LETTER

doi:10.1038/nature23473

Mutation predicts 40 million years of fly wing evolution

David Houle[1], Geir H. Bolstad[1,2], Kim van der Linde[1,3] & Thomas F. Hansen[4]

Mutation enables evolution, but the idea that adaptation is also shaped by mutational variation is controversial[1–4]. Simple evolutionary hypotheses predict such a relationship if the supply of mutations constrains evolution[5,6], but it is not clear that constraints exist, and, even if they do, they may be overcome by long-term natural selection[7]. Quantification of the relationship between mutation and phenotypic divergence among species will help to resolve these issues. Here we use precise data on over 50,000 Drosophilid fly wings to demonstrate unexpectedly strong positive relationships between variation produced by mutation, standing genetic variation, and the rate of evolution over the last 40 million years. Our results are inconsistent with simple constraint hypotheses because the rate of evolution is very low relative to what both mutational and standing variation could allow. In principle, the constraint hypothesis could be rescued if the vast majority of mutations are so deleterious that they cannot contribute to evolution, but this also requires the implausible assumption that deleterious mutations have the same pattern of effects as potentially advantageous ones. Our evidence for a strong relationship between mutation and divergence in a slowly evolving structure challenges the existing models of mutation in evolution.

Arguments in favour of the idea that mutation may constrain evolution rest on indirect evidence. Theory demonstrates that the rate of evolutionary change is potentially constrained by the availability of mutational variation[8]. There is empirical evidence that the amount of variation that mutation produces varies widely among traits[9,10]. Several studies have found correlations between mutational variation and the amount of standing genetic variation within a population[11–13], and there is often, but not always, rough similarity between standing genetic variation and divergence among species[7]. Several simple models of evolution, summarized in Table 1, assume that mutation constrains evolution[5,7,14,15]. These models all predict that divergence will tend to be greater in directions with more mutational variation, but differ in their predictions of the rate of divergence among species, scaling of divergence rate with mutation, and phylogenetic heritability (the degree of phenotypic resemblance between related species).

Equally plausible arguments can be made that evolution is not constrained by mutation. The influence of natural selection in evolution is evident in the fit of phenotypes to environments throughout the tree of life; observed mutational variation can fuel orders of magnitude more evolutionary change than is observed on geological time scales, even in rapidly evolving organs such as the human skull[16], suggesting that phenotypes can be rapidly optimized by natural selection. Diversifying natural selection often arises from external ecological circumstances, such as the presence or absence of competitors, predators or resources[17], suggesting that the direction of adaptive evolution could be both idiosyncratic and independent of mutational variation.

Here, we directly estimate the relationship between mutation, rate of species divergence and within-population genetic variation. We summarize variation using variance matrices, consisting of the variances and covariances of traits. Mutation is quantified as the amount of phenotypic variance created by mutation per generation, captured in M matrices. The rate at which differences among species have evolved is summarized in an R matrix. Lande[5] showed that within-population additive genetic variance, summarized in a G matrix, mediates the possible relationship between M and R. G predicts the short-term response to selection and potentially long-term evolution, and is shaped by the interactions among selection, mutation and other evolutionary forces.

We estimated variation in the locations of 12 landmark vein intersections on wings in the Dipteran family Drosophilidae (Extended Data Fig. 1a), summarized as wing size and 20 independent wing shape variables. Each shape variable represents a specific pattern of relative movement in the locations of vein landmarks, while holding size constant (Extended Data Fig. 1b–d).

Our first result is that all 21 measured aspects of wing form have evolved among Drosophilid flies, consistent with our previous finding that G captures variation in all measured aspects of wing form in *Drosophila melanogaster*[18]. Our estimates of spontaneous mutational variation provide evidence for variation in at least 18 orthogonal trait combinations when measured in homozygous condition (M_{hom}), and 8 in heterozygous condition (M_{het})[10]. The discrepancy is likely to be due to insufficient statistical power in the mutation experiment, as the

Table 1 | Macroevolutionary predictions of mutational constraint models

Evolutionary model[7]	Fitness function	Divergence rate*	Scaling exponent	Phylogenetic heritability
Neutral evolution	Flat	High ($2V_M$†)	**1**	**High**
Fluctuating directional selection	Linear	High	2	**High**
Divergent selection‡	Linear	Very high $>2V_M$	2	Intermediate
BM§ slow‖	Moving optimum¶	**Low**	~0	**High**
BM§ fast‖	Moving optimum	Very high	~0	**High**
White noise**	Moving optimum	**Low**	0	0
OU†† slow‖	Moving optimum	**Low**	~0	Intermediate
OU†† intermediate‖	Moving optimum	**Low**	**0–1**	Low
OU†† fast‖	Moving optimum	**Low**	~0	~0
Observed		**Low**	**~1**	**High**

Results for fly wings are given in the last row ('observed'). Predictions that match the observed results are in bold.
*Per-generation divergence rate averaged over macroevolutionary timescales.
†V_M is the per-generation increase in variance in a single trait due to mutation.
‡Each taxon is subject to selection in a different direction.
§Optima move according to Brownian motion in an unlimited random walk.
‖Speed is the movement of optimum relative to the rate at which the population evolves towards the optimum.
¶All moving optimum models assume that fitness declines with the square of the distance from the multivariate optimum.
**Optima move in a temporally uncorrelated pattern.
††Optima move according to an Ornstein–Uhlenbeck process, which combines Brownian motion with a tendency to return to a single global optimum. OU parameters chosen to approximate the low divergence rate observed.

[1]Department of Biological Science, Florida State University, Tallahassee, Florida 32306, USA. [2]Norwegian Institute for Nature Research (NINA), NO-7485 Trondheim, Norway. [3]Animal Genetics Inc., 1336 Timberlane Road, Tallahassee, Florida 32312, USA. [4]Department of Biology, CEES & EVOGENE, University of Oslo, 0316 Oslo, Norway.

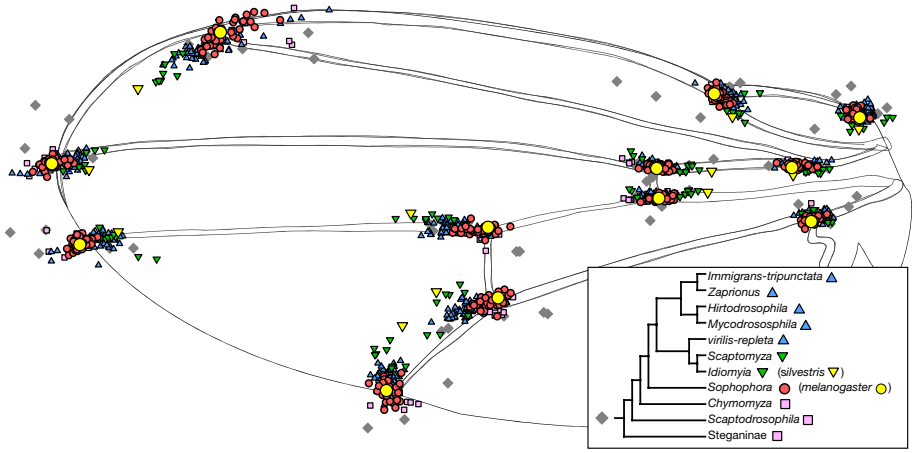


Figure 1 | Mean wing shapes of 112 Drosophilid species, plus five outgroup taxa. Each point represents the Procrustes-aligned position of a single vein intersection in one taxon. The inset shows the phylogeny of the major Drosophilid taxa represented in this study. Grey diamonds indicate phenotypes of outgroup taxa.

existence of genetic variation within and among species in all aspects of wing form must have a basis in mutation. The estimates of R, G, M_{hom} and M_{het} are in Supplementary Tables 2–5.

Our second result is that wings evolve very slowly. Figure 1 shows that, once scaled to the same size, all fly wings are similar, despite diverging over at least 40 million years. Extended Data Fig. 2 shows that shape and size show modest levels of divergence at short time scales but greatly increased divergence over tens of millions of years. The rate of species divergence is remarkably slow compared to the variation within species (Extended Data Table 2). Assuming that the average number of generations per year for *Drosophila* species is 10, the rate of divergence is less than 10^{-4} of the rate that would be generated by genetic drift in the absence of selection. The variance generated by just 187 generations of mutations, as summarized in M_{hom}, is equal to the average variation created by one million years (Myr) of wing evolution. This corresponds to 14 mutational standard deviations. Just half the variance in the G matrix is equal to the evolutionary rate per Myr, an average change of just 0.7 genetic standard deviations. For comparison, artificial-selection experiments on wing shape readily achieve multiple genetic standard deviations of response in tens of generations[19–22].

Our third result is that the phylogenetic relationship between species predicts 90% of their phenotypic similarity, as measured by the phylogenetic heritability[23] (Extended Data Table 3). When variation among species is captured in the two variables that best show divergence, major Drosophilid clades occupy restricted regions of phenotype space (Fig. 2).

Finally, the patterns of mutational, within-species and among-species variation are remarkably similar. This is evident in the relative variation around each landmark position on the wing (Fig. 3). To investigate the similarities in matrix structure more comprehensively, we

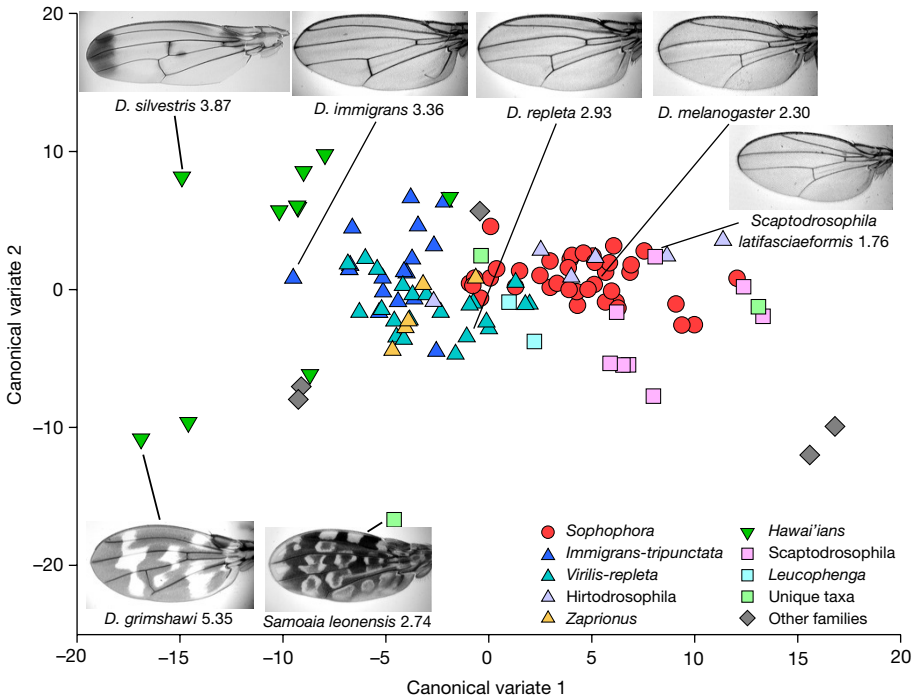


Figure 2 | Species scores on the two directions in shape phenotype space (canonical variables) that explain the most variance among species means. Canonical variate 1 and Canonical variate 2 capture 57% and 14% of the variance among species means, respectively. A representative wing from each species group is shown (numbers are mean centroid size in millimetres). Symbol shapes correspond to the groups in Fig. 1. Hirtodrosophila includes the genera *Hirtodrosophila* and *Mycodrosophila*. Scaptodrosophila includes the genera *Scaptodrosophila* and *Chymomyza*. Unique taxa are *D. busckii*, *Samoaia leonensis*, and *Dettopsomyia nigrovittata*. Grey diamonds indicate phenotypes of outgroup taxa.

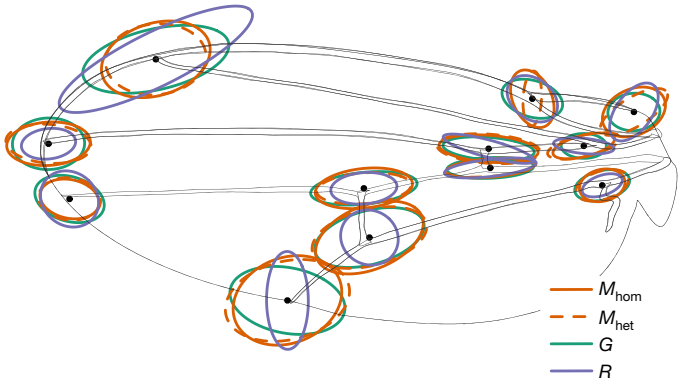


Figure 3 | Ellipses representing the variation around each landmark. Variance matrices were scaled to the same size, so areas of ellipses are relative to the total variance within each matrix.

compared the variance along the same set of directions for each pair of matrices, with the results for shape shown in Fig. 4, and for shape and size in Extended Data Fig. 3. The precision of these relationships is striking, as the proportion of variance explained (R^2) by almost all of these relationships is greater than 90%. The scaling parameters for the influence of mutation on divergence, b, are all within a factor of 0.3 of the proportional value of $b = 1$.

We thus have four main results: all measured wing traits evolve; the rate of evolution is low; phylogenetic heritability is high; and mutation is highly correlated with evolutionary divergence. The largely proportional relationship between mutation and divergence gives apparent support to the idea that mutation constrains evolutionary divergence, but examination of the predictions listed in Table 1 make it clear that none of the mutational constraint models can explain all of our results. For example, the neutral model predicts high phylogenetic heritability and $b = 1$, but a rate of evolution over four orders of magnitude higher than what we observe[5,15,24].

The mutational constraint hypothesis can be rescued if most mutations cause deleterious pleiotropic effects that render them irrelevant to adaptation, and, more importantly, the proportion of mutational variation that is deleterious is similar for all traits. To examine the consistency of this hypothesis, we used conditional evolvability[25], the evolvability achieved when all trait combinations not subject to directional selection are subject to stabilizing selection. This is equivalent to assuming that mutations with effects that are not in the precise direction favoured by selection are deleterious. Conditioning on the wing traits that we measured lowers the predicted evolvability by a factor of about 10 (Extended Data Fig. 4). The relationships based on conditional variation within the wing are still fairly precise, consistent with deleterious pleiotropy affecting all traits to a similar degree. The potential impact of deleterious pleiotropy could be even larger because there is certain to be pleiotropy between wings and traits we did not measure.

Alternatively, our results could be explained if the pattern of mutation is the outcome, rather than the driver, of macroevolutionary processes[26,27]. If phenotypic optima change faster along dimensions of phenotype space that are normally under weaker stabilizing selection, it is conceivable that the genetic architecture itself would evolve so that M matches R. Genetic variation for M exists[28], and some studies have found selective conditions under which M will plausibly be reshaped to more closely match the fitness landscape[29–31]. More explicit models generally cast doubt on this possibility, as M can also evolve counter to the adaptive expectation, depending on the pattern of epistasis[32–34].

We do not know whether the striking pattern of similarity between mutational and among-species variation we observe is found in other traits and taxa. It is possible that the low rate of evolution in fly wings makes them particularly susceptible to whichever processes favour matching mutational and among-species variation. In some cases in

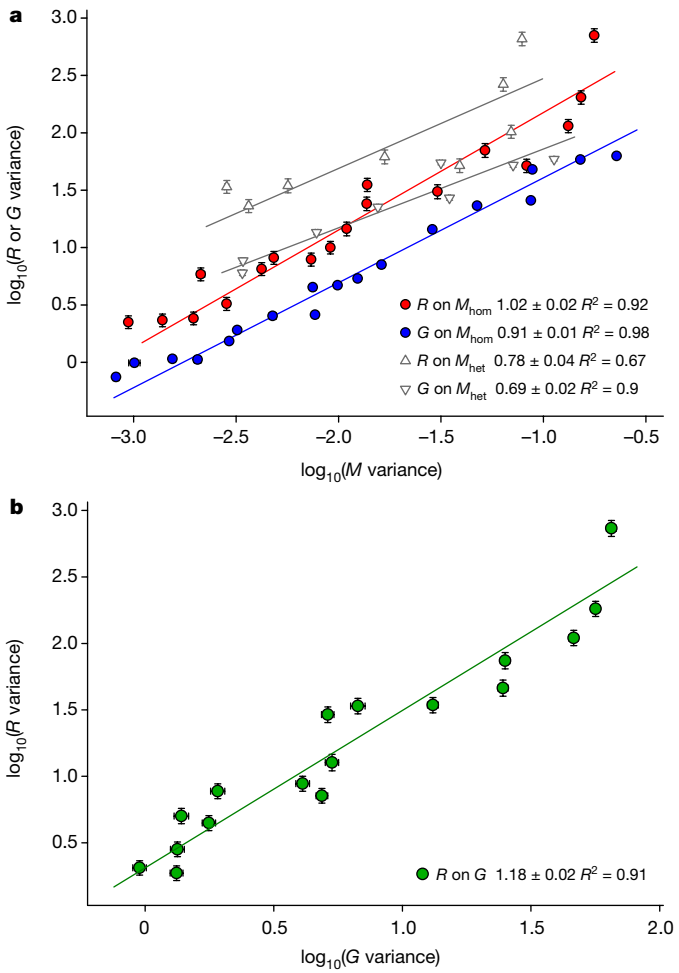


Figure 4 | Relationships between variance in wing shape in M, G and R. Points are $\log_{10}$(variances) ± 2 s.e. along orthogonal directions in phenotype space. Key gives the allometric scaling parameter, b ± s.e. and R^2 for the regression. Lines pass through the bivariate means of the points, and have slopes equal to b. **a**, R and G on M_{hom} and M_{het}. **b**, R on G.

which rapid evolution has taken place there is evidence that within- and among-population variation can be decoupled (for example, ref. 35). It would be particularly interesting to study mutation and genetic variation in relation to divergence in phenotypes with a history of evolutionary innovation.

Our finding of striking similarities among mutational, genetic and among-species variation coupled with a low rate of evolution are not readily explained by any of the available models of evolution. This is an important challenge for evolutionary theorists.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 November 2016; accepted 4 July 2017.
Published online 9 August 2017.

1. Arnold, S. J., Pfrender, M. E. & Jones, A. G. The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* **112–113,** 9–32 (2001).
2. Hansen, T. F. & Houle, D. in *The Evolutionary Biology of Complex Phenotypes* (eds M. Pigliucci & K. Preston) 130–150 (Oxford Univ. Press, 2004).
3. Blows, M. W. & Hoffmann, A. A. A reassessment of genetic limits to evolutionary change. *Ecology* **86,** 1371–1384 (2005).
4. Futuyma, D. J. Evolutionary constraint and ecological consequences. *Evolution* **64,** 1865–1884 (2010).
5. Lande, R. Quantitative genetic analysis of multivariate evolution applied to brain:body size allometry. *Evolution* **33,** 402–416 (1979).
6. Wagner, G. P. & Altenberg, L. Perspective: complex adaptations and the evolution of evolvability. *Evolution* **50,** 967–976 (1996).

7. Bolstad, G. H. *et al.* Genetic constraints predict evolutionary divergence in *Dalechampia* blossoms. *Phil. Trans. R. Soc. Lond. B* **369,** 20130255 (2014).
8. Hill, W. G. Rates of change in quantitative traits from fixation of new mutations. *Proc. Natl Acad. Sci. USA* **79,** 142–145 (1982).
9. Houle, D., Morikawa, B. & Lynch, M. Comparing mutational variabilities. *Genetics* **143,** 1467–1483 (1996).
10. Houle, D. & Fierst, J. Properties of spontaneous mutational variance and covariance for wing size and shape in *Drosophila melanogaster*. *Evolution* **67,** 1116–1130 (2013).
11. Lynch, M., Latta, L., Hicks, J. & Giorgianni, M. Mutation, selection, and the maintenance of life-history variation in a natural population. *Evolution* **52,** 727–733 (1998).
12. Houle, D. How should we explain variation in the genetic variance of traits? *Genetica* **102–103,** 241–253 (1998).
13. Farhadifar, R., Ponciano, J. M., Andersen, E. C., Needleman, D. J. & Baer, C. F. Mutation is a sufficient and robust predictor of genetic variation for mitotic spindle traits in *Caenorhabditis elegans*. *Genetics* **203,** 1859–1870 (2016).
14. Schluter, D. Adaptive radiation along genetic lines of least resistance. *Evolution* **50,** 1766–1774 (1996).
15. Hansen, T. F. & Martins, E. P. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* **50,** 1404–1417 (1996).
16. Lynch, M. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *Am. Nat.* **136,** 727–741 (1990).
17. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
18. Houle, D. & Meyer, K. Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood. *J. Evol. Biol.* **28,** 1542–1549 (2015).
19. Weber, K. E. Selection on wing allometry in *Drosophila melanogaster*. *Genetics* **126,** 975–989 (1990).
20. Weber, K. E. How small are the smallest selectable domains of form? *Genetics* **130,** 345–353 (1992).
21. Houle, D., Mezey, J., Galpern, P. & Carter, A. Automated measurement of *Drosophila* wings. *BMC Evol. Biol.* **3,** 25 (2003).
22. Bolstad, G. H. *et al.* Complex constraints on allometry revealed by artificial selection on the wing of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **112,** 13284–13289 (2015).
23. Lynch, M. Methods for the analysis of comparative data in evolutionary biology. *Evolution* **45,** 1065–1080 (1991).
24. Lynch, M. & Hill, W. G. Phenotypic evolution by neutral mutation. *Evolution* **40,** 915–935 (1986).
25. Hansen, T. F. & Houle, D. Measuring and comparing evolvability and constraint in multivariate characters. *J. Evol. Biol.* **21,** 1201–1219 (2008).
26. Riedl, R. A systems-analytical approach to macro-evolutionary phenomena. *Q. Rev. Biol.* **52,** 351–370 (1977).
27. Cheverud, J. M. Quantitative genetics and developmental constraints on evolution by selection. *J. Theor. Biol.* **110,** 155–171 (1984).
28. Pavličev, M. & Cheverud, J. M. Constraints evolve: context dependency of gene effects allows evolution of pleiotropy. *Annu. Rev. Ecol. Evol. Syst.* **46,** 413–434 (2015).
29. Draghi, J. & Wagner, G. P. Evolution of evolvability in a developmental model. *Evolution* **62,** 301–315 (2008).
30. Pavličev, M., Cheverud, J. M. & Wagner, G. P. Evolution of adaptive phenotypic variation patterns by direct selection for evolvability. *Proc. R. Soc. Lond. B* **278,** 1903–1912 (2011).
31. Jones, A. G., Bürger, R. & Arnold, S. J. Epistasis and natural selection shape the mutational architecture of complex traits. *Nat. Commun.* **5,** 3709 (2014).
32. Hansen, T. F., Alvarez-Castro, J. M., Carter, A. J. R., Hermisson, J. & Wagner, G. P. Evolution of genetic architecture under directional selection. *Evolution* **60,** 1523–1536 (2006).
33. Hansen, T. F. in *Epigenetics: Linking Genotype and Phenotype in Development and Evolution* (eds B. Hallgrimsson & B. K. Hall) 357–376 (Univ. California, 2011).
34. Le Rouzic, A., Álvarez-Castro, J. M. & Hansen, T. F. The evolution of canalization and evolvability in stable and fluctuating environments. *Evol. Biol.* **40,** 317–340 (2013).
35. Grabowski, M. W., Polk, J. D. & Roseman, C. C. Divergent patterns of integration and reduced constraint in the human hip and the origins of bipedalism. *Evolution* **65,** 1336–1356 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Galpern, J. Birdsley, F. S. Hollis, Y. Ng, and L. Carpenter for rearing and imaging Drosophilid species; C. Boake, J. David, G. Gilchrist, J. Hey, A. Hoikalla, L. Reed and J. True for contributing stocks, specimens or images; K. Meyer for assistance with Wombat; S. Steppan for help with r8s and phylogenetic analyses; J. Merilä and M. Kirkpatrick for comments; and to the many students who measured fly wings. This work was supported by US NSF DEB grants 0129219 and 0950002, and NSERC grants to D.H.

Author Contributions D.H. conceived the project and supervised the gathering of data, K.v.d.L. wrote software and assembled the species dataset, T.F.H., G.H.B. and D.H. derived the major theoretical framework for the analyses, D.H. and G.H.B. analysed the data, and D.H., T.F.H. and G.H.B. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to D.H. (dhoule@bio.fsu.edu).

Reviewer Information *Nature* thanks J. Cheverud and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Datasets. We combine three datasets in this analysis: a new species dataset, R, used to estimate the rate of among-species divergence, *R*; and two previously published datasets, M and G. The 12,075-wing M dataset was used to estimate the homozygous (M_{hom} total) and heterozygous (M_{het} non-segregational) mutational variance for wing shape arising in two different inbred *D. melanogaster* genotypes[10]. The 17,323-wing G dataset was used to estimate the additive genetic variance, *G*, in a single outbred *D. melanogaster* population from Wabasso, Florida, USA[18,36].

For the R dataset, we measured 21,138 wings from 117 taxa. These include the 111 samples of taxa listed in Supplementary Table 6 in ref. 22, plus an additional six taxa listed in Extended Data Table 1. Of these taxa, 112 are from the family Drosophilidae, and the remaining five are from the Acalyptrate families Ephydridae, Lauxaniidae and Chloropidae. Our Drosophilids are overwhelmingly drawn from the subfamily Drosphilinae, with just two taxa from the subfamily Steganinae. Species were collected from the wild or obtained from the *Drosophila* Species Stock Center, or from other collectors. Flies were reared using combinations of food, temperature, and rearing environments 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 laboratory. Four Drosophilid species are represented by more than one subspecies. The mean sample size per taxon was 181. Eighty-five species had samples of at least 181 wings. The minimum sample size was nine wings.

No statistical methods were used to predetermine sample size. Observers were not blinded to the identity of specimens.

Phylogenetic hypothesis. Our phylogenetic hypothesis for these taxa is the partitioned maximum-likelihood phylogram given in Fig. 3 of ref. 37, including *Ceratitis capitata* (Tephritidae), *Ephydra riparia, Neogriphoneura sordida* and *Discocerina obscurella* (Ephydridae) as outgroups. We omitted the *N. sordida* wing data from the comparative analysis because this species diverged well before the other taxa. All species in our phylogeny are members of the Dipteran section Schizophora.

To convert this phylogeny to estimates of divergence time, we estimated an ultrametric tree using penalized likelihood in the program r8s[38]. We estimated the best smoothing parameter to be 10 using cross-validation, allowing considerable variation in rates across the phylogeny. We analysed evolutionary rates assuming a minimum age of 23 Myr for the divergence between *Scaptomyza* and Hawi'ian *Drosophila (Idiomyia)*, and a maximum age of the 'picture-winged' Hawi'ian clade of 5.1 Myr[39]. These constraints by themselves yielded unrealistically old divergence times. There has been considerable disagreement as to the age of the Drosophilidae[39–41] and its containing clade, the Schizophorans, so we conducted separate analyses assuming that the last common ancestor of the Schizophorans lived a maximum of either 75 or 150 Myr ago, yielding divergence times for the family Drosophilidae of 39.5 or 66.4 Myr, respectively. While the overall rate of evolution was influenced by this difference in assumed divergence time, all other results of this study were otherwise very similar. All results presented assume the more recent divergence time of 39.5 Myr for the family, based on the evidence that the radiation of the Schizophoran clade was post-Cretaceous[42,43].

To obtain the final relationship matrix, we omitted taxa not in the family Drosophilidae, then re-estimated the optimal smoothing parameter to be 1. The r8s estimation was repeated on this smaller dataset, constraining the divergence time of the family to 39.5 Myr. The final estimates of divergence times are given in Supplementary Table 1.

We supplemented the phylogeny by inserting three species based on other evidence. *D. athabasca* was added based on the phylogeny of ref. 44. *D. arizonae* and *D. anceps* were added based on our super-tree analysis[45], supplemented with other published phylogenies[46–48]. Six taxa for which wing data are shown in Fig. 1 are missing from this phylogeny: one species in the Drosophilid subfamily Drosophilinae (*Samoaia leonensis)*, one of our two species in the Drosophilid subfamily Steganinae (*Leucophenga* sp.) and the four taxa from families Lauxaniidae and Chloropidae.

Wing measurements. We used a semi-automated procedure for imaging wings and for estimating vein locations and corresponding landmarks[21]. The user immobilizes the left wing of a live CO_2-anaesthetized fly with a suction device and takes a digital image of the wing. We fit a series of cubic B-splines to the wing veins in the resulting image in the program Wings 3.72 (ref. 49). The *x* and *y* coordinates of 12 landmark vein intersections were extracted from the fitted splines (Extended Data Fig. 1a).

Shape data were geometrically aligned to eliminate variation due to differences in the rotation and translation of the images using Procrustes least-squares superimposition[50] within each species, then subjected to the robust Minimum Volume Ellipsoid[51,52] algorithm to detect multivariate outliers. The fit of the splines to the corresponding image was examined for all outliers, and corrected when necessary using an interactive module in Wings 3.72. Damaged wings and truly aberrant wings that were more than four robust s.d. away from the mean were excluded. All but one taxon in our sample have the same vein topology with five longitudinal veins and two crossveins. The exception is the Hawi'ian species *D. silvestris*, which has an additional crossvein, as shown in Fig. 2, as well as dark clouds on the wings that made automated splines fit poorly. Splines for this species were substantially fit by hand, starting from the approximate fits obtained in Wings 3.72.

The previously published M[10] and G[36] datasets were gathered using the same landmark definitions and techniques.

Analysis of wing shapes. The M, G and R datasets were then simultaneously aligned[50] with an even larger dataset of 83,000 wings that also includes wings from experiments that are not part of this study. Although the superimposed data are still in the form of 12 pairs of *x* and *y* coordinates, three degrees of freedom are used for superimposition, so the resulting variance matrices have a maximum rank or dimensionality of 21—centroid size and 20 shape variables. Centroid size was measured in millimetres, then natural log-transformed before analysis[53]. All shape variables and ln(size) were then multiplied by 1,000. We carried out analyses in the 21-dimensional shape-size space then back-transformed parameter estimates into the wing shape space for interpretation and visualization. Examples of these back-transformed shape variables are shown in Extended Data Fig. 1b–d.

Canonical variate analysis. To visualize the relationship of phenotypes among species in Fig. 2, we used canonical discriminant analysis in the procedure CANDISC in SAS 9.3 (ref. 54). We used taxon as the classification variable, then scored the phenotype of individuals on the first two canonical variables. The first two canonical variables were nearly orthogonal (vector correlation $r = 0.03$) and similar to the first two principal components of the among-species covariance matrix ($r = 0.83$ PC1 vs CV1; $r = 0.84$ PC2 vs CV2).

Phylogenetic heritability. We partitioned the among-species variance into that which can be explained by a Brownian-motion model of evolution V_B, a phylogenetically uncorrelated residual V_U, and within-species variance Vw using the the maximum-likelihood mixed-model program Wombat[55]. The full multivariate model would not converge, so we analysed each of the 21 traits in univariate analyses. Phylogenetic heritability[23], H^2, calculated as $V_B/(V_B + V_U)$, is a measure of the expected correlation of related taxa. To compute an average H^2 over sets of traits, we weighted each variable by its among-species variance. H^2 estimates are shown in Extended Data Table 3.

Estimating variance matrices. We estimated variance matrices in Wombat[55]. In each case, we estimated pooled-sex variance matrices by including the sex of the fly as a fixed effect. Estimation of matrices was carried out for both full- and reduced-rank models[56–58], and we selected the best-fitting rank model on the basis of Akaike's Information Criterion corrected for small sample size (AICc). We obtained 1,000 matrix estimates from the sampling distribution of each matrix using the restricted maximum likelihood, multivariate normal (REML-MVN) approach[18]. Matrices were sampled in a transformed space (the L-scale[18]) to ensure that each sampled estimate was of the same rank as the best estimate. To estimate the rate of divergence across our phylogeny, we used a phylogenetic mixed model[23,59] implemented in Wombat using a relationship matrix based on the divergence times in Supplementary Table 1. A full-rank model of species divergence in all 21 possible phenotype dimensions fits more than 2,000 AICc units better than lower-dimension models.

The homozygous (M_{hom} total) and heterozygous (M_{het} non-segregational) mutational variance matrices were estimated using a single analysis of the data from two different inbred genotypes, fitting a fixed effect for genotype and sex in each case[10].We re-estimated the standing additive genetic variance matrix *G* as previously described[18], with the addition of natural log-transformed centroid size as a 21st variable. Block and sex were fit as fixed effects.

Comparison of variation across the phenotype space. M_{hom} and M_{het} are both less than full rank, so we used 'common subspace analysis'[10] to find a subspace in which to compare each *M* matrix with the better-estimated *G* or *R* matrix. Common-subspace analysis simultaneously maximizes the proportion of variance of both matrices that is captured in the common subspace, and ensures that the variance in all directions in the common subspace is greater than the minimum eigenvalue that was statistically supported for each matrix[10]. We used the trace (sum of the diagonal elements) of each matrix in the common subspace as a measure of overall variation. Proportionality constants were calculated as the ratio of the traces.

To compare variation in each pair of matrices, we scaled them to the same size by multiplying one of them by the appropriate proportionality constant, then computed the average matrix, *H*. We decomposed *H* into its eigenvectors, and took the resulting eigenvectors K_H as directions along which to compare variation. The vector of variances along each of these eigenvectors was calculated from each matrix estimate, *C*, as the diagonal of $K_H{}^T C K_H$, where T indicates transposition. The directions K_H are not in themselves particularly meaningful, but collectively

provide a set of orthogonal directions that cover the entire phenotype space. For G, the values of this vector are evolvabilities in the sense of ref. 25. For the M matrices, these variances are related to the evolvability that would arise in a population segregating for the mutations that arise each generation. Because our estimates of M were not full rank, we include just the r directions in which the corresponding M matrix was well-estimated.

Conditional variance in a direction is uncorrelated with variation in a set of orthogonal directions[25,60]; for these calculations, we assumed that when one direction was under directional selection, all other orthogonal directions were constrained by stabilizing selection. Conditional variance is zero in a matrix of less than full rank, so we calculated it in the r rank space common to the compared matrices as follows. The conditional variances are the inverses of the diagonal elements of $C^{-1}_{[r,\,r]}$, where

$$C_{[r,\,r]} = K_{H[n,\,r]}{}^{T} C K_{H[n,\,r]}$$

and $K_{H[n,\,r]}$ is an $n \times r$ matrix consisting of the first r columns of K_H.

The scaling relationships between R and M or G were estimated using a phylogenetic mixed model assuming that Brownian motion explains all of the phylogenetic signal. The r supported orthogonal directions of K_H defined the traits in this analysis, and the mean of trait i in species j was given by

$$y_{ij} = \mu_i + d_{ij} + m_{ij}$$

where μ is the (weighted) average of the trait across the phylogeny, d is the phylogenetic effect and m combines sampling and measurement errors. The vector of measurement errors was assumed to have a multivariate Gaussian distribution $\boldsymbol{m} \sim \mathcal{N}(\mathbf{0}, P \otimes S)$, where $\mathcal{N}$ represents normal distribution, $\sim$ denotes 'is distributed as', $\otimes$ is the Kronecker product, P is the variance matrix of the average within-species phenotypic distribution, and S is a diagonal matrix with the inverse of species sample sizes minus one along the diagonal. The vector of phylogenetic effects was assumed to be distributed as $\boldsymbol{r} \sim \mathcal{N}(\mathbf{0}, R \otimes A)$,where A is a matrix with elements equal to the branch lengths shared between species. Our analysis differs from standard phylogenetic mixed models in that the rate matrix has off-diagonal elements set to zero, and the ith diagonal element specified by the log-linear relationship $\log_{10}(R_{ii}) = \log_{10}(k) + b\log_{10}(Q_{ii})$, where $\log_{10}(k)$ is the intercept and b is the slope of the scaling relationship, R_{ii} is the among-species evolutionary rate in the ith direction, and Q_{ii} is the predictor variance (conditional or unconditional, depending on the analysis) in the same direction. The scaling relationship between G and M was estimated by the same model, in which full-sib family means were replaced by the species means, A was the additive genetic relatedness among full-sib families, P was the average within-family variance matrix, and N was based on the sample size of each full-sib family. To fully implement the half-sib breeding design, which was carried out in 36 temporal blocks[36], we included block as a fixed effect in this analysis.

We corrected for sex differences before the analyses by centring the raw data on global sex means. All scaling-relationship models were fitted using the Template Model Builder[61] implemented in R[62]. Values were $\log_{10}$-transformed before analysis. The magnitude of variation in parameters was calculated from 1,000 randomly paired replicate estimates of each matrix. For comparison, we calculated the corresponding least-squares regressions (results not shown), which gave slopes consistent with, but generally slightly lower than, the scaling relationship fit with the mixed model.

The slope estimates are slightly biased owing to the variance in estimating the predictor variances. Under standard linear model assumptions and the classical model of measurement error[63,64], the bias will be

$$\frac{\hat{\beta} - \beta}{\beta} \approx -\frac{\sigma^2_{\mathrm{u}}}{\sigma^2_{\mathrm{X}}}$$

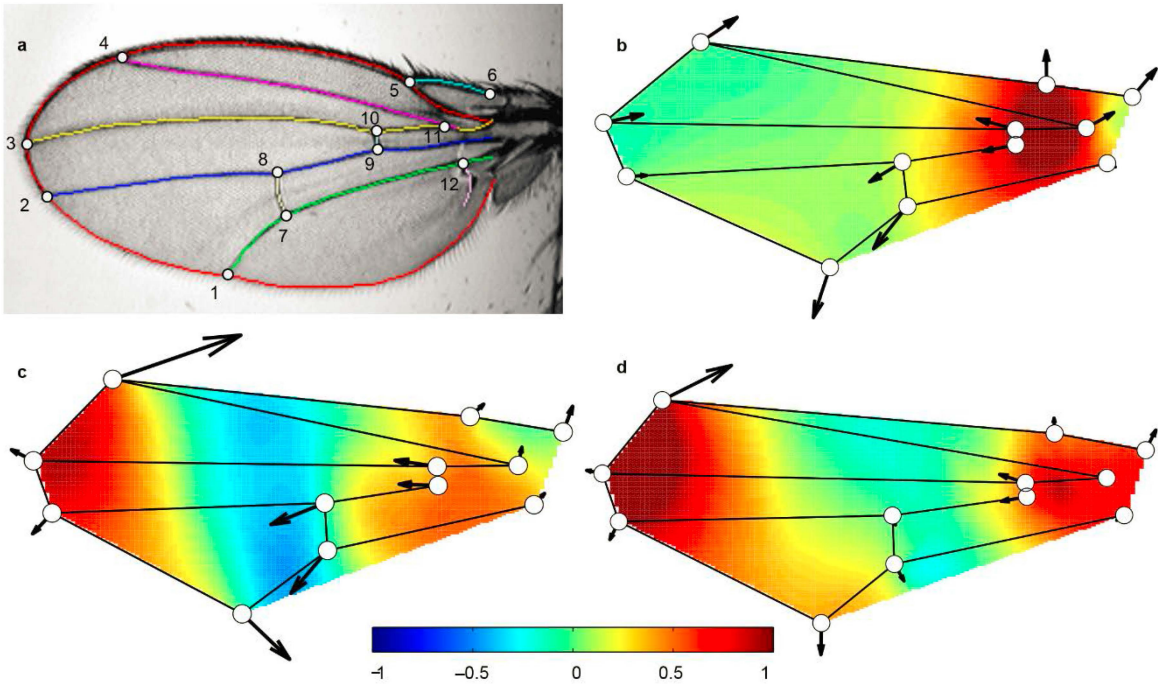
where β is the true value of the slope, $\hat{\beta}$ is our estimate, σ^2_{u} is the pooled error variance in the predictors, and σ^2_{X} is the underlying variance of the predictors themselves. When the unconditional variances are used as predictors the bias is trivial for M (-0.04% for M_{hom} and -0.01% for M_{het}) and -1.5% for G. For regressions on conditional variances, the bias is -2.2% for M_{hom}, -0.7% for M_{het} and -0.4% for G.

We attempted to fit multiple regressions in which both M and G were used to explain R. Unconditional variances in M and G were highly collinear (variance inflation factor 74), so multiple regression yielded slope estimates with high standard errors (results not shown).

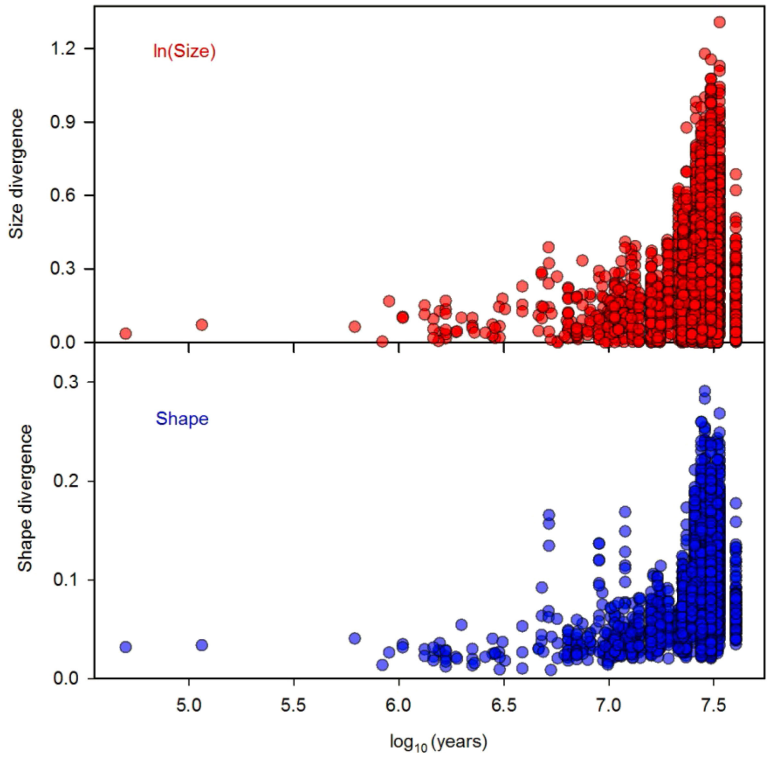
Code availability. Code used for extraction of wing data, and for statistical analyses in SAS, R and Template Model Builder is available upon request from the authors. Executable versions of the Wings program are available at http://www.bio.fsu.edu/~dhoule/software.html.

Data availability. The data on which this study is based are archived in Dryad (doi:10.5061/dryad.3b7g5, doi:10.5061/dryad.hb37q and doi:10.5061/dryad.5fr18). Source Data for Figs 1–4 are provided with the online version of the paper.

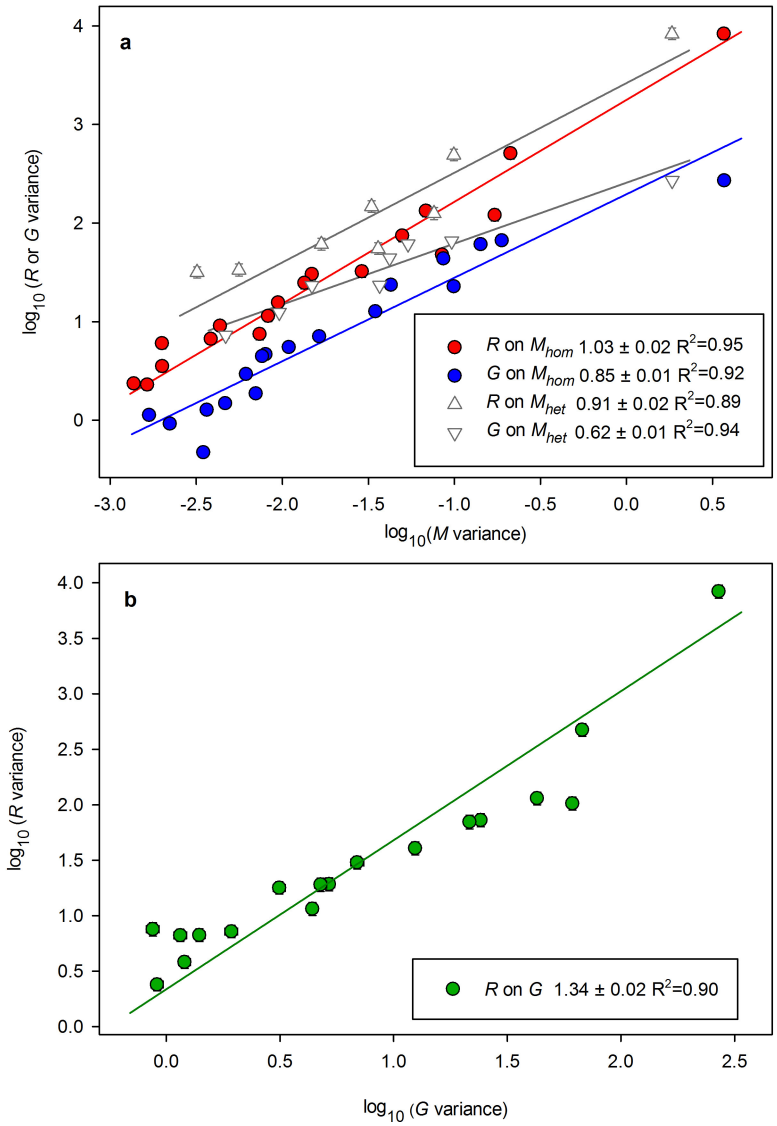
36. Mezey, J. G. & Houle, D. The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. *Evolution* **59,** 1027–1038 (2005).
37. van der Linde, K., Houle, D., Spicer, G. S. & Steppan, S. J. A supermatrix-based molecular phylogeny of the family Drosophilidae. *Genet. Res.* **92,** 25–38 (2010).
38. Sanderson, M. J. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* **19,** 101–109 (2002).
39. Obbard, D. J. *et al.* Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Mol. Biol. Evol.* **29,** 3459–3473 (2012).
40. Russo, C. A. M., Takezaki, N. & Nei, M. Molecular phylogeny and divergence times of drosophilid species. *Mol. Biol. Evol.* **12,** 391–404 (1995).
41. Tamura, K., Subramanian, S. & Kumar, S. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol. Biol. Evol.* **21,** 36–44 (2004).
42. Grimaldi, D. & Engel, M. S. *Evolution of the Insects* (Cambridge Univ. Press, 2005).
43. Wiegmann, B. M. *et al.* Episodic radiations in the fly tree of life. *Proc. Natl Acad. Sci. USA* **108,** 5690–5695 (2011).
44. O'Grady, P. M. Reevaluation of phylogeny in the *Drosophila obscura* species group based on combined analysis of nucleotide sequences. *Mol. Phylogenet. Evol.* **12,** 124–139 (1999).
45. Van der Linde, K. & Houle, D. A supertree analysis and literature review of the genus *Drosophila* and closely related genera (Diptera, Drosophilidae). *Insect Syst. Evol.* **39,** 241–267 (2008).
46. Reed, L. K., Nyboer, M. & Markow, T. A. Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*. *Mol. Ecol.* **16,** 1007–1022 (2007).
47. Stensmyr, M. C., Stieber, R. & Hansson, B. S. The Cayman crab fly revisited—phylogeny and biology of *Drosophila endobranchia*. *PLoS One* **3,** e1942 (2008).
48. Oliveira, D. C. S. G. *et al.* Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the *Drosophila repleta* species group. *Mol. Phylogenet. Evol.* **64,** 533–544 (2012).
49. *Wings: Automated Capture of* Drosophila *Wing Shape* v.3.72 http://bio.fsu.edu/~dhoule/wings.html (2004–2009).
50. Rohlf, F. J. & Slice, D. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst. Zool.* **39,** 40–59 (1990).
51. Rousseeuw, P. J. & van Zomeren, B. C. Unmasking multivariate outliers and leverage points. *J. Am. Stat. Assoc.* **85,** 633–639 (1990).
52. *MVE*: *Minimum Volume Ellipsoid Estimation for Robust Outlier Detection in Multivariate Space* v. Java http://www.kimvdlinde.com/professional/mve.html (2004).
53. Mitteroecker, P. & Bookstein, F. The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans. *Evolution* **63,** 727–737 (2009).
54. *The SAS System for Windows* Release 9.3 (SAS Institute, 2011).
55. *Wombat*: *A Program for Mixed Model Analyses by Restricted Maximum Likelihood* (Animal Genetics and Breeding Unit, University of New England 2010–2015).
56. Kirkpatrick, M. & Meyer, K. Direct estimation of genetic principal components: simplified analysis of complex phenotypes. *Genetics* **168,** 2295–2306 (2004).
57. Meyer, K. & Kirkpatrick, M. Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices. *Genet. Sel. Evol.* **37,** 1–30 (2005).
58. Meyer, K. & Kirkpatrick, M. Perils of parsimony: properties of reduced-rank estimates of genetic covariance matrices. *Genetics* **180,** 1153–1166 (2008).
59. Hadfield, J. D. & Nakagawa, S. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *J. Evol. Biol.* **23,** 494–508 (2010).
60. Hansen, T. F., Armbruster, W. S., Carlson, M. L. & Pélabon, C. Evolvability and genetic constraints in *Dalechampia* blossoms: genetic correlations and conditional evolvability. *J. Exp. Zool. B Mol. Dev. Evol.* **296B,** 23–39 (2003).
61. Kristensen, K., Nielsen, A., Berg, C. & Skaug, H. Template model builder TMB. *J. Stat. Softw.* **70,** 1–21 (2015).
62. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2016).
63. Fuller, W. A. *Measurement Error Models* (Wiley, 1987).
64. Hansen, T. F. & Bartoszek, K. Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. *Syst. Biol.* **61,** 413–425 (2012).



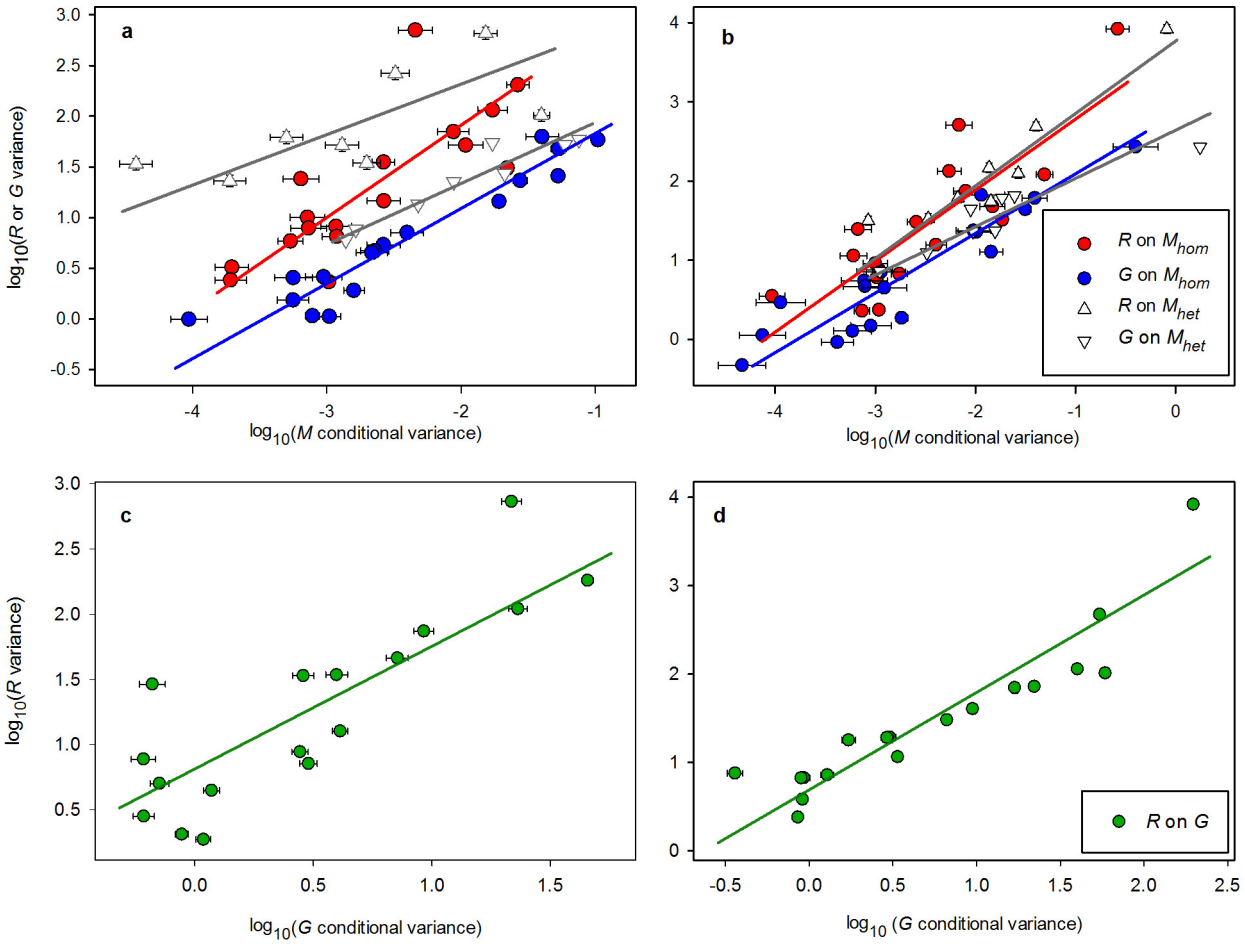
Extended Data Figure 1 | Wing landmarks measured and representation of shape changes. a, Vein model fitted to a *D. melanogaster* wing. The coordinates of the twelve vein intersections shown are the data for this study. **b**–**d**, Shape vectors for PC1 of the matrices. **b**, M_{hom} matrix. **c**, *G* matrix. **d**, *R* matrix. Each vector represents a pattern of changes in the locations of landmark intersections, represented by the arrows. The colours represent the pattern of landmark movements as local expansions and contractions that can explain the movements of the landmarks. The scale of local changes is in $\log_2$ units, so −1 represents a halving of local area, and +1 a local doubling.



Extended Data Figure 2 | Divergence of wing phenotypes as a function of the time since the most common ancestor. Size divergence (top) is the absolute value of the difference in log-transformed centroid size. Shape divergence (bottom) is measured as Procrustes distance.



Extended Data Figure 3 | Relationships between variance in *M*, *G* and *R* in wing size and shape. a, *R* and *G* on M_{hom} and M_{het}. **b**, *R* on *G*. See legend of Fig. 4 for additional explanation.



Extended Data Figure 4 | Relationships between conditional variance in *M* or *G* and variance at higher levels. See Legend of Fig. 4 for explanation. Values of the scaling slopes and R^2 are given in Extended Data Table 4. **a**, *R* and *G* on M_{hom} and M_{het} for shape. **b**, *R* and *G* on M_{hom} and M_{het} for shape–size. **c**, *R* on *G* for shape. **d**, *R* on *G* for shape–size.

Extended Data Table 1 | Taxa included in the species dataset, but not included in ref. 22

Genus	Species	Location of collection	Source†	Flies‡	N
Drosophila	*anceps*	Infiernillo, Michoacan, Mexico	SS 15081-1261	Lab	69
Drosophila	*ezoana*	Oulanka, Finland	A. Hoikkala	Lab	148
Drosophila	*lutescens*	Honshu, Japan	SS 14022-0271.00	Lab	79
Drosophila	*montana*	Oulanka, Finland, Colorado USA, Vancouver, Canada	A. Hoikkala	Lab	288
Drosophila	*silvestris*	Hawai'i, USA	C. Boake	Lab F1, F2	49
Scaptodrosophila	*galloi*	Sao Paulo, Brazil	SS 11010-0051.1	Lab	209

†SS, species obtained from the US *Drosophila* Species Stock Center. Number following is the stock number. A. Hoikkala and C. Boake provided mounted wings.
‡Laboratory, flies measured after multiple generations in laboratory culture. F1 and F2, flies measured in first or second generation in the laboratory.

Extended Data Table 2 | Ratios of evolvabilities of matrix 1 to matrix 2

Matrix 1	Matrix 2	Dim.*	Shape ± S.E.	Shape and ln(size) ± S.E.
R †	M_{hom}	13	187.00 ± 15.69	216.55 ± 26.77
R	M_{het}	5	429.98 ± 39.08	443.02 ± 54.78
R	G	20	0.49 ± 0.04	1.77 ± 0.23
G	M_{hom}	15	377.88 ± 7.08	121.92 ± 3.56
G	M_{het}	6	839.02 ± 16.85	238.42 ± 6.85
R	conditional M_{hom}	13	1231.06 ± 123.42	5269.75 ± 652.03
R	conditional M_{het}	5	657.23 ± 59.06	3391.68 ± 400.18
R	conditional G	20	6.86 ± 0.74	40.57 ± 5.82
G	conditional M_{hom}	15	3252.17 ± 179.93	5826.21 ± 437.26
G	conditional M_{het}	6	2181.22 ± 59.78	3444.40 ± 150.26

*Dimension of common sub-space used to compare matrices[10]. Dimensions of the comparisons chosen to be in the full-rank common subspace that best fits the two matrices.
†R is expressed as the rate per Myr.

Extended Data Table 3 | Univariate estimates of phylogenetic heritability (H^2) for twenty shape traits, plus ln(centroid size)

	Among-species variance			
trait	Total	Brownian motion	Residual	Phylogenetic H^2
1	2295.65	2239.80	55.85	0.976
2	1351.00	1137.80	213.20	0.842
3	1042.61	963.14	79.47	0.924
4	188.85	167.56	21.29	0.887
5	150.49	95.64	54.85	0.636
6	229.15	220.98	8.17	0.964
7	107.52	82.00	25.52	0.763
8	42.44	17.34	25.10	0.409
9	37.12	24.21	12.91	0.652
10	32.86	27.27	5.59	0.830
11	15.81	6.74	9.07	0.426
12	22.56	18.14	4.43	0.804
13	14.13	6.52	7.61	0.461
14	21.79	20.28	1.51	0.931
15	10.19	6.45	3.74	0.633
16	12.97	10.65	2.32	0.821
17	7.86	5.92	1.94	0.753
18	12.55	11.22	1.33	0.894
19	5.38	4.30	1.08	0.800
20	2.82	1.93	0.88	0.686
ln(size)	60039	57951	2088	0.965

Extended Data Table 4 | Regression of $\log_{10}$ variances in *R* (per Myr) or *G* on $\log_{10}$ conditional variances in *G* or *M* matrices

Data	Predictor	Matrix	Dim.*	Slope	S.E.	R^2
Shape	M_{hom}	*R*	17	0.911	0.028	0.61
Shape	M_{het}	*R*	8	0.498	0.036	0.47
Shape	M_{hom}	*G*	17	0.741	0.008	0.91
Shape	M_{het}	*G*	8	0.602	0.016	0.93
Shape	*G*	*R*	17	0.941	0.022	0.71
Shape-size	M_{hom}	*R*	18	0.898	0.018	0.69
Shape-size	M_{het}	*R*	8	0.916	0.027	0.88
Shape-size	M_{hom}	*G*	18	0.753	0.007	0.85
Shape-size	M_{het}	*G*	8	0.610	0.013	0.88
Shape-size	*G*	*R*	18	1.103	0.016	0.85

*Number of dimensions compared.

nature research

Corresponding author(s): Houle

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined. Counts of individuals, counts of taxa studied.

2. Data exclusions

Describe any data exclusions. Described in text.

3. Replication

Describe whether the experimental findings were reliably reproduced. No experiments to replicate.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups. Not applicable.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis. Not applicable

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Descriptions of software are in the Methods section.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No materials.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Not applicable.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	N/A
b. Describe the method of cell line authentication used.	N/A
c. Report whether the cell lines were tested for mycoplasma contamination.	N/A
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	N/a

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	Insects not treated as "animals".

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	N/A