Submitted Manuscript: Confidential
Template revised November 2022

Title: Evolvability predicts macroevolution under fluctuating selection

Authors: Agnes Holstad[1]*, Kjetil L. Voje[2], Øystein H. Opedal[3], Geir H. Bolstad[4], Salomé Bourg[1], Thomas F. Hansen[5], Christophe Pélabon[1]

Affiliations:

[1]Department of Biology, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology; Trondheim, Norway.

[2]Natural History Museum, University of Oslo; Oslo, Norway.

[3]Department of Biology, Biodiversity Unit, Lund University; Lund, Sweden.

[4]Norwegian Institute for Nature Research (NINA); Trondheim, Norway

[5]Department of Biology, Centre for Ecological and Evolutionary Synthesis, University of Oslo; Oslo, Norway.

*Corresponding author. Email: agnes.holstad@ntnu.no

Abstract: Heritable variation is a prerequisite for evolutionary change, but the relevance of genetic constraints on macroevolutionary timescales is debated. By using two datasets on fossil and contemporary taxa we show that evolutionary divergence among populations, and to a lesser extent among species, increases with microevolutionary evolvability. We evaluate and reject several hypotheses to explain this relationship and propose that an effect of evolvability on population and species divergence can be explained by the influence of genetic constraints on the ability of populations to track rapid, stationary environmental fluctuations.

One-Sentence Summary: Macroevolution depends on microevolutionary potential for evolution.

Main Text: A key insight from the modern synthesis of evolutionary biology was that response to selection, and therefore adaptation, depends on the presence of genetic variation. This led to the idea that more genetically diverse populations and species should be able to adapt faster when their environment changes. Simpson (*1*) in the 1940s was the first to test this prediction, but he failed to detect differences in levels of variation between slow- and fast-evolving lineages. Thirty years later, Kluge and Kerfoot (*2*) reported a positive correlation between within-population variation and among-population differentiation in seven vertebrate species. Their study was criticized on methodological grounds (*3, 4*), casting doubt on what became known as the "Kluge-Kerfoot phenomenon". Following the emergence of evolutionary quantitative genetics in the 1980s, the focus shifted from phenotypic to genetic variation, and studies started relating patterns of multivariate additive genetic variance (evolvability) to patterns of phenotypic divergence. The results have been mixed, however (*5-18*), and marred by persistent methodological problems related to the quantification of evolvability, divergence and their relationship (*19-21*). Amending some of these problems, two recent synthetic studies (*17, 21*) have concluded that multivariate microevolutionary evolvability relates to evolution on longer timescales within many study systems, but how general this relationship is and what generates it remain unknown.

To investigate the divergence-evolvability relationship and evaluate various biological and methodological hypotheses put forward to explain it, we gathered two extensive datasets on contemporary and fossil taxa. For the contemporary taxa, evolvability estimated as within population mean-scaled additive genetic variance (*22, 23*), is combined with trait divergence based on 2011 population means from 280 traits in 33 species and 676 species means from 130 traits in 96 different species. For the fossil taxa, evolvability estimated from mean-scaled within-sample variance is combined with changes in trait means across time in 589 fossil time series from 150 independent lineages for a total of 10594 samples. The time of divergence between fossil samples ranges from 10 years to 7.6 million years.

The two datasets offer complementary strengths and unprecedented insights into the mechanisms that can generate a relationship between evolvability and divergence among taxa. The contemporary data provide direct measures of evolvability based on additive genetic variance in a wide variety of traits, but the exact history of population divergence is usually not known, which limits information about the causal interplay between evolvability and divergence. The fossil time-series data allow investigation of the dynamic relationship between divergence and evolvability through time but lack direct information about genetic variation. We mitigate this shortcoming by documenting a strong and near isometric scaling relationship between measures of additive and phenotypic variance observed across traits in the contemporary data (Fig. 1) and use this to translate phenotypic variation within fossil samples into estimates of evolvability.

Relationship between divergence and evolvability

Higher evolvability is systematically associated with more divergence among populations, species, and fossil samples (Fig. 2). Variation in evolvability explains 30% of the among-population divergence, 12% of the among-species divergence and 37% of the fossil sample divergence. The scaling between divergence and evolvability is similar in the contemporary-population and fossil data, with a 1% increase in evolvability predicting a 0.46±0.05% increase in among-population divergence and a 0.42±0.04% increase in divergence among fossil samples. For the contemporary-species data, a 1% increase in evolvability predicts a 0.36±0.08% increase in among-species divergence.

The fossil time series further allow for causal analysis, in which the evolvability of each fossil sample is used to predict the evolutionary change to the next sample (Fig. 3A). As for

divergence measured for whole time series, morphological distance between consecutive samples scales positively with evolvability (Fig. 3B). Our results show that an increase in evolvability of 1% is associated with a 0.40±0.02% increase in the magnitude of divergence between consecutive samples, explaining 17% of the variance. This relationship is not driven by differences among time series, as the average within-time-series relationship was of comparable magnitude: 0.36±0.02%. It could be an artifact, however, of within-sample variation (i.e., evolvability) being confounded by microevolution within the samples. To account for this, we fitted an Ornstein-Uhlenbeck process to each time series and used this to predict and remove within-sample variance due to microevolution of the trait mean. Because we did not have the exact duration of each fossil sample, we considered a worst-case scenario in which the within-sample time interval equals 50% of the maximum possible duration, that is, when there is no temporal gap between successive samples (see supplementary material (*24*)). Under such conditions a 1% increase in corrected evolvability estimates is still associated with a 0.30±0.02% increase in the magnitude of changes in the trait mean, explaining 11% of the variance (fig. S1D). A more realistic, yet still exaggerated, scenario in which the fossil samples are assumed to span 10% of the maximum duration gives an increase of 0.37±0.02% explaining 15% of the variance (fig. S1C), which is almost identical to the uncorrected result.

We compiled and developed various hypotheses to investigate potential mechanisms that could generate a correlation between evolvability and divergence on micro- and macroevolutionary timescales (table S1). After rejecting several non-causal hypotheses, involving statistical artifacts, gene flow, plasticity and selection shaping genetic variation to align with lines of population divergence (concordant selection), we turn to causal hypotheses. We first argue that the relationship is not a simple consequence of lack of genetic variation limiting evolution under directional selection or genetic drift, before we propose a new hypothesis based on genetic constraints limiting evolutionary responses to rapidly fluctuating selection.

Rejection of non-causal explanations for divergence-evolvability correlations

Spurious correlation

The regression of inter- on intrapopulation variation can be subject to statistical artifacts arising from *i*) the use of the same or related variables at both levels (*4*), *ii*) a correlation between within-sample variance, which is related to estimated evolvability, and estimation variance in the means, which will bias estimates of among-population variance, iii) a correlation between estimation variances of evolvabilities and means, and iv) heterogeneity in the data. The first problem, which was a source of criticisms of Kluge and Kerfoot, does not apply to our analysis, because we used population means to scale measures of evolvability and log transformation to measure divergence on a proportional scale.

The second problem can be eliminated by correcting among-population divergence for bias. Analyzing a subset of the data for which bias correction was possible revealed no qualitative differences from our main results (figs. S2 and S3). As for the third problem, we show in figure S4 that correlated estimation variances in means and variances do not cause a spurious correlation.

Correlations may also arise because both inter- and intrapopulation variances depend on trait type and dimensionality (*25*). This can be rejected as an explanation for our general result because divergence-evolvability relationships are similar within homogenous trait categories (Fig. 4A-F).

Gene flow

Gene flow among populations may generate a divergence-evolvability relationship because introgression among more divergent populations could generate more genetic variation within populations (*26*), and because high evolvability could counteract maladaptation and homogenization of populations due to gene flow (*27, 28*). Although this mechanism could contribute to the stronger relationship between divergence and evolvability at the population level than at the species level, we reject the gene-flow hypothesis as a general explanation because it cannot generate the divergence-evolvability relationship observed across temporally separated fossil samples.

Phenotypic plasticity

Non-genetic responses to environmental changes (plasticity) often constitute a substantial component of population differences (*29*) that may correlate with evolvability across traits within populations (*30-32*). We reject this as a general explanation, because a strong evolvability-divergence relationship remains in data from common-garden designs that reduce plasticity (Fig. 4G-I; see also Opedal et al. 17). We cannot exclude a minor role for plasticity, however, because the relationship is slightly shallower in the common-garden data.

Concordant selection

A divergence-evolvability correlation could arise if the episodes of directional selection that drive divergence are concordant with patterns of stabilizing selection molding genetic variance within populations (*7, 33-35*). In this case, differences in population evolvability would not be the cause of differences in divergence and would therefore not be causally relevant to macroevolution. Although relevant empirical estimates of stabilizing selection are lacking, this hypothesis can be rejected based on inconsistencies with some of our empirical findings and on theoretical grounds.

First, molding genetic architecture by stabilizing selection is likely to take time and we would therefore expect the scaling relationship among species to be as strong – if not stronger – than that among populations. The weaker divergence-evolvability relationship at the among-species level contradicts this prediction.

Second, stabilizing selection is likely to reduce additive genetic variation more than environmental variation because the latter is not transmitted through generations. We would therefore expect a relationship between heritability and divergence, which is absent (fig. S5).

The concordance hypothesis is also theoretically problematic because it requires an implausible range of variation in strengths of stabilizing selection. Assuming a standard model for maintenance of genetic variance in a balance between mutation and stabilizing selection (*36*), we show that the strength of stabilizing selection needs to vary over at least four orders of magnitude, and likely more, to account for the observed scaling between evolvability and divergence (fig. S6). This would cover a range from inefficiently weak to unrealistically strong selection. Such range in stabilizing selection is even more unlikely to explain the within-trait scaling relationship between evolvability and divergence observed in the fossil time series (Fig. 3).

An alternative version of the concordance hypothesis is that within-population variation is shaped not by mutation-selection balance, but by canalizing selection changing the effects of alleles (rather than their frequencies) to match the fitness landscape. This is even less plausible, however, because canalizing selection is weak, nonlinearly related to patterns of selection on the phenotype, and largely determined by patterns of epistasis (*37-39*). Most salient, strong stabilizing selection makes canalization less effective than intermediate strengths of selection, rendering a close match between evolvability and the curvature of the fitness landscape unlikely even under ideal conditions. Combining this with the similarly

complex and nonlinear relationship between mutational effects and segregating genetic variance discussed with figure S6, we reject canalization as a general explanation for a regular relationship between divergence and evolvability.

Causal explanations for divergence-evolvability correlations

Neutral evolution

In the absence of selection, trait divergence is expected to scale proportionally with evolvability and linearly with time due to genetic drift (*40, 41*). We reject this hypothesis because divergence in our data does not accumulate with time (fig. S7A) and on longer timescales rates of evolution becomes much too slow to be explained by either the standard drift model (*40*) or by the mutation-drift model (*41, 42*) (Fig. 5A; see also *33, 43-48*).

Directional selection and genetic constraints

If population divergence reflects patterns of directional selection, the genetic-constraints hypothesis (*7, 49*) predicts more divergence along directions with more genetic variation, and therefore a positive relationship between divergence and evolvability (*19*). Given enough time, however, populations would reach their optima, and the relationship between divergence and evolvability should vanish (*7*). This prediction is supported by our finding of a weaker relationship at the species level than at the population level, as well as the strong signal in the fossil data for which the timescale is comparable to population divergence. It is also supported by the fact that low-evolvability traits have diverged less among populations than among species, while divergence of high-evolvability traits is similar at both levels (Fig. 2).

Nevertheless, estimated evolvabilities are too large to substantially constrain directional selection on the timescales considered. Indeed, only 29 generations would be necessary to generate the median divergence magnitude of 5% observed in the fossil data for a trait with a moderate evolvability under moderate selection (*24, 50*). This time span is much shorter than the median of 35115 years between our fossil samples.

Pleiotropic constraints

The genetic-constraints hypothesis could be rescued if the true potential for evolution were much lower than indicated by estimated evolvabilities. This could come about through constraining selection on genetically correlated traits (*51*). Such constraints can be quantified with conditional evolvability, that is, the evolvability of a focal character when other genetically correlated characters are kept constant (*52, 53*). If conditional evolvabilities are much smaller than unconditional evolvabilities but show similar patterns of variation across traits, then genetic constraints may influence evolution on longer time scales and explain the relationship between divergence and unconditional evolvabilities. We provide a partial test of this hypothesis by conditioning evolvabilities on overall size of the organism in a subset of twenty-five G-matrices from animal species in the contemporary data. Conditioning on size reduces the median evolvability by 43% and reveals a strong correlation between conditional and unconditional evolvabilities (fig. S8, R^2 = 87%). Although a reduction of 43% is by itself insufficient to cause substantial genetic constraints under directional selection, it is possible that conditioning on more aspects of the organism than just size would reduce the evolvability of focal traits more drastically. This would increase the viability of the genetic-constraints hypothesis beyond microevolutionary timescales and could explain some of the observed divergence-evolvability relationship.

Fluctuating selection and genetic constraints

Genetic constraints could also influence divergence if the evolutionary timescale is very short. Evolution is not a rectilinear process, and on timescales below a million years or so it mostly takes the form of bounded fluctuations with a constant (stationary) distribution so that evolutionary changes do not accumulate with time (*43*, *44,* fig. S7A). This is likely caused by populations tracking adaptive optima that fluctuate within a limited range. Under this scenario, traits with low evolvabilities would lag further behind their optima and change less than traits with higher evolvabilities as illustrated in Figure 5B. To explore if such dynamics could rescue a version of the genetic-constraints hypothesis we show (*24*; based on *14*) that with stationary fluctuations, the predicted variance in the log trait mean would be $D = Vr/(r + \alpha)$, where r is the tracking rate of adaptation, and V and α are the stationary variance and the return rate of the fluctuating optimum, respectively. Because the tracking rate is proportional to evolvability, we predict a positive relationship between divergence and evolvability when the tracking rate is equal or slower than the rate of environmental fluctuations (*14, 24*).

Fitting this model to the combined fossil time-series data revealed predominantly stationary dynamics and returned high rates of both tracking and fluctuations in the optimum, with plausible half-lives in the range from one and up to a hundred years at most (Fig. 5A, fig. S7B). As can be seen from the likelihood surface (fig. S7B), the fitted model is symmetric for r and α, and cannot tell if r is smaller or larger than α. Nevertheless, as illustrated in Figure 5B, our estimated evolvabilities combined with reasonable strengths of selection will often generate tracking rates in the range of tens and hundreds of generations making it plausible that adaptation is slower than at least part of the environmental fluctuations of the optimum and able to influence the extent of fluctuations in the trait mean.

Hence, tracking fast stationary fluctuations in optima can plausibly account for an effect of evolvability on evolutionary divergence in both extant and extinct populations. On above million-year timescales divergence may start to accumulate, and evolution may include rare bursts of change to new adaptive zones (*1*, *44*). Even so, rapid stationary fluctuations would still constitute a component of the among-species variance (*44*) and could therefore explain an influence of evolvability also on this level. This is in line with the weaker relationship between divergence and evolvability observed in the contemporary species data. Note, however, that this model would not generate a phylogenetic signal by itself, and it cannot explain the strong association between evolvability and rates of evolution across million-year timescales found in studies of homogenous morphological traits such as dipteran wings, which are dominated by non-stationary Brownian-motions-like evolution (*15, 18*).

Conclusion

With two large and independent datasets we have established the existence of a positive scaling relationship between evolutionary divergence and evolvability, thus providing a link between micro- and macroevolution. After eliminating alternative explanations, we conclude that this pattern most plausibly results from genetic constraints on evolution under rapid stationary fluctuations. We have shown that if stabilizing selection around optima is not too strong, even high observed evolvabilities may cause constraints limiting trait fluctuations. If pleiotropic constraints further reduce evolvability along directions of selection, stronger selection and/or slower fluctuations of optima can be accommodated, leaving genetic constraints on stationary fluctuations a robust explanation for the divergence-evolvability relationship.

References and Notes

1. G. G. Simpson, *Tempo and Mode in Evolution.* (Columbia University Press, 1944).

2. A. G. Kluge, W. C. Kerfoot, The predictability and regularity of character divergence. *Am. Nat.* **107**, 426-442 (1973).
3. R. R. Sokal, The Kluge-Kerfoot phenomenon reexamined. *Am. Nat.* **110**, 1077-1091 (1976).
4. F. J. Rohlf, A. J. Gilmartin, G. Hart, The Kluge-Kerfoot phenomenon—a statistical artifact. *Evolution* **37**, 180-202 (1983).
5. D. Lofsvold, Quantitative genetics of morphological differentiation in *Peromyscus*. II. Analysis of selection and drift. *Evolution* **42**, 54-67 (1988).
6. D. L. Venable, M. A. Búrquez, Quantitative genetics of size, shape, life-history, and fruit characteristics of the seed heteromorphic composite *Heterosperma pinnatum*. II. Correlation structure. *Evolution* **44**, 1748-1763 (1990).
7. D. Schluter, Adaptive radiation along genetic lines of least resistance. *Evolution* **50**, 1766-1774 (1996).
8. S. Andersson, Genetic constraints on phenotypic evolution in *Nigella* (Ranunculaceae). *Biol. J. Linn. Soc. Lond.* **62**, 519-532 (1997).
9. A. V. Badyaev, G. E. Hill, The evolution of sexual dimorphism in the house finch. I. Population divergence in morphological covariance structure. *Evolution* **54**, 1784-1794 (2000).
10. G. Marroig, J. M. Cheverud, Size as a line of least evolutionary resistance: diet and adaptive morphological radiation in New World monkeys. *Evolution* **59**, 1128-1142 (2005).
11. S. F. Chenoweth, M. W. Blows, Q_{ST} meets the G matrix: the dimensionality of adaptive divergence in multiple correlated quantitative traits. *Evolution* **62**, 1437-1449 (2008).
12. P. A. Hohenlohe, S. J. Arnold, MIPoD: a hypothesis-testing framework for microevolutionary inference from patterns of divergence. *Am. Nat.* **171**, 366-385 (2008).
13. C. B. Kimmel *et al.*, Independent axes of genetic variation and parallel evolutionary divergence of opercle bone shape in threespine stickleback. *Evolution* **66**, 419-434 (2012).
14. G. H. Bolstad *et al.*, Genetic constraints predict evolutionary divergence in Dalechampia blossoms. *Phil. Trans. R. Soc. B* **369**, 20130255 (2014).
15. D. Houle, G. H. Bolstad, K. van der Linde, T. F. Hansen, Mutation predicts 40 million years of fly wing evolution. *Nature* **548**, 447-450 (2017).
16. J. W. McGlothlin *et al.*, Adaptive radiation along a deeply conserved genetic line of least resistance in *Anolis* lizards. *Evol. Lett.* **2**, 310-322 (2018).
17. Ø. H. Opedal *et al.*, Evolvability and trait function predict phenotypic divergence of plant populations. *PNAS* **120**, e2203228120 (2023).
18. P. T. Rohner, D. Berger, Developmental bias predicts 60 million years of wing shape evolution. *PNAS* **120**, e2211210120 (2023).
19. T. F. Hansen, D. Houle, Measuring and comparing evolvability and constraint in multivariate characters. *J. Evol. Biol.* **21**, 1201-1219 (2008).
20. T. F. Hansen, K. L. Voje, Deviation from the line of least resistance does not exclude genetic constraints: a comment on Berner et al.(2010). *Evolution* **65**, 1821-1822 (2011).
21. K. L. Voje *et al.*, Does lack of evolvability constrain adaptation? If so, on what timescales. In *Evolvability: A Unifying Concept in Evolutionary Biology?,* T. F. Hansen, D. Houle, M. Pavličev, C. Pélabon, Eds. (MIT press, 2023), chap. 14, pp. 289-306.

22. D. Houle, Comparing evolvability and variability of quantitative traits. *Genetics* **130**, 195-204 (1992).
23. T. F. Hansen, C. Pélabon, W. S. Armbruster, M. L. Carlson, Evolvability and genetic constraint in *Dalechampia* blossoms: components of variance and measures of evolvability. *J. Evol. Biol* **16**, 754-766 (2003).
24. Materials and methods are available as supplementary materials.
25. R. Lande, On comparing coefficients of variation. *Syst. Zool.* **26**, 214-217 (1977).
26. F. Guillaume, M. C. Whitlock, Effects of migration on the genetic covariance matrix. *Evolution* **61**, 2398-2409 (2007).
27. M. G. Bulmer, *The mathematical theory of quantitative genetics*. (Clarendon Press, 1980).
28. J. Tufto, Quantitative genetic models for the balance between migration and stabilizing selection. *Genetics Research* **76**, 285-293 (2000).
29. M. A. Stamp, J. D. Hadfield, The relative importance of plasticity versus genetic differentiation in explaining between population differences; a meta-analysis. *Ecology Letters* **23**, 1432-1441 (2020).
30. D. W. Noble, R. Radersma, T. Uller, Plastic responses to novel environments are biased towards phenotype dimensions with high additive genetic variation. *PNAS*, 201821066 (2019).
31. R. Radersma, D. W. Noble, T. Uller, Plasticity leaves a phenotypic signature during local adaptation. *Evolution letters* **4**, 360-370 (2020).
32. F. Johansson, P. C. Watts, S. Sniegula, D. Berger, Natural selection mediated by seasonal time constraints increases the alignment between evolvability and developmental plasticity. *Evolution* **75**, 464-475 (2021).
33. S. J. Arnold, Phenotypic Evolution: The Ongoing Synthesis. *Am. Nat.* **183**, 729-746 (2014).
34. W. Armbruster, K. Schwaegerle, Causes of covariation of phenotypic traits among populations. *J. Evol. Biol.* **9**, 261-276 (1996).
35. S. J. Arnold, M. E. Pfrender, A. G. Jones, The adaptive landscape as a conceptual bridge between micro-and macroevolution. *Genetica* **112/113**, 9-32 (2001).
36. R. Lande, The genetic covariance between characters maintained by pleiotropic mutations. *Genetics* **94**, 203-215 (1980).
37. T. F. Hansen, J. M. Álvarez-Castro, A. J. R. Carter, J. Hermisson, G. P. Wagner, Evolution of genetic architecture under directional selection. *Evolution* **60**, 1523-1536 (2006).
38. A. Le Rouzic, J. M. Álvarez-Castro, T. F. Hansen, The evolution of canalization and evolvability in stable and fluctuating environments. *Evol. Biol.* **40**, 317-340 (2013).
39. T. F. Hansen, G. P. Wagner, The evolution of evolvability. In *Evolvability: A Unifying Concept in Evolutionary Biology?,* T. F. Hansen, D. Houle, M. Pavličev, C. Pélabon, Eds. (MIT press, 2023), chap. 7, pp. 121-145.
40. R. Lande, Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314-334 (1976).
41. M. Lynch, W. G. Hill, Phenotypic evolution by neutral mutation. *Evolution* **40**, 915-935 (1986).
42. M. Lynch, The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *Am. Nat* **136**, 727-741 (1990).
43. S. Estes, S. J. Arnold, Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *Am. Nat* **169**, 227-244 (2007).

44. J. C. Uyeda, T. F. Hansen, S. J. Arnold, J. Pienaar, The million-year wait for macroevolutionary bursts. *PNAS* **108**, 15908-15913 (2011).
45. P. Gingerich, Rates of evolution: effects of time and temporal scaling. *Science* **222**, 159-162 (1983).
46. P. D. Gingerich, Rates of evolution on the time scale of the evolutionary process. *Genetica* **112/113**, 127-144 (2001).
47. P. D. Gingerich, Rates of evolution. *Annu. Rev. Ecol. Evol. Syst.* **40**, 657-675 (2009).
48. L. J. Harmon *et al.*, Causes and consequences of apparent timescaling across all estimated evolutionary rates. *Annu. Rev. Ecol. Evol. Syst.* **52**, 587-609 (2021).
49. S. J. Arnold, Constraints on phenotypic evolution. *Am. Nat.* **140**, S85-S107 (1992).
50. J. Hereford, T. F. Hansen, D. Houle, Comparing strengths of directional selection: how strong is strong? *Evolution* **58**, 2133-2143 (2004).
51. T. F. Hansen, D. Houle, Evolvability, stabilizing selection, and the problem of stasis. In *Phenotypic integration: Studying the ecology and evolution of complex phenotypes,* M. Pigliucci, K. Preston, Eds. (Oxford University Press, 2004), pp. 130-150.
52. T. F. Hansen, W. S. armbruster, M. L. Carlson, C. Pélabon, Evolvability and genetic constraint in *Dalechampia* blossoms: genetic correlations and conditional evolvability. *J. Exp. Zool. B Mol. Dev. Evol.* **296**, 23-39 (2003).
53. T. F. Hansen, T. M. Solvin, M. Pavlicev, Predicting evolutionary potential: A numerical test of evolvability measures. *Evolution* **73**, 689-703 (2019).
54. M. Grey, Z. V. Finkel, P. K. Pufahl, L. M. Reid, Evolutionary mode of the ostracod, Velatomorpha altilis, from the Joggins Fossil Cliffs UNESCO World Heritage Site. *Lethaia* **45**, 615-623 (2012).
55. T. F. Hansen, C. Pélabon, D. Houle, Heritability is not evolvability. *Evol. Biol.* **38**, 258-277 (2011).
56. G. Matthews, S. Hangartner, D. G. Chapple, T. Connallon, Quantifying maladaptation during the evolution of sexual dimorphism. *Proc. Royal Soc. B* **286**, 20191372 (2019).
57. Ø. H. Opedal, The evolvability of animal-pollinated flowers: towards predicting adaptation to novel pollinator communities. *New Phytol.* **221**, 1128-1135 (2019).
58. K. L. Voje, Phenotypic Evolution Time Series (PETS) Database, version 1.0 (2023). https://pets.nhm.uio.no/
59. D. Bates, M. Mächler, B. M. Bolker, S. C. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1-48 (2015).
60. K. Barton, MuMIn: multi-model inference. R package version 1.43.6 (2009). http://r-forge.r-project.org/projects/mumin/
61. T. F. Hansen, K. Bartoszek, Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. *Syst. Biol.* **61**, 413-425 (2012).
62. M. Lynch, B. Walsh, Genetics and analysis of quantitative traits. (Sinauer Associates, 1998).
63. T. F. Hansen, Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**, 1341-1351 (1997).
64. T. F. Hansen, Variation, inheritance, and evolution: A primer on evolutionary quantitative genetics. In *Evolvability: A Unifying Concept in Evolutionary Biology?,* T. F. Hansen, D. Houle, M. Pavličev, C. Pélabon, Eds. (MIT press, 2023), chap. 5, pp. 73-100.
65. M. Lynch, The rate of polygenic mutation. *Genet. Res.* **51**, 137-148 (1988).

Acknowledgments: We thank the Centre of Advanced Study (CAS) at the Norwegian Academy of Science and Letters for hosting us during the academic year of 2019/2020 when part of this project was conducted.

Funding:

Norwegian Research Council grant 287214 (CP)

Swedish Research Council grant 2021-04777 (ØHO)

The Crafoord Foundation grant 20210661 (ØHO)

European Research Council StG 948465 (KLV)

Norwegian Research Council grant 275862 (GHB)

Author contributions:

Conceptualization: AH, TFH, CP, GHB, KLV, ØHO

Data curation: KLV, ØHO, AH

Formal Analysis: AH, KLV, GHB, ØHO, SB

Methodology: AH, TFH, CP, GHB, KLV, ØHO

Visualization: AH, GHB

Funding acquisition: CP, KLV, ØHO, GHB

Project administration: CP

Writing – original draft: AH

Writing – review & editing: all authors

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Data and R code underlying the analyses available at Dryad: https://doi.org/10.5061/dryad.4j0zpc8hx

Supplementary Materials

Materials and Methods

Figs. S1 to S8

Tables S1 to S3

References (*55*–*65*)

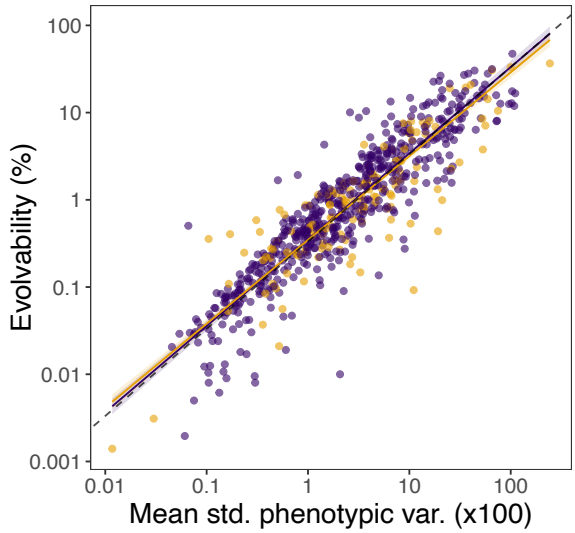

Fig. 1. Isometric scaling between phenotypic and genetic variation. Mean-scaled genetic variance, evolvability (e_μ), expressed as percent evolutionary change regressed against mean-scaled phenotypic variance ($p_\mu \times 100$). For the morphological traits (purple line, n = 527), $\ln(e_\mu)$ = -1.10 (±0.08) + 0.99 (±0.02) $\ln(p_\mu)$ and R^2 = 83%. For all traits (orange line, n = 669), $\ln(e_\mu)$ = -1.24 (±0.08) + 0.96 (±0.02) $\ln(p_\mu)$ and R^2 = 81%. The coefficients are obtained from least-squares regressions fitted to log-transformed variables, and the slopes±SE are corrected for attenuation (= 0.6%) due to estimation error in the phenotypic variance. The dashed line shows isometric scaling. We considered the exponent of the intercept (-1.1) from the morphological traits as an estimate of heritability, $h^2 = e^{-1.1} = 0.33$, and used this to predict evolvabilities from phenotypic variances in the fossil data. The near isometry and high R^2 makes this a good prediction over the range of the data.

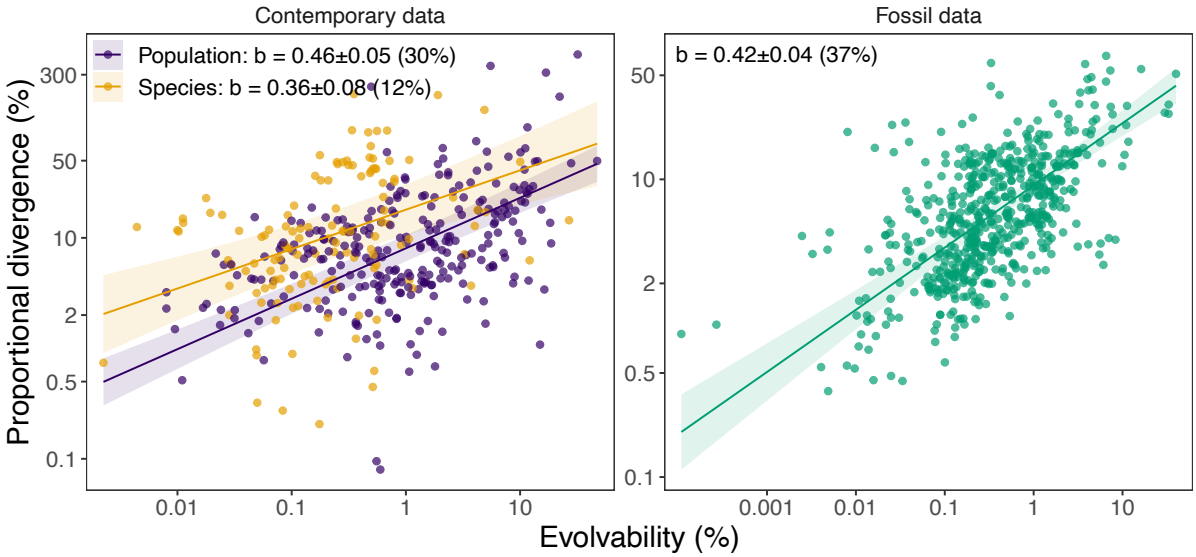

Fig. 2. Evolvability predicts divergence among populations, species, and fossil samples. Divergence (d) is the expected percent change in magnitude from the trait grand mean (*24*). Evolvability (e_μ) is the mean-scaled additive genetic variance expressed as percent evolutionary change under unit selection. For the fossil data evolvability is predicted by multiplying the sample variance by the heritability ($h^2 = 0.33$) obtained in Fig.1. The scaling exponents (b±SE) and marginal R^2 (%) are obtained from mixed-effect models fitted to log-transformed variables and are corrected for attenuation bias of 13%, 17% and 12% for the population ($n = 271$), species ($n = 130$), and fossil data ($n = 589$), respectively.

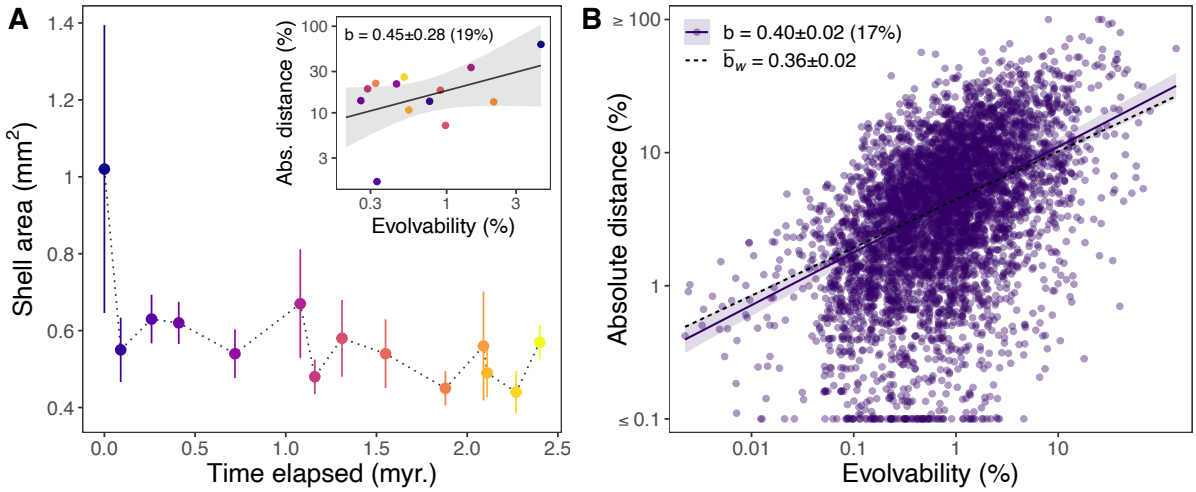

Fig. 3. Evolvability predicts divergence between consecutive fossil samples. (**A**) An example of a fossil time series from Grey et al. (*54*), for which sample means of shell area of the ostracod *Velatomorpha altilis* are shown across time elapsed since the oldest sample (bars indicate sample standard deviation). We converted sample variances into estimates of evolvability (e_μ) as explained in Fig. 1 and used these to predict the absolute morphological distance to the next sample on log scale (inset figure). Sample 1 (dark blue) has a large variance and there is a large difference between the mean shell area of sample 1 and sample 2. Hence, the evolvability of sample 1 and the absolute morphological distance from sample 2 correspond to the point in the right upper corner of the inset figure. Plot (**B**) shows the relationship between evolvability and absolute morphological distance to the next sample for all cases with a sample size of at least 30 specimens (n = 5009). The slope (b±SE) and marginal R^2 (%) are obtained from a mixed-effect model fitted to log-transformed variables. The dashed line shows the average within-time-series slope ($\bar{b}_w$±SE). Both slopes±SE are corrected for a 2.1% attenuation bias.

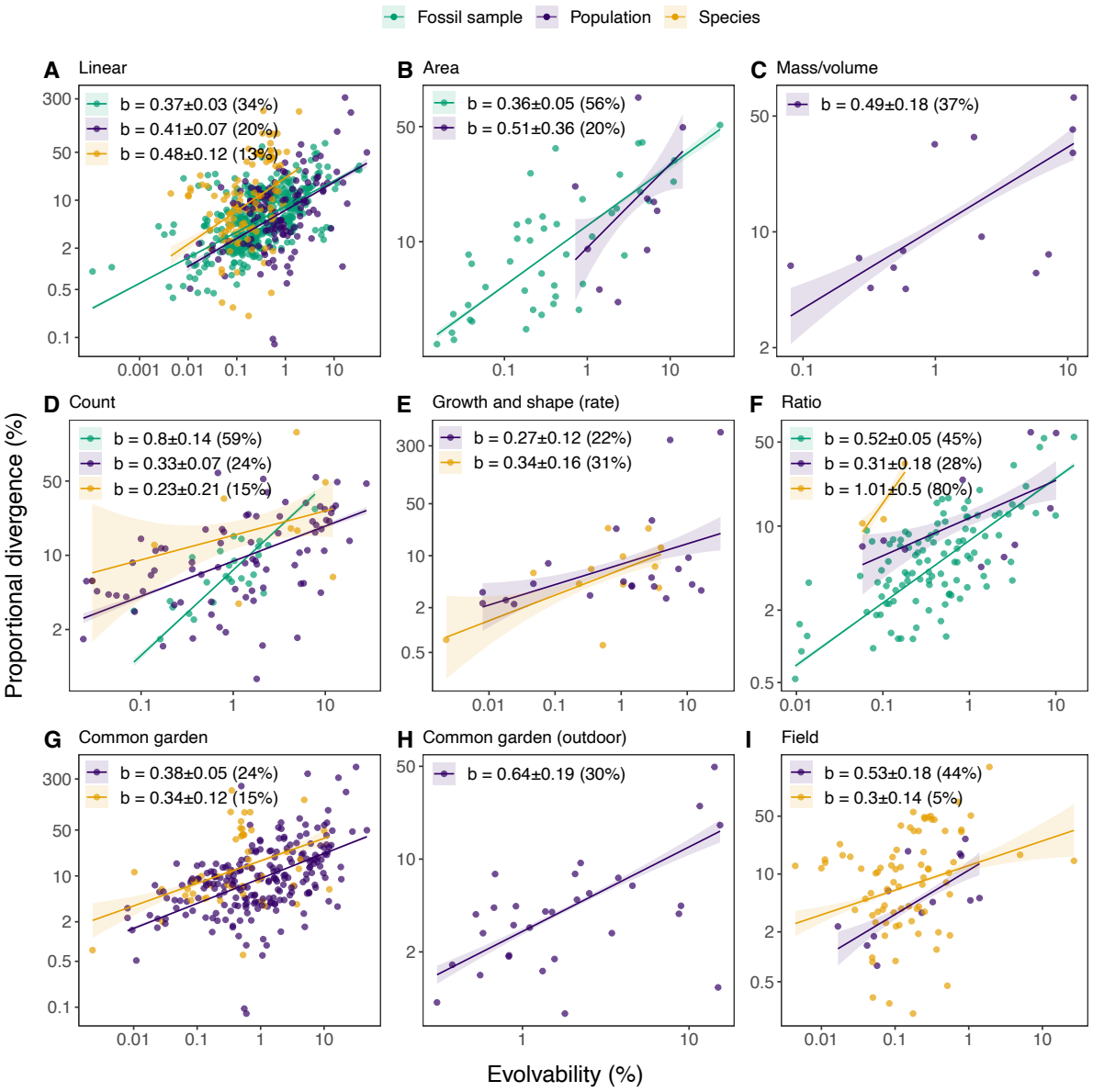

Fig. 4. The divergence-evolvability relationship across different trait types and environments. The magnitude of divergence (d) is the expected percent change of a population, species, or fossil sample from the trait grand mean. Evolvability (e_μ) is the mean-scaled additive genetic variance expressed as percent evolutionary change under unit selection. For the fossil data evolvability is predicted by multiplying the sample variance by the heritability ($h^2 = 0.33$) obtained in Fig.1. The scaling exponents (b±SE) are the slopes from the log-log regression from each group corrected for attenuation bias. The R^2 (%) is given in parenthesis.

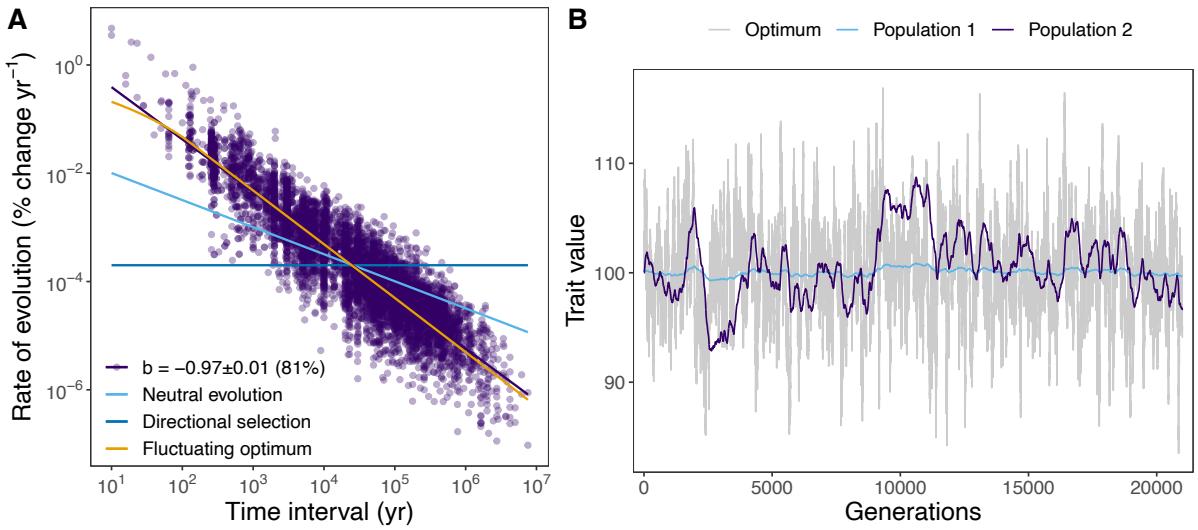

Fig. 5. Stationary fluctuations in optima and traits. (**A**) Rate of evolution (magnitude of change per year) against the time interval between two consecutive samples in the fossil data (n = 6231). The magnitude of change (d_c) is corrected for estimation error in the sample means (*24*). The scaling exponent, b±SE (R^2), is the slope of the least-squares regression on log-transformed data. Also represented are regression lines predicted by neutral evolution (light blue, scaling exponent = -0.5) and directional selection (dark blue, scaling exponent = 0) that are fitted to the data with an intercept at the grand mean of both variables and a slope set to the theoretical prediction (*43, 45-47*). The prediction for the fluctuating-selection model (orange) is based on an Ornstein-Uhlenbeck model with parameters fitted using a grid search (*24,* fig. S6B), where the half-lives of the return rate of the optimum and the tracking are $\ln(2)/a$ = 1 year and $\ln(2)/r$ = 30.5 years, respectively. (**B**) Simulation showing evolutionary changes in the trait mean of two populations with different evolvabilities tracking a fluctuating optimum (grey curve). The half-life of the return rate of the optimum to its central value is set to 30 years (fig. S6B). If we assume one generation equals one year, then keeping $r \leq \alpha$ requires $r < 0.023$ per generation. Because r equals the evolvability times the mean-scaled quadratic selection gradient around the optimum, γ, we can compute that an evolvability of 0.56% (the median of our data) would require $|\gamma| < 4.1$ to keep $r < 0.023$. With this value of γ, a 10% trait shift from the optimum reduces fitness with 2%. Population 1 (light blue curve) has an evolvability of 0.01% and hence a slower tracking rate ($r = \gamma e_\mu$) of the optimum and smaller fluctuations in the trait mean than population 2 (e_μ = 0.1%, purple curve).

Science
AAAS

Supplementary Materials for

Evolvability predicts macroevolution under fluctuating selection

Agnes Holstad, Kjetil L. Voje, Øystein H. Opedal, Geir H. Bolstad, Salomé Bourg, Thomas F. Hansen, Christophe Pélabon

Correspondence to: agnes.holstad@ntnu.no

This PDF file includes:

Materials and Methods

Data collection

Contemporary data: We searched the primary scientific literature to compile a meta-database on within-population genetic variation and among-population or -species phenotypic variation. Only studies in which populations or species had diverged in the wild were included. This excluded studies in which divergence resulted from artificial selection, populations kept and bred in the lab (except for the generations in common garden for estimating genetic divergence) or populations constructed from inbred lines. The statistics required for these data were at least one estimate of additive genetic variance with corresponding mean and the mean for two or more populations or species for the same trait. In some cases, different studies (papers) were combined to obtain these data. This collection was based on (i) studies gathered by GHB, TFH and CP for a review on genetic constraints (*14*), (ii) studies included in the analyses of Hansen et al. (*55*), Matthews et al. (*56*), Noble et al. (*30*), Opedal (*57*) and Opedal et al. (*17*) and (iii) an additional search conducted with ISI Web of Science (date: 04/12/2019) with the key words "additive genetic variance" and "divergence" for all collections and all years.

In total, the database contains 411 unique traits from 124 different species from 55 published studies. The number of populations or species per unique trait ranged from 2 to 41. A total of 2061 phenotypic variances, 1044 additive genetic variances, and 2687 trait means were retrieved.

Fossil data: Fossil data come from the Phenotypic Evolution Time Series (PETS) Database version 1.0 (*58*) curated by KLV. Contemporary time series were excluded, and only fossil time series with traits on a ratio scale were included to allow for estimation of mean-scaled variance. This resulted in a total of 589 evolutionary (fossil) time series from 150 independent lineages and 101 studies. Each time series represents changes across consecutive fossil samples for a single trait. Data for each fossil sample ("population") in a time series include mean trait size, sample variance, sample size and the geological age of the fossil sample.

Analyses of contemporary data

We measured evolvability on a proportional scale as the additive genetic variance divided by the square of the mean. This measure is derived from the mean-standardized Lande equation (*23, 55*):

$$\frac{\Delta\bar{z}}{\bar{z}} = \frac{V_A}{\bar{z}^2}\beta\bar{z},$$

where $\Delta\bar{z}/\bar{z}$ is the proportional change in trait mean, $\bar{z}$, over one generation, $V_A/\bar{z}^2$ is the mean-scaled additive genetic variance (i.e. evolvability, e_μ, on a proportional scale) and $\beta\bar{z}$ is the mean-scaled selection gradient. On this scale, evolvability can be interpreted as the expected proportional change in the trait mean under a mean-scaled selection gradient of 1, that is, under directional selection as strong as selection on fitness itself (*23, 50*). We computed mean-scaled evolvabilities as additive genetic variance estimates divided by the trait mean squared or by using additive genetic variance estimates computed on log-transformed trait values. These standardizations are asymptotically equivalent with negligible loss in accuracy when $e_\mu < 0.1$ (10%), which is the case for 93% of our estimates. When more than one estimate of evolvability were available per trait, we used the unweighted mean of these. We did not calculate a weighted mean because only 32% of the collected additive variance estimates were reported with error.

Divergence variance (D) between populations or species was also calculated on a proportional scale as the variance in the log-transformed trait means across populations or species, D = var[log($\bar{z}_j$)], where $\bar{z}_j$ is the trait mean in population j. We also estimated a divergence variance corrected for sampling error in the mean for a subset of the data with given sample sizes as $D_c = D - \mathrm{E}[\mathrm{SE}^2/\bar{z}^2]$. To facilitate interpretation, we converted D into a magnitude of divergence (d) expressed as expected proportional change of a population or species from the trait grand mean, $d = \exp(\sqrt{2D\pi^{-1}})$. The multiplication of $\sqrt{D}$ by $\sqrt{2\pi^{-1}}$ gives the expected value of a folded normal distribution and taking the exponent returns the estimate on an arithmetic scale. By subtracting one and multiplying by a hundred we expressed the estimates as percent change. We estimated the corrected magnitude of divergence (d_c) by substituting D_c for D.

To analyze the relationship between evolvability and divergence we fitted mixed-effect models using the "lme4" package in R (*59*), with divergence as the response variable, evolvability as the fixed effect and closest shared taxon (i.e., species or genus) as the random effect. The evolvability and divergence variables were log-transformed prior to the analyses. Marginal coefficients of determination (R^2) were computed using the "MuMIn" R package (*60*). Models were fitted for among-population variance and among-species variance separately in the extant-species data. We also fitted models for different subcategories of the data, such as trait type (morphological, life-history and physiological), trait dimension (linear, area, mass or volume, counts, rate and unitless ratio), or taxa (plants and animals) to investigate if the scaling relationship between divergence and evolvability remained similar within groups.

To account for the attenuation of the slope due to measurement error in the predictor variable (evolvability) we divided fixed-effect slopes and their standard errors with an attenuation factor, K, computed following Hansen and Bartoszek (*61*). The attenuation factor was computed as

$$K = 1 - \frac{\mathbf{x}^{\mathrm{T}}\mathbf{V}_u\mathbf{V}_x^{-1}\mathbf{x}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}},$$

where $\mathbf{x}$ is the vector of predictor variables (evolvabilities) centered on the mean, $\mathbf{V}_u$ and $\mathbf{V}_x$ are (in our case) diagonal variance matrices containing respectively measurement and total variances of the predictor variable. In Figure 2, the diagonal elements of $\mathbf{V}_u$ is the variance among evolvability estimates of different populations/species/samples of the same trait (σ^2_{ui}), as we assume this is larger than the square of the standard errors of the evolvability estimates. The diagonal elements of $\mathbf{V}_x$ are $\sigma_x^2 - \overline{\sigma_{\mu i}^2} + \sigma_{\mu i}^2$, where σ^2_x is the empirical variance of the predictor variable. In Figures 1 and 3B, the diagonal elements of $\mathbf{V}_u$ is the measurement error variance ($\sigma^2_{ui} = 2/(n + 2)$) (*62* p. 815) of the mean-scaled phenotypic variance and evolvability estimates, respectively.

To assess to what extent evolvability measures decrease when at least one correlated trait is not allowed to evolve, we estimated evolvabilities conditioned on size using the conditional evolvability from Hansen et al. (*19, 52*):

$$c(y|x) = G_y\left(1 - r_{xy}^2\right),$$

where G_y is the variance of the focal trait y and r_{xy}^2 is the squared genetic correlation coefficient between focal trait y and trait x on which y is conditioned on.

To assess the effect of phenotypic plasticity on the evolvability-divergence relationship, we compared the relationship obtained with divergence data collected in either controlled indoor, controlled outdoor or field environment. If the evolvability-divergence relationship were mostly due to correlations between evolvability and phenotypic plasticity, we would expect the relationship between evolvability and divergence to vanish for the data obtained under controlled environments where the effect of plasticity has been strongly reduced.

To quantify the constraints generated by univariate evolvabilities, we considered the number of generations that would be necessary to produce a divergence of a given magnitude (d) under a typical mean-scaled selection gradient (β_μ) per generation (*50*) with a given evolvability e_μ. This can be calculated as, $n_{gen} = \ln(1+d)/\ln(1+ e_\mu\beta_\mu)$ (*53*). Here, we used $\beta_\mu = 0.3$ that correspond to the median selection gradient reported by Hereford et al. (*50*), and the median eμ = 0.56% observed in the contemporary data.

Analyses of fossil data

In a first analysis equivalent to the analysis for the extant data, we investigated if the mean sample variance per time series predicted the divergence among all the fossil samples in that time series, thus estimating a single measure of divergence per time series. To reduce the bias from unreliable estimates of individual sample variances, we estimated a sample-size-weighted average of the sample variance (p_μ) for a time series as:

$$p_\mu = \frac{\sum_j \sigma_j^2 (n_j - 1)}{\sum_j (n_j - 1)},$$

where σ_j^2 is the proportional (mean-scaled) variance of sample j and n_j is the size of sample j.

Using the contemporary species data, we showed that mean-scaled additive genetic (e_μ) and phenotypic variances (p_μ) for morphological traits scaled isometrically on a log-log scale (slope = 0.99±0.02 corrected for 0.6% attenuation, R^2 = 83%). On the original scale, the intercept of this relationship (-1.10±0.08 when $\log(p_\mu) = 0$, hence, $p_\mu = 1$) represents the average heritability, $h^2 = e^{-1.1} = 0.33$, across all measures, as

$$h^2 = \frac{V_A}{V_p} = \frac{e_\mu}{p_\mu}.$$

We used this heritability to predict evolvabilities from phenotypic variances in the fossil data as $e_\mu = h^2 p_\mu$. Divergence was computed as the variance in sample ("population") means, D = var[ln($\bar{z}_j$)] per time series ("trait"), and the magnitude of divergence, $d = \exp(\sqrt{2D\pi^{-1}})$, as in the contemporary data. The corrected divergence (D_c) and magnitude of divergence (d_c) estimates were also calculated as in the contemporary data. Mixed-effect models were then fitted with divergence as the response variable, evolvability as the fixed effect and species and study as the random effects.

The fossil data also allow the use of evolvability measured for each sample ("population") to predict the divergence during the next timestep in the timeseries, i.e., the phenotypic distance in the trait mean between two consecutive samples. To decrease the variation in sampling variance resulting from the irregular number of specimens measured per sample in this analysis, we only computed the evolvability from samples with at least 30 measured specimens. This resulted in a dataset of 5009 fossil samples across 382 time series from 96 independent lineages. Proportional variance for a fossil "population" was calculated as var[log(z_i)], where z_i is specimen i in the sample. We used the observed heritability from the contemporary data (Fig. 1) to estimate evolvability of the fossil samples as $e_\mu = h^2$var[log(z_i)]. Morphological distance between two consecutive sample means defined the magnitude of divergence and was calculated as, $d = |\ln(\bar{z}_j) - \ln(\bar{z}_{j+1})|$, where $\bar{z}_j$ is the trait mean in sample j. We also calculated a magnitude of divergence corrected for sampling error variance as $d_c = \sqrt{d^2 - \mathrm{E}[\mathrm{SE}^2/\bar{z}^2]}$. Mixed-effect models were then fitted with divergence as the response variable, evolvability as the fixed effect and time series nested within study, and species, as the random effects.

Microevolution during the accumulation of a fossil sample may inflate estimates of evolvability. To quantify this effect, we predicted the amount of variance that could be explained by evolution during the time interval covered by the sample. We first fitted an Ornstein-Uhlenbeck model of evolution to each of the 382 time series and estimated the expected variance in the trait mean ($\bar{z}_i$) as a function of time, t, following Hansen (*63*): $\mathrm{var}[\bar{z}_i|\bar{z}_a] = \sigma^2/2\alpha(1 - e^{-2\alpha t})$, where $\bar{z}_a$ is the estimated ancestral state of the trait mean, σ^2 is the variance of the perturbations in the trait mean, and α determines the rate of adaptation toward the optimum. The sample variance corrected for anagenetic microevolution in the trait mean is then given by var[ln(z_i)] – var[$\bar{z}_i|\bar{z}_a$; t_{sample}], where t_{sample} is the time interval covered by the sample. The analyzed time series lacked information on the within-sample time interval. We thus calculated the maximum duration that could be covered by the samples as the duration of the entire time series divided by the number of samples, assuming that there was no temporal gap between samples in the time series. As a "worst-case scenario", we then used 50% of this maximum duration as the t_{sample} to calculate the sample variance corrected for anagenetic microevolution. We also calculated corrected sample variance with t_{sample} as 10% and 1% of max duration (fig. S1).

From each fitted Ornstein-Uhlenbeck model, we also calculated the distance from each fossil sample mean to the primary optimum to assess the effects of maladaptation on the predicted trait divergence. To test if the scaling relationship between evolvability and morphological distance changes when traits evolve across longer time span, we fitted mixed-effect models with log morphological distance as response variable and log evolvability as fixed effect. Time separating the sample means in millions of years, distance to the optimum, and sample size were fitted as covariates. The interaction terms of the covariates with evolvability were included. Trait dimension or type (linear, ratio, count), size (microfossil, macrofossil) and growth pattern (indeterminate, determinate) were included as fixed factors. Study was fitted as a random effect. Model fit was assessed using AICc and coefficients of determination (R^2). The best-fitting model included time, distance to the optimum, sample size (and their interaction terms) and trait type as fixed effects.

Tracking fluctuating selection

To formalize and test the hypothesis that genetic constraints limit the ability of populations to track rapid environmental fluctuations, we use a model from Bolstad et al. (*14*) in which the trait mean, z, tracks a moving optimum, θ, at a rate determined by the product of the evolvability, e_μ, and a linear selection gradient proportional to the distance of the trait mean to the optimum, $\beta = -\gamma(z - \theta)$. The optimum itself moves around an arbitrary central point (zero) according to an Ornstein-Uhlenbeck process. This is described by the stochastic differential equations:

$$dz = e_\mu \beta = -r(z - \theta)dt,$$
$$d\theta = -\alpha\theta\, dt + \sigma\, dB,$$

where $r = \gamma e_\mu$ measures how fast the trait tracks the optimum, dt is time differential, dB is the differential of a Brownian-motion process (i.e., a white noise), and σ is the square root of the average magnitude of random fluctuations per time. The parameter α measures how fast the optimum returns to its central value and is an inverse measure of autocorrelation in the movements of the optimum.

From the moments of this process, derived in Bolstad et al. (*14*), we compute the expected rate of evolution of the trait as

$$R = \frac{E[|z(t) - z(0)|]}{t} = \frac{2}{t}\sqrt{\frac{rV}{\pi(r + \alpha)}}\sqrt{1 - \frac{re^{-\alpha t} - \alpha e^{-rt}}{r - \alpha}},$$

where $V = \sigma^2/2\alpha$ is the stationary (stochastic equilibrium) variance of the fluctuations in the optimum. This expression is derived by assuming that the initial values of the trait and the optimum are random variables drawn from the stationary distribution of the process, and uses the fact that the expectation of the absolute value of the deviation of a normal variate from its mean is $\sqrt{2/\pi}$ times the standard deviation of the variate.

In the limit of $t \to 0$, the rate is a constant $R = \sqrt{2\alpha r^2 V/\pi(r + \alpha)}$, and for short time intervals the scaling exponent with time is zero, which means that the rate is unaffected by the time interval. As the time interval increases a negative scaling appears and for long intervals a scaling exponent of -1 is reached (i.e. the process behaves as white noise).

We fitted this rate function to the combined fossil time-series data by assuming that the parameters r, α and V were the same for all the time series (Fig. 5A, $n = 6231$). We used a mixed model with the deviance of the observed $\ln R$ from the predicted $\ln R$ as response variable with time series nested within study as random variables and assuming the residuals were gaussian. The rate of evolution, R, is calculated as $R = d_c/t$, where d_c is the magnitude of divergence corrected for sampling error variance as $d_c = \sqrt{(|\ln(\bar{z}_j) - \ln(\bar{z}_{j+1})|)^2 - \mathrm{E}[\mathrm{SE}^2/\bar{z}^2]}$ and t is time in years. We performed a grid search to calculate restricted likelihood over values of the parameters r, α and V. Results from the grid search are shown in Figure S7B.

When the process is stationary, the variance of the trait fluctuations reduces to $Vr/(r + \alpha)$. This expression is also an accurate approximation of the variance in the fluctuations of log trait values

(D), provided the variance of the optimum, V, is scaled with the trait mean squared. Assuming that the strength of stabilizing selection around the optimum is not related to the amplitude of fluctuations, the scaling exponent between divergence (D) and evolvability (proportional to r) would be unity when $r \ll \alpha$, about 1/2 when $r \approx \alpha$, and approach zero as $r \gg \alpha$. If we take our divergence measures to be the magnitude of the difference between two independent random observations, z_1 and z_2, from this process as $d = |\ln(z_1) - \ln(z_2)|$, then the expectation of d will be $2\sqrt{Vr/\pi(r+\alpha)}$. We can write this as

$$\ln(d) = \tfrac{1}{2}\ln(V) + \tfrac{1}{2}\ln(r) - \tfrac{1}{2}\ln(r+\alpha) + \mathrm{C},$$

where C is a constant. Assuming that $r << \alpha$, and decomposing r into its components e_μ and γ, we can write this as

$$\ln(d) = \tfrac{1}{2}\ln(V) + \tfrac{1}{2}\ln(e_\mu) + \tfrac{1}{2}\ln(\gamma) - \tfrac{1}{2}\ln(\alpha) + \mathrm{C}.$$

This shows that variation in divergence has four equally contributing components due to variation in the position of the optimum, V, due to evolvability, e_μ, due to strength of stabilizing selection around the optimum, γ, and due to the rate of fluctuation in the optimum, α. This predicts that one quarter of the variance in evolvability should appear as variance in divergence. Values presented in table S2 reveal that the variance in evolvability is larger than half the variance of divergence for the contemporary population and fossil data and a bit less than half for the species data. This fits well with evolvability explaining 20-40% of the variance in contemporary population and fossil divergence and a bit less for species divergence.

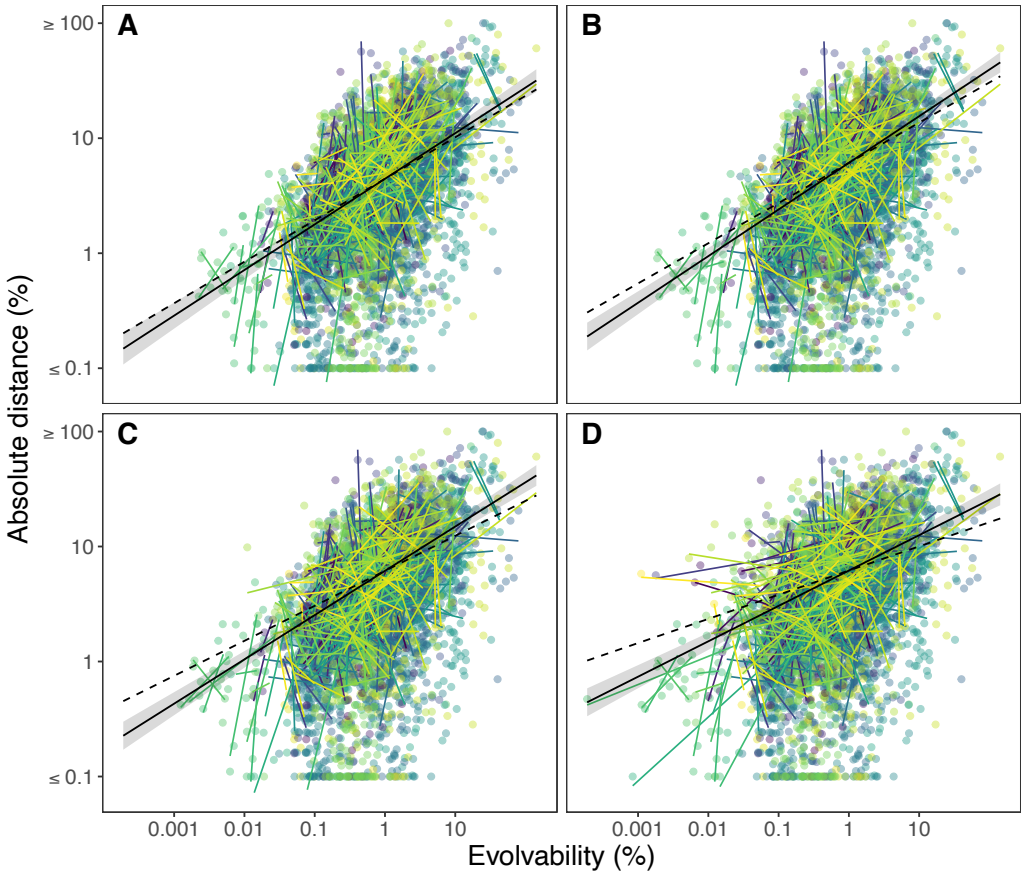

Fig. S1. Testing the effects of microevolution within fossil sample on the scaling relationship between evolvability and sample divergence.

Microevolution occurring in the time interval during which fossil samples accumulate may inflate our estimates of evolvability by confounding standing variation with evolutionary change and may affect the scaling relationship between evolvability and sample divergence. Here, we correct estimates of evolvability (sample variance) for anagenetic microevolution in the trait mean. Because most time series lacked information on the time interval within samples (t_{sample}), we estimated this time interval as a percent of the maximum duration covered by the fossil samples computed as the duration of the entire time series divided by the number of samples. Hence, 50% of max duration is considered as a "worst-case scenario", as this gives the average estimate of microevolution in the trait mean with a within-sample time interval similar to the temporal gap between the fossil samples. Solid lines represent the slope (b±SE) of the mixed-effect model (n = 5009) and the dashed line represent the average within time-series slopes ($\bar{b}_w$±SE), corrected for 2.1% attenuation bias. Colors represent different time series. (**A**) Original data (same as Fig 3B): Absolute distance over evolvability (within sample variance), b = 0.40±0.02%, R^2 = 17%, $\bar{b}_w$ = 0.36±0.02%. (**B**) Evolvability corrected with t_{sample} as 1% of max duration: b = 0.39±0.02, R^2 = 16%, $\bar{b}_w$ = 0.35±0.02%. (**C**) Evolvability corrected with t_{sample} as 10% of max duration: b = 0.37±0.02, R^2 = 15%, $\bar{b}_w$ = 0.30±0.01%. (**D**) Evolvability corrected with t_{sample} as 50% of max duration: b = 0.30±0.02, R^2 = 11%, $\bar{b}_w$ = 0.21±0.01%.

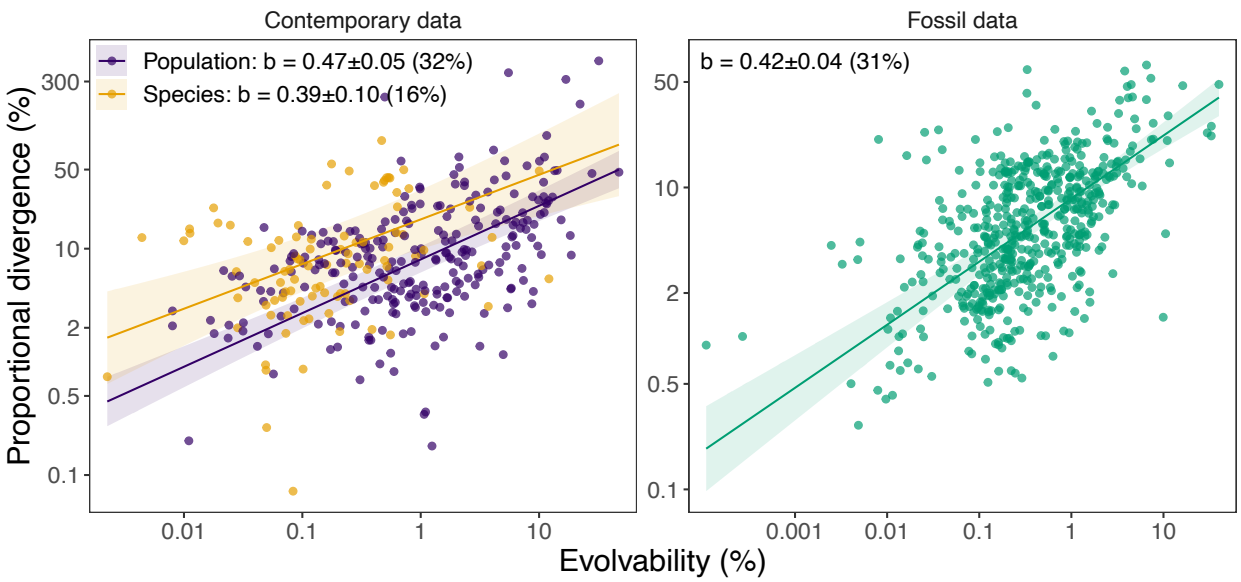

Fig. S2. The divergence-evolvability relationship among populations, species and fossil samples with divergence corrected for sampling error variance.

The magnitude of divergence (d_c) is the expected percent change of a population, species, or fossil sample from the trait grand mean, corrected for sampling error variance (see methods). Evolvability (e_μ) is the mean-scaled additive genetic variance expressed as percent evolutionary change under unit selection. For the fossil data evolvability is predicted by multiplying the sample variance by the heritability ($h^2 = 0.33$) obtained in Figure 1. The scaling exponents (b±SE) and marginal R^2 (%) are obtained from mixed-effect models fitted to log-transformed variables and the b±SE are corrected for attenuation bias of 14%, 15% and 13% for the population ($n = 257$), species ($n = 93$), and fossil data ($n = 544$), respectively.

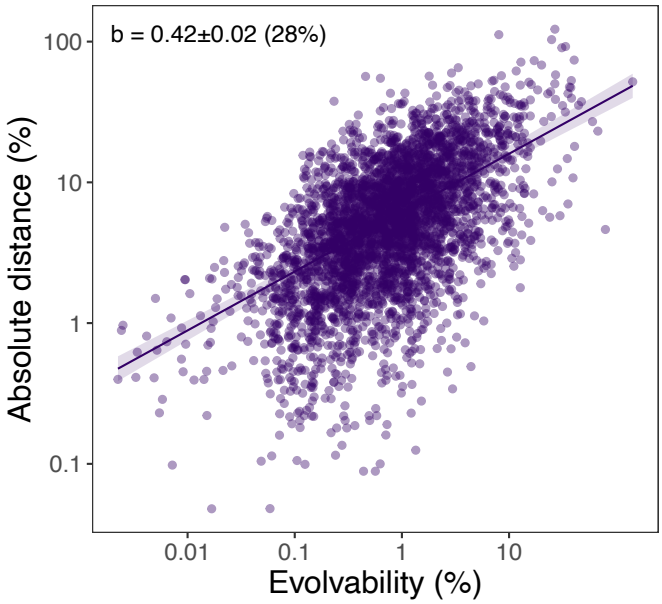

Fig. S3. The divergence-evolvability relationship among consecutive fossil samples with divergence corrected for sampling error variance.

The magnitude of divergence (d_c) is the absolute morphological distance to the next sample on log-scale, corrected for sampling error variance (see methods). Evolvability (e_μ) is predicted by multiplying the sample variance by the heritability ($h^2 = 0.33$) obtained in Fig.1 and expressed as percent evolutionary change under unit selection. The scaling exponent (b±SE) and marginal R^2 (%) are obtained from a mixed-effect model fitted to log-transformed variables ($n = 3572$) and the b±SE is corrected for 2.2% attenuation bias.

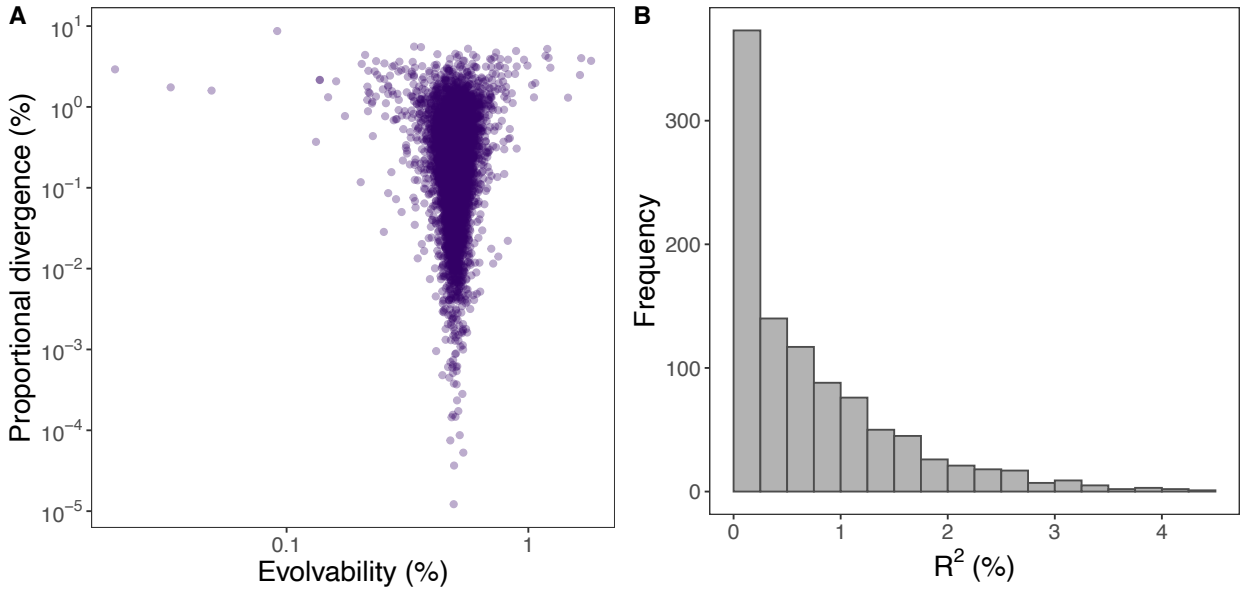

Fig. S4. Effect of correlated sampling error on the evolvability-divergence scaling.
Correlated sampling errors in means and variances could generate correlated inflation of within- and among-population variances if the sampling error has a directional bias in both variables. Assuming no true divergence among populations (i.e., same population means) and no differences in evolvability among populations, sampling errors in divergence estimates will be positively biased, while estimates of evolvability will have a symmetrical sampling error (i.e., no bias). We tested this by regressing population divergence against evolvability for 500 simulated traits. Each trait was simulated with two populations with the same trait mean and standard deviation (i.e., the two population means only differ due to sampling error). Correlated error was introduced by varying the sample size of the two populations from 2 to 500 with the paired populations having the same sample size. The evolvability (e_μ) and divergence (D) were estimated for each trait as $e_\mu = \mathrm{E}[\frac{\mathrm{var}[z_i]}{\bar{z}^2}]$ and $D = \mathrm{var}[\log(\bar{z}_j)]$, for individual i in population j. (**A**) An example of one regression of evolvability and divergence. The sampling error in evolvability is symmetric around the true evolvability while the sampling error in the divergence estimates is positively biased. $R^2 = 0.1\%$ and $n = 10000$. (**B**) Correlated errors did not generate any correlation between divergence and evolvability estimates, as illustrated by the distribution of the R^2 from the 1000 replicates of divergence-evolvability regressions with 500 traits each which barely exceeds 4%.

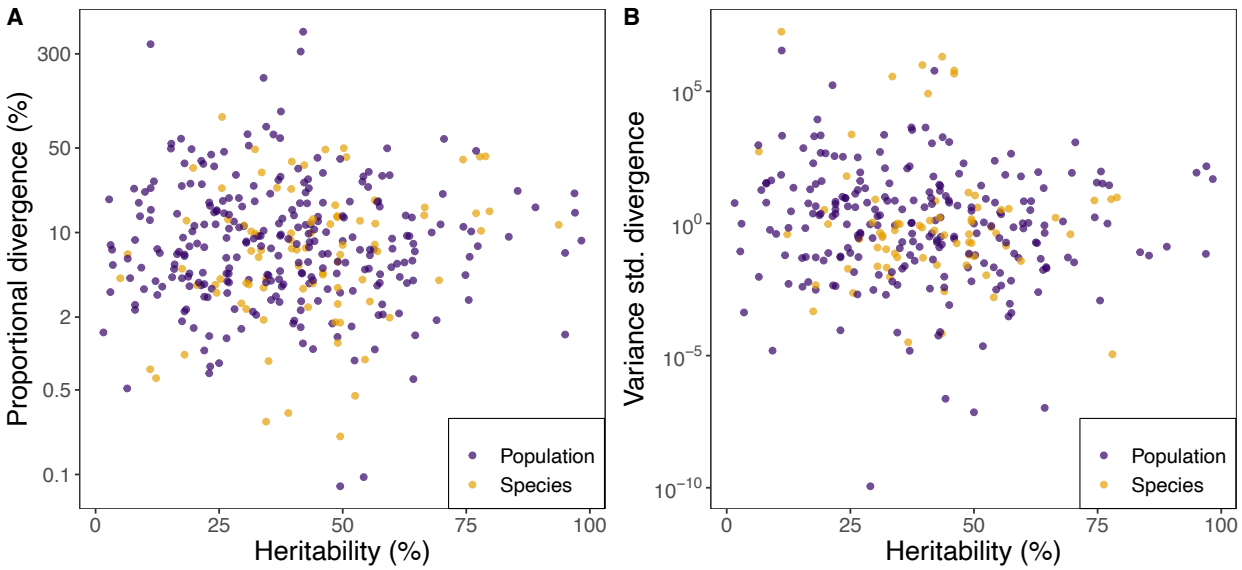

Fig. S5. No relationship between heritability and divergence.

Heritability is the proportion of additive genetic variance in the total phenotypic variance, expressed in percent. In (**A**) divergence (D) is computed as the variance in logarithmic trait means of the measured populations or species, expressed as expected proportional divergence in percent change of the trait grand mean ($\sqrt{2D/\pi}$). The least-square regression slope for the population data is b = -0.15±0.67 and R^2 = 0%, and for the species data b = 3.21±1.49 and R^2 = 5%. In (**B**) divergence is computed as the variance of the trait mean standardized by the phenotypic variance (var[$\bar{z}_i/V_{Pi}$]) of the measured populations or species, i.e., similar to the heritability standardization. The least-square regression slope for the population data is b = -2.20±1.63 and R^2 = 1%, and for the species data b = -3.85±4.35 and R^2 = 1%.

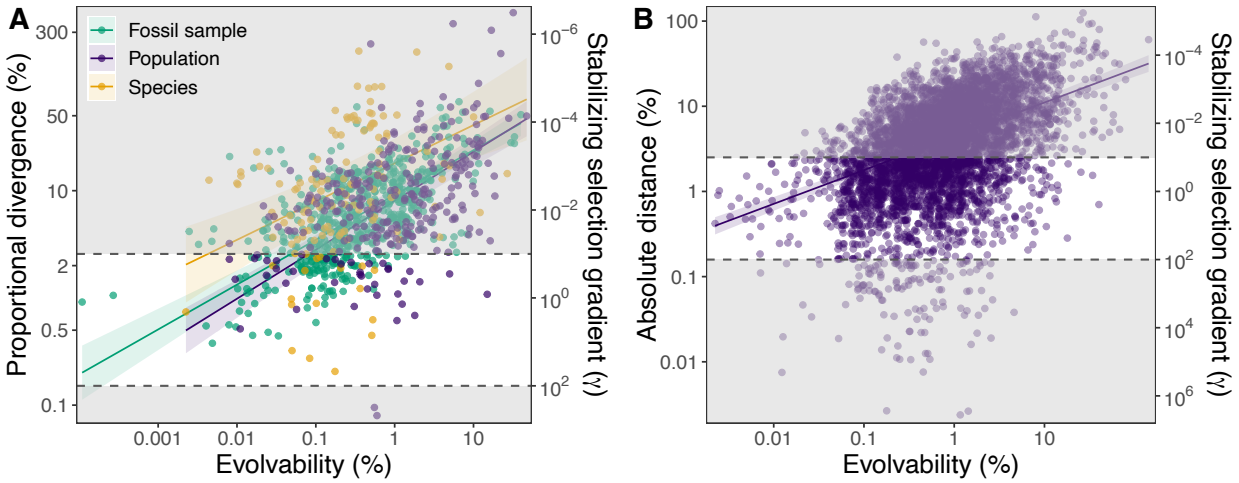

Fig. S6. Theoretical difficulties with the concordant-selection hypothesis. This hypothesis posits that patterns of stabilizing selection that shapes evolvability within populations are concordant with patterns of directional selection that cause population divergence. We now illustrate that this would require extreme variation in strengths of stabilizing selection to be effective. We first assume that the magnitude of divergence in a trait or direction **x** in phenotype space scales inversely with the strength of stabilizing selection in direction **x** to a power k, $d(\mathbf{x}) \propto \gamma(\mathbf{x})^{-k}$. Here **x** is taken to be eigenvectors of the second-order selection gradient matrix with corresponding eigenvalues $\gamma(\mathbf{x})$. The gaussian model of mutation-selection balance (*36*) combined with observed near isometric scaling between mutational variance and evolvability (*15*) yields $e_{\mu}(\mathbf{x}) \propto \gamma(\mathbf{x})^{-1}$ (see *64*, p. 85). Therefore, the magnitude of divergence should scale with evolvability as $d(\mathbf{x}) \propto e_{\mu}(\mathbf{x})^{k}$. This implies that k has the same range as the scaling exponents in Figures 2 and 3B, ranging from 0.36 to 0.46. Thus, explaining 2-3 orders of magnitude differences in divergence with differences in stabilizing selection would require the strength of stabilizing selection to vary with at least 4, and maybe as much as 9, orders of magnitude. This range goes from extremely strong (lower shaded area) to insufficient weak (upper shaded area) stabilizing selection. The lower and upper dashed lines mark selection strengths corresponding to a 10% trait shift from the optimum reducing fitness by 50% and 0.05%, respectively. (**A**) Data from Figure 2, and we use $\bar{b} = k = 0.41$. (**B**) Data from Figure 3B (without cutting the axis), $k = 0.40$.

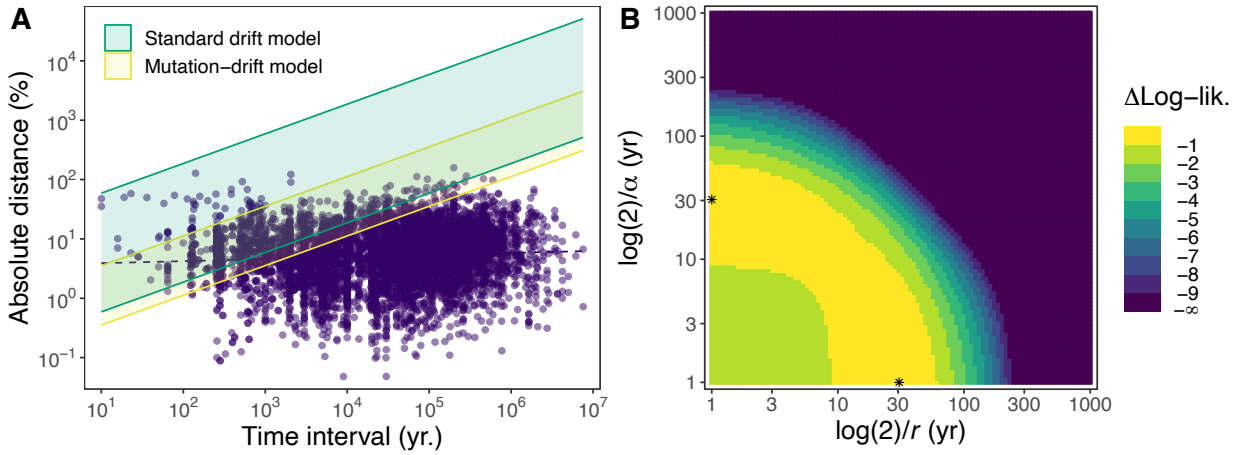

Fig. S7. Stationary fluctuations.

(**A**) The magnitude of divergence (d_c) corrected for sampling error variance (see methods) over the time interval between two samples. The magnitude of divergence is the absolute morphological distance to the next sample on log-scale and does not accumulate over time (dashed line: slope = 0.03±0.01%, R^2 = 0.5%, n = 6231). The green and yellow ribbons show the expectations from standard genetic-drift (green) and mutation-drift (yellow) models. The range of effective population sizes for the standard drift model is 10 – 10^5 individuals, taken from Estes and Arnold (*43*). The range of mutational variance for the neutral model is 10^{-2} % – 1% of the environmental variance (V_E) found by Lynch (*65*) in *Drosophila, Tribolium,* mice, and several crop species. We assume V_E is 65% of the median sample variance. (**B**) Support surface, in log-likelihood deviance (ΔLog-lik.) from the maximum, for the α and r parameters represented as half-lives (in years) obtained from a grid search performed on the derived Ornstein-Uhlenbeck-process function for rate of evolution (R) over time (t). The surface is symmetrical for values of α and r, and the two stars represent the best estimate combinations where the maximum log-likelihood resides.

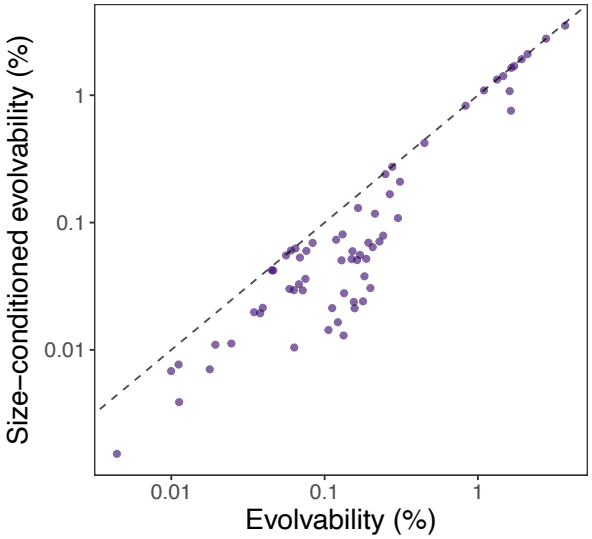

Fig. S8. Relationship between unconditional and conditional evolvability.

Conditional evolvabilities were obtained by conditioning the additive genetic variance of 66 traits on the size of the organisms (see methods). The dashed line represents the 1:1 line. The regression is $\ln(e_{\mu}|_{size}) = 0.01\ (\pm 0.36) + 1.11\ (\pm 0.05)\ \ln(e_{\mu})$, $R^2 = 87\%$. Mean unconditional and conditional evolvabilities per trait within study is shown in table S3.

Table S1. Hypotheses to explain the evolvability-divergence relationship and associated predictions.

Predictions for the causal and non-causal hypotheses and empirical results on the scaling relationship between the magnitude of divergence and evolvability ($d \propto e_\mu$) or heritability ($d \propto h^2$). Predictions also concern the scaling between divergence and stabilizing selection ($d \propto \gamma$) or time ($d \propto t$). Predictions supported by the empirical results are written in bold text.

	$d \propto e_\mu$ *among:*			$d \propto e_\mu$ *within:*		$d \propto h^2$	$d \propto \gamma$	$d \propto t$
	Populations	Species	Fossil samples	Common garden	Trait categories			
Non-causal hypotheses								
Heterogeneous data	**Positive**	Positive	**Positive**	**Positive**	No scaling	—	—	—
Gene flow	**Positive**	No scaling†	No scaling	**Positive**	**Positive**	**No scaling**	No scaling	—
Phenotypic plasticity	**Positive**	Positive	**Positive**	No scaling	**Positive**	—	No scaling	—
Concordant selection	**Positive**	Positive§	**Positive**	**Positive**	**Positive**	Positive	Negative	—
Causal hypotheses								
Neutral evolution	**Positive**	**Positive***	**Positive**	**Positive**	**Positive**	**No scaling**	No scaling	Positive
Directional selection under genetic constrains	**Positive**	**Positive***	**Positive**	**Positive**	**Positive**	**No scaling**	No scaling	Positive
Pleiotropic constraints	**Positive**	**Positive***	**Positive**	**Positive**	**Positive**	**No scaling**	No scaling	—
Fluctuating selection and genetic constraints	**Positive**	**Positive***	**Positive**	**Positive**	**Positive**	**No scaling**	No scaling	**No scaling**
Empirical results								
Contemporary data	**Positive**	**Positive***	—	**Positive**	**Positive**	**No scaling**	No evidence	—
Fossil data	—	—	**Positive**	—	**Positive**	—	No evidence	**No scaling**

† Assuming no gene flow among species

* A positive, but weaker scaling relationship than among populations and fossil samples predicted or observed.

§ A positive, but stronger scaling relationship than among populations and fossil samples predicted.

Table S2. Median and variance for evolvability and divergence on log-scale in the different data sets.

The tracking fluctuating selection hypothesis predicts that one quarter of the variance in evolvability should appear as variance in divergence (see methods). Here, we present these different variances for the different datasets.

Estimate	Population data	Species data	Fossil data (time-series level)
Median[e_μ]	0.01	0.002	0.0034
Var[ln(e_μ)]	3.08	2.41	2.59
Median[D]	0.0086	0.014	0.0045
Var[ln(D)]	4.92	6.38	3.16

e_μ = mean-scaled evolvability (not in percent)
D = var[ln(z)]
z = trait mean

Table S3. Evolvability and conditional evolvability.

Conditional evolvability is calculated for 66 traits from 25 G-matrices of animal species. Trait: focal trait; Conditioned on: genetically correlated trait representing the organism size that is used to calculate conditional evolvability; *n*: number of populations or species with available G matrices per trait; e_μ: mean evolvability per trait in percent; *c*: mean conditional evolvability in percent; *a*: mean autonomy; *Red*: percent reduction in evolvability when conditioned on size; *N* gen.: number of generations necessary for doubling the trait value with a mean standardized directional selection gradient (β) of 0.1 for both evolvability and evolvability conditioned on size. The last two rows show the overall median and mean for all columns.

Trait	Conditioned on	Shared taxa	*n*	e_μ	*c*	*a*	Red.	*N* gen. (e_μ)	*N* gen. (*c*)
wing length	tarsus length	*Cyanistes caeruleus*	4	0.017%	0.015%	0.89	10%	41 382	46 077
body mass	tarsus length	*Cyanistes caeruleus*	4	0.081%	0.067%	0.82	18%	8 584	10 418
tarsus length	wing length	*Cyanistes caeruleus*	4	0.042%	0.037%	0.89	11%	16 602	18 648
bill length	tarsus length	*Cyanistes caeruleus*	4	0.057%	0.054%	0.95	5%	12 160	12 793
subcaudal scales (F)	ventral scales (F)	*Thamnophis elegans*	2	0.140%	0.126%	0.89	10%	4 964	5 495
dorsal scale rows (F)	ventral scales (F)	*Thamnophis elegans*	2	0.033%	0.033%	0.99	0%	20 836	20 923
infralabial scales (F)	ventral scales (F)	*Thamnophis elegans*	2	0.054%	0.021%	-0.31	62%	12 798	33 748
supralabial scales (F)	ventral scales (F)	*Thamnophis elegans*	2	0.085%	0.082%	0.97	3%	8 132	8 420
postocular scales (F)	ventral scales (F)	*Thamnophis elegans*	2	0.269%	0.266%	0.99	1%	2 578	2 602
subcaudal scales (M)	ventral scales (M)	*Thamnophis elegans*	2	0.128%	0.122%	0.95	5%	5 411	5 689
dorsal scale rows (M)	ventral scales (M)	*Thamnophis elegans*	2	0.069%	0.067%	0.97	3%	10 040	10 372
infralabial scales (M)	ventral scales (M)	*Thamnophis elegans*	2	0.044%	0.037%	0.74	17%	15 616	18 734
supralabial scales (M)	ventral scales (M)	*Thamnophis elegans*	2	0.105%	0.079%	0.85	25%	6 625	8 810
eye span	femur length	*Diasemopsis dubia*	1	0.018%	0.007%	0.39	61%	38 897	98 545
eye stalk width	femur length	*Diasemopsis dubia*	1	0.063%	0.010%	0.16	84%	10 934	66 428
head length	femur length	*Diasemopsis dubia*	1	0.035%	0.020%	0.57	43%	19 993	35 097
femur length	tibia length	*Diasemopsis dubia*	1	0.004%	0.002%	0.35	65%	156 963	454 281
tibia length	femur length	*Diasemopsis dubia*	1	0.011%	0.004%	0.35	65%	61 630	178 368
tarsus length (1st tars)	femur length	*Diasemopsis dubia*	1	0.025%	0.011%	0.46	54%	28 129	61 773
wing length 1	femur length	*Diasemopsis dubia*	1	0.019%	0.011%	0.57	43%	35 729	63 157
wing length 2	femur length	*Diasemopsis dubia*	1	0.011%	0.008%	0.69	31%	62 121	90 355
wing length 3	femur length	*Diasemopsis dubia*	1	0.010%	0.007%	0.68	32%	69 488	101 735
femur length	prothorax width	Gryllidae	7	0.117%	0.030%	0.30	74%	5 946	22 812
head width	femur length	Gryllidae	7	0.130%	0.040%	0.32	69%	5 326	17 277
prothorax length	femur length	Gryllidae	7	0.161%	0.061%	0.41	62%	4 309	11 322
prothorax width	femur length	Gryllidae	7	0.142%	0.037%	0.30	74%	4 896	18 596
ovipositor length	femur length	Gryllidae	7	0.211%	0.118%	0.60	44%	3 283	5 868
body mass	tarsus length	*Cyanistes caeruleus*	3	0.325%	0.312%	0.96	4%	2 136	2 223
Cold recovery time (F)	wing size (F)	*Drosophila melanogaster*	3	1.690%	1.361%	0.80	19%	410	509
Desiccation rate (F)	wing size (F)	*Drosophila melanogaster*	3	1.600%	1.380%	0.86	14%	433	502
Heat knockdown time (F)	wing size (F)	*Drosophila melanogaster*	3	3.096%	3.032%	0.98	2%	224	229
Cold recovery time (M)	wing size (M)	*Drosophila melanogaster*	3	1.338%	1.323%	0.99	1%	518	524
Desiccation rate (M)	wing size (M)	*Drosophila melanogaster*	3	1.665%	1.646%	0.99	1%	416	421
Heat knockdown time (M)	wing size (M)	*Drosophila melanogaster*	3	1.883%	1.858%	0.99	1%	368	373
Overall median				**0.083%**	**0.047%**	**0.81**	**43%**	**8 358**	**15 031**
Overall mean				**0.402%**	**0.361%**	**0.69**	**10%**	**19 938**	**42 151**