



Robustness revisited: On the neutral evolution of centrality and localization

Yehonatan Sella^{a,b} and Aviv Bergman^{a,b,c,1}

Edited by Marcus Feldman, Stanford University, Stanford, CA; received October 13, 2024; accepted April 23, 2025

This study investigates the intricate interplay among neutral landscape structure, mutation rate, recombination rate, and population dynamics in shaping evolutionary robustness. We provide a comprehensive framework that elucidates how different evolutionary forces interact to influence genotypic robustness and localization within haploid and diploid populations. We demonstrate that in haploid populations, high mutation rates relative to recombination typically drive the population toward regions of increased eigencentrality, a graph-theoretic measure of centrality which is correlated while not identical to mutational robustness. On the other hand, recombination increases the localization of the population to a smaller region of genotypic space, while high values of recombination relative to mutation can introduce shifts in distribution away from regions of high eigencentrality and toward attractors of the recombination dynamics. Diploid dynamics further complicate these interactions, showing reduced alignment with eigencentrality under both high mutation and recombination rates, with the exception of structured diploid landscapes where dynamics are still aligned with increasing eigencentrality. Our findings underscore the nuanced dependencies of evolutionary outcomes on both local and global landscape structures as well as evolutionary parameters.

neutral evolution | fitness landscapes | robustness | eigencentrality | localization

The evolution of genetic robustness has long fascinated evolutionary biologists. Robustness, originally termed by Waddington “canalization,” was introduced to account for “the very general observation ... that the wild type of an organism, that is to say, the form which occurs in Nature under the influence of natural selection, is much less variable in appearance than the majority of the mutant races.” Waddington’s concept of canalization describes how developmental processes are buffered against genetic and environmental perturbations, leading to genetic and environmental robustness by ensuring consistent phenotypic outcomes despite underlying variability (1). Ever since then, numerous studies took upon themselves to corroborate and expand on this phenomenon (2–5). While robustness has often been studied under the lens of adaptive evolution and selective pressures, a growing body of work suggests that neutral evolutionary processes—where most mutations are either neutral or slightly deleterious—can also play a significant role in shaping robustness. Specifically, more recent studies, incorporating development in the context of gene regulatory networks, have shown that robustness can evolve at the absence of stabilizing selection and that selection for developmental stability suffices to induce phenotypic robustness (6–9).

The importance of studying evolution independently of stabilizing selection (which is typically understood as selecting for optimal values of continuous traits) is underscored by Kimura’s neutral theory of evolution (10), which posits that, at least at the biomolecular level, most mutations are either roughly neutral or deleterious— involving a significant impairment of function—and that it is these neutral mutations which are responsible for most gene substitutions in a population, so that evolution at this level proceeds not by a continuous improvement or calibration of the phenotype but by an exploration of the space of neutral mutations.

The framework of neutral evolution can be profitably combined with the powerful metaphor of fitness landscapes introduced by Sewall Wright in the early 1930s and later revisited in the late 1980s (11–13). A fitness landscape refers to a network of potential genotypes for a population, connected via mutation, together with a fitness value assigned to each genotype. The metaphor is suggestive of a population that walks along the landscape of genotypes in a direction of increasing fitness, climbing toward a peak (14). However, in the neutral setting, the idea of climbing toward peaks is de-emphasized given the flat nature of the landscape. Neutral landscapes, characterized by a set of genotypes that all share a similar fitness level, provide a unique perspective on

Significance

Our work advances the theory of neutral evolution, paying particular attention to how the holistic fitness landscape structure shapes the process of evolution and gives rise to emergent evolutionary phenomena. Since neutral evolution does not depend on fitness differences among viable genotypes, its ramifications can be both subtle, as they depend on network-wide properties, and ubiquitous, as they are not tied to context-specific adaptations. Our study provides a theoretical framework that connects the structure of neutral fitness landscapes with the dynamics of mutation and recombination rates, and the distinct behaviors of haploid and diploid populations. We establish general heuristic principles regarding how evolutionary outcomes, such as robustness and localization, are influenced by the interplay of these factors.

Author affiliations: ^aDepartment of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, NY 10461; ^bInstitute for Advanced Study in the Life Sciences, Albert Einstein College of Medicine, Bronx, NY 10461; and ^cSanta Fe Institute, Santa Fe, NM 87501

Author contributions: Y.S. and A.B. designed research; performed research; analyzed data; and wrote the paper. The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2025 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: aviv@einsteinmed.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2421006122/-DCSupplemental>.

Published May 29, 2025.

evolutionary dynamics. Gavrilets (15–17), who studies such landscapes at length, has evocatively termed them “holey landscapes” to emphasize that they can be viewed as flat surfaces riddled with holes of lower fitness, and that the dynamics of evolution are dominated not by climbing higher peaks but by staying away from the holes. In these landscapes, the movement of populations is not solely directed by fitness gradients but rather influenced by the connectivity of genotypic networks and the interplay of evolutionary mechanisms like mutation and recombination. Understanding how these factors interact is crucial for comprehending the evolutionary pathways that lead to robustness.

Previous studies have begun to chart the implications of evolution on neutral landscapes. Key papers (5, 18) have shown that evolution of large populations on neutral landscapes can be well-modeled by assuming for simplicity that all deleterious mutations are fatal, so that the fitness is either 0 or 1, allowing us to restrict our attention to the set of neutral genotypes with fitness 1. Under this model, assuming infinite haploid populations and only mutation (segregation without crossing over), the population will always converge to a distribution described by an eigenvector of the adjacency matrix of the graph of viable genotypes known as the *Perron eigenvector*, as long as this graph is connected and aperiodic. This convergence is independent of the choice of mutation rate or of the initial population distribution. Given the importance of this fact, we include a separate proof in the Supporting information. Moreover, the eigenvalue of this matrix measures the average mutational robustness of the stationary distribution.

The Perron eigenvector of a graph’s adjacency matrix is the unique (up to scaling) eigenvector associated to the maximal eigenvalue of the matrix, so we also term it the *principal eigenvector* of the network. Its uniqueness is guaranteed by Perron–Frobenius (19), which also ensures it is the unique eigenvector (up to scaling) all of whose entries are positive. We can scale the principal eigenvector so that it sums to 1 and can be interpreted as a probability distribution.

The principal eigenvector has both dynamical and a static interpretation. Dynamically, in the context of evolution on neutral landscapes, it is the stationary distribution of a mutation–selection process, regardless of initial distribution or mutation rate. However, since the principal eigenvector is an invariant of the graph it also has static interpretations: it has been interpreted in the field of graph theory as assigning to each node a score measuring a kind of centrality within the network called “eigencentrality” or “prestige score” (20).

That is, for a given vertex v , we can interpret the v th component of the principal eigenvector as a score, where a higher score indicates greater eigencentrality for that vertex. Letting $\lambda \in \mathbb{R}$ be the top eigenvalue of A , and letting \mathbf{ec} denote the principal eigenvector, the eigenvector equation $A\mathbf{ec} = \lambda\mathbf{ec}$ means that for each vertex v in the graph,

$$\lambda \mathbf{ec}_v = \sum_{v' \sim v} \mathbf{ec}_{v'},$$

where $v' \sim v$ represents the set $\{v'\}$ of one-step neighbors of v . Thus eigencentrality rewards a node for having many neighbors, but only if those neighbors themselves are eigencentral —a seemingly recursive condition which nevertheless has a stable interpretation in the form of an eigenvector.

We will take the convention that, when referring to the overall distribution of the principal eigenvector of the adjacency matrix of the graph of viable nodes, we will term it “the

principal eigenvector,” while when referring to an individual node’s entry under this eigenvector, we will term it the node’s “eigencentrality.” The observation that the frequency of a genotype in the stationary distribution of mutation–selection dynamics can be interpreted as a kind of centrality agrees with the intuition that such evolutionary dynamics drive the population to central regions of the landscape.

Centrality is closely related to robustness. In a neutral landscape, one may define the mutational robustness of a genotype as the fraction of its mutational neighbors that are viable, which is proportional to its degree in the graph of viable genotypes. While degree- and eigen- centrality can markedly diverge depending on the structure of the graph, eigencentral nodes tend to have large degree, thus explaining the evolution of robustness on neutral landscapes under mutational dynamics (i.e., the operation of mutation in the absence of recombination). We further explore the interplay between different measures of centrality on different families of landscapes in the present paper.

A notable property of principal eigenvectors on graphs that are sufficiently irregular, to be discussed further in the main body of the paper, is *localization*. Namely, for these networks, high eigencentrality scores are limited to a small set of nodes which are close together (21–23), while exponentially decaying away from this localized region. The precise degree of localization depends on the structure of the network. This property is noteworthy since it suggests that, even under the paradigm of neutral evolution, the course of evolution is not entirely diffuse, but may be steered to a particular localized region of genotypic space.

While the stationary distribution under mutation-only dynamics corresponds to the principal eigenvector of the graph, once recombination is introduced, the dynamics become more complicated, and the stationary distribution depends on the balance between the rates of mutation and recombination, as well as on the specific initial distribution. Recent papers (24–26) analyzed the mutation–recombination dynamics on holey landscapes, concluding that recombination tends to further increase population robustness. These papers also suggest the importance of a recombination weight in determining the most frequent genotype in the stationary distribution, when recombination dominates over mutation.

In this paper, we further elucidate and complicate these observations. Specifically, we show that while high mutation relative to recombination tends to drive populations toward regions of high eigencentrality, correlating with robustness to genetic perturbations, increased recombination can either enhance or disrupt this robustness depending on the underlying landscape structure and population dynamics.

We place particular emphasis on studying localization along with robustness and its more refined counterpart, eigencentrality. We show that increased recombination consistently leads to increased localization of the stationary distribution but that the impact of recombination on robustness is more complicated. We show that, for certain families of landscapes, there are distinct phase shifts associated with the balance between mutation and recombination: if mutation is large enough compared to recombination, then the stationary distribution resembles that of the pure mutation–selection case, only more localized around the high-eigencentrality region, whereas if mutation falls below a certain threshold, the stationary distribution often shifts to a different part of the graph, which partly correlates with eigencentrality and robustness, but much less reliably so. We separate the nature of increased population robustness into two distinct phenomena: increased localization and increased robustness (or eigencentrality) of the mode. While recombination

reliably increases the former, mutation more reliably increases the latter.

We also systematically study the influence of the initial population on the stationary distribution. In addition, we study the dynamics under both haploid and diploid models, and show that the evolutionary dynamics substantially differ under the two models, with diploid populations' stationary distributions showing no clear relationship to the principal eigenvector of the mutational structure of the landscape even when mutation is high relative to recombination, with the exception of specially structured fitness landscapes as discussed further below.

Evolution on Abstract Fitness Landscapes

We describe the process of evolution on fitness landscapes abstractly, in terms of both mutation and recombination. We define a fitness landscape to be the data of a graph $G = (V, E)$ together with a fitness function $w : V \rightarrow \mathbb{R}^{\geq 0}$ valued in nonnegative numbers. For simplicity, we assume the graph is undirected and unweighted. We will further make the simplifying assumption that the graph G is regular, so that each vertex has the same degree N for some $N \geq 1$.

Each vertex $v \in V$ stands for an entire genotype, a point in genotypic space. There is an edge $v_1 \sim v_2$ between two vertices $v_1, v_2 \in V$ if a single mutation would transform v_1 to v_2 and vice versa. The fitness function $w(v)$ describes the probability of an individual with genotype v surviving to reproduce. We assume uniform fecundity for all genotypes.

The standard biological example to keep in mind is the graph of all sequences of length n over a given alphabet \mathcal{A} , where two sequences are neighbors if there is a single edit that transforms one to the other. In the case that \mathcal{A} is an alphabet of size two, this graph is the n -dimensional hypercube, denoted C_n . However, for the sake of abstractness and generality, we do not restrict ourselves to this biological case. Notice that if \mathcal{A} has size k , then the degree is $N = n \cdot (k - 1)$. [Please note that for simplicity, all simulations were performed using a diallelic case, $\mathcal{A} \in \{0, 1\}$.]

Define a node v to be *viable* if $w(v) > 0$ and *inviable* if $w(v) = 0$. Since inviable nodes are thrown out by evolution, it is useful to consider the restricted graph $G' = (V', E')$, where $V' \subset V$ consists only of the viable nodes, and E' is the restriction of E to V' . We will make the assumption that the graph G' is connected. If it had multiple connected components, one can study them separately since they never interact, at least assuming segregation without crossing over and only one-step mutations.

Under this simple abstract framework, we can model evolution as a stochastic (Markovian) process on this graph G' . We begin with the case of asexual reproduction, and later treat the more involved case of sexual reproduction. We also make the simplifying assumption of a very large population, so that we can ignore stochastic effects.

Let \mathbf{x}_t denote the expected population distribution at time t , which we think of as a vector in $\mathbb{R}^{|V'|}$; that is $x_{t,v}$ is the expected number of individuals with genotype v at time t . Now, let \mathbf{p}_t denote the probability distribution of the population at time t , which is simply a normalized version of \mathbf{x}_t , namely, we have $p_{t,v} = \frac{x_{t,v}}{\sum_{v \in V'} x_{t,v}}$.

Mutational Dynamics. We first model evolution with the following asexual model, which depends on the specification of a mutation rate $0 < \mu < 1$. Each individual with genome v has probability $w(v)$ of surviving and reproducing. If it survives, there is a probability $(1 - \mu)$ that the offspring's genotype is still

v , and a probability of μ that it has incurred a mutation. In this event, the new genotype is randomly, uniformly selected from among the neighbors of v in the graph G . The mutant is deleted if it is inviable.

The evolutionary process is specified by recurrence equations for \mathbf{x}_{t+1} in terms of \mathbf{x}_t , as follows:

$$\mathbf{x}_{t+1,v} = (1 - \mu)w(v)x_{t,v} + \mu \sum_{v' \sim v} \frac{1}{N}w(v')x_{t,v'}. \quad [1]$$

We can rewrite this as a matrix equation. Let A be the $|V'| \times |V'|$ matrix recording the edges E' between *viable* nodes. That is, we let $A_{vu} = 1$ if v, u are neighbors which are both viable, and $A_{vu} = 0$ otherwise. Let W be the $|V'| \times |V'|$ diagonal matrix with $W_{vv} = w(v)$. Then we have

$$\mathbf{x}_{t+1} = (1 - \mu)W\mathbf{x}_t + \frac{\mu}{N}AW\mathbf{x}_t = BW\mathbf{x}_t \quad [2]$$

for $B = (1 - \mu)I + \frac{\mu}{N}A$.

Note that the vectors \mathbf{x}_{t+1} will decay exponentially since each individual produces on average less than one offspring. However, we are implicitly assuming a population whose size approaches infinity and are therefore more interested in the evolution of the normalized vectors \mathbf{p}_t describing the probability distribution of the population among the viable nodes.

Reproductive Dynamics. Reproductive dynamics are specified by a recombination, crossover, parameter c , and can apply in either haploid setting, where genotypes are modeled as sequences of length n over an alphabet \mathcal{A} , or diploid setting where genotypes are modeled as pairs of sequences of such sequences. The parameter s describes the probability of a crossover event; if it occurs, then a random site along the sequence is picked where the crossover occurs. A more precise mathematical description is given in *SI Appendix*.

Holey Landscapes. We will primarily focus on the case of “holey” landscapes. Namely, those in which all viable individuals have the same fitness (for simplicity we set it at 1, though that does not affect our results), and so the only fitness distinction is between viable and nonviable genotypes.

In the case of holey landscapes, the matrix W , which records the fitness of viable genotypes, is the identity. Thus the mutation process becomes

$$\mathbf{x}_{t+1} = B\mathbf{x}_t$$

And together with recombination, it becomes

$$\mathbf{x}_{t+1} = B \cdot Q(\mathbf{x}_t)$$

where $Q : \mathbb{R}^{|V'|} \rightarrow \mathbb{R}^{|V'|}$ is defined as a symmetric quadratic function (*SI Appendix*).

van Nimwegen et al's result (18), mentioned in the introduction, shows that mutational dynamics (in the absence of recombination) converge to the leading eigenvector of the neutral landscape's adjacency matrix, follows from the fact that the eigenvectors of B are the same as those of A , together with the Perron–Frobenius theorem. See *SI Appendix* for a proof [this result is also shown in chapter 7 of van Nimwegen thesis (27)].

Measures of Centrality, Robustness, and Localization

Next we study how robustness relates to eigencentrality. On holey landscapes, we equate viability with phenotype. The robustness

of a genotype is then the proportion of its mutational neighbors which are viable. As a vector equation, the robustness vector $\mathbf{r} \in \mathbb{R}^{|V'|}$, where \mathbf{r}_v is the robustness of $v \in V'$, is given by $\mathbf{r} = \frac{1}{N} A \mathbf{1}$. Here, $\mathbf{1}$ is the vector of ones.

We can generalize this notion of robustness to define the k th-order robustness of a node $v \in V'$ as $\mathbf{r}_k = (\frac{1}{N^k} A^k \mathbf{1})$. The number $r_k(v)$ measures the probability, given a path of length k in the graph G starting at v (representing a sequence of mutations), that all the nodes along the path are viable. The vector \mathbf{r}_k records these probabilities across all v . Of course, \mathbf{r}_1 equals the above measure of robustness \mathbf{r} , and the \mathbf{r}_k , after renormalizing, converge to eigencentrality \mathbf{ec} .

We can integrate the different k th-order robustnesses in one measure by computing the average length of a random walk in G before reaching an inviable node. That is, for each $v \in V'$, consider the random walk which starts at v and where each time, the next node is chosen uniformly at random from among the neighbors of v in V . The random walk is terminated as soon as an inviable node is visited. Let L_v be the random variable measuring the length of this random walk, that is, the length of the path taken in V' , not including the terminal inviable node. Then define $l(v) = E(L_v)$, the expected length of this random walk. A low value of $l(v)$ would indicate that v is precariously located on the edge of the region of viability, while a high value of $l(v)$ would indicate v is securely within the region of viability. Let \mathbf{l} denote the vector recording $l(v)$ across all $v \in V'$.

The measure $l(v)$ is related to the k th-order robustnesses by the formula $l(v) = \sum_{k=0}^{\infty} r_k(v)$, where we define $r_0(v) = 1$. Indeed, this follows from the fact that $r_k(v) = P(L_v \geq k)$. The vector \mathbf{l} can then be calculated by $\mathbf{l} = \sum_{k=0}^{\infty} (\frac{1}{N} A)^k \mathbf{1} = (I - \frac{1}{N} A)^{-1} \mathbf{1}$. Note that the matrix $I - \frac{1}{N} A$ is invertible as long as N is not an eigenvalue of A , which it is not unless all states are viable. Also note that the $(I - \frac{1}{N} A)^{-1}$ is the *fundamental matrix* in absorbing Markov chains.

We can obtain expressions for the average values of \mathbf{r}_k and \mathbf{l} across the stationary population distribution, under mutational dynamics. Let $\tilde{\mathbf{p}}$ be the stationary distribution of evolution under mutation-selection dynamics, which we know is the principal eigenvector, normalized to sum to 1. van Nimwegen et. al. have shown that the average population robustness, that is, $\mathbf{r} \cdot \tilde{\mathbf{p}}$, is equal to $\frac{\rho(A)}{N}$, where $\rho(A)$ is the spectral radius of A , which is the top eigenvalue of A . More generally, it is a simple derivation that the average population k th-order robustness, given by $\mathbf{r}_k \cdot \tilde{\mathbf{p}}$, equals $(\frac{\rho(A)}{N})^k$ and the average value of \mathbf{l} is $\frac{1}{1 - \frac{\rho(A)}{N}}$ ([SI Appendix](#))

Next we argue that evolution under mutation tends to favor robust nodes, whether this robustness is measured by $r(v)$, $r_k(v)$ or $l(v)$. We can express this fact in two different, but equivalent ways:

First, that the average population robustness of the stationary distribution will always be greater than the population robustness of the uniform distribution, representing a “random” distribution; second, that the correlation between the robustness of a genotype and its frequency in the stationary distribution is positive [see also theorem 5 of Altenberg (4)]. See [SI Appendix](#) for proofs of these claims.

Localization. A notable property of eigencentrality in certain landscapes (or graphs) is localization. Namely, high eigencentrality scores are limited to a small set of nodes which are close together (21–23).

The precise degree of localization depends on the structure of the network. To measure localization in a given population distribution \mathbf{p} , One can consider the use of two measures:

First, the Simpson diversity index $SD(\mathbf{p}) = \sum_{v \in V'} (\mathbf{p}_v)^2 = \mathbf{p} \cdot \mathbf{p}$, which measures the probability that two randomly chosen individuals in the population share the same genotype. We have $\frac{1}{|V'|} \leq SD \leq 1$, where a larger value of SD indicates greater localization, the minimum being attained at a uniform distribution, and the maximum of $SD = 1$ attained at an isogenic distribution.

Second, one can measure the expected distance (shortest-path distance in the graph G , which translates to Hamming distance in the biological setting) between two randomly chosen individuals, $ED(\mathbf{p}) = \sum_{v_1, v_2 \in V'} \mathbf{p}_{v_1} \mathbf{p}_{v_2} d(v_1, v_2) = \mathbf{p}^T D \mathbf{p}$, where D is the $|V'| \times |V'|$ distance matrix. A lower value of ED indicates increased localization, with $ED = 0$ for isogenic distributions, and where ED is bounded above by the diameter of the graph.

Normalized versions of these localization measures, in which we normalize by the localization of the uniform distribution, can also be used: $SD_n(\mathbf{p}) = \frac{SD(\mathbf{p})}{SD(\mathbf{u})}$ and $ED_n(\mathbf{p}) = \frac{ED(\mathbf{p})}{ED(\mathbf{u})}$, where \mathbf{u} denotes the uniform distribution on V' .

Below we investigate the above measures of robustness and localization for stationary distributions on specific families of holey landscapes (to illustrate our results, for the localization measure, we employed the Simpson diversity measure).

Families of Holey Landscapes

We present a few examples of families of holey landscapes, which we will study throughout this paper.

Example 1: Geometric examples. To provide geometric intuition, we first provide examples of holey landscapes on geometric grids. Though these are not directly applicable to networks of genomes, they illustrate the concepts under discussion and may be easily visualizable for one- and two-dimensional grids. Moreover, these graphs can in principle be embedded as subgraphs of hypercubes and thus attain a genetic context.

Example 1.1: Box in \mathbb{Z}^d Consider the d -dimensional grid \mathbb{Z}^d , with the edge relation given by adjacency on the grid. Let n_1, \dots, n_d be positive numbers, and say that a vertex (a_1, \dots, a_d) is viable if $1 \leq a_i \leq n_i$ for all i . (For a visual example, refer to [SI Appendix, Fig. S11](#).)

Example 1.2: Russian Roulette on a grid

The Russian Roulette holey landscape, so-named by Gavrilets (15–17, 28), is a grid as in Example 1.1, in which nodes have a certain independent probability P of being viable. This process of removing points at random from a network is also known as *site percolation* in percolation theory.

Percolation theory studies the connected components of the resulting graph in the infinite setting of \mathbb{Z}^d as opposed to finite grids. In particular, there is a critical threshold p_c such that if $P > p_c$, the resulting graph has an infinite connected component, whereas if $P < p_c$, then the resulting graph has no infinite connected components (almost surely). For example, for \mathbb{Z}^2 , the critical probability is known to be approximately $p_c \approx 0.593$ (29). We employ values of P larger than p_c in order to obtain large connected components, and we will focus on the largest connected component.

Example 2: Genetic examples (hypercube subgraphs) While geometric examples, in particular in two or three dimensions, can help by providing visualizable examples, this geometric intuition

can be misleading, as Wright himself observed about his concept of the fitness landscape (11, 12). It is therefore best to check any intuition gained by the geometric examples against landscapes that are grounded in the genetic context, namely subgraphs of the hypercube $\{0, 1\}^n$.

Example 2.1 : Mesa landscapes. A Mesa landscape is specified by a dimension n , a choice of central node $v \in \{0, 1\}^n$ and a threshold number a , where $0 < a < n$. A node $v' \in \{0, 1\}^n$ is then viable if its Hamming distance from v is at most a . Biologically, viability is maintained as long as the genotype lies within a certain threshold of an “optimal” genotype. By symmetry, we can always assume for simplicity that v is the string $(0, 0, \dots, 0)$, so that a node is viable if the number of ones in the corresponding binary is bounded above by a .

Mesa landscapes are in some sense analogous to the boxes of Example 1.1, in that there is a clear central node, and a clear boundary lying a given distance from that central node.

Example 2.2: Genetic Russian Roulette In the genetic Russian Roulette, each node in the hypercube $\{0, 1\}^n$ has independent probability p of being viable. We then take the largest connected component.

Eigencentrality, Robustness, and Localization for Mutation-Only Dynamics

We now investigate the properties of the stationary distribution of mutational dynamics—that is, the principal eigenvector of the adjacency matrix of viable nodes—on the above families of landscapes. In this section, we will denote this distribution by $\tilde{\mathbf{p}}$.

First, for the sake of visualization, we plot the principal eigenvector for representative landscapes of each of the families described above; see Fig. 1.

In the geometric box as well as the Mesa landscape, there is a clearly defined center, and the principal eigenvector peaks at that center and diminishes away from the center, approaching

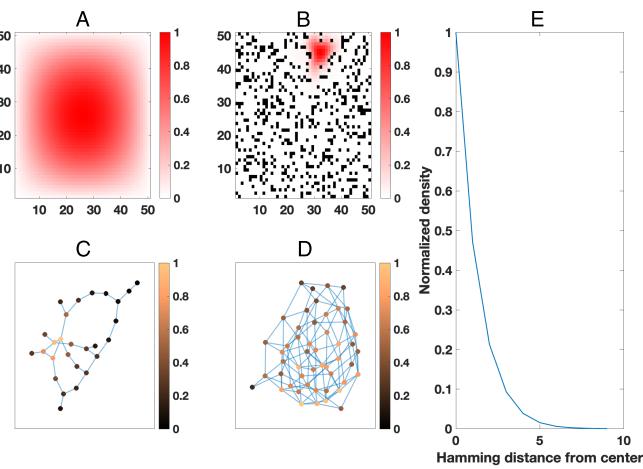


Fig. 1. Plots of principal eigenvector distributions for representative landscapes of different types: (A) a two-dimensional geometric grid with boundary interpreted as inviable, and mutations being neighbors on the grid ($n_1 = n_2 = 50$). (B) A two-dimension geometric landscape with holes ($n_1 = n_2 = 50, P = 0.8$). (C) A genetic landscape with low probability of viability ($n = 7, P = 0.25$). (D) A genetic landscape with high probability of viability ($n = 6, P = 0.75$). (E) A “Mesa” landscape ($n = 100, a = 10$). Note that landscapes (A) and (B) serve only for geometric intuition, not being biologically motivated, while (C) and (D) are more biologically relevant. The phenomenon of localization is illustrated in examples (B) and (C), where we see that the high-eigencentrality portion of genotypic space is limited to a small area, rather than being diffuse. The other examples are not characteristic of localization.

0 toward the boundary. In the geometric case of an $n_1 \times \dots \times n_d$ box, one can determine analytically that the stationary distribution is (up to rescaling)

$$\tilde{\mathbf{p}}((a_1, \dots, a_m)) = \prod_{i=1}^d \sin\left(\frac{a_i}{n_i + 1}\pi\right)$$

(see *SI Appendix* for a proof).

On the other hand, for the geometric and genetic Russian Roulette families, there is no clear center to the landscape. In the geometric Russian Roulette landscapes in two dimensions in which p is not too close to 1, the principal eigenvector is localized to a small region of the space, a pocket which is a relatively high-viability region. Similar localization is observed for the genetic Russian Roulette if P is small. However, for larger P the principal eigenvector distribution is more diffuse.

To support these results systematically, we calculate the localization measures SD_n for these families of landscapes over varying parameters. The following image shows normalized Simpson diversity (SD_n) for different families of landscapes, under varying parameters.

For the stochastic families, we generated 50 landscapes for each parameter choice, and took the mean SD_n across these landscapes. In the geometric Russian roulette, we see that a lower probability of viability increases localization, as does increased grid size. Similarly in the genetic Russian roulette, localization tends to increase with lower probability of viability and higher dimension. The Mesa landscape serves as a control. We do not consider the Mesa landscapes as exhibiting localization, since the population is diffused through the entire landscape as there are no internal barriers to diffusion. Indeed, while SD_n for the Mesa landscapes increases with dimension, it is still only around twice the base-level SD of a uniform distribution, whereas the Russian Roulette landscapes attain much higher values of SD_n for certain parameters.

As a control for the geometric landscapes, one can prove using an integral approximation (*SI Appendix*) that SD_n of an interval approaches $\frac{\pi^2}{8} \approx 1.2$ as the size of the interval approaches infinity, and SD_n of a square grid approaches $(\frac{\pi^2}{8})^2 \approx 1.5$ as the size of the grid approaches infinity.

The phenomenon of eigenvector localization in both the adjacency matrix and the Laplacian of a graph has been extensively studied. Hata et al. (30) have shown that eigenvectors of Laplacians of random graphs tend to be localized and that this localization is greater for graphs with heterogeneous degree distributions. Notably, the principal eigenvector often concentrates among clusters of nodes with similarly high degrees. Arnold et al. (22) theoretically studied localization of eigenfunctions in the continuous context, with similar results. Indeed, localization can be understood to emerge when a cluster of nodes with high degrees is surrounded by nodes with significantly lower degrees, which act as a barrier to diffusion. These observations are consistent with our observations on the Russian roulette landscapes, since a lower probability p of viability entails a higher variance in the degree distribution relative to the mean degree (Fig. 2).

Next we consider robustness. As we have noted, average population robustness always increases upon evolution if the initial population distribution is uniform. We see below (Fig. 3) that the increase is more pronounced, in both geometric and genetic versions of the Russian roulette landscapes, when the probability of viability is low, that is, when localization is present. This agrees with the intuition that a localized distribution is

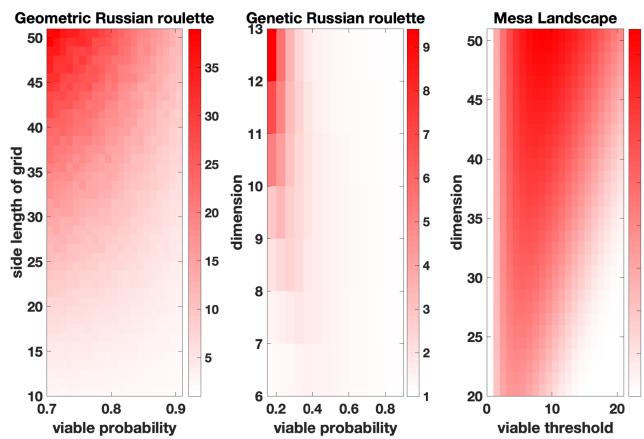


Fig. 2. A systematic “heatmap” of the localization measure for different families of landscapes. This shows the trend that a lower proportion of viable genotypes, as well as higher dimension, tends to increase localization, and that there seems to be a sharp transition from low-localization to high-localization regimes.

concentrated on a pocket of high-degree (i.e., high-robustness) genotypes, thus buffering against the low average robustness of the landscape as a whole. On the other hand, nonlocalized stationary distributions, while correlated with robustness, are diffused throughout the entire landscape. For another perspective on this observation, we plot eigencentrality scores (normalized such that the maximum is 1) of genotypes in a landscape against the number of viable neighbors, aggregated over 50 landscapes for each landscape type and visualized as box plots. We note that in the case of low probability of viability, exhibiting localization, the correlation between degree- and eigencentrality is less predictable, with many outliers in which lower-degree genotypes have high eigencentrality and vice versa. In the case of high probability of viability, the correlation, though noisy, is much more straightforward.

Eigencentrality, Robustness, and Localization Under Mutation and Recombination

We now explore the general differences between the dynamics of evolution under mutation versus recombination. We begin by studying the haploid case.

Haploid Recombination Dynamics. While asexual evolution on neutral landscapes reaches the same stationary distribution regardless of the mutation rate and the initial distribution, the stationary distribution reached by sexual reproduction with recombination depends on the rates of mutation and recombination as well as on the initial population distribution. In this section, we investigate these dependencies and the relationship of the resulting stationary distribution to the principal eigenvector distribution (31).

For the purposes of this section, we focus on genetic Russian Roulette landscapes, in both the low- P regime, where we see localization of the principal eigenvector, and in the high- P regime, where we do not. We collect data from 50 different randomly generated genetic Russian Roulette landscapes in each regime, and for each of these landscapes sweep the mutation rates between 0 and 0.5 (we sample more closely the smaller mutation rates, but include large ones for a full picture), while keeping recombination rate at 0.1.

Our choice of $c = 0.1$ is arbitrary—our main interest is in exploring the impact of the relative rates of mutation and recombination, giving rise to distinct dynamics in regimes where mutation dominates over recombination and vice versa. In the Supporting Information, we show that different values of the recombination rate still lead to qualitatively similar dynamics (*SI Appendix*, Figs. S3 and S4 for $c = 0.05$ and *SI Appendix*, Figs. S5 and S6 for $c = 0.5$). Nevertheless, we acknowledge that a more granular exploration across the full range of recombination rates ($0 \leq c \leq 0.5$), which we leave for future work, is desirable as it could further refine our understanding of mutation–recombination interactions.

We infer the stationary distribution by consecutively iterating mutation, recombination, and selection dynamics until stability is reached: namely, if applying the dynamics for 10 consecutive steps results in a total variation in the distribution less than ϵ , where we set $\epsilon = 10^{-6}$. For the initial distribution, we focus on the two extremes: either a uniform distribution over the entire viable genotypic space or an isogenic distribution at a single genotype.

The first observation is that the introduction of recombination increases the stationary distribution’s localization; the lower the rate of mutation is relative to recombination, the more pronounced this localization becomes (see also refs. 26 and 32). This qualitative relationship is observed regardless of p and regardless of the initial distribution, though it is more pronounced for low P and isogenic initial distribution.

As seen in Fig. 4, the transition to high localization occurs rapidly as the mutation rate decreases past a certain inflection point, analogous to the “error threshold” introduced by Eigen (33). Thus, when the effect of recombination dominates over that of mutation, the resulting stationary distribution is substantially localized.

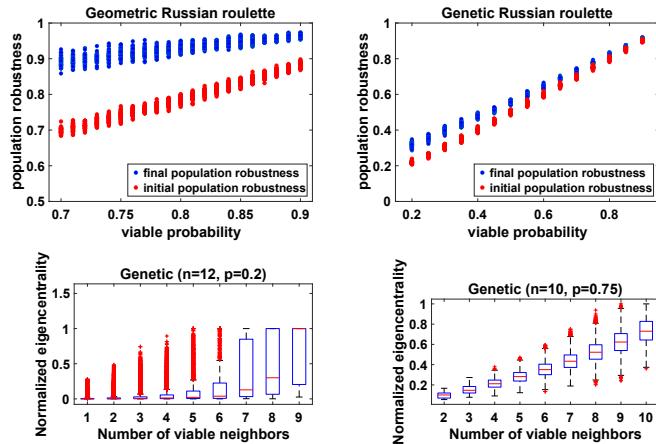


Fig. 3. Robustness and localization. The top two plots show the final versus initial measures of population robustness for random landscapes of both the geometric and genetic Russian Roulette families of landscapes. We observe that in both cases, the trend is of increasing robustness over the course of evolution, while the magnitude of change is more pronounced in the geometric landscapes as well as when the viable probability is lower. The bottom plots show the distribution of normalized eigencentrality (normalized so that 1 is the maximum), broken down by a node’s number of viable neighbors. We see that the distributions have starkly different shapes between the low- p and high- p cases of genetic Russian roulette landscapes. In both there is a correlation between number of neighbors and eigencentrality, but in the low- p case, there are many more outliers, nodes where the number of viable neighbors is relatively low while eigencentrality remains high. This is indicative of localization: being located in the right region of the network is more important than the immediate number of neighbors.

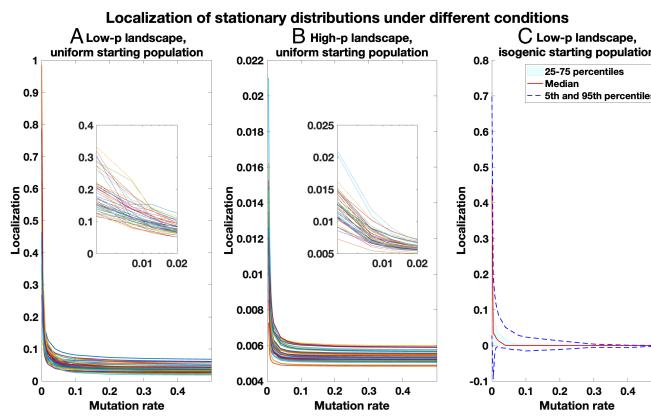


Fig. 4. Illustrating the dependence of localization on parameters in the haploid regime with mutation and recombination. Here, the rate of recombination ($c = 0.1$) is held fixed and the rate of mutation is allowed to vary ($\mu = 0.0 \dots 0.5$). The insets are presented on a log scale with μ ranging from 0.004 to 0.02). In (A) and (B), the localization is measured for the stationary distributions of 50 randomly generated landscapes (in A with low $P = 0.2$ and in B with high $P = 0.8$) under a uniform starting population. As the figure shows, when mutation is low, localization drastically increases. On the other hand, when mutation increases past a certain threshold, the localization measure stabilizes. Similar patterns are observed for low and high P , but the baseline localization is higher for low P . In (C), we measure for each of the 50 low- P landscapes its stationary distributions under all possible initial isogenic distributions. We then indicate several quantiles of the distribution of these stationary distributions' localization measure, normalized so that 0 represents the localization under uniform starting conditions. We notice that in the median case, localization tends to be even higher for isogenic starting populations when mutation is low.

However, in low P landscapes (Fig. 4A) the stationary distribution's localization is higher (as we have seen is the case for mutation only), and the transition to high localization occurs at a higher mutation inflection point, as compared to high P landscapes (Fig. 4B). Moreover, in low P landscapes, localization tends to be, but is not always, higher when the initial population is isogenic as opposed to uniform (Fig. 4C). In high P landscapes, we find that the initial population distribution has a limited effect on the stationary distribution.

We next observe that, when mutation rate is high relative to recombination, the stationary distribution tends to be very close to the principal eigenvector distribution, for both p regimes but especially for high P landscapes, indicating this distribution's stability to the presence of (relatively low) recombination, and the potential relevance of eigencentrality even in the presence of recombination. However, as mutation rates decrease, the stationary distribution's distance to the eigencentrality distribution increases due to its increased localization.

In addition to the gradual process of increased localization, the stationary distribution may experience abrupt shifts for certain values of the mutation parameter μ . These shifts can be measured by tracking the eigencentrality score of the mode of the stationary distribution (SI Appendix, Fig. S2) which at high values of μ is precisely the peak of eigencentrality for the vast majority of landscapes, but which abruptly falls for several landscapes as μ decreases, indicating sudden shifts to less eigencentral regions of the graph. These shifts are known as “error catastrophes” in the quasi-species literature (34, 35). These abrupt falls in the eigencentrality of the mode are more prevalent, more substantial, and begin at higher values of μ for low- P landscapes when compared to high- P landscapes. For the latter, even when the eigencentrality of the mode falls it is still much higher than the baseline (average) eigencentrality of nodes in the landscape.

Accordingly, the average eigencentrality of the stationary distribution reliably increases for high- P (Fig. 5B) landscapes as mutation decreases, due to the effect of localization—localizing around a high-eigencentrality region increases the stationary distribution's average eigencentrality even if its mode is not at peak eigencentrality. In contrast, for the low- P landscapes (Fig. 5A), while we observe a general trend of increasing average eigencentrality of the stationary distribution as mutation decreases, several landscapes see a fall in stationary distribution average eigencentrality, which can be explained by increased localization around low-eigencentrality regions. For another view on the interplay of eigencentrality and localization, SI Appendix, Fig. S1 tracks the total distance (measured as total variation distance between probability densities) between the stationary distribution and eigencentrality, a distance which can increase either due to increased localization or due to a shift in location.

We note that a similar though somewhat less pronounced relationship can be observed when measuring the stationary distribution's average mutational robustness instead of eigencentrality (SI Appendix, Fig. S10): it tends to reliably increase as mutation decreases (due to localization), but then, for low- P landscapes, might suddenly drop as mutation falls past a certain threshold. Such similar behavior can be explained by the correlation between eigencentrality and robustness. While the average mutational robustness over the 50 landscapes increases with decreased mutation, for some landscapes mutational robustness drops as recombination dominates. This complicates the observation by Klug et al. (24) that recombination tends to promote mutational robustness.

Under low- P landscapes with isogenic starting populations (Fig. 5C), most but not all stationary distributions are equally or less eigencentral in comparison to a uniform starting population, whether eigencentrality is measured by distance to

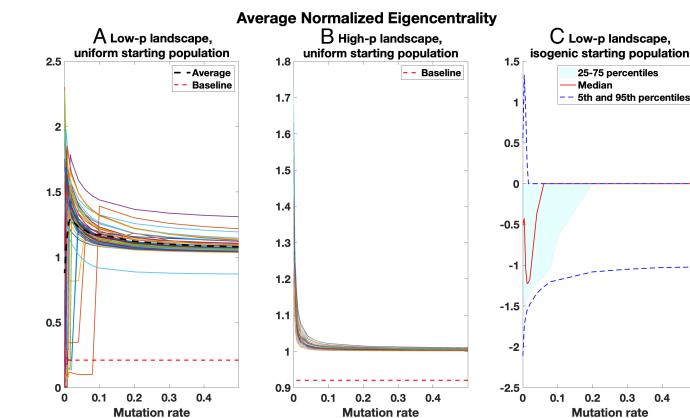


Fig. 5. In (A) and (B), the average eigencentrality is measured for stationary distributions of 50 randomly generated landscapes (in A, with low $P = 0.2$ and in B with high $P = 0.8$) under a uniform starting population, then normalized by dividing by the corresponding measure of the principal eigenvector distribution itself. A value greater than 1 implies a population distribution which is both located along eigencentral nodes and localized. While in the high P case, average eigencentrality reliably increases with decreased mutation rate due to increased localization, in the low P case as we decrease mutation rate, the average eigencentrality of the stationary distribution first increases due to localization, but then for certain landscapes it sharply decreases, as recombination dominates and results in shifting the region of genotype space occupied by the population to a less eigencentral region. In (C), we plot, for a given landscape the distribution of average eigencentralities for each possible isogenic starting population, normalized so that 0 represents the average eigencentrality under uniform starting conditions. We see that the result depends on initial condition and, more often than not, results in a lower eigencentrality due to getting stuck in a low-eigencentrality region.

the eigencentrality distribution, average eigencentrality, or the eigencentrality of the mode. We interpret this to be an effect of increased local influence of the starting state of the population. Indeed, we observe that in almost all cases in which the mode of the stationary distribution differs when we change from a uniform to an isogenic starting distribution, it moves closer to the starting point of that isogenic distribution.

In the case of isogenic starting distributions, we observe a bifurcation process. For high mutation rates, under the majority of starting points, the stationary distribution for that isogenic starting population is the same as with a uniform starting population. As mutation rate lowers, however, the possible stationary distributions bifurcate into an increasing number of different possibilities, representing basins of attraction. Thus, the local influence of the starting point on the final distribution increases as mutation rate decreases relative to recombination [see also Higgs (36)]. This reflects a transition away from the principal eigenvector (the only equilibrium associated to mutational dynamics) and toward the equilibria of the recombination dynamics, which can depend on initial condition. We now investigate these equilibria.

Stable Equilibria for Recombination. In order to appreciate the role of recombination in the dynamics of neutral evolution, it is instructive to examine the extreme case of recombination only, with no mutation. In this case, we make some observations about stable equilibria.

In our observations, in all the instances, we have looked at stable equilibria of the recombination dynamics are always *closed under recombination*: that is, any two genotypes with nonzero frequency in these equilibria will produce a viable genotype when exposed to any recombination event. It remains an open question to mathematically prove this observation, or to give general conditions under which it holds. In fact, a family of genotypes closed under recombination is necessarily a subhypercube of the n -dimensional hypercube, where some alleles are fixed and the other loci may assume any combination of alleles. Thus, the process of recombination only becomes localized to a small family of this type. This gives a theoretical explanation of the observation that recombination entails higher localization, since in the extreme case of no mutation, we see localization to a closed family of viable genotypes, which is necessarily small in a typical random network.

We emphasize that for each landscape there can be stable equilibria on multiple distinct closed subregions of the graph; and for each such subregion, there can be infinitely many different equilibria (37) (as the densities assigned to the genotypes may be continuously adjusted as long as they satisfy linkage equilibrium). A generic initial distribution will converge, when put through evolutionary dynamics, to one of these equilibria, depending on the initial condition.

When a small mutation rate is added, in the case where recombination dominates over mutation, the population will tend to converge to a distribution which is close to one of the above-mentioned equilibria, but somewhat more diffuse due to the presence of mutation. Again, the specific attractor it converges to depends on the initial condition. In Fig. 6, we plot a sample landscape (genetic Russian roulette with $n = 10, P = 0.2$) and four of its basins of attraction. For each genotype, evolve to stationarity a population beginning with an initial isogenic population of that genotype. We observe that the genotypes are partitioned into seven classes depending on which stationary distributions these isogenic initial distributions converge to. We

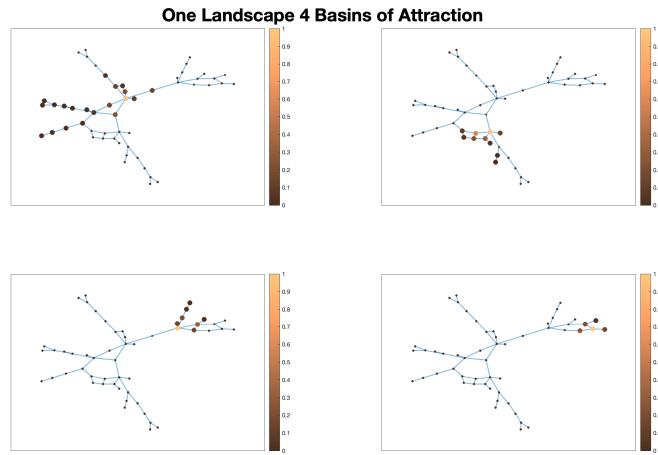


Fig. 6. The four largest basins of attraction of the given landscape (genetic Russian roulette with $n = 10, P = 0.2$). The basins of attraction are calculated by partitioning the network into equivalence classes according to the stationary distribution resulting from initial isogenic populations at each node. For each panel, a node is depicted larger if it belongs to the given basin of attraction, and the color represents each node's frequency in the attractor's stationary distribution.

consider these seven stationary distributions *attractors*, and the set of genotypes which converge to a given attractor its *basin of attraction*. In Fig. 6, we plot the four largest basins of attraction for the given landscape. Notice that each basin of attraction tends to be a connected region of the graph, which overlaps with the high-frequency nodes of its attractor, but also includes several low-frequency nodes.

The Case of Diploid Population

Next, we consider a diploid population. We assume n loci with two alleles each (the case of a diallelic model).

We consider the following two paradigms for generating families of holey landscapes. In the first paradigm, we think of a “locus” as actually signifying a single nucleotide. Thus the “genotype” under consideration is actually a gene or even a portion of a gene. In this view, a fitness value of 1 indicates maintenance of the wild-type functionality of this gene, whereas a fitness value of 0 indicates a deleterious effect on this functionality.

In the second paradigm, we think of a locus as representing an entire gene. In this view, a fitness value of 1 indicates viability of the genotype as a whole.

In this paper, we favor the first paradigm in applying our framework, since it more closely matches the model: each nucleotide position has a small number of possible values, whereas a gene has a vast number of possibilities. In order to work with our model under the second paradigm, one would have to restrict attention to only a small select set of mutants, which may be unrealistic. However, in *SIA Appendix, Fig. S9*, we analyze the second paradigm as well, showing qualitatively similar behavior as in the first paradigm, at least in the unstructured case explained below.

As such, working in the first paradigm, we require the following symmetry: the fitness of a pair of gametes is the same regardless of the order of gametes. (In contrast, if one were working in the second paradigm, one may require a stronger symmetry: the fitness of a diploid remains the same when transposing the alleles at a specific locus. In particular, the second paradigm does not distinguish between *cis* (e.g., $\frac{11}{00}$) and *trans* (e.g., $\frac{10}{01}$) genotypes,

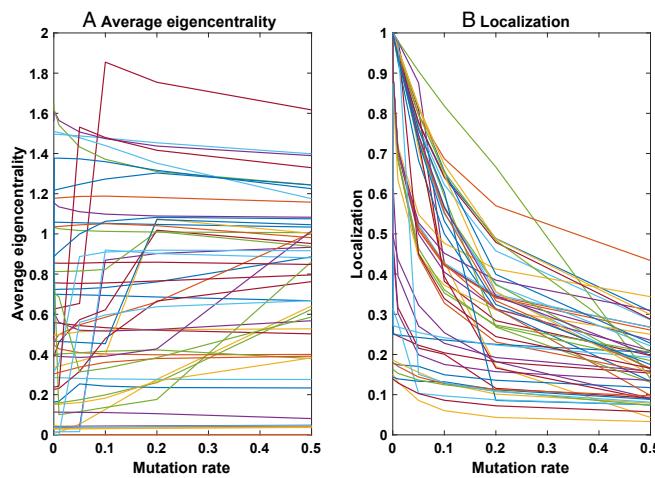


Fig. 7. (A) Average Eigencentrality (normalized so that 1 is the average eigencentrality of the principal eigenvector distribution itself) and (B) localization for 50 random diploid (unstructured) Russian roulette landscapes.

which are understood as the same heterozygous genotype, while under the first paradigm they are understood as distinct).

Next, we consider different viability functions that determine neutral diploid landscapes. A first approach, which we term *unstructured* diploid landscapes, is analogous to the Russian roulette in the haploid case: each (unordered) pair of gametes is randomly and independently deemed viable with probability P . We note that the mutational network structure of n -locus diploids is equivalent to that of $2n$ -locus haploids. However, even in the absence of recombination, the diploid case is drastically different due to the random assortment of gametes. Indeed, segregation can be seen as a de facto recombination-like event (38). To put it concretely, diploid reproduction with mutation and segregation without crossing over is very similar to—though not quite the same as—haploid reproduction with mutation and a single recombination event at a 100% rate (the reason that these two are different is that diploid recombination is indifferent to which gamete is picked).

Due to our understanding of haploid recombination, which dominates over mutation when the recombination rate is high and drives the population away from eigencentrality, we would expect the same to hold for diploid populations even in the absence of explicit recombination. Indeed, we observe that, regardless of recombination and even with a high mutation rate, the population does not approach high eigencentrality. Indeed, in Fig. 7A we plot average eigencentrality of the stationary distributions of 50 random landscapes under the diploid dynamics with no recombination, only random assortment (normalized so that 1 represents the average eigencentrality of the principal eigenvector distribution itself). The figure shows that the average eigencentrality widely varies among landscapes, with no clear trend (compared to the haploid case). On the other hand, increased mutation does still lead to consistently lower localization by spreading out the population (Fig. 7B) (no qualitative differences are observed across varying recombination rates; see *SI Appendix*, Figs. S7 and S8, for recombination rate $c = 0.5$ and $c = 0.05$ respectively).

Adding in recombination (crossing over) does not have a clear effect on the distribution or its localization. The stationary distributions are largely similar with or without recombination. In the high-mutation case, 90% of the landscapes had the same mode when recombination was added, and in the zero-mutation case, 80% of the landscapes had the same mode.

An alternative approach, which we refer to as *structured* diploid landscapes, is to define the viability of a genotype as a function of the individual gametes' viabilities. Here, we consider two possible ways of combining a viability function on gametes to form a viability function on diploids: dominance of lethality, requiring that both gametes be viable in order for the genotype to be viable; and recessiveness of lethality, requiring that at least one gamete be viable in order for the genotype to be viable.

In the case of dominance, we note that under the dynamics of mutation and recombination, the gamete distribution evolves independently in a manner nearly identical to the haploid dynamics, and the diploid distribution is simply the product of the gamete distribution. Indeed, when we defined our model of diploid recombination dynamics, we noted that the recombination tensor for diploid distributions behaves as the product of corresponding haploid recombination tensors for the individual gamete distributions. The same is not quite true for the mutation dynamics, due to our simplifying modeling assumption that there can only be at most one mutation in every generation. Thus, a diploid under our model can be exposed to at most one mutation per generation whereas two independent haploids could be exposed to two. However, as mutation rate is generally small, the chance of two mutations can be treated as negligible. Therefore, the diploid mutation dynamics is approximately the same as independently exposing the gametes to haploid dynamics and then taking the product distribution (we could change our model to make this approximation exact, but currently choose not to for simplicity). Finally, the selection dynamics for diploids also behave according to independent selection of gametes, by virtue of the underlying assumption of dominance of lethality which guarantees a diploid is viable exactly when both its gametes are. Therefore, the entire evolutionary dynamics behave approximately as the product of haploid dynamics of the gametes. Indeed, simulations show that, for a diploid landscape with dominance of lethality of size 100, the difference between the stationary distribution of the diploid dynamics and the product of haploid dynamics is very small: a difference of less than 10^{-4} for each diploid genotype.

Due to the above theoretical observation, we do not analyze and simulate this case further as it is approximately reducible to the haploid dynamics already mentioned. In particular, in the mutation-only case, we see that the population distribution will converge to the product of the principal eigenvector distributions of the individual gametes. This product is in fact identical to the principal eigenvector distribution of the diploid landscape itself under mutation (see *SI Appendix* on product structures for more detail).

In the case of recessiveness of lethality, the diploid stationary distribution, even in the mutation-only case, does not have such a simple description, and is not given in terms of the eigencentrality of the gametes. Indeed, in a recessive model, a viable diploid can have a gamete which is on its own inviable, and thus does not have a notion of eigencentrality.

For these landscapes, unlike the Russian Roulette diploid landscapes, we observe that the stationary distribution does tend to closely align with its eigencentrality (defined in terms of the principal eigenvector of the adjacency matrix of diploids under mutation). We first consider the case of only mutation, and no recombination (only the random assortment of gametes). Fig. 8 shows relevant statistics for a typical landscape:

In Fig. 8, we order the landscape's genotypes in order of increasing eigencentrality (whose value is indicated by the green line), while the blue and red lines indicate the stationary distributions under high (0.5) and low (0.001) values of mutation,

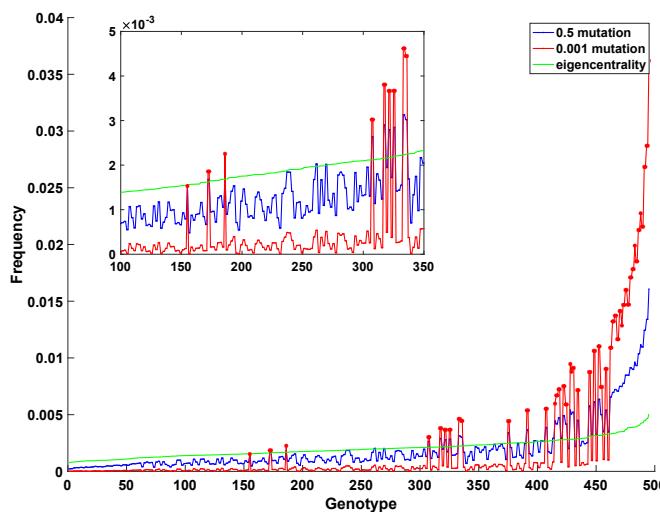


Fig. 8. For a fixed landscape in the recessive lethality family of diploid landscapes, we calculate stationary distributions under high mutation (0.5) and low mutation (0.001). In both cases, there is no explicit recombination, only random assortment of gametes. The blue and red lines show the stationary distributions of the evolutionary dynamics under the two cases, respectively. Genotypes are sorted in ascending order of their eigencentrality measure, and eigencentrality is plotted in the solid green line. The insert at the Top Left zooms in on a portion of the plot. Asterisks are placed over those genotypes both of whose gametes are viable.

respectively. The asterisks are placed along those genotypes where both gametes are viable. We can make several observations: first, we notice that the eigencentrality distribution is characterized by low localization—the distribution is spread out over all genotype space. Moreover, eigencentrality itself tends to reward those double-viable genotypes, which makes sense given that those genotypes will tend to have many more viable neighbors as any single mutation preserves the viability of the diploid. However, notably there are a few exceptions to this pattern: some double-viable genotypes have lower eigencentrality than some single viabiles. This can be explained by the fact that eigencentrality rewards not just the number of viable neighbors, but those neighbors' centrality as well. Thus, a single-viable in a central region of the graph can be rewarded over a double-viable whose location in the graph is less central. Next, we can observe that, especially in the higher-frequency genotypes of the distributions, there is a strong relationship between their ranking in eigencentrality and their ranking in frequency, confirming that the stationary distribution aligns with eigencentrality. We observe that the key effect of decreasing mutation rate is to boost the frequency of the double-viables relative to the single-viables. However, within each of these pools the relative relationship of frequencies remains similar, as can be seen in the insert, comparing the red line and blue line among the single-viable genotypes. This effect of boosting the double-viables can be understood by considering the limit case when mutation rate is brought down to 0. In this case, it can be shown mathematically (*SI Appendix*) that only the diploids where both gametes are viable survive in the stationary distribution, at a proportional frequency to their initial distribution.

Relatedly, decreasing mutation leads to increased localization of the stationary distribution, consistent with the tendency of the distribution toward the small set of double-viable genotypes. If we think of random assortment as de facto recombination, this behavior is also consistent with the dynamics we have observed of landscapes under recombination, tending toward

higher localization when mutation is low. Specifically, the set of double-viables acts as an attractor which is closed under the random-assortment operator.

Next, we consider the case in which recombination is added in addition to mutation. Here, we observe the same broad patterns as above: the stationary distribution tends to align with eigencentrality (although there are now more instances of breaking away from eigencentrality when mutation is decreased), and localization increases as the mutation rate decreases. Moreover, the stationary distribution is more localized when compared to the analogous stationary distribution with segregation without crossing over but the same mutation rate.

Given the importance of eigencentrality to the stationary distributions of these recessive lethality landscapes, it is an interesting question to relate the eigencentrality of a diploid genotype to its component gametes and their position within the fitness landscape of gametes. One preliminary observation is that, while for the most part the eigencentrality favors diploids where both gametes are viable, nonetheless some inviable gametes have a higher total frequency in the principal eigenvector distribution than some viable gametes. This can be explained by the observation that some inviable gametes may still have a high number of highly eigencentral viable gametes, which will make them more favored when compared with viable gametes that do not have many viable or central neighbors. This observation reinforces the importance of position within the network, which can in some cases prove more important than a gamete's immediate fitness. However, as explained above, the frequency of such inviable gametes is further lessened as the mutation rate is decreased.

Discussion

To summarize, our study details and examines the richness of phenomena associated with neutral evolution of large populations, with an emphasis on the importance of *network structure* in shaping these evolutionary dynamics, while also examining different regimes in parameter space that lead to qualitatively distinct evolutionary behaviors. Our findings highlight the complex interplay between mutation, recombination, and the structure of fitness landscapes in shaping evolutionary trajectories, even in the absence of fitness differences among viable genotypes. They underscore the importance of considering the global structure of these landscapes to understand evolutionary outcomes.

Two properties of population distributions that we pay particular attention to are *eigencentrality*, referring to the location of the distribution among genotypic nodes which are highly represented in the landscape's principal eigenvector and can be viewed as more central (or robust) in the network, and *localization*, a measure of the homogeneity of a distribution, where most of the population is concentrated along a small region of genotypic space.

Though knowledge of these concepts' role in neutral evolution is not new, they remain underexplored, and a full understanding of the range of implications is still lacking. Toward a more robust understanding, we carry out a systematic study of these properties under different regimes in parameter space and their dependence on network structure.

To summarize our findings broadly: under a regime of only mutation in haploid populations, the population converges to the principal eigenvector distribution regardless of mutation rate or initial conditions. This distribution may or may not be localized depending on the network structure: for random landscapes,

the localization increases as the probability of viability decreases. Once recombination is added to the evolutionary dynamics, these dynamics become more complex: they depend both on the initial conditions and on the balance between the parameters. Broadly speaking, there are two regimes, with a phase shift between them depending on the balance between mutation and recombination rates: if mutation dominates, then the stationary distribution resembles the principal eigenvector, while being more localized due to the influence of recombination, whereas if recombination dominates, then the distribution no longer resembles the principal eigenvector but instead approaches attractors of the recombination dynamics, where the specific attractor that the distribution goes toward depends on initial conditions. These attractors tend to be very localized, explaining the tendency of recombination to increase the distribution's localization. We note that while (24) made the general observation that the presence of recombination tends to increase average population robustness, we complicate this observation, as high recombination can often lead a population away from the most eigencentral region of the graph, while compensating for this by increasing the population's localization.

Finally, diploid populations present a different case. Random diploid landscapes, even without explicit recombination, behave qualitatively similarly to haploid landscapes where recombination dominates—that is, they bear no relation to eigencentrality. However, if the diploid landscape is structured, such that diploid viability is a function of the component gametes' viabilities, then in fact the stationary distribution does bear a close relationship to eigencentrality, showing that eigencentrality has the potential to be a relevant factor in the neutral evolution of diploid populations as well.

The potential broader implications of our findings are myriad. We suggest that the role of neutral evolution should be taken seriously as a possible explanation of phenomena that are typically attributed to fitness differentials. We note that, while the neutral landscapes we study have no fitness differentials between viable genotypes, nevertheless the mutation–selection dynamics will drive the population toward an equilibrium where some viable genotypes are systematically favored over others, due to their position within the neutral network.

For example, conserved regions within the genome may not necessarily reflect a selective advantage. Instead, they may be the result of neutral evolutionary dynamics in which recombination rates predominate, which can yield very high population localization for the neutral landscape of that portion of the genome, even without direct selection. Similarly, in cases where separate populations or even distinct species appear to exhibit evolutionary convergence, it may not be accurate to attribute this convergence solely to selective pressures favoring a particular sequence. Such convergent evolution might instead be a manifestation of the structure of the neutral fitness landscape. This could be consistent with neutral evolution in either the mutation-dominant or recombination-dominant regimes, where neutral evolutionary processes guide different initial populations toward similar genotypic regions without the direct influence of selection.

1. C. H. Waddington, Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565 (1942).
2. S. L. Rutherford, S. Lindquist, Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
3. C. Queitsch, T. A. Sangster, S. Lindquist, Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).
4. L. Altenberg, Fundamental properties of the evolution of mutational robustness. arXiv [Preprint] (2015). <https://arxiv.org/abs/1508.07866> (Accessed 31 August 2015).

When analyzing the patterns of evolutionary trajectories, it is essential to recognize that different regions of the genome may not evolve in qualitatively identical ways. Our study suggests that the qualitative properties of these evolutionary trajectories are highly dependent on the structure of a particular genetic site's fitness landscape, or even on the relative rates of mutation and recombination which can vary over the genome. This highlights the need for a nuanced approach to studying evolutionary processes, where the specific dynamics of mutation and recombination as well as structural characteristics of the fitness landscape are carefully considered before drawing conclusions about the presence and relevance of selection in shaping these trajectories (8, 34, 35, 39–46).

The fact that qualitatively distinct evolutionary dynamics can arise depending on the balance between mutation and recombination also has potential implications, which we have not explored here, on the evolution of recombination and the regulation of these rates.

Many questions remain which can be tackled in future work. We assumed for simplicity an infinite population and a binary (0 or 1) fitness function. It is understood that relaxing the fitness function so that deleterious mutations do not lead to 0 fitness but rather to a sharp drop in fitness, does not significantly alter the general evolutionary dynamics (18), but a more thorough testing of this understanding is merited. Furthermore, the study of finite (even if large) populations adds the potential of random drift as an additional complicating factor, whose interaction with the dynamics we have identified requires further study. While it is understood that, when drift dominates, the impact of network topology is lessened and the population will disperse uniformly over the neutral network (18, 27), there may be a “middle ground” regime where neither mutation nor drift completely dominates.

Moreover, there is potential in applying mathematical tools from random graph theory and spectral graph theory to lend a theoretical foundation to the observations we have made.

Finally, our simulations could be augmented by considering multiple populations with a migration rate or adding a spatial component, which can shed light on the implications of neutral evolution on speciation and the maintenance of genetic divergence between populations despite gene flow. Specifically, as we have seen, recombination dynamics can drive populations to different attractors depending on the initial conditions. We hypothesize that these attractors, even if not fully incompatible with each other, could in some cases exert enough pull that distinct populations at different attractors can maintain their divergence even in the presence of limited amounts of gene flow between them.

Data, Materials, and Software Availability. There are no data underlying this work. The code for simulations is publicly available at <https://github.com/AvivLab/Robustness> (47).

ACKNOWLEDGMENTS. This research was supported by internal funding from the Albert Einstein College of Medicine. A.B. gratefully acknowledges the generous hospitality and support of the Santa Fe Institute during his sabbatical.

5. E. Bornberg-Bauer, H. S. Chan, Modeling evolutionary landscapes: Mutational stability, topology, and superfunnels in sequence space. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10689–10694 (1999).
6. M. L. Siegal, A. Bergman, Waddington's canalization revisited: Developmental stability and evolution. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10528–10532 (2002).
7. A. Bergman, M. L. Siegal, Evolutionary capacitance as a general feature of complex gene networks. *Nature* **424**, 549–552 (2003).
8. T. MacCarthy, A. Bergman, Coevolution of robustness, epistasis, and recombination favors asexual reproduction. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12801–12806 (2007).

9. M. L. Siegal, D. E. Promislow, A. Bergman, Functional and evolutionary inference in gene networks: Does topology matter? *Genetica* **129**, 83–103 (2007).
10. M. Kimura, The neutral theory of molecular evolution. *Sci. Am.* **241**, 98–129 (1979).
11. S. Wright, "The roles of mutation, inbreeding, crossbreeding, and selection in evolution" in *Proceedings of the Sixth International Congress of Genetics* (1932).
12. S. Wright, Surfaces of selective value revisited. *Am. Nat.* **131**, 115–123 (1988).
13. M. Pigliucci, Sewall Wright's adaptive landscapes: 1932 vs. 1988. *Biol. Philos.* **23**, 591–603 (2008).
14. A. Bergman, D. B. Goldstein, K. E. Holsinger, M. W. Feldman, Population structure, fitness surfaces, and linkage in the shifting balance process. *Genet. Res.* **66**, 85–92 (1995).
15. S. Gavrilets, Evolution and speciation on holey adaptive landscapes. *Trends Ecol. Evol.* **12**, 307–312 (1997).
16. S. Gavrilets, A dynamical theory of speciation on holey adaptive landscapes. *Am. Nat.* **154**, 1–22 (1999).
17. S. Gavrilets, *Fitness Landscapes and the Origin of Species* (MPB-41) (Princeton University Press, 2004), vol. 41.
18. E. Van Nimwegen, J. P. Crutchfield, M. Huynen, Neutral evolution of mutational robustness. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9716–9720 (1999).
19. G. Strang, *Introduction to Linear Algebra* (Wellesley-Cambridge Press, 2023).
20. M. E. Newman, The mathematics of networks. *New Palgrave Encycl. Econ.* **2**, 1–12 (2008).
21. K. J. Sharkey, Localization of eigenvector centrality in networks with a cut vertex. *Phys. Rev. E* **99**, 012315 (2019).
22. D. N. Arnold, G. David, M. Filoche, D. Jerison, S. Mayboroda, Localization of eigenfunctions via an effective potential. *Comm. Partial Differ. Equations* **44**, 1186–1216 (2019).
23. M. Smerlak, Neutral quasispecies evolution and the maximal entropy random walk. *Sci. Adv.* **7**, eabb2376 (2021).
24. A. Klug, S. C. Park, J. Krug, Recombination and mutational robustness in neutral fitness landscapes. *PLoS Comput. Biol.* **15**, e1006884 (2019).
25. A. Klug, J. Krug, Conflicting effects of recombination on the evolvability and robustness in neutrally evolving populations. *PLoS Comput. Biol.* **18**, e1010710 (2022).
26. G. J. Szöllősi, I. Derényi, The effect of recombination on the neutral evolution of genetic robustness. *Math. Biosci.* **214**, 58–62 (2008).
27. E.J.v. Nimwegen, "The statistical dynamics of epochal evolution," PhD thesis, Utrecht University, AMS (1999).
28. S. Gavrilets, Perspective: Models of speciation: What have we learned in 40 years? *Evolution* **57**, 2197–2215 (2003).
29. M. Levinstein, B. Shklovskii, M. Shur, A. Efros, The relation between the critical exponents of percolation theory. *Sov. J. Exp. Theor. Phys.* **42**, 197 (1975).
30. S. Hata, H. Nakao, Localization of Laplacian eigenvectors on random networks. *Sci. Rep.* **7**, 1121 (2017).
31. T. Paixão, K. E. Bassler, R. B. Azevedo, Emergent speciation by multiple Dobzhansky-Muller incompatibilities. *bioRxiv* [Preprint] (2014). <https://doi.org/10.1101/008268> (Accessed 7 July 2015).
32. Y. Xia, M. Levitt, Roles of mutation and recombination in the evolution of protein thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10382–10387 (2002).
33. M. Eigen, Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465–523 (1971).
34. J. J. Bull, L. A. Meyers, M. Lachmann, Quasispecies made simple. *PLoS Comput. Biol.* **1**, e61 (2005).
35. P. Schuster, "Quasispecies on fitness landscapes" in *Quasispecies: From Theory to Experimental Systems*, E. Domingo, P. Schuster, Eds. (Springer International Publishing, Cham, Switzerland, 2016), pp. 61–120.
36. P. G. Higgs, Compensatory neutral mutations and the evolution of RNA. *Genetica* **102**, 91–101 (1998).
37. M. Pontz, M. W. Feldman, Loss of genetic variation in the two-locus multiallelic haploid model. *Theor. Popul. Biol.* **136**, 12–21 (2020).
38. S. P. Otto, The advantages of segregation and the evolution of sex. *Genetics* **164**, 1099–1118 (2003).
39. M. Nei, Modification of linkage intensity by natural selection. *Genetics* **57**, 625 (1967).
40. M. W. Feldman, Selection for linkage modification: I. Random mating populations. *Theor. Popul. Biol.* **3**, 324–346 (1972).
41. U. Liberman, M. W. Feldman, Modifiers of mutation rate: A general reduction principle. *Theor. Popul. Biol.* **30**, 125–142 (1986).
42. A. Bergman, M. W. Feldman, Recombination dynamics and the fitness landscape. *Physica D* **56**, 57–67 (1992).
43. M. C. Boerlijst, S. Bonhoeffer, M. A. Nowak, Viral quasi-species and recombination. *Proc. R. Soc. Lond. Ser. B: Biol. Sci.* **263**, 1577–1584 (1996).
44. R. Bürger, Linkage and the maintenance of heritable variation by mutation-selection balance. *Genetics* **121**, 175–184 (1989).
45. S. C. Park, J. Krug, Bistability in two-locus models with selection, mutation, and recombination. *J. Math. Biol.* **62**, 763–788 (2011).
46. M. N. Jacobi, M. Nordahl, Quasispecies and recombination. *Theor. Popul. Biol.* **70**, 479–485 (2006).
47. Y. Sella, A. Bergman, Code for "Robustness revisited: On the neutral evolution of centrality and localization." Github. <https://github.com/AvivLab/Robustness>. Accessed 10 May 2025.