Integrative and Comparative Biology

Integrative and Comparative Biology, volume 59, number 5, pp. 1429–1440
doi:10.1093/icb/icz099

Society for Integrative and Comparative Biology

SYMPOSIUM

Why does allometry evolve so slowly?

David Houle,[1] Luke T. Jones, Ryan Fortune and Jacqueline L. Sztepanacz

Department of Biology, Florida State University, Tallahassee, FL, USA

From the symposium "Allometry, Scaling and Ontogeny of Form" presented at the annual meeting of the Society for Integrative and Comparative Biology, January 3–7, 2019 at Tampa, Florida.

[1]E-mail: dhoule@bio.fsu.edu

Synopsis Morphological allometry is striking due to its evolutionary conservatism, making it an example of a certain sort of evolutionary stasis. Organisms that vary in size, whether for developmental, environmental, or evolutionary reasons, adopt shapes that are predictable from that size alone. There are two major hypotheses to explain this. It may be that natural selection strongly favors each allometric pattern, or that organisms lack the development and genetic capacity to produce variant shapes for selection to act on. Using a high-throughput system for measuring the size and shape of *Drosophila* wings, we documented an allometric pattern that has been virtually unchanged for 40 million years. We performed an artificial selection experiment on the static allometric slope within one species. In just 26 generations, we were able to increase the slope from 1.1 to 1.4, and decrease it to 0.8. Once artificial selection was suspended, the slope rapidly evolved back to a value near the initial static slope. This result decisively rules out the hypothesis that allometry is preserved due to a lack of genetic variation, and provides evidence that natural selection acts to maintain allometric relationships. On the other hand, it seems implausible that selection on allometry in the wing alone could be sufficiently strong to maintain static allometries over millions of years. This suggests that a potential explanation for stasis is selection on a potentially large number of pleiotropic effects. This seems likely in the case of allometry, as the sizes of all parts of the body may be altered when the allometric slope of one body part is changed. Unfortunately, hypotheses about pleiotropy have been very difficult to test. We lay out an approach to begin the systematic study of pleiotropic effects using genetic manipulations and high-throughput phenotyping.

Introduction

The study of the evolution of allometry has historically been dominated by comparative and theoretical approaches that focused on understanding why allometric slopes evolve so slowly. In the last few years, many groups have started to apply genetic tools to understand the proximal and ultimate causes of allometry. The increased use of genetic and physiological manipulations to alter allometry within species (Truman et al. 2006; Shingleton et al. 2009; Tang et al. 2011; Testa and Dworkin 2016), has led to a natural focus on those processes that are or could bring about evolved changes in allometry. There are now several biological systems in which the genetic bases of trans-specific changes in allometry are being unraveled, as described in other contributions to this symposium. These genetic tools have been applied less frequently to understand stasis in allometry. Here we focus on the potential reasons why allometry evolves so slowly, and not on the genetic, developmental, or physiological machinery that allows it to evolve.

Many students of allometry have emphasized the relative constancy of allometric slope in contrast to the intercept (Huxley 1932; Cock 1966; Gould 1971; Voje et al. 2014). As a consequence, many have favored the general hypothesis that allometry constrains long-term evolutionary change. More specifically, some have hypothesized that the covariance among traits inherent in allometric relationships drives evolutionary trends in traits that are not themselves under strong selection (e.g., Simpson 1944; Rensch 1959; Gould and Lewontin 1979; Lande 1979). In the extreme case where allometry is imagined to be incapable of evolving, these correlated responses have been offered as explanations for the presumed maladaptive state of traits. The antlers of 'Irish elk' are the best known example

Advance Access publication June 14, 2019

of this conflation of maladaptation and allometry (Simpson 1953, 286–7; reviewed in Gould 1974). A recent review of allometry among individuals at the same developmental stage (static allometry), provides quantitative evidence for relative stasis. It found little evidence that static allometric slopes evolve on contemporary timescales, but found evidence for modest changes on million-year time scales (Voje et al 2014). They showed that 11% of the interspecific variation in static allometry of morphological traits could be attributed to changes in allometric slope in six taxa. Changes in mean size or overall shape (the intercept of the allometric equation) accounted for the majority of the rest of the interspecific variation. These results show that allometric slopes can evolve in natural populations, but do so more slowly than non-allometric changes. Given this background we use the term 'allometry' only in reference to the allometric slope.

One example of this pattern that we will focus on here is the conservatism of the relationship between the length of wing vein L2 and wing size in the family Drosophilidae (Bolstad et al. 2015). Bolstad et al. (2015) evaluated the static and evolutionary allometry (allometry among species means) of this trait combination in 111 species of Drosophilid flies, most of which were in the clade defined by the polyphyletic genus *Drosophila*. Static allometric slopes are remarkably similar among species, considering that the family is at least 40 million years old. More precisely, the rate of evolution of slope and intercept on the phylogeny is just 3% per million years for both slope and intercept.

The evolutionary conservatism of allometry is a striking case of a more general pattern of evolutionary stasis that is observed across a variety of traits and geological timescales. The causes of evolutionary stasis can be separated into mechanisms of constraint and selection. Determining which of these mechanisms underlies evolutionary stasis is a central goal of evolutionary biologists. Here we pose the question: is there something fundamental about the genetic basis of allometric slope that leads to its evolutionary conservatism? To begin to answer this question we draw on three sets of data from the fruit fly, *Drosophila melanogaster*—a selection experiment that sought to change the allometric slope of wing vein L2, and unpublished datasets that estimate genetic variation in appendage form, and that characterize the effects of targeted manipulations of those forms. First, we document the pattern of genetic variation in allometry for fly wings using unpublished data from our lab. Second, we discuss the results of our artificial selection experiments on allometry (Bolstad et al. 2015) in the context of genetic and selective constraints. Finally, we demonstrate how the manipulation of gene expression can be used to reveal the genetic and pleiotropic architecture that underlies allometry.

Constraints on the evolution of allometry

Explanations for the evolutionary stasis of allometry fall on a continuum that, at one extreme, entirely results from genetic or developmental constraints, and at the other extreme, entirely results from selection. As discussed above, allometry is often assumed to be subject to genetic or developmental constraints. The key prediction of the pure constraint hypothesis is that there is effectively no genetic variation to enable an evolutionary response in slope. Most work on the evolution of allometry is based on the standard static allometric equation $\log(Y) = a_s + \beta_s \times \log(X - X_{\text{int}})$, where X is some measure of size, Y is the trait allometrically related to size, a_s is the value of Y at the intercept value of size, X_{int}, and β_s is the static allometric slope. Allometric relationships are generally well fit by this linear equation, so most work on the evolution of allometry reifies the terms in this model, and think in terms of genetic variation and covariation in its three parameters, size, X, slope, β_s and intercept, a_s. There are a number of different ways that slope and intercept can (co)vary, which we illustrate in Fig. 1.

Unfortunately, there are few studies that have appropriately estimated the additive genetic variance of allometric slope on a log scale and in a quantitative genetic framework (Pélabon et al 2014), leaving us with few parameter estimates for genetic variation in slope. Pavlicev et al. (2011) used this approach to estimate the heritability of allometry between body size and four long-bone traits and four internal organ weight traits in inbred mouse lines. They found the heritability of allometric slope to range from 0.068 to 0.195 among traits, which is on the low end of heritability for most quantitative traits.

We have estimated genetic variation in allometry ourselves, by reanalyzing data from of a recently published a broad-sense genetic variance (**G**) matrix for wing shape and size (Pitchers et al. 2019) for a population of inbred lines from the *Drosophila* Genetic Reference Panel (DGRP) (Mackay et al. 2012). We reanalyzed these data for the Huxley allometric parameters for the relationship between the length of wing veins and the square root of wing size. To estimate this **G** matrix, we took advantage

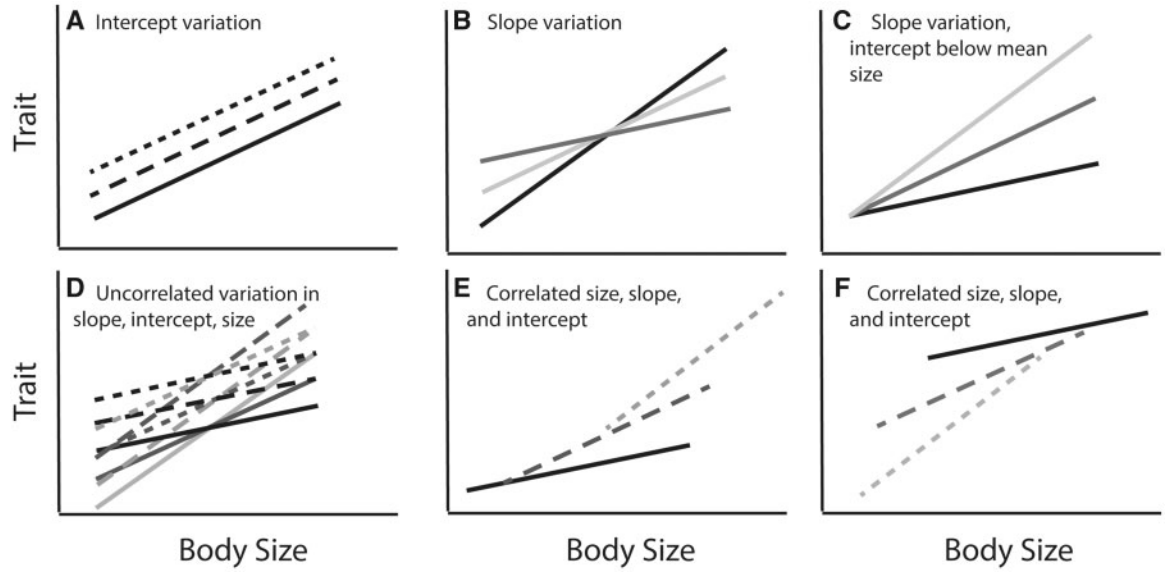

Fig. 1 Variation assuming that the parameters of the Huxley growth model cause variation in trait value. (**A**) Variation in intercept, no variation in slope. (**B**) Variation in slope with an intercept fixed at the population mean. (**C**) Variation in slope with an intercept fixed below the population mean. (**D**). Uncorrelated variation in both slope and intercept, with the intercept fixed at the population mean. (**E**, **F**). Variation with correlated size, slope, and intercept parameters.

of the fact that all flies within each DGRP line are extremely closely related, so the allometric relationship achieved among a cohort of flies directly estimates the allometry of that genotype. In total, we measured more than 24,000 flies from 190 inbred lines, in up to four experimental blocks in two different labs. For each one of the 510 line–block combinations, we estimated the slope and intercept for four different long veins, as well as mean wing area. These were then analyzed using a mixed model using restricted maximum likelihood in the program Wombat (Meyer 2006–2018; Meyer 2007). We fit the genetic component by assuming that all DGRP lines were unrelated, and fit a fixed effect for lab.

The **G** matrix for log L2 length and wing size using the allometric model is shown in Table 1. Results for the lengths of the other three long veins on the fly wing are similar (data not shown). There is ample genetic variation in slope intercept and mean wing size. Genetic correlations among these three traits are low. The correlations of slope with size and intercept are not significantly different from 0, but there is a modest negative correlation between intercept and wing size. Figure 2 represents a sample of the most extreme slopes and intercepts. Slopes tend to cross across the broad middle of the size distribution. Despite the substantial genetic variance in slope relative to intercept, the variation in slope is too small to have much impact on the spread of the data. The actual variation in allometric parameters resembles that in Fig. 1D, although the variation in slope is smaller relative to the variation in intercept. These results show that the change in residual variation with size is minimal, ruling out the "broomstick" and "speedometer" patterns that Dreyer et al. (2016) suggest are possible. An important caveat to our results is that the genetic variation we detected is among inbred genotypes, and does not necessarily reflect additive genetic variation that could respond to selection in an outbred population. The real test of the absolute constraint hypothesis is whether allometry can actually respond to selection.

Table 1 Genetic correlations, covariances, and variance ratios for the allometry between vein L2 length and wing size in DGRP inbred lines.

	Intercept	Slope	Mean wing size
Intercept	**0.90 ± 0.01***	0.03 ± 0.09	−0.31 ± 0.09*
	5.42 ± 0.59*		
Slope	0.32 ± 0.99	**0.44 ± 0.05***	−0.14 ± 0.09
		21.23 ± 3.42*	
Mean wing size			**0.38 ± 0.05***
	−1.96 ± 0.62*	−0.46 ± 1.45	7.03 ± 1.24*

Notes: Bold-faced values on the diagonal are the proportion of variation in the parameters that are explicable by line effects. Plain font values on the diagonal are line variances. Values below the diagonal are the genetic covariances, and those above the diagonal are the genetic correlations. Variances and covariances are multiplied by 10,000. All values shown ± approximate sampling errors. Sampling errors calculated using the REML-MVN method (Houle and Meyer 2015).
*Estimate significantly different from 0.

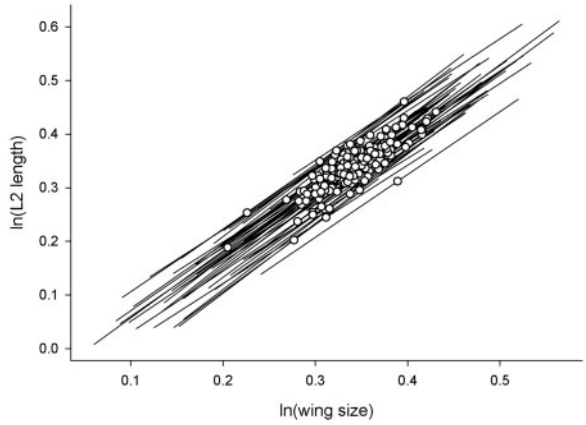

Fig. 2 Allometric relationships for the log of the length of vein L2 in relation to log wing size for 48 DGRP lines. Estimates are the least-squares means for line from a model with line and lab as factors. Estimates were calculated in SAS Proc GLM (SAS Institute, 2016). The 48 lines whose slopes are represented are those with the 10 most extreme estimates of slope, intercept and wing size in each direction. Open circles show the mean wing sizes and vein lengths at that mean size for all 190 DGRP genotypes.

Can allometry respond to selection?

A relatively simple way to test whether the evolution of allometry is constrained by a lack of additive genetic variation, is to use artificial selection to attempt to change it. For organisms that can be reared in controlled conditions, it is relatively easy to apply strong artificial selection to target traits, while suspending many aspects of natural selection. If a targeted trait can respond to artificial selection, then selectable genetic variation in the trait exists, and the constraint hypothesis can be rejected. Despite the conceptual simplicity of this approach, Egset et al. (2012) were the first to make the attempt. The varied usage of the term allometry has contributed to the lack of previous experiments, as many experiments have claimed to be selecting on allometry while selecting either on the intercept of the allometric relationship, or some indeterminate mixture of slope and intercept (reviewed by Houle et al. 2011; Pélabon et al. 2014). The logical paradox of directly applying selection to a static allometric slope may also have limited such selection experiments. An individual organism does not express a slope, just a combination of their size and the trait allometrically related to size. The observed size and trait combination is but one realization of an infinite combination of trait values that they could have expressed if a different combination of the factors that influence organism size had arisen (e.g. Dreyer et al. 2016). Allometry, is thus an example of a function-valued trait (Stinchcombe et al. 2012). Unfortunately, the formalism that has recently been developed for such traits has not yet been applied to allometry.

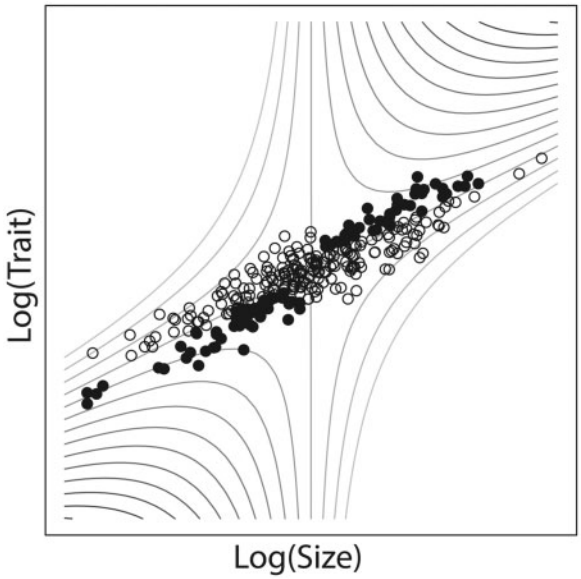

Fig. 3 Selection scheme used to select on slope. To increase slope, one would select the individuals represented by filled circles. *Source*: Adapted from Fig. 2 in Pélabon et al. (2014).

In an artificial selection experiment there are two ways of dealing with the fact that individuals do not express slope. First, one could measure the slope within a group of relatives, analogous to what we did in the **G** matrix experiment above, and then select individuals from those groups whose slopes deviate in targeted direction. Family level selection is a common approach to artificial selection in general (e.g., Hine et al 2011, 2014, Sztepanacz and Blows 2017), but has not been applied to the study of allometry. The second approach is to apply individual-level selection based on a particular model of how the inheritance of the sizes of the two parts works. Three artificial selection papers that have targeted static allometry have taken the latter approach (Egset et al. 2012; Bolstad et al. 2015; Stillwell et al. 2016), and all used a similar scheme to select on slope, represented in Fig. 3. This procedure defines the selection index conditional on size relative to the sex-specific mean size. The simplest use of this procedure was by Egset et al. (2012), where individuals that were above average in the size trait were selected if they also had positive residuals from the current allometric relationship, while individuals with below average sizes were selected if they had negative residuals. This scheme will be particularly effective when size, slope, and intercept are uncorrelated, as in Fig. 1D, and the **G** matrix represented in Table 1. In these distributions, individuals close to the mean size can be above or below the regression

Table 2 Design of and results from allometry selection experiments

Study	Species	Size	Trait	Generations	Rearing conditions	Significant response	
						Intercept	Slope
Egset	*Poecilia reticulata*	Body area	Tail area	3	Good	Yes	No
Bolstad	*Drosophila melanogaster*	Square root of wing area	Wing vein L2 length	Slope 25, 26 Intercept 7	Good	Yes	Yes
Bolstad	*Drosophila melanogaster*	Square root of wing area	Wing vein L2 length	19	Mixed	—	No
Stillwell	*Drosophila melanogaster*	Pupal area	Wing area	18	Mixed	—	Yes

because of variation in either slope or intercept, while individuals far from the mean size, are more likely to have a slope that deviates in the direction selected. This suggests that favoring individuals that deviate more from the mean size will be more effective at targeting slope. To take advantage of this, Stillwell et al. (2016) only selected individuals in the first and fourth quartiles of size. Bolstad et al. (2015) similarly favored individuals that were far from the mean size but also included stabilizing selection on size and the trait.

The three papers that have selected on static allometry include a total of six experiments, four targeted at slope, and two targeted at intercept. The characteristics of these selection experiments are summarized in Table 2. Bolstad et al.'s (2015) unmanipulated size experiment and Stillwell et al.'s (2016) experiment generated significant responses in slope to artificial selection. Bolstad et al. (2015) achieved a response of 1% per generation in their unmanipulated size treatment. The rate of change in slope cannot be directly calculated from the data in the paper by Stillwell et al. (2016). To get a crude idea of the actual rate of response, we assumed that the original slope was 1, then used Fig. 2 in their paper to estimate that slope changed at the rate of about 0.4% per generation on average. Overall, the results of these experiments reject the absolute genetic constraint hypothesis as a general explanation for the cause of evolutionary stasis in allometric slope.

In contrast to these experiments, Egset et al. (2012) did not generate a statistically significant difference in slope. However, they only selected for three generations. The mixed rearing condition treatment in Bolstad et al. (2015) also did not show significant response in slope. Bolstad et al. (2015) suggested that this was because the genetic variation among their starved flies reflected only variation in early growth, while the size range in well-fed flies was weighted toward variation in late growth. Alternatively, these results could reflect different mechanisms of sensitivity to environmental variation (Debat and David 2001) that result in size variation. In an experiment in *Drosophila serrata*, most of the genetic variation in environmental sensitivity of wing shape was dominance and not additive variance (Sztepanacz et al 2017). Consequently, the starved selection treatments may have primarily targeted dominance variation, or environmentally induced covariance (Rausher 1992) between traits, rather than an additive genetic covariance that could respond to selection.

In two of these studies, the authors also selected on the allometric intercept to contrast the rate of response in slope and intercept. Egset et al. (2012) selected for three generations to increase or decrease residuals from the least-squares regression. Bolstad et al. (2015) selected for seven generations to increase or decrease the residuals, while also performing stabilizing selection on size. These treatments are similar to earlier selection experiments on the residuals from the relationship between two traits (Weber 1990, 1992; Wilkinson 1993; Frankino et al. 2005). All of these experiments achieved rapid, highly significant changes in intercept. However, the actual rate of response of intercept was similar to the rate of response in slope. In the well-fed treatment, Bolstad et al. (2015), on average, obtained a 1.6% change in intercept per generation, slightly greater than the response they obtained for slope. Tail area increased by 2% per generation and decreased by 1% per generation in guppies (Egset et al. 2012). Overall, these results from slope selection experiments suggest that there is often substantial genetic variation for slope that can yield a rate of evolution in slope comparable to that on traits like intercept that can be measured at the individual level.

Stasis and natural selection

In just seven generations of artificial selection by Bolstad et al. (2015), the height of the allometric line approached the outermost range of heights that 40 million years of evolution produced.

Similarly, after just 25 or 26 generations of selection, the allometric slopes approached the outer limits found across the family Drosophilidae. Importantly, neither the slope nor the intercept appeared to plateau in response, suggesting that both traits could have become more extreme had artificial selection continued. The fact that more extreme trait values do not exist in nature, despite the genetic capacity to produce such phenotypes, suggests that selection limits their sustained exaggeration.

Natural selection consistent enough to generate stasis can be generated in different ways. For example, temporally fluctuating selection acting over short timescales can, over longer timescales, result in an overall pattern of stabilizing selection (Lynch 1990; Hendry and Kinnison 1999; Eldredge et al. 2005; Estes and Arnold 2007; Uyeda et al. 2011). Stabilizing selection acting directly on allometric slope may also be conserved over evolutionary time, so that the slope is continually returned to its starting state whenever it is perturbed away from that state. Strong stabilizing selection acting on many individual traits, however, is unlikely. If there was direct stabilizing selection acting on many traits, even its strength was 100 times weaker than current estimates suggest, population mean fitness would be impossibly low (Barton 1990). In most cases, stabilizing selection will arise through a multivariate process, where some multivariate trait combinations are subject to strong stabilizing selection and most are subject to much weaker stabilizing selection (Johnson and Barton 2005, Blows and Brooks 2003). A conserved pattern of multivariate stabilizing selection is a popular explanation for the observation of stasis in many types of traits over long timescales. One prediction of this hypothesis is that multivariate patterns of genetic variation will become aligned with the fitness surface (Lande 1980; Cheverud 1984; Arnold et al. 2001), for which there is some empirical evidence in experimental systems (Hunt et al. 2007; Roff and Fairbairn 2012, Jones et al. 2003, Revell et al. 2010, Brooks et al. 2005). An allometric slope would then lie along a ridge of high fitness.

While unchanging stabilizing selection can explain stasis, Hansen and Houle (2004) have emphasized that even though the potential biological reasons for unchanging stabilizing selection are plausible, they are not necessarily true. The many hypotheses that could explain stable patterns of selection essentially shift the problem of stasis to some other sort of stability, such as stability of the ecological niche, or tracking of a fixed niche by compensatory evolution, among other possibilities (Holt and Gaines 1992; Ackerly 2003; Hansen 2012). In the context of allometry, such stable niche parameters do seem plausible for many traits. For example, the demands of locomotion may enforce optimal body-size to appendage relationships that depend on the unchanging properties of the medium through (or on) which an organism moves.

Pleiotropic genetic constraints

More dramatic evidence of the role of selection in maintaining the pattern of allometry was provided by the change in phenotype after artificial selection was relaxed in the Bolstad et al. (2015) experiment. Each slope-selected population rapidly returned toward its original slope, reverting faster than they had evolved in response to artificial selection, losing 75–80% of their response in just 16 generations. About 20% of this return could, in principle, be explained by the breakup of linkage disequilibrium between genetic variation in size and intercept. Therefore, the vast majority of the reversion in slope was due to opposing natural selection. Artificial selection experiments often observe a reversion of traits toward their initial values after selection is relaxed (Reeve and Robertson 1953; Falconer and Mackay 1996; Hine et al. 2011), like Bolstad et al. observed. Furthermore, agricultural populations that have experienced sustained directional selection for economically important traits also tend to have depressed fitness that generates opposing selection (Rauw et al. 1998; Havenstein et al. 2003; Hill and Kirkpatrick 2010). Opposing selection appears to be a common cause of evolutionary limits in a variety of contexts.

The behavior of the intercept-selected populations was very different when artificial selection was relaxed. The down-selected population maintained the new intercept with minimal return towards the original mean. The up-selected population showed a significant but small decrease toward the original intercept of about 15% of the total response. This relative lack of reversion in intercept suggests that selection on the genetic variants that allowed the rapid increase in intercept is quite weak in the laboratory environment. Examination of Fig. 1 in Bolstad et al. (2015) shows that evolution in the intercept-selected lines actually produced flies that deviated more from the starting pattern of allometry than in the slope-selected lines. If the opposing natural selection to return to the original allometric relationship acted directly on static allometry, we would predict that the intercept-selected lines would return faster than the slope-selected lines. The fact

that they do not demonstrates that direct selection on allometry in the wing cannot explain the stasis of this allometric relationship.

If we reject both the absolute genetic constraint hypothesis, and the direct stabilizing selection hypothesis, what is left? The combination of results suggests that the detrimental fitness effects of the alleles that underlie allometry are not principally on wing-shape, but must also affect other aspects of organism form and function. The variation that affects slope may have pleiotropic effects on more traits, or have pleiotropic effects on traits that are more strongly selected than the variation that affects intercept. This leads to a hybrid hypothesis where evolution is limited by a combination of multivariate constraints and multivariate stabilizing selection. At the core of both the multivariate constraint and selection hypotheses, is the notion that allometry is highly polygenic, and that the genes that underlie allometry also underlie many other traits that determine an organism's fitness. The qualitative and quantitative extent of pleiotropy, in general, is an open question. Quantitative genetic approaches indicate that pleiotropy is widespread (Houle and Fierst 2013, McGuigan et al 2014, Blows et al 2015), with somewhere between 10 (Walsh and Blows 2009) and 20 (Hine et al 2018) independent trait combinations underlying all of the phenotypic variation in an organism. Quantitative trait locus (QTL) and genomic analyses come to somewhat different conclusions with a relatively small number of QTLs or alleles typically found to affect many traits (Zou et al 2008, Wagner et al 2008, Albert et al 2008, Durham et al 2014, Wagner and Zhang 2011). The latter analyses have low power, however, and the degree of pleiotropy is likely to be more extensive than suggested by these studies (Paaby and Rockman 2013).

In a multivariate context, we can redefine the question of genetic constraint on a focal trait to how much of its genetic variation is not shared with other traits that are also under strong selection. This is the concept of conditional genetic variance (Hansen et al. 2003; Hansen and Houle 2008), which describes the variation that is independent of all other traits. Conditional evolvability predicts the selection response of traits when all other traits are held constant. It has long been appreciated that multivariate genetic covariances can redistribute genetic variation into a relatively small number of trait combinations (Walsh and Blows, 2009), leaving others with little or no genetic variation (Gomulkiewicz and Houle 2009; Kirkpatrick 2009). These trait combinations tend to suffer from evolutionary responses that are stochastic (Hine et al. 2014), slow, or biased toward directions of high genetic variation (Schluter 1996; Chenoweth et al. 2010). In reality it is likely that only some other traits are under enough selection to impact the evolution of a focal trait, and it is the partial conditional evolvability of the focal trait generated by selection on that unknown subset of all traits that will generate pleiotropic constraints. If the conditional genetic variance is close to zero, the evolvability will also approach zero unless the changes in selection are both strong and maintained for long enough to allow the genetic architecture itself to evolve.

Why are there pleiotropic constraints?

The hypothesis of pleiotropic constraints is related to the relatively unfamiliar concept of burden (Riedl 1977). Riedl proposed that systems that arise early in development or that are functionally essential for an organism are both highly pleiotropic to begin with but also evolve to become more highly interconnected as any accretion of new traits and organismal functions must depend on the previously existing functions (Riedl, 1977). He considered these systems to be "burdened" because their genetic interdependence would make them, in principle, unalterable. Allometry may be one example of a burdened system as envisioned by Riedl, particularly if it arises through a response to common and fundamental growth factors. There are many candidates for such fundamental systems to underlie allometry (Shingleton and Frankino 2018), in particular the insulin signaling pathway that is a primary integrator of growth in animals.

The existence of pleiotropic constraints or burden is an attractive hypothesis that can explain stasis in allometry, but is also a difficult one to test. Logically it rests on the existence of pleiotropic effects that have not been observed directly. It further depends on the nature of natural selection on those unknown traits. Identifying which other traits suffer detrimental consequences and therefore mediate opposing natural selection is straightforward in some cases (Ryan et al. 1982; Godin and McDonough 2003; Fernandez and Morris 2008). In many cases, however, opposing selection arises through pleiotropic effects of alleles on a number of unknown traits. Two recent studies in *D. serrata* and one in *D. bunnanda*, have attributed the lack of contemporary evolution of male sexual displays, whose phenotypes are subject to directional sexual selection, to pleiotropic alleles that generate a net effect of stabilizing selection on the major axes of genetic variance that underlie these traits (McGuigan and Blows 2009;

McGuigan et al. 2011; Delcourt et al. 2012; Sztepanacz and Rundle 2012). The key result of these experiments was that it was not possible to detect selective constraints by studying phenotypes themselves. Multivariate stabilizing selection was only detected by studying the genetic covariance between multivariate traits and fitness.

In order to validate the existence of burden and pleiotropic constraints, we need to find at least some of these pleiotropically related traits and identify their costs. This is a classic example of a genotype–phenotype map problem, suggesting that both genetic and phenotypic approaches are possible. Here we focus on a genetic approach that centers on manipulations of candidate pathways.

Several key pathways have been proposed for master regulators of relative growth and therefore allometry in flies (Shingleton and Frankino 2018). The first is the insulin pathway. Levels of circulating insulin-like peptides are sensitive to the overall nutritional state, and activate a common pathway in proliferating cells, helping to coordinate the overall level of growth across the body. Manipulation of this sensitivity to insulin can alter allometry (Tang et al. 2011), and some naturally evolved differences in allometric relationships can be traced to differences in such sensitivity (Emlen et al. 2012). In addition, there is increasing evidence that the key insect growth hormones ecdysone and juvenile hormone that regulate molt and maturation also interact with the insulin pathway to generate overall effects on growth (Shingleton and Frankino 2018). It is also clear that there are other ways of altering the relative growth of body parts that can operate locally, for example, within certain tissues. Organs behave as if they have a target size, and will often undergo compensatory growth if they are manipulated to retard their growth early in development, a process that also seems to involve insulin-like peptides (Shingleton and Frankino 2018). With candidate pathways in hand, the ability to manipulate their expression greatly simplifies the completion of a genotype–phenotype map. Large manipulations of function may have phenotypic effects well beyond the range of natural variation, making it relatively easy to detect the pleiotropic effects of those manipulations even without an a priori set of target phenotypes.

On the other hand, the notion that a small number of master regulators are responsible for the majority of natural phenotypic variation has not fared well as our ability to map genome-wide variation has accelerated. Therefore, to answer the question about the genetic basis for the evolution of allometry, we should consider candidate allometric pathways, but also allow for the possibility that other processes may be involved. For example, the set of canonical developmental pathways involved in growth and differentiation (Matamoro-Vidal et al. 2015) operate throughout the fly body, could affect allometry, and are also likely to have pleotropic effects. Such an investigation would depend on the ability to genetically manipulate a wide array of genes. With the toolkit available in *D. melanogaster*, it is now possible to envision such experiments on a scale sufficient to determine the relative ability of candidate pathways and other genes to alter allometry. For several years we have been carrying out experiments looking for effects of gene expression knockdowns on wing shape in *D. melanogaster* that meet some of these requirements, which we call the Dictionary of Genetic Effects (Pitchers et al. 2019; D. Houle, unpublished data). Unfortunately for the study of allometry, we designed these experiments to measure general shape changes, and not to study allometry. Nevertheless, we have now reanalyzed these experiments to look for effects on allometry. The results are not particularly convincing, but serve to illustrate a viable approach.

In the Dictionary project, we quantitatively manipulate gene expression using organism-wide RNAi knockdowns and measure the phenotypic effects on appendage morphology (Pitchers et al. 2019; D. Houle, unpublished data). We have manipulated expression of over 150 genes thus far, and these have been successful at determining genes capable of affecting appendage size and shape, and provide a direction for those effects that can be related to gene function (Pitchers et al. 2019). All these experiments have assessed wing variation, and in the past year we have also added leg morphology to the list of phenotypes measured (D. Houle, unpublished data). To measure appendage size, we calculated wing area, and the total length of all three legs on one side of the body. We then tested whether the allometric relationship between log appendage size and the logs of the lengths of the four major long veins of the wing, or the lengths of each leg segment were altered by the knockdowns. The results of these analyses are shown in Supplementary Table S1 for wing only experiments, and Supplementary Table S2 for the more recent estimates of wing and leg experiments. Overall, RNAi knockdowns seem to increase the variance among treatment slopes, although in some cases controls also show increases in slope variance among treatments. The wing-only experiments offered fairly convincing evidence that the knockdowns increased the slope variance among

treatments. The average quantile of the observed slope variance was higher in the knockdowns than the controls for all phenotypes except the length of vein 2, and for the average quantiles across all traits. On average 1.52 of the four measured allometric relationships showed significantly increased variance in the RNAi knockdowns, while 0.75 showed significantly increased variance in the four control experiments. Genes in a wide variety of pathways had highly significant effects on the variance in slope, including representatives of the Dpp, wingless, hippo pathways, as did some of the genes with unknown function, such as those with a 'CG' designation.

In the combined leg–wing experiments, knockdown experiments also resulted in increased slope variance, while variance increased somewhat less in most control crosses. One control set showed highly significant increases in slope variance in allometry in four leg segments, despite the lack of a gene knockdown. While the mean number of significant changes in slope was 2.8 per knockdown experiment, and 1.7 in controls, in this set of experiments, the average quantile of the observed variances was not significantly different between knockdowns and controls. If there were a global allometric process, we would expect that knockdowns that increased slope variance in legs would also increase slope variance in wings. The observed correlation was 0.32 ($P = 0.09$).

Our results from these experiments are consistent with the possibility that many different genes and pathways can contribute genetic variation in allometry. The paradox raised by these results is that the strength of the pleiotropic constraints suggested by the rapid recovery toward the naturally evolved allometric slope would be readily explained if there were a simple master regulator of allometry throughout the body. Perturbing such a process would have predictable and widespread pleiotropic effects that would explain how costly to fitness such a perturbation appears to be. Our knockdown experiments instead seem consistent with many developmental processes having potential effects on allometry, as does the apparently highly polygenic basis for genetic variation in allometry. In this case, we then have to imagine that each of the many alleles capable of changing allometry is deleterious due to its own unique set of pleiotropic effects. In that case, the task of understanding the pleiotropic constraints will not be accomplished by the study of particular phenotypes.

Conclusions

Our work with fly appendages, and in particular the allometry of the L2 wing vein captures many of the issues concerning the evolution of allometry. Allometry of the L2 vein evolves very slowly, as does allometry of many other traits in a wide variety of taxa. The available explanations for this relative stasis involve some combination of genetic constraints and natural selection.

Using large scale quantitative genetic experiments in the *Drosophila* wing, we can decisively reject the pure constraint hypothesis that stasis is due to an absolute lack of genetic variation. Inbred line analyses indicate among line variance in slope and intercept, and we have caused the uncharacteristically rapid evolution of L2 allometry to the outer limits of that found in the family Drosophilidae with just 1 year of artificial selection. When we relaxed artificial selection on intercept-selected and allometry-selected populations, we obtained strikingly different results: the allometry-selected lines raced back toward the starting allometry, while the intercept selected lines returned toward the starting intercept very slowly. If direct opposing selection on allometry of the wing was the primary constraint, perturbing either intercept or slope should have roughly equivalent fitness costs.

These combined results suggest that unknown pleiotropic costs of changes to allometry are responsible for its striking evolutionary stasis. The remaining mystery is what generates those pleiotropic costs. We can imagine them arising from the effects of a few master regulatory processes, or from the diffuse effects of perturbations to a wide variety of developmental events. The results of our genetic manipulations were equivocal, but suggest that many developmental processes have effects on allometry. We believe that well-designed high-throughput genetic manipulations that target allometric relationships are well worth pursuing in the future. Such experiments will have the ability to resolve some of the pressing questions about the evolution of allometry.

Acknowledgments

We thank Rosa Moscarella, Eladio J. Márquez, and Kevin Doheny for their leading roles in the RNAi knockdown experiments, the many undergraduates who aided in rearing flies and imaging and splining of wings and now legs. We also thank Geir Bolstad for providing the R code to produce Fig. 3, and Christophe Pélabon and the reviewers for their comments on the manuscript.

Funding

This work was supported by National Institute of General Medical Sciences 1R01GM094424-01 to D.

Houle and Ian Dworkin, U.S. National Science Foundation Division of Environmental Biology grants 0129219 to D.H. and E. J. Márquez and 1556774 to D.H., and by Research Council of Norway, Grant 196494/V40 to Christophe Pélabon. Stocks obtained from the Bloomington Drosophila Stock Center (NIH P40OD018537) were used in this study.

Supplementary data

Supplementary data available at *ICB* online.

References

Ackerly DD. 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. Int J Plant Sci 164:S165–84.

Albert AY, Sawaya S, Vines TH, Knecht AK, Miller CT, Summers BR, Balabhadra S, Kingsley DM, Schluter D. 2008. The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. Evolution 62:76–85.

Arnold SJ, Pfrender ME, Jones AG. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. Genetica 112/113:9–32.

Blows MW, Allen SL, Collet JM, Chenoweth SF, McGuigan K. (2015). The phenome-wide distribution of genetic variance. Am Nat 186:15–30.

Blows MW, Brooks R. 2003. Measuring nonlinear selection. Am Nat 162:815–20.

Blows MW, McGuigan K. 2015. The distribution of genetic variance across phenotypic space and the response to selection. Mol Ecol 24:2056–72.

Bolstad GH, Cassara JA, Márquez E, Hansen TF, Van Der Linde K, Houle D, Pélabon C. 2015. Complex constraints on allometry revealed by artificial selection on the wing of *Drosophila melanogaster*. Proc Natl Acad Sci U S A 112:13284–9.

Brooks R, Hunt J, Blows MW, Smith MJ, Bussiere LF, Jennions MD. 2005. Experimental evidence for multivariate stabilizing sexual selection. Evolution 59:871–80.

Chenoweth SF, Rundle HD, Blows MW. 2010. The contribution of selection and genetic constraints to phenotypic divergence. Am Nat 175:186–96.

Cheverud JM. 1984. Quantitative genetics and developmental constraints on evolution by selection. J Theor Biol 110:155–71.

Cock A. 1966. Genetical aspects of metrical growth and form in animals. Q Rev Biol 41:131–90.

Debat V, David P. 2001. Mapping phenotypes: canalization, plasticity and developmental stability. Trends Ecol Evol 16:555–61.

Delcourt M, Blows MW, Aguirre JD, Rundle HD. 2012. Evolutionary optimum for male sexual traits characterized using the multivariate Robertson-Price Identity. Proc Natl Acad Sci U S A 109:10414–9.

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 70:1703–16.

Durham MF, Magwire MM, Stone EA, Leips J. 2014. Genome-wide analysis in *Drosophila* reveals age-specific effects of SNPs on fitness traits. Nat Commun 5:4338.

Egset CK, Hansen TF, Le Rouzic A, Bolstad GH, Rosenqvist G, Pelabon C. 2012. Artificial selection on allometry: change in elevation but not slope. J Evol Biol 25:938–48.

Eldredge N, Thompson JN, Brakefield PM, Gavrilets S, Jablonski D, Jackson JB, Lenski RE, Lieberman BS, McPeek MA, Miller W. 2005. The dynamics of evolutionary stasis. Paleobiology 31:133–45.

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 337:860–4.

Estes S, Arnold SJ. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. Am Nat 169:227–44.

Falconer DS, Mackay T. 1996. Introduction to quantitative genetics. Essex: Addison Wesley Longman.

Fernandez AA, Morris MR. 2008. Mate choice for more melanin as a mechanism to maintain a functional oncogene. Proc Natl Acad Sci USA 105:13503–7.

Frankino WA, Zwaan BJ, Stern DL, Brakefield PM. 2005. Natural selection and developmental constraints in the evolution of allometries. Science 307:718–20.

Godin JGJ, McDonough HE. 2003. Predator preference for brightly colored males in the guppy: a viability cost for a sexually selected trait. Behav Ecol 14:194–200.

Gomulkiewicz R, Houle D. 2009. Joint demographic and genetic constraints on evolution. Am Nat 174:E218–E2229.

Gould SJ. 1971. Geometric similarity in allometric growth: a contribution to the problem of scaling in the evolution of size. Am Nat 105:113–36.

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.

Gould SJ, Lewontin RC. 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc Lond B Biol Sci 205:581–98.

Hansen TF. 2012. Adaptive landscapes and macroevolutionary dynamics. In: Svensson E, Calsbeek R, editors. The adaptive landscape in evolutionary biology. Oxford: Oxford University Press. p. 205–26.

Hansen TF, Houle D. 2004. Evolvability, stabilizing selection, and the problem of stasis. In: Pigliucci M, Preston K, editors. The Evolutionary Biology of Complex Phenotypes. Oxford: Oxford University Press. p. 130–50.

Hansen TF, Houle D. 2008. Measuring and comparing evolvability and constraint in multivariate characters. J Evol Biol 21:1201–19.

Hansen TF, Pélabon C, Armbruster WS, Carlson ML. 2003. Evolvability and genetic constraint in *Dalechampia* blossoms: components of variance and measures of evolvability. J Evol Biol 16:754–66.

Havenstein G, Ferket P, Qureshi M. 2003. Growth, livability, and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. Poult Sci 82:1500–8.

Hendry AP, Kinnison MT. 1999. Perspective: the pace of modern life: measuring rates of contemporary microevolution. Evolution 53:1637–53.

Hill WG, Kirkpatrick M. 2010. What animal breeding has taught us about evolution. Annu Rev Ecol Evol Syst 41:1–19.

Hine E, McGuigan K, Blows MW. 2011. Natural selection stops the evolution of male attractiveness. Proc Natl Acad Sci U S A 108:3659–64.

Hine E, McGuigan K, Blows MW. 2014. Evolutionary constraints in high-dimensional trait sets. Am Nat 184:119–31.

Hine E, Runcie DE, McGuigan K, Blows MW. 2018. Uneven distribution of mutational variance across the transcriptome of *Drosophila serrata* revealed by high-dimensional analysis of gene expression. Genetics 209:1319–28.

Holt RD, Gaines MS. 1992. Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. Evol Ecol 6:433–47.

Houle D, Fierst J. 2013. Properties of spontaneous mutational variance and covariance for wing size and shape in *Drosophila melanogaster*. Evolution 67:1116–30.

Houle D, Meyer K. 2015. Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood. J Evol Biol 28:1542–9.

Houle D, Pelabon C, Wagner GP, Hansen TF. 2011. Measurement and meaning in biology. Q Rev Biol 86:3–34.

Hunt J, Blows MW, Zajitschek F, Jennions MD, Brooks R. 2007. Reconciling strong stabilizing selection with the maintenance of genetic variation in a natural population of black field crickets (*Teleogryllus commodus*). Genetics 177:875–80.

Huxley JS. 1932. Problems of relative growth. New York: Dover.

Johnson T, Barton N. 2005. Theoretical models of selection and mutation on quantitative traits. Philos Trans R Soc B Biol Sci 360:1411–25.

Jones AG, Arnold SJ, Bürger R. 2003. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. Evolution 57:1747–60.

Kirkpatrick M. 2009. Patterns of quantitative genetic variation in multiple dimensions. Genetica 136:271–84.

Lande R. 1979. Quantitative genetic analysis of multivariate evolution applied to brain: body size allometry. Evolution 33:402–16.

Lande R. 1980. The genetic covariance between characters maintained by pleiotropic mutations. Genetics 94:203–15.

Lynch M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. Am Nat 136:727–41.

Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu DH, Casillas S, Han Y, Magwire MM, Cridland JM, et al. 2012. The *Drosophila melanogaster* genetic reference panel. Nature 482:173–8.

Matamoro-Vidal A, Salazar-Ciudad I, Houle D. 2015. Making quantitative morphological variation from basic developmental processes: where are we? The case of the *Drosophila* wing. Dev Dyn 244:1058–73.

McGuigan K, Blows MW. 2009. Asymmetry of genetic variation in fitness-related traits: apparent stabilizing selection on g_{max}. Evolution 63:2838–47.

McGuigan K, Collet JM, Allen SL, Chenoweth SF, Blows MW. 2014. Pleiotropic mutations are subject to strong stabilizing selection. Genetics 197:1051–62.

McGuigan K, Rowe L, Blows MW. 2011. Pleiotropy, apparent stabilizing selection and uncovering fitness optima. Trends Ecol Evol 26:22–9.

Meyer K. 2006–2018. Wombat: a program for mixed model analyses by restricted maximum likelihood. Armidale (NSW): Animal Genetics and Breeding Unit, University of New England.

Meyer K. 2007. WOMBAT–A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J Zhejiang Univ Sci B 8:815–21.

Paaby AB, Rockman MV. 2013. The many faces of pleiotropy. Trends Genet 29:66–73.

Pavlicev M, Norgard EA, Fawcett GL, Cheverud JM. 2011. Evolution of pleiotropy: epistatic interaction pattern supports a mechanistic model underlying variation in genotype-phenotype map. J Exp Zool B Mol Dev Evol 316B:371–85.

Pélabon C, Firmat C, Bolstad GH, Voje KL, Houle D, Cassara JA, Le Rouzic A, Hansen TF. 2014. Evolution of morphological allometry. Ann N Y Acad Sci 1320:58–75.

Pitchers W, Nye J, Márquez EJ, Kowalski A, Dworkin I, Houle D. 2019. A multivariate genome-wide association study of wing shape in *Drosophila melanogaster*. Genetics published online (https://doi.org/10.1534/genetics.118.301342).

Rausher M. 1992. The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. Evolution 46:616–26.

Rauw W, Kanis E, Noordhuizen-Stassen E, Grommers F. 1998. Undesirable side effects of selection for high production efficiency in farm animals: a review. Livest Prod Sci 56:15–33.

Reeve EC, Robertson FW. 1953. Studies in quantitative inheritance II. Analysis of a strain of *Drosophila melanogaster* selected for long wings. J Genet 51:276–316.

Rensch B. 1959. Evolution above the species level. New York: Columbia University Press.

Revell L, Mahler D, Sweeney J, Sobotka M, Fancher V, Losos J. 2010. Nonlinear selection and the evolution of variances and covariances for continuous characters in an anole. J Evol Biol 23:407–21.

Riedl R. 1977. A systems analytical approach to macroevolutionary phenomena. Q Rev Biol 52:351–70.

Roff DA, Fairbairn DJ. 2012. A test of the hypothesis that correlational selection generates genetic correlations. Evolution 66:2953–60.

Ryan MJ, Tuttle MD, Rand AS. 1982. Bat predation and sexual advertisement in a neotropical anuran. Am Nat 119:136–9.

Schluter D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766–74.

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*. Proc R Soc Lond B Biol Sci 276:2625–33.

Shingleton AW, Frankino WA. 2018. The (ongoing) problem of relative growth. Curr Opin Insect Sci 25:9–19.

Simpson GG. 1944. Tempo and mode in evolution. New York: Columbia University.

Simpson GG. 1953. The major features of evolution. New York: Columbia University Press.

Stillwell RC, Shingleton AW, Dworkin I, Frankino WA. 2016. Tipping the scales: evolution of the allometric slope independent of average trait size. Evolution 70:433–44.

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 Ecol Evol 27:637–47.

Sztepanacz JL, Blows MW. 2017. Artificial selection to increase the phenotypic variance in $\mathbf{g}_{max}$ fails. Am Nat 190:707–23.

Sztepanacz JL, McGuigan K, Blows MW. 2017. Heritable micro-environmental variance covaries with fitness in an outbred population of *Drosophila serrata*. Genetics 206:2185–98.

Sztepanacz JL, Rundle HD. 2012. Reduced genetic variance among high fitness individuals: inferring stabilizing selection on male sexual displays in *Drosophila serrata*. Evolution 66:3101–10.

Tang HY, Smith-Caldas MS, Driscoll MV, Salhadar S, Shingleton AW. 2011. FOXO regulates organ-specific phenotypic plasticity in *Drosophila*. PLoS Genet7:e1002373.

Testa ND, Dworkin I. 2016. The sex-limited effects of mutations in the EGFR and TGF-β signaling pathways on shape and size sexual dimorphism and allometry in the *Drosophila* wing. Dev Genes Evol 226:159–71.

Truman J, Hiruma K, Allee J, MacWhinnie S, Champlin D, Riddiford L. 2006. Juvenile hormone is required to couple imaginal disc formation with nutrition in insects. Science 312:1385–8.

Uyeda JC, Hansen TF, Arnold SJ, Pienaar J. 2011. The million-year wait for macroevolutionary bursts. Proc Natl Acad Sci U S A 108:15908–13.

Voje KL, Hansen TF, Egset CK, Bolstad GH, Pelabon C. 2014. Allometric constraints and the evolution of allometry. Evolution 68:866–85.

Wagner GP, Kenney-Hunt JP, Pavlicev M, Peck JR, Waxman D, Cheverud JM. 2008. Pleiotropic scaling of gene effects and the 'cost of complexity'. Nature 452:470–2.

Wagner GP, Zhang JZ. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nat Rev Genet 12:204–13.

Walsh B, Blows MW. 2009. Abundant genetic variation + strong selection = multivariate genetic constraints: a geometric view of adaptation. Annu Rev Ecol Evol Syst 40:41–59.

Weber KE. 1990. Selection on wing allometry in *Drosophila melanogaster*. Genetics 126:975–89.

Weber KE. 1992. How small are the smallest selectable domains of form? Genetics 130:345–53.

Wilkinson GS. 1993. Artificial sexual selection alters allometry in the stalk-eyed fly *Cyrtodiopsis dalmanni* (Diptera, Diopsidae). Genet Res 62:213–22.

Zou L, Sriswasdi S, Ross B, Missiuro PV, Liu J, Ge H. 2008. Systematic analysis of pleiotropy in *C. elegans* early embryogenesis. PLoS Comp Biol 4:e1000003.