Eigen's error threshold: A concise derivation

Jacob Fine

Nov 27th, 2024

The purpose of this exercise is to provide a concise demonstration of how Eigen's error threshold is obtained based on the mutation rate and length of a polymer undergoing self-replication. This was proposed in physical chemist Manfred Eigen's landmark paper 'Self-organization of matter and evolution of biological macromolecules' (Eigen 1971).

Suppose we have a polymer of N nucleotides replicating with an mutation rate μ. At some replication step, the probability of replicating the sequence without error is given by

$$q = (1 - \mu)^N$$

If μ is small, we can assume that, based on the Taylor series for $\ln(1 - x)$, that

$$\ln(1 - \mu) \approx -\mu$$

Multiple both sides by N to obtain

$$N\ln(1 - \mu) \approx -N\mu$$

$$(1 - \mu)^N \approx e^{-N\mu}$$

We can assume that if μ is small, than the probability of replicating a sequence with no error at a given step is therefore given by

$$q = e^{-N\mu}$$

This can also be interpreted as the expected ratio in abundance of 'master sequences' (the sequence the process began with) to 'mutated sequences' (any sequence that deviates by at leasat one base from the master sequence) that will remain after the master sequence undergoes an arbitrarily large number of replication events. Therefore

$$q = \frac{\text{number of master sequences after many reps}}{\text{number of mutated sequences after many reps}}$$

The fitness of the master sequence σ can be defined as the number of master sequences produced by the master sequence relative to mutated sequences, and is an intrinsic property of the copying of the sequence (i.e., its ability to produce offsrping), therefore

$$\sigma = \frac{\text{number of master sequences per rep}}{\text{number of mutated sequences per rep}}$$

An 'error catastrophe' occurs when the mutation rate is large enough, such that q is too small to maintain the population of master sequences. Here, the master sequence becomes outcompeted by mutant sequences. To prevent an error catastrophe, the fitness of the master sequence must be larger than the ratio of mutated sequences to master sequences that would be obtained after many replication events. Therefore

$$\sigma > 1/q$$

And since $q = e^{-N\mu}$, we can express the 'error threshold' as

$$\sigma > e^{N\mu}$$

which gives an upper bound on mutation rate as a function of fitness and polymer length

$$\mu < \frac{ln(\sigma)}{N}$$

This result has key implications on the origin of life, which likely places constraints on the length of the first 'genes' as well as the mutation rate of the first replicases.