Escape from Typicality: Why Evolution Operates Near One Mutation per Genome per Generation

Ishanu Chattopadhyay[a]

This manuscript was compiled on February 17, 2026

Across organisms and viruses, the per-site mutation rate scales inversely with genome length, yielding an approximately constant number of mutations per genome per generation (Drake's rule). Existing explanations emphasize biochemical fidelity constraints or population-genetic error-threshold arguments that limit mutation from above. Here we provide a complementary, mechanism-independent account based on variation supply. Treating mutation as a blind local perturbation process, we ask which mutation rates maximize the expected rate of producing variants that are atypical relative to the mutation kernel itself, and therefore available for selective amplification rather than drift into mutational background. Using counting bounds for the fraction of such mutation-atypical outcomes within Hamming neighborhoods and combining them with an independent per-site mutation model, we obtain an explicit discovery-rate expression that is maximized when the per-genome mutation intensity $T = n\mu$ is $O(1)$, with a peak near $T \approx 1$. Consequently, the optimal per-site rate scales as $\mu^\star = (1 + o(1))/n$. These results derive Drake's rule from a generic supply–destruction tradeoff under local stochastic perturbations, complementing fidelity- and stability-based perspectives.

biological evolution | mutation rate | algorithmic complexity | Drake's rule

Mutation injects randomness; selection preserves and amplifies structure. Evolution therefore operates under a fundamental tension between entropy production and entropy suppression. A striking empirical regularity across diverse taxa is that the per-site mutation rate μ scales approximately as $1/n$, where n is genome length, so the per-genome mutation rate $U = n\mu$ is roughly constant. This phenomenon, observed from RNA viruses to eukaryotes, is known as Drake's rule (1–4).

Standard explanations of Drake's rule fall into two broad categories. One emphasizes biochemical constraints and tradeoffs in replication fidelity, proofreading, and energetic cost (2). The other emphasizes the population-genetic consequences of excessive mutation, most prominently Eigen's error-threshold framework (5), in which high mutation loads destabilize inherited information and limit the maintenance of adapted structure. These approaches provide important insight and useful upper bounds, but they primarily explain why mutation rates cannot be too large. They do not, by themselves, explain why evolution repeatedly operates near a particular scaling regime with $U = O(1)$.

Here we pursue a complementary route that is upstream of selection and does not invoke biochemical detail, explicit fitness landscapes, or equilibrium population genetics. We ask: given a blind local mutation mechanism, at what mutation load is the *supply rate* of statistically exceptional variants maximized? The point is not that evolution seeks exceptionality per se, but that selection can only act on what mutation supplies, namely for a parent genome x:

$$x \xrightarrow{\text{mutation}} \text{Large set of random mutants} \xrightarrow{\text{selection}} \{y_1, \ldots, y_k\} \qquad [1]$$

If mutation predominantly produces outcomes that are typical under its own kernel, then repeated mutation pushes lineages toward the kernel's typical set, and accumulated organization is continually eroded into mutational background; we descend into noise (See Fig.1b). Any sustained adaptive process therefore requires a nonzero supply of outcomes that are atypical relative to mutation alone, since these are the only candidates that can be preferentially retained rather than washed out. Thus, we are not proposing any explicit evolutionary objective function; we are analyzing structural constraints on variation supply induced by local blind search.

Our starting observation is that mutation induces a natural statistical model for an offspring relative to its parent. For a parent genome $x \in [q]^n$, an m-mutation restricts the offspring y to the Hamming sphere $H_m(x)$, the set of sequences at Hamming distance m from x. Under the mutation-induced model that is uniform over $H_m(x)$, a typical draw requires $\log |H_m(x)|$ bits to specify once the radius is

Author affiliations: [a]University of Kentucky

IC conceived of research, implemented the algorithm, wrote the paper and procured suppport.

Authors have no competing interests.

[2]To whom correspondence should be addressed. E-mail: ishanu_ch@uky.edu

fixed, and $|H_m(x)| = \binom{n}{m}(q-1)^m$ grows rapidly with m. This immediately implies a notion of typicality: at a given radius, most outcomes resemble typical draws from the mutation kernel, and only a small fraction are exceptional relative to that kernel. In particular, the larger the neighborhood, the thinner the tail of mutation-atypical outcomes within it.

We formalize this atypicality or exceptionality relative to the mutation-induced model on $H_m(x)$ using randomness deficiency (Defn. S1), which measures how much rarer an outcome is than a typical draw from the same model. Rarity here is not a proxy for fitness; rather, being rare relative to the mutation kernel is what makes an outcome statistically distinguishable from mutational background and therefore *available*, in principle, for selective amplification. A mutant is called a *net structure discovery* if its deficiency exceeds a margin Δ (Def. S1). A two-part coding/counting argument (Thm. S1, Lem. S1) implies that, for typical parents, the fraction of Δ-exceptional outcomes at radius m scales as $2^{-\Delta}/|H_m(x)|$ up to polylogarithmic factors; this "typical" qualifier is in a counting/combinatorial sense (formalized in the SI Appendix) and is not an empirical claim that biological genomes are algorithmically random. (However, real genomes often lie close to the maximal entropy rate for a four-letter alphabet, consistent with the combinatorial typicality regime, see Fig.1a.) We then combine this typical-case rarity with an independent per-site mutation model $M \sim \mathrm{Binomial}(n, \mu)$ to define an expected discovery rate $\Phi(\mu)$ that averages the Δ-exceptional probability over mutation radii (Eq. (S14)). Optimizing $\Phi(\mu)$ yields an interior optimum at per-genome mutation intensity $T = n\mu = O(1)$ (Theorem S2), and in particular,

$$\mu^\star = \frac{1 + o(1)}{n}.$$

Thus Drake's rule emerges as a mechanism-independent supply optimum under blind local perturbation, complementing fidelity- and stability-based perspectives such as the error-threshold viewpoint.

Discussion

This work provides a mechanism-independent route to Drake's rule from the standpoint of variation supply. Under a minimal model in which mutation acts as a local stochastic perturbation on a discrete sequence space, we derived a typical-case bound on the probability that an m-mutant is statistically exceptional relative to its mutation neighborhood. Combining this rarity scaling with an independent per-site mutation model yields an explicit expression for the expected rate at which mutation supplies such exceptional variants, and shows that this rate is maximized when the expected number of mutations per genome per generation is $O(1)$. In the usual per-site parametrization, this corresponds to the inverse-length scaling $\mu^\star = \Theta(1/n)$.

A useful way to express the optimum is through the per-genome mutation intensity

$$T \;:=\; n\mu \;=\; \mathbb{E}[M], \qquad M \sim \mathrm{Binomial}(n, \mu),$$

which acts as an intensive "mutation temperature" controlling the typical perturbation size. Figure 1 illustrates the scaling directly. In Fig. 1c, $\Phi(\mu)$ plotted against the per-site rate μ exhibits a single interior maximum whose location shifts left as n increases, consistent with $\mu^\star \propto 1/n$. In Fig. 1d, plotting the same discovery rate against $T = n\mu$ collapses curves for different n onto a common profile with a maximum near $T \approx 1$, matching the predicted Te^{-T} dependence up to a multiplicative constant. The collapse emphasizes that the scaling law is most naturally stated as a constant optimal per-genome perturbation size. More precisely, the exact finite-n expression involves the factor $(1-\mu)^n$, so that the leading behavior is proportional to $T(1-\mu)^n$; the Te^{-T} profile is its small-μ asymptotic form. The existence of an interior optimum at $T = O(1)$ does not depend on the Poisson approximation itself but on the independence and thin-tailed nature of the mutation count distribution.

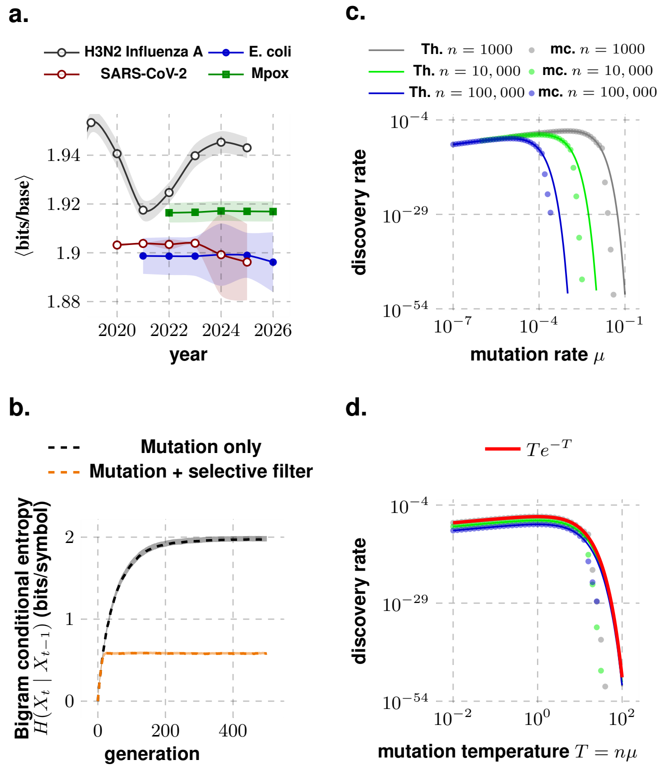

Fig. 1. Scaling and collapse of the discovery rate under independent per-site mutation. a. Genome compression values over time (viral genomes from GISAID; *E. coli* from NCBI; see SI) illustrating empirical boundedness of sequence complexity over evolution and the near-incompressibility of genomes under a reference-free compression proxy. The Shannon ceiling for a four-letter alphabet is 2 bits/base, achieved by perfectly incompressible sequences. Shaded bands indicate 95% confidence intervals across genomes within each year. **b.** Toy evolution on $[4]^n$ under mutation alone versus mutation with a viability filter that accepts offspring only if they remain within a fixed Hamming distance of a structured template. The plotted statistic is a reference-free compression proxy (compressed bits per symbol using `zlib`) along replicate lineages; mutation alone drives trajectories toward a max-entropy regime, while the viability filter maintains bounded complexity. **c.** The expected discovery rate $\Phi(\mu)$ as a function of the per-site mutation rate μ for increasing genome lengths n (Th. = theory; mc. = Monte Carlo estimates). The maximizer shifts to smaller μ as n increases, consistent with the inverse-length scaling $\mu^\star \propto 1/n$. **d.** The same discovery rate plotted against the per-genome mutation intensity $T = n\mu$. Curves for different n collapse onto a common profile with a maximum near $T \approx 1$; we illustrate the predicted Te^{-T} dependence up to a multiplicative constant. The collapse demonstrates that the optimal regime corresponds to an $O(1)$ number of mutations per genome per generation.

Our notion of discovery is defined relative to the distribution induced by the mutation mechanism itself. This reference measure is not an external modeling choice but is canonically determined by the perturbation process that supplies variation. Selection can only act on outcomes drawn from this kernel; thus atypicality must be measured relative to it. We do not claim that the Hamming sphere is the globally optimal statistical model for an offspring, only that it is the model canonically induced by a local mutation mechanism on sequence space. The substantive claim is upstream: since mutation is the sole source of variation, selection can only act on what mutation supplies. If offspring were always typical draws from the mutation kernel, repeated perturbation would concentrate lineages toward the kernel's typical set, eroding accumulated organization into mutational background. Positive randomness deficiency provides a quantitative, mechanism-independent way to express the minimal sense in which a variant is not merely background mutational noise.

The inverse-length scaling is robust to the choice of novelty margin Δ, provided Δ does not grow with n fast enough to dominate the one-step indexing cost $c(1) = \Theta(\log n)$. In particular, any fixed Δ or $\Delta = o(\log n)$ yields the same $\mu^\star = \Theta(1/n)$ scaling, with Δ affecting only the overall rate through the factor $2^{-\Delta}$.

Why focus on discovery probability?. Why should evolution optimize the probability of structure discovery rather than a more direct quantity such as expected fitness gain (2–4)? Mutation has no access to fitness: it is a blind local perturbation mechanism that generates variants without foresight, and selection acts only after these variants have been produced. Any objective defined in terms of fitness therefore lives downstream of mutation and cannot be optimized at the level of the generative process itself. The discovery probability instead isolates an upstream constraint imposed by mutation on the supply of selectable variants.

It is also useful to situate this result relative to classical "error threshold" arguments. Eigen-type frameworks constrain mutation from above by requiring that inherited information remain stable under copying errors (5). Our result is different in kind: it identifies an interior optimum for the upstream *supply* of mutation-atypical variants under blind local perturbations, without invoking a fitness landscape or equilibrium population genetics. The fact that both perspectives emphasize the $U = T = n\mu = O(1)$ regime suggests that Drake's rule may reflect the intersection of two generic pressures: fidelity constraints that limit loss of existing information and supply constraints that limit how often mutation produces variants sufficiently atypical to be preferentially amplified. In this sense, operating near one mutation per genome per generation can be read as a regime in which both retention and exploration remain nontrivial.

If the perturbation kernel becomes strongly non-local (e.g., recombination, large indels) or highly overdispersed in mutation count (*e.g.*, mutator mixtures or burst-like replication errors), the detailed functional form of $\Phi(T)$ can change and the maximizing value may shift. However, for any local, thin-tailed perturbation kernel parameterized by an intensive scale T, the supply/erosion tradeoff generically produces an interior optimum at order-one T.

Beyond biological specificity. Although motivated by biological evolution, the derivation depends only on three ingredients: a discrete configuration space, a local stochastic perturbation mechanism, and the requirement that adaptive progress requires access to outcomes that are atypical relative to the perturbation mechanism itself. Any system that explores a high-dimensional space via local random perturbations (a blind search unaware of downstream fitness functions) faces an analogous tradeoff: perturbations that are too small rarely generate statistically exceptional outcomes, while perturbations that are too large overwhelm any low-complexity deviation by pushing outcomes into exponentially large neighborhoods in which the probability of exceptionality is sharply suppressed. The resulting optimum selects an $O(1)$ intensive perturbation scale, made visible here by the collapse of discovery-rate curves under the parameter T in Fig. 1.

In summary, the combinatorial growth of mutation neighborhoods suppresses the probability of mutation-atypical outcomes at large radii, and optimizing the resulting discovery rate under independent per-site mutation yields the observed inverse scaling $\mu^\star \sim 1/n$, equivalently an $O(1)$ mutation load per genome.

Materials and Methods

Approach. We model mutation as a blind local perturbation on $[q]^n$: conditional on m mutated sites, offspring are drawn uniformly from the Hamming sphere $H_m(x)$ around a parent x (formal setup in the SI Appendix, Sec. 1). For any finite model class $S \ni y$, the two-part code bound gives

$$K(y) \;\le\; K(S) + \log|S| + O(1),$$

so describing y via S incurs an intrinsic indexing cost $\log|S|$. Applying this with $S = H_m(x)$ yields

$$K(y) \;\approx\; K(x) + \log|H_m(x)|, \qquad |H_m(x)| = \binom{n}{m}(q-1)^m,$$

so the combinatorial growth of the mutation neighborhood determines the typical description length of an m-mutant (details in the SI Appendix, Eq. Eq. (S2)–Eq. (S3)).

We quantify mutation-atypicality using randomness deficiency (Eq. (S5)), which measures how much rarer an outcome is than a typical draw from the mutation-induced model. A mutant is called a *net structure discovery* if its deficiency exceeds a margin Δ (Def. S1). A counting argument (Thm. S1, Lem. S1) shows that only a $2^{-\Delta}/|H_m(x)|$ fraction of outcomes at radius m can be Δ-exceptional, up to polylogarithmic factors.

We then embed this local rarity within an independent per-site mutation model $M \sim \mathrm{Binomial}(n, \mu)$, averaging over mutation radii to obtain an expected discovery rate $\Phi(\mu)$ (Cor. S1). Analyzing the small-μ asymptotics of $\Phi(\mu)$ (Lem. S2) and optimizing with respect to μ (Thm. S2) yields an interior optimum at $n\mu = O(1)$ (Eq. (S20)).

Code availability. Code to generate the plots are available https://github.com/zeroknowledgediscovery/drake

ACKNOWLEDGMENTS. This work was supported in part by the Defense Advanced Research Projects Agency (DARPA) under the ARC program (HR0011-26-3-E016). Views, opinions, and findings expressed are solely those of the authors.

1. JW Drake, A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci.* **88**, 7160–7164 (1991).
2. M Lynch, Evolution of the mutation rate. *Trends Genet.* **26**, 345–352 (2010).
3. P Sniegowski, Y Raynes, Mutation rates: how low can you go? *Curr. Biol.* **23**, R147–R149 (2013).
4. R Hershberg, Mutation—the engine of evolution: studying mutation and its role in the evolution of bacteria. *Cold Spring Harb. Perspectives Biol.* **7**, a018077 (2015).
5. M Eigen, Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465–523 (1971).
6. M Li, PMB Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications.* (Springer), 4 edition, (2019).
7. P Gács, JT Tromp, PMB Vitányi, Algorithmic statistics. *IEEE Transactions on Inf. Theory* **47**, 2443–2463 (2001).

Supplementary Methods

Sequence Space, Mutation Geometry and Two-Part Codes. Let $[q]^n$ denote the set of length-n strings over an alphabet of size $q \geq 2$. For $x, y \in [q]^n$, let $d_{\mathrm{H}}(x,y)$ denote the Hamming distance. For $m \in \{0, 1, \ldots, n\}$, define the Hamming sphere

$$H_m(x) = \{y \in [q]^n : d_{\mathrm{H}}(x,y) = m\}. \quad \text{[S1]}$$

with cardinality $|H_m(x)| = \binom{n}{m}(q-1)^m$, where m represents the number of mutated sites in one generation. Let $K(\cdot)$ denote prefix-free Kolmogorov complexity with respect to a fixed universal Turing machine (6). If $x \in S$ and S is a finite set, then the two-part code bound is as follows:

$$K(x) \leq K(S) + \log|S| + O(1), \qquad K(x \mid S) \leq \log|S| + O(1). \quad \text{[S2]}$$

which expresses the optimality condition of two-part codes (6, 7): first describe the model S, then describe the index of x within S.

Next, we note that the mutation process induces a restricted model class. For fixed parent x and mutation radius m, the Hamming sphere $H_m(x)$ is the set of sequences accessible by an m-mutation. Crucially, $K(H_m(x) \mid x, m) = O(1)$, which implies $K(H_m(x)) = K(x) + O(\log n)$, where the logarithmic term accounts for encoding (n, m, q). This leads to the interpretation of *mutation indexing cost* as

$$c(m) = \log|H_m(x)| = \log\binom{n}{m} + m\log(q-1). \quad \text{[S3]}$$

Thus, random mutation induces the two-part code

$$K(y) \ \approx \ K(x) + c(m), \quad \text{[S4]}$$

for y uniformly drawn from $H_m(x)$. Finally, for a model $S \ni y$, the randomness deficiency (6, 7) is

$$\Delta_y(S) = K(S) + \log|S| - K(y). \quad \text{[S5]}$$

Thus, admissible models are restricted to mutation-accessible sets $H_m(x)$.

Definition S1 (Net structure discovery)**.** *Fix $\Delta \geq 1$. A mutant $y \in H_m(x)$ constitutes a* net structure discovery *if $\Delta_y\big(H_m(x)\big) \geq \Delta$, or equivalently, $K(y) \leq K(x) + c(m) - \Delta$.*

Net structure discovery formalizes novelty as being exceptional relative to what mutation alone would typically produce.

Theorem S1 (Universal upper bound)**.** *For any $x \in [q]^n$, any m, and any $\Delta \geq 1$,*

$$\Pr_{y \sim \mathrm{Unif}(H_m(x))}\Big[\Delta_y(H_m(x)) \geq \Delta\Big] \leq O\left(\frac{2^{-\Delta}}{|H_m(x)|}\right). \quad \text{[S6]}$$

Proof. If $\Delta_y(H_m(x)) \geq \Delta$, then $K(y) \leq K(H_m(x)) + \log|H_m(x)| - \Delta$. By the Kraft inequality for prefix-free complexity, at most 2^{T+1} strings satisfy $K(y) \leq T$. Substituting $T = K(H_m(x)) + \log|H_m(x)| - \Delta$ and dividing by $|H_m(x)|$ yields the bound. □

Lemma S1 (Typical tightness for incompressible parents)**.** *Let x satisfy $K(x) \geq n\log q - O(1)$. Then for y drawn uniformly from $H_m(x)$,*

$$K(y \mid x, m) = \log|H_m(x)| \pm O(1), \quad \text{[S7]}$$

with probability $1 - O(1/|H_m(x)|)$. Consequently, up to polylogarithmic factors,

$$\Pr\Big[\Delta_y(H_m(x)) \geq \Delta\Big] = \Theta\left(\frac{2^{-\Delta}}{|H_m(x)|}\right). \quad \text{[S8]}$$

Proof. Fix x and m. There exists a computable bijection

$$\pi_{x,m} : \{1, \ldots, |H_m(x)|\} \to H_m(x) \quad \text{[S9]}$$

such that given (x, m, i) one can compute $y = \pi_{x,m}(i)$. Hence

$$K(y \mid x, m) \leq \log|H_m(x)| + O(1) \quad \text{[S10]}$$

for all $y \in H_m(x)$. For the lower tail, define

$$A_\Delta = \{y \in H_m(x) : K(y \mid x, m) \leq \log|H_m(x)| - \Delta\}. \quad \text{[S11]}$$

Each $y \in A_\Delta$ has a prefix-free description of length at most $\log|H_m(x)| - \Delta$, so $|A_\Delta| \leq |H_m(x)| 2^{-\Delta + O(1)}$. Thus

$$\Pr[y \in A_\Delta] \leq 2^{-\Delta + O(1)}. \quad \text{[S12]}$$

Since x is incompressible, the two-part code through $H_m(x)$ is optimal up to $O(\log n)$ terms. Therefore

$$\Delta_y(H_m(x)) = \log|H_m(x)| - K(y \mid x, m) \pm O(\log n), \quad \text{[S13]}$$

which yields the stated asymptotic bound. □

Corollary S1 (Expected discovery under binomial mutation)**.** *Let $X \sim \mathrm{Unif}([q]^n)$ be the parent sequence. Let $M \sim \mathrm{Binomial}(n, \mu)$ denote the number of mutated sites, and conditional on $(X, M = m)$ let $Y \sim \mathrm{Unif}(H_m(X))$. Define the expected discovery probability*

$$\Phi(\mu) := \mathbb{P}\left(\Delta_Y(H_M(X)) \geq \Delta\right) = \sum_{m=1}^{n} \Pr(M = m) \Pr_{y \sim H_m(X)}\left[\Delta_y(H_m(X)) \geq \Delta\right]. \quad \text{[S14]}$$

Then, for $\Delta = o(\log n)$,

$$\Phi(\mu) \asymp 2^{-\Delta} \sum_{m=1}^{n} \Pr(M = m)\frac{1}{|H_m(X)|}. \quad \text{[S15]}$$

Proof. For $X \sim \mathrm{Unif}([q]^n)$, the standard incompressibility bound

$$\Pr\left[K(X) \leq n\log q - t\right] \leq q^{-t} \quad \text{[S16]}$$

implies that X satisfies the incompressibility condition of Lemma S1 with overwhelming probability. On this event, Lemma S1 yields

$$\Pr_{y \sim H_m(X)}\left[\Delta_y(H_m(X)) \geq \Delta\right] \asymp \frac{2^{-\Delta}}{|H_m(X)|}. \quad \text{[S17]}$$

The contribution of the exceptional set $\{X : K(X) \leq n\log q - t\}$ is at most q^{-t} and is negligible for $t = \omega(1)$. Substituting the tight bound into the definition of $\Phi(\mu)$ gives the stated expression. □

Note that upon substituting the binomial mutation model and the sphere cardinality $|H_m(X)| = \binom{n}{m}(q-1)^m$, the combinatorial factor $\binom{n}{m}$ appearing in $\Pr(M = m)$ cancels with the identical factor in $|H_m(X)|$. Thus the multiplicity of mutation-location choices does not amplify discovery probability: it is already accounted for by the mutation process itself. What remains is the symbol-choice entropy $(q-1)^m$ and the exponential suppression arising from the rarity of exceptional strings within each mutation-accessible model. We can therefore state the small-μ asymptotics of the discovery rate as follows.

Remark on mutation-count distributions. The Binomial model arises directly from independent per-site mutation. The Poisson(T) form used for intuition is the classical large-n, small-μ limit with $T = n\mu$ fixed. The exponential factor in the discovery rate originates from the probability of zero additional perturbations under this independent model. More generally, replacing the Binomial kernel by another thin-tailed mutation-count distribution replaces $(1-\mu)^n$ by the corresponding small-step mass of that kernel; the existence of an interior optimum at $T = O(1)$ persists under such replacements, although its precise location may shift.

Lemma S2 (Small-μ asymptotics of discovery rate)**.** *For $\mu = o(1)$,*

$$\sum_{m=1}^{n}\left(\frac{\mu}{q-1}\right)^m = \frac{\mu}{q-1} + O(\mu^2), \qquad (1-\mu)^n = e^{-n\mu + O(n\mu^2)}. \quad \text{[S18]}$$

Consequently,

$$\Phi(\mu) \asymp C\, 2^{-\Delta}\, \frac{\mu}{q-1}\, e^{-n\mu + O(n\mu^2)}, \quad \text{[S19]}$$

where C absorbs polylogarithmic factors.

Proof. The first expansion is the truncated geometric series. Substituting $n\log(1-\mu) = -n\mu + O(n\mu^2)$ into the expression for $\Phi(\mu)$ gives the claim. □

Theorem S2 (Optimal mutation rate)**.** *In the regime $\mu = o(1)$ with $n\mu^2 = o(1)$, the leading term $\mu e^{-n\mu}$ is maximized at $n\mu = 1$. Consequently,*

$$\mu^\star = \frac{1 + o(1)}{n}. \quad \text{[S20]}$$

Corollary S2 (Drake's rule)**.** *Optimizing information-theoretic structure discovery under random mutation yields an inverse scaling of per-site mutation rate with genome length.*

Robustness. The maximizer $n\mu = 1$ is obtained under the regime $\mu = o(1)$ with $n\mu^2 = o(1)$, corresponding to independent, local perturbations. For mutation kernels exhibiting strong overdispersion or correlated multi-site bursts, the detailed maximizing value can deviate, but the emergence of an interior optimum driven by the competition between supply ($\propto T$) and erosion ($\propto e^{-T}$) remains structurally intact for thin-tailed kernels.

Genome Dataset Description and Compression-Based Complexity Estimation. For the genome-compressibility analysis (Fig. 1a), we analyzed complete genomes from four organisms: H3N2 influenza A (2010–2025), SARS-CoV-2 (hCoV-19; 2020–2025), mpox virus (2022–2026), and *Escherichia coli* (2021–2026). All viral genomes (H3N2, SARS-CoV-2, mpox) were obtained from the GISAID database, and *E. coli* genomes were obtained from the NCBI GenBank/RefSeq genomes repository. In total, we analyzed 54,904 H3N2 genomes across 16 annual bins (2010–2025), 4,986 SARS-CoV-2 genomes across 6 annual bins (2020–2025), 3,429 mpox genomes across 5 annual bins (2022–2026), and 8,785 *E. coli* genomes across 6 annual bins (2021–2026). Sequences were grouped by collection year and filtered to retain high-coverage, near-complete assemblies; sequences with substantial ambiguity (extended runs of non-ACGT symbols), obvious truncation, or pervasive gaps were excluded prior to analysis. All retained genomes were processed through an identical compression pipeline to ensure cross-organism comparability.

To estimate compressibility we computed a reference-free "bits per base" statistic from lossless compression. Each genome sequence was mapped to an ASCII byte string over $\{A, C, G, T\}$ after the filtering step and then compressed using `zlib` with a fixed compression level and identical settings across all organisms and years. Let $C_{\mathrm{real}}(x)$ denote the compressed length in bits of a genome x of length n. We first computed the naive compressed bits per base $b_{\mathrm{raw}}(x) = C_{\mathrm{real}}(x)/n$. Because lossless compressors incur finite overhead and windowing effects that can bias short sequences, we applied a length-matched overhead correction: for each distinct length n observed within an organism-year bin, we generated a set of IID uniform random DNA sequences of length n, compressed them with the same `zlib` settings, and estimated the overhead term

$$\widehat{O}(n) = \mathrm{median}\Big(C_{\mathrm{rand}}(n)\Big) - 2n,$$

where $2n$ is the Shannon limit (in bits) for IID uniform DNA over a four-letter alphabet. The corrected bits-per-base value reported in Fig. 1a was then

$$b_{\mathrm{corr}}(x) = \frac{C_{\mathrm{real}}(x) - \widehat{O}(n)}{n},$$

with $b_{\mathrm{corr}}(x)$ truncated to the interval $[0, 2]$ for interpretability. Annual means were computed by averaging $b_{\mathrm{corr}}(x)$ over all genomes in the corresponding organism-year bin. Uncertainty bands in Fig. 1a correspond to 95% confidence intervals computed from the within-bin replicate distribution using the same procedure for all organisms and years.

Generation of Figure 1b: Toy Evolution Under Mutation With and Without a Viability Filter. Panel b illustrates a controlled simulation on sequence space $[4]^n$ designed to contrast mutation-driven entropy increase with mutation constrained by a simple structural viability criterion. We fix an alphabet of size $q = 4$ and a genome length n (as specified in the accompanying code). A single "structured template" sequence $x^\star \in [4]^n$ is sampled once and held fixed throughout the experiment. Independent replicate lineages are initialized at this template and evolved forward for a fixed number of generations.

At each generation, offspring are generated from the current sequence by independent per-site mutation with probability μ. Conditional on mutation at a site, the new symbol is chosen uniformly from the remaining $q - 1$ alternatives. This defines the mutation-only dynamics. For the mutation-plus-filter condition, the same mutation step is applied, but the offspring is accepted only if its Hamming distance from the template $x^\star$ does not exceed a fixed threshold r. If the proposed offspring violates this constraint, it is rejected and the parent sequence is retained for that generation.

For each lineage and generation, we compute a reference-free complexity proxy defined as the zlib-compressed length (in bits) divided by genome length, yielding a bits-per-symbol statistic. Compression is applied directly to the raw symbolic sequence without alignment or model fitting. Multiple independent replicate lineages are simulated under each regime, and the plotted trajectories represent the mean across replicates as a function of generation.

Under mutation alone, trajectories approach the maximal entropy regime for a 4-letter alphabet (near 2 bits per symbol), reflecting convergence toward statistical typicality under repeated local perturbation. Under mutation with the viability filter, trajectories remain bounded away from this maximum, as the Hamming constraint prevents diffusion into the exponentially large typical set. The panel therefore visualizes the supply–erosion tension discussed in the main text: unconstrained mutation drives sequences toward maximal entropy, whereas even a simple structural constraint maintains bounded complexity over time.

Generation of Figure 1c,d: Simulation and theory for the discovery rate under independent per-site mutation. Panels c–d plot the expected discovery rate $\Phi(\mu)$ under the independent per-site mutation model, together with Monte Carlo (MC) estimates for selected genome lengths. We fix alphabet size q (taken as $q = 4$ in the simulations) and a novelty margin Δ (as specified in the code). For each genome length n, mutation is modeled as follows: the number of mutated sites M is distributed as $M \sim \mathrm{Binomial}(n, \mu)$, and conditional on $M = m$ and a parent sequence $X \in [q]^n$, an offspring Y is drawn uniformly from the Hamming sphere $H_m(X)$, implemented by choosing m distinct sites uniformly without replacement and replacing each selected symbol by a uniformly chosen alternative from the remaining $q - 1$ symbols. The discovery event is defined as $\Delta_Y(H_M(X)) \geq \Delta$, i.e., positive randomness deficiency at least Δ relative to the mutation-induced model class $H_m(X)$.

Theory curves. The theoretical $\Phi(\mu)$ plotted as solid lines in panel c are computed from the closed-form expression obtained by substituting the typical-case bound into the binomial mixture (Eq. (S15) in the SI), yielding

$$\Phi(\mu) \;\approx\; C\, 2^{-\Delta}(1-\mu)^n \sum_{m=1}^{n} \left(\frac{\mu}{q-1}\right)^m,$$

where C absorbs polylogarithmic factors (set to $C = 1$ for plotting), and the geometric sum is evaluated exactly up to $m = n$. Panel c displays $\Phi(\mu)$ as a function of the per-site rate μ for multiple n, showing the interior maximizer shifting as $\mu^\star \propto 1/n$.

Monte Carlo estimates. For MC points (shown where available), we estimate $\Phi(\mu)$ directly by sampling R independent replicates at each (n, μ). In each replicate we sample a parent $X \sim \mathrm{Unif}([q]^n)$, sample $M \sim \mathrm{Binomial}(n, \mu)$, and generate Y by applying an m-site mutation as above. We then evaluate the discovery indicator $\mathbf{1}\{\Delta_Y(H_M(X)) \geq \Delta\}$ using the SI criterion $\Delta_Y(H_m(X)) = K(H_m(X)) + \log|H_m(X)| - K(Y)$ with $K(H_m(X)) = K(X) + O(\log n)$ and with $K(\cdot)$ operationalized by the same code-length surrogate used throughout the manuscript (as implemented in the provided code). Averaging the indicator over replicates yields the MC estimate $\widehat{\Phi}(\mu)$.

Collapse in mutation temperature. Panel d re-parameterizes the horizontal axis by the per-genome mutation intensity ("mutation temperature") $T = n\mu = E[M]$. The same theoretical curves are replotted as $\Phi(T/n)$ versus T for each n, producing a collapse onto a common profile with a maximum near $T \simeq 1$. The dashed guide curve shows the small-μ asymptotic shape $\Phi(\mu) \propto \mu e^{-n\mu}$, i.e., $\Phi(T/n) \propto Te^{-T}$ up to a multiplicative constant, derived by expanding $(1-\mu)^n \approx e^{-n\mu}$ and truncating the geometric sum to leading order in μ (SI Lemma S2). This demonstrates that the optimal regime corresponds to an $O(1)$ expected number of mutations per genome per generation and yields the scaling $\mu^\star \approx 1/n$.