

Phenotypic divergence predicts transgressive segregation in recombinant hybrids

Ken A. Thompson¹✉

¹Department of Zoology & Biodiversity Research Centre, University of British Columbia, Canada

The mutations fixed during divergent adaptation can have deleterious pleiotropic effects on traits under stabilizing selection. Compensatory mutations fix within populations to counteract this deleterious pleiotropy, and these compensatory mutations can segregate in interpopulation hybrids. Theory predicts if alleles are pleiotropic and meaningfully large in effect, then the segregation variance in non-divergent traits should be positively correlated with the amount of divergence in divergently selected traits. I systematically searched the literature for studies that measured phenotypic traits in two parent taxa and their F₁ and F₂ hybrids in a common environment, then evaluated whether parental divergence predicts the segregation variance in traits putatively under stabilizing selection. The results support the theoretical prediction: populations that have a greater degree of phenotypic divergence have a larger amount of segregation variance in phenotypes that do not differ between them. My results suggest that the genes used during divergent adaptation are pleiotropic and that potentially deleterious segregation variance accumulates systematically as populations diverge.

Fisher's geometric model | pleiotropy | hybridization | theory

Correspondence: ken.thompson@zoology.ubc.ca

Introduction

Populations adapt to novel environments by increasing the frequency of mutations that change the values of traits under natural selection. Models of this process, such as Fisher's (1) geometric model of adaptation, are useful because they generate explicitly testable predictions that can be tested in experimental evolution studies (2) and natural populations (3, 4). In such models, alternative outcomes can be observed depending on the parameter values assumed in the model. For example, some predictions only emerge when universal pleiotropy is assumed (5), when parallel genetic evolution is allowed (6) or not (7), when population size is fixed or not (8), or when a particular form of selection is assumed (9). Accordingly, testing the explicit and conditional predictions of models allows us to learn about the conditions under which adaptation tends to proceed in nature.

In models assuming universal pleiotropy, mutations have the potential to affect all of an organism's traits. Barton (10) simulated Fisher's geometric model in a case where two populations of an organism with 10 traits underwent divergent adaptation for a single trait while the others were under stabilizing selection. These simulations showed that, following hybridization of the two populations, the segregation variance in the 9 traits under stabilizing selection was positively associated with the distance of the parental phenotypic

optima from one another. If Barton had simulated a species with a high degree of modularity (low pleiotropy) or infinitely small mutations, there would be no such relationship. A conceptual model and hypothetical prediction is shown in Figure 1 with a full explanation given in the caption.

Barton's observation – a positive correlation between the amount of divergence in trait x and the segregation variance released by hybridization in the non-divergent trait y – has not yet been tested empirically. In this article I test this theoretical prediction using data collated from experimental crossing studies. Whether or not it receives support will serve to calibrate our expectation of whether pleiotropy is a major feature of the alleles incorporated during divergent evolution.

Methods

As a part of an ongoing and separate project, I conducted a systematic literature search with the goal of identifying studies that measured phenotypic traits and variances in two parent taxa and their F₁ hybrids in a common environment. Details of the literature search are given in the Sup. Note 1. In total I (with help) screened over 11,000 studies and collected data from 200 studies. I analyze a subset of these studies in the present article. In order to be included, a study had to meet all of the following criteria:

- Studies had to measure parents, F₁, and F₂ hybrids
- At least two non-fitness ('ordinary' (11)) traits
- At least one trait where the parents were significantly different and at least one trait where they were not
- Parents wild or in lab for 10 or fewer generations

After screening, thirteen studies (one with two crosses) were eligible for inclusion (12–24) in the present analysis.

For each study I divided traits into two groups: those that differed between the parents and were presumably subject to divergent selection, and those that did not and were more likely subject to stabilizing selection. For each divergent selection, I calculated the degree of phenotypic divergence in units of parental phenotypic SDs using the smaller of the two parental values. I then determined the segregation variance of each trait as the ratio of F₂ to F₁ standard deviations – the ratio is used because I do not have the raw data to scale each study and units are not comparable across traits or studies. For each cross, I extracted two data points by averaging across all traits: the mean value parental divergence and of segregation variance.

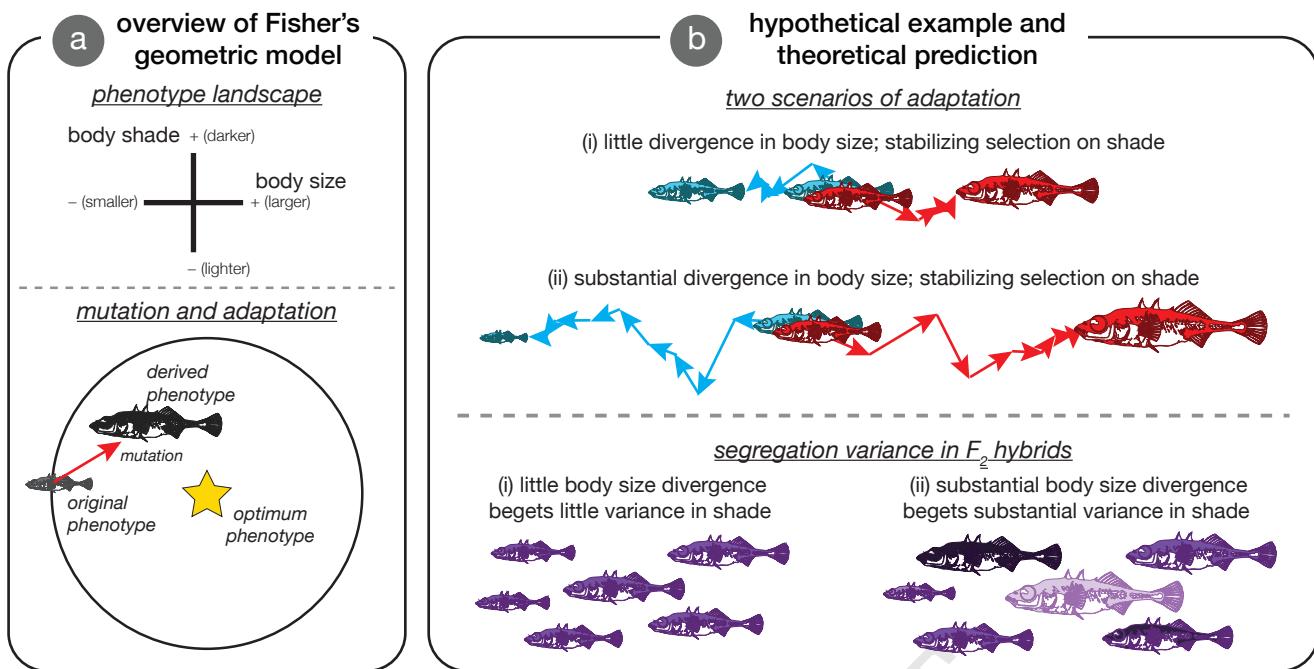


Fig. 1. Overview of Fisher's geometric model and theoretical prediction. Panel A shows a general overview of Fisher's geometric model. The upper section shows the phenotype landscape under consideration, wherein the horizontal axis is body size and the vertical axis is body shade. The lower section illustrates the fixation of a pleiotropic allele during adaptation. The original phenotype is medium in size and medium in shade. The optimal phenotype is larger, but the same shade. A mutation arises that greatly increases size and has a deleterious pleiotropic effect to darken shade (it could just as easily lighten it). Since the mutation is beneficial (inside the circle), it has a high probability of fixation in spite of the deleterious side-effect. Panel B illustrates the theoretical prediction in two diverging populations—red and blue—with the same initial phenotype for size and shade. The upper section illustrates to alternative adaptive scenarios for comparison. In both scenarios (i) and (ii), shade is under stabilizing selection in the two populations. Scenario (i) is a case where the two populations diverge little in body size and scenario (ii) represents substantial divergence in body size. The lower section of the panel illustrates the outcome of hybridization. The key insight is that the segregation variance in shade is greater in (ii) than (i). Body size segregates as well, but it would do so in a model without pleiotropy whereas shade would not necessarily.

Fisher's model makes an explicitly testable prediction: there should be a positive relationship between parental divergence – for divergently selected traits – and segregation variance – for traits that do not differ between the parents. I tested this prediction using a simple linear model with parental divergence as the independent variable and segregation variance as the dependent variable. Although I by necessity classified traits as 'non-divergent' they were typically not identical and have some amount of phenotypic difference between them that could be associated with the segregation variance. To test whether the non-significant but present difference between the so-called non-divergent trait predicted phenotypic segregation variance, I conducted a separate multiple regression with both divergent trait SDs and non-divergent trait SDs as predictors. The relationships in the raw data were non-linear and so I used logarithmic transformations on all summary statistics. To evaluate the relationships in the raw data I used Spearman's rank-order correlations.

A similar pattern to what is predicted above could emerge if phenotypic divergence is correlated with genetic divergence (i.e., time) in the data. Unfortunately I could only obtain gene sequences for 6 of 14 studies, insufficient for a robust test of genetic divergence. I instead used two alternative approaches to evaluate whether observed patterns could be due to time. First, I use a simple *t*-test to determine if intraspecific crosses ($n = 6$) have lower phenotypic divergence than interspecific crosses – the assumption here is that in-

traspecific pairs have diverges more recently than interspecific pairs. Second, because these studies were drawn from a larger subset of studies, I could evaluate whether genetic distance between the parental species is typically associated with phenotypic divergence in a representative set of studies selected for the same reasons. It might not be the case if, for example, particularly divergent parents were selected for intraspecific crosses.

Results

My analysis lends support to the prediction that compensatory alleles fixed during divergent adaptation cause transgressive segregation variance. Specifically, in a simple linear model I observed a positive correlation between divergence and segregation variance (log-log scale; $F_{1,12} = 9.24$, $P = 0.0103$, $r^2 = 0.435$). When I conducted a multiple regression with both phenotypic divergence in divergent and in non-divergent traits as predictors, divergent traits remains a significant predictor ($P = 0.0437$) but non-divergent traits do not significantly predict segregation variance ($P = 0.281$). When analyzing the raw (non-transformed) data with a Spearman's rank-order correlation, the phenotypic of divergent traits predicts segregation variance of non-divergent traits ($\rho = 0.635$, $P = 0.0171$), whereas the divergence of non-divergent traits does not ($\rho = 0.486$, $P = 0.0809$). Our qualitative conclusions are unchanged if we include three additional studies with data from first-generation back-crosses but not F_2 s (not

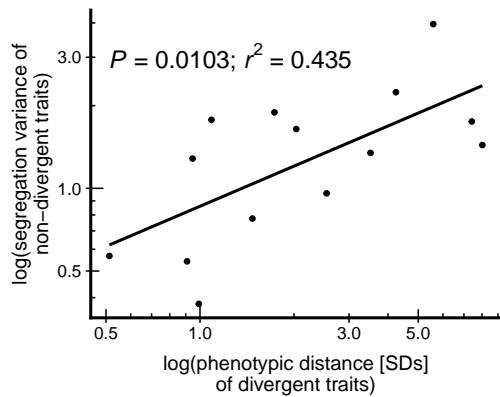


Fig. 2. Scatterplot depicting the empirical result. Each data-point represents a unique cross between two populations or species. Note the log-log scale.

shown here but included in analysis script [see Data accessibility].

As mentioned in the methods, genetic drift can cause an association between divergence time and segregation variance. A *t*-test did not detect a difference between intra-specific and inter-specific crosses in phenotypic divergence ($F_{1,12} = 1.027, P = 0.331$). We examined the larger data set from which the present studies were drawn and found that genetic and phenotypic distance were not correlated ($F_{1,33} = 0.0458, P = 0.832, r^2 = 0.001$).

Discussion

I tested the hypothesis that divergent selection is associated with transgressive phenotypic variation in recombinant hybrids. This prediction holds if the genes underlying divergent adaptation are pleiotropic and does not if they are not or if they are pleiotropic but have infinitely small effects (25). The consistency between the results presented here and the theoretical prediction therefore suggests that the genes used during adaptation are indeed pleiotropic and of appreciably large effect. The results also inform a general understanding of the mode of adaptation. Even if mutations can be large and pleiotropic, models with slowly moving fitness optima predict that very small alleles will be used during adaptation (26). The analyses above suggest that optima in nature move fast enough for alleles of non-trivial effects alleles to be incorporated.

Previous studies have studied similar processes to what we have shown herein. Stelkens and colleagues (27, 28) showed that genetic divergence but not phenotypic divergence between species predicts transgressive segregation. Their approach largely focussed on the mean phenotype rather than the variance, however, and so the underlying theoretical predictions are not identical. In studies that implement simulations with Fisher's geometric model, the accumulation of segregation variance grinds to a near-halt after the adaptation phase of simulations has ended (6, 10), because the mutations fixed during the 'optimum circling' stage fix due to drift or to compensate for it and therefore are of very small effect (5). In any case the analyses of inter- & intra-specific crosses and genetic distance suggest that phenotypic and ge-

netic divergence are de-coupled in the compiled data and that the observed patterns are not caused by drift.

Although my analysis supports the prediction of models of adaptation with respect to segregation variance for traits, I cannot test predictions about hybrid fitness. In Barton's (10) simulations, segregation variance in non-divergent traits is deleterious and accordingly hybrid fitness declines as segregation variance increases. It would be useful to test this prediction in an experimental system – for example parental lines can be selected for divergence to varying degrees (e.g., differing concentrations of a stressor) and then hybridized with a common ancestor. F_1 and F_2 hybrid fitness could easily be compared in a common (intermediate or ancestral) environment and the clear prediction is that the loss in fitness of F_2 hybrids (due to segregating breakup of co-adapted compensatory alleles) compared to F_1 s will be greater in more divergently selected lines. One might also expect that selection for heterozygosity (for a given genomic hybrid index) will increase with parental divergence (29).

It is worth taking the patterns described herein with a pinch of salt. Although I illustrate a correspondence between theory and data, I did so using a correlational approach and with a small sample size of 14 crosses. Nevertheless, the significant result and appreciable variance-explained should bolster our assessment that models such as Fisher's (1) geometric model is a robust and useful abstraction of the evolutionary process.

Acknowledgements

Feedback from D. Irwin, M. Osmond, S. Otto, and D. Schlüter benefited the manuscript. M. Urquhart-Cronish assisted with the literature search and data collection. We are grateful to all authors who responded to requests for data.

Data accessibility

All data used in this article will be deposited in a repository following publication.

Bibliography

1. Fisher, R. F. (1930) *The Genetical Theory of Natural Selection*. (Oxford University Press, Oxford, UK), p. 272.
2. Tenailion, O. (2014) The utility of Fisher's geometric model in evolutionary genetics. *Annual Review of Ecology, Evolution, and Systematics* **45**, 179–201.
3. Rogers, S. M., Tamkee, P., Summers, B., Balabahadra, S., Marks, M., Kingsley, D. M., & Schlüter, D. (2012) Genetic signature of adaptive peak shift in threespine stickleback. *Evolution* **66**, 2439–2450.
4. Stearns, F. W. & Fenster, C. B. (2016) Fisher's geometric model predicts the effects of random mutations when tested in the wild. *Evolution* **70**, 495–501.
5. Yamaguchi, R & Otto, S. P. (2019) Insights from Fisher's geometric model on the likelihood of speciation under different histories of environmental change. *bioRxiv*.
6. Chevin, L. M., Decoercent, G., & Lenormand, T. (2014) Niche dimensionality and the genetics of ecological speciation. *Evolution* **68**, 1244–1256.
7. Thompson, K. A., Osmond, M. M., & Schlüter, D. (2019) Parallel genetic evolution and speciation from standing variation. *Evolution Letters* **3**, 129–141.
8. Osmond, M. M & Otto, S. P. (2015) Fitness-valley crossing with generalized parent-offspring transmission. *Theoretical Population Biology*.
9. Fraïsse, C., Gunnarsson, P. A., Roze, D., Bierne, N., & Welch, J. J. (2016) The genetics of speciation: Insights from Fisher's geometric model. *Evolution* **70**, 1450–1464.
10. Barton, N. H. (2001) The role of hybridization in evolution. *Molecular Ecology* **10**, 551–568.
11. Orr, H. A. (2001) The genetics of species differences. *Trends in Ecology and Evolution* **16**, 343–350.
12. Bradshaw, H. D., Otto, K. G., Frewen, B. E., McKay, J. K., & Schemske, D. W. (1998) Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). *Genetics*.

13. Bratteler, M., Lexer, C., & Widmer, A. (2006) Genetic architecture of traits associated with serpentine adaptation of *Silene vulgaris*. *Journal of Evolutionary Biology* **19**, 1149–1156.
14. Fishman, L., Kelly, A. J., & Willis, J. H. (2002) Minor quantitative trait loci underlie floral traits associated with mating system divergence in *Mimulus*. *Evolution* **56**, 2138–2155.
15. Hermann, K., Klahre, U., Venail, J., Brandenburg, A., & Kuhlemeier, C. (2015) The genetics of reproductive organ morphology in two *Petunia* species with contrasting pollination syndromes. *Planta*.
16. Husenmann, M., Tobler, M., McCauley, C., Ding, B., & Danley, P. D. (2017) Body shape differences in a pair of closely related Malawi cichlids and their hybrids: Effects of genetic variation, phenotypic plasticity, and transgressive segregation. *Ecology and Evolution* **7**, 4336–4346.
17. Jacquemyn, H., Brys, R., Honnay, O., & Roldán-Ruiz, I. (2012) Asymmetric gene introgression in two closely related *Orchis* species: Evidence from morphometric and genetic analyses. *BMC Evolutionary Biology* **12**.
18. Koeling, V. A. & Mauricio, R. (2010) Genetic factors associated with mating system cause a partial reproductive barrier between two parapatric species of *Leavenworthia* (Brassicaceae). *American Journal of Botany* **97**, 412–422.
19. MACNAIR, M. R., MACNAIR, V. E., & MARTIN, B. E. (1989) Adaptive speciation in *Mimulus*: an ecological comparison of *M. cuprifilus* with its presumed progenitor, *M. guttatus*. *New Phytologist* **112**, 269–279.
20. Pritchard, V. L., Knutson, V. L., Lee, M., Zieba, J., & Edmonds, S. (2013) Fitness and morphological outcomes of many generations of hybridization in the copepod *Tigriopus californicus*. *Journal of Evolutionary Biology* **26**, 416–433.
21. Raeymaekers, J. A., Delaire, L., & Hendry, A. P. (2009) Genetically based differences in nest characteristics between lake, inlet, and hybrid threespine stickleback from the misty system, British Columbia, Canada. *Evolutionary Ecology Research* **11**, 905–919.
22. Selz, O. M., Thommen, R., Maan, M. E., & Seehausen, O. (2014) Behavioural isolation may facilitate homoploid hybrid speciation in cichlid fish. *Journal of Evolutionary Biology* **27**, 275–289.
23. Shore, J. S. & Barrett, S. C. (1990) Quantitative genetics of floral characters in homostylous *Turnera ulmifolia* var. *Angustifolia* Willd. (Turneraceae). *Heredity* **64**, 105–112.
24. Vallejo-Marin, M. & Hiscock, S. J. (2016) Hybridization and hybrid speciation under global change. *New Phytologist* **211**, 1170–1187.
25. Barton, N. H., Etheridge, A. M., & Véber, A. (2017) The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology* **118**, 50–73.
26. Matuszewski, S., Hermisson, J., & Kopp, M. (2014) Fisher's geometric model with a moving optimum. *Evolution* **68**, 2571–2588.
27. Stelkens, R. & Seehausen, O. (2009) Genetic distance between species predicts novel trait expression in their hybrids. *Evolution* **63**, 884–897.
28. Stelkens, R. B., Schmid, C., Selz, O., & Seehausen, O. (2009) Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. *BMC Evolutionary Biology* **9**, 283.
29. Simon, A., Bierne, N., & Welch, J. J. (2018) Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. *Evolution Letters*.
30. Zirkle, C. (2002) *Genetics and the Origin of Species*. Theodosius Dobzhansky. (Columbia University Press, New York) Vol. 30, pp. 128–131.
31. Hubbs, C. L. (1955) Hybridization between fish species in nature. *Systematic Zool* **4**, 1–20.
32. Mayr, E. (1963) *Animal species and evolution*. (The Belknap Press, Cambridge, MA).
33. Grant, V. (1981) *Plant speciation*. (New York: Columbia University Press xii, 563p.-illus., maps, chrom. nos.. En 2nd edition. Maps, Chromosome numbers. General (KR, 198300748)).
34. Lande, R. (1981) The minimum number of genes contributing to quantitative variation between and within populations. *Genetics* **99**, 541–553.
35. Tave, D. (1986) Genetics for fish hatchery managers. *Tave, D. Genetics for Fish Hatchery Managers. Xv+299p.*
36. Churchill, G. A & Doerge, R. W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971.
37. Bradshaw, H. D., Wilbert, S. M., Otto, K. G., & Schemske, D. W. (1995) Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature* **376**, 762–765.
38. Lynch, M & Walsh, B. (1998) *Genetics and analysis of quantitative traits*. (Sinauer, Sunderland, MA), 1 edition, p. 980.
39. Hatfield, T. & Schlüter, D. (1999) Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution* **53**, 866–873.
40. Schlüter, D. (2000) *The Ecology of Adaptive Radiation*. (Oxford University Press, New York).
41. Coyne, J. A & Orr, H. A. (2004) *Speciation*. (Sinauer), p. 545.
42. Crnokrak, P. & Roff, D. a. (1995) Dominance variance: associations with selection and fitness. *Heredity* **75**, 530–540.
43. McGee, M. D., Reustle, J. W., Oufiero, C. E., & Wainwright, P. C. (2015) Intermediate Kinematics Produce Inferior Feeding Performance in a Classic Case of Natural Hybridization. *The American Naturalist* **186**, 807–814.
44. Merilä, J. & Sheldon, B. C. (1999) Genetic architecture of fitness and nonfitness traits: Empirical patterns and development of ideas.
45. Schlüter, D., Price, T. D., & Rowe, L. (1991) Conflicting selection pressures and life history trade-offs.
46. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA. (2009) Preferred Reporting Items for Systematic Reviews and Meta-Analyses :. *Annals of Internal Medicine* **151**, 264–269.
47. Winter, D. J. (2017) rentrez: An R package for the NCBI eUtils API. *The R journal*.
48. Wilkinson, S. P. (2018) The 'aphid' package for analysis with profile hidden Markov models. R package version 1.1.0. <https://cran.r-project.org/package=aphid>.
49. Paradis, E & Schliep, K. (2018) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*.

Word Counts

This section is *not* included in the word count.

Statistics on word count.

PRE-PRINT

Supplementary Note 1: Database assembly

A. Search strategy. We searched the literature for studies that made measurements of traits in F₁ hybrids and their parents. To identify studies for possible inclusion, we conducted a systematic literature search using Web of Science (www.webofknowledge.com). We included all papers that resulted from a general topic search of “Castle-Wright”, and from a topic search of “F1 OR hybrid OR inherit*” in articles published in Evolution, Proceedings of the Royal Society B, Journal of Evolutionary Biology, Heredity, or Journal of Heredity. These journals were selected because a preliminary search indicated that approximately half of all suitable studies were published in these journals. These searches returned 106 studies deemed suitable after screening. To be more comprehensive, we conducted additional systematic searches by conducting similar topic searches among articles citing influential and highly-cited publications (30–41) The full literature search results are available in the archived data. Our initial search returned 14048 studies, and after removing duplicates this left 11287 studies to be screened for possible inclusion. This literature search was primarily done for another unpublished study.

B. Evaluation of studies. We required studies to meet several criteria to merit inclusion in our database. First, the study organisms had to originate recently from a natural (i.e., ‘wild’) population. This is because dominance patterns in domestic species differ substantially from non-domesticated species (42) and because we are explicitly interested in patterns as they occur in nature. We excluded studies using crops, domestic animals, laboratory populations that were > 10 (sexual) generations removed from the wild, or where populations were subject to artificial selection in the lab, or was a lab ‘strain’. If populations were maintained in a lab for more than 10 generations but were found by comparison to still strongly resemble the source population, we included the study ($n = 2$). We also excluded studies where the origin of the study populations was ambiguous. Hybrids had to be formed via the union of gametes from parental taxa, so we excluded studies using techniques like somatic fusion. Second, the ancestry of hybrids had to be clear. Many studies reported phenotypes of natural hybrids, for example in hybrid zones. We did not include these studies unless the hybrid category (i.e., F₁, F₂, backcross) was confidently (typically over 95 % probability, unless the authors themselves used a different cut-off) determined with molecular markers or knowledge that hybrids were sterile and thus could not be beyond the F₁.

Third, because we were interested in the inheritance of traits that are proximally related to organismal performance (43), we required studies to report measurements of at least one ‘non-fitness’ trait (‘ordinary’ traits (11)). Non-fitness traits (hereafter simply ‘traits’) are those that are likely under stabilizing selection, whereas ‘fitness’ traits are those that are likely under directional selection (44, 45). In most cases it was possible to evaluate this distinction objectively because authors specifically referred to traits as components of fitness, reproductive isolating barriers, or as being affected by non-ecological hybrid incompatibilities. In some cases, however, we made the distinction ourselves. If particular trait values could be interpreted as resulting in universally low fitness, for example resistance to herbivores or pathogens, this trait was not included. The majority of cases were not difficult to assess, but we have included reasons for excluding particular studies or traits in our database screening notes (see Data Accessibility).

Traits had to be measured in a quantitative manner to be included in our dataset. For example, if a trait was reported categorically (e.g., ‘parent-like; or ‘intermediate’), we did not include it. Some traits such as mate choice must often be scored discretely (in the absence of multiple trials per individual), even though the trait can vary independent trials. Accordingly, we included discretely scored traits—like mate choice—when it was possible in principle to obtain a different outcome on independent trials. Such traits are recorded as 0s and 1s, but hybrids can be intermediate if both outcomes occurred with equal frequencies. We included traits where authors devised their own discrete scale for quantification. When suitable data were collected by the authors but not obtainable from the article, we wrote to the authors and requested the data. If the author cited a dissertation as containing the data, we attempted to locate the data therein because dissertations are not indexed by Web Of Science. We included multivariate trait summaries (e.g., PC axis scores) when reported. If traits reported both the raw trait values and the PC axis scores for a summary of those same traits, we collected both sets of data but omitted the PCs in our main analyses. Below we report results for a subset of analyses where we analyzed only PCs or only uncorrelated traits [**TBA!**]

Using these criteria, we screened each article for suitability. As a first pass, we quickly assessed each article for suitability by reading the title and abstract and, if necessary, consulting the main text. After this initial search, we retained 407 studies. The lead author (KAT) conducted an in-depth evaluation of each study. If deemed suitable, we next evaluated whether the necessary data could be obtained. After this second assessment, 198 studies remained. The reasons for exclusion of each study are documented the archived data (see Data Accessibility).

C. Data collection. For each study, we recorded several types of data. First, we recorded the mean, sample size, and an estimate of uncertainty (if available) for each measured trait for all parental crosses and hybrid categories. In most cases, these data were included in tables or could be extracted from figures. In some cases, we contacted authors for the raw data or summary data. Each study conducted a minimum of three records to our database: one trait measured in each parent and the F₁ generation. Traits were categorized as one of: behaviour, chemical, life history, or morphological. If the same traits were measured over ontogeny, we used only the final data point. When data were measured in multiple ‘trials’ or ‘sites’ we pooled them. If data were reported for different cross directions or sexes we recorded data for each sex separately.

For each paper we recorded whether the phenotypes were measured in the lab or field, if in the lab the number of generations of captivity, and whether a correlation matrix (preferably in recombinant—F₂ or BC—hybrids; see below) was available or calculable from the raw data or figures. Each study thus contributed a minimum of three observations to our dataset: one observation for both parents and their F₁ hybrid. Occasionally, different studies analysed different traits from individuals from the same crosses. In these cases, we simply grouped them as being the same study before analysis. The final database is available from Dryad (see Data Accessibility).

C.1. Comments on systematic nature of review. We attempted to follow PRISMA (46) guidelines to the best of our ability. Most of the criteria have been addressed above but a few other comments are warranted. First, we have no reason to suspect that any bias was introduced in our estimates of dominance. This is because no studies seemed to have *a* *priori* hypotheses about dominance patterns. Accordingly, we do not believe that our estimates suffer from a file drawer problem, since detecting dominance was not the stated goal of contributing studies. We also were forced to be subjective at times when excluding variables. We noticed that many 'fitness' variables were transgressive in the sense that—for example—individuals were larger than their parents or failed to obtain mates. Excluding these variables is therefore expected to largely render our estimates more conservative (i.e., closer to 0). We present an analysis as in Fig. 2 of the main text but with characters coded as 'ambiguous fitness' excluded in **Fig. SX [TBA]**.

D. Calculating genetic distance. We estimated genetic distance for pairs of species where data were available for both parents. A preliminary screening revealed that the internal transcribed spacer (ITS I and II) was the most commonly available gene for plants and cytochrome b was the most available gene for animals in our dataset. We downloaded sequences in R using the 'rentrez' package (47), and retained up to 40 sequences per species. Sequences were then aligned with the profile hidden Markov models implemented in the 'align' function in the package, 'aphid' (48). After aligning sequences we calculated genetic distance by simply counting the number of sites that differed between two aligned sequences, implemented using the 'raw' model option in the 'dist.dna' within 'ape' (49).