

Why Evolution Operates Near One Mutation per Genome per Generation

xxx^a and Ishanu Chattopadhyay^a

This manuscript was compiled on January 31, 2026

Across organisms and viruses, per-site mutation rates scale inversely with genome length, yielding an approximately constant number of mutations per genome per generation. Existing explanations for this phenomenon, known as Drake's rule, invoke biochemical fidelity constraints or population-genetic error-threshold arguments that primarily bound mutation rates from above. Here we show that Drake's rule emerges from a mechanism-independent information-theoretic constraint on blind local perturbation. Treating mutation as a generative process upstream of selection, we formalize evolutionary novelty as statistical exceptionality relative to mutation alone. We show that the rate at which mutation supplies such exceptional variants is maximized when the expected number of mutations per genome per generation is order one, implying an optimal per-site mutation rate scaling as $1/n$, where n is genome length. This result does not rely on fitness landscapes, population structure, or biochemical details, and reflects a general principle governing structure discovery under local random perturbations.

entropy rate | AI-generated text detection | large language models | algorithmic complexity | probabilistic automata

Mutation introduces randomness; selection preserves structure. Evolution therefore operates under a fundamental tension: mutation must generate novelty without overwhelming inheritable organization. A striking empirical regularity across biological systems is that the per-site mutation rate μ scales approximately as $1/n$, where n is genome length, so that the per-genome mutation rate

$$U = n\mu$$

is approximately constant. This phenomenon, observed across DNA- and RNA-based organisms spanning many orders of magnitude in genome size, is known as Drake's rule.

Prevailing explanations for Drake's rule fall into two broad classes. Mechanistic accounts attribute mutation rates to biochemical replication fidelity and energetic costs, while population-genetic theories emphasize error thresholds beyond which inherited information becomes unstable. These approaches provide important constraints, but they primarily explain why mutation rates cannot be too large. They do not explain why evolution consistently operates near a particular scaling regime across diverse biological implementations.

Here we pursue a complementary, mechanism-independent explanation by treating mutation as a blind generative process upstream of selection. Mutation has no access to fitness; it samples variants locally around a parent genome. Selection can only act on the statistical properties of the variants mutation supplies. This motivates a minimal question: which mutation rates maximize the supply of variants that are exceptional relative to mutation alone, and therefore capable of serving as candidates for selection-driven amplification?

Mutation geometry and exceptionality

Let $x \in [q]^n$ denote a length- n genome over an alphabet of size $q \geq 2$. Mutation at Hamming radius m produces variants in the Hamming sphere

$$H_m(x) = \{y \in [q]^n : d_H(x, y) = m\},$$

whose cardinality is

$$|H_m(x)| = \binom{n}{m} (q-1)^m.$$

Author affiliations: ^aUniversity of Kentucky

RS carried out experimental runs and wrote the paper, IC conceived of research, implemented the algorithm, wrote the paper and procured support.

Authors have no competing interests.

²To whom correspondence should be addressed. E-mail: ishanu.ch@uky.edu

As the number of mutated sites increases, the mutation-accessible neighborhood grows combinatorially, diluting any structured or special outcome among an exponentially expanding set of possibilities.

Random mutation induces a canonical two-part description of its outcomes: first specify the mutation-accessible model $H_m(x)$, then specify an index within that model. The associated indexing cost is

$$c(m) = \log |H_m(x)|.$$

For a uniformly drawn mutant $y \sim \text{Unif}(H_m(x))$, the typical description length satisfies, up to lower-order terms,

$$K(y) \approx K(x) + c(m),$$

where $K(\cdot)$ denotes prefix-free Kolmogorov complexity (Supporting Information).

We formalize evolutionary novelty as exceptionality relative to this mutation-induced model. A mutant y constitutes a *net structure discovery* at margin $\Delta \geq 1$ if

$$K(y) \leq K(x) + c(m) - \Delta.$$

Equivalently, y has positive randomness deficiency at least Δ relative to $H_m(x)$. This definition captures a minimal requirement for adaptive potential upstream of selection: a variant must be rare under mutation alone in order to plausibly serve as a candidate for selective amplification.

A key consequence is that the probability of such exceptionality decays inversely with the size of the mutation-accessible neighborhood. For typical genomes (Supporting Information),

$$\Pr_{y \sim \text{Unif}(H_m(x))} [\text{net discovery at margin } \Delta] \propto \frac{2^\Delta}{|H_m(x)|},$$

up to polylogarithmic factors. Thus, as mutation radius increases, exceptional outcomes become rapidly vanishingly rare.

Optimal mutation rate

In biological evolution, the mutation radius itself is random. Let

$$M \sim \text{Binomial}(n, \mu)$$

denote the number of mutated sites per generation under independent per-site mutation probability μ . For fixed novelty margin Δ , define the expected discovery rate

$$\Phi(\mu) = \sum_{m=1}^n \Pr(M = m) \Pr_{y \sim \text{Unif}(H_m(x))} [\text{net discovery at margin } \Delta]$$

Substituting the inverse-neighborhood scaling above and the identity

$$\Pr(M = m) = \binom{n}{m} \mu^m (1 - \mu)^{n-m},$$

the combinatorial factor $\binom{n}{m}$ cancels with the corresponding term in $|H_m(x)|$, yielding the leading-order form (Supporting Information)

$$\Phi(\mu) \propto (1 - \mu)^n \sum_{m \geq 1} \left(\frac{\mu}{q - 1} \right)^m \approx C \mu e^{-n\mu},$$

for $\mu \ll 1$ and fixed q , where C absorbs constants and lower-order factors.

The optimization is immediate: the function $\mu e^{-n\mu}$ is maximized at

$$n\mu = 1.$$

Thus the optimal per-site mutation probability scales as

$$\mu^* = \frac{1 + o(1)}{n},$$

equivalently yielding an optimal per-genome mutation rate $U^* \approx 1$. This recovers Drake's rule as a consequence of maximizing the supply of mutation-generated exceptional variants under blind local perturbation.

Discussion

Our analysis reframes Drake's rule as an upstream supply constraint imposed by mutation, rather than as a downstream equilibrium determined by selection. Mutation cannot optimize fitness directly; it can only optimize the statistical properties of the variants it supplies. Maximizing the rate at which mutation produces variants that are exceptional relative to its own generative mechanism provides a natural objective under this constraint.

This explanation does not rely on biochemical replication fidelity, fitness landscapes, population structure, or mutation-selection balance. Instead, it depends only on the geometry of local random perturbations in high-dimensional discrete spaces and on the requirement that adaptive novelty be rare under mutation alone. Although motivated by genetic evolution, the argument applies broadly to any adaptive system driven by blind local exploration, in which perturbations that are too small fail to generate novelty and perturbations that are too large destroy exploitable structure.

Conclusion

We show that Drake's rule emerges as a general information-theoretic constraint on blind local exploration. Mutation optimally supplies selection with novel, structured variants when the expected number of mutations per genome per generation is order one. This result transcends biological detail and reflects a universal principle governing structure discovery under uncertainty.

Materials and Methods

Data acquisition. Human-authored texts were obtained from Project Gutenberg (English-language compositions) and technical manuscripts from arXiv via the official OAI-PMH interface. AI-generated texts were produced via webform and API access to the LLMs listed in Table ?? . Details on prompt design, API access, and NERO computation are provided in the Supplementary Methods.

Code availability. A reproducible implementation of the entropy-rate estimator and end-to-end evaluation pipeline is available under an MIT license at <https://github.com/zeroknowledgediscovery/nero>.

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