Drift on holey landscapes as a dominant evolutionary process

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

Affiliations:
[a] Department of Biological Sciences; North Dakota State University
[b] ned.dochtermann@gmail.com
[c] Department of Biology; University of California, Riverside
[d] French National Institute for Agriculture, Food, and Environment (INRAE), Versailles Cedex, France
* these authors are listed in alphabetical order

Abstract: Phenotypes typically display integration, i.e. correlations between traits. For quantitative traits—like many behaviors, physiological processes, and life-history traits—patterns of integration are often assumed to have been shaped by the combination of linear, non-linear, and correlated selection, with trait correlations representative of optimal combinations and reflective of the adaptive landscapes that have shaped a population. Unfortunately, this assumption has rarely been critically tested, in part due to a lack of clear alternatives. Here we show that trait integration across 6 phyla and 60 species (including both Plantae and Animalia) is consistent with evolution across high dimensional "holey landscapes" rather than classical models of selection. 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.

One-Sentence Summary: Patterns of correlations among traits are inconsistent with dominant models of evolution and suggest, instead, that quantitative traits have predominantly evolved via drift of populations across high dimensional holey landscapes.

A common attribute of most organisms is that they display trait integration. For example, life-history traits are often correlated according to a slow-fast continuum [1,2]. This trait integration is commonly understood in terms of trade-offs and fitness maximization [3-8] and is frequently modeled as populations moving across adaptive landscapes toward peaks of higher fitness. However, this adaptive perspective has rarely been evaluated due to a lack of clear alternatives. Consequently, much of our understanding of when and why quantitative traits are correlated might be shaped by adaptive just-so-stories [9].

Competing evolutionary processes

Our understanding of selection has been strongly shaped by Sewall Wright's conceptualization of an adaptive landscape, with populations moving from areas of low fitness to areas of higher fitness [10,11]. While the simple one and two trait landscapes Wright originally detailed have been criticized as unrealistic, including by Wright himself [10], the general metaphor has nonetheless guided much of evolutionary thought [12].

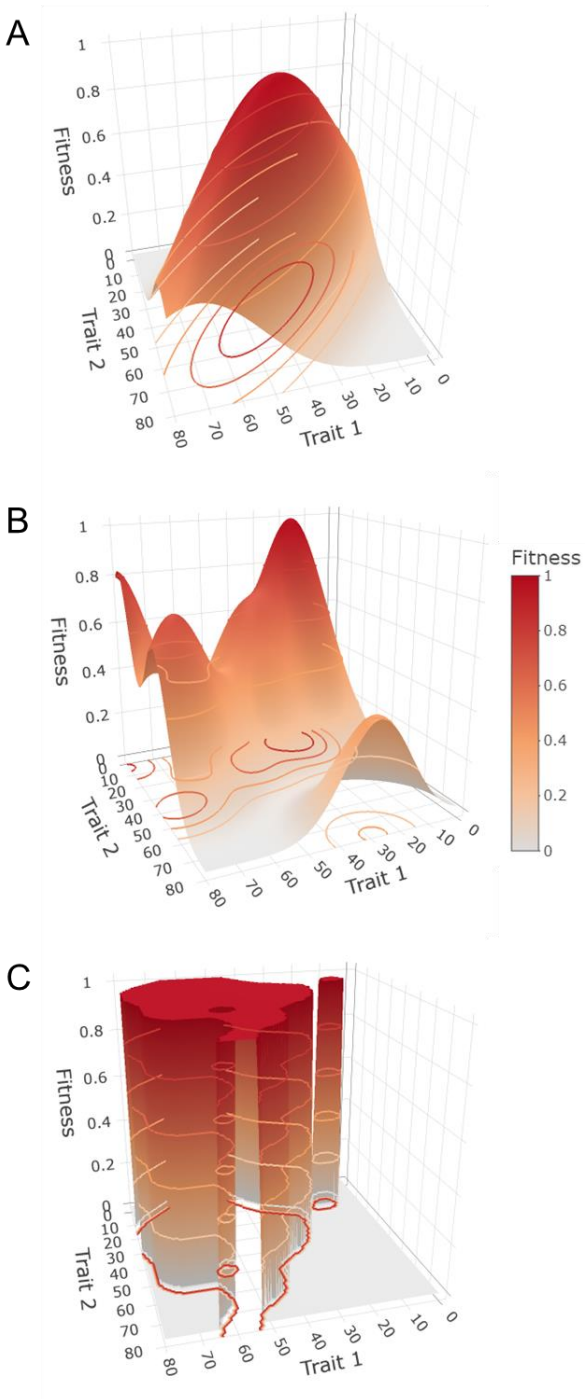


Figure 1. Example fitness landscapes. Redder colors correspond to higher fitness. A. A simple Gaussian, single peak Fujiyama landscape with a single optimum (1). B. A more rugged landscape with multiple local optima and a single global optimum. C. A simplified Holey landscape where particular combinations of values correspond to high, average, fitness (1) or low (0) fitness.

For quantitative traits, like many aspects of physiology, behavior, and morphology, Wright's metaphor has been mathematically extended to complex topographies with ridges or tunnels of high fitness [13-15]. Applying these adaptive landscape topologies in mathematical models has led to insights into how evolutionary change may not be monotonic and how correlations may evolve [15]. Simulations have similarly led to the prediction that landscapes with complex topographic features like fitness ridges result in populations with genetic correlations aligned with these ridges [3-5]. This has led, for example, to an ability to predict how the pace of evolution might be constrained [16].

Concurrent to the study of quantitative trait variation, the question of how the topography of fitness landscapes affects sequence evolution at the genomic level has garnered similar interest [17]. Whereas theoreticians interested in quantitative trait variation have focused on relatively simple landscapes e.g. [3,4,5,18-20], theoretical research regarding sequence evolution has spanned simple single peak Gaussian "Fujiyama landscapes", to "badlands landscapes" (Fig 1A & 1B [21]), to abstract high-dimensional "holey landscapes" (Fig 1C [22]). Among other topics, this research has explored how topographies of varying complexity (Fig 1) affect the ability of populations to reach optima [17]. An important conclusion from this research is that evolutionary dynamics on simple landscapes often fail to properly predict evolution on landscapes of higher dimensionality.

Of these landscapes, perhaps most conceptually unfamiliar and unintuitive to researchers focused on quantitative trait evolution are Gavrilets' (1997) holey landscapes (Fig 1C). Holey landscapes consist of trait combinations conferring either average fitnesses or zero. The general concept of holey landscapes is that, because phenotypes are made up of a large number of traits, phenotypes are abstract high dimensional constructs and corresponding landscapes will consist of either trait combinations that are of average fitness or trait combinations that confer low fitness or are inviable [22,23]. This results in flat landscapes with holes at inviable or low fitness phenotypes (Fig 1C). The flat landscape can be understood as stemming from the full multivariate nature of the phenotype: while there may be clear fitness differences in two dimensions, strong gradients will create holes in the landscape and peaks will average out when additional traits are considered. This is conceptually related to more recent discussions of the Pareto optimization of traits [24,25].

Under Pareto optimization across just three traits a flat fitness surface—the Pareto front—connects single trait × environment optima (i.e. "archetypes" [25]). Likewise, rugged landscapes can create steep fitness declines and consequent holes in the overall landscape. Unfortunately, predictions about quantitative trait evolution on holey landscapes are not clear.

Even more broadly, it is not clear what the topography of landscapes typically is for natural populations. While portions of selection surfaces and fitness landscapes can be directly estimated [26,27], these estimates may differ from the underlying full landscape due to several factors. These include: the omission of fitness affecting traits [28], incomplete estimation of fitness [29,30], and insufficient power to estimate non-linear selection coefficients [31]. An alternative to direct estimation of adaptive 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 [32], and additive genetic correlations are expected to emerge from correlational selection and fitness ridges in a landscape, e.g. [13,14]. Thus, an ability to gain an understanding of the topography of adaptive landscapes based on observed trait variation would aid our understanding as to how selection is realized in natural populations.

Here we used a simulation model to examine how evolution on different landscapes contributes to patterns of trait integration. We modeled populations that evolved solely via drift, that evolved via adaptation on simple Gaussian fitness landscapes stemming from Wright's metaphor, or that evolved on holey landscapes. This allowed us to generate testable predictions for how the structure of additive genetic variances and covariances (**G**) are shaped by different landscape topographies. We next compared these modeled outcomes to 181 estimates of **G**, representing 60 species from 6 phyla, including both plants and animals, to determine if observed trait integration is consistent with any of the modeled processes.

Model Construction

We developed an individual variance components model (Methods, Fig S1 [33]) 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 [22,23], 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. 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). This metric 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 confirmed the robustness of the findings reported below.

Model outcomes

When evolving on holey landscapes, 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). 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 likely is due to the observation elsewhere that, when p is greater than $1/2^k$, a cluster of viable phenotypic values—and therefore phenotypic space exists—through which a population can drift [34,35]. 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 driven by the high power available with 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).

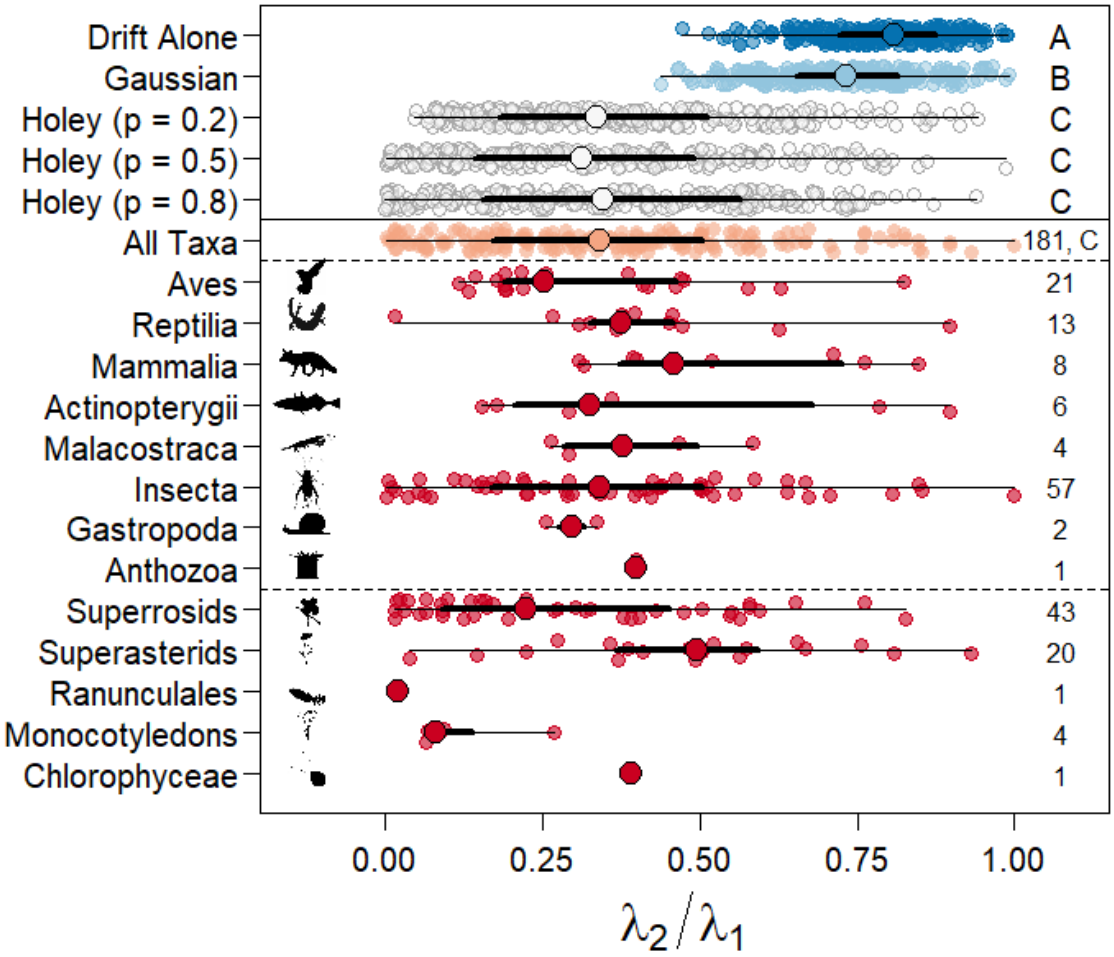


Figure 2. Modified "Orchard plot" of λ_2/λ_1 values for simulated (above solid line) and observed **G** matrices. *Trunks* (large points) are the medians for the specified group (e.g. Gaussian landscapes or Insecta), *branches* (thick lines) are interquartile ranges, *twigs* (thin lines) give the full 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).

Observed patterns of trait integration

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.

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 simulated populations that evolved on Gaussian landscapes or via drift alone ($t_{df:17.275}$ = -12.42, -14.55 respectively, p < 0.001 (both)).

While some individual estimates at the species level exhibited high λ_2/λ_1 values (Fig 2), phylogeny explained little variation in these values (phylogenetic heritability = 0.05; Table S9). As was the case across all taxa, median λ_2/λ_1 values for each taxonomic Class (or comparable level clade) were consistently lower than expected if evolution occurred on Gaussian landscapes or via drift alone (Fig 2). Instead, these results are strongly consistent with evolution on Holey landscapes.

Conclusions

The observation that traits linked to fitness are frequently correlated has been a major driver of research across evolutionary ecology. Research in life-history, physiology, and behavior has frequently been structured around such observations, arguing that this integration stems from optimization in the face of trade-offs [1,2,36-38]. However, because selection is frequently acting on many traits, patterns of integration quickly diverge from simple expectations, even under conventional models of evolution. *However, our results suggest something substantively different is occurring: the observed pattern of variation across taxa suggests that classic models of the evolution of quantitative traits—e.g. stabilizing and correlational selection—are not what have predominantly shaped trait integration.* Instead, drift across holey landscapes [22,23] is more consistent with observed quantitative genetic variation (Fig 2).

Much of the theoretical development of holey landscapes focused on the ability of populations to traverse genomic sequence differences via drift, with some sequences being inviable (e.g. due to missense differences in coding regions). How this extended to quantitative traits was less clear. Our simulation model provides one approach to applying the holey landscape framework to quantitative traits, treating each trait as a threshold character [39]. Other approaches to modeling quantitative traits on holey landscapes and evolution in response to these versions, such as the generalized Russian roulette model [23], may produce different outcomes. It is also important to recognize that the broad support for evolution on holey landscapes does not preclude subsets of traits from having evolved on Gaussian landscapes. Indeed, stabilizing selection has been observed in natural populations [31], though understanding its general strength even on a case-by-case basis is confounded with methodological problems [40,41]. Regardless, our finding that observed patterns of quantitative genetic variation across taxonomic groups are not consistent with traditional evolutionary models stands.

This disconnect between observed patterns of multivariate variation and expectations under conventional models of selection suggests that Wright's metaphor of fitness landscapes and the subsequent implementation of this metaphor as Gaussian surfaces may have contributed to an improper, or at least incomplete, understanding of

how selection has shaped phenotypes. A potential contributor to this problem has been the lack of clear alternative explanations besides a simple null hypothesis of drift with no selection. Moving forward, clear development of alternative models of the action of selection and evolution in multivariate space are needed.

References Cited

1 Wright, J., Solbu, E. B. & Engen, S. Contrasting patterns of density-dependent selection at different life stages can create more than one fast–slow axis of life-history variation. *Ecol. Evol.* **10**, 3068-3078 (2020).

2 Ricklefs, R. E. & Wikelski, M. The physiology/life-history nexus. *Trends in Ecology & Evolution* **17**, 462-468 (2002).

3 Arnold, S. J., Burger, R., Hohenlohe, P. A., Ajie, B. C. & Jones, A. G. Understanding the evolution and stability of the G-matrix. *Evolution* **62**, 2451-2461, doi:10.1111/j.1558-5646.2008.00472.x (2008).

4 Jones, A. G., Arnold, S. J. & Borger, R. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution* **57**, 1747-1760 (2003).

5 Jones, A. G., Arnold, S. J. & Burger, R. Evolution and stability of the G-matrix on a landscape with a moving optimum. *Evolution* **58**, 1639-1654 (2004).

6 Roff, D. A. *Life history evolution*. (Sinauer Associates, Inc., 2002).

7 Van Noordwijk, A. J. & de Jong, G. Acquisition and allocation of resources: their influence on variation in life history tactics. *American Naturalist*, 137-142 (1986).

8 Reznick, D., Nunney, L. & Tessier, A. Big houses, big cars, superfleas and the costs of reproduction. *Trends in Ecology & Evolution* **15**, 421-425 (2000).

9 Gould, S. J. & Lewontin, R. C. Spandrels Of San-Marco And The Panglossian Paradigm - A Critique Of The Adaptationist Program. *Proceedings Of The Royal Society Of London Series B-Biological Sciences* **205**, 581-598 (1979).

10 Wright, S. Surfaces of selective value revisited. *The American Naturalist* **131**, 115-123 (1988).

11 Wright, S. 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* (1932).

12 Olson, M. E., Arroyo-Santos, A. & Vergara-Silva, F. A user's guide to metaphors in ecology and evolution. *Trends in ecology & evolution* (2019).

13 Armbruster, W. & Schwaegerle, K. Causes of covariation of phenotypic traits among populations. *Journal of Evolutionary Biology* **9**, 261-276 (1996).

14 Armbruster, W. S. Estimating and testing the shapes of adaptive surfaces: the morphology and pollination of Dalechampia blossoms. *The American Naturalist* **135**, 14-31 (1990).

15 Phillips, P. C. & Arnold, S. J. Visualizing multivariate selection. *Evolution* **43**, 1209-1222 (1989).

16 Dochtermann, N. A. & Dingemanse, N. J. Behavioral syndromes as evolutionary constraints. *Behavioral Ecology* **24**, 806-811 (2013).

17 Kauffman, S. & Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology* **128**, 11-45 (1987).

18 Bulmer, M. The genetic variability of polygenic characters under optimizing selection, mutation and drift. *Genetics Research* **19**, 17-25 (1972).
19 Reeve, J. P. Predicting long-term response to selection. *Genetics Research* **75**, 83-94 (2000).
20 Turelli, M. Phenotypic evolution, constant covariances, and the maintenance of additive variance. *Evolution* **42**, 1342-1347 (1988).
21 Kauffman, S. A. & Weinberger, E. D. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *Journal of theoretical biology* **141**, 211-245 (1989).
22 Gavrilets, S. Evolution and speciation on holey adaptive landscapes. *Trends in ecology & evolution* **12**, 307-312 (1997).
23 Gavrilets, S. *Fitness Landscapes and the Origin of Species*. (Princeton University Press, 2004).
24 Petrie, K. L. *et al.* Destabilizing mutations encode nongenetic variation that drives evolutionary innovation. *Science* **359**, 1542-1545 (2018).
25 Shoval, O. *et al.* Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science* **336**, 1157-1160 (2012).
26 Lande, R. & Arnold, S. J. The measurement of selection on correlated characters. *Evolution* **37**, 1210-1226 (1983).
27 Morrissey, M. B. In search of the best methods for multivariate selection analysis. *Methods Ecol. Evol.* **5**, 1095-1109 (2014).
28 Delcourt, M., Blows, M. W., Aguirre, J. D. & Rundle, H. D. Evolutionary optimum for male sexual traits characterized using the multivariate Robertson–Price Identity. *Proceedings of the National Academy of Sciences* **109**, 10414-10419 (2012).
29 Arnold, S. J. & Wade, M. J. On the measurement of natural and sexual selection: theory. *Evolution* **38**, 709-719 (1984).
30 Shaw, R. G. & Geyer, C. J. Inferring fitness landscapes. *Evolution* **64**, 2510-2520 (2010).
31 Kingsolver, J. G. *et al.* The strength of phenotypic selection in natural populations. *The American Naturalist* **157**, 245-261 (2001).
32 Mousseau, T. A. & Roff, D. A. Natural selection and the heritability of fitness components. *Heredity* **59**, 181-197 (1987).
33 Roff, D. *Modeling evolution: an introduction to numerical methods*. (Oxford University Press, 2010).
34 Gavrilets, S. & Gravner, J. Percolation on the fitness hypercube and the evolution of reproductive isolation. *Journal of theoretical biology* **184**, 51-64 (1997).
35 Gravner, J., Pitman, D. & Gavrilets, S. Percolation on fitness landscapes: effects of correlation, phenotype, and incompatibilities. *Journal of theoretical biology* **248**, 627-645 (2007).
36 Roff, D. A. & Fairbairn, D. J. The evolution of trade-offs: where are we? *Journal Of Evolutionary Biology* **20**, 433-447 (2007).
37 Réale, D. *et al.* Personality and the emergence of the pace-of-life syndrome concept at the population level. *Philos. Trans. R. Soc. B-Biol. Sci.* **365**, 4051-4063, doi:10.1098/rstb.2010.0208 (2010).
38 Houle, D. Genetic covariance of fitness correlates--what genetic correlations are made of and why it matters. *Evolution* **45**, 630-648, doi:10.1111/j.1558-5646.1991.tb04334.x (1991).
39 Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits*. 980 (Sinauer Associates, 1998).
40 Morrissey, M. B. Meta-analysis of magnitudes, differences and variation in evolutionary parameters. *Journal of Evolutionary Biology* **29**, 1882-1904 (2016).

41 Stinchcombe, J. R., Agrawal, A. F., Hohenlohe, P. A., Arnold, S. J. & Blows, M. W. Estimating nonlinear selection gradients using quadratic regression coefficients: double or nothing? *Evolution: International Journal of Organic Evolution* **62**, 2435-2440 (2008).
42 Lewandowski, D., Kurowicka, D. & Joe, H. Generating random correlation matrices based on vines and extended onion method. *Journal of multivariate analysis* **100**, 1989-2001 (2009).
43 Lande, R. Genetic variation and phenotypic evolution during allopatric speciation. *The American Naturalist* **116**, 463-479 (1980).
44 Hansen, T. F. & Houle, D. Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology* **21**, 1201-1219, doi:10.1111/j.1420-9101.2008.01573.x (2008).
45 Kirkpatrick, M. Patterns of quantitative genetic variation in multiple dimensions. *Genetica* **136**, 271-284, doi:10.1007/s10709-008-9302-6 (2009).
46 Nakagawa, S. & Santos, E. S. Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology* **26**, 1253-1274 (2012).
47 Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823* (2014).
48 Zar, J. H. *Biostatistical Analysis*. 4th edn, (Pearson Education, Inc., 1999).
49 Weisberg, M. & Reisman, K. The robust Volterra principle. *Philosophy of science* **75**, 106-131 (2008).

Acknowledgements:

The authors thank A.J. Wilson and B. de Bivort for helpful conversations. This work was supported by US NSF IOS grant 1557951 to N.A.D.

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.

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

This PDF file includes:

Materials and Methods

Simulation Models

Model Construction

We developed an individual variance components model Fig S1; sensu [33] 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

371 converted to 0's and 1's as above. This approach to generating subsequent generations
372 follows the structure of individual variance components models described by Roff [33]. We
373 used this individual variance components approach rather than an agent-based approach
374 as the latter combined with the computational requirements of matching phenotypes to
375 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

376 *Gaussian* (*Wrightian*) *adaptive landscapes*

377 For simulations evaluating Gaussian landscapes, we generated genotypes and phenotypes
378 as above but without the categorical conversion (Table S1). We then generated random
379 landscapes such that the optima (θ) for all traits was set to zero. The topography of the
380 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
381 previous simulation studies examining the evolution of quantitative traits reviewed by [3].
382 This approach corresponds to single peak landscapes in any two dimensions. The forty-five
383 $\omega_{i,j}$ values that fully describe the landscape were generated using the LKJ onion method for
384 constructing random correlation matrices with a pseudo-normal distribution of
385 correlations where the average correlation is 0 (η = 1; Lewandowski et al. 2009; Fig S2).
386 Using the LKJ onion method ensures that the full description of the landscape (ω) is
387 positive semi-definite with feasible partial correlations. We then calculated each
388 individual's fitness based on a Gaussian surface [43]:

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

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 [33]. 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 [44]. Evolutionary potential throughout multivariate space
5. ā: average reduction in evolvability due to trait covariance [44]. 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 $\mathbf{G}$ (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 [42] with $\eta = 1$.
	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 [42] with $\eta = 1$.
	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 [42] with $\eta = 1$.
	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 [45], 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 [46]. 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 [47]. 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 [48]:

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

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

Simulation Models

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

499 landscapes (Figures S6 & S7) and the compression of variation for populations evolving on
500 holey landscapes (Figures 2 & S5).
501

502 Table S3. ANOVA and Tukey HSD results for λ_2/λ_1. Significantly greater genetic variation
503 was maintained across all dimensions when populations evolved on Gaussian landscapes
504 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

505

506

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 – S12). This parameter robustness [49] supports the generality of our modeling. Unfortunately, we were not able to investigate other forms of robustness [49] due to computational limitations.

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

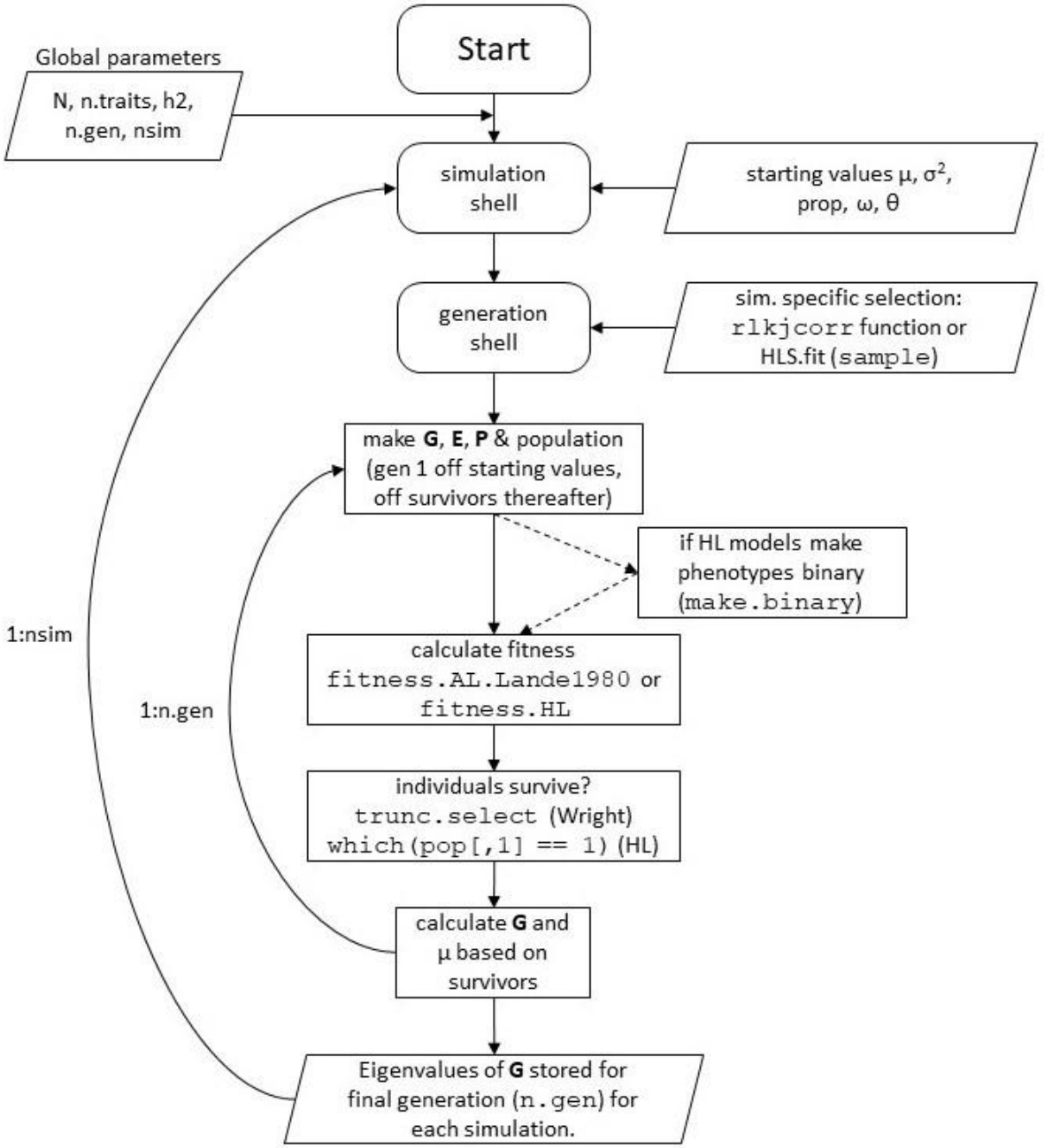
Comparison of Observed Results to Simulation Results

Observed results did not significantly differ from simulated populations that evolved on holey landscapes (Figure 2; Table S10).

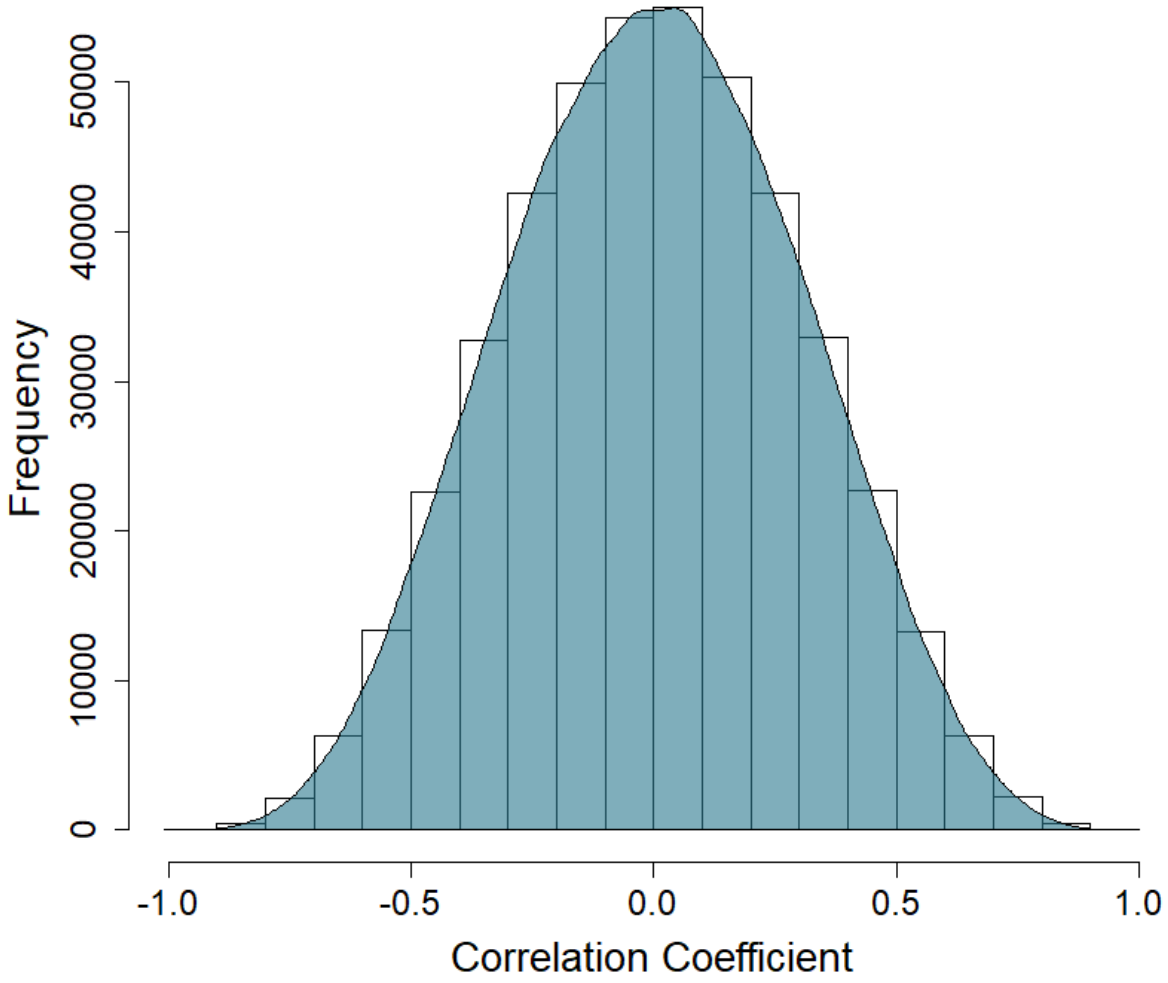
Table S10. t values and associated p values for the comparison of the observed average of λ_2/λ_1 versus the average λ_2/λ_1 for each set of simulations. The observed average and its standard error was taken from a taxonomic mixed-effects model.

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

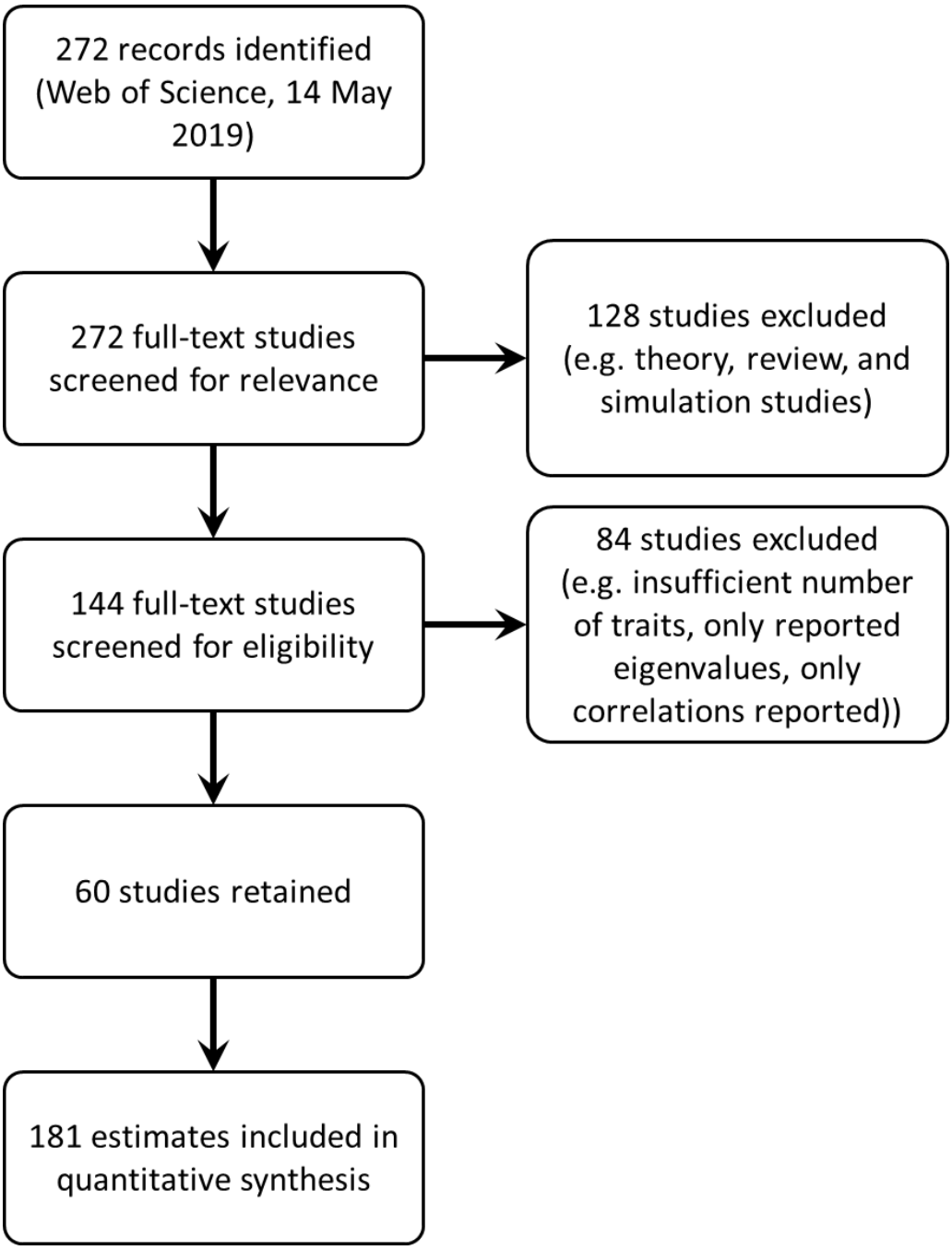
* degrees of freedom = 17.275



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



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



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

566

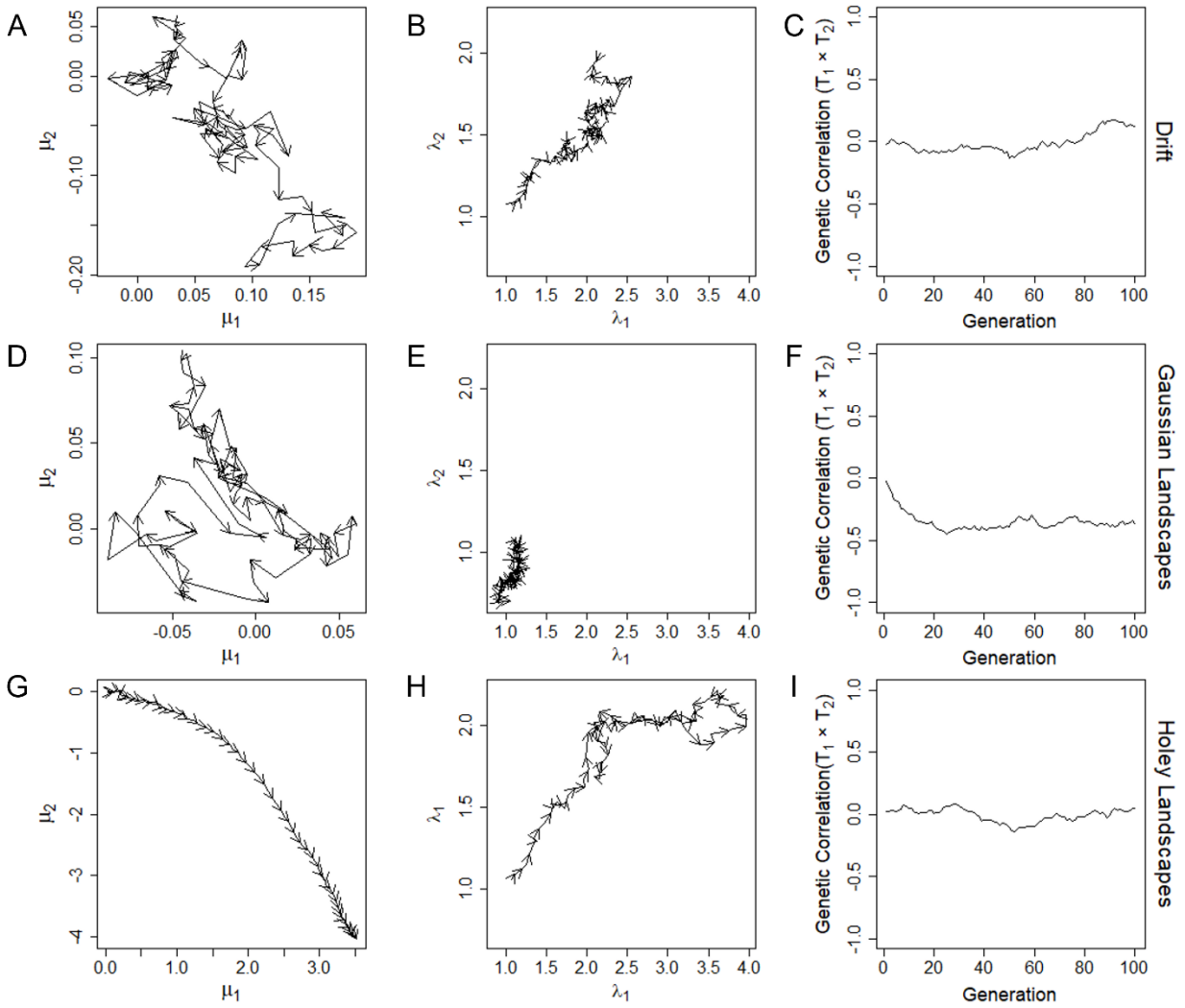
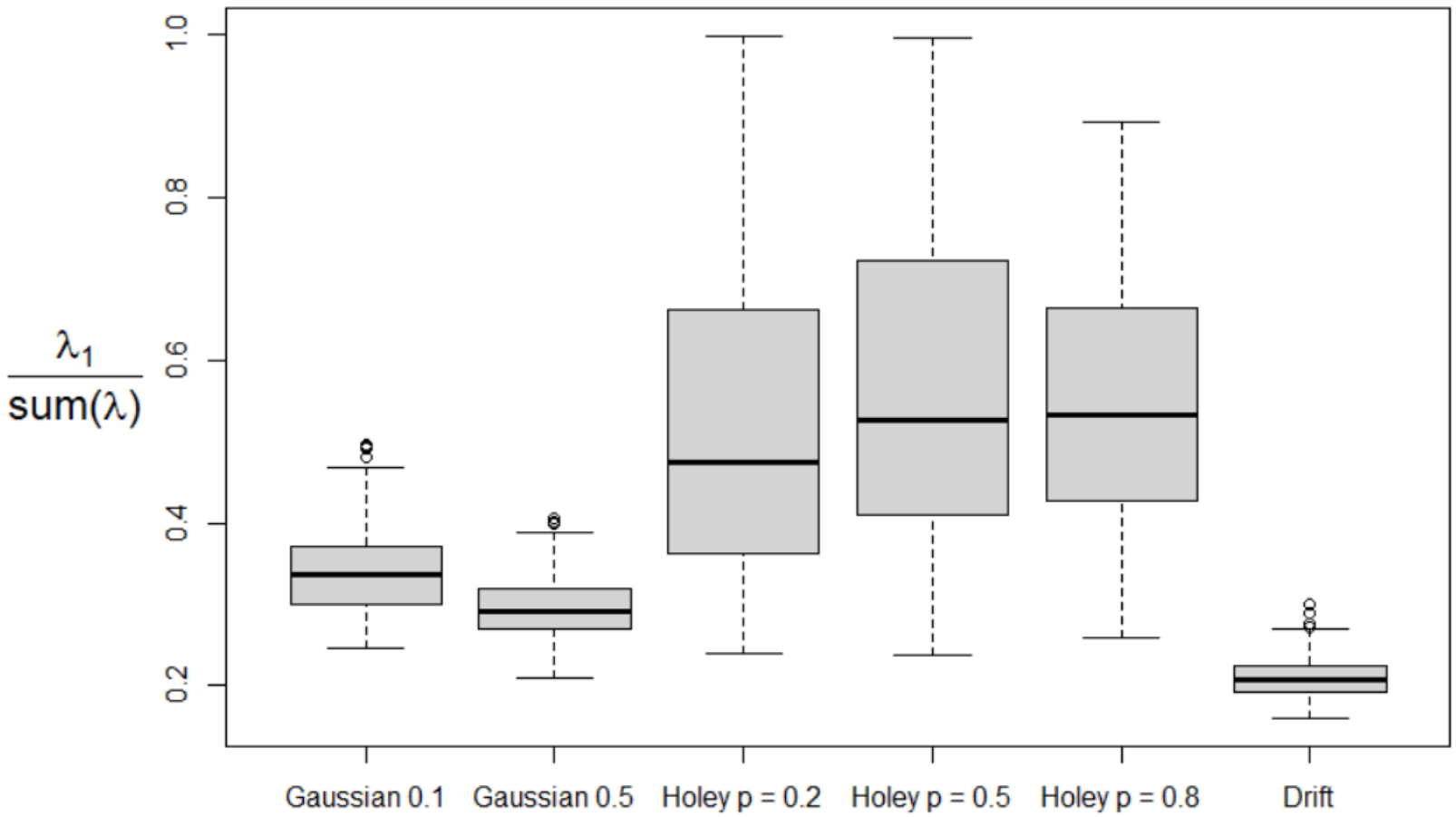
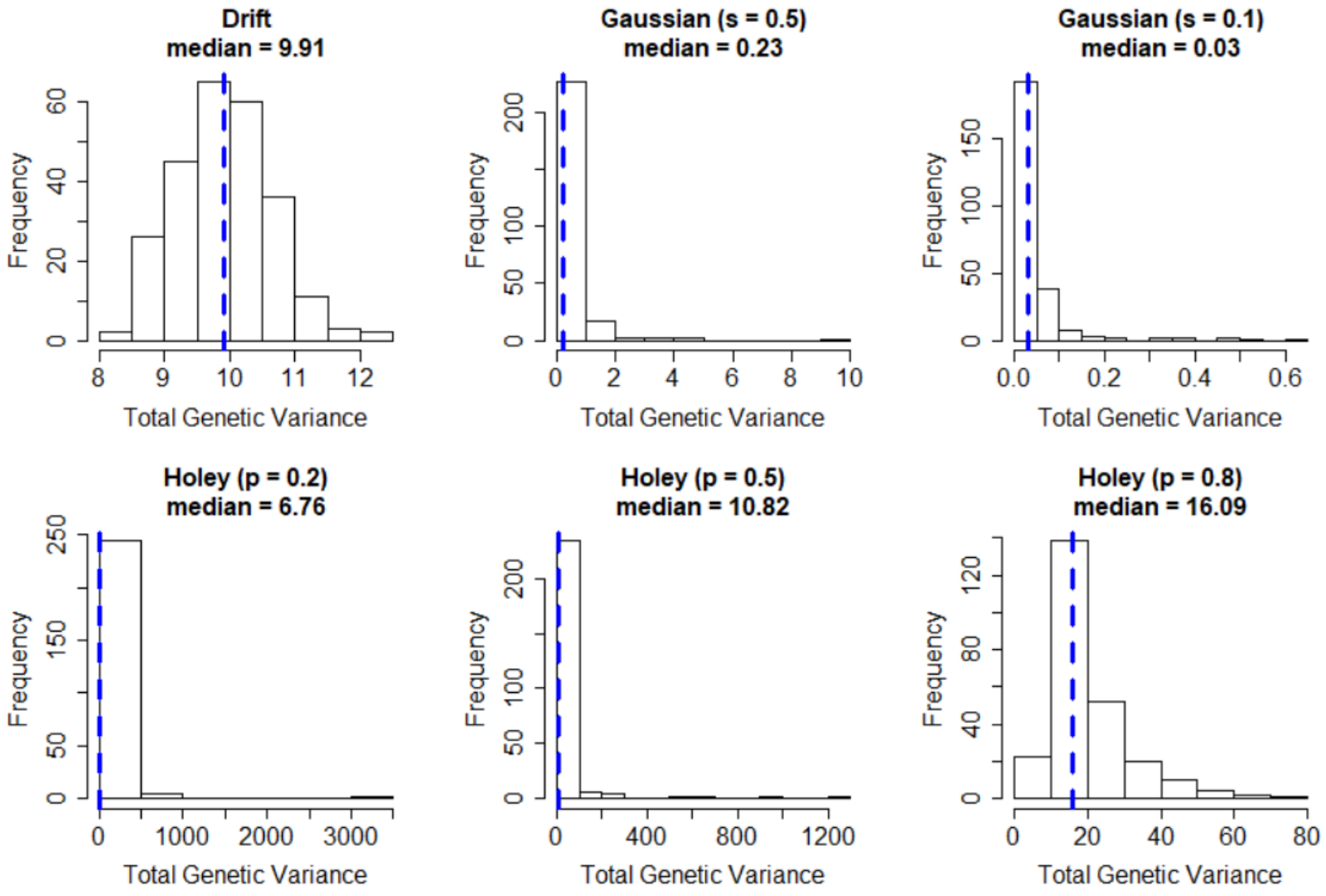


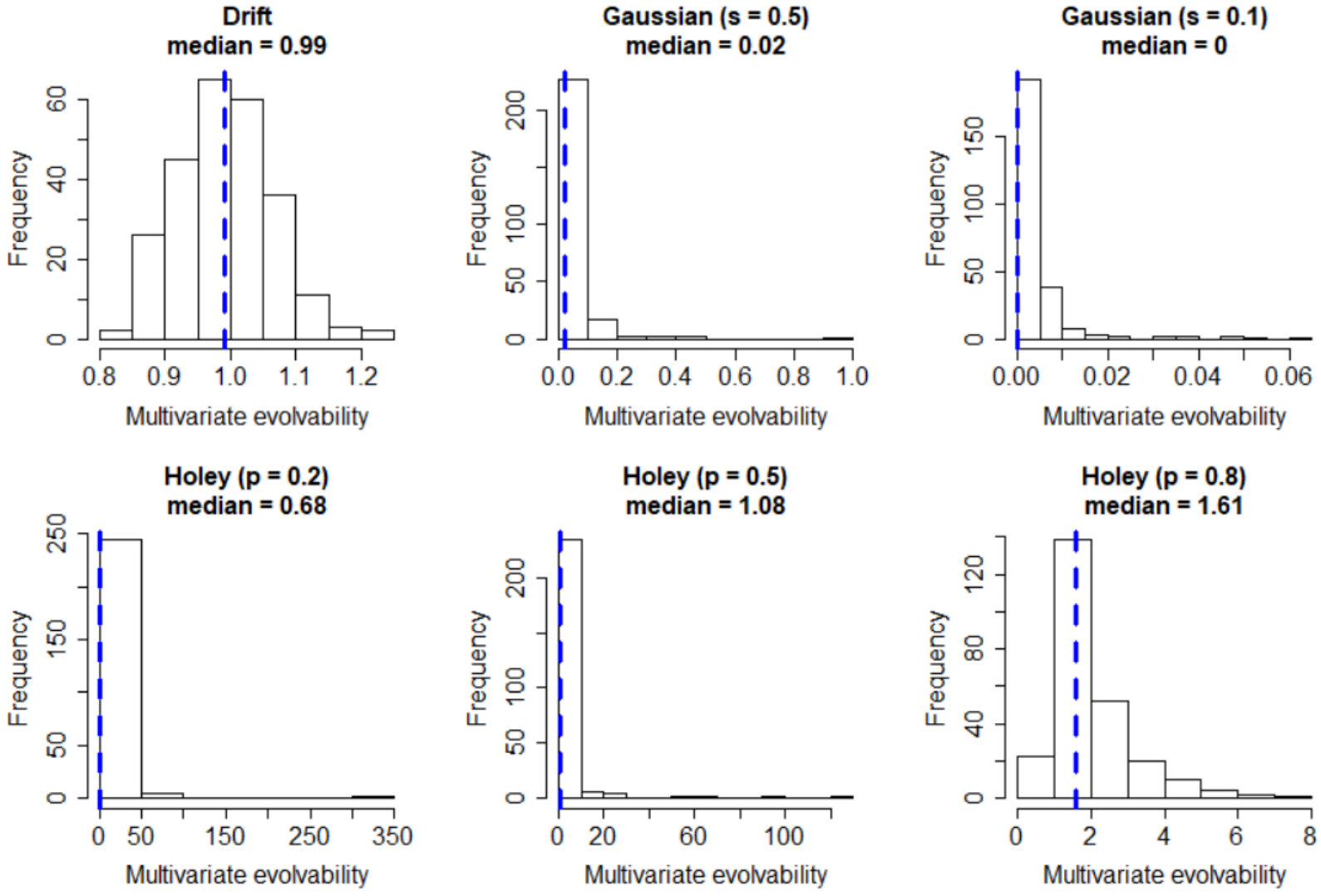
Figure S4. Single population comparisons of population evolution over 100 generations under drift (A – C), on a Gaussian landscape (D – F), and on a holey landscape (G – I). Arrow heads in A, B, D, E, G, and H indicate the direction of evolutionary change at every second generation. Evolutionary change 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.



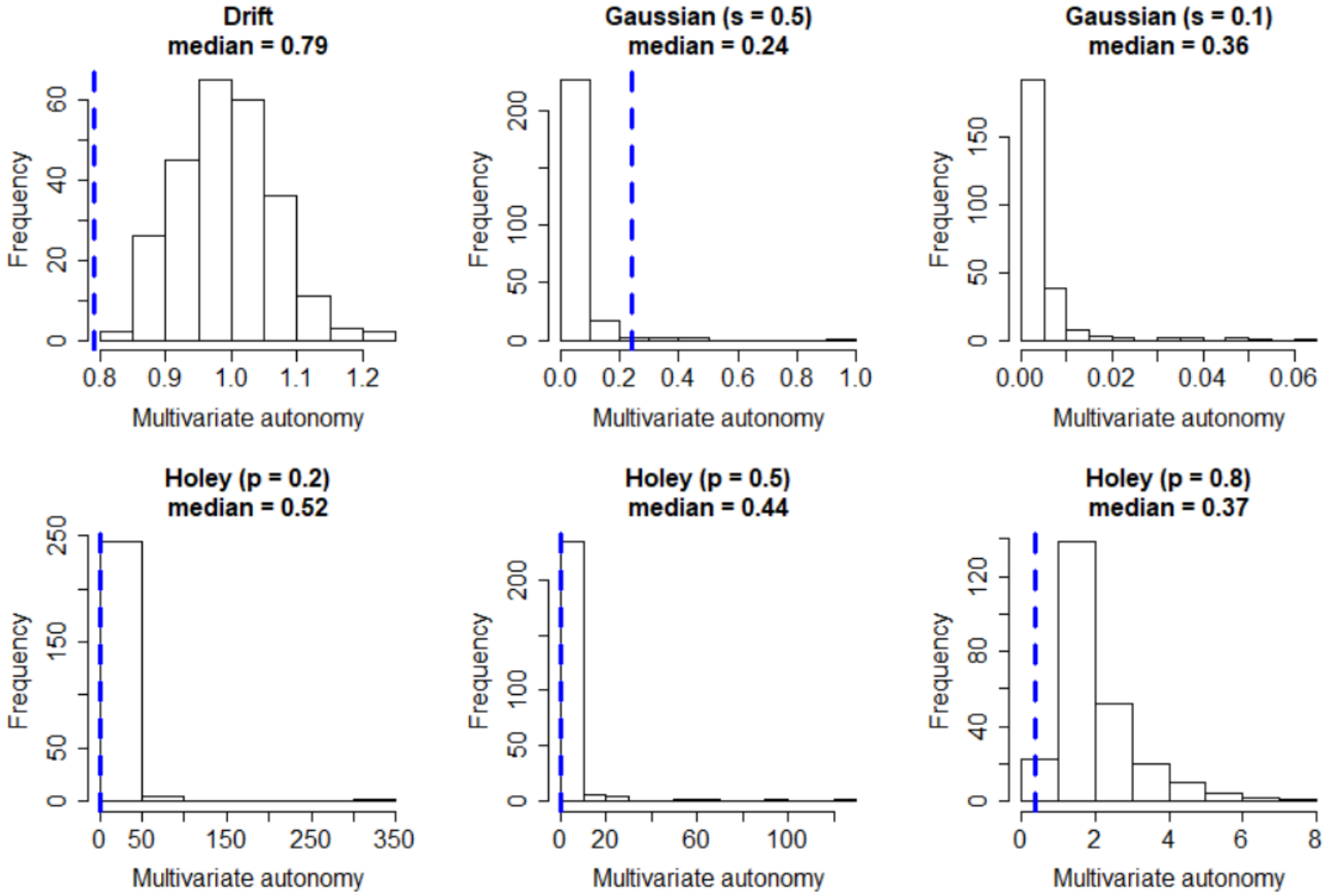
584 Figure S5. Variation was more evenly distributed across dimensions when populations evolved on Gaussian landscapes or due solely to
585 drift. Consequently, less total variation was present in the first dimension (Table S4).



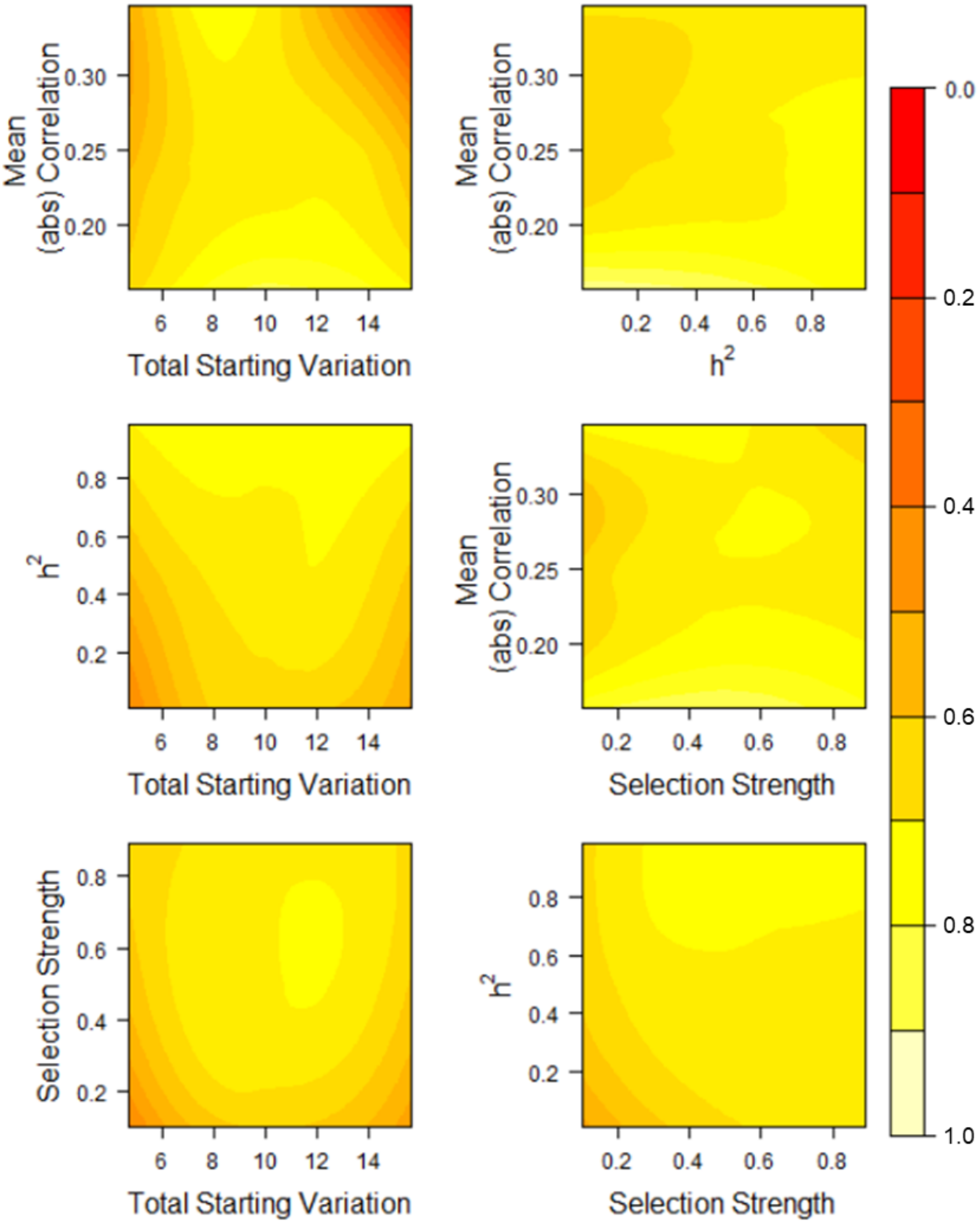
586 Figure S6. The total genetic variation present after 100 generations in each of six modeling conditions and across 250 simulations.
587 Selection on Gaussian surfaces led to a significant reduction in the amount of variation present (Table S5).



588 Figure S7. Multivariate evolvability after 100 generations in each of six modeling conditions and across 250 simulations. Selection on
589 Gaussian surfaces led to a significant reduction in evolvability (Table S6).



590 Figure S8. Multivariate autonomy after 100 generations in each of six modeling conditions and across 250 simulations. Selection on
591 Gaussian surfaces led to a significant reduction in autonomy (Table S7).



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

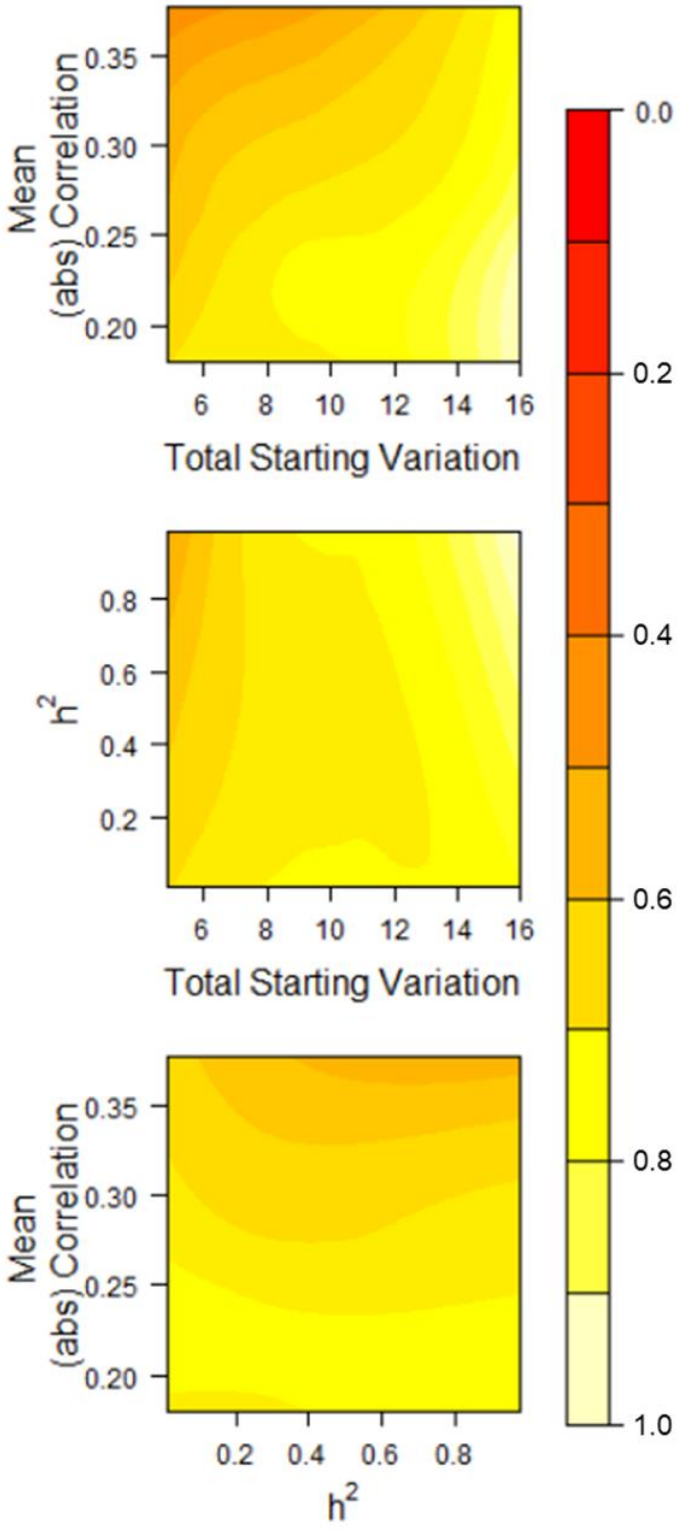
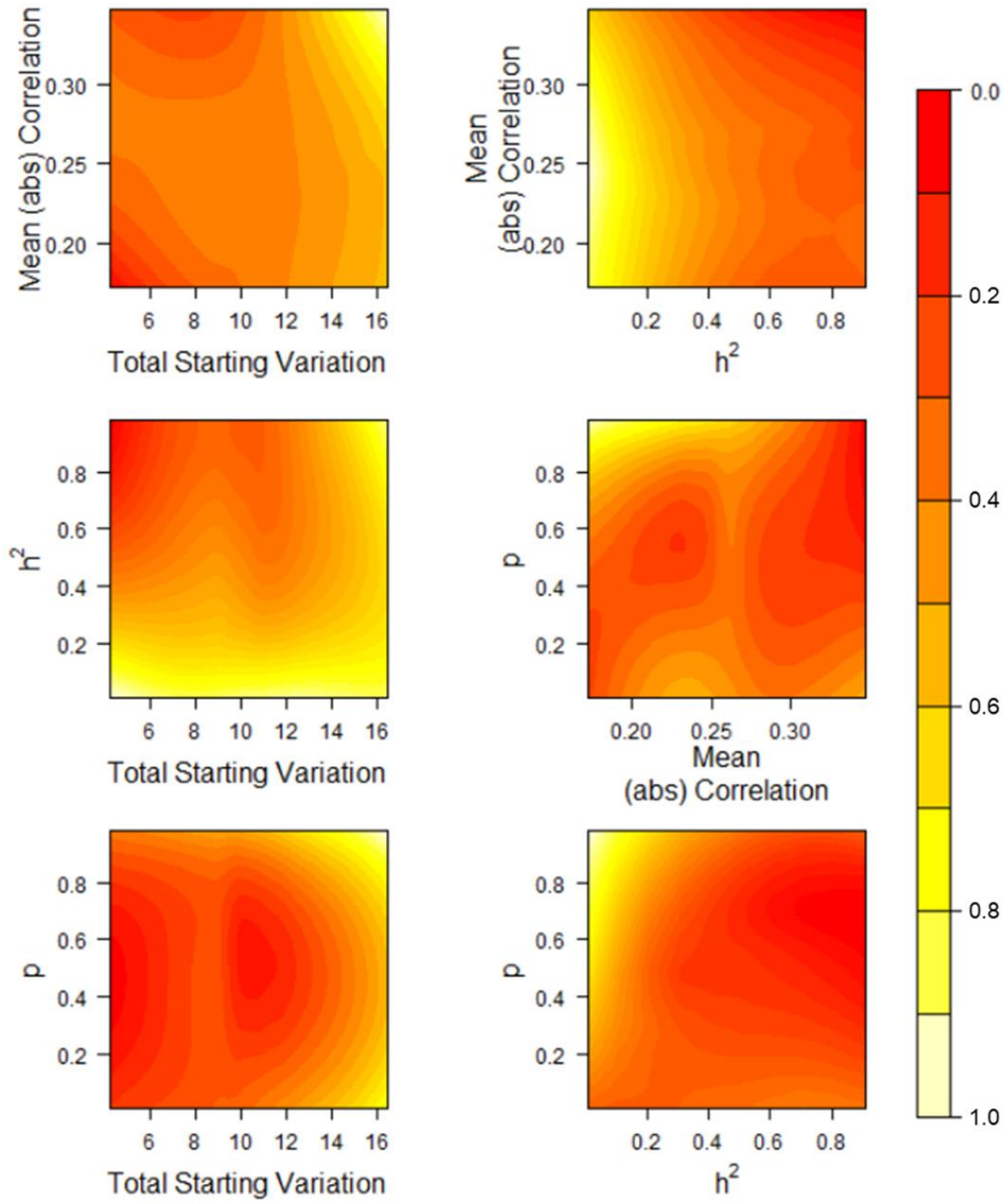


Figure S10. λ_2/λ_1 after evolution due to drift remained high regardless of starting parameters (Table S9).



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