In Vitro Evolution

Spiegelman: Evolution in the Qβ-virus model

Spiegelman used Qβ-virus that attacks E.coli to demonstrate evolution in the lab. Upon infection, the virus first translates its RNA into a complementary strand which is used to get the original RNA. This replication could be shrunk in a test tube with only three components: the RNA of the virus, the Qβ-Replicase enzyme and activated nucleotides.

Spiegelman: Evolution in the Qβ-virus model

Spiegelman used Qβ-virus that attacks E.coli to demonstrate evolution in the lab. Upon infection, the virus first translates its RNA into a complementary strand which is used to get the original RNA. This replication could be shrunk in a test tube with only three components: the RNA of the virus, the Qβ-Replicase enzyme and activated nucleotides.

It could be shown:

- Shorter and shorter RNA was created when the selection pressure was on fast replication: sequence information needed to infect E.coli was removed over time.
- Resistance against high temperatures and RNA degrading enzymes could be bred.
- Replication could be even triggered without an initial RNA template (Eigen et.al.).

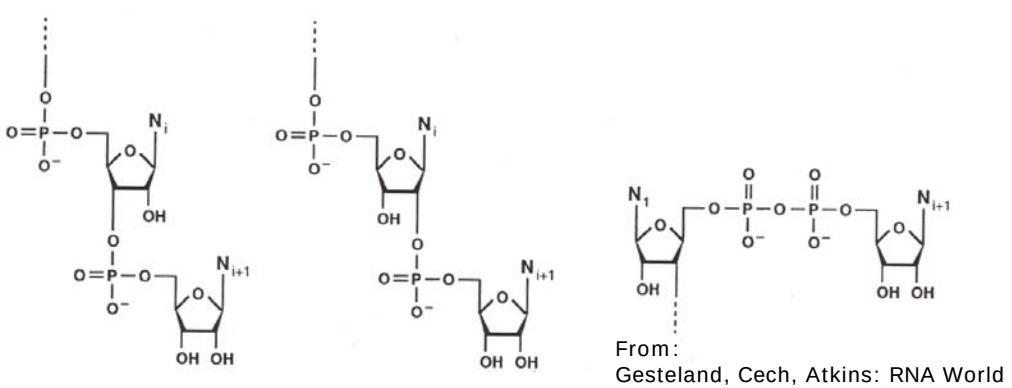
Spiegelman: Evolution in the Qβ-virus model

Spiegelman used Qβ-virus that attacks E.coli to demonstrate evolution in the lab. Upon infection, the virus first translates its RNA into a complementary strand which is used to get the original RNA. This replication could be shrunk in a test tube with only three components: the RNA of the virus, the Qβ-Replicase enzyme and activated nucleotides.

It could be shown:

- Shorter and shorter RNA was created when the selection pressure was on fast replication: sequence information needed to infect E.coli was removed over time.
- Resistance against high temperatures and RNA degrading enzymes could be bred.
- Replication could be even triggered without an initial RNA template (Eigen et.al.).

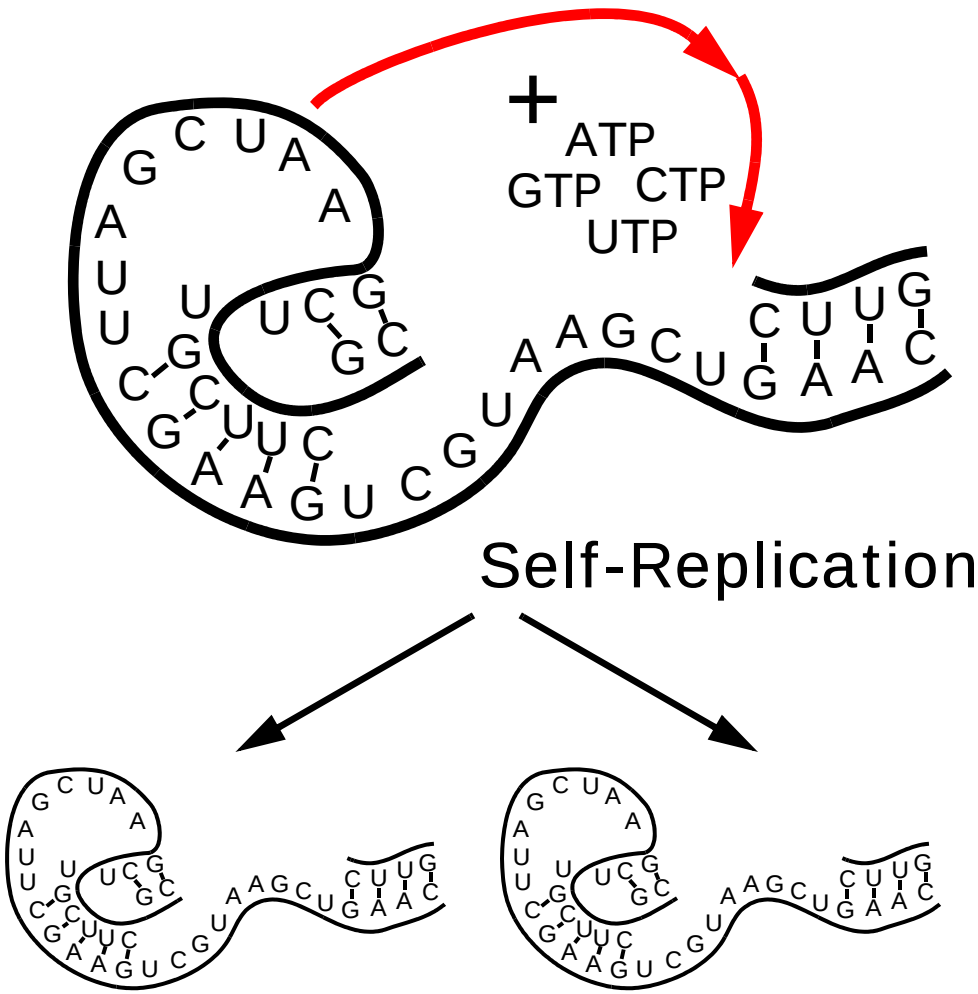
Orgel in the 70s pushed it further by dropping the enzyme. Its existence is anyhow unrealistic for an origin of life scenario. He searched for direct replication on a single stranded RNA. Some short complementary strands could be found in favorable and peculiar environmental conditions, but the polymerization did not produce a correct 3'-5' backbone of the complementary RNA molecule:



From:
Gesteland, Cech, Atkins: RNA World

Eigen established a model for approaches where the RNA itself catalyses the replication

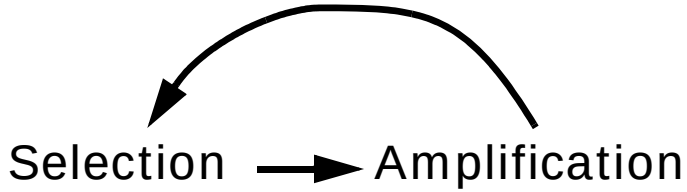
RNA World

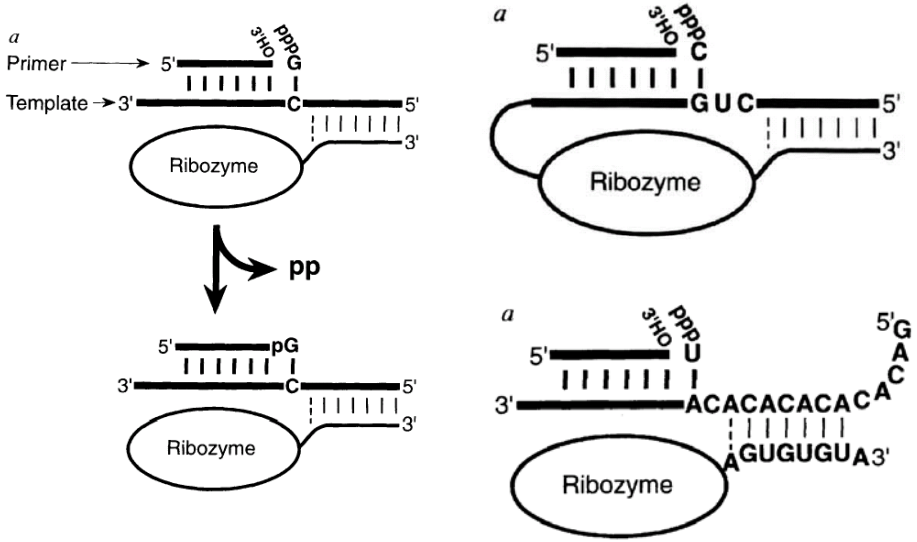


Based on the antiquity of Rybozymes it is argued that RNA might have been the earliest “protein” to replicate, i.e. to replicate itself.

The idea is that dNTPs can be polymerized into a polymerase molecule, allowing for self sustained darwinian evolution within one molecule.

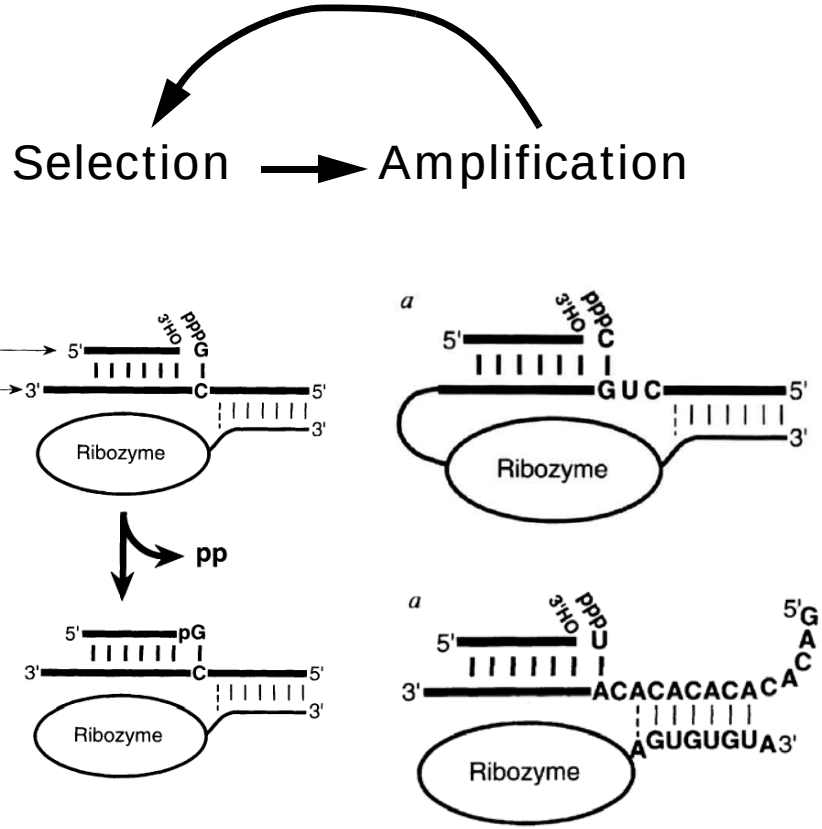
Testing the RNA world: in vitro selection





Replicating for the first base has a fidelity of >95%. However not much more than three bases can be added. The complementary strand shows no catalytic potential.

Testing the RNA world: in vitro selection



Replicating for the first base has a fidelity of >95%. However not much more than three bases can be added. The complementary strand shows no catalytic potential.

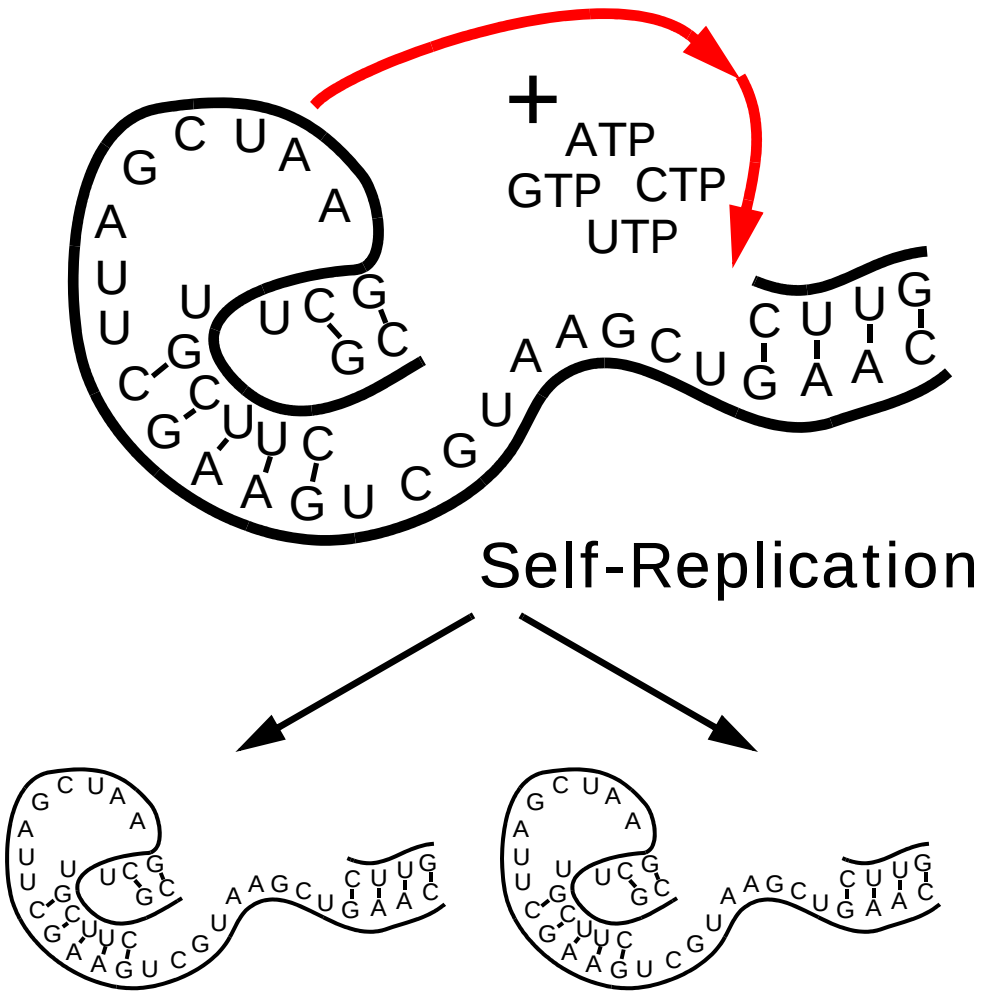
Starting from an already engineered pool of sequences, one can indeed find RNA molecules that can at 1-3 activated base pairs to the end of an already double stranded RNA piece (Ekland&Bartel, 1996) (Szostak&Bartel 1993).

The reaction goes through many cycles of selection and amplification. Note that no high mutation is built into the process. This process can amplify the advantageous molecule exponentially. It needs good balance of selection pressure and amplification strength. This scheme was used in many different areas of biochemistry to 'breed' molecules (note there often is no mutation) like we bred animals. Check

www.pubmed.org for publications

Models of Chemical Evolution

Eigen & Schuster: Selfreplicative Molecules



Eigen and Schuster applied population dynamics to self-replicating molecules, for example to RNA strands.

The idea is that there might exist RNA molecules that can duplicate themselves from activated (and unstable) Tri-Phosphate nucleotides (dNTP). Eigen&Schuster studied these systems in theory.

The equations are easily understood based on standard population dynamics.

Basic Population Dynamics

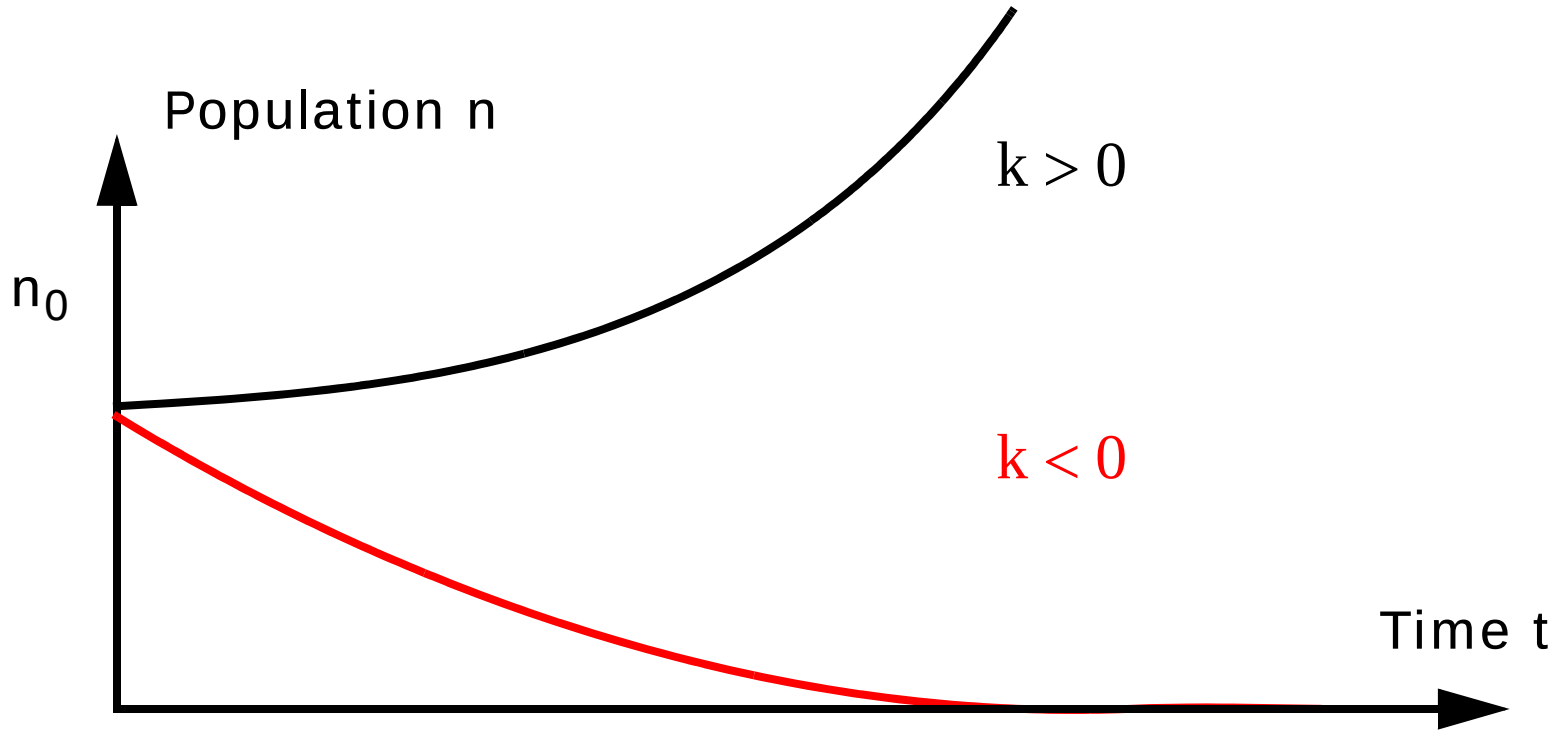
$$\dot{n} = \frac{dn}{dt} = kn$$

$$k = k_p - k_m$$ Units of k: 1/s

$$n = n_0 e^{kt}$$

Basics:

Most basic is the exponential growth or decay law with a rate k. Based on a propagation rate k_p and a mortality rate k_m the population of a species either grows or shrinks exponentially.

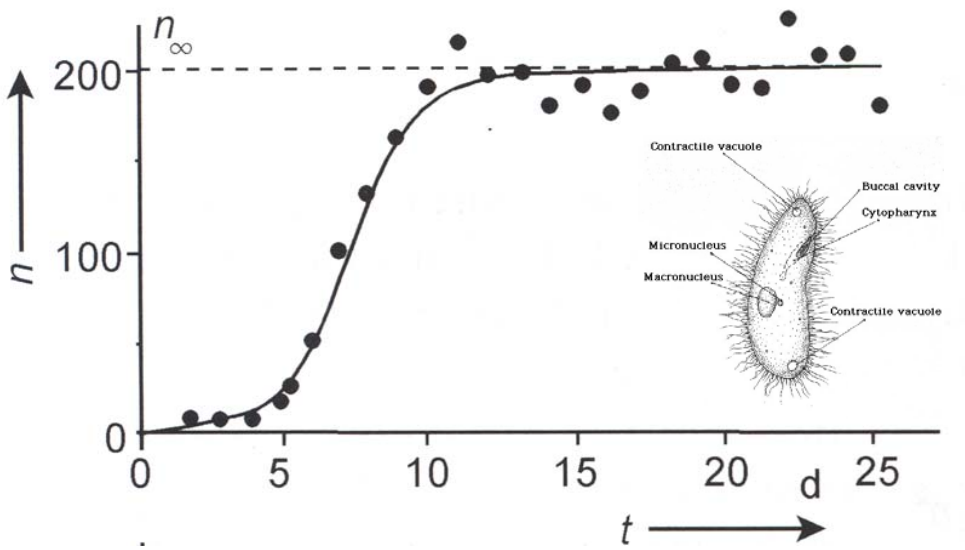


Basic Population Dynamics

$$\dot{n} = (k_p - k_m n)n$$

$$n = \frac{n_0 k_p e^{k_p t}}{k_p + k_m n_0 (e^{k_p t} - 1)}$$

$$n \to \frac{k_p}{k_m}$$



From: Roland Glaser: Biophysics

Basics:

Most simple models of populations start with a first order rate equation of number of entities n with a propagation rate k_p. It leads to exponential growth for positive k_p and to exponential decline for negative k_p.

Most simple way to implement a limitation of the growth, for example by limiting the food available, is to introduce a negative, quadratic term. It implements a mortality term k_m*n which depends on the number of entities.

This most simple approach has the logistic function as solution and can be used for example to model the growth of paramecium, a single cellular organism (Fig. left).

Basic Population Dynamics

Food limited growth

$$\dot{n} = (k_p - k_m n)n$$

Competition for food

$$\dot{n}_1 = (k_{p1} - k_{m1}(n_1 + n_2))n_1$$

$$\dot{n}_2 = (k_{p2} - k_{m2}(n_1 + n_2))n_2$$

Competition Models

Basic competition models involve competing for food or a hunter-prey relationship.

Competition for food is the simple expansion of the previous limited growth model.

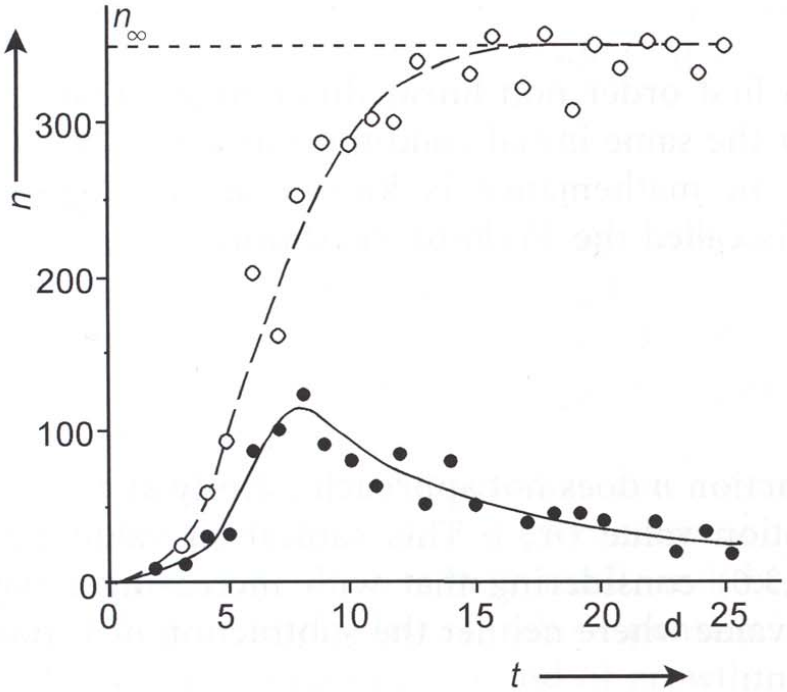
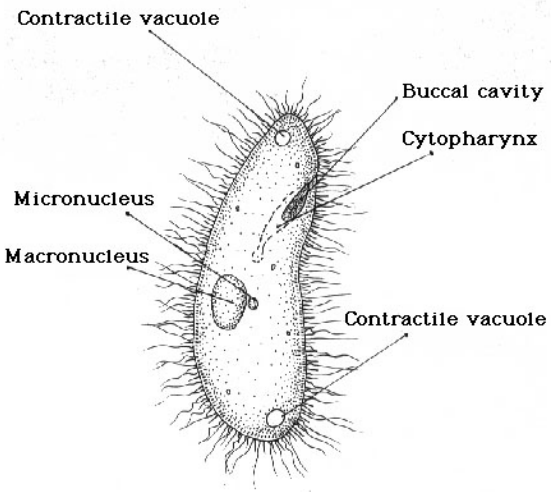


Fig. 5.18. The time dependence of the growth of a pure population of *Paramecium caudatum* (•——• above) and mixed together with a population of *Paramecium aurelia* (○- - -○). The ordinate indicates the number of individuals (n) per 0.5 ml medium. The abscissa gives the time in days (d). (Data from Gause 1935)



From:
Roland Glaser: Biophysics

Basic Population Dynamics

Hunter-Prey relationship (Volterra-Lotka)

$$\dot{n}_P = (k_{pP} - k_{mP} n_H) n_P$$

$$\dot{n}_H = (k_{pH} n_P - k_{mH}) n_H$$

The same logic is applied to hunter-prey relationships, studied in the 1930 by Volterra and Lotka. The mortality rate of the prey depend on the number of hunters and the propagation of hunters depends on the number of prey (see left). These systems can show oscillations in population also recorded in nature (below).

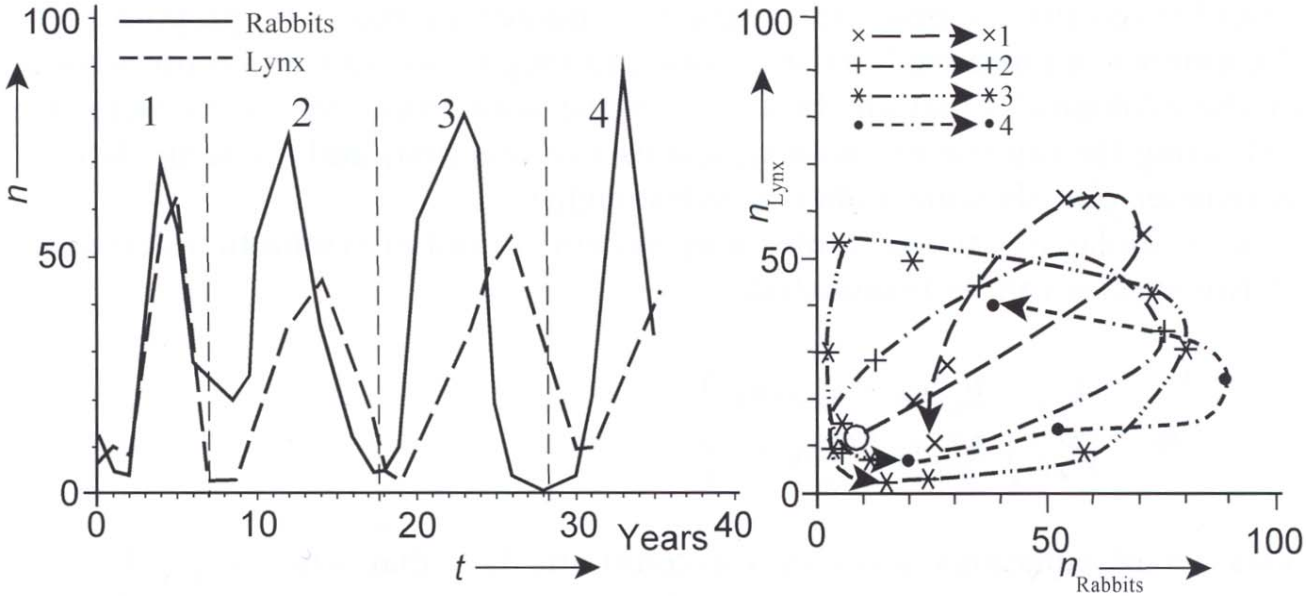


Fig. 5.19. A particular example of population kinetics: the number of pelts, obtained from hunted lynx and snow rabbits in Canada after 1900 (numbers in thousands). On the *left*: time course; on the *right*: the same numbers in a Volterra-plot. The population waves 1–4 of the *left* part of this figure appear as Volterra cycles in the graph on the *right*, starting with the point ○. (Data from Haken 1983)

Simplifications are for example:

- no hunter-hunter interaction
- everything is linearized
- no time delays due to birth/development

From:
Roland Glaser: Biophysics

Eigen & Schuster: Selfreplicative Molecules

$$\dot{n}_i = (k_{pi} q_i - k_{mi}) n_i + \sum_{i \neq j} k_{m,ji} n_j$$

Take molecules the number of molecules n_i of sequence i (also called species) and introduce the following population dynamics:

k_{pi} Propagation rate,
i.e. the ability to self-replicate

k_{mi} Mortality rate and rate to leave the area

$q_i < 1$ Replication fidelity

$k_{m,ji}$ Probability to mutate from another species j.

From: Gesteland, Cech, Atkins: RNA World

Eigen & Schuster: Selfreplicative Molecules

$$\dot{n}_i = (k_{pi} q_i - k_{mi}) n_i + \sum_{i \neq j} k_{m,ji} n_j$$

Take molecules the number of molecules n_i of sequence i (also called species) and introduce the following population dynamics:

k_{pi} Propagation rate,
i.e. the ability to self-replicate

k_{mi} Mortality rate and rate to leave the area

$q_i < 1$ Replication fidelity

$k_{m,ji}$ Probability to mutate from another species j.

$$\sum_i n_i = \text{const}$$

Note that we enforce additionally a limited total number of molecules.

To make things more realistic, one can group the species into similar and non-competing groups, called quasi-species

From: Gesteland, Cech, Atkins: RNA World

Eigen & Schuster: Error Threshold

$$\text{Average} \quad \langle x \rangle = \frac{\sum_{i \neq j} x n_j}{\sum_{i \neq j} n_j}$$

$$\sigma_i = \frac{k_{pi}}{\langle k_p \rangle - \langle k_m \rangle + k_{mi}} > \sigma_j$$

From the Eigen model one can derive the replication fidelity of the best molecule to be able to outgrow all the others. It depends crucially on the length of RNA molecule used.

One can derive that the sequence i is able to outgrow its competitors, if it has a superiority σ_i larger than all others.

From: Gesteland, Cech, Atkins: RNA World

Eigen & Schuster: Error Threshold

$$\text{Average} \quad \langle x \rangle = \frac{\sum_{i \neq j} x n_j}{\sum_{i \neq j} n_j}$$

From the Eigen model one can derive the replication fidelity of the best molecule to be able to outgrow all the others. It depends crucially on the length of RNA molecule used.

$$\sigma_i = \frac{k_{pi}}{\langle k_p \rangle - \langle k_m \rangle + k_{mi}} > \sigma_j$$

One can derive that the sequence i is able to outgrow its competitors, if it has a superiority σ_i larger than all others.

$$q_i = q^N > \frac{1}{\sigma_i}$$

This superiority is linked to the replication fidelity of a single base q by the formula to the left.

$$N < \frac{-\ln \sigma_i}{\ln q}$$

This gives a length limitation N depending on the fidelity of replication of the best molecule.

From: Gesteland, Cech, Atkins: RNA World

Eigen & Schuster: Targeting the best

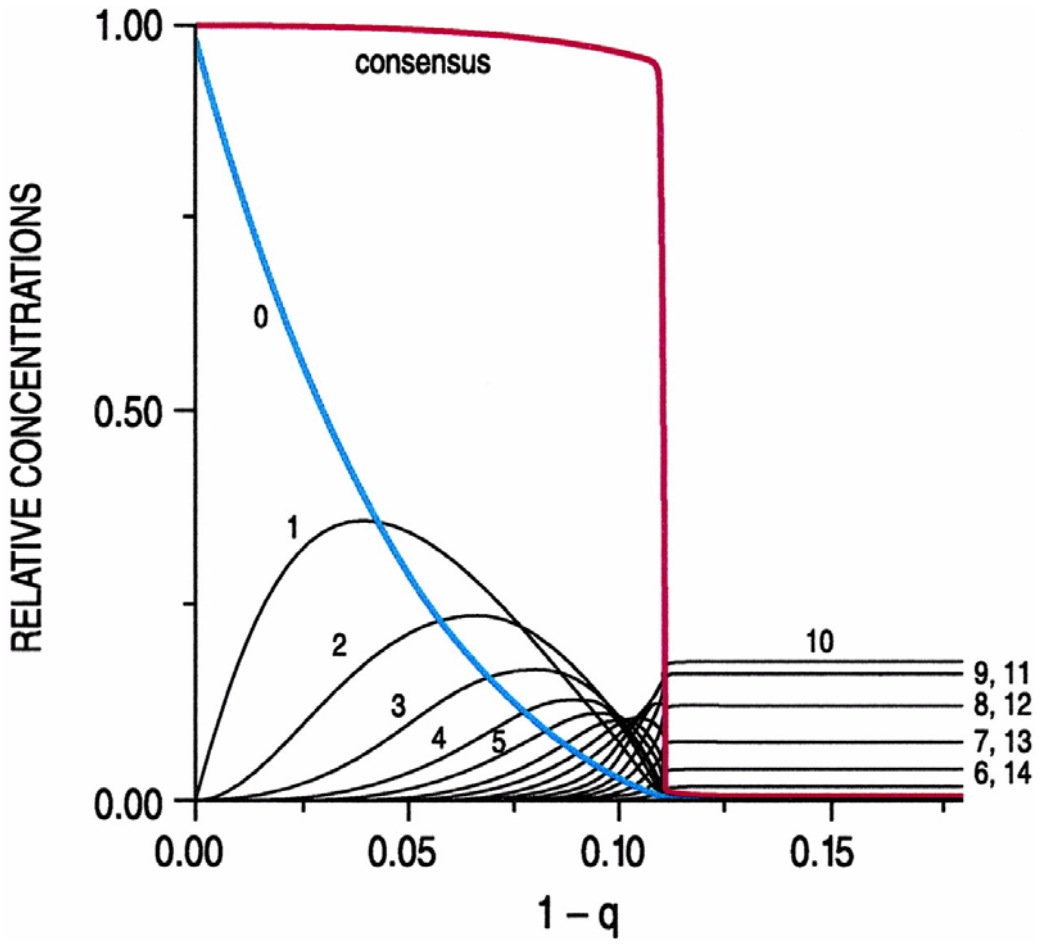


Fig. 1. Relative population numbers of binary sequences S_k (ordinate) as functions of single-digit error rate (1-q) (abscissa). The length of all binary sequences is N = 20. All $2^N=10^6$ sequences are degenerate in their reproductivity except for one "master" sequence S_m, which reproduces 10 times more efficiently than the rest. The resulting quasispecies distribution is centered at the master sequence ("0" errors). The numbers 1, 2, ... 20 refer to the sum of all 1-, 2-, ... 20-error mutants. The red curve refers to the consensus sequence, which shows a sharp first-order phase transition at the error threshold.. From PNAS 99: 13374-13376 (2001).

Replication Fidelity and sequence length

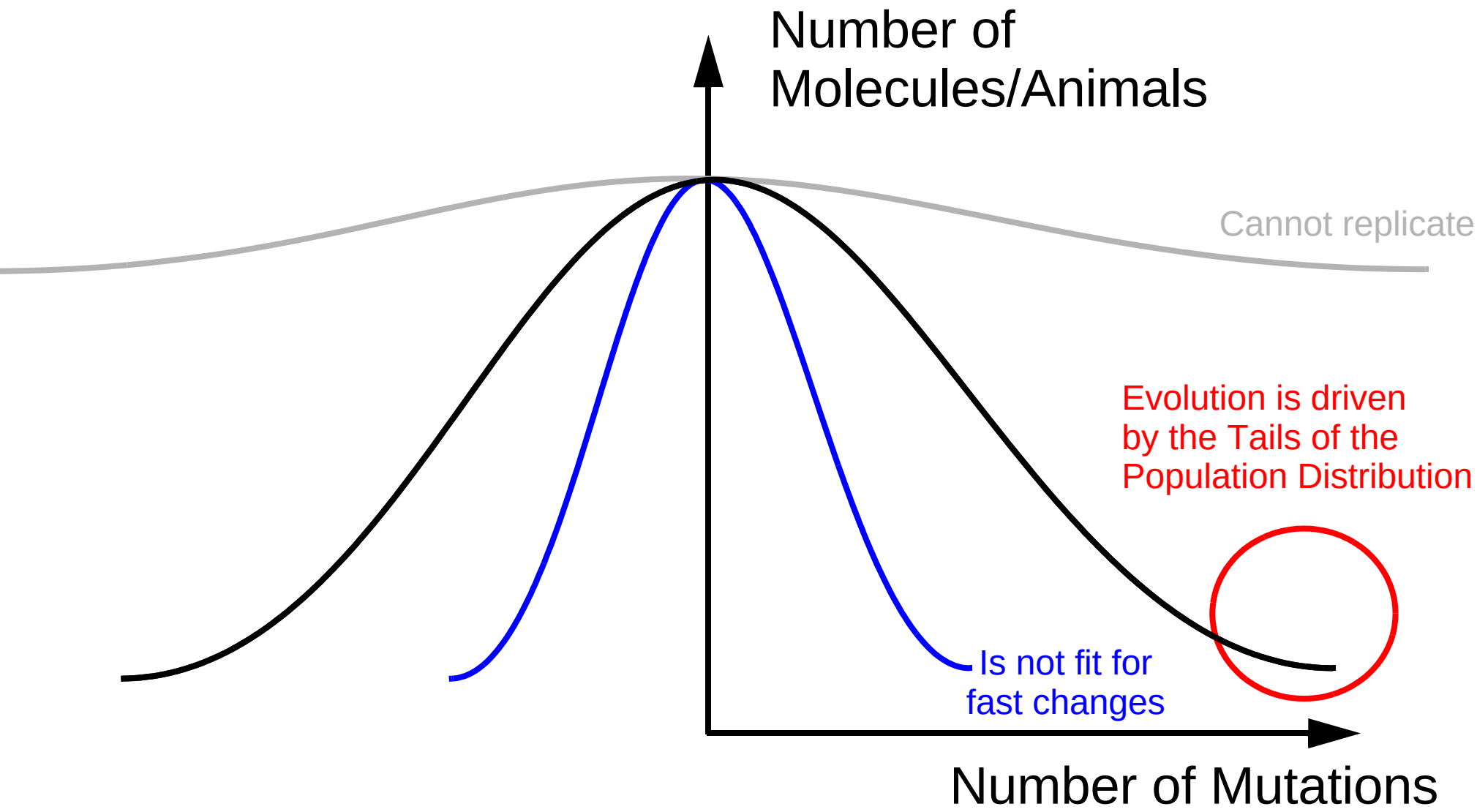
The law that the error of replication depends on the sequence length can be found also for microbes and other animals

Table 4.1 Rates of spontaneous mutation in DNA-based microbes

Organism	Genome size (bp)	Target	Mutation rate: Per bp (μ_{bp})	Mutation rate: Per genome (μ_g)
Bacteriophage M13	6.41×10^3	*lacZα*	7.2×10^{-7}	0.0046
Bacteriophage λ	4.85×10^4	*cI*	7.7×10^{-8}	0.0038
Bacteriophage T2	1.60×10^5	*rII*	2.7×10^{-8}	0.0043
Bacteriophage T4	1.66×10^5	*rII*	2.0×10^{-8}	0.0033
Escherichia coli	4.70×10^6	*lacI*	4.1×10^{-10}	0.0019
			6.9×10^{-10}	0.0033
		his GDCBHAFE	5.1×10^{-10}	0.0024
Saccharomyces cerevisiae	1.38×10^7	*URA3*	2.8×10^{-10}	0.0038
		SUP4	(7.9×10^{-9})	(0.11)
		CANI	1.7×10^{-10}	0.0024
Neurospora crassa	4.19×10^7	*ad-3AB*	4.5×10^{-11}	0.0019
		mtr	(4.6×10^{-10})	(0.019)
			1.0×10^{-10}	0.0042

From: Gesteland, Cech, Atkins: RNA World

Population Genetics



Survival of the Flattest

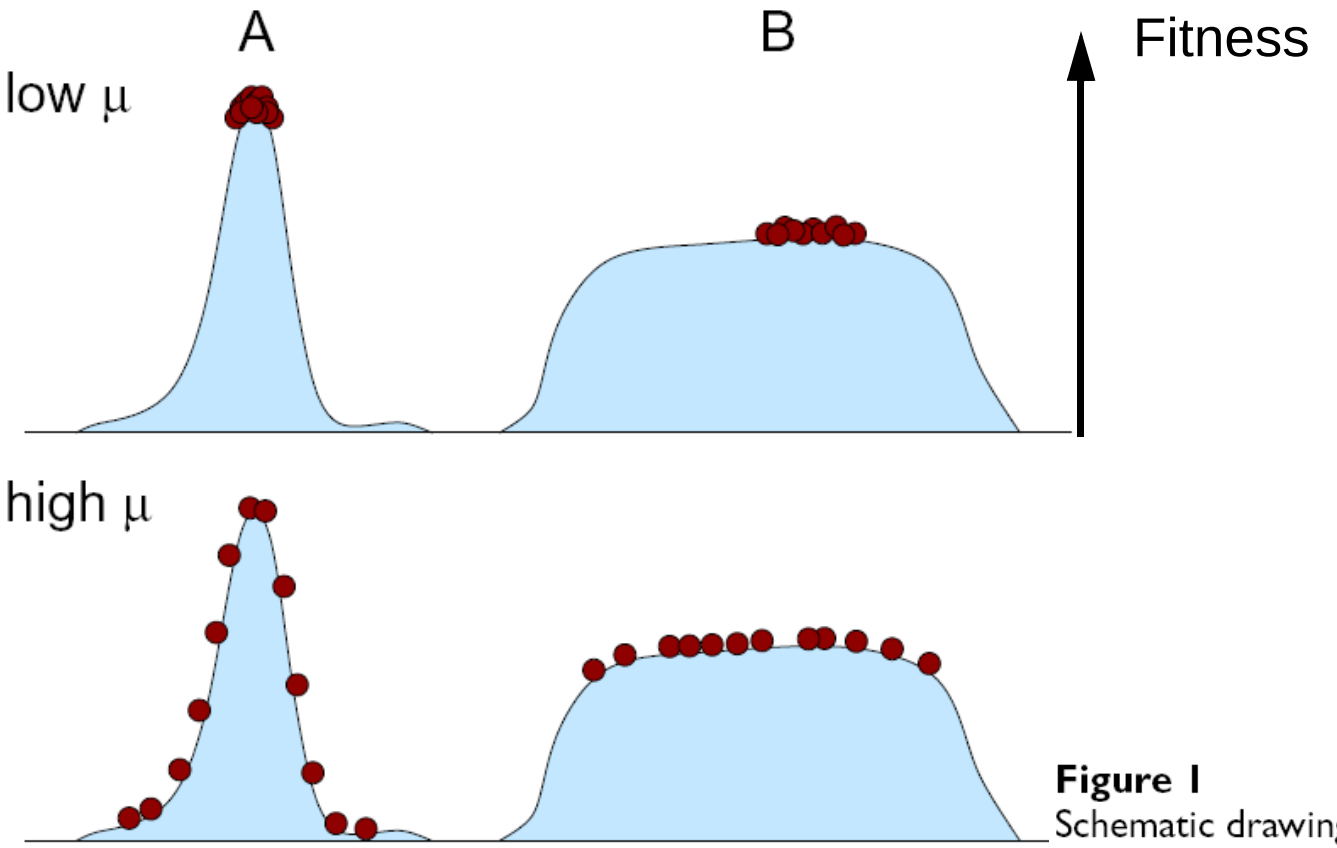


Figure 1
Schematic drawing of the survival-of-the-flattest effect. At low mutation rate μ, all individuals accumulate close to the top of the local fitness peak, and hence the individuals on peak A outcompete the individuals on peak B. At high mutation rate, most individuals on the steep peak A are located at low fitness values, while the individuals on the flat peak B remain close to the local optimum. As a consequence, the mean fitness of the individuals on peak B exceeds that of the individuals on peak A, and thus the former outcompete the latter.

From Wilke, http://www.biomedcentral.com/1471-2148/5/44

Critisism: Combinational Explosion

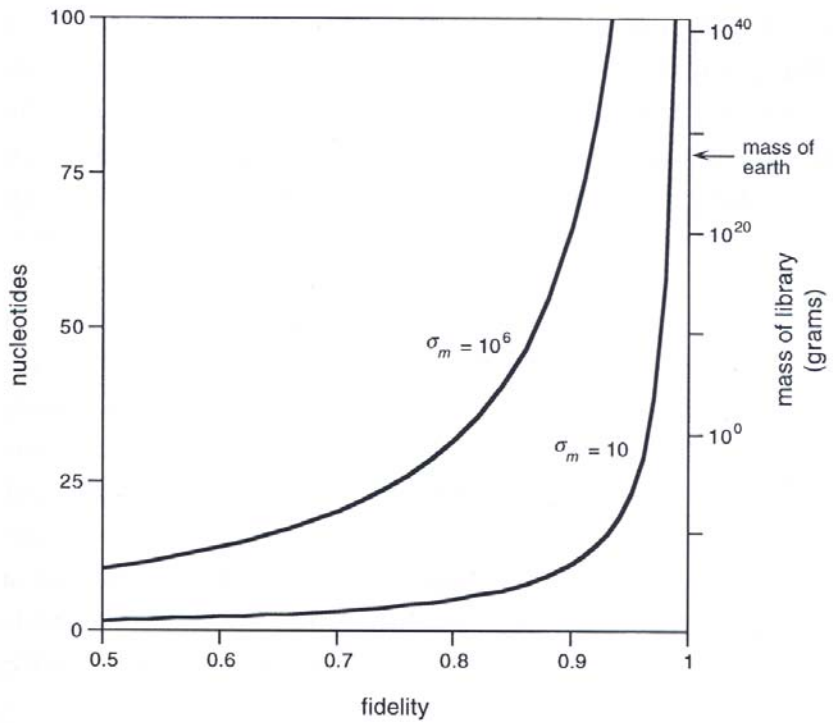


Figure 3 The Eigen error threshold, relating fidelity of replication to the maximum number of nucleotides, for two different values of superiority (σ_m). Fidelity (q) is that of the component condensation reactions in the replication cycle. The number of condensation reactions (ν) is related to the number of nucleotides (n) by: $\nu = 2n - 2$. Scale at right indicates the mass of a combinatorial library containing one copy each of the 4^n possible sequences of length n.

From: Gesteland, Cech, Atkins: RNA World

Let us start with molecules of 500 bases, i.e. with 1000bits of information. The starting pool to have each molecule at least once are $2^{1000}=10^{300}$ molecules. However we have only 10^{32} molecules on the whole hydrosphere of the earth!

So let us be optimistic and assume a 40mer can already do it with a superiority of $\sigma_i=10^3$ and a replication fidelity of q=0.9. The complete library are 1kg. However the complementary part has also to be catalytic and also replicate itself!

This argument leads to the idea of hypercycles.

Eigen & Schuster: Hypercycles

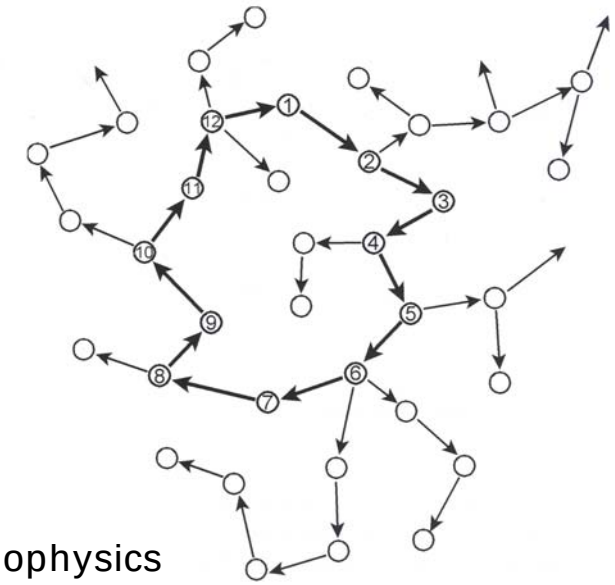
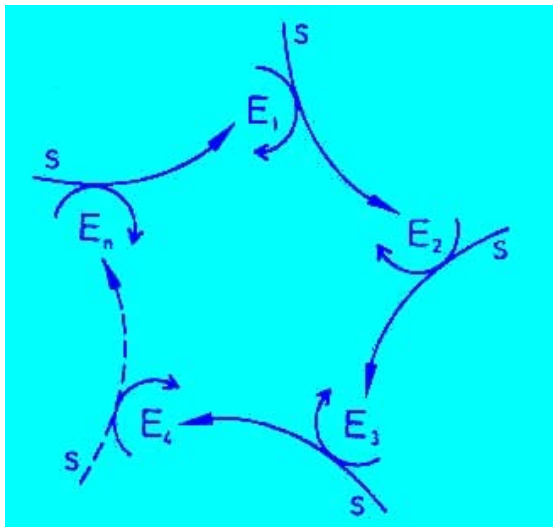
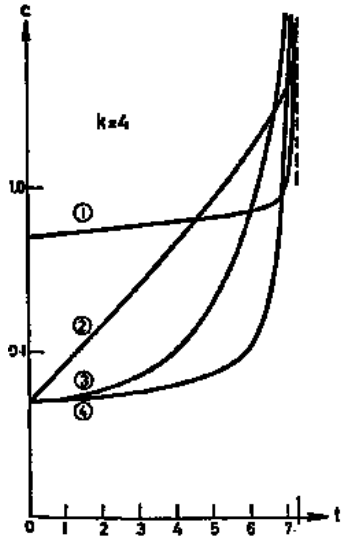


Fig. 5.21. The network of catalytic correlations in a mixture of polymers. The points with numbers connected by *thick arrows* indicate an auto-catalytic cycle

From:
Roland Glaser: Biophysics





Hypercycles are formed from molecules which do not replicate themselves, but other molecules. More specifically, RNA i codes for a protein or Rybozyme to replicate the RNA i+1.

If they **form a loop that eventually also replicate the first molecule**. Comparable to a bee and a flower: each helps replicating the other. It is beneficial to isolate each hypercycle into “chambers”.

Depending on the parameters, hypercycles can lead to linear, exponential and hyperexponential growth, yielding ’once-for-all’ selection.

However:

- The cycles are unstable for >4 cycles
- Problem of ’selfish’ viral mole-

From: Erich Sackmann, Biophysics Script, Ch. 19

Eigen's Paradox: Proteins cannot be reached

Hypercycles show the strong potential of coupled catalysis. Especially heterogeneous cycles (proteins & nucleic acids) appear to have an advantage in the simulations. However we are far away from establishing such heterogeneous hypercycles since the sequence length is much too long. Or formulated as Eigen's paradox:

No enzyme can be produced without a large genome
No large genome can be replicated without the help of enzymes.

References for Hypercycles:

M. Eigen, P.Schuster, "The Hypercycle", Springer-Verlag, 1979
M. Eigen, "Steps towards life: A Perspective on Evolution", Oxford Univ Press, 1996
J. Hofbauer, K. Sigmund, "The theory of evolution and dynamical system", Cambrige Univ Press, 1988
J. M. Smith, "Hypercycles and the origin of life", Nature 20:445-6, 1979
J. M. Smith, E. Szathmary "The major transitions in evolution", W.H. Freeman, 1995

.... yet still the starting pool gives a lot of headaches.