studies in quantitative inheritance. *Journal of Genetics* 50:414-448.
- Schwenk K, Wagner GP. 2003. Constraint. In: Hall BK, Olson W, editors. *Keywords in Evolutionary Developmental Biology*. Cambridge MA: Harvard University Press. p. 52-61.
- Shingleton AW. 2019. Which line to follow? The utility of different line-fitting methods to capture the mechanism of morphological scaling. *Comparative and Integrative Biology*.
- Shingleton AW. 2011. The regulation and evolution of growth and body size. In: Flatt T, Heyland A, editors. *Mechanisms of Life History Evolution*. Oxford: OUP.
- Shingleton AW, Estep CM, Driscoll MV, Dworkin I. 2009. Many ways to be small: Different environmental regulators of size generate distinct scaling relationships in *Drosophila melanogaster*. *Proceedings of Royal Society B: Biological Sciences* 276:2625-2633.
- Shingleton AW, Frankino WA. 2018. The (ongoing) problem of relative growth. *Current Opinion in Insect Science* 25:9-19.
- Shingleton AW, Frankino WA. 2013. New perspectives on the evolution of exaggerated traits. *BioEssays* 35:100-107.
- Shingleton AW, Frankino WA, Flatt T, Nijhout HF, Emlen DJ. 2007. Size and shape: The developmental regulation of static allometry in insects. *BioEssays* 29:536-548.
- Shingleton AW, Mirth CK, Bates PW. 2008. Developmental model of static allometry in holometabolous insects. *Proceedings of Royal Society B: Biological Sciences* 275:1875-1885.
- Shingleton AW, Tang HY. 2012. Plastic flies: The regulation and evolution of trait variability in *Drosophila*. *Fly* 6:147-152.
- Stewart AD, Rice WD. 2018. Arrest of sex-specific adaptation during the evolution of sexual dimorphism in *Drosophila*. *Nature Ecology and Evolution* 2:1507-1513.
- Stillwell RC, Dworkin I, Shingleton AW, Frankino WA. 2011. Experimental manipulation of body size to estimate morphological scaling relationships in *Drosophila*. *Journal of Visualized Experiments* (56):e3162.

- Stillwell RC, Shingleton AW, Dworkin I, Frankino WA. 2016. Tipping the scales: Evolution of the allometric slope independent of average trait size. *Evolution* 70:433-444.
- Stinchcombe JR, Beder J, Carter PA, Gilchrist GW, Gervini D, Gomulkiewicz R, Hallgrímsson B, Heckman N, Houle D, Kingsolver JG et al. . 2012. Genetics and evolution of function-valued traits: Understanding environmentally responsive phenotypes. *Trends in Ecology & Evolution* 27:637-647.
- Swallow, JG, Wallace, LE, Christianson, SJ, Johns, PM and Wilkinson, GS. 2005 Genetic divergence does not predict change in ornament expression among populations of stalk-eyed flies. *Molecular Ecology* 14:3787-3800.
- Tang HY, Smith-Caldas MS, Driscoll MV, Salhadar S, Shingleton AW. 2011. FOXO regulates organ-specific phenotypic plasticity in *Drosophila*. *Public Library of Science Genetics* 7:e1002373.
- Tobler A, Nijhout HF. 2010. Developmental constraints on the evolution of wing-body allometry in *Manduca sexta*. *Evolution and Development* 12:592-600.
- Vega-Trejo R, Head ML, Jennions MD, Kruuk LEB. 2018. Maternal-by-environment but not genotype-by-environment interactions in a fish without parental care. *Heredity* 120:154-167.
- Voje KL, Hansen TF. 2014. Evolution of static allometries: Adaptive change in allometric slopes of eyespan in stalk-eyed flies. *Evolution* 67:453-467.
- Voje KL, Hansen TF, Egset CK, Bolstad GH, Pelabon C. 2013. Allometric constraints and the evolution of allometry. *Evolution* 68:866-885.
- Warren IA, Gotoh H, Dworkin IM, Emlen DJ, Lavine LC. 2013. A general mechanism for conditional expression of exaggerated sexually- selected traits. *BioEssays* 35:889–899.
- Weber KE. 1990. Selection on wing allometry in *Drosophila melanogaster*. *Genetics* 126:975-989.
- Wilkinson GS. 1993. Artificial sexual selection alters allometry in the stalk-eyed fly *Cyrtodiopsis dalmanni* (Diptera: Diopsidae). *Genetics Research* 62:213-222.
- Wilkinson GS, Dodson GN. 1997. Function and evolution of antlers and eye stalks in flies. In: Choe J, Crespi B, editors. *Evolution of Mating Systems in Insects and Arachnids*. Cambridge: Cambridge University Press. p. 310-328.
- Wilkinson GS, Reillo PR. 1994. Female choice response to artificial selection on an exaggerated male trait in a stalk-eyed fly. *Proceedings of the Royal Society: Biological Sciences* 255:1-6.
- Zeh DW, Zeh JA. 1988. Condition-dependent sex ornaments and field tests of sexual-selection theory. *The American Naturalist* 132:454–459.

Figure Captions

Figure 1. Relationship among individual cryptic scaling relationships, the population-level static allometry and the mean point of intersection. (A) Individual cryptic scaling relationships (thin lines) between two traits (t_1 and t_2 ; e.g., wing and body size respectively) for a set of genotypes. Variation in an environmental factor that affects growth (e.g., access to nutrition) determines which size phenotype (round markers) each genotype expresses on its individual cryptic scaling relationship. The population-level static allometry (thick line) is fit using all expressed phenotypes in the population. (B-D) Individual cryptic scaling relationships can have three distributions, depending on the location of their mean point of intersection (MPI; open circles) relative to the potentially expressed size phenotypes of individuals in the population: (B) If the MPI is at the lower end of the size range, the pattern is referred to as a 'speedometer', (C) if the MPI is in the higher end of the range of size, the pattern is called a 'broomstick', and (D) if it is near the center of the size range, the pattern is a 'seesaw'. In reality, natural populations likely exist along the continuum represented by these three idealized distributions, and a given distribution might exhibit variation in intersection points around the MPI (insets in panel D).

Figure 2: Individual cryptic scaling relationships, MPIs and size distributions for isogenic fly lineages. The individual wing- and pupal-size cryptic scaling relationships for 85 male (A) and 87 female (B) isogenic lineages of *Drosophila melanogaster*, fit using major axis regression. The white marker indicates the mean point of intercept (MPI) among all individual cryptic scaling relationships for that sex. (C-D) Density plots (contours) for the intercepts of all pairwise combinations of individual scaling relationships for (C) males and (D) females. The histograms show the distribution of wing and pupal sizes among low-starvation flies (d0; dark grey) and all (light grey) flies. The length of the individual scaling relationships (A and B) shows the range of observed trait sizes. The individual scaling relationships as plotted do not, therefore, allow accurate categorization of their pattern visually.

Figure 3: The relationship between the slope of the individual cryptic scaling relationship for males and females of the same genotype from the DGRP lineages. Each lineage is represented by the slope for females (filled circle) and males (open circles) connected by a line. Lineage slopes are sorted by value for females under (A) MA Type II regression and this lineage order is retained for the (B) OLS Type I regression slopes.

Figure 4: Inferring the pattern of the individual scaling relationships. (A) For speedometer distributions of individual cryptic scaling relationships (dashed lines) there should be a positive correlation between mean relative wing size for a lineage, based on the largest individuals (black lines), and the slope of the lineage's individual scaling relationship, whereas this relationship should be weaker or absent among the smallest individuals (grey lines). (B) For seesaw distributions, the largest individuals (black lines) should exhibit a positive correlation between mean relative wing size for a lineage and the slope of the lineage's individual scaling relationship (dashed lines), whereas the smallest individuals (grey lines) should have a negative correlation between mean relative wings size and slope. (C) For DGRP females, there is no significant relationship between mean relative wing size for a lineage, based on the smallest 25% of all individuals regardless of lineage, and the slope of the lineage's individual scaling relationship (Linear Model: slope = relative wing size, $R^2 = 0.0066$, $F_{(1,70)} = 10.59$, $P = 0.4969$), (D) whereas this relationship is negative for DGRP males (Linear Model: slope = relative wing size, $R^2 = 0.3617$, $F_{(1,69)} = 39.1$, $P < 0.0001$). For both females (E) and males (F), there is a significant positive relationship between mean relative wing size for a lineage, based on the largest 25% of all individuals regardless of lineage, and the slope of the lineage's individual scaling relationship (Linear Model: slope = relative wing size, Male: $R^2 = 0.1513$, $F_{(1,70)} = 12.48$, $P = 0.0007$; Female: $R^2 = 0.1242$, $F_{(1,75)} = 7.45$, $P = 0.0007$). Thus, the shading in panel A corresponds to that of the symbols in C and E, and the shading in panel B corresponds to that of the symbols in D and F.

Figure 5. Consequences of environmental effects on phenotypic variation. (A) Under benign environmental conditions, individuals from three genotypes (differently shaped symbols) will approach their genetically determined maximal size (bold symbols), which is larger than the size expressed in less benign environments (light symbols). A population-level static allometry fit only to individuals from the benign environment (heavy line) may therefore differ substantially from that fit to a population of individuals reared across all environments (light line). (B) Individual cryptic scaling relationships fit to these genotypes may also differ depending on whether they are fit to individuals reared in a single benign environment (solid lines), or to individuals reared across all environments (dashed lines). (C) The range of trait sizes expressed in a population will determine the realized pattern of individual cryptic scaling relationships and thereby affect how individuals are exposed to selection as well as the response to selection itself. Here, the pattern of individual scaling relationships is a seesaw across the full range of sizes (lines), but a speedometer for populations in benign environments where all individuals

are large (dark portion of lines). Selecting to rotate the slope of population scaling relationship around the bivariate mean (e.g. as in Stillwell and others (2016)), would select individuals from regions a and b^1 in the benign environment, but would select individuals from regions a and b^2 for populations expressing the full range of trait sizes. Individuals in regions b^1 and b^2 have very different individual scaling relationships, thus, the response to the same form of selection will be different under the different ranges of environmentally induced phenotypic variation. (D) A population of individuals (circles) and their population-level static allometry (line). Individuals in the shaded area would be selected under proportional selection to increase the intercept. (E) This shows the same individuals and pattern of selection as in the previous panel, but here the individual cryptic scaling relationships are plotted (grey lines), distributed as a speedometer. (F) Again, the same individuals and pattern of selection in the previous two panels is shown, however, the pattern of individual scaling relationships here is a seesaw. Note that the selected individuals (filled circles) in both E and F would contribute differently to evolution of the intercept because they differ in the pattern of their individual cryptic scaling relationships (dashed lines).

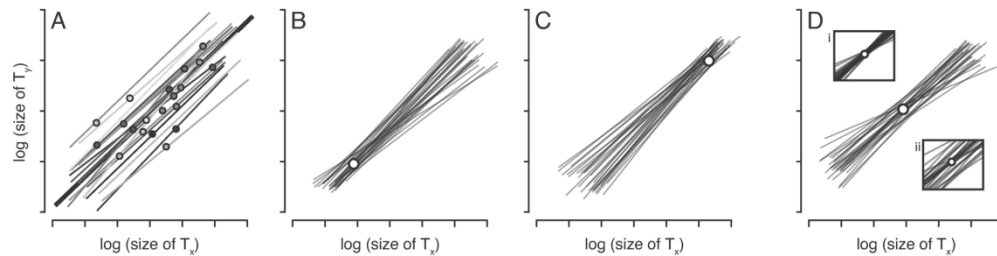


Figure 1. Relationship among individual cryptic scaling relationships, the population-level static allometry and the mean point of intersection. (A) Individual cryptic scaling relationships (thin lines) between two traits (t_1 and t_2 ; e.g., wing and body size respectively) for a set of genotypes. Variation in an environmental factor that affects growth (e.g., access to nutrition) determines which size phenotype (round markers) each genotype expresses on its individual cryptic scaling relationship. The population-level static allometry (thick line) is fit using all expressed phenotypes in the population. (B-D) Individual cryptic scaling relationships can have three distributions, depending on the location of their mean point of intersection (MPI; open circles) relative to the potentially expressed size phenotypes of individuals in the population: (B) If the MPI is at the lower end of the size range, the pattern is referred to as a 'speedometer', (C) if the MPI is in the higher end of the range of size, the pattern is called a 'broomstick', and (D) if it is near the center of the size range, the pattern is a 'seesaw'. In reality, natural populations likely exist along the continuum represented by these three idealized distributions, and a given distribution might exhibit variation in intersection points around the MPI (insets in panel D).

180x80mm (300 x 300 DPI)

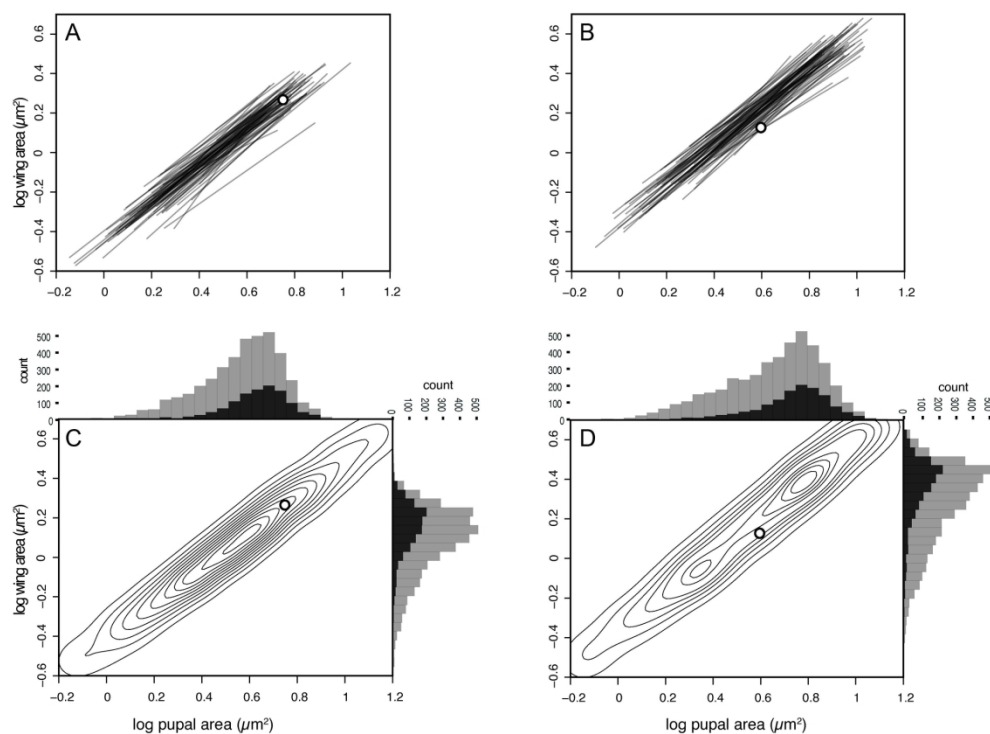


Figure 2: Individual cryptic scaling relationships, MPIs and size distributions for isogenic fly lineages. The individual wing- and pupal-size cryptic scaling relationships for 85 male (A) and 87 female (B) isogenic lineages of *Drosophila melanogaster*, fit using major axis regression. The white marker indicates the mean point of intercept (MPI) among all individual cryptic scaling relationships for that sex. (C-D) Density plots (contours) for the intercepts of all pairwise combinations of individual scaling relationships for (C) males and (D) females. The histograms show the distribution of wing and pupal sizes among low-starvation flies (d0; dark grey) and all (light grey) flies. The length of the individual scaling relationships (A and B) shows the range of observed trait sizes. The individual scaling relationships as plotted do not, therefore, allow accurate categorization of their pattern visually.

179x152mm (300 x 300 DPI)

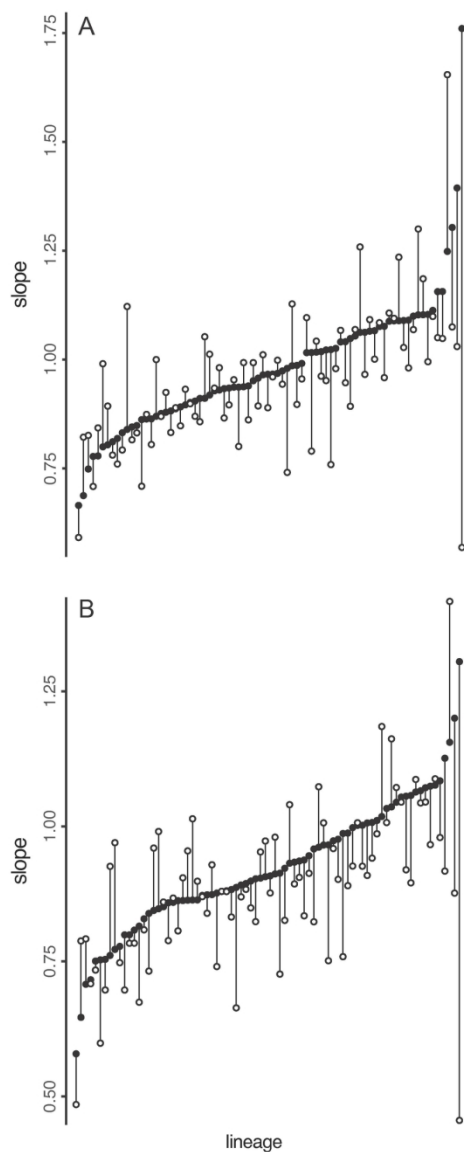


Figure 3: The relationship between the slope of the individual cryptic scaling relationship for males and females of the same genotype from the DGRP lineages. Each lineage is represented by the slope for females (filled circle) and males (open circles) connected by a line. Lineage slopes are sorted by value for females under (A) MA Type II regression and this lineage order is retained for the (B) OLS Type I regression slopes.

86x203mm (300 x 300 DPI)

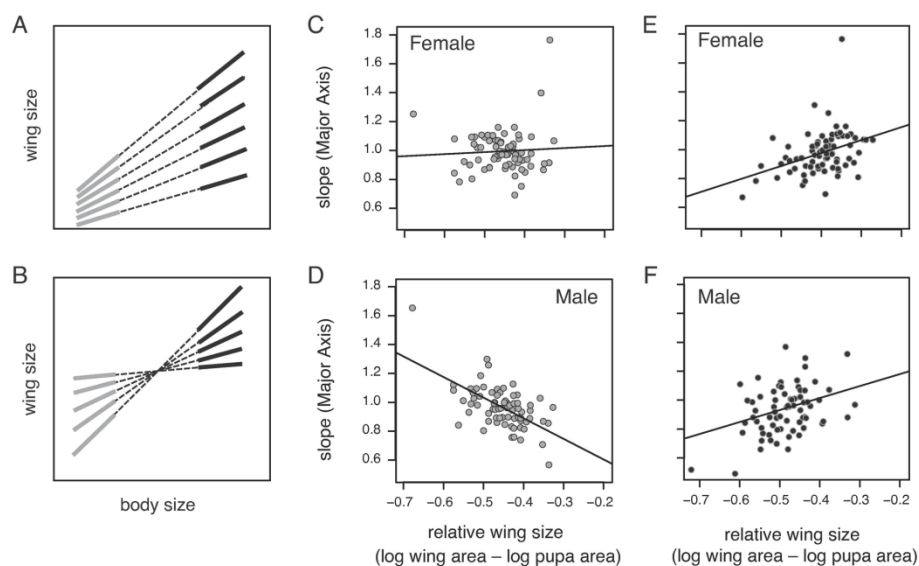


Figure 4: Inferring the pattern of the individual scaling relationships. (A) For speedometer distributions of individual cryptic scaling relationships (dashed lines) there should be a positive correlation between mean relative wing size for a lineage, based on the largest individuals (black lines), and the slope of the lineage's individual scaling relationship, whereas this relationship should be weaker or absent among the smallest individuals (grey lines). (B) For seesaw distributions, the largest individuals (black lines) should exhibit a positive correlation between mean relative wing size for a lineage and the slope of the lineage's individual scaling relationship (dashed lines), whereas the smallest individuals (grey lines) should have a negative correlation between mean relative wings size and slope. (C) For DGRP females, there is no significant relationship between mean relative wing size for a lineage, based on the smallest 25% of all individuals regardless of lineage, and the slope of the lineage's individual scaling relationship (Linear Model: slope = relative wing size, $R^2 = 0.0066$, $F(1,70) = 10.59$, $P = 0.4969$), (D) whereas this relationship is negative for DGRP males (Linear Model: slope = relative wing size, $R^2 = 0.3617$, $F(1,69) = 39.1$, $P < 0.0001$). For both females (E) and males (F), there is a significant positive relationship between mean relative wing size for a lineage, based on the largest 25% of all individuals regardless of lineage, and the slope of the lineage's individual scaling relationship (Linear Model: slope = relative wing size, Male: $R^2 = 0.1513$, $F(1,70) = 12.48$, $P = 0.0007$; Female: $R^2 = 0.1242$, $F(1,75) = 7.45$, $P = 0.0007$). Thus, the shading in panel A corresponds to that of the symbols in C and E, and the shading in panel B corresponds to that of the symbols in D and F.

177x114mm (300 x 300 DPI)

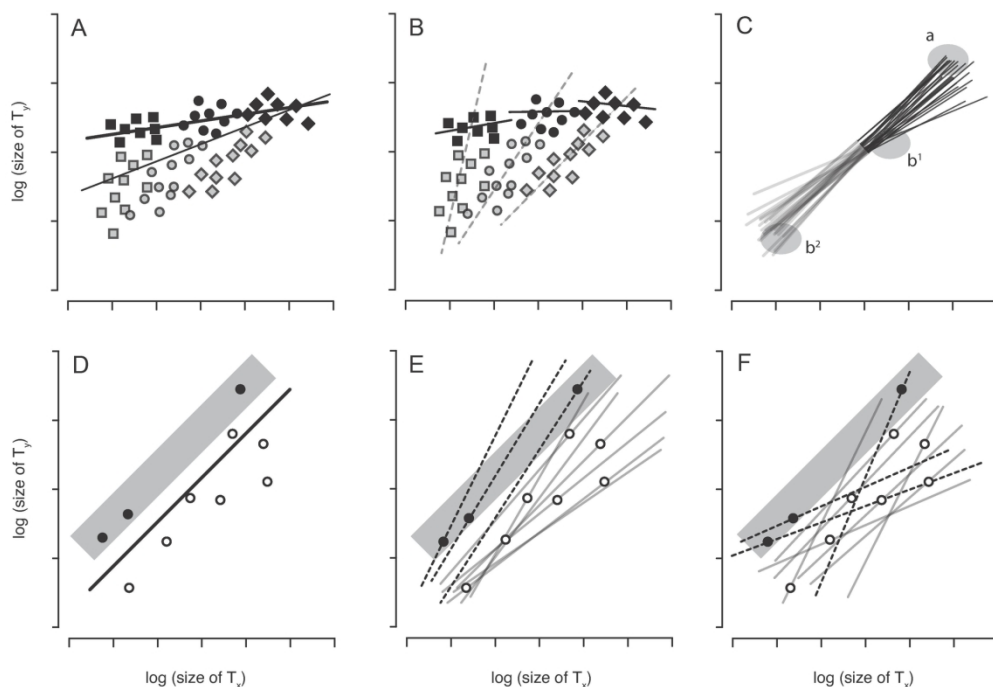


Figure 5. Consequences of environmental effects on phenotypic variation. (A) Under benign environmental conditions, individuals from three genotypes (differently shaped symbols) will approach their genetically determined maximal size (bold symbols), which is larger than the size expressed in less benign environments (light symbols). A population-level static allometry fit only to individuals from the benign environment (heavy line) may therefore differ substantially from that fit to a population of individuals reared across all environments (light line). (B) Individual cryptic scaling relationships fit to these genotypes may also differ depending on whether they are fit to individuals reared in a single benign environment (solid lines), or to individuals reared across all environments (dashed lines). (C) The range of trait sizes expressed in a population will determine the realized pattern of individual cryptic scaling relationships and thereby affect how individuals are exposed to selection as well as the response to selection itself. Here, the pattern of individual scaling relationships is a seesaw across the full range of sizes (lines), but a speedometer for populations in benign environments where all individuals are large (dark portion of lines). Selecting to rotate the slope of population scaling relationship around the bivariate mean (e.g. as in Stillwell and others (2016)), would select individuals from regions a and b1 in the benign environment, but would select individuals from regions a and b2 for populations expressing the full range of trait sizes. Individuals in regions b1 and b2 have very different individual scaling relationships, thus, the response to the same form of selection will be different under the different ranges of environmentally induced phenotypic variation. (D) A population of individuals (circles) and their population-level static allometry (line). Individuals in the shaded area would be selected under proportional selection to increase the intercept. (E) This shows the same individuals and pattern of selection as in the previous panel, but here the individual cryptic scaling relationships are plotted (grey lines), distributed as a speedometer. (F) Again, the same individuals and pattern of selection in the previous two panels is shown, however, the pattern of individual scaling relationships here is a seesaw. Note that the selected individuals (filled circles) in both E and F would contribute differently to evolution of the intercept because they differ in the pattern of their individual cryptic scaling relationships (dashed lines).

177x127mm (300 x 300 DPI)

Table 1: Effect of pupal case size and sex on wing size

INDEPENDENT ¹	SS	MS	DF ^{NUM}	DF ^{Den}	F	P
sex	1.9333	1.933	1	7774.2	508.13	<0.0001
pupa	29.0883	29.0883	1	68.1	7645.50	<0.0001
sex:pupa	0.0528	0.0528	1	7776.9	13.87	0.0002

¹ Effect based on the linear mixed model $W_{ijkl} = S_i + P_j + S_i P_j + B_k + P_i L_l + \varepsilon_{ijkl}$ where W is wing size, S is sex, P is pupal case size, B is block (random effect), L is lineage (random effect) and ε is normally-distributed error. Model was fit using *lme4* package in *R*.

Table 2: Effect of including random variation in slope on the fit of the relationship between wing and pupal case size in males and females

SEX	MODEL ¹	DF ²	AIC ³	BIC ³	LIKELIHOOD RATIO ³	P ⁴	Variance of slope among lineages ⁵
male	Model 1: $W_{ijkl} = P_j + B_k + P_i \cdot L_l$	7	-10068.47	-10024.90	68.63	0.0010	0.0075 (0.0033-0.0117)
	Model 2: $W_{ijkl} = P_j + B_k + L_l$	5	-10003.84	-9972.72			
female	Model 1: $W_{ijkl} = P_j + B_k + P_i \cdot L_l$	7	-10860.59	-10816.30	69.98	0.0010	0.0061 (0.0029-0.094)
	Model 2: $W_{ijkl} = P_j + B_k + L_l$	5	-10794.61	-10762.98			

¹ W is wing size, P is pupal case size, B is block (random effect), L is lineage (random effect). The models differ by having random slopes and intercepts (model 1) versus random intercepts (model 2) among lineages.

² Estimated degrees of freedom for each model.

³ AIC, BIC and Likelihood ratio calculated using ML fit.

⁴ P value calculated by parametric bootstrapping using ML fit.

⁵ Variance of slope among lineages calculated using Bayesian fit, with 95% confidence interval.

Table 3: Hotelling's T comparing MPI for the wing-pupa individual scaling relationships to the bivariate mean wing-pupa size for unstarved flies and all flies.

	Model used to fit Individual Scaling Relationships	Mean Point of Intersection	Bivariate mean of unstarved flies	Bivariate mean of all flies
			(0.63 _{pupa} , 0.15 _{wing}) SE: 0.004 _{upa} 0.003 _{wing}	(0.52 _{pupa} , 0.04 _{wing}) SE: 0.003 _{upa} 0.003 _{wing}
Male	MA (Model II)	(0.75 _{pupa} , 0.26 _{wing}) SE: 0.48 _{pupa} 0.43 _{wing}	$T^2 = 0.044$ $F_{2,7556} = 0.074$, $P = 0.963$	$T^2 = 0.956$ $F_{2,10208} = 0.478$ $P = 0.620$
	OLS (Model I)	(0.58 _{pupa} , 0.12 _{wing}) SE: 0.26 _{pupa} 0.25 _{wing}	$T^2 = 0.141$ $F_{2,7656} = 0.071$ $P = 0.932$	$T^2 = 0.595$ $F_{2,10208} = 0.298$, $P = 0.742$
			(0.74 _{pupa} , 0.36 _{wing}) SE: 0.004 _{upa} 0.004 _{wing}	(0.61 _{pupa} , 0.23 _{wing}) SE: 0.003 _{upa} 0.003 _{wing}
Female	MA (Model II)	(0.59 _{pupa} , 0.14 _{wing}) SE: 0.21 _{pupa} 0.22 _{wing}	$T^2 = 2.139$ $F_{2,8758} = 1.070$ $P = 0.343$	$T^2 = 6.736$ $F_{2,10837} = 3.368$ $P = 0.035$
	OLS (Model I)	(0.17 _{pupa} , -0.13 _{wing}) SE: 0.24 _{pupa} 0.22 _{wing}	$T^2 = 1.082$ $F_{2,8758} = 0.541$ $P = 0.582$	$T^2 = 1.950$ $F_{2,11628} = 0.975$ $P = 0.377$

¹ All units are log(mm²)

² Significant difference is shown in bold