Drift on holey landscapes as a dominant evolutionary process

Ned A. Dochtermann [a, b] Brady Klock [a,*], Derek A. Roff [c,d*], & Raphaël Royauté [e,f,*]

Author Affiliations:
[a] Department of Biological Sciences; North Dakota State University
[b] orcid ID: 0000-0002-8370-4614
[c] Department of Biology; University of California, Riverside
[d] orcid ID: 0000-0002-9337-7303
[e] French National Institute for Agriculture, Food, and Environment (INRAE), Versailles Cedex, France
[f] orcid ID: 0000-0002-5837-633X
* these authors are listed in alphabetical order

Corresponding Author: N.A. Dochtermann; ned.dochtermann@gmail.com;
Stevens Hall, 1340 Bolley Ave, Fargo, ND, USA

Preprint Servers: This manuscript is available at bioRXIV under a CC-BY-NC-ND 4.0 license (doi: 10.1101/2021.10.22.465488)

Classification: Biological Sciences – Evolution

Keywords: G matrix, fitness landscapes, evolutionary constraints

Author Contributions:

NAD conceived of the project and developed the first version of the model. BK collected published **G** matrices and calculated matrix summary estimates. RR contributed to model development and analyses as well as article screening. DAR contributed to model development and developed the parameter exploration scheme. All authors contributed to the writing of the manuscript.

Abstract

An organism's phenotype has been shaped by evolution but the specific processes have to be indirectly inferred for most species. For example, correlations among traits imply the historical action of correlated selection and, more generally, the expression and distribution of traits is expected to be reflective of the adaptive landscapes that have shaped a population. However, our expectations about how quantitative traits—like most behaviors, physiological processes, and life-history traits—should be distributed under different evolutionary processes is not clear. Here we show that genetic variation in quantitative traits is not distributed as would be expected under dominant evolutionary models. Instead, we found that genetic variation in quantitative traits across 6 phyla and 60 species (including both Plantae and Animalia) is consistent with evolution across high dimensional "holey landscapes". This suggests that the leading conceptualizations and modeling of the evolution of trait integration fail to capture how phenotypes are shaped and that traits are integrated in a manner contrary to predictions of dominant evolutionary theory. Our results demonstrate that our understanding of how evolution has shaped phenotypes remains incomplete and these results provide a starting point for reassessing the relevance of existing evolutionary models.

Significance Statement

We found that empirical estimations of how quantitative genetic variation is distributed do not correspond to typical Gaussian representations of fitness landscapes. These Gaussian landscapes underpin major areas of evolutionary biology and how selection is estimated in natural populations. Rather than being consistent with evolution on Gaussian landscapes, empirical estimates of genetic variation are, instead, consistent with evolution on high-dimensional "holey" landscapes. These landscapes represent situations where specific combinations of trait values are either viable or not and populations randomly drift among the viable combinations. This finding suggests that we have substantially misunderstood how selection actually shapes populations and thus how evolution typically proceeds.

Introduction

Our understanding of selection has been strongly shaped by Sewall Wright's conceptualization of an evolutionary landscape, with populations moving from areas of low fitness to areas of higher fitness (1, 2). While the one and two trait landscapes Wright originally described have been criticized as unrealistic, including by Wright himself (1), the general metaphor has nonetheless guided much of evolutionary thought (3).

For quantitative traits, like many aspects of physiology, behavior, and morphology, Wright's metaphor has been extended to complex topographies with ridges or tunnels of high fitness (4-6). Mathematical implementations of these landscapes have led to insights into how topography shapes the distribution of genetic (co)variance (7-9). Gaussian landscapes influenced by Wright's metaphor also underpin the statistical approaches that are used to estimate the action of stabilizing selection in natural populations (10).

How the topography of evolutionary landscapes affects sequence evolution at the genomic level has garnered similar interest (11). While research into quantitative traits has focused on relatively simple landscapes (e.g. 7, 8, 9, 12-14), research regarding sequence evolution spans simple single peak Gaussian "Fujiyama landscapes", to "badlands landscapes" (Fig 1A & 1B (15)), to abstract high-dimensional "holey landscapes" (Fig 1C (16)). An important conclusion from this research is that evolutionary dynamics on simple landscapes often fail to properly predict evolution on higher dimensional landscapes. Empirical research into quantitative traits has been slow to incorporate this need for a higher-dimensional perspective.

Perhaps most conceptually unfamiliar and unintuitive to researchers focused on quantitative traits are holey landscapes (Fig 1C (16)). Holey landscapes are high-dimensional evolutionary landscapes that consist of trait combinations that are either of average fitness or that are inviable (16, 17). This results in flat landscapes with holes at inviable or low fitness phenotypes (Fig 1C). This topography stems from the multivariate nature of phenotypes: while there may be continuous fitness differences in two dimensions, fitness gradients will create holes in the landscape and peaks will average out when additional traits are considered. Unfortunately, predictions about quantitative trait evolution on holey landscapes are not clear and have rarely been pursued, (e.g. 18).

Even more importantly, it is unknown what the topography of landscapes is for natural populations. While portions of selection surfaces and landscapes can be directly estimated (10, 19), these estimates may differ from the full landscape due to several factors. These include: the omission of fitness affecting traits (20), incomplete estimation of fitness (21, 22), and insufficient power to estimate non-linear selection coefficients (23).

An alternative to direct estimation of landscape topography is to infer landscape topography from observed trait (co)variances. For example, low additive genetic variation is suggestive of stabilizing or directional selection (24) and additive genetic correlations are expected to emerge from correlational selection (e.g. 4, 5). An ability to gain an understanding of the topography of landscapes based on observed variation would greatly further our understanding of how selection is realized in natural populations.

Here we used a simulation model to examine how evolution on different landscapes shapes genetic (co)variation. We modeled populations as evolving solely via drift, evolving via adaptation on Gaussian fitness landscapes stemming from Wright's metaphor, or as evolving on holey landscapes. While many additional implementations of Wright's metaphor exist (reviewed by 25), we focused on Gaussian landscapes with single peaks. We made this choice because this implementation of the metaphor connects to statistical approaches to estimating selection in natural populations (10, 26, 27) and to models of how genetic (co)variances evolve (7, 9, 28). These models allowed us to generate testable predictions for how the structure of additive genetic variances and covariances (**G**) are shaped by different landscape topographies. We then compared these modeled outcomes to 181 estimates of **G**, representing 60 species from 6 phyla, including both plants and animals. We show that the observed distribution of genetic variation is ultimately more consistent with evolution on holey landscapes than on Wrightian landscapes, indicating our models of phenotypic evolution should be re-evaluated.

Results

Model outcomes

When evolving on holey landscapes, simulated populations lost greater relative variation in the non-dominant dimensions as compared to when evolving on simple Gaussian

landscapes or when subject solely to drift (Fig 2; Fig S5). The ratio of genetic variation in the second versus first dimension (λ_2/λ_1) significantly differed depending on selection regime ($F_{4,1245} = 368$, $p << 0.01$; Fig 2). Populations experiencing either just drift or evolving on Gaussian landscapes maintained a more even amount of variation across dimensions compared to those evolving on holey landscapes (i.e. higher λ_2/λ_1 all post-hoc comparisons $p < 0.001$; Fig 2, Table S3). All populations evolving on holey landscapes exhibited similar λ_2/λ_1 ratios regardless of p (all post-hoc comparisons of outcomes for holey landscapes: $p > 0.05$; Fig 2, Table S3). This similarity is likely due to the observation elsewhere that, when p is greater than $½^k$, a cluster of viable phenotypic values—and therefore phenotypic space exists—through which a population can drift (18, 29). Given that k here was 10, this condition was satisfied.

While a modest difference, populations evolving due to drift alone also exhibited a significantly greater ratio than populations evolving on Gaussian landscapes (difference = 0.06, $p = 0.002$; Fig 2, Table S3). This magnitude of a difference is unlikely to be biologically important or detectable in natural populations and instead is likely only detectable here due to the high power afforded by simulations. These differences were consistent across multiple approaches to summarizing **G** and are robust to conditions of the simulations (Tables S4 – S7, Figs S6 – S8). Interestingly, examination of single population outcomes suggests that the outcomes observed for populations evolving on Gaussian landscapes stem from the populations becoming trapped at local optima (e.g. Fig S4).

These modeling results produce the general prediction that greater relative variation in multiple dimensions is maintained when populations evolve on Gaussian landscapes than when evolving on holey landscapes. Put another way, evolving on holey landscapes is predicted to result in a large decrease in variation from the dominant to subsequent dimensions and, consequently, a lower λ_2/λ_1 value (Fig 2).

Observed outcomes

Across all taxa, average λ_2/λ_1 was 0.36 (sd: 0.23, Fig 2). This estimate is consistent with and statistically indistinguishable from those observed for simulated populations evolving on holey landscapes ($t_{df:17.275} = 0.32, 1.20, -0.05$, $p > 0.2$ (all) versus holey landscapes with $p = 0.2, 0.5$, and 0.8 respectively; Fig 2, Table S10) and substantially less than observed for

141 simulated populations that evolved on Gaussian landscapes or via drift alone ($t_{df:17.275}$ = -
142 12.42, -14.55 respectively, $p < 0.001$ (both)).
143 While some individual estimates at the species level exhibited high λ_2/λ_1 values (Fig
144 2), phylogeny explained little variation in these values (phylogenetic heritability = 0.05;
145 Table S9). As was the case across all taxa, median λ_2/λ_1 values for each taxonomic Class (or
146 comparable level clade) were consistently lower than expected if evolution occurred on
147 Gaussian landscapes or via drift alone (Fig 2). Instead, these results are strongly consistent
148 with evolution on holey landscapes.

149 **Discussion**

150 Wright's conceptual model of adaptive landscapes has been key to our understanding of
151 how selection acts. However, our results suggest something substantively different is
152 occurring for many populations and species: the observed pattern of variation across taxa
153 suggests that classic models of the evolution of quantitative traits are not the predominant
154 force that has shaped the distribution of genetic variation. Instead, observed patterns of
155 quantitative genetic variation are more consistent with drift across holey landscapes (Fig
156 2).
157 Our implementation of Wright's metaphor represents only one of many possible
158 evolutionary models (25). It is possible that unmodeled alternative landscapes may
159 produce populations for which variation is distributed in a manner similar to holey
160 landscapes and empirical estimates. As such, despite the strong similarity between
161 empirical estimates of the structure of **G** and those from simulated populations that have
162 evolved on holey landscapes, questions remain as to whether alternative landscapes are
163 similarly likely. Importantly, and as mentioned previously, much of the exploration of
164 evolution of quantitative traits has focused on simple landscapes like we have implemented
165 here. Thus, it also is an open question what different models of selection "look" like when
166 implemented for higher-dimensional phenotypes. For example, rugged landscapes of high-
167 dimensionality may give rise to holey landscapes as peaks average out and valleys are
168 inviable. Alternatively, but not mutually exclusively, Pareto-optimality may contribute to
169 fitness planes (see below). Nonetheless, the close correspondence between empirical data

170 and populations simulated as evolving on holey landscapes suggest that our understanding
171 of quantitative trait evolution remains incomplete.
172 Much of the theoretical development of holey landscapes focused on the ability of
173 populations to traverse genomic sequence differences via drift, with some sequences being
174 inviable (e.g. due to missense differences in coding regions). How this extended to
175 quantitative traits was less clear. Our simulation model provides one approach to applying
176 the holey landscape framework to quantitative traits, treating each trait as a threshold
177 character (30). Other approaches to modeling quantitative traits on holey landscapes and
178 evolution in response to these versions, such as the generalized Russian roulette model
179 (17), may produce different outcomes. Moreover, the topography of our modeled holey
180 landscapes may also come about through other mechanisms. The flat surfaces of holey
181 landscapes might instead emerge through similar processes as recent discussions of the
182 Pareto optimization of traits (31, 32). Under Pareto optimization across just three traits a
183 flat surface—the Pareto front—connects single trait × environment optima (i.e.
184 "archetypes" (32)). Likewise, holes may come about not due to simple inviable trait
185 combinations but, instead, rugged landscapes can create steep fitness declines and
186 consequent holes in the overall landscape.
187 It is also important to recognize that the support for evolution on holey landscapes
188 that we found does not preclude subsets of traits from having evolved on Gaussian
189 landscapes. Indeed, stabilizing selection has been observed in natural populations (23),
190 though understanding its general strength even on a case-by-case basis is confounded with
191 methodological problems (33, 34). Regardless, our finding that observed patterns of
192 quantitative genetic variation across taxonomic groups are not consistent with traditional
193 evolutionary models stands.
194 This disconnect between observed patterns of multivariate variation and
195 expectations under conventional models of selection suggests that Wright's metaphor of
196 landscapes—and the subsequent implementation of this metaphor as Gaussian surfaces—
197 may have contributed to an incomplete understanding of how selection has shaped
198 phenotypes. A potential contributor to this problem has been the lack of clear alternative
199 explanations besides a simple null hypothesis of drift with no selection. Moving forward,

clear development of additional alternative models of the action of selection and evolution in multivariate space are needed. This will allow the comparison of simulated populations to empirical data as we have done here.

Ultimately, our findings suggest that evolutionary biologists need to better consider the effects of high dimensionality as simple standard evolutionary models are not consistent with available data for quantitative data. Our findings also emphasize a need to consider *how* different landscapes might emerge. We are cautiously optimistic that novel development in high-throughput phenotyping (35, 36) will provide data that is better suited to answer the crucial questions we have raised.

Materials & Methods

Model Construction

We developed an individual variance components model (Methods, Fig S1 (37)) wherein individuals had phenotypes comprised of 10 traits (k), with each trait being highly heritable ($h^2 = 0.8$), and initial genetic covariances between traits set at zero. Populations of individuals evolved on one of five landscapes: (i) a flat landscape where no selection occurred (i.e. drift alone), (ii) Gaussian landscapes where fitness for each pair of traits was characterized by a single peak but with correlational selection, and three (iii – v) implementations of holey landscapes differing by p (16, 17), the proportion of viable phenotypes in a holey landscape (p = 0.2, 0.5, and 0.8). Each of the modeling scenarios was simulated 250 times for populations of 7500 individuals and for 100 generations for each population. Full modeling details are provided in the Methods and all modeling code is available at https://github.com/DochtermannLab/Wright_vs_Holey.

Model analysis

Following these simulations, the eigen structures of the resulting 1250 population genetic covariance matrices were compared. Because the simulated phenotypes consisted of 10 traits, it was the overall multivariate pattern of variation that was of interest rather than any specific single trait or pairwise combination (38). To do so, we calculated the ratio of each matrix's second eigen value (λ_2) to its dominant eigen value (i.e. λ_2/λ_1). The choice of

using λ_2/λ_1 was based on preliminary observations of modeling results. These results demonstrated that variation in populations that have evolved on holey landscapes was compressed into a primary dimension to a greater degree than observed for populations evolving under different conditions. λ_2/λ_1 provides a better estimate of the compression of variance into a leading dimension than do other common metrics like the variation of the first eigen value to the sum of eigen values (i.e. $\lambda_1/\sum\lambda$). For example, $\lambda_1/\sum\lambda$ could be low if the variation not captured by λ_1 is equally distributed across all other dimensions, even if all other dimensions contained relatively little variation. The same scenario would produce a high value for λ_2/λ_1.

λ_2/λ_1 was then compared across the modeling scenarios using analysis of variance and Tukey post-hoc testing. Four alternative metrics for characterizing covariance matrices were consistent with the results for λ_2/λ_1 (see Supplementary Results). We also present the results of analyses of a broad range of starting conditions and model conditions in the Supplementary Results. These supplemental analyses covered much of the possible parameter space (Figures S12 – S14) and confirmed the robustness of the findings reported below.

Observed patterns of genetic covariation

We next wanted to determine which of the modeled processes produced results consistent with observed patterns of trait integration. To do so, we conducted a literature review wherein we used Web of Science to search the journals American Naturalist, Ecology and Evolution, Evolution, Evolutionary Applications, Evolutionary Ecology, Genetics, Heredity, Journal of Evolutionary Biology, Journal of Heredity, Nature Ecology and Evolution, and the Proceedings of the Royal Society (B). We searched these journals using the terms "G matrix" on 14 May 2019, yielding a total of 272 articles. Each article was reviewed and estimated **G** matrices extracted if the article met inclusion criteria. For inclusion, an estimated **G** matrix must have been estimated for more than 2 traits (i.e. > 2 × 2), must have been reported as variances and covariances (i.e. not genetic correlations), and must not have been estimated for humans. Based on these inclusion criteria, we ended up with a dataset of 181 estimated **G** matrices from 60 articles (Fig S2). For each published **G** matrix, we estimated λ_2/λ_1. Study ID was included as a random effect to accommodate the

estimation of different **G** matrices by group. This may not have fully captured the possibility of alternative fitness peaks by unknown grouping in underlying data.

As discussed above, the choice to use λ_2/λ_1 was an *a priori* decision based on examination of initial simulation runs. By focusing on a single predetermined metric we were able to conduct a confirmatory test of the concordance between published estimates of **G** and simulation outcomes. However, because $\lambda_1/\sum\lambda$ is more commonly reported, we also provide a graphical summary of simulation results and published estimates using this metric in the supplemental material (Fig S15). This summary, as with the primary results, demonstrate that published estimates of **G** more closely match populations evolving on holey landscapes than they do populations evolving on Gaussian landscapes.

Acknowledgements:

The authors thank B. de Bivort, A.J. Wilson, and J. Wright for helpful conversations. This work was supported by the US National Science Foundation grant 1557951 to N.A.D.

References Cited

1. Wright S (1988) Surfaces of selective value revisited. *The American Naturalist* 131(1):115-123.
2. Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution, pp. 356–366 in Proceedings of the Sixth International Congress of Genetics, edited by D. *Jones. Ithaca, NY*.
3. Olson ME, Arroyo-Santos A, & Vergara-Silva F (2019) A user's guide to metaphors in ecology and evolution. *Trends in ecology & evolution*.
4. Armbruster W & Schwaegerle K (1996) Causes of covariation of phenotypic traits among populations. *Journal of Evolutionary Biology* 9(3):261-276.
5. Armbruster WS (1990) Estimating and testing the shapes of adaptive surfaces: the morphology and pollination of Dalechampia blossoms. *The American Naturalist* 135(1):14-31.
6. Phillips PC & Arnold SJ (1989) Visualizing multivariate selection. *Evolution* 43(6):1209-1222.
7. Arnold SJ, Burger R, Hohenlohe PA, Ajie BC, & Jones AG (2008) Understanding the evolution and stability of the G-matrix. *Evolution* 62(10):2451-2461.
8. Jones AG, Arnold SJ, & Borger R (2003) Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution* 57(8):1747-1760.

9. Jones AG, Arnold SJ, & Burger R (2004) Evolution and stability of the G-matrix on a landscape with a moving optimum. *Evolution* 58(8):1639-1654.
10. Lande R & Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution* 37(6):1210-1226.
11. Kauffman S & Levin S (1987) Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology* 128(1):11-45.
12. Bulmer M (1972) The genetic variability of polygenic characters under optimizing selection, mutation and drift. *Genetics Research* 19(1):17-25.
13. Reeve JP (2000) Predicting long-term response to selection. *Genetics Research* 75(1):83-94.
14. Turelli M (1988) Phenotypic evolution, constant covariances, and the maintenance of additive variance. *Evolution* 42(6):1342-1347.
15. Kauffman SA & Weinberger ED (1989) The NK model of rugged fitness landscapes and its application to maturation of the immune response. *Journal of theoretical biology* 141(2):211-245.
16. Gavrilets S (1997) Evolution and speciation on holey adaptive landscapes. *Trends in Ecology & Evolution* 12(8):307-312.
17. Gavrilets S (2004) *Fitness Landscapes and the Origin of Species* (Princeton University Press, Princeton) p 432.
18. Gravner J, Pitman D, & Gavrilets S (2007) Percolation on fitness landscapes: effects of correlation, phenotype, and incompatibilities. *Journal of theoretical biology* 248(4):627-645.
19. Morrissey MB (2014) In search of the best methods for multivariate selection analysis. *Methods Ecol. Evol.* 5(10):1095-1109.
20. Delcourt M, Blows MW, Aguirre JD, & Rundle HD (2012) Evolutionary optimum for male sexual traits characterized using the multivariate Robertson–Price Identity. *Proceedings of the National Academy of Sciences* 109(26):10414-10419.
21. Arnold SJ & Wade MJ (1984) On the measurement of natural and sexual selection: theory. *Evolution* 38(4):709-719.
22. Shaw RG & Geyer CJ (2010) Inferring fitness landscapes. *Evolution* 64(9):2510-2520.
23. Kingsolver JG, *et al.* (2001) The strength of phenotypic selection in natural populations. *The American Naturalist* 157(3):245-261.
24. Mousseau TA & Roff DA (1987) Natural selection and the heritability of fitness components. *Heredity* 59:181-197.
25. Fragata I, Blanckaert A, Louro MAD, Liberles DA, & Bank C (2019) Evolution in the light of fitness landscape theory. *Trends in ecology & evolution* 34(1):69-82.
26. Lande R (1979) Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* 33(1):402-416.
27. Lande R (1980) Genetic variation and phenotypic evolution during allopatric speciation. *The American Naturalist* 116(4):463-479.
28. Arnold SJ, Pfrender ME, & Jones AG (2001) The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* 112:9-32.
29. Gavrilets S & Gravner J (1997) Percolation on the fitness hypercube and the evolution of reproductive isolation. *Journal of theoretical biology* 184(1):51-64.

30. Lynch M & Walsh B (1998) *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, Sunderland, MA).
31. Petrie KL*, et al.* (2018) Destabilizing mutations encode nongenetic variation that drives evolutionary innovation. *Science* 359(6383):1542-1545.
32. Shoval O*, et al.* (2012) Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science* 336(6085):1157-1160.
33. Morrissey MB (2016) Meta-analysis of magnitudes, differences and variation in evolutionary parameters. *Journal of Evolutionary Biology* 29(10):1882-1904.
34. Stinchcombe JR, Agrawal AF, Hohenlohe PA, Arnold SJ, & Blows MW (2008) Estimating nonlinear selection gradients using quadratic regression coefficients: double or nothing? *Evolution: International Journal of Organic Evolution* 62(9):2435-2440.
35. Houle D, Govindaraju DR, & Omholt S (2010) Phenomics: the next challenge. *Nature reviews genetics* 11(12):855-866.
36. Lürig MD, Donoughe S, Svensson EI, Porto A, & Tsuboi M (2021) Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology. *Frontiers in Ecology and Evolution* 9:642774.
37. Roff D (2010) *Modeling evolution: an introduction to numerical methods* (Oxford University Press, Oxford) p 352.
38. Kirkpatrick M (2009) Patterns of quantitative genetic variation in multiple dimensions. *Genetica* 136(2):271-284.

Figure Legends

Figure 1. Example fitness landscapes. Redder colors correspond to higher fitness. A. A simple Gaussian, single peak Fujiyama landscape with a single optimum. B. A more rugged landscape with multiple local optima and a single global optimum. C. A simplified Holey landscape where combinations of values correspond to high, average, fitness (1) or low (0) fitness. This landscape has a large cluster of viable phenotypes with a hole and irregular borders and a smaller cluster of viable phenotypes at low trait values.

Figure 2. Modified "Orchard plot" of λ_2/λ_1 values for simulated (above solid line) and observed **G** matrices. *Trunks* (large points) are the medians for the specified group (e.g. Gaussian landscapes or Insecta), *branches* (thick lines) are interquartile ranges, *twigs* (thin lines) give the full range of values, and *fruits* (smaller points) are individual estimates within a simulation or taxonomic group. Rightmost letters correspond to statistical significance—or lack thereof—of comparisons of ratios among simulations. Datasets sharing letters did not significantly differ (Table S3). Populations evolving due to drift alone had a significantly higher ratio than observed for either stabilizing selection or evolution on any of the holey landscapes. Populations evolving on holey landscapes also had lower ratios than those experiencing stabilizing selection but did not differ from each other. Rightmost numbers are the number of estimates available via literature search. (organism silhouettes courtesy of phylopic.org, Public Domain Mark 1 licenses or CCA 3.0; Chlorophyceae: S.A. Muñoz-Gómez, Superrosid: D.J. Bruzzese, Superasterid: T.M. Keesey & Nadiatalent).

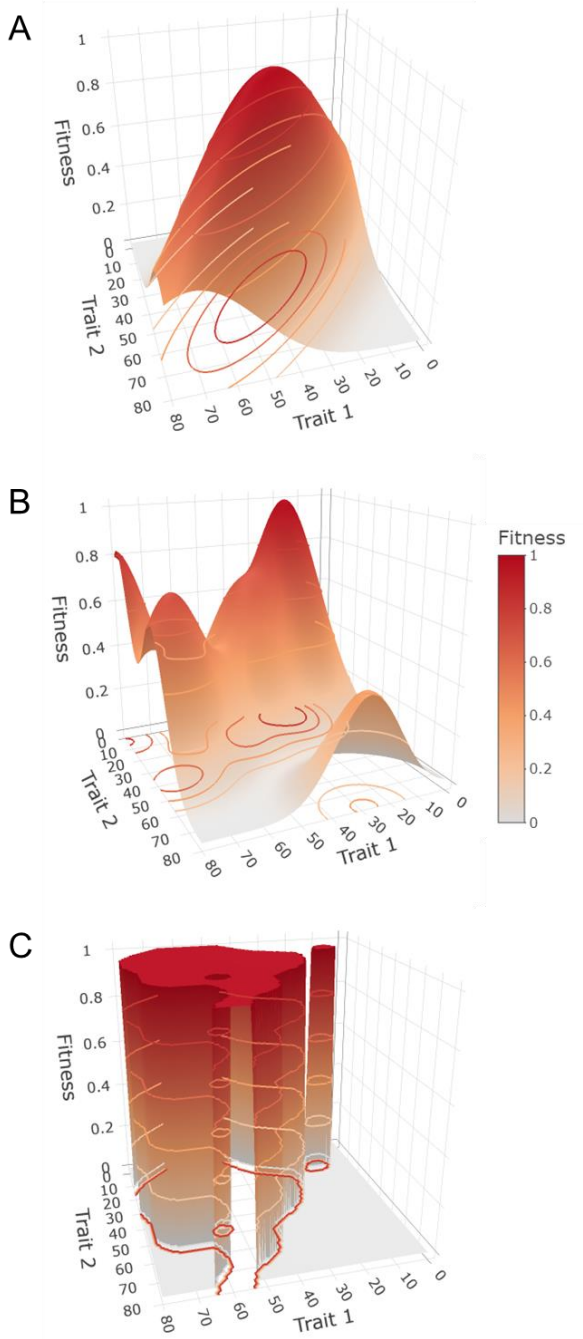


Figure 1.

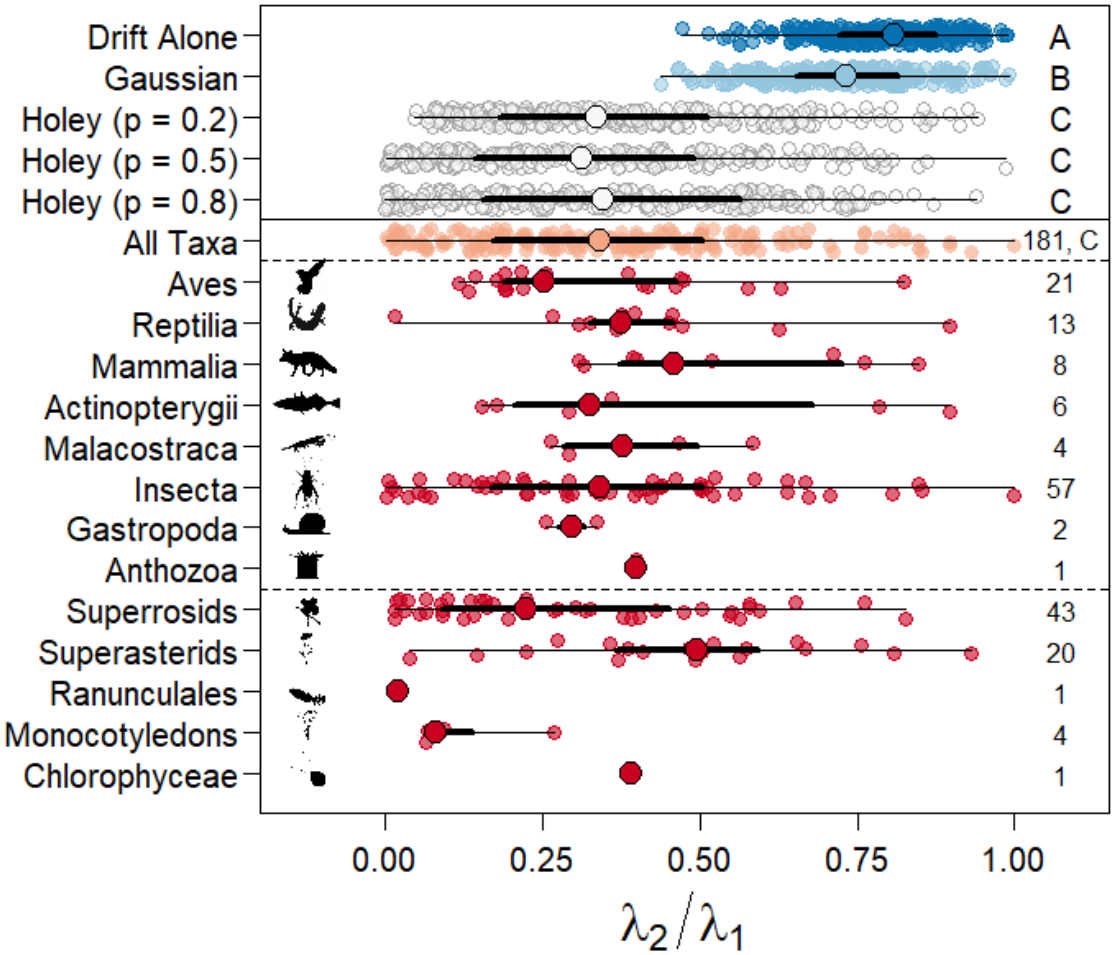

Drift Alone
Gaussian
Holey (p = 0.2)
Holey (p = 0.5)
Holey (p = 0.8)
All Taxa
Aves
Reptilia
Mammalia
Actinopterygii
Malacostraca
Insecta
Gastropoda
Anthozoa
Superrosids
Superasterids
Ranunculales
Monocotyledons
Chlorophyceae
A
B
C
C
C
181, C
21
13
8
6
4
57
2
1
43
20
1
4
1
0.00
0.25
0.50
0.75
1.00
λ2/λ1

Figure 2.

Supplementary Materials for: **Drift on holey landscapes as a dominant evolutionary process**

Ned A. Dochtermann, Brady Klock, Derek A. Roff, & Raphaël Royauté

Correspondence to: ned.dochtermann@gmail.com

Supplementary materials include:

- Detailed Methods
- Supplemental Results
- Figs. S1 to S15
- Tables S1 to S10

Detailed Methods

Simulation Models

Model Construction

We developed an individual variance components model (Fig S1; sensu 1) wherein individuals had phenotypes comprised of 10 traits (k) and with each trait being highly heritable ($h^2 = 0.8$) and initial genetic covariances between traits of 0. A high heritability was initially used to reduce the number of generations needed to determine the response of populations to selection. Genetic covariances were set to an initial value of zero to simulate a population under linkage equilibrium. Viability selection was applied based on fitness, which was determined either by location on a ten-dimensional holey landscape or on simple Gaussian landscapes with a single optimum per trait pair.

Holey Landscapes

For simulations evaluating holey landscapes, we simulated populations in which traits were inherited as though continuous but expressed categorically as one of two phenotypic variants (e.g. phenotype 0 versus 1 for trait 1). Specifically, at the start of simulations, we drew genotypes for each individual from a normal distribution with a mean of zero and standard deviation of 1. To these normally distributed genotypes, we added "environmental" values ($\mu = 0$, all covariances = 0) to generate a phenotype with a heritability of 0.8. These continuously distributed phenotypic values were then transformed as one implementation of the holey landscape is based on the fitness of specific and discrete *combinations*. Specifically, the continuously distributed values were transformed to be a phenotype of 0 or 1, with a genotype < 0 being "0" and a genotype > 0 being "1" (Table S1).

The holey landscape for a specific simulation was then constructed by randomly assigning a fitness of 0 or 1 to the 1024 possible phenotypes (2^k) trait combinations based on the parameter p. "p" was the probability that a trait combination had a fitness of 1 and corresponds to Gavrilets' (2004) percolation parameter. We used three values of p in our simulation ranging from weak ($p = 0.2$), moderate ($p = 0.5$) and high ($p = 0.8$). p can vary between 0 and 1, with values of 1 corresponding to a landscape where all trait combinations are viable and have a fitness of 1. As p approaches 0, few trait combinations are viable.

After the first generation, genotypes were drawn from a multivariate normal distribution based on the means and genetic variance-covariance matrix of the population that survived selection. Environmental contributions again had an average of 0 and no environmental correlation with a variance set to keep heritability at 0.8 (or other values during parameter exploration, below). The resulting phenotypic values were then

converted to 0's and 1's as above. This approach to generating subsequent generations follows the structure of individual variance components models described by Roff (1). We used this individual variance components approach rather than an agent-based approach as the latter combined with the computational requirements of matching phenotypes to fitness under the holey landscape model was not amenable to simulation analysis.

Table S1. Example conversion of an underlying genotype to a phenotype under the two modelling scenarios. The same individual has a genotypic value for each of the 10 traits simulated (e.g. -0.918 for trait 10). To this, "environmental" contributions are added, taking heritability to 0.8. For Holey Landscape simulations, these phenotypic values are then converted to either 0 or 1 based on whether the phenotype is negative or positive.

	Trait									
	1	2	3	4	5	6	7	8	9	10
Genotype	0.008	0.770	0.477	0.112	-0.512	0.751	-1.752	-0.944	0.030	-0.918
Environmental Contribution	0.402	-0.221	0.023	0.053	0.082	-0.25	0.63	0.285	-0.007	0.271
Holey Landscape Phenotype	1	1	1	1	0	1	0	0	1	0
Gaussian Landscape Phenotype	0.410	0.549	0.500	0.165	-0.430	0.501	-1.122	-0.659	0.023	-0.647

Gaussian (*Wrightian*) *adaptive landscapes*

For simulations evaluating Gaussian landscapes, we generated genotypes and phenotypes as above but without the categorical conversion (Table S1). We then generated random landscapes such that the optima (θ) for all traits was set to zero. The topography of the landscape for each pair of traits (e.g. $\omega_{i,j}$) was defined as $\begin{bmatrix} 1 & \omega_{i,j} \\ \omega_{i,j} & 1 \end{bmatrix}$ consistent with previous simulation studies examining the evolution of quantitative traits (reviewed by 2). This approach corresponds to single peak landscapes in any two dimensions. The forty-five $\omega_{i,j}$ values that fully describe the landscape were generated using the LKJ onion method for constructing random correlation matrices with a pseudo-normal distribution of correlations where the average correlation is 0 ($\eta = 1$; Lewandowski et al. 2009; Fig S2) . Using the LKJ onion method ensures that the full description of the landscape (ω) is positive semi-definite with feasible partial correlations. We then calculated each individual's fitness based on a Gaussian surface (4):

$$w_h = \exp(-.5(z_h-\theta)^T\omega^{-1}(z_h-\theta))$$

where w_h is the fitness of individual h, z_h is a vector of the observed phenotypic values for individual h, ω is the selection surface, and θ is the optima for traits (0). Truncation selection was applied based on fitness, with the 50% of individuals possessing the highest fitness surviving (main results). In an additional set of simulations, stronger truncation selection was applied and only 10% of the population survived.

Following selection in either framework, the next generation was constructed using an individual variance components approach (1). Specifically, the next generation was generated as described above based on the trait means, variances and covariances of survivors. Selection therefore acted via changes in means and variances and drift during the selection simulations was due to sampling error from the selection shaped phenotypic distributions.

Drift alone

For populations evolving via drift alone phenotypes were generated as for Gaussian adaptive landscapes. Composition of subsequent generations was likewise generated based on the means and variances of the prior generation, without selection. The drift model therefore was simply a model of sampling error.

Each of five modeling scenarios (simple landscapes, drift alone, three Holey landscapes with p = 0.2, 0.5, or 0.8) was simulated 250 times for populations of 7500 individuals and for 100 generations for each population. All modeling code is available at https://github.com/DochtermannLab/Wright_vs_Holey.

Statistical Comparison of Evolutionary Metrics

To clarify differences in evolutionary outcomes across modeling scenarios, we summarized evolutionary outcomes at the level of **G** matrices based on several metrics:

1. λ_2/λ_1; results for this metric are presented in the main text
2. $\lambda_1/\sum\lambda$; this is a commonly used summary value and represents the proportion of variation captured by dominant eigenvalue. This can be interpreted as the proportion variation in the main dimension of covariance
3. $\sum\lambda$; matrix trace, the total variation present. For simulations this is informative as to whether a particular process results in the loss of more or less variation
4. ē: average evolvability across dimensions (5). Evolutionary potential throughout multivariate space
5. ā: average reduction in evolvability due to trait covariance (5). Can be interpreted as how constrained evolutionary responses are based on correlations. At the extreme, an average autonomy of 0 would indicate absolute constraints on responses to

selection and an average autonomy of 1 indicates evolutionary independence. Values between 0 and 1 represent quantitative constraints.

We compared these metrics across drift, Gaussian, and holey landscape simulations, following the main text, based on ANOVA followed by post-hoc comparisons based on calculation of Tukey's Honest Significant Differences (HSD).

Post-hoc Parameter Exploration

The above modeling scenarios were used for our overall general analyses and for comparison to observed values. However, to explore whether our modeling outcomes were due to fundamentally different and generalizable outcomes or instead emerged from peculiarities of initial parameters, we expanded our analyses in two ways.

First, in addition to the moderate/weak strength of truncation selection modeled above (0.5), we also modeled stronger selection where only 10% of individuals survived. For this stronger strength of selection we again conducted 250 simulations of 7500 individuals for 100 generations. These simulations were included in the above analyses.

Second, to more broadly examine the sensitivity of our results to different starting values, we conducted simulation studies for our selection model, our model of drift, and our model of evolution on flat holey landscapes. For each modeling scenario (Gaussian surfaces, drift, Holey landscapes) we conducted 1000 simulations where both the magnitude of initial genetic variation in each trait varied and h^2 varied (h^2 was defined independently). For each scenario we then explored how other changes in starting parameters affected the eigenstructure of **G** (Table S2).

We then quantitatively assessed the relevance of each varied parameter on λ_2/λ_1—within modeling scenario—using linear models. All two-way interactions were included in analyses and variables (model parameters) were mean centered but unscaled. We then qualitatively compared λ_2/λ_1 across modeling scenarios based on heat plots.

Table S2. Parameters varied across simulation iterations by modeling scenario and range of possible values

Modeling Scenario	Parameter varied	Parameter values
Gaussian surfaces	Genetic variation present in traits	Single trait variabilities were independently drawn from uniform distributions ranging from 0.1 to 1.9.
	Correlations among traits	Initial genetic correlations were drawn according to the LKJ onion method (3) with η = 1.
	h^2	Heritabilities were drawn from a uniform distribution ranging from 0.01 to 0.99
	Selection strength	Proportion of individuals surviving to reproduce was drawn from a uniform distribution ranging from 0.1 to 0.9.
Drift	Genetic variation present in traits	Single trait variabilities were independently drawn from uniform distributions ranging from 0.1 to 1.9.
	Correlations among traits	Initial genetic correlations were drawn according to the LKJ onion method (3) with η = 1.
	h^2	Heritabilities were drawn from a uniform distribution ranging from 0.01 to 0.99
Holey landscapes	Genetic variation present in traits	Single trait variabilities were independently drawn from uniform distributions ranging from 0.1 to 1.9.
	Correlations among traits	Initial genetic correlations were drawn according to the LKJ onion method (3) with η = 1.
	h^2	Heritabilities were drawn from a uniform distribution ranging from 0.01 to 0.99
	p	Proportion of inviable phenotypes, Gavrilets' percolation parameter

Empirically Estimated **G** Matrices

Observed patterns of multivariate genetic variation

We conducted a literature review with Web of Science to search the journals American Naturalist, Ecology and Evolution, Evolution, Evolutionary Applications, Evolutionary Ecology, Genetics, Heredity, Journal of Evolutionary Biology, Journal of Heredity, Nature Ecology and Evolution, and the Proceedings of the Royal Society (B). These journals were searched using the terms "G matrix" on 14 May 2019, yielding a total of 272 articles. Each article was reviewed to determine if the article met inclusion criteria. Our inclusion criteria were:

1. A **G** matrix must have been estimated for more than 2 traits (i.e. > 2 × 2)
2. Must have been reported as variances and covariances (i.e. not genetic correlations)
3. Must not have been estimated for humans.

Based on these inclusion criteria, we ended up with 181 estimated **G** matrices (Fig S3). For each published **G** matrix, we calculated λ_2/λ_1 using a purpose-built R Shiny App (link).

For each estimate we recorded the paper from which it was drawn (recorded as a unique study ID), taxonomic information (Kingdom through species epithet), trait category (life-history,

physiology, morphology, behavior or mixed), the number of traits in the matrix, λ_1, λ_2, λ_2/λ_1, number of dimensions (6), number of dimensions divided by the number of traits, and all bibliographic information.

Phylogenetic Signal in λ_2/λ_1

To test for phylogenetic signal we fit a simple taxonomic mixed-effects model. This modeling approach incorporates the hierarchical non-independence due to taxonomic relationships but does not require a full phylogeny (7). Essentially, at each node of a phylogeny, relationships are modeled according to a star relationship. Each taxonomic grouping was included as a random effect, as was study ID, and the resulting model fit with the `lme4` package in R (8). From this model we estimated phylogenetic signal as the proportion of variation attributable to taxonomy, the variation attributable to study ID, and the residual variance. Confidence intervals were then estimated based on likelihood profile likelihoods.

Comparison of Observed Results to Simulation Results

Finally, we compared the observed values to the average for each of the simulation using the intercept coefficient of the above linear model. For this, t was calculated as (9):

$$t = \frac{\hat{\beta} - \beta_{H_0}}{s.e.(\hat{\beta})}$$

where $\hat{\beta}$ was the estimated intercept from the taxonomic model (above) and β_{H_0} was a simulation average. p was calculated with degrees of freedom estimated using Satterthwaite's method (df = 17.275).

Supplemental Results

<u>Simulation Models</u>

Statistical Comparison of Evolutionary Metrics

Populations that evolved on different landscapes (drift alone, Gaussian, or holey) significantly differed from each other in the structure of **G** after 100 generations (Tables S3 – S7). Holey landscapes were characterized by a compression of most variation into the dominant dimension in multivariate space (Tables S3 & S4; Figures 2 & S5). Populations evolving on Gaussian landscapes were characterized by a drastic reduction in the total variation present, which was also reflected in reduced evolvability (Tables S6 & S7; Figures S5 & S6). The combination of high standing genetic variation and this variation being distributed across dimensions led to populations that evolved solely due to drift to exhibit significantly greater autonomy than observed in any of the other modeling scenarios (Table S7; Figure S8). This greater constraint in populations evolving on either Gaussian or holey landscapes is likely due to the loss of variation for populations evolving on Gaussian

landscapes (Figures S6 & S7) and the compression of variation for populations evolving on holey landscapes (Figures 2 & S5).

Table S3. ANOVA and Tukey HSD results for λ_2/λ_1. Significantly greater genetic variation was maintained across all dimensions when populations evolved on Gaussian landscapes or due to drift than when evolving on holey landscapes (Figure 2, main text).

ANOVA Results					
	df	SS	MSS	F	p
Simulation type	5	54.98	10.996	343.5	<0.01
Residual	1494	47.82	0.032		
Tukey HSD					
Simulation Comparison		Difference	Lower	Upper	p
Holey p = 0.5-Holey p = 0.2		-0.026	-0.071	0.020	0.589
Holey p = 0.8-Holey p = 0.2		0.011	-0.035	0.057	0.984
Wright 0.1-Holey p = 0.2		0.293	0.248	0.339	<0.01
Wright 0.5-Holey p = 0.2		0.374	0.329	0.420	<0.01
Drift-Holey p = 0.2		0.437	0.391	0.483	<0.01
Holey p = 0.8-Holey p = 0.5		0.037	-0.009	0.082	0.198
Wright 0.1-Holey p = 0.5		0.319	0.273	0.365	<0.01
Wright 0.5-Holey p = 0.5		0.400	0.354	0.446	<0.01
Drift-Holey p = 0.5		0.463	0.417	0.508	<0.01
Wright 0.1-Holey p = 0.8		0.282	0.237	0.328	<0.01
Wright 0.5-Holey p = 0.8		0.363	0.318	0.409	<0.01
Drift-Holey p = 0.8		0.426	0.380	0.472	<0.01
Wright 0.5-Wright 0.1		0.081	0.035	0.127	<0.01
Drift-Wright 0.1		0.144	0.098	0.189	<0.01
Drift-Wright 0.5		0.063	0.017	0.108	<0.01

Table S4. ANOVA and Tukey HSD results for $\lambda_1/\sum\lambda$. Significantly greater proportional genetic variation was retained in the dominant multivariate direction for populations that evolved on Gaussian landscapes or via drift than when evolving on holey landscapes (Figure S5).

ANOVA Results					
	df	SS	MSS	F	p
Simulation type	5	29.49	5.90	325.4	<0.01
Residual	1494	27.08	0.02		
Tukey HSD					
Simulation Comparison		Difference	Lower	Upper	p
Holey p = 0.5-Holey p = 0.2		0.044	0.010	0.079	<0.01
Holey p = 0.8-Holey p = 0.2		0.019	-0.015	0.054	0.594
Wright 0.1-Holey p = 0.2		-0.188	-0.223	-0.154	<0.01
Wright 0.5-Holey p = 0.2		-0.233	-0.268	-0.199	<0.01
Drift-Holey p = 0.2		-0.320	-0.354	-0.285	<0.01
Holey p = 0.8-Holey p = 0.5		-0.025	-0.059	0.009	0.307
Wright 0.1-Holey p = 0.5		-0.232	-0.267	-0.198	<0.01
Wright 0.5-Holey p = 0.5		-0.278	-0.312	-0.243	<0.01
Drift-Holey p = 0.5		-0.364	-0.398	-0.330	<0.01
Wright 0.1-Holey p = 0.8		-0.208	-0.242	-0.173	<0.01
Wright 0.5-Holey p = 0.8		-0.253	-0.287	-0.218	<0.01
Drift-Holey p = 0.8		-0.339	-0.374	-0.305	<0.01
Wright 0.5-Wright 0.1		-0.045	-0.080	-0.011	<0.01
Drift-Wright 0.1		-0.132	-0.166	-0.097	<0.01
Drift-Wright 0.5		-0.086	-0.121	-0.052	<0.01

Table S5. ANOVA and Tukey HSD results for the total genetic variation in populations at the end of simulations $\sum\lambda$. The amount of total variation significantly varied across simulation types. Populations that evolved on Gaussian landscapes lost considerably more genetic variation than those evolving on other landscapes (Figure S6).

ANOVA Results					
	df	SS	MSS	F	p
Simulation type	5	357826	71565	6.23	<0.01
Residual	1494	17167085	11491		
Tukey HSD					
Simulation Comparison		Difference	Lower	Upper	p
Holey p = 0.5-Holey p = 0.2		-1.050	-28.408	26.308	1.000
Holey p = 0.8-Holey p = 0.2		-18.153	-45.511	9.205	0.407
Wright 0.1-Holey p = 0.2		-37.651	-65.009	-10.293	<0.01
Wright 0.5-Holey p = 0.2		-37.237	-64.595	-9.879	<0.01
Drift-Holey p = 0.2		-27.791	-55.149	-0.433	0.044
Holey p = 0.8-Holey p = 0.5		-17.103	-44.461	10.255	0.477
Wright 0.1-Holey p = 0.5		-36.601	-63.959	-9.243	<0.01
Wright 0.5-Holey p = 0.5		-36.187	-63.545	-8.830	<0.01
Drift-Holey p = 0.5		-26.741	-54.099	0.617	0.060
Wright 0.1-Holey p = 0.8		-19.498	-46.856	7.860	0.324
Wright 0.5-Holey p = 0.8		-19.084	-46.442	8.274	0.348
Drift-Holey p = 0.8		-9.638	-36.996	17.720	0.916
Wright 0.5-Wright 0.1		0.414	-26.944	27.771	1.000
Drift-Wright 0.1		9.860	-17.498	37.218	0.908
Drift-Wright 0.5		9.446	-17.912	36.804	0.923

Table S6. ANOVA and Tukey HSD results for evolvability, ē. Because more genetic variation was maintained when populations evolved on holey landscapes or drift (Table S5), evolvability was significantly lower when populations evolved on Gaussian landscapes (Figure S7). (evolvability is just the matrix trace divided by the number of traits)

ANOVA Results					
	df	SS	MSS	F	p
Simulation type	5	3578	715.7	6.23	<0.01
Residual	1494	171671	114.9		
Tukey HSD					
Simulation Comparison		Difference	Lower	Upper	p
Holey p = 0.5-Holey p = 0.2		-0.105	-2.841	2.631	1.000
Holey p = 0.8-Holey p = 0.2		-1.815	-4.551	0.921	0.407
Wright 0.1-Holey p = 0.2		-3.765	-6.501	-1.029	<0.01
Wright 0.5-Holey p = 0.2		-3.724	-6.460	-0.988	<0.01
Drift-Holey p = 0.2		-2.779	-5.515	-0.043	0.044
Holey p = 0.8-Holey p = 0.5		-1.710	-4.446	1.025	0.477
Wright 0.1-Holey p = 0.5		-3.660	-6.396	-0.924	<0.01
Wright 0.5-Holey p = 0.5		-3.619	-6.355	-0.883	<0.01
Drift-Holey p = 0.5		-2.674	-5.410	0.062	0.060
Wright 0.1-Holey p = 0.8		-1.950	-4.686	0.786	0.324
Wright 0.5-Holey p = 0.8		-1.908	-4.644	0.827	0.348
Drift-Holey p = 0.8		-0.964	-3.700	1.772	0.916
Wright 0.5-Wright 0.1		0.041	-2.694	2.777	1.000
Drift-Wright 0.1		0.986	-1.750	3.722	0.908
Drift-Wright 0.5		0.945	-1.791	3.680	0.923

Table S7. ANOVA and Tukey HSD results for autonomy, ā. Significantly greater variation was maintained across all dimensions when populations evolved on Gaussian landscapes or due to drift than when evolving on holey landscapes (Figure S8).

ANOVA Results					
	df	SS	MSS	F	p
Simulation type	5	43.61	8.72	518.3	<0.01
Residual	1494	25.14	0.02		
Tukey HSD					
Simulation Comparison		Difference	Lower	Upper	p
Holey p = 0.5-Holey p = 0.2		-0.021	-0.054	0.012	0.479
Holey p = 0.8-Holey p = 0.2		-0.042	-0.075	-0.009	<0.01
Wright 0.1-Holey p = 0.2		-0.148	-0.181	-0.115	<0.01
Wright 0.5-Holey p = 0.2		0.395	0.362	0.428	<0.01
Drift-Holey p = 0.2		0.077	0.044	0.110	<0.01
Holey p = 0.8-Holey p = 0.5		-0.022	-0.055	0.011	0.424
Wright 0.1-Holey p = 0.5		-0.127	-0.160	-0.094	<0.01
Wright 0.5-Holey p = 0.5		0.415	0.382	0.448	<0.01
Drift-Holey p = 0.5		0.098	0.064	0.131	<0.01
Wright 0.1-Holey p = 0.8		-0.106	-0.139	-0.072	<0.01
Wright 0.5-Holey p = 0.8		0.437	0.404	0.470	<0.01
Drift-Holey p = 0.8		0.119	0.086	0.152	<0.01
Wright 0.5-Wright 0.1		0.543	0.509	0.576	<0.01
Drift-Wright 0.1		0.225	0.192	0.258	<0.01
Drift-Wright 0.5		-0.318	-0.351	-0.285	<0.01

Post-hoc Parameter Exploration

For populations evolving on Gaussian landscapes, compression of genetic variation into the leading dimension decreased with increasing heritability and an increasing strength of selection (Table S8, Figure S9). No two-way interaction was statistically significant. Put another way, λ_2/λ_1, increased with heritability and the strength of selection and average λ_2/λ_1 was 0.68 for average parameter values (Table S8).

For populations evolving solely due to drift, λ_2/λ_1 increased with greater initial total genetic variation (Table S9). However, the strength of this effect was minimal. More dramatically, λ_2/λ_1 significantly and strongly decreased with increasing average initial absolute genetic correlation (Table S9). At the extreme, λ_2/λ_1 approached 0 as the average initial absolute correlation approaches 1. No two-way interaction was statistically significant. Average λ_2/λ_1 was 0.69 for average parameter values (Table S9).

When evolving on holey landscapes, and consistent with prior simulation comparisons, λ_2/λ_1 was lower for average parameter values (0.42, Table S10). Compression into a single dimension also increased with increasing heritability and increasing average absolute initial correlations (Table S10).

Genetic variation was more strongly compressed into a primary dimension when populations evolved on holey landscapes versus when they evolved due to drift or due to selection on Gaussian surfaces (Tables S8 – S10; Figures S9 – S11). This was a surprisingly robust result regardless of the starting parameters of a simulation (Figures S9 – S11). This parameter robustness (10) supports the generality of our modeling. Unfortunately, we were not able to investigate other forms of robustness (10) due to computational limitations.

Our parameter exploration approach covered a wide range of possible values, making results from this exploration generalizable (Figures S12 – S14).

Table S8. Linear modeling results for Gaussian landscape parameter exploration. All covariates were modeled while centered (but not variance standardized).

Covariate	Estimate	Standard Error	t*	p
Intercept (average)	0.680	0.004	157.94	<0.01
Total variation (tot. var)	0.004	0.003	1.33	0.182
Mean |correlation| (mean cor)	-0.256	0.170	-1.51	0.132
h^2	**0.103**	**0.015**	**6.70**	**<0.01**
Selection strength (ss)	**0.069**	**0.019**	**3.60**	**<0.01**
tot.var × mean cor	-0.070	0.107	-0.66	0.513
tot.var × h^2	-0.008	0.010	-0.75	0.454
tot.var × ss	0.012	0.013	0.95	0.344
mean cor × h^2	0.806	0.601	1.34	0.180
mean cor × ss	0.384	0.733	0.52	0.600
h^2 × ss	-0.098	0.070	-1.39	0.164

*p values are based on this t value with 989 degrees of freedom

Table S9. Linear modeling results for parameter exploration of the drift model. All covariates were modeled while centered (but not variance standardized).

Covariate	Estimate	Standard Error	t*	p
Intercept (average)	0.689	0.004	165.12	<0.01
Total variation (tot. var)	**0.009**	**0.003**	**3.62**	**<0.01**
Mean |correlation| (mean cor)	**-0.867**	**0.160**	**-5.41**	**<0.01**
h^2	0.004	0.015	0.27	0.786
tot.var × mean cor	-0.034	0.103	-0.33	0.740
tot.var × h^2	0.015	0.009	1.63	0.103
mean cor × h^2	-0.209	0.555	-0.38	0.706

*p values are based on this t value with 993 degrees of freedom

Table S8. Linear modeling results for holey landscape parameter exploration. All covariates were modeled while centered (but not variance standardized).

Covariate	Estimate	Standard Error	t*	p
Intercept (average)	0.423	0.007	61.03	<0.01
Total variation (tot. var.)	0.007	0.004	1.56	0.119
Mean |correlation| (mean cor)	-0.195	0.265	-0.74	0.462
h^2	**-0.272**	**0.024**	**-11.22**	**<0.01**
p	0.011	0.024	0.47	0.640
tot.var × mean cor	0.070	0.150	0.47	0.640
tot.var × h^2	0.012	0.015	0.78	0.435
tot.var × *p*	0.007	0.015	0.48	0.631
mean cor × h^2	0.547	0.940	0.58	0.561
mean cor × *p*	-0.239	0.941	-0.26	0.799
h^2 × *p*	**-0.593**	**0.086**	**-6.93**	**<0.01**

*p values are based on this t value with 989 degrees of freedom

Empirically Estimated **G** Matrices

Phylogenetic Signal in λ_2/λ_1

Table S9. Variances for λ_2/λ_1—with associated 95% confidence intervals—at each taxonomic level, for study ID, and residual. Proportion of variation for taxonomy, study ID, and residual are also provided

Variance component	Estimate (95% CI)	Proportion of variance
Study ID	0.026 (0.013 : 0.048)	0.45
Taxonomy	0.003	0.05
species	0 (0 : 0.01)	
Genus	0 (0 : 0.016)	
Family	0.003 (0 : 0.02)	
Order	0 (0 : 0.018)	
Class	0 (0 : 0.008)	
Phylum	0 (0 : 0.007)	
Kingdom	0 (0 : 0.011)	
Residual	0.029 (0.023 : 0.037)	0.50

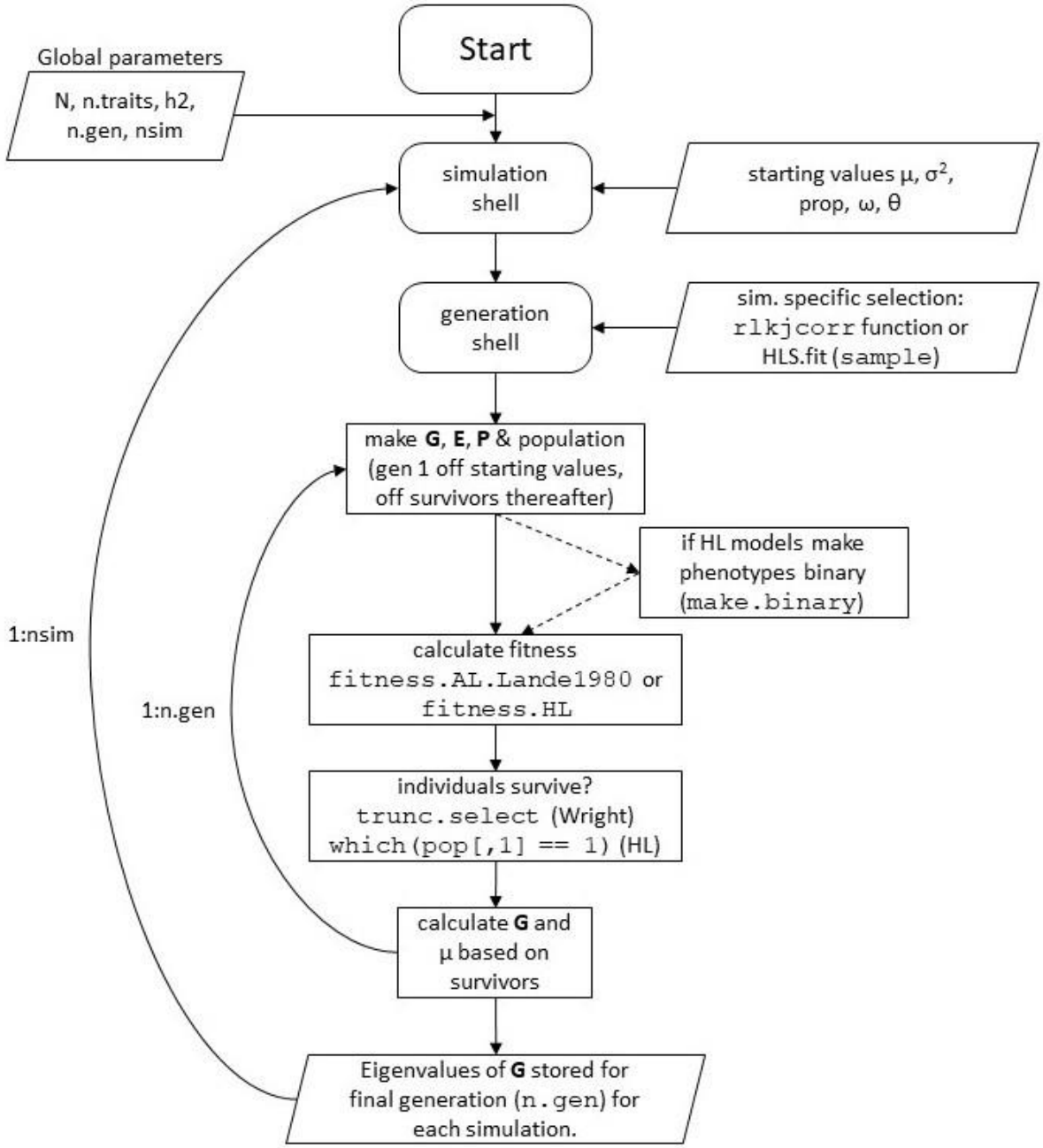
Comparison of Observed Results to Simulation Results

Observed results did not significantly differ from simulated populations that evolved on holey landscapes (Figure 2; Table S10). This conclusion based on λ_2/λ_1 is consistent when values of $\lambda_1/\sum\lambda$ are compared between simulated and empirical datasets (Figure S15)

Table S10. t values and associated p values for the comparison of the observed average of λ_2/λ_1 versus the average λ_2/λ_1 for each set of simulations. The observed average and its standard error was taken from a taxonomic mixed-effects model.

Average observed λ_2/λ_1	Simulation	Simulation average λ_2/λ_1	t	p
0.366 vs: (se: 0.03)	Holey (p = 0.2)	0.357	0.320	0.753
	Holey (p = 0.5)	0.331	1.199	0.247
	Holey (p = 0.8)	0.368	-0.050	0.961
	Gaussian (surv. prob. = 0.1)	0.650	-9.66	<0.01
	Gaussian (surv. prob. = 0.5)	0.731	-12.416	<0.01
	Drift	0.794	-14.552	<0.01

* degrees of freedom = 17.275



249 Figure S1. Model flow diagram for HL and gaussian landscapes

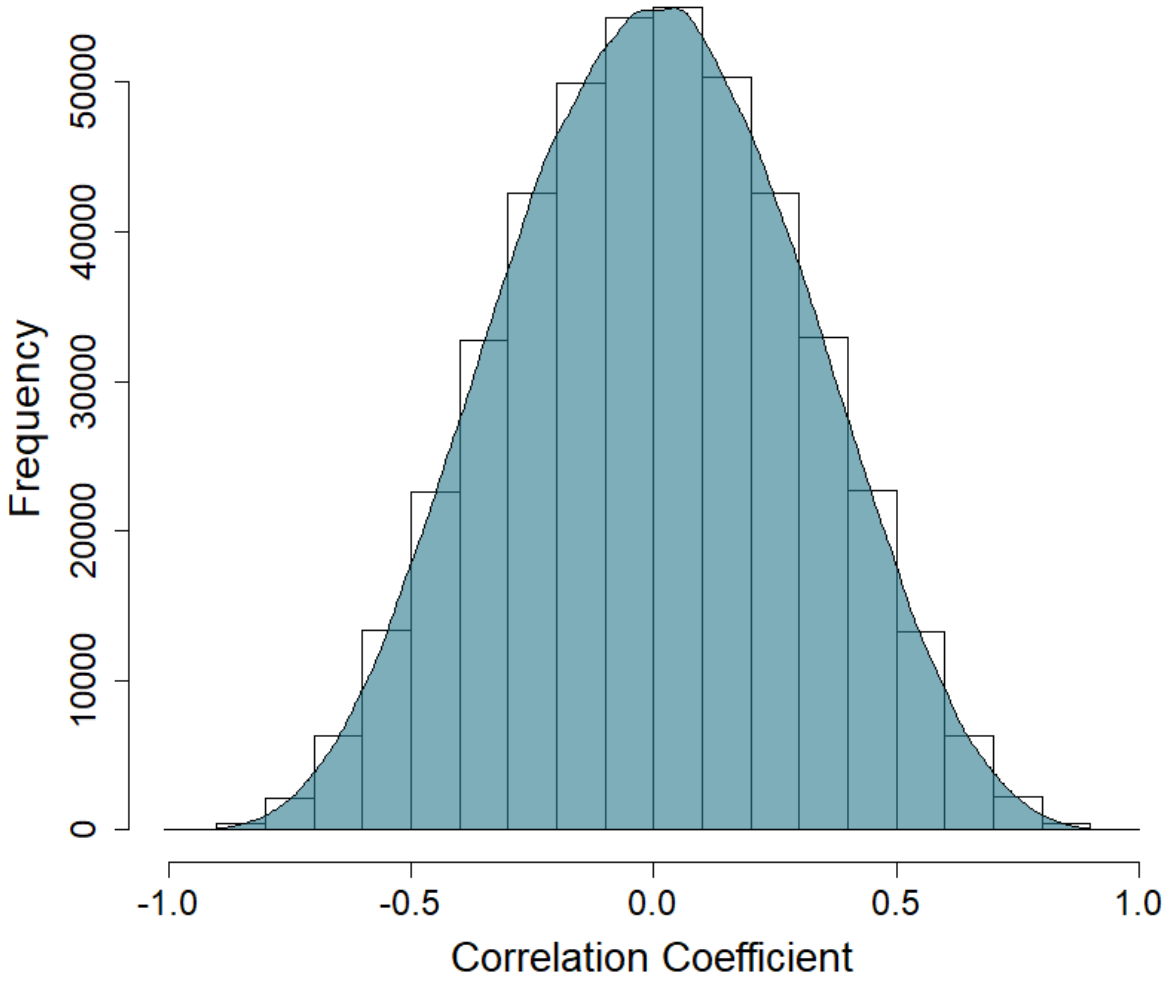


Figure S2. Distribution of 450000 random correlations generated using the LKJ Onion method with k = 10.

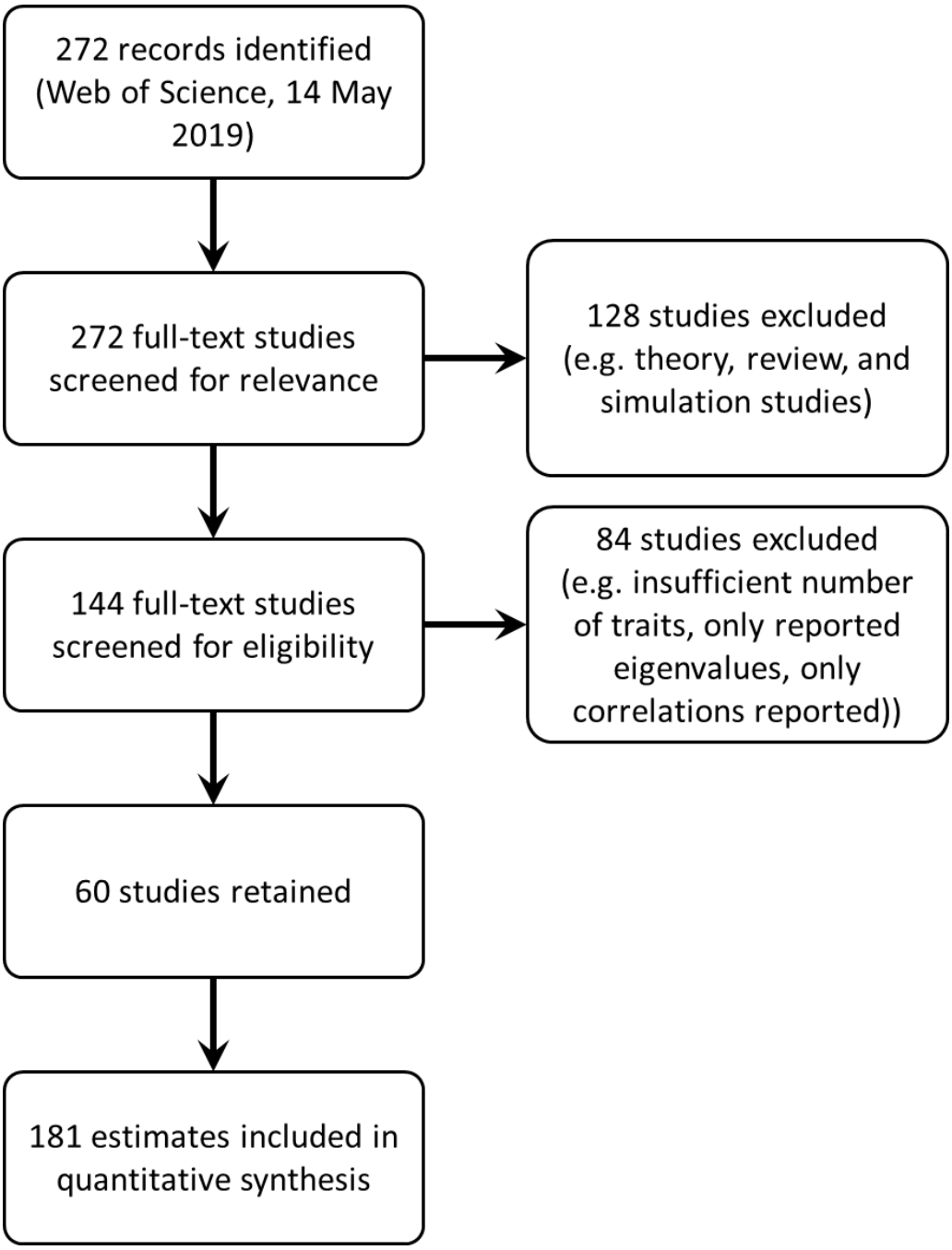


Figure S3. PRISMA diagram for studies and estimates included in taxonomic analyses.

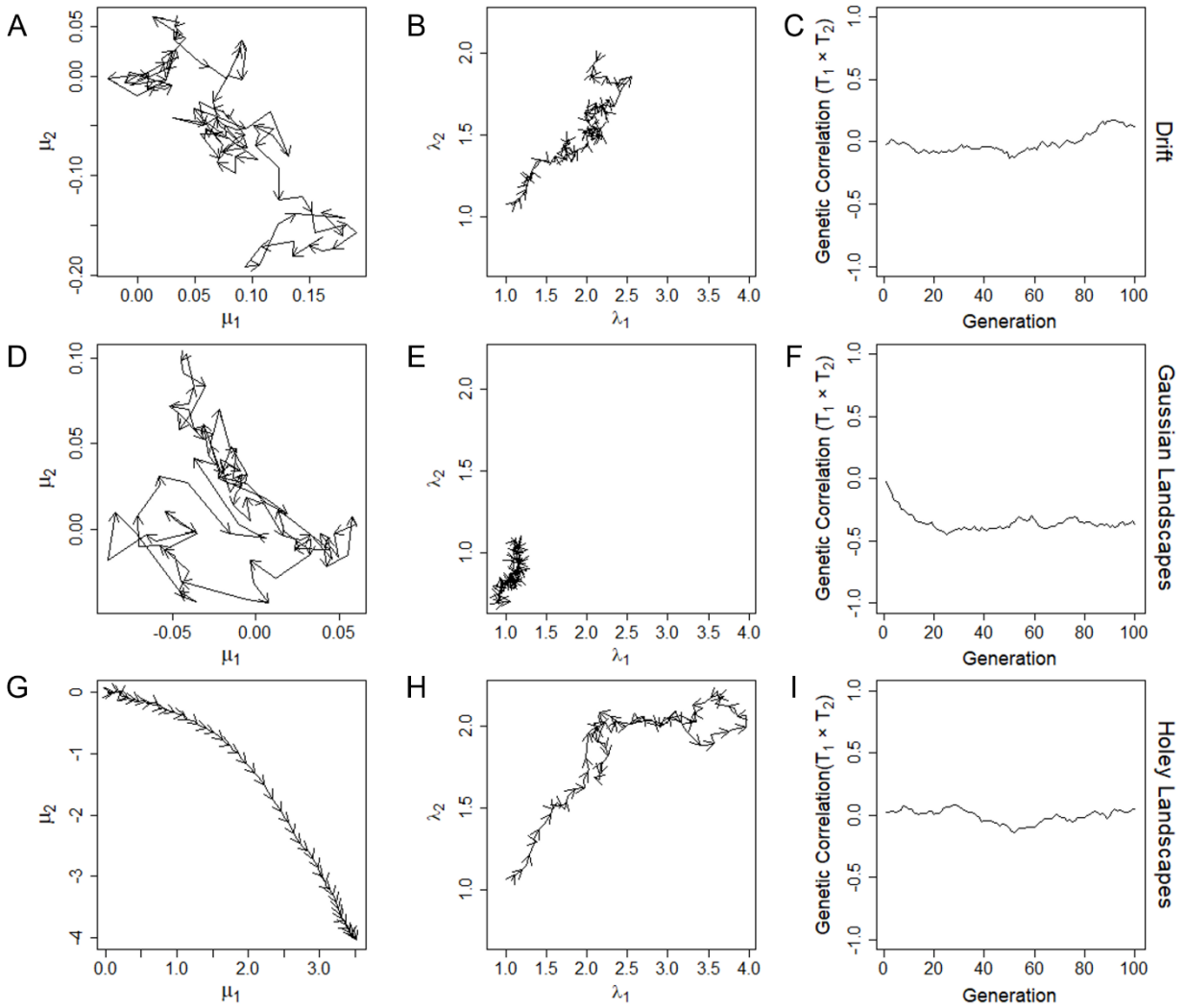


Figure S4. Single population comparisons of population evolution over 100 generations under drift (A – C), on a Gaussian landscape (D – F), and on a holey landscape (G – I). Arrow heads in A, B, D, E, G, and H indicate the direction of evolutionary change at every second generation. Evolutionary change in the average values for traits 1 and 2 (A, D, and G; note the different scales for axes) show little change for either drift and on a Gaussian landscape. In contrast, the population shows substantial and directional change in trait values on a holey landscape (G). This suggests the population is moving between holes in G but is restricted to a local optimum in D. The first and second eigenvalues (λ_1 & λ_2) show little total change due to drift (B), consistent decreases on a Gaussian landscape (E), and larger changes—including overall increases—on a holey landscape (H). This is consistent with the overall compression of variance reported elsewhere in the main and supplemental results. The bivariate genetic correlation between the first two traits shows little directional change under either drift of on holey landscapes (C and I) but rapid absolute increase followed by becoming static on a Gaussian landscape (F). As was the case for eigenvalues (E), this is consistent with stabilizing selection at a local optimum. These results are consistent across multiple runs, though exact trajectories vary and the sign of genetic correlations is equally likely to be positive as negative.

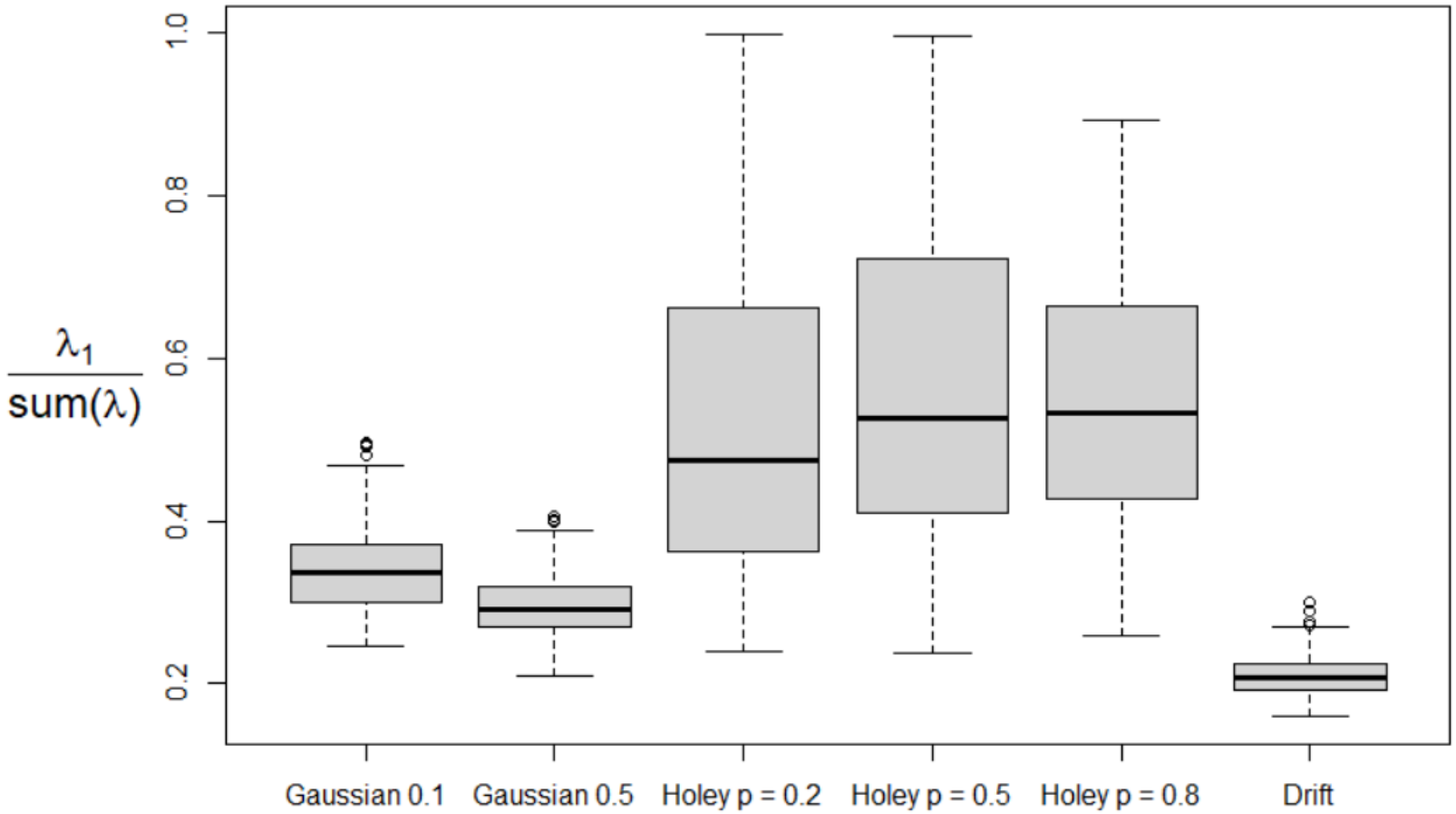


Figure S5. Variation was more evenly distributed across dimensions when populations evolved on Gaussian landscapes or due solely to drift. Consequently, less total variation was present in the first dimension (Table S4).

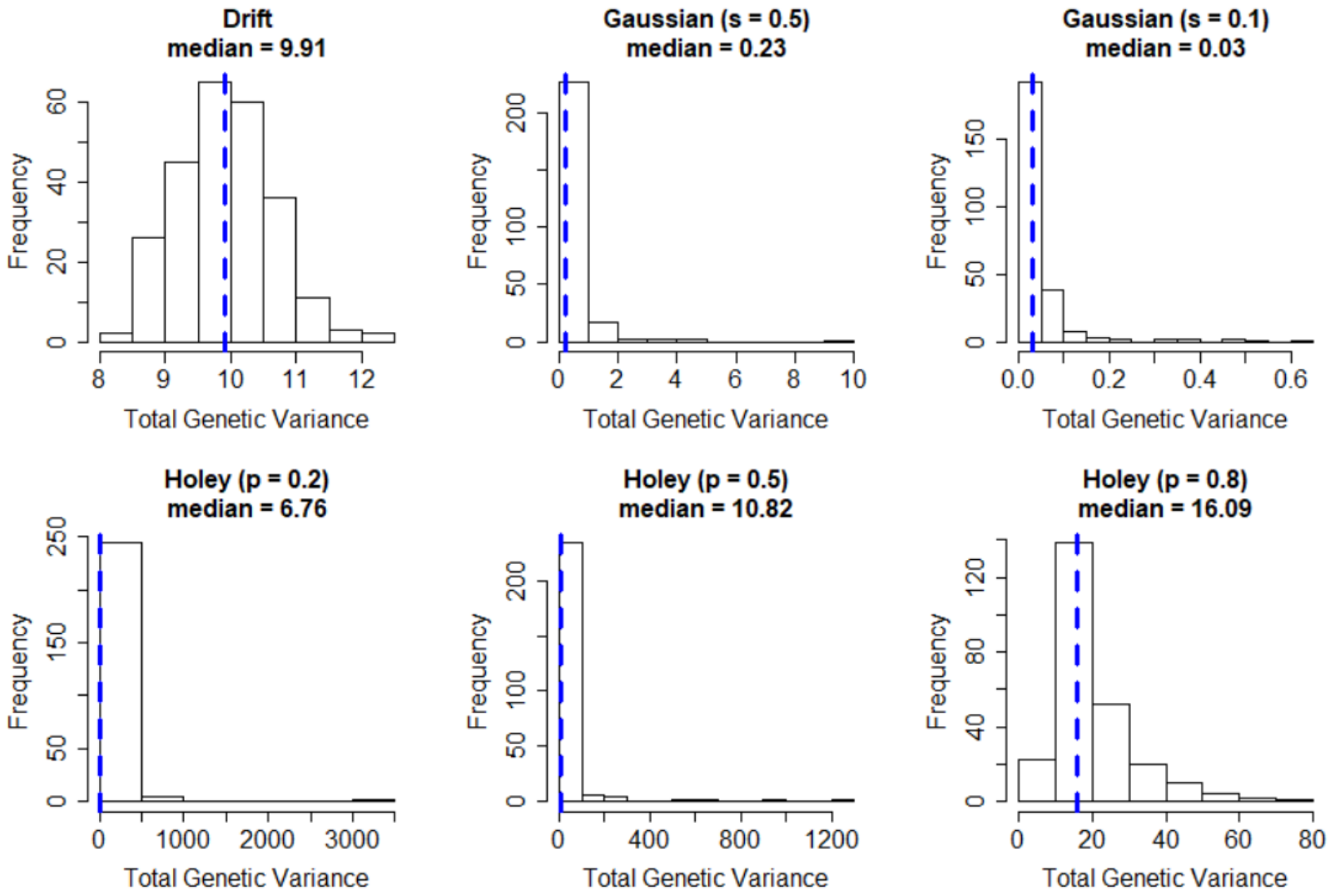


Figure S6. The total genetic variation present after 100 generations in each of six modeling conditions and across 250 simulations. Selection on Gaussian surfaces led to a significant reduction in the amount of variation present (Table S5). Dashed blue bars indicate the median value.

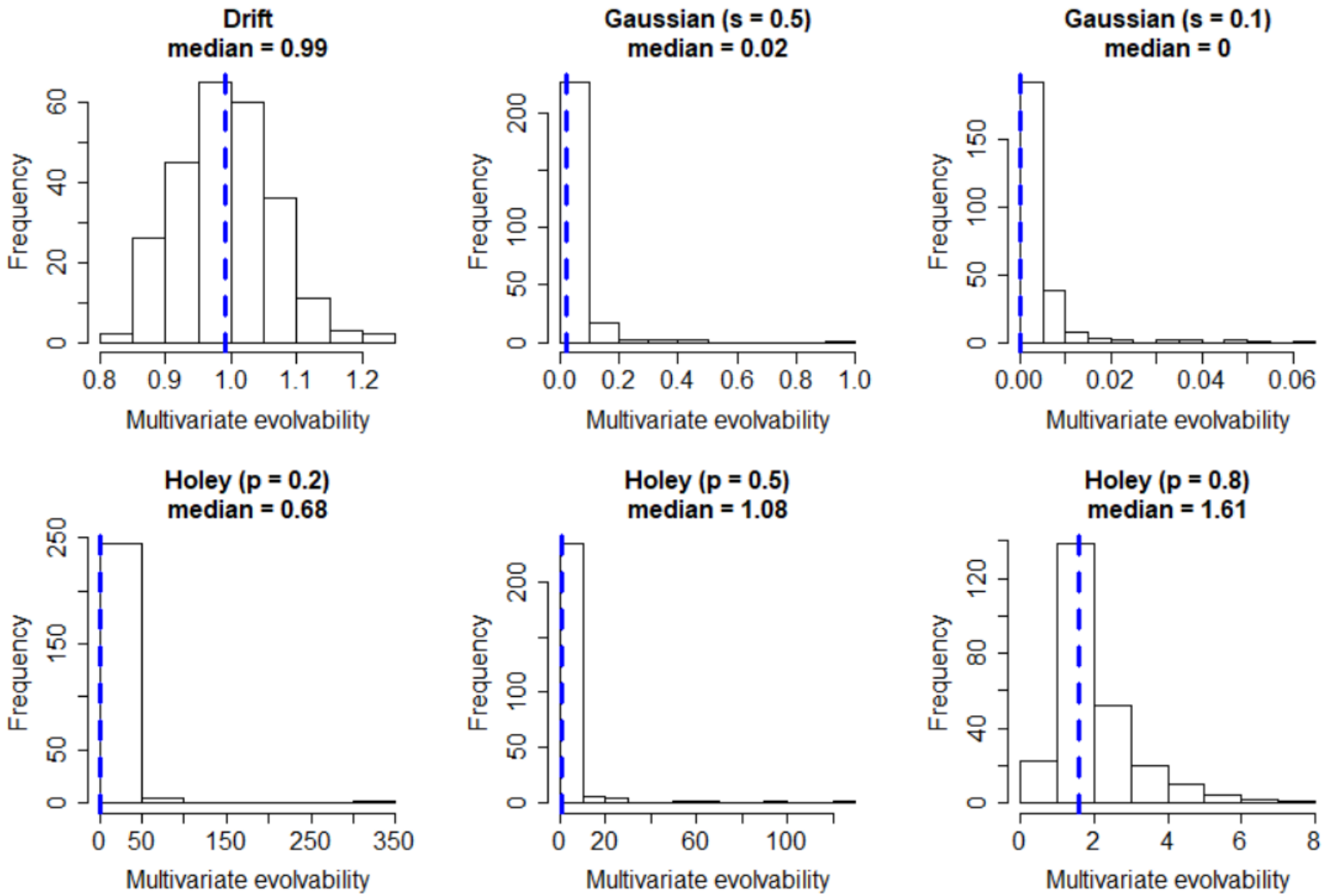


Figure S7. Multivariate evolvability after 100 generations in each of six modeling conditions and across 250 simulations. Selection on Gaussian surfaces led to a significant reduction in evolvability (Table S6). Dashed blue bars indicate the median value.

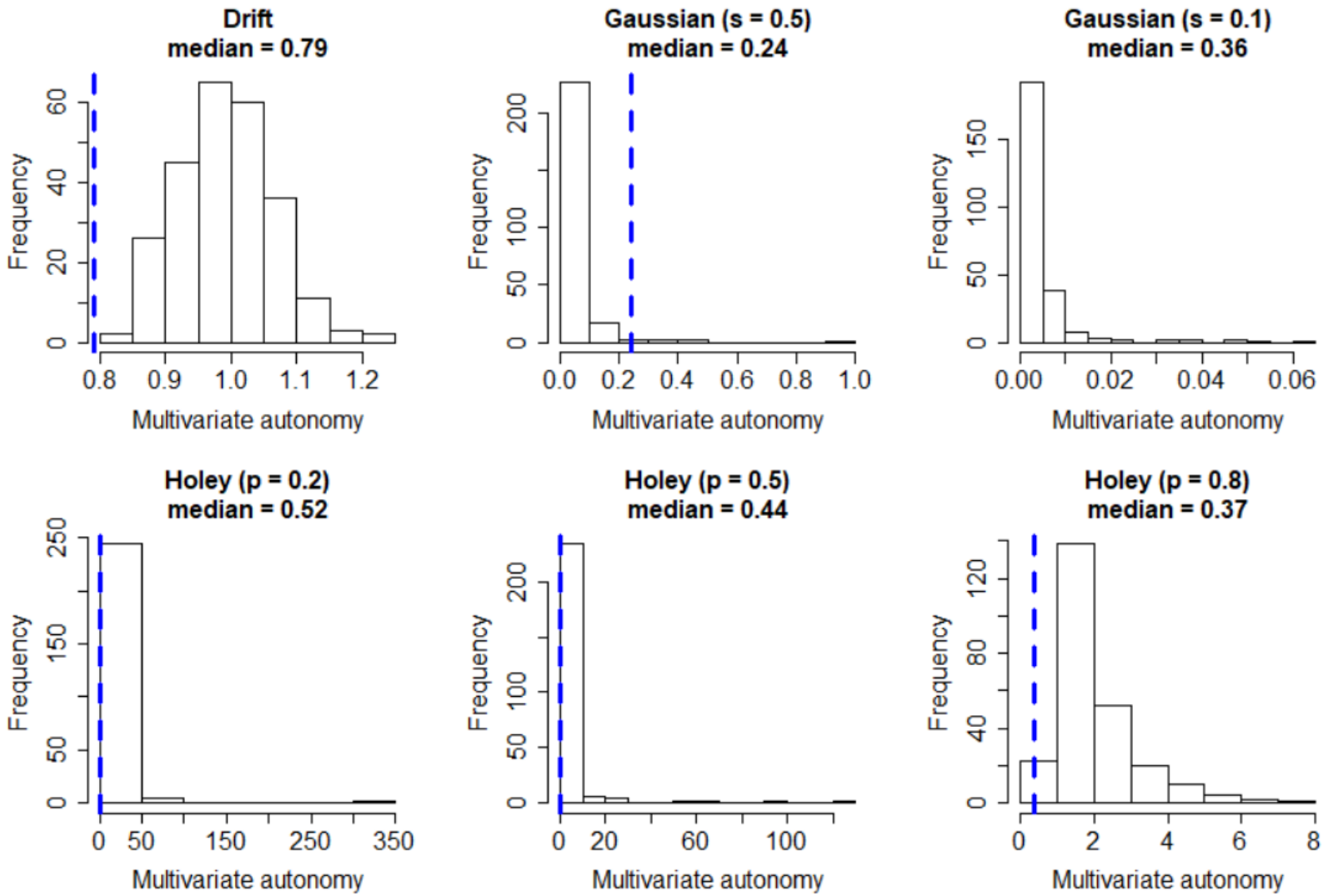


Figure S8. Multivariate autonomy after 100 generations in each of six modeling conditions and across 250 simulations. Selection on Gaussian surfaces led to a significant reduction in autonomy (Table S7). Dashed blue bars indicate the median value.

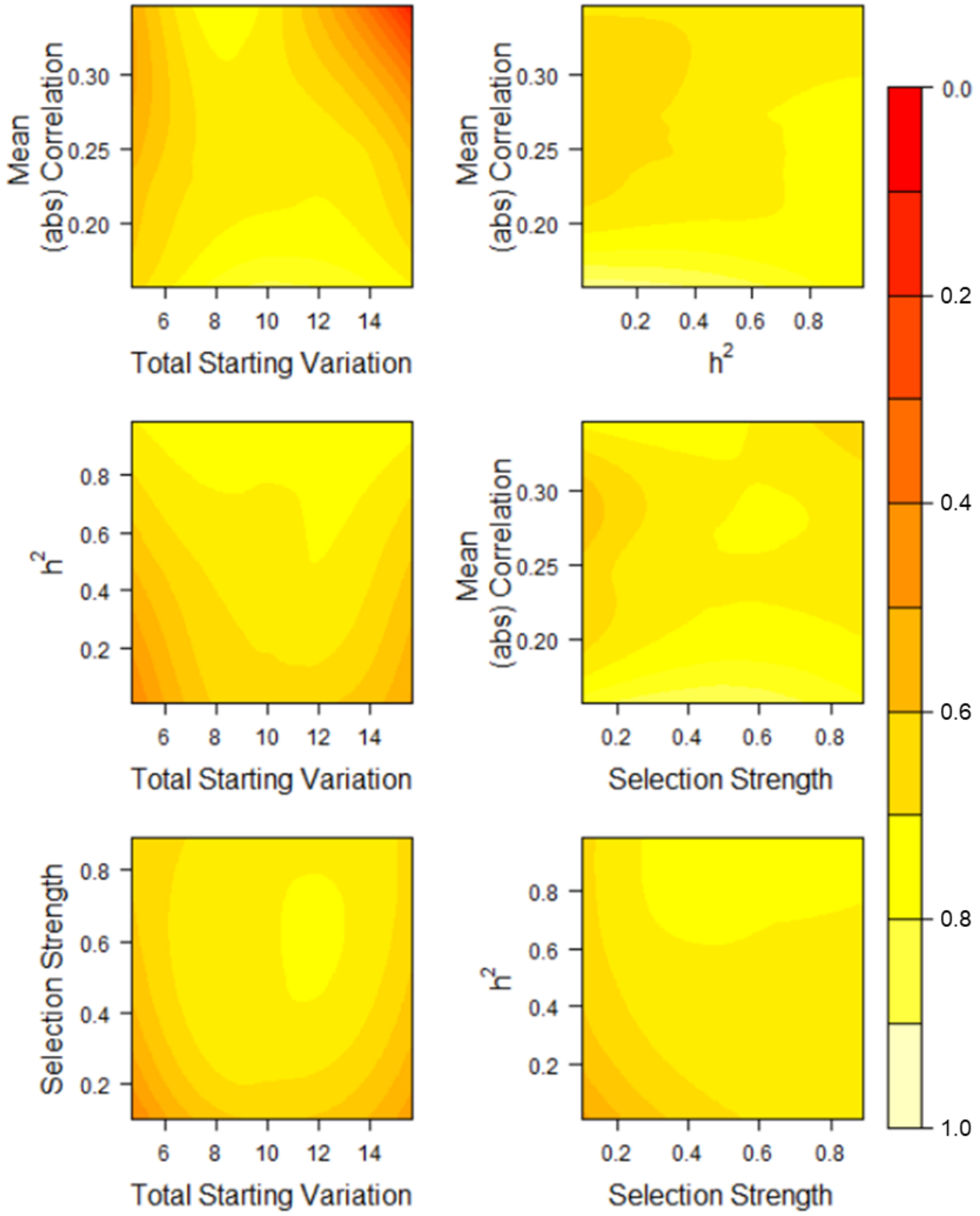


Figure S9. λ_2/λ_1 after selection on Gaussian surfaces remained high regardless of starting parameters (Table S8).

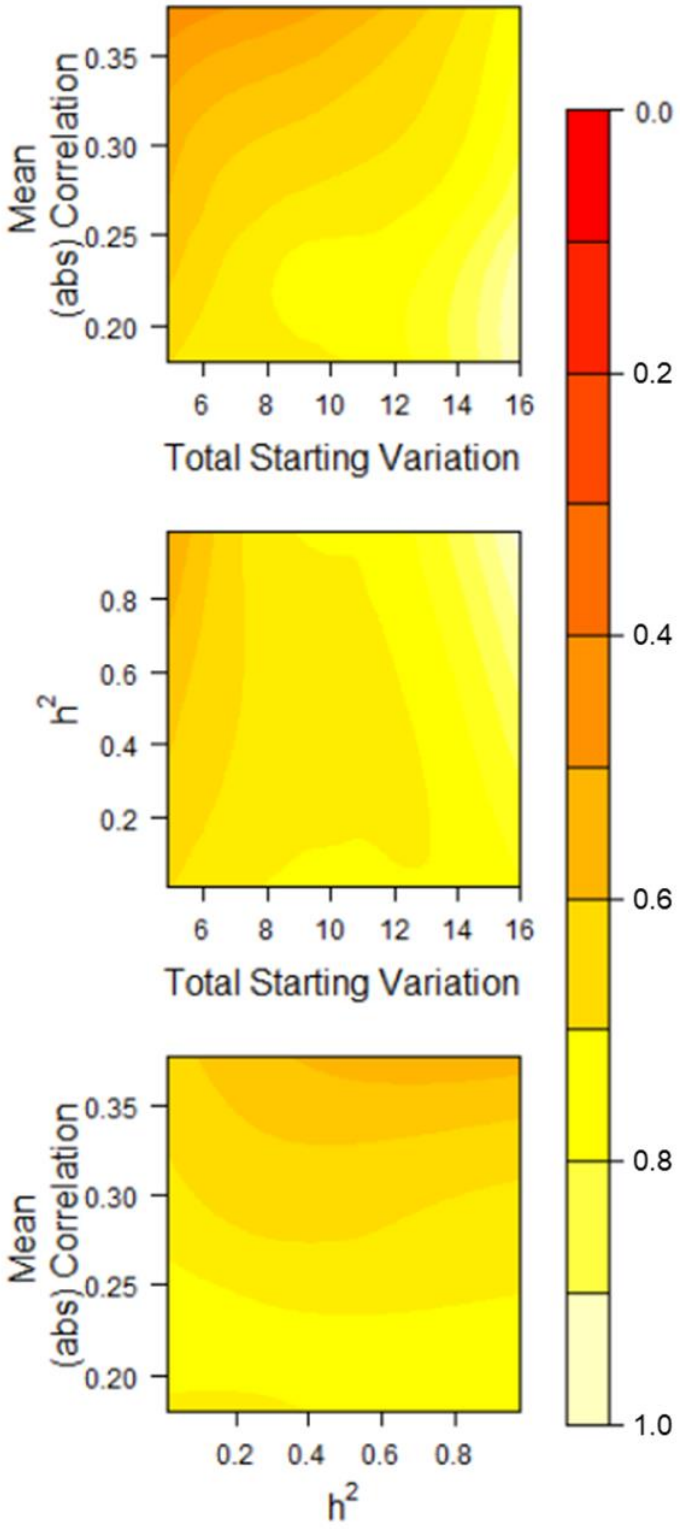


Figure S10. λ_2/λ_1 after evolution due to drift remained high regardless of starting parameters (Table S9).

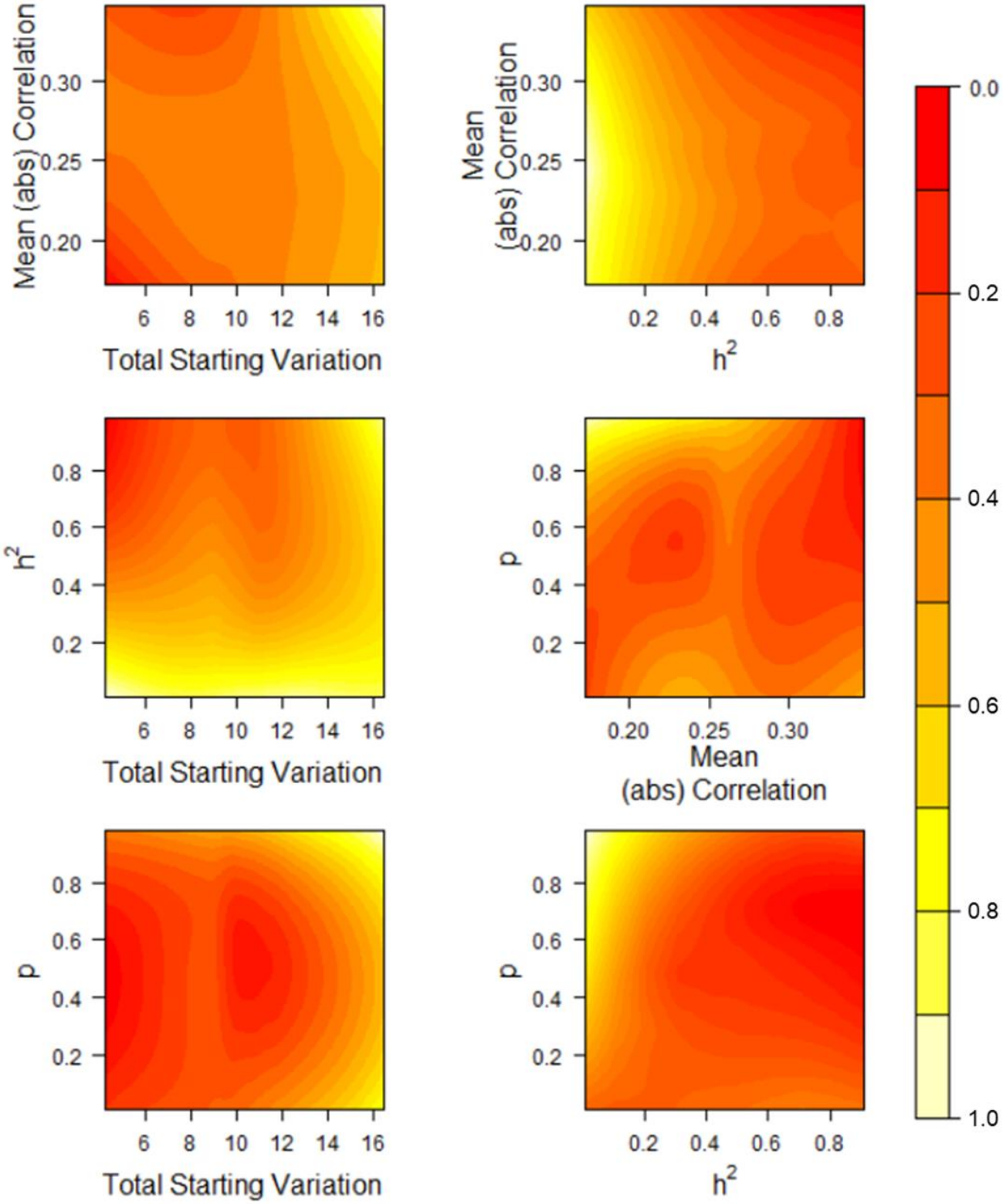


Figure S11. λ_2/λ_1 after evolution on holey landscapes remained low regardless of starting parameters (Table S10).

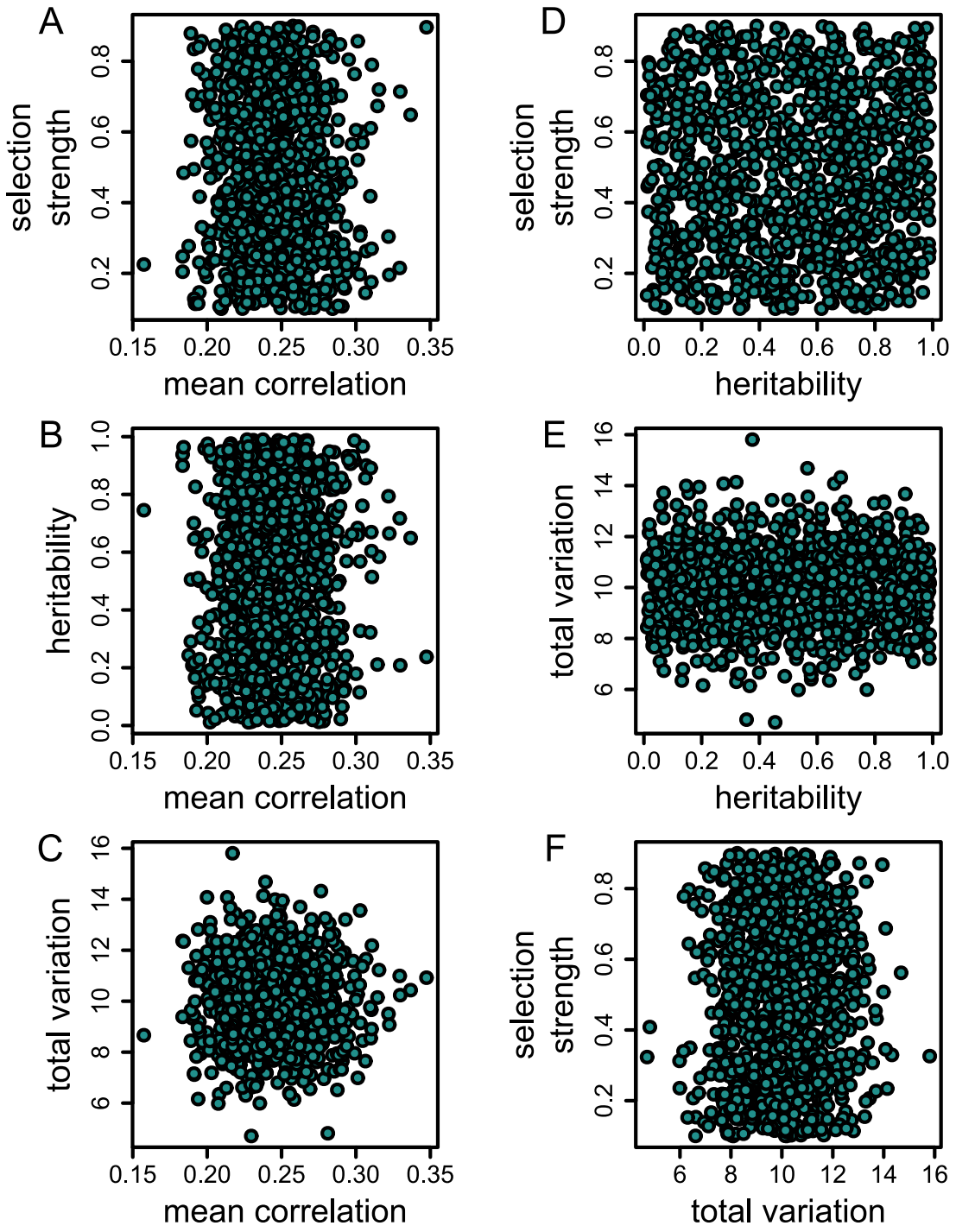


Figure S12. The parameter space across 1000 random sets of Gaussian landscapes covered a wide-range of plausible values of genetic variation, genetic correlations, and selection strengths. "Mean correlation" is the average *absolute* correlation (mean|*r*|) of an iteration's 45 total correlations. The distribution of mean|*r*| is constrained due to values needing to satisfy the requirements to be a correlation matrix (e.g. being positive definite).

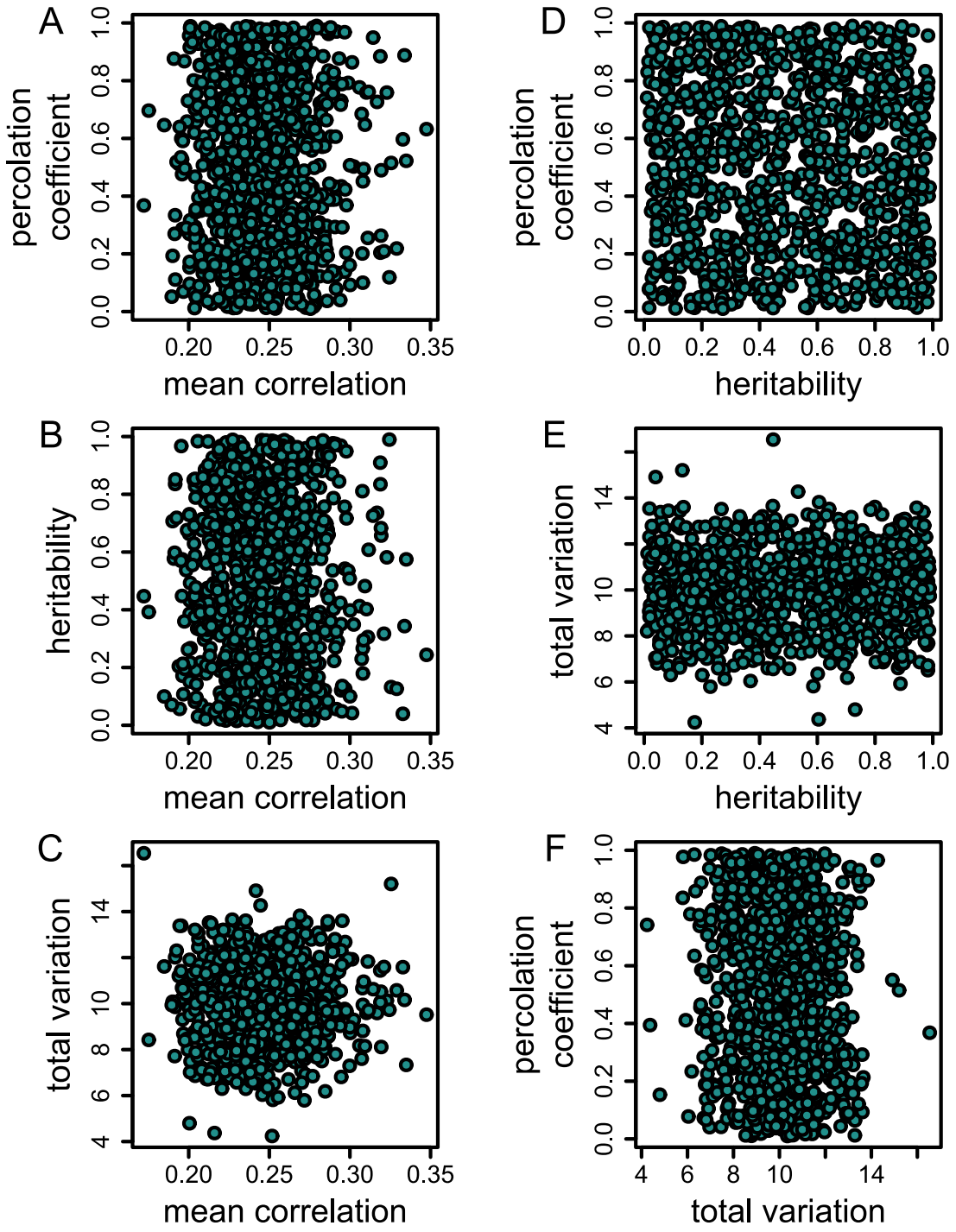


Figure S13. The parameter space across 1000 random sets of holey landscapes covered a wide-range of plausible values of genetic variation, genetic correlations, and proportion of viable phenotypes (i.e. the percolation coefficient). "Mean correlation" is the average *absolute* correlation (mean|r|) of the iteration's 45 total correlations. The distribution of mean|r| is constrained due to values needing to satisfy the requirements to be a correlation matrix (e.g. being positive definite). The percolation coefficient (i.e. p) determines the probability that a phenotypic combination is viable.

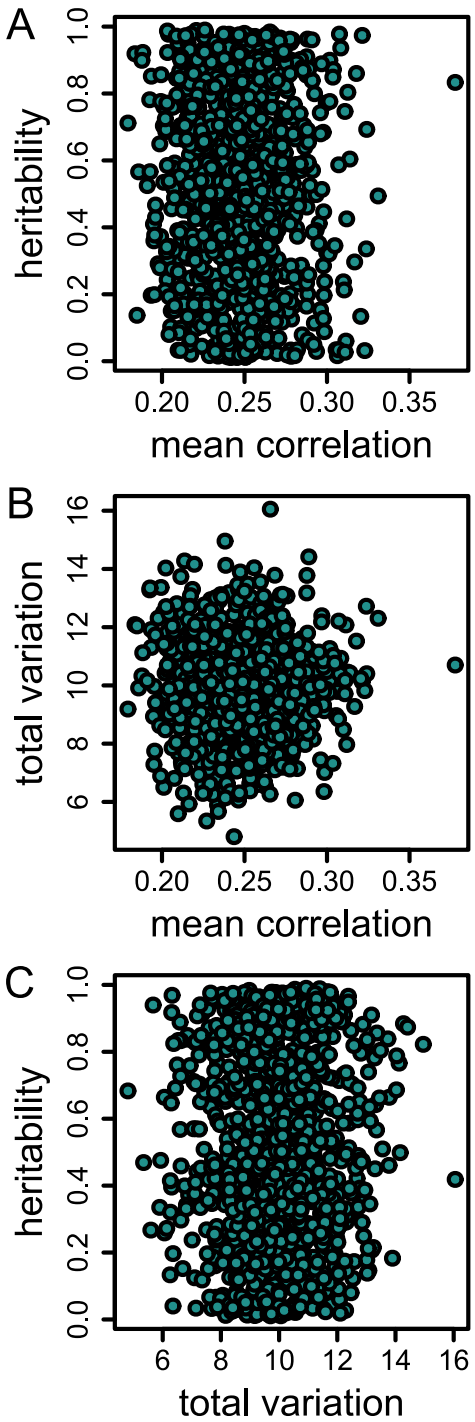


Figure S14. The parameter space across 1000 random sets of flat, drift landscapes covered a wide-range of plausible values of genetic variation and genetic correlations. "Mean correlation" is the average *absolute* correlation (mean|*r*|) of the iteration's 45 total correlations. The distribution of mean|*r*| is constrained due to values needing to satisfy the requirements to be a correlation matrix (e.g. being positive definite).

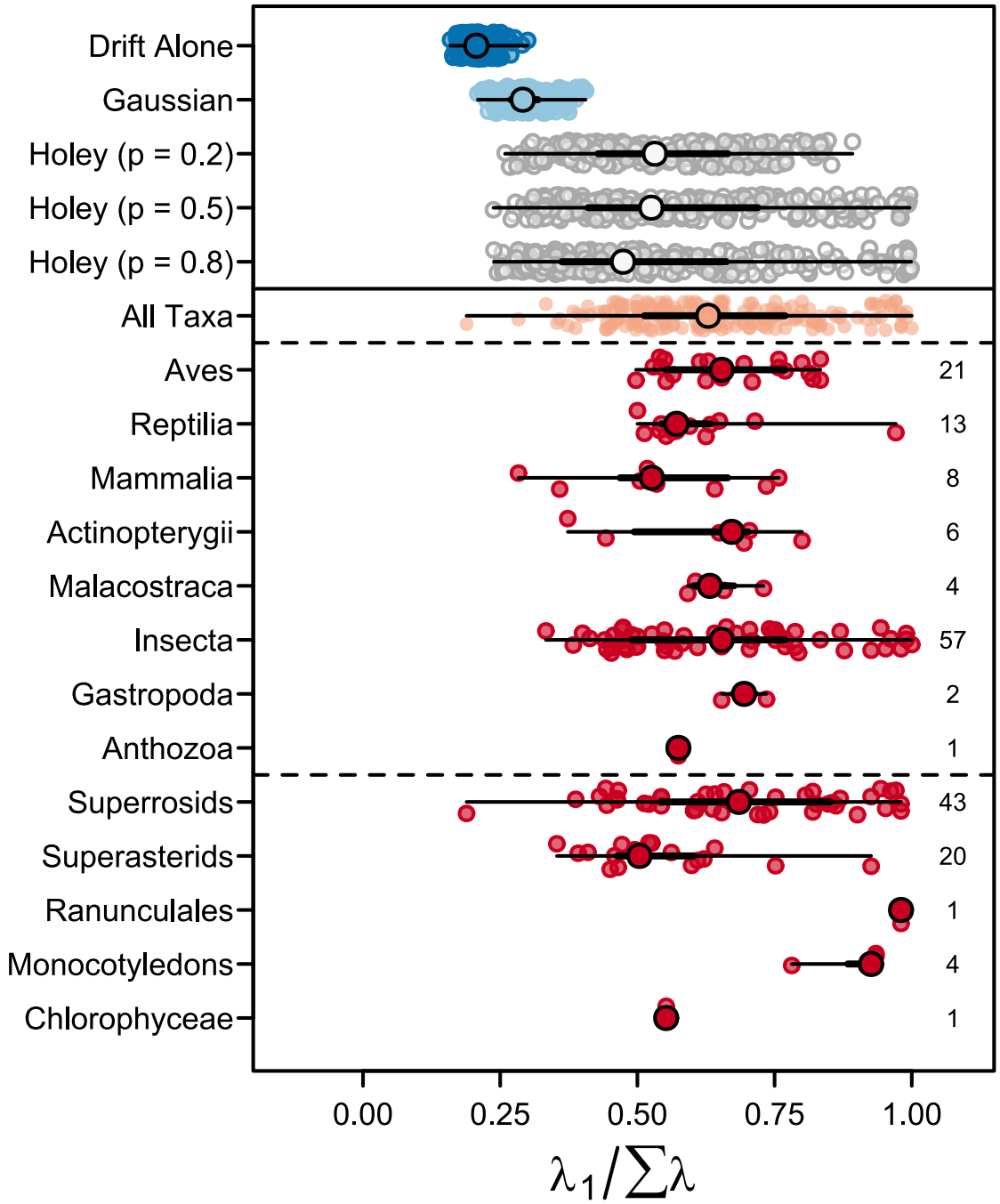


Figure S15. Modified "Orchard plot" of λ1/Σλ values for simulated (above solid line) and observed **G** matrices. *Trunks* (large points) are the medians for the specified group (e.g. Gaussian landscapes or Insecta), *branches* (thick lines) are interquartile ranges, *twigs* (thin lines) give the full range of values, and *fruits* (smaller points) are individual estimates within a simulation or taxonomic group. Rightmost numbers are the number of estimates available via literature search.

References Cited

1. Roff D (2010) *Modeling evolution: an introduction to numerical methods* (Oxford University Press, Oxford) p 352.
2. Arnold SJ, Burger R, Hohenlohe PA, Ajie BC, & Jones AG (2008) Understanding the evolution and stability of the G-matrix. *Evolution* 62(10):2451-2461.
3. Lewandowski D, Kurowicka D, & Joe H (2009) Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* 100(9):1989-2001.
4. Lande R (1980) Genetic variation and phenotypic evolution during allopatric speciation. *The American Naturalist* 116(4):463-479.
5. Hansen TF & Houle D (2008) Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology* 21(5):1201-1219.
6. Kirkpatrick M (2009) Patterns of quantitative genetic variation in multiple dimensions. *Genetica* 136(2):271-284.
7. Nakagawa S & Santos ES (2012) Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology* 26(5):1253-1274.
8. Bates D, Mächler M, Bolker B, & Walker S (2014) Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
9. Zar JH (1999) *Biostatistical Analysis* (Pearson Education, Inc., Upper Saddle River, NJ) 4th Ed.
10. Weisberg M & Reisman K (2008) The robust Volterra principle. *Philosophy of science* 75(1):106-131.