Why Evolution Operates Near One Mutation per Genome per Generation

Chattopadhyay

Abstract

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.

INTRODUCTION

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; without a continuing supply of mutation-atypical outcomes, repeated perturbation would drive lineages toward the typical set of the mutation kernel.

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 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.

We formalize exceptionality relative to the mutation-induced model on $H_m(x)$ using randomness deficiency (Definition S1 in the SI Appendix). A two-part coding/counting argument (Theorem S1 and Lemma 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. 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 plays the role of an intensive "mutation temperature" controlling the typical perturbation size. Figure 1 illustrates the scaling directly. In the left panel, $\Phi(\mu)$ 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 the right panel, plotting the same discovery rate against $T = n\mu$ collapses the 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 more naturally stated as a constant optimal per-genome perturbation size.

Our notion of discovery is defined relative to the mutation-induced model. We do not claim that the Hamming sphere is the globally optimal statistical model for an offspring, only that it is the model naturally induced by the mutation mechanism. Since mutation is the sole upstream source of variation, selection can only act on what mutation supplies. In particular, sustained accumulation of biological organization requires a nonzero supply of variants that are atypical relative to mutation alone. If offspring were always typical draws from the mutation kernel, then over many generations the population would be driven toward the typical set of that kernel, and previously accumulated structure would be eroded. Positive randomness deficiency provides a quantitative, mechanism-independent way to express this 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 $2^{-\Delta}$.

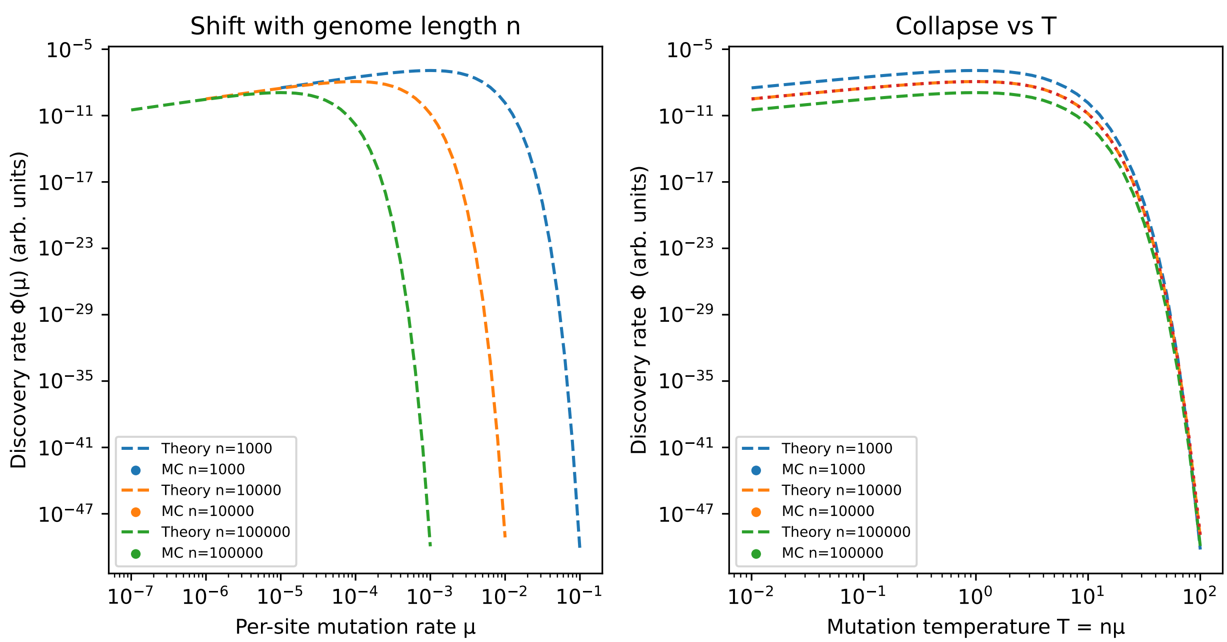

Fig. 1. Scaling and collapse of the discovery rate under independent per-site mutation. **Left:** The expected discovery rate $\Phi(\mu)$ as a function of the per-site mutation rate μ for increasing genome lengths n (theory and Monte Carlo simulation). The location of the maximum shifts left as n increases, consistent with the inverse-length scaling $\mu^* \propto 1/n$. **Right:** 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 $T \approx 1$, matching the predicted Te^{-T} dependence up to a multiplicative constant. This collapse demonstrates that the optimal regime corresponds to a constant $O(1)$ number of mutations per genome per generation.

Why focus on discovery probability?: A natural objection is to ask why evolution should be viewed as optimizing the probability of structure discovery, rather than a more direct quantity such as expected fitness gain[2–4]. We emphasize that our goal is not to model fitness, nor to assert that natural selection explicitly maximizes the objective we write down. Rather, we use discovery probability to isolate an upstream constraint imposed by mutation itself on the supply of potentially selectable variants.

Fitness effects of mutations are inherently unknown *a priori*. From the perspective of the mutation process, there is no access to a fitness landscape. Mutation generates variants blindly, and selection subsequently amplifies or suppresses them. In this setting, one can ask a counterfactual but well-posed question: given a local mutation kernel, which per-site mutation rate maximizes the expected rate at which mutation produces variants that are atypical relative to that kernel. Randomness deficiency relative to the mutation-induced model captures precisely this notion of atypicality: such variants are rare under mutation alone and therefore constitute candidates that selection can preferentially amplify if they confer advantage in a given environment.

Finally, it is 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.

This framing connects naturally to classical biology concepts. Population-genetic error-threshold arguments, such as Eigen's, constrain mutation from above by requiring heritable information to remain stable under copying errors[5]. Our result is different in kind: it identifies an *optimal* scaling for the upstream supply of rare variants under blind local perturbations. The fact that both viewpoints emphasize the $1/n$ regime suggests a convergence of two pressures, fidelity constraints that prevent loss of existing information and supply constraints that prevent evolutionary stasis. In this sense, operating near one mutation per genome per generation can be interpreted as a compromise between maintaining inherited structure and generating a steady trickle of mutation-atypical variants on which selection can act.

Beyond biological specificity: Although motivated by biological evolution, the argument presented here is not inherently biological. No assumptions are made about molecular replication, detailed population structure, or selection dynamics beyond the minimal requirement that selection acts on variants supplied by mutation. 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 variants that are atypical relative to the perturbation mechanism itself.

In this sense, Drake's rule emerges as a special case of a more general principle governing exploration under constrained generative processes. Any system in which novel structure must be discovered through local random perturbations faces an analogous tradeoff: perturbations that are too small fail to generate exceptionality at an appreciable rate, while perturbations that are too large destroy exploitable structure. The resulting optimum selects a regime in which the expected perturbation size is $O(1)$ relative to system dimension. Figure 1 makes this high-dimensional scaling visible by showing that the optimal regime is characterized by a constant intensive perturbation parameter T.

This perspective suggests that the inverse scaling of mutation rate with genome length is not an idiosyncratic feature of biology, but a manifestation of a general information-theoretic constraint on structure discovery under blind local exploration. Biological evolution realizes this constraint through molecular mutation and selection, but the same logic applies to any adaptive process operating under local stochastic perturbations.

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.

References

[1] J. W. Drake, "A constant rate of spontaneous mutation in DNA-based microbes," *Proceedings of the National Academy of Sciences*, vol. 88, no. 16, pp. 7160–7164, 1991.

[2] M. Lynch, "Evolution of the mutation rate," *Trends in Genetics*, vol. 26, no. 8, pp. 345–352, 2010.

[3] P. Sniegowski and Y. Raynes, "Mutation rates: how low can you go?" *Current Biology*, vol. 23, no. 4, pp. R147–R149, 2013.

[4] R. Hershberg, "Mutation—the engine of evolution: studying mutation and its role in the evolution of bacteria," *Cold Spring Harbor Perspectives in Biology*, vol. 7, no. 9, p. a018077, 2015.

[5] M. Eigen, "Selforganization of matter and the evolution of biological macromolecules," *Naturwissenschaften*, vol. 58, no. 10, pp. 465–523, 1971.

[6] M. Li and P. M. B. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, 4th ed. Springer, 2019.

[7] P. Gács, J. T. Tromp, and P. M. B. Vitányi, "Algorithmic statistics," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 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

$$H_m(x) = \{y \in [q]^n : d_{\mathrm{H}}(x, y) = m\}. \tag{1}$$

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:

$$K(x) \leq K(S) + \log|S| + O(1), \qquad K(x \mid S) \leq \log|S| + O(1). \tag{2}$$

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

$$c(m) = \log|H_m(x)| = \log\binom{n}{m} + m\log(q-1). \tag{3}$$

Thus, random mutation induces the two-part code

$$K(y) \;\approx\; K(x) + c(m), \tag{4}$$

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

$$\Delta_y(S) = K(S) + \log|S| - K(y). \tag{5}$$

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

Definition 1 (Net structure discovery)**.** *Fix $\Delta \geq 1$. A mutant $y \in H_m(x)$ constitutes a* net structure discovery *if $\Delta_y(H_m(x)) \geq \Delta$, 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 1 (Universal upper bound)**.** *For any $x \in [q]^n$, any m, and any $\Delta \geq 1$,*

$$\Pr_{y \sim \mathrm{Unif}(H_m(x))}\left[\Delta_y(H_m(x)) \geq \Delta\right] \leq O\left(\frac{2^{-\Delta}}{|H_m(x)|}\right). \tag{6}$$

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

$$K(y \mid x, m) = \log|H_m(x)| \pm O(1), \tag{7}$$

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

$$\Pr\left[\Delta_y(H_m(x)) \geq \Delta\right] = \Theta\left(\frac{2^{-\Delta}}{|H_m(x)|}\right). \tag{8}$$

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

$$\pi_{x,m} : \{1, \ldots, |H_m(x)|\} \to H_m(x) \tag{9}$$

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

$$K(y \mid x, m) \leq \log|H_m(x)| + O(1) \tag{10}$$

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

$$A_\Delta = \{y \in H_m(x) : K(y \mid x, m) \le \log |H_m(x)| - \Delta\}. \tag{11}$$

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

$$\Pr[y \in A_\Delta] \le 2^{-\Delta + O(1)}. \tag{12}$$

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

$$\Delta_y(H_m(x)) = \log |H_m(x)| - K(y \mid x, m) \pm O(\log n), \tag{13}$$

which yields the stated asymptotic bound. ■

Corollary 1 (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*

$$\Phi(\mu) := \mathbb{P}\left(\Delta_Y(H_M(X)) \ge \Delta\right) = \sum_{m=1}^{n} \Pr(M = m) \Pr_{y \sim H_m(X)} \left[\Delta_y(H_m(X)) \ge \Delta\right]. \tag{14}$$

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

$$\Phi(\mu) \asymp 2^{-\Delta} \sum_{m=1}^{n} \Pr(M = m) \frac{1}{|H_m(X)|}. \tag{15}$$

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

$$\Pr\left[K(X) \le n \log q - t\right] \le q^{-t} \tag{16}$$

implies that X satisfies the incompressibility condition of Lemma 1 with overwhelming probability. On this event, Lemma 1 yields

$$\Pr_{y \sim H_m(X)} \left[\Delta_y(H_m(X)) \ge \Delta\right] \asymp \frac{2^{-\Delta}}{|H_m(X)|}. \tag{17}$$

The contribution of the exceptional set $\{X : K(X) \le n \log q - t\}$ 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.

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

$$\sum_{m=1}^{n} \left(\frac{\mu}{q-1}\right)^m = \frac{\mu}{q-1} + O(\mu^2), \qquad (1-\mu)^n = e^{-n\mu + O(n\mu^2)}. \tag{18}$$

Consequently,

$$\Phi(\mu) \asymp C\, 2^{-\Delta} \frac{\mu}{q-1} e^{-n\mu + O(n\mu^2)}, \tag{19}$$

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 2 (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,*

$$\mu^\star = \frac{1 + o(1)}{n}. \tag{20}$$

Corollary 1 (Drake's rule). *Optimizing information-theoretic structure discovery under random mutation yields an inverse scaling of per-site mutation rate with genome length.*