1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136

Complex constraints on allometry revealed by artificial selection on the wing of *Drosophila melanogaster*

Geir H. Bolstad[1,2,3], Jason A. Cassara[3], Eladio Márquez[3], Thomas F. Hansen[4], Kim van der Linde[3], David Houle[3], Christophe Pélabon[2].

[1]Norwegian Institute for Nature Research, 7485 Trondheim, Norway. [2]Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, 7491 Trondheim, Norway. [3]Department of Biological Science, Florida State University, Tallahassee, FL, USA. [4]Department of Biology, Centre for Ecological and Evolutionary Synthesis, University of Oslo, 0316 Oslo, Norway.

Precise exponential scaling with size is a fundamental aspect of phenotypic variation. These allometric "power laws" are often invariant across taxa and have long been hypothesized to reflect developmental constraints. Here we test this hypothesis by investigating the evolutionary potential of an allometric scaling relationship in drosophilid wing shape that is nearly invariant across 111 species separated by at least 50 million years of evolution. In only 26 generations of artificial selection in a population of *Drosophila melanogaster,* we were able to drive the allometric slope to the outer range of those found among the 111 sampled species. This response was rapidly lost when selection was suspended. Only a small proportion of this reversal could be explained by breakup of linkage disequilibrium, and direct selection on wing shape is also unlikely to explain the reversal, because the more divergent wing shapes produced by selection on the allometric intercept did not revert. We hypothesize that the reversal was instead caused by internal selection arising from pleiotropic links to unknown traits. Our results also suggest that the observed selection response in the allometric slope was due to a component expressed late in larval development and that variation in earlier development did not respond to selection. Together, these results are consistent with a role for pleiotropic constraints in explaining the remarkable evolutionary stability of allometric scaling.

allometry | artificial selection | comparative analyses | developmental constraints | pleiotropy

INTRODUCTION

Allometric scaling is an ubiquitous aspect of biological variation that is often strongly conserved across evolutionary time and typically explains a large fraction of observed variation in morphology, physiology or life history (1-7). This evolutionary conservatism can be explained either by stabilizing selection or by fundamental developmental or physiological constraints (8-17). Allometric "power laws" have been thought to reflect developmental constraints for nearly a century (6). Arguments of allometric constraints were used to explain patterns of macroevolution by architects of the modern synthesis such as Huxley (5), Simpson (18) and Rensch (19), and played a major role in Gould and Lewontin's (20) criticism of the "adaptationist programme". The idea of allometric constraints may, at least partially, have originated from the multiplicative growth model underlying Huxley's derivation of the allometric "power law" for morphological traits.

Julian Huxley (5, 21) showed that when a trait is under common growth regulation with size, the relationship between the trait Y and a size measure X is a power function of the form $Y = aX^b$, where a and b are constants. On a log-log scale, power functions become linear, with $\log(a)$ representing the intercept and b the slope of the allometric relationship $\log(Y) = \log(a) + b \log(X)$. Allometric "power laws" summarize variation among developmental stages (ontogenetic allometry), individuals in a population (static allometry) and populations or species (evolutionary allometry) (22). The three levels of allometry are related, and a higher level of allometry can be expressed as a function

of allometry at lower levels (23). Limited potential for evolution at a lower level in this allometric hierarchy would then cause constraints at all higher levels (6, 23).

Because of their fundamental importance and their relation to developmental constraints, there has been interest in testing the evolvability of scaling relationships in general and in particular the evolvability and evolutionary invariance of the static allometric slope. In a recent review, Voje et al. (6) argued that while there is abundant evidence for evolvability of intercepts of static allometric relations (i.e. mean shape), there are few clear demonstrations of additive genetic variance or microevolutionary changes in allometric slopes. This is mostly due to various conceptual and methodological issues (see 6, 24-26).

One exception is Pavlicev et al.'s (27) finding of small but significant heritabilities for several allometric exponents in an intercross between mouse strains selected for large and small body size. Another case comes from Tobler and Nijhout (28) where ten generations of selection on wing mass in the moth *Manduca sexta* produced a small change in wing-body scaling (see 6 for a reanalysis). This was, however, an indirect response, so it is unclear how free the slope was to evolve on its own. The result is also based on observing the relationship in a single generation, calling its replicability into question (see 26 for further discussion). In the only study that performed artificial selection separately and directly on the allometric slope and intercept, Egset et al. (29) found a clear response in the intercept but not in the slope of a tail-body allometry in guppies (*Poecilia reticulata*). The power of this study was also limited, however, because it extended only over three generations. From comparative analyses, there is clear


Significance

Many traits scale precisely with size, but it is unknown whether this is due to selection for optimal function, or due to evolutionary constraint. We used artificial selection to demonstrate that wing-shape scaling in fruit flies can respond to selection. This evolved response in scaling was lost during a few generations after selection ended, but other selected changes in wing shape persisted. Shape-size scaling in fly wings is therefore evolvable, but adaptation is apparently constrained by selection that may not be on wings. This may explain why scaling relationships are often evolutionarily conserved.
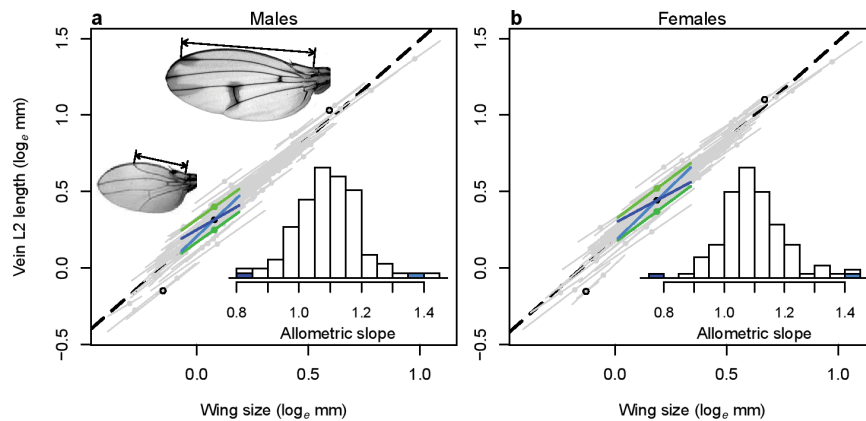
Fig. 1. Allometric relationships of males (**a**) and females (**b**) within and among 111 drosophilid species. Among-species (evolutionary) allometry (dashed line; slope males = 1.282±0.029, R^2 = 0.95; slope females = 1.291±0.029, R^2 = 0.95), within-species (static) allometries (grey lines; average slope males = 1.091±0.095, average R^2 = 0.89; average slope females = 1.095±0.096, average R^2 = 0.92), and the selection responses (green and blue lines, see Fig. 2 for explanation of color coding). The pictures illustrate our measure of L2-vein length (double-headed arrows) in a small (*Dettopsmyia nigrovittata*) and a large (*Idomyia mimica*) species (encircled dots). Wing size is square root of wing area. The inset histogram shows the among-species distribution of slopes with the final slopes of the up- and down-selected populations in blue.
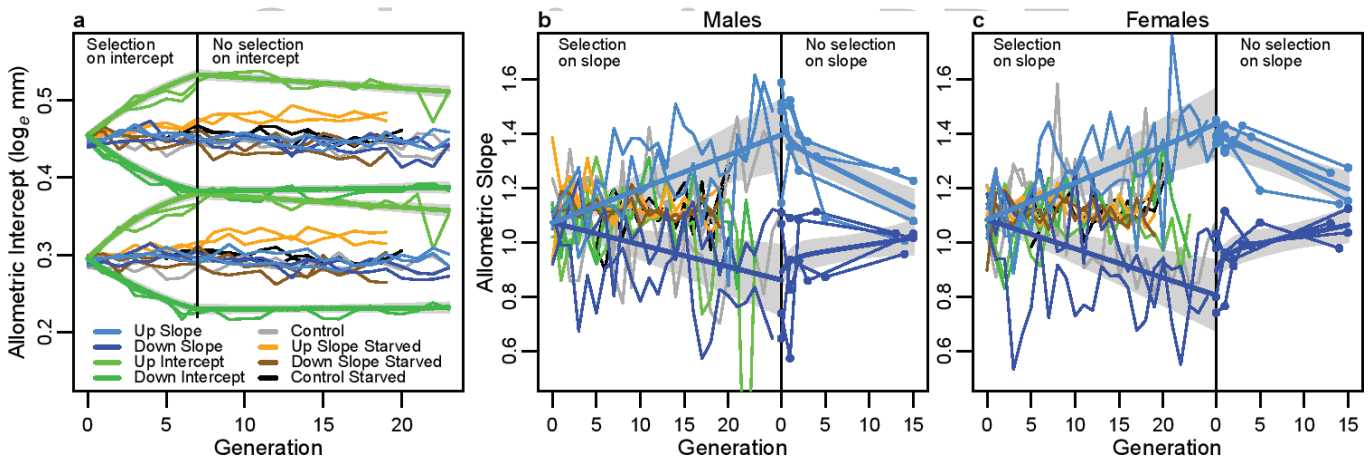
Fig. 2. Change in allometric intercept (**a**), allometric slope for males (**b**) and females (**c**) for the two replicates of each selection regime. The model fit for the different selection regimes are given by the thick lines with ± standard error in grey, see Tables S2, S3, and S4 for parameter estimates. In a, males have the lower intercept (trait mean). In b, the outlier "Up-Intercept" generation 22 had a slope of -0.17. The selection stopped at generation 25 and 26 in the slope-selected populations. In addition, we budded off a population from each replicate at generation 23 and maintained these new populations under relaxed selection. Therefore, the generation axis has two scales in b and c, with the right scale denoting generations under relaxed selection. The populations were measured at each generation except during relaxed selection on slope where measurements are indicated by circles.
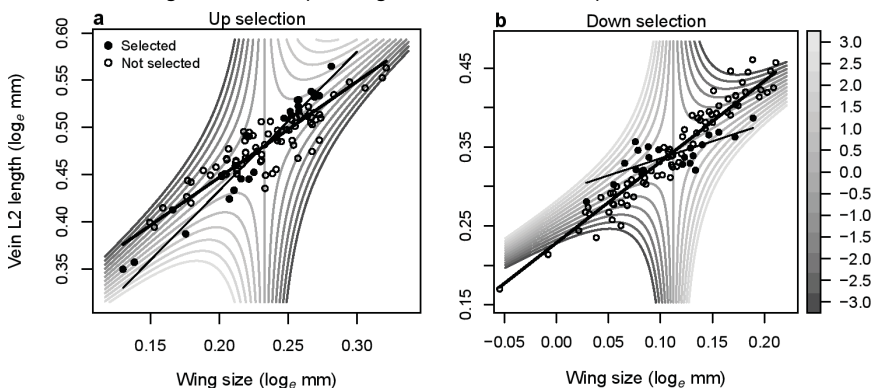
Fig. 3. Saddle function used as selection index (grayscale legend). Examples of selection to increase (**a**) and decrease (**b**) the allometric slope (thick regression lines). The 20 individuals (filled circles), out of 100, with the highest selection indices were selected. The thin regression lines give the allometric slope among the selected individuals. Only females are shown.

evidence of evolution of static slopes on long time scales, but no clear cases of substantial change over less than a million years (6).

Here we test the evolutionary potential of both the static allometric intercept and slope of an aspect of wing-shape allom-

etry in *D. melanogaster*. To do this, we use a series of large-scale artificial-selection experiments (58,046 measured flies in total), and compare the selection response to the natural variation in

Table 1. Response to selection on slope in the un-starved populations (c.f. Table S2)*, given in contrasts between treatments (with standard error). The units are change in allometric slope per generation. A model with only a common linear effect for each sex instead of an effect of each treatment (Up, Down, and Control) for each sex increase the AIC-score by 38.

Treatment contrast	Difference
Males	
Up – Down	0.0212±0.0078
Up – Control	0.0112±0.0078
Down – Control	-0.0100±0.0078
Females:	
Up – Down	0.0245±0.0078
Up – Control	0.0114±0.0078
Down – Control	-0.0131±0.0078

*The same statistical model is used to produce the results of this table and Table S2, only that the relevant fixed effects are here expressed as contrasts.

wing-shape allometry obtained from a comparative study of 111 drosophilid species (20,345 measured flies).

RESULTS AND DISCUSSION

In the Dipteran family Drosophilidae, the length of wing-vein L2 shows tight and positive evolutionary and static allometries with wing size (Fig. 1). The L2 vein intersects the leading edge of the wing (vein L1) in a relatively proximal position in small wings and more distally in larger wings (Fig. 1). Our comparative analysis shows that both the intercept and the slope of the static allometry have evolved, but at very low rates (Table S1). For the intercept, the estimated standard deviation of change from a Brownian-motion model of evolution ranges from 0.026 to 0.042 $\log_e$ mm per Myr depending on how deep the phylogeny is considered to be. This corresponds to average changes in relative L2 length of less than 4% per million years. Similarly, the slope changed with an estimated standard deviation of less than 0.037 slope units per Myr. Hence, wing-shape allometry is strongly conserved in drosophilids.

It proved easy to change the allometric intercept (i.e. mean shape) by selecting on the residuals of the log-log regression (Fig. 2a). After seven generations of selection on the intercept, the average length of vein L2, had increased by ∼8.5% (from 1.34 to 1.46 mm in males and from 1.58 to 1.70 mm in females) in the "up-selected" populations and decreased by ∼6.5% (from 1.34 to 1.26 mm in males and from 1.58 to 1.47 in females) in the "down-selected" populations. Hence, the average length of vein L2 had become as large in the up-selected males as in the down-selected females (Fig 2a) while the sexual size dimorphism remained unchanged. This selection response of about 1% per generation is rapid evolution on a macroevolutionary scale. Only four generations of artificial selection were sufficient to produce a change comparable to the one observed over a million years of evolution, as judged from the average change in the comparative data. This change was also stable. During 16 generations without artificial selection, the intercept returned less than 0.15% per generation towards the original value (Fig. 2a). Hence, natural selection on wing shape was weak in the lab environment. These results show that mean shape is highly evolvable in this population, as it is in many insects (30-35). The stasis of wing shape among drosophilids must therefore be due to other factors than the lack of genetic variation.

Selecting on the static allometric slope is challenging because it is a property of a population that is not expressed at the individual level. To construct an individual-based selection index we used Huxley's (5, 21) allometric model, which is based on

individual growth parameters (see *SI Materials and Methods*). These individual growth parameters can be expressed as individual allometries that relate to the static allometry in the same way as ontogenetic allometries (see 23). Under this model, individuals that are near the bivariate phenotypic mean (i.e. average wing size and vein length) could have any breeding value for slope, and selection of those individuals would contribute nothing to the selection differential for slope. Therefore, while natural selection on slope may operate through favoring a particular trait-size relationship and generate high fitness for individuals at the bivariate phenotypic mean, a ridge-like selection index would be inefficient for generating selection on slope. Instead, we derived a saddle-shaped "fitness" function that simultaneously maximizes selection on slope while minimizing selection on size, variance in size, and the allometric intercept (Fig. 3, *SI Materials and Methods*). The result was that clusters of larger and smaller than average flies were selected within each sex in each generation.

The responses of the allometric slopes were erratic, but after 26 generations the male and female slopes had changed from 1.08 and 1.09 in the original populations to 1.39 and 1.42 in the up-selected populations and to 0.84 and 0.78 in the down-selected populations (Fig. 2b and c). These are large changes, resulting in allometric slopes in the outer range of those found in the sampled species (Fig. 1). The allometry in the original population implies that a 10% increase in wing size increased L2 length by 10.9% (for females). After 26 generations of selection this had changed so that a 10% increase in wing size would now increase L2 length by 14.2% in the up-selected females and 7.8% in the down-selected females. Figures S1 and S2 give a visual representation of the response to selection on the wing outline.

The statistical support for selection responses in the static allometric slope is strong, as judged from the Akaike Information Criterion (AIC; Table 1). The average responses of the up- and the down-selected lines were significantly different from each other, although not from the average of the two controls (Table 1). The relationship between the change in slope and a measure of cumulative selection strength was also statistically significant for all comparisons except for the males in one of the down-selected populations (Fig 4), and several generations after selection ended, the average allometric slopes of the up- and down-selected lines still differed significantly from each other and from the starting value (Fig 2b and c). We noticed an erratic and diverging behavior of the allometric slopes in the control lines (Fig. 2b and c). This strange behavior may, at least partly, be due to smaller sample size in these lines, resulting in less precise estimates of the slopes.

The precision with which the allometric slope can be estimated depends on the range of sizes investigated. We therefore expected that an increased range of sizes would improve our ability to select on the allometric slope by making it easier to detect flies with extreme breeding values for allometric slope. We thus performed the same selection experiment on independent populations in which size variation was increased by starving half of the larvae over the last two days of development. These populations produced adult flies with a much wider range of sizes. Contrary to our expectation, the allometric slope did not respond to selection in these "starved" populations (Fig. 2b and c, Table S5).

To understand the lack of selection response in the starved populations, we subjected flies from the un-starved slope-selected populations to the same starvation treatment one generation after selection ended. Surprisingly, the entire evolved difference in the allometric slope disappeared in these starved flies. The difference in static allometric slope between the up- and the down-selected populations was only 0.022±0.061 when the flies were starved, as compared to 0.271±0.061 in their un-starved siblings in the same generation (averaged over replicates and sexes). This genotype-
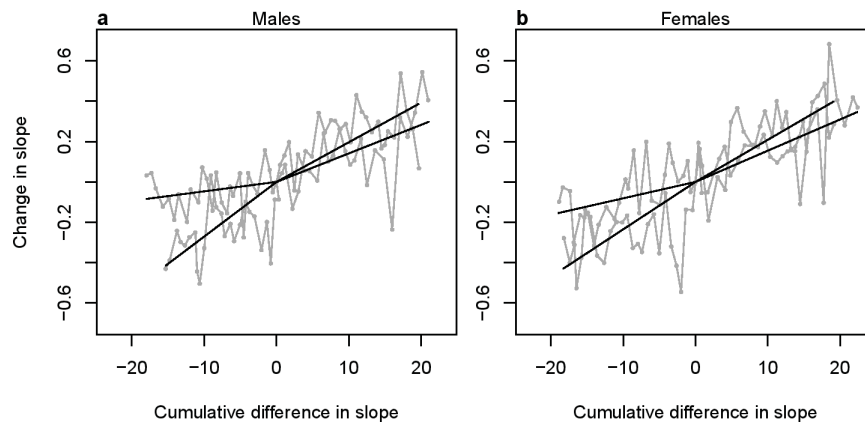
Fig. 4. The selection response of the allometric slope as a function of a measure of cumulative "selection strength" in males (**a**) and females (**b**). The cumulative selection strength is measured as the cumulative difference between the allometric slope of all individuals under selection and the allometric slope of the selected individuals (i.e. the difference between the thick and the thin regression line in Fig. 3). Note that this is not a proper measure of the selection differential because it does not reflect the difference in the underlying slope of each individual. Therefore, the regression lines in the figure do not represent true realized heritabilities, but still give some information about the response to selection. The regression lines, slope ± standard error, are 0.0196±0.0022 and 0.0142±0.0022 for up-selected males, 0.0047±0.0025 and 0.0271±0.0029 for down-selected males, 0.0209±0.0024 and 0.0155±0.0022 for up-selected females, and 0.0081±0.0025 and 0.0235±0.0026 for down-selected females.

by-environment interaction and the erratic selection response underscores the need for an experimental design with replicated controlled environments when investigating genetic differences in static slopes (see also 26). We hypothesize that the developmental processes responsible for the observed selection response in allometric slope are acting during the late third instar when growth is precluded in the starved flies. The allometric relations may thus be the result of developmental "palimpsests" (36), where subsequent developmental processes are written on top of each other to partially mask variation created at different stages. Our results suggest that some of these processes are evolvable, causing a selection response in the slope, while others constrain small flies to remain on the original static allometric line.

A potentially confounding source of response to selection is the creation of linkage disequilibrium between alleles that affect wing size and alleles that affect L2 length. Such linkage disequilibrium could have generated a change in the allometric slope without changes in allele frequency. For example, an association between alleles that increase wing size with alleles that increase L2 length would increase the allometric slope between wing size and L2 length. To minimize linkage disequilibrium we used disassortative mating when selecting on the static allometric slope. Female flies in the cluster above mean wing size were mated with male flies from below and vice versa (see Methods), so that recombination would be maximally effective in breaking up linkage disequilibrium. However, this does not completely prevent association between alleles, and we need to consider if the reversal of the response could have been due to breakup of linkage disequilibrium. Assuming an average recombination fraction of $r = 0.365$ between random loci in *D. melanogaster* (37), we estimated that the maximum fraction of the response that could be due to linkage disequilibrium averaged only 19.9±22.6% and 14.3±24.0% in males and females, respectively.

Non-genetic inheritance, such as parental (e.g. maternal) effects, may in principle affect the response to selection (38), but is not likely to be important in this case, as there are no indications of heritable non-genetic effects on wing shape in drosophilids.

We consider natural selection to be the most likely cause of the reversal towards the ancestral allometric slope. However, natural selection for restoring optimal wing function, for example due to selection for flight or courtship behavior, does not seem to be strong in the lab environment. In the presence of such selection we expect that the intercept-selected populations would also have

rapidly returned to their starting value. These populations instead showed little reversal of the selection response over an even longer time period, despite being more different from the initial wing shape than the slope-selected populations. Therefore, the more likely alternative is that the evolutionary change in allometric slope generated deleterious pleiotropic responses in other aspects of the phenotype, resulting in strong natural selection in the lab environment.

We have shown that a phylogenetically invariant allometric slope can evolve rapidly under selection, but that this seems to generate countervailing natural selection to return the allometric slope to its initial value. This suggests that conserved allometric scaling may be best explained by pleiotropic constraints (12). Riedl (39, 40) proposed that fundamental developmental processes may become increasingly constrained, or burdened, by other processes that interact with or depend on them. Under this hypothesis, aspects of the developmental system that lead to precise allometric scaling also affect other aspects of organismal form and function, leading to deleterious pleiotropic effects when they are altered. If true, this hypothesis could provide a general explanation for the striking evolutionary conservatism of allometric power laws, while still leaving open the possibility for allometries to be optimized by natural selection over long time scales through compensatory mutations.

METHODS

Comparative data. Species were obtained by collection from the wild, from the *Drosophila* Species Stock Center, or from other collectors. The full list of 111 taxa, specifications, their collection locations, and sample sizes are reported in Table S6. Flies were reared using combinations of food, temperature, and rearing environments suggested to be optimal for each particular species based on the instructions from the source or from published sources. Wild-collected specimens were measured when we were unable to rear the flies in the lab. Most of the 111 taxa are currently classified in the paraphyletic genus *Drosophila* or to genera in the subfamily Drosophilinae (41, 42). In addition, we included two species of steganine drosophilids and five outgroup species from other families. Four drosophilid species are represented by more than one subspecies. Most species had sample sizes of around 200 measured flies and the total sample size was 20,345.

Wing Measurements. The full procedure for imaging wings and for estimating vein locations and corresponding landmarks has been described by Houle et al. (43). In short, the left wing of a live CO_2-anaesthetized fly is immobilized in a suction device, a digital image of the wing is obtained, and the program Wings3.8 (44) is used to fit cubic B-splines to the wing veins, from which the coordinates of landmarks and semilandmarks are extracted. The length of vein L2 was estimated as the straight-line distance between the humeral break in the costa and the distal end of L2. The square root of wing

area was used as a measure of wing size. See Fig. S3 for detailed information on the wing measurements.

Derivation of populations for the selection experiment. The initial population was made by intercrossing 30 inbred lines obtained from the *Drosophila* Genetic Reference Panel (45). This population was allowed to mate freely for one generation before being divided into the selected populations. We maintained 16 different populations in this experiment: two selected for increase and two for decrease in allometric intercept, two for increase and two for decrease in allometric slope, two for increase and two for decrease in slope with the starvation treatment, and two un-starved and two starved controls. Details are given in *SI Materials and Methods*.

Rearing of selected populations. Details are given in *SI Materials and Methods*.

Justification of the allometric-slope selection index. Details are given in *SI Materials and Methods*.

Selection procedure. In the two replicate *control populations*, 20-25 virgin females and 20-25 males were chosen haphazardly and imaged before being divided into two vials to produce the next generation.

In the two replicate *starved control populations*, 50 virgin females and 50 males were imaged, then divided haphazardly into four groups of 25 flies each (12 or 13 of each sex). Two groups were placed in vials and their larvae fed normally; the other two were placed in "egg layers" and their larvae underwent the starvation treatment (see *SI Materials and Methods*). Note that these two populations were started later in the experiment (at generation 5 of the other populations).

Two replicate populations were maintained for each direction (up and down) of *selection on the allometric intercept*. Each generation in each population 100 virgin females and 100 males were chosen haphazardly and imaged. From these, 20-25 of each sex were selected to produce the next generation using the selection index:

$$W_{int} = \beta_1(x - \bar{x}) + \beta_2(y - \log_e[a] - bx),$$

where x is $\log_e$ wing size (square-root wing area), $\bar{x}$ is the average of x, y is $\log_e$ L2 length, $\log_e[a]$ is the allometric intercept, b is the allometric slope, and the term $(y - \log_e[a] - bx)$ is the residuals of the allometric regression. For the up selection $\beta_2 = 1$ and for the down selection $\beta_2 = -1$, while β_1 was optimized to reduce selection on size (see *SI Materials and Methods*). The selection index was fitted independently each generation within each selected population, and the flies were stored in individual vials from imaging until the selection-index score of each fly was calculated. Selection on the intercept was maintained in both replicates (A and B) for seven generations for both the up and the down directions. After this, the A replicates were discarded, while up- and down-selected populations in replicate B were maintained with relaxed selection (i.e. the same mating regime as in the control populations) for an additional 16 generations.

For each direction (up and down) of selection on the allometric slope, we kept two replicate populations (A and B). Each generation in each population, 100 virgin females and 100 males were imaged, of which 24 were selected using the selection index:

$$W_{slope} = \beta_x(x - \bar{x}) + \beta_y(y - \bar{y}) + \frac{1}{2}\gamma_x(x - \bar{x})^2 + \gamma_{xy}(x - \bar{x})(y - \bar{y}),$$

where $\gamma_{xy} = 1$ for the up selection (increase in slope) and -1 for the down selection (decrease in slope). This function was optimized by choosing values of the parameters β_x, β_y and γ_x, to minimize selection on the means of x and y and on the variance of x (see *SI Materials and Methods*). The selection index was fitted independently each generation within each selected population, and the flies were stored in individual vials after imaging until the selection-index score of each fly was calculated. In order to prevent linkage disequilibrium, we enforced disassortative mating on size by first dividing the selected flies into two groups, one containing the largest 12 females and smallest 12 males, the other with the smallest 12 females and largest 12 males. Each of these groups was then split at random into two groups of six males and six females, and each such group was placed in vials with another group of opposite-sex flies of contrasting size. These flies were allowed to mate and lay eggs overnight. Selection was carried out for 26 generations in replicate A and for 25 generations in replicate B. After this we maintained the populations for an additional 13 generations in replicate A and 14 generations in replicate B under relaxed selection (i.e. the same mating regime as the control populations). At generation 23 we split off a population from each of the replicates and maintained these new populations under relaxed selection for 15 generations. At several points during the period of relaxed selection we imaged 100 males and 100 females from each of the populations (see Fig. 2b).

We also maintained two replicate *starved slope-selected populations* for each direction of slope selection that were subjected to the same selection regime as the un-starved slope-selected populations, but reared using the starvation treatment. Selected was carried out for 19 generations in these populations. Details are given in *SI Materials and Methods*.

Statistical analyses. To estimate the rate of evolution for the allometric intercept and slope we fitted a Brownian-motion model of evolution using a phylogenetic mixed model (46). Because the depth of the phylogeny is highly uncertain (47), we used both a "best guess" of 133 Myr and a conservative estimate of 50 Myr to estimate the rate.

For the selection experiment, changes and differences in allometric intercept and slope of $\log_e$ vein-L2 length on $\log_e$ wing size (square-root wing area) were analyzed by a series of linear mixed-effects models because of the hierarchical structure of the data. Details are given in *SI Materials and Methods*.

Data storage. The data is available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.s270f.

1. Charnov EL (1993) *Life history invariants* (Oxford University Press, New York).
2. Schmidt-Nielsen K (1984) *Scaling: why is animal size so important?* (Cambridge University Press, Cambridge).
3. West GB, Brown JH, & Enquist BJ (1997) A general model for the origin of allometric scaling laws in biology. *Science* 276(5309):122-126.
4. Brown JH & West GB (2000) *Scaling in biology* (Oxford University Press, Oxford).
5. Huxley JS (1932) *Problems of relative growth* (L. MacVeagh, New York).
6. Voje KL, Hansen TF, Egset CK, Bolstad GH, & Pélabon C (2014) Allometric constraints and the evolution of allometry. *Evolution* 68(3):866-885.
7. Gould SJ (1966) Allometry and size in ontogeny and phylogeny. *Biol Rev Camb Philos Soc* 41(4):587-640.
8. Bradshaw AD (1991) Genostasis and the limits to evolution. *Phil Trans R Soc B* 333(1267):289-305.
9. Björklund M (1996) The importance of evolutionary constraints in ecological time scales. *Evol Ecol* 10(4):423-431.
10. Schluter D (2000) *The ecology of adaptive radiation* (Oxford University Press, Oxford).
11. Arnold SJ, Pfrender ME, & Jones AG (2001) The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* 112:9-32.
12. Hansen TF & Houle D (2004) Evolvability, stabilizing selection, and the problem of stasis. *The evolutionary biology of complex phenotypes*, eds Pigliucci M & Preston K (Oxford University Press, Oxford).
13. Brakefield PM & Roskam JC (2006) Exploring evolutionary constraints is a task for an integrative evolutionary biology. *Am Nat* 168(6):S4-S13.
14. Polly PD (2008) Developmental dynamics and G-matrices: can morphometric spaces be used to model phenotypic evolution? *Evol Biol* 35(2):83-96.
15. Futuyma DJ (2010) Evolutionary constraint and ecological consequences. *Evolution* 64(7):1865-1884.
16. Hansen TF (2012) Adaptive landscapes and macroevolutionary dynamics. *The adaptive landscape in evolutionary biology*, eds Svensson EI & Calsbeek R (Oxford University Press, Oxford).

17. Bolstad GH, *et al.* (2014) Genetic constraints predict evolutionary divergence in *Dalechampia* blossoms. *Phil Trans R Soc B* 369(1649):20130255.
18. Simpson GG (1944) *Tempo and mode in evolution* (Colombia Univ. Press, New York).
19. Rensch B (1959) *Evolution above the species level* (Columbia University Press, New York).
20. Gould SJ & Lewontin RC (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc B* 205(1161):581-598.
21. Huxley JS (1924) Constant differential growth-ratios and their significance. *Nature* 114:895-896.
22. Cheverud JM (1982) Relationships among ontogenetic, static, and evolutionary allometry. *Am J Phys Anthropol* 59(2):139-149.
23. Pélabon C, *et al.* (2013) On the relationship between ontogenetic and static allometry. *Am Nat* 181(2):195-212.
24. Houle D, Pélabon C, Wagner GP, & Hansen TF (2011) Measurement and meaning in biology. *Q Rev Biol* 86(1):3-34.
25. Hansen TF & Bartoszek K (2012) Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. *Syst Biol* 61(3):413-425.
26. Pélabon C, *et al.* (2014) Evolution of morphological allometry. *Ann N Y Acad Sci* 1320:58-75.
27. 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(5):371-385.
28. Tobler A & Nijhout HF (2010) Developmental constraints on the evolution of wing-body allometry in *Manduca sexta*. *Evol Dev* 12(6):592-600.
29. Egset CK, *et al.* (2012) Artificial selection on allometry: change in elevation but not slope. *J Evol Biol* 25(5):938-948.
30. Frankino WA, Zwaan BJ, Stern DL, & Brakefield PM (2005) Natural selection and developmental constraints in the evolution of allometries. *Science* 307(5710):718-720.
31. Frankino WA, Zwaan BJ, Stern DL, & Brakefield PM (2007) Internal and external constraints in the evolution of morphological allometries in a butterfly. *Evolution* 61(12):2958-2970.

32. Weber KE (1990) Selection on wing allometry in *Drosophila melanogaster*. *Genetics* 126(4):975-989.

33. Wilkinson GS (1993) Artificial sexual selection alters allometry in the stalk-eyed fly *Cyrtodiopsis dalmanni* (Diptera, Diopsidae). *Genet Res* 62(3):213-222.

34. Emlen DJ (1996) Artificial selection on horn length-body size allometry in the horned beetle *Onthophagus acuminatus* (Coleoptera: Scarabaeidae). *Evolution* 50(3):1219-1230.

35. Okada K & Miyatake T (2009) Genetic correlations between weapons, body shape and fighting behaviour in the horned beetle *Gnatocerus cornutus*. *Anim Behav* 77(5):1057-1065.

36. Hallgrimsson B, *et al.* (2009) Deciphering the palimpsest: studying the relationship between morphological integration and phenotypic covariation. *Evol Biol* 36(4):355-376.

37. Lynch M & Walsh B (1998) *Genetics and analysis of quantitative traits* (Sinauer Associates, Sunderland).

38. Kirkpatrick M & Lande R (1989) The Evolution of Maternal Characters. *Evolution* 43(3):485-503.

39. Riedl R (1977) Systems-analytical approach to macro-evolutionary phenomena. *Q Rev Biol* 52(4):351-370.

40. Riedl R (1978) *Order in living organisms* (Wiley, Brisbane).

41. van der Linde K & Houle D (2008) A supertree analysis and literature review of the genus *Drosophila* and closely related genera (Diptera, Drosophilidae). *Insect Syst Evol* 39(3):241-267.

42. van der Linde K, Houle D, Spicer GS, & Steppan SJ (2010) A supermatrix-based molecular phylogeny of the family Drosophilidae. *Genet Res, Camb* 92(1):25-38.

43. Houle D, Mezey J, Galpern P, & Carter A (2003) Automated measurement of *Drosophila* wings. *BMC Evol Biol* 3(1):25.

44. van der Linde K (2004-2013) Wings: automated capture of *Drosophila* wing shape, version 3.8.

45. Mackay TFC, *et al.* (2012) The *Drosophila melanogaster* genetic reference panel. *Nature* 482(7384):173-178.

46. Lynch M (1991) Methods for the analysis of comparative data in evolutionary biology. *Evolution* 45(5):1065-1080.

47. Obbard DJ, *et al.* (2012) Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Mol Biol Evol* 29(11):3459-3473.

48. Stillwell RC, Dworkin I, Shingleton AW, & Frankino WA (2011) Experimental manipulation of body size to estimate morphological scaling relationships in *Drosophila*. *J Vis Exp* (56):e3162.

49. O'Grady PM (1999) Reevaluation of phylogeny in the *Drosophila obscura* species group based on combined analysis of nucleotide sequences. *Mol Phylogen Evol* 12(2):124-139.

50. R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.

51. Bates D, Maechler M, Bolker B, & Walker S (2013) lme4: linear mixed-effects models using Eigen and S4. R package version 1.0-5. http://CRAN.R-project.org/package=lme4.

52. Bates D & Vazquez AI (2013) pedigreemm: Pedigree-based mixed-effects models. R package version 0.3-1. http://CRAN.R-project.org/package=pedigreemm.