SUPPORTING ONLINE INFORMATION

SUPPORTING MATERIALS AND METHODS.

General Rearing. All lineages were established from the same stock population which was established from ~80 gravid females and has been in culture for ~90 generations (*S1*). All rearing was at 27°C and 12:12 L:D. Unless noted otherwise, larvae were reared in groups not exceeding 450 larvae/cage and fed maize *ad libitum*. Adults were maintained on banana.

Artificial Selection Experiments and Estimation of Genetic Correlation Between Traits

Artificial Selection on Absolute Trait Size. To estimate the heritabilities of body size and forewing area as well as the genetic correlation between them, we undertook two experiments in which we selected to change the absolute value of each of these traits.

Selection on absolute size, i) pupal mass. Three selection categories were established from ~2000 stock pupae: increased pupal mass (+PM), decreased pupal mass (-PM), and an unselected control treatment (Control). Larvae were reared at an initial density of 400 / fabric cage. One day after pupation, individuals were sexed, weighed (0.1 mg), and then housed individually until eclosion. Before selection, pupal weights in each rearing container were standardized by sex to minimize environmental variation, after which the values per lineage were pooled over the containers. Standardized pupal weights were ranked and the 30-35 individuals of each sex with the most extreme masses were selected in each lineage in each generation. Control lineage animals were treated similarly, except that individuals were selected randomly. After generation 12, we performed truncation

selection using the mass of the 20th percentile of the previous generation as the threshold point, but always ensuring at least 30 breeding pairs per lineage. To make estimates comparable to those for forewing area (described below), realized heritabilities were calculated using only the first 6 generations. Individuals from generation 27 (not shown) were used to estimate the genetic correlation between female size and forewing area (FW) (below).

Selection on absolute size, ii) forewing area. Lineages were established by collecting ~2000 eggs from the stock population. Forewing areas (FW) of 750 newly eclosed females were estimated as the area within a polygon defined by one proximal and three distal landmarks (Figure S1). Measurements were made using a camera lucida attached to a digitizing tablet. FW was highly repeatable (0.91) (*S2, S3*). One forewing of each female was marked with a unique number for identification. Females were ranked according to FW and the 30 individuals with the most extreme phenotypes were selected to establish one of two unreplicated selection lines, large-FW (+FW) and small-FW (-FW). A control lineage was established by selecting females at random. After each round of selection, females were caged for ~4 days with 35-40 males from the same lineage for mating after which eggs were collected for 3-5 days. Approximately 10 days after hatching, populations were culled to ~450 larvae each.

Estimation of Genetic Correlation Between Body Size and Forewing Area. Genetic correlations (*r)* between FW and pupal mass were estimated in each selected direction by

calculating the square root of the products of the ratios of the indirect and direct response of the traits to selection (equation 19.7, (*S4*)).

Artificial Selection on Relative Trait Size: Forewing-Body Size Allometry. Our goal was to select for changes in the area of the forewing relative to that of body mass, moving the intercept but not the slope, of the scaling relationship. For logistical reasons, we were unable to measure pupal mass of individuals, therefore we used fresh mass at eclosion as our measure of total body size (BS). Freshmass is highly correlated with pupal mass (Pearson Correlation Coefficient = 0.86, N=208 stock females, p=0.0001). Females were weighed (± 0.01mg) within 24h of eclosion, following discharge of the meconial fluid but before feeding was allowed. Within 4 days, female FW was estimated as in the single trait selection experiment. Following FW measurement, females were placed with males from the same lineage for mating.

To identify individuals for selection, we used Type II Regression (orthogonal regression), which fits a regression line through the bivariate data by minimizing the deviations in both axes. This line is essentially the axis of the first principal component and we interpreted it as representing the mean allometry for a given lineage in each generation. We then quantified the deviation of all female's static allometry from the mean allometry of the population by calculating the perpendicular distance from this regression line to each data point in the cloud. These distributions are essentially the loadings for the second principal component. Females were ranked based on these deviations from the mean allometry and those with the most extreme values in the appropriate direction were

selected in each generation. In this manner, we established and perpetuated lineages selected for large FW relative to BS (+FW/-BS), and for small FW relative to BS (-FW/+BS). Each selected lineage was replicated once. We also established an unreplicated control lineage for which both traits were measured but from which females were selected at random.

Lineages were established by collecting ~2000 eggs from the lab stock population, from which we were able to use 780 females. Forty females were selected to establish the control lineage and each replicated selection lineage. To establish the initial lineages, females were ranked according to their static FW/BS allometries and the 80 females with the most extreme phenotypes (most extreme residuals) in the appropriate direction selected. Alternating females from this ranking were placed into each replicate. Subsequent generations comprised ~300-400 adults/lineage with 30 or 35 females selected each generation. Each generation, the mean allometry for each population was identified by fitting a Type II regression to the standardized bivariate (FW, BS) data. Females were identified for selection by the ranking of their perpendicular distance from the mean allometry of its population; this provides a multivariate measure of the deviation in relative wing and body size for each individual from the population average.

Each generation, the response to selection on the scaling relationships was quantified by calculating the perpendicular distance from the bivariate mean of each replicate to the mean allometry of the Control Lineage. Realized heritabilities of the allometry in each

selected direction were calculated by regressing the response to selection of the replicates in a given direction against the cumulated selection differential.

The independent contribution of each trait to the evolution of the allometry was identified by regression of the univariate means of each lineage against the cumulative indirect selection pressure applied to that trait. Non-zero slopes were interpreted as indicating a change in the size of a trait relative to the control mean. We felt regression was more appropriate than comparison of means in the final generation (13) because our approach shows the overall trend in evolution of the traits (comparison of means at any individual generation provides highly variable results in our experiment because of considerable inter-generational noise in the data).

Regression of the individual trait means across the cumulated indirect selection differential they experienced revealed slopes for FW that differed significantly from zero in all cases (analyses not shown). Because FW and BS did not diverge in lineage F until generation 4 (Figure 3), regressions were conducted after excluding generations 0-2 for this population.

Natural Selection Experiment: Male Phenotype Competition

To estimate the form and strength of external natural selection on the FW/MG allometry, we measured the mating success of competing male butterflies from three phenotype classes, one with typical (wild-type) allometries and the two extreme phenotype classes (+FW/-MW, -FW/+MG).

Creation of male phenotype classes. To minimize effects of female preference for a particular genotype (*S5*), reciprocal crosses were performed (5 males and 5 females per cross) from each of the replicate lineages selected for novel scaling relationships. FW/MG static allometries of the F_1 male hybrids were quantified using the same methodology for lineages as for selection on the allometries. To define the wild-type, control phenotype allometry (see below), males from the control lineage were reared and measured similarly. Males with desired phenotypes were identified for inclusion in the experiment by calculating the difference between their static allometries and the mean allometry of the control lineage males. The 30 males with the most extreme static allometries were placed into one of two phenotype categories: +FW/-MG or -FW/+MG (referred to collectively as 'treatment males'). Fifteen males from each hybrid lineage that had static allometries close to that of the mean allometry of the control lineage males were placed in a third, wild-type phenotype category. In this manner, we generated a control group that had scaling relationships as close as possible to those of wild-type males, but with genotypes drawn equally from the same genetic backgrounds as those of the treatment males.

In each trial, phenotypes were randomly assigned one of three colors: blue, orange, or yellow. Fluorescent powder of this color was applied to the genitalia of each male in that class. This powder is transferred during copulation and thus can be used to identify the class of male with which a female mated (*S5*).

Fitness test in a natural setting. After marking, males were released (Day 0) into a spacious, tropical, naturally-planted greenhouse in which they interacted as if they were in the wild. Six feeding stations of water and banana were placed at regular intervals within the greenhouse. Ninety unmated females from the control lineage were released on Day 1, and an additional 30 on Day 2, creating a realistic butterfly density (*S6*). All butterflies were recaptured on the afternoon of Day 3 and the morning of Day 4. Butterflies were frozen at capture and inspected under blacklight (*S5*) to identify the phenotype class of recaptured males, and the class with which individual females had mated. The experiment was replicated once. Male survival (recapture frequency) and fitness (proportions of mating) were compared among phenotype classes by G-test with the expectation of no differences among groups.

SUPPORTING TEXT

Diversity in Lepidoptera Wing Loading

Despite very high wing loading relative to other insects, Lepidoptera still exhibit substantial diversity among lineages in relative wing size (Figure S2).

Ecological Morphology of Phenotype Classes

Preliminary studies of flight performance using high speed photography indicate that butterflies from the three phenotype classes differ in wing beat frequency and flight efficiency speed during take-off. unpublished data, W. A. Frankino, M. Kortoff, W. Kreiken, J. J. Videler, and P. M. Brakefield.

SUPPORTING FIGURES & LEGENDS

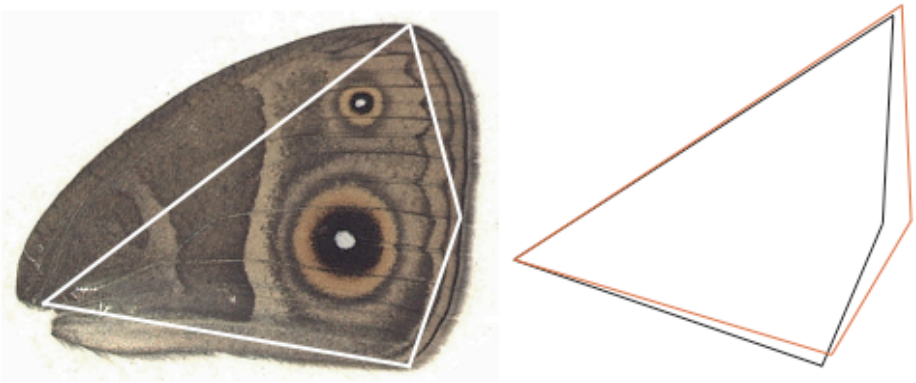

Figure S1. Estimation of forewing area and mean polygons of forewings from each selected direction. A, wing sizes were estimated by calculating the area within the polygons (white lines), defined by four landmarks (vertices of polygons). B, mean polygons from selected directions, scaled relative to control lineage polygon. Orange polygon is +FW/-MG direction and black polygon is -FW/+MG.

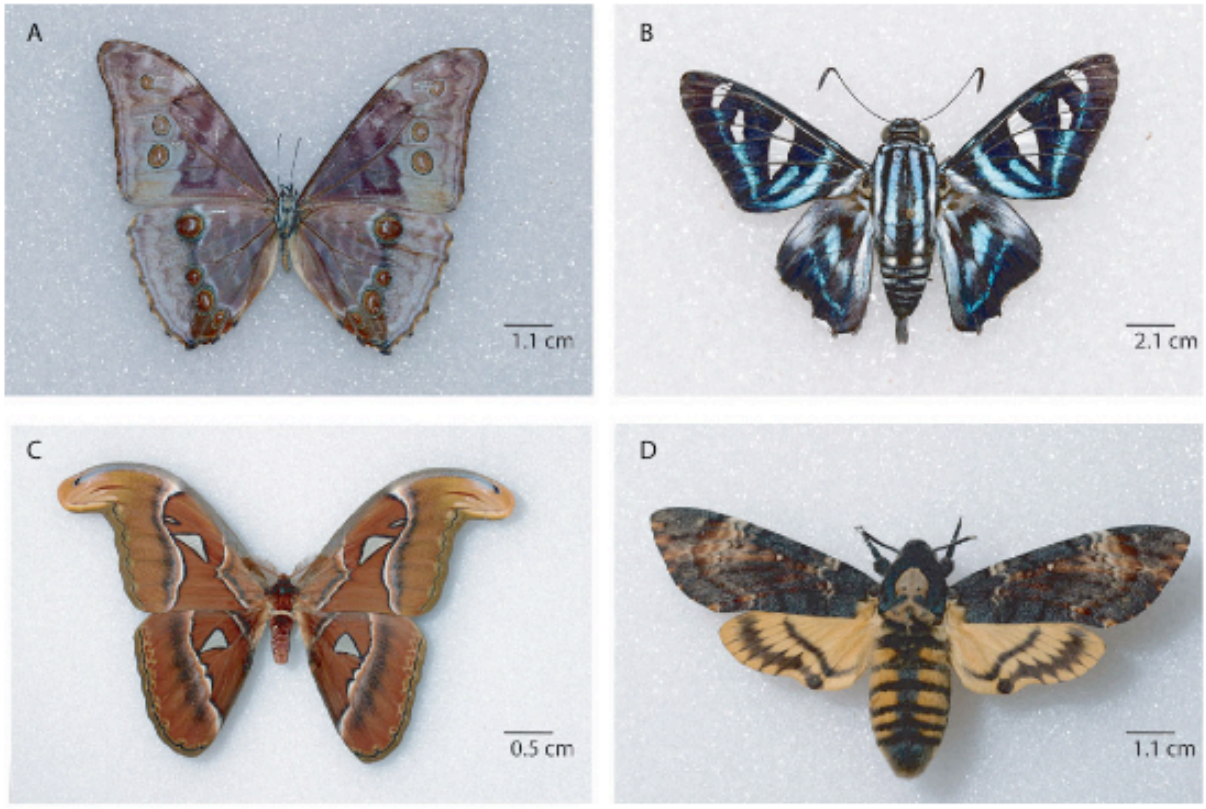

Figure S2. Evolution of wing loading in Butterflies (A, B) and moths (C, D). A, *Morpho aurora* (ventral view); B, *Attacus taprobanis*; C, *Jemadia fallax*; D, *Acherantia atropos*. A and C have low wing loading whereas B and D have very high wing loading. Photocredit: H. Berkhoudt.

SUPPORTING REFERENCES

S1. P. M. Brakefield *et al*., *Journal of Evolutionary Biology* **14**, 148 (2001).
S2. W. A. Becker, *Manual of Quantitative Genetics* (Academic Enterprises, Pullman, 1992), pp.
S3. C. M. Lessels, P. T. Boag, *Auk* **104**, 116 (1987).
S4. D. S. Falconer, T. F. C. MacKay, *Introduction to Quantitative Genetics* (Addison Wesley Longman Limited, Essex, ed. 4, 1997), pp.
S5. M. Joron, P. M. Brakefield, *Nature* **424**, 191 (2003).
S6. J. J. Windig, P. M. Brakefield, P. M. Reitsma, J. G. M. Wilson, *Ecological Entomology* **19**, 285 (1994).