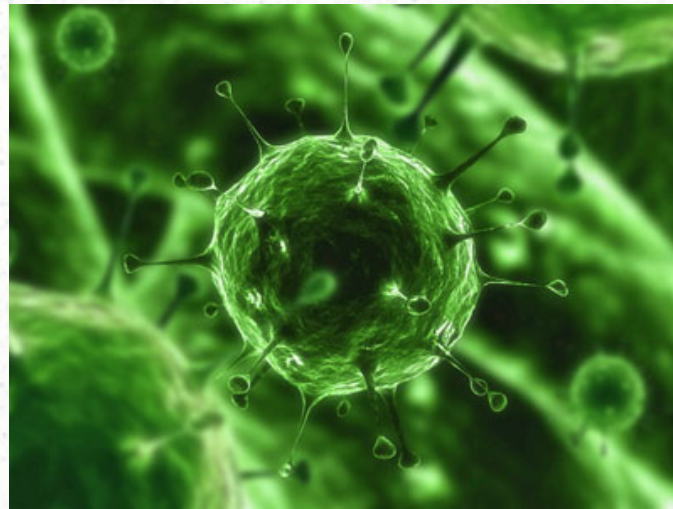


Quasi-species models of viral evolution



José A. Cuesta

Grupo Interdisciplinar de Sistemas Complejos
Universidad Carlos III de Madrid
Spain



Universidad
Carlos III de Madrid

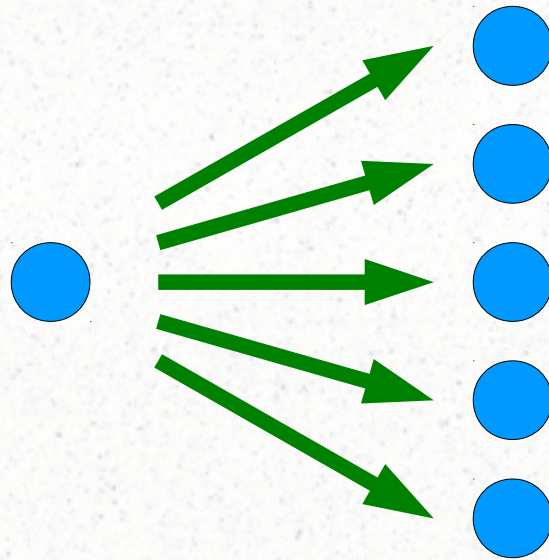
Building blocks of evolution

Replication

Selection

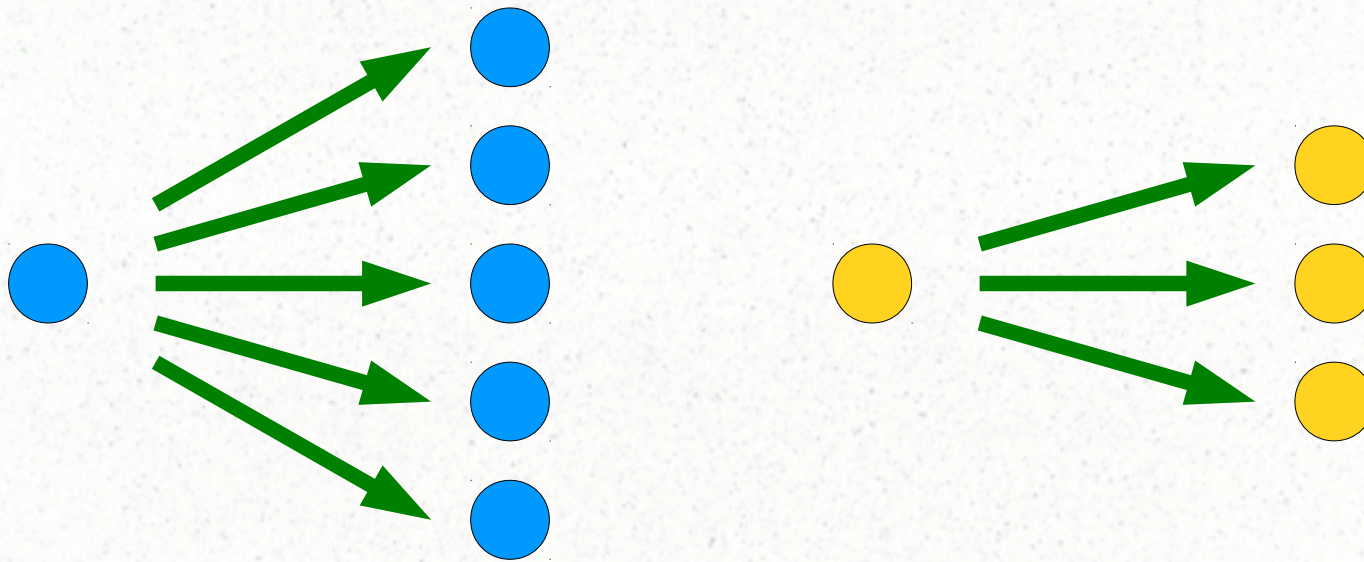
Mutation

Replication



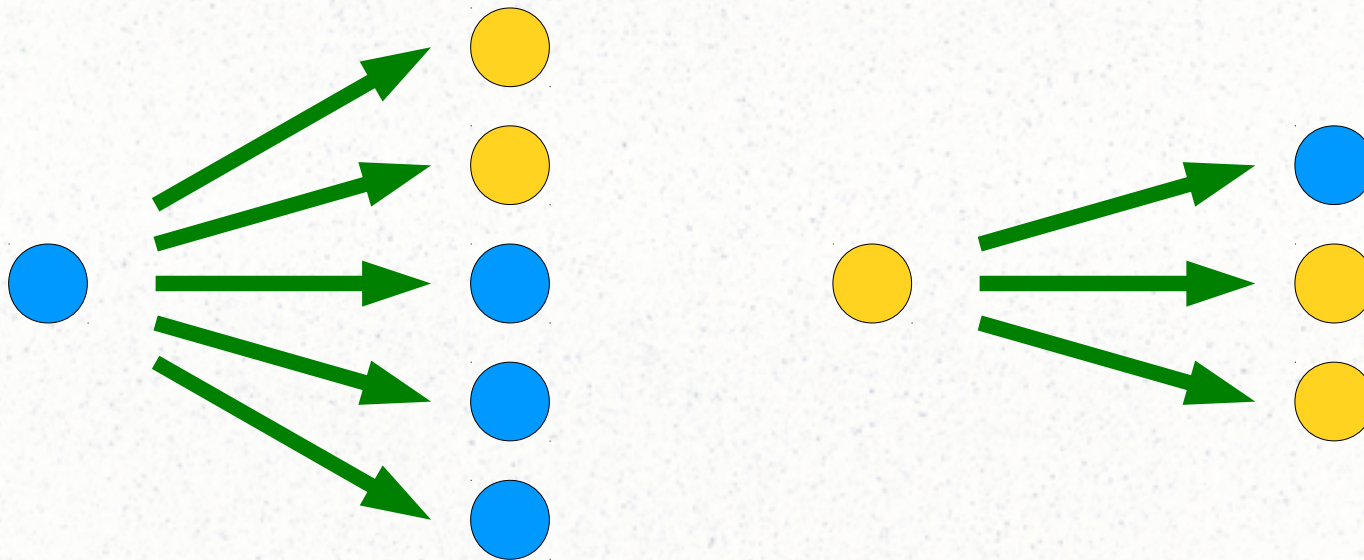
$$P\{X = k\} = p(k) \quad k = 0, 1, 2, \dots$$

Selection



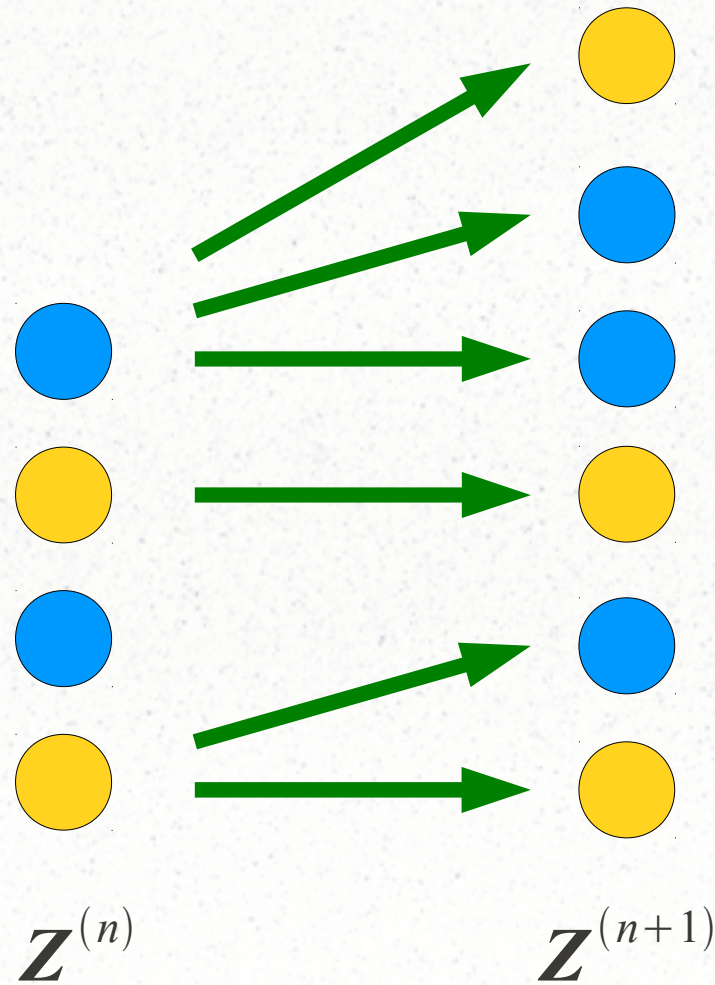
$$P_i \{X = k\} = p_i(k) \quad k = 0, 1, 2, \dots$$

Mutation



$$P_i \{ \mathbf{X} = \mathbf{k} \} = p_i(\mathbf{k}) \quad k_i = 0, 1, 2, \dots$$

Branching process



Mutation-selection matrix

$$E_i \{X_j\} = w_{ij} = r_i q_{ij}$$

$$W \equiv R Q$$

mutation-selection matrix

$$R = \begin{pmatrix} r_1 & & 0 \\ & \ddots & \\ 0 & & r_n \end{pmatrix}$$

replication matrix

$$Q = (q_{ij})$$

mutation matrix

$$Q \mathbf{u}^T = \mathbf{u}^T \quad \mathbf{u} = (1, \dots, 1)$$

Evolution equation for the mean

$$N^{(n)} = E \{ Z^{(n)} \}$$

$$N^{(n+1)} = N^{(n)} W$$

quasi-species equation

Some properties

1. Exponential growth

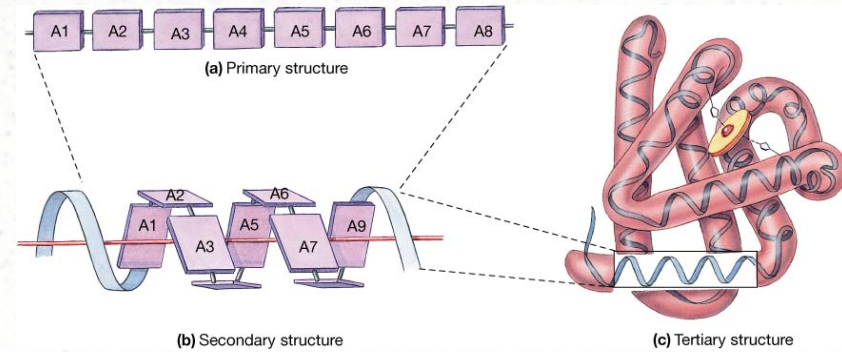
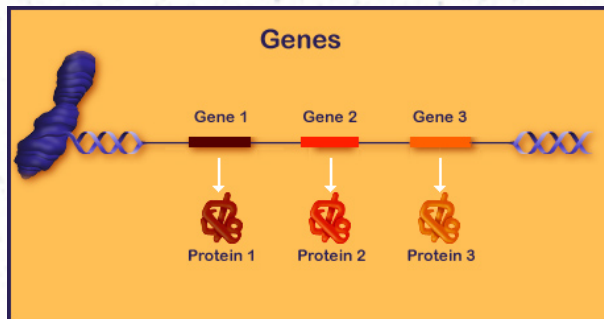
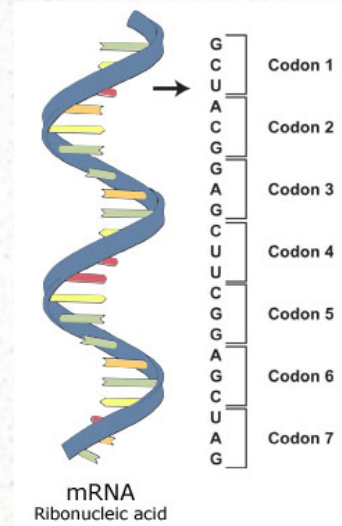
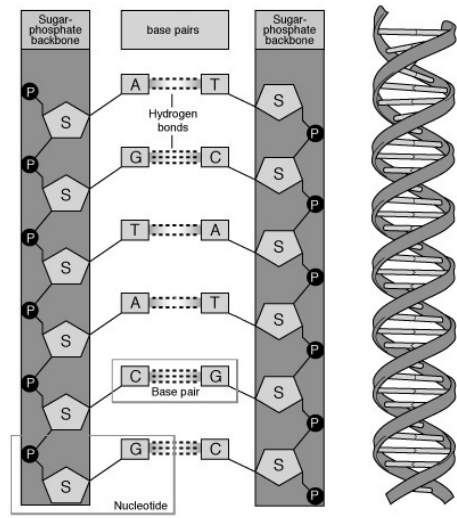
$$N^{(n)} \sim \lambda_{max}^n \mathbf{v}_{max} \quad \mathbf{v}_{max} W = \lambda_{max} \mathbf{v}_{max}$$

2. Survival of the fittest

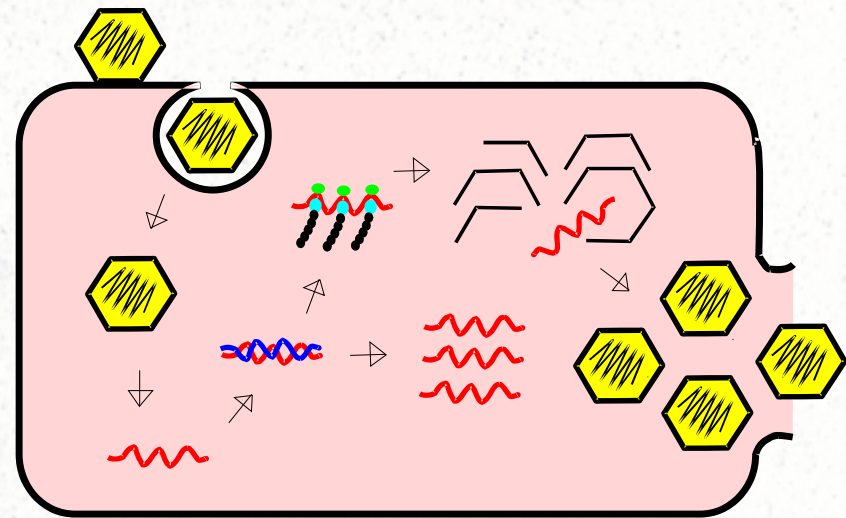
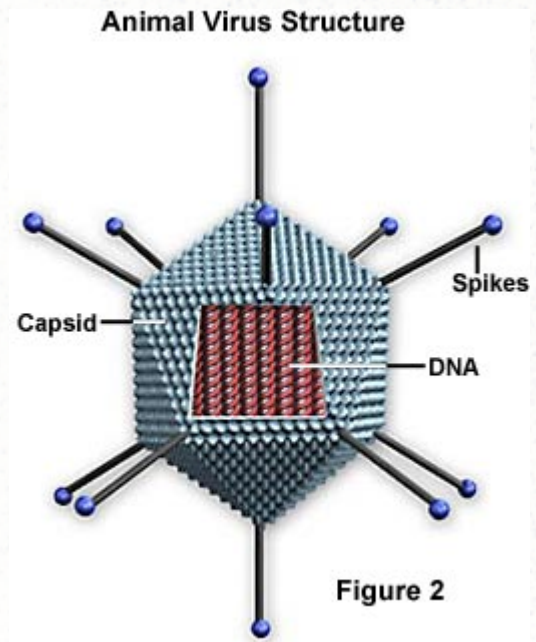
$$W=R \quad \lambda_{max} = r_i \quad \mathbf{v}_{max} = \mathbf{e}_i \quad r_i > r_j$$

3. Mutations maintain variability in population

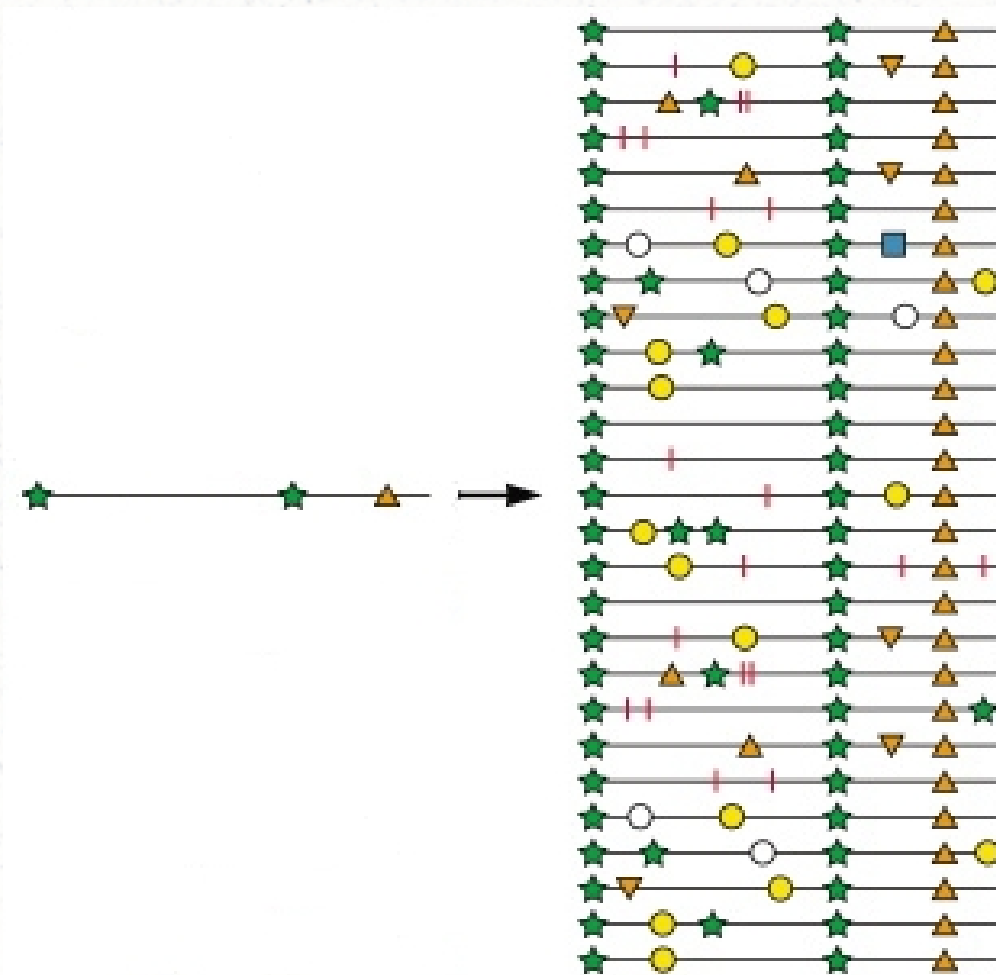
Evolution acts on sequences



Virus



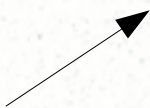
Viral replication



Quasi-species

ATTGGAAATGCCGCAATTTACGGGA
ACTTGCAAATTCGCAAATTCGGGG
AGTTGGAACCTTCGCAATTCCTCGGGA
ACTTGGACATTCCGATATTCTCGGGA
GGTTGGAATAACCCAATTTTCGGGA
ACTTTGAAATTCGCAACGGTCGGGA
ACATGGAAATTCGCAATTTTCGGGA

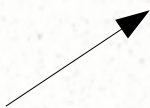
ACTTGGAAATTCGCAATTTTCGGGA


consensus sequence

Quasi-species

11001010100110011011000100
10001110101110011111100101
11001010001110011010100100
10001011101110101010100100
01001010100111011011100100
1000010101110011100100100
10101010101110011011100100

10001010101110011011100100


consensus sequence

Steady state

quasi-species equation

$$\mathbf{x}^{(n+1)} = \frac{\mathbf{x}^{(n)} \mathbf{W}}{\mathbf{x}^{(n)} \mathbf{W} \mathbf{u}^T} \quad \mathbf{x}^{(n)} \mathbf{u}^T = 1$$

steady state

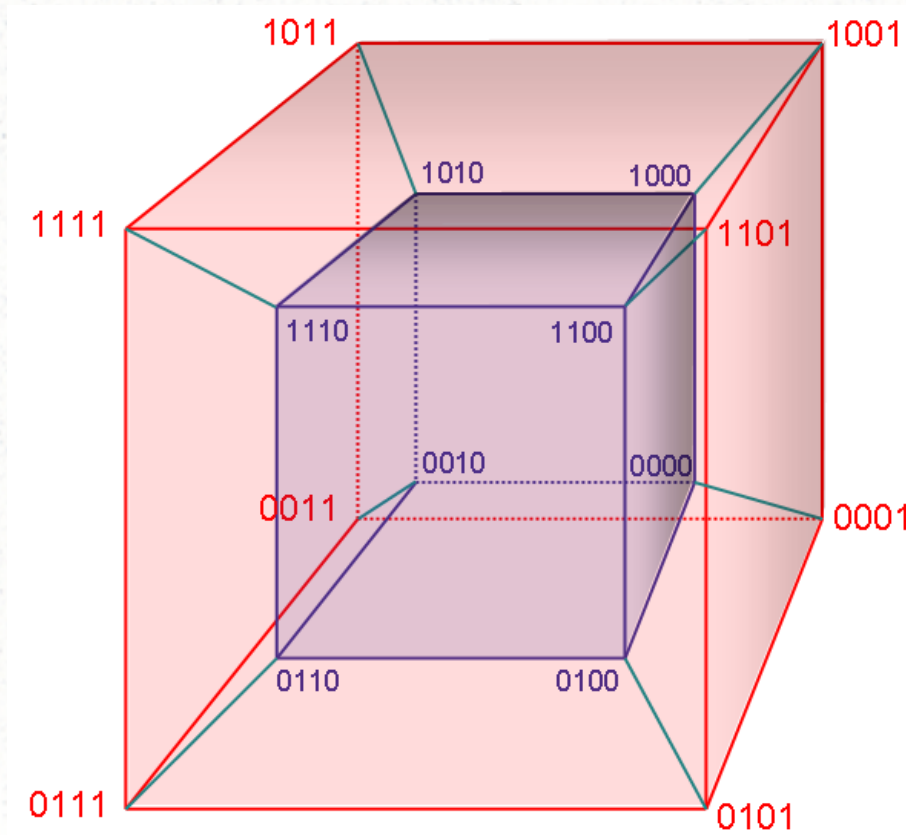
$$\mathbf{x} \mathbf{W} = \phi \mathbf{x}$$

$$\phi = \mathbf{x} \mathbf{W} \mathbf{u}^T = \mathbf{x} \mathbf{R} \mathbf{u}^T = \sum_i r_i x_i$$

average fitness

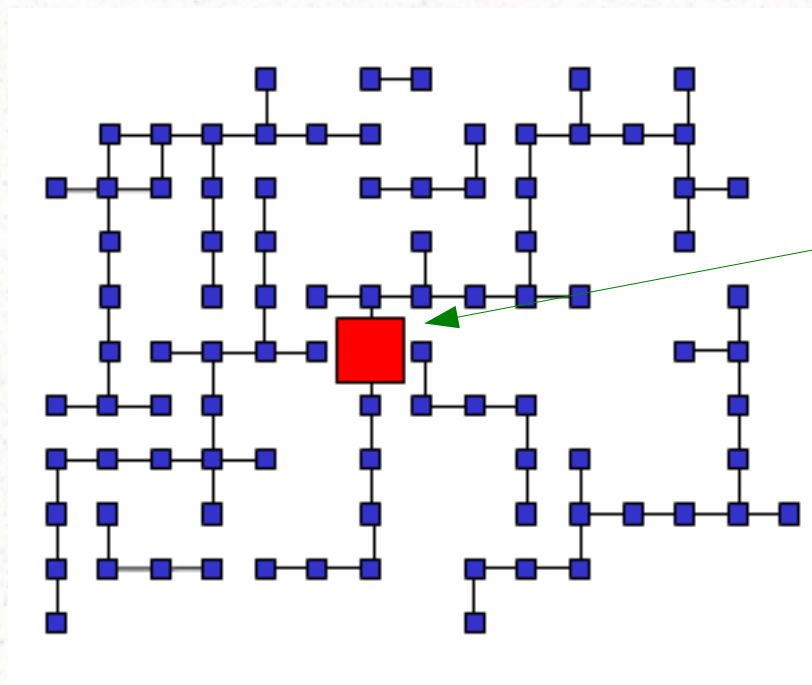


Point mutations



probability of a point mutation: $\mu \ll 1$

Single-peak landscape



master sequence

$$r_0 = r > 1 = r_1 = r_2 = \dots = r_n$$

$$x_0 = x \quad x_1 + x_2 + \dots + x_n = 1 - x$$

Error catastrophe

$$L \gg 1$$

$$\sum_{j=0}^n x_j r_j q_{ji} = \phi x_i \quad \phi = r x + 1 - x$$

$$x [r(1 - \mu L) + O(\mu)] = x [1 + (r - 1)x]$$

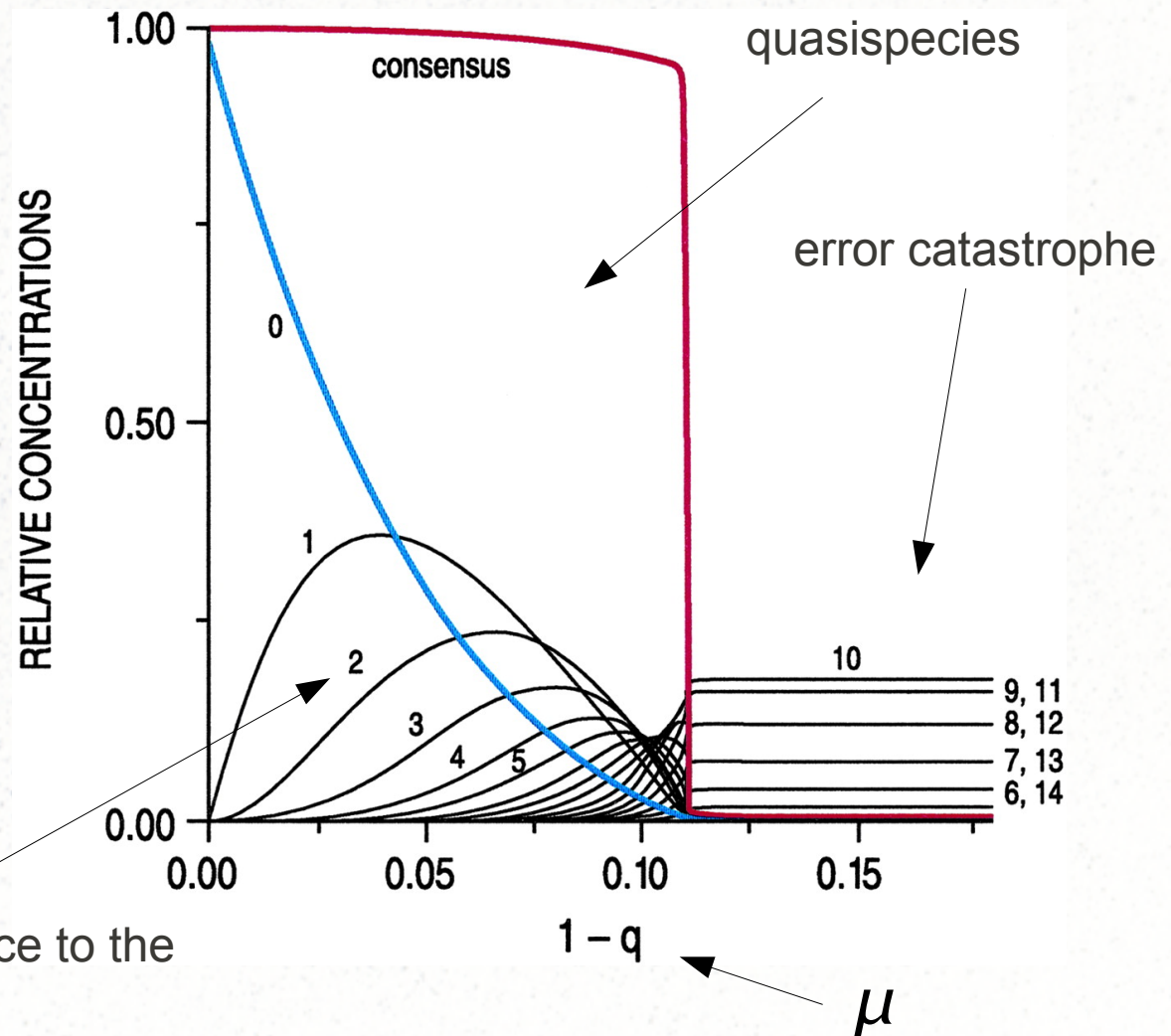
$$x^* \approx 1 - \frac{r}{r-1} \mu L \quad \text{if } \mu L < 1 - \frac{1}{r}$$

$$x^* = O(\mu) \quad \text{if } \mu L > 1 - \frac{1}{r}$$

Error catastrophe



Manfred Eigen (1971)

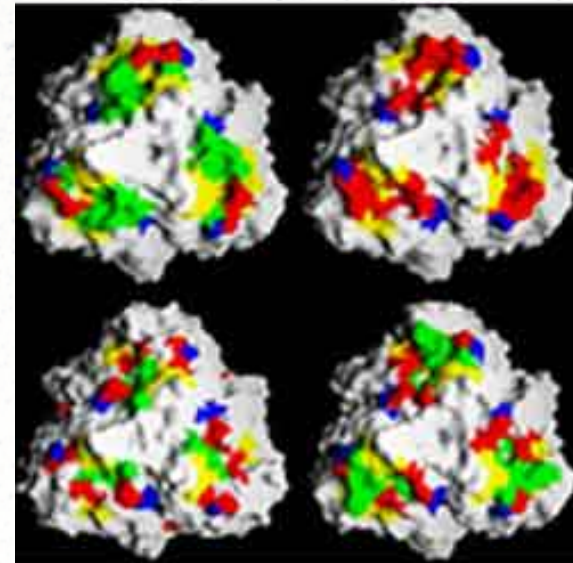
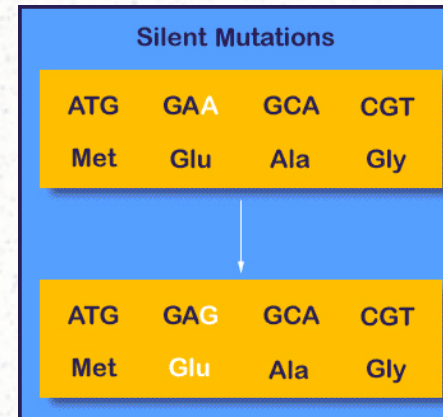


Hamming distance to the fittest sequence

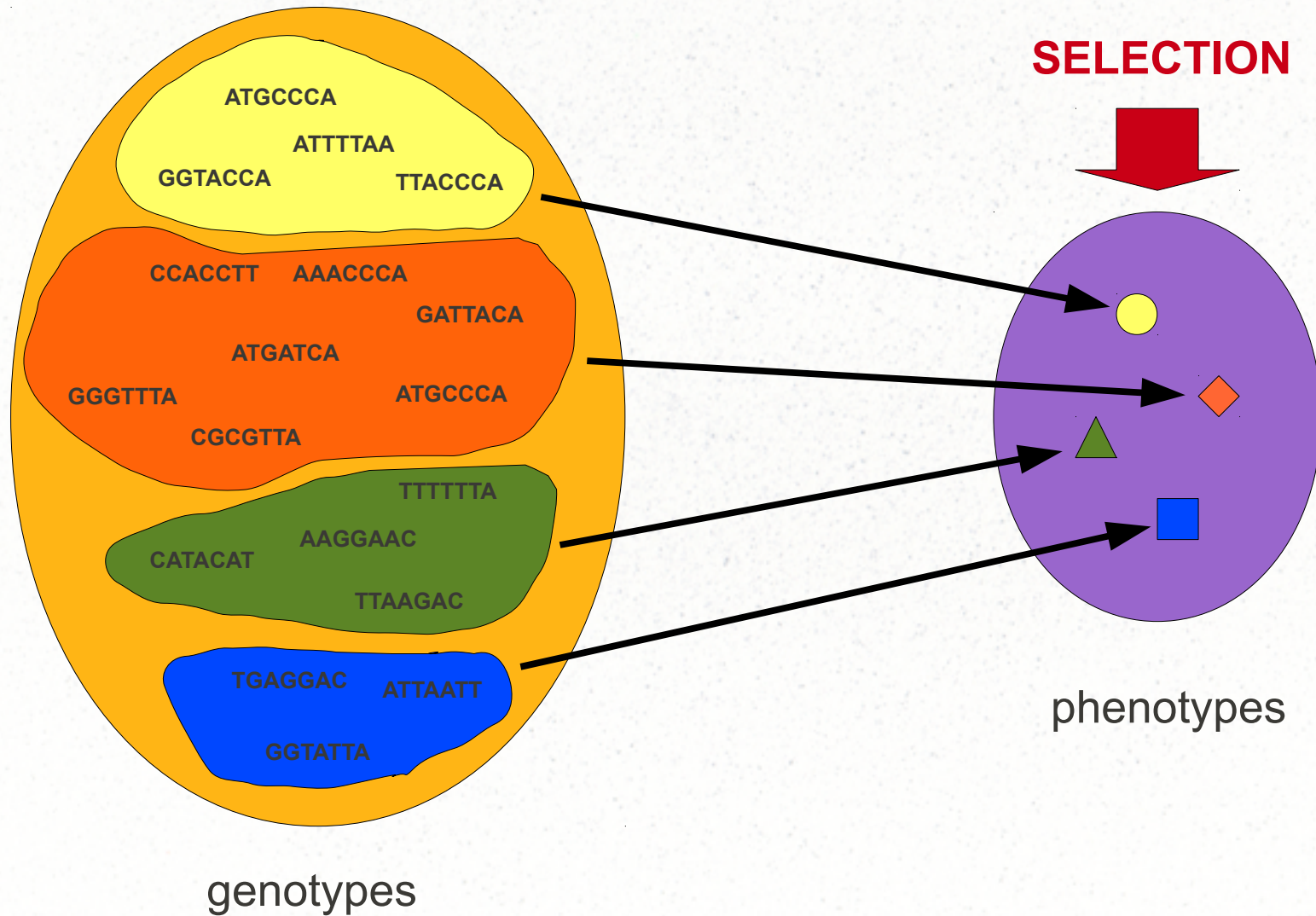
*How reliable is the single-peak
landscape?*

High redundancy

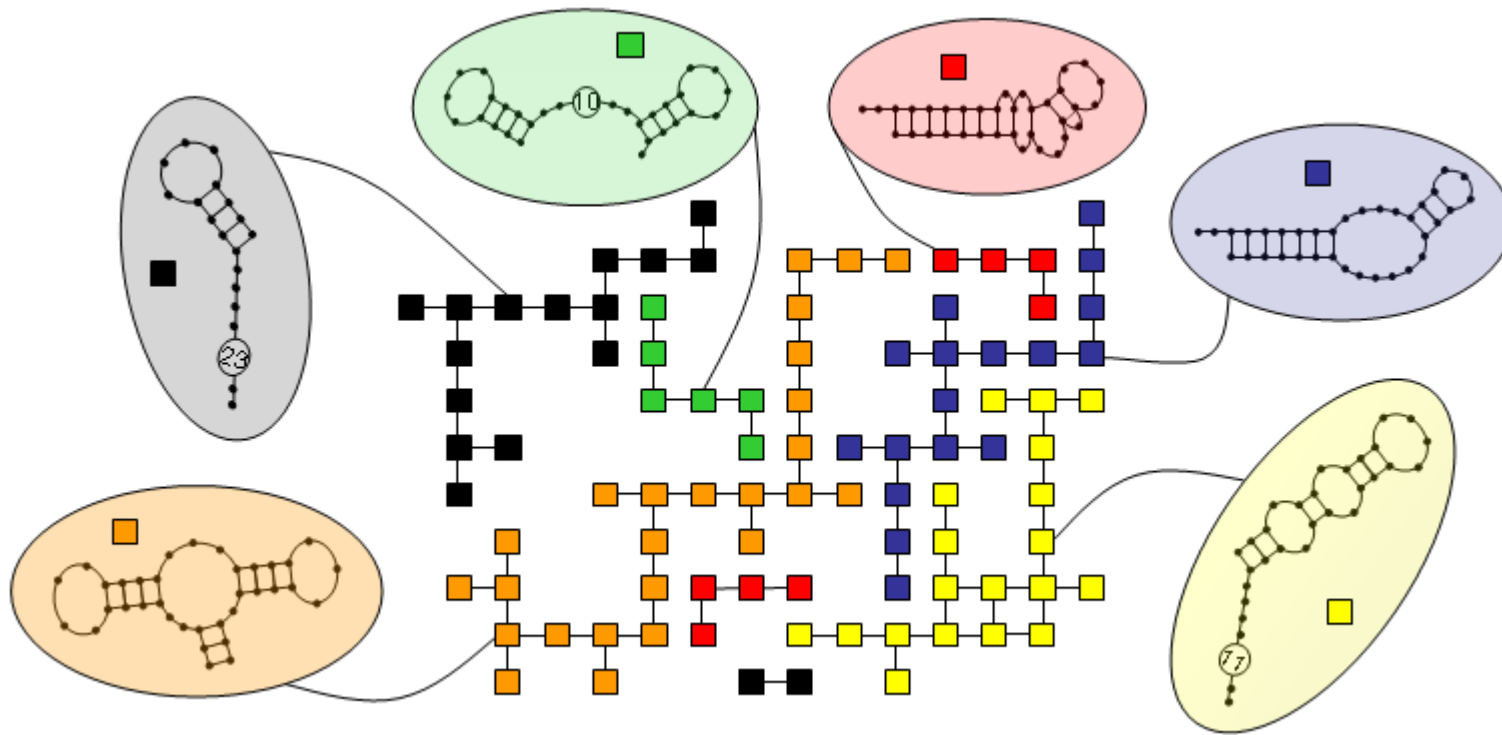
		Second Base				Third Base
		U	C	A	G	
U	U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	U	UUC	UCC	UAC	UGC	C
	U	UUA Leu	UCA	UAA Stop	UGA Stop	A
	U	UUG	UCG	UAG Stop	UGG Trp	G
C	C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	C	CUC	CCC	CAC	CGC	C
	C	CUA	CCA	CAA Gln	CGA	A
	C	CUG	CCG	CAG	CGG	G
A	A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	A	AUC	ACC	AAC	AGC	C
	A	AUA	ACA	AAA Lys	AGA Arg	A
	A	AUG Met / Start	ACG	AAG	AGG	G
G	G	GUU Val	GCU Ala	CAU Asp	GGU Gly	U
	G	GUC	GCC	GAC	GGC	C
	G	GUA	GCA	GAA Glu	GGA	A
	G	GUG	GCG	GAG	GGG	G



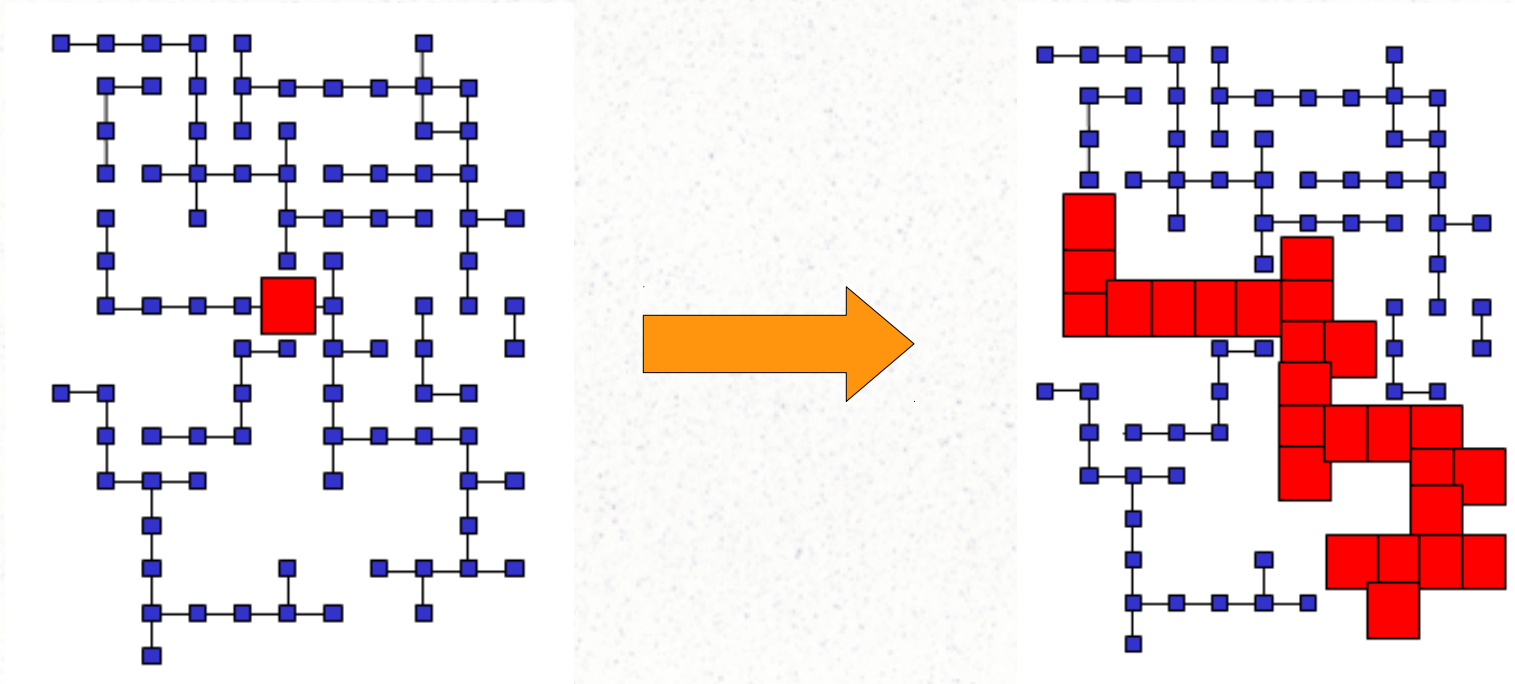
Genotype-phenotype mapping



Distribution of phenotypes

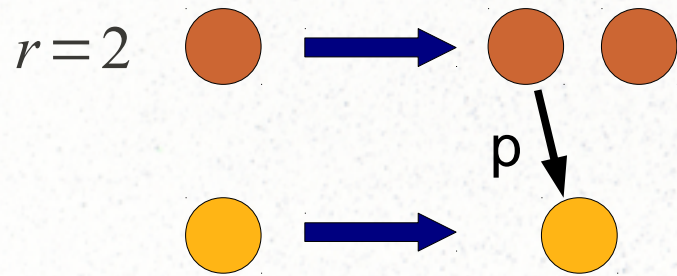


There's not single-peak landscape

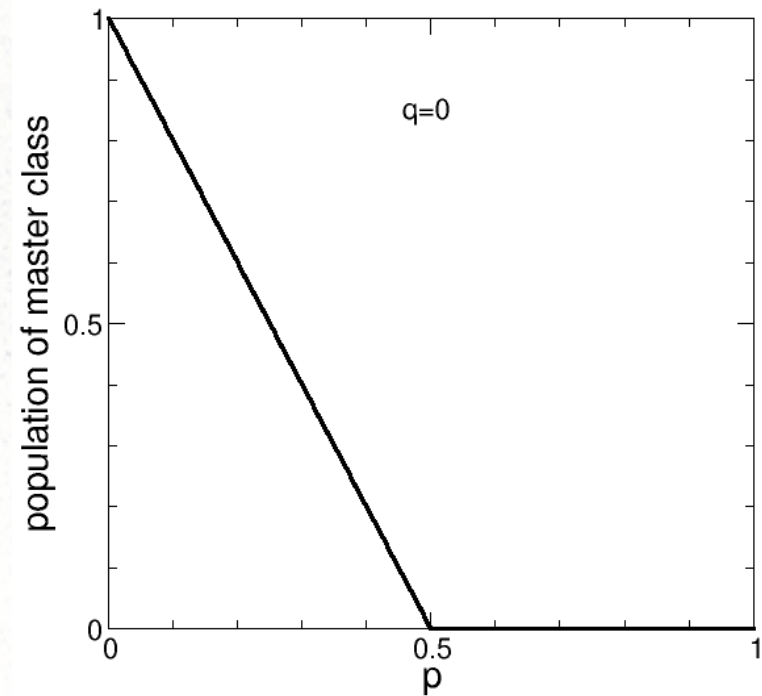
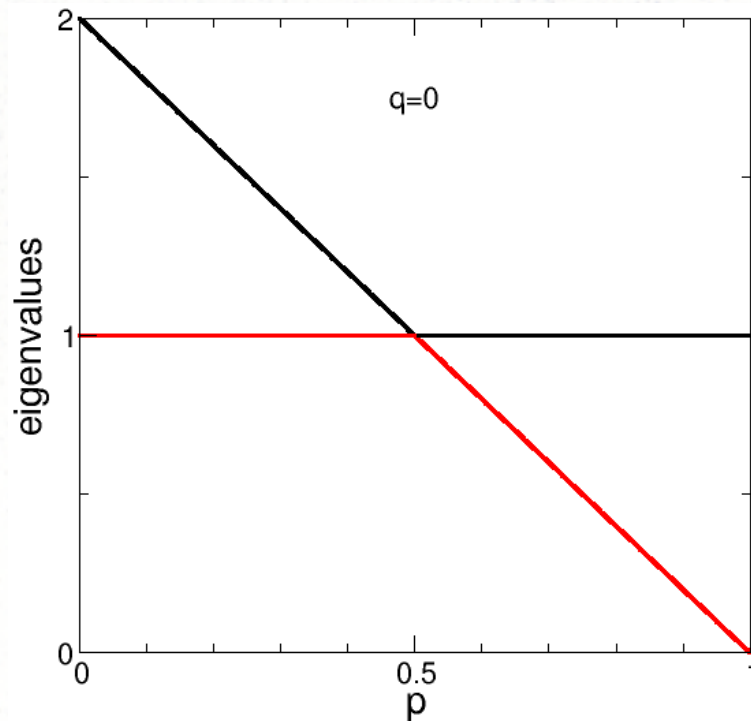


compensatory mutations
cannot be neglected

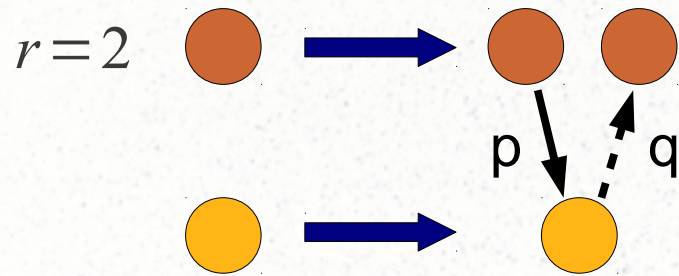
Error catastrophe?



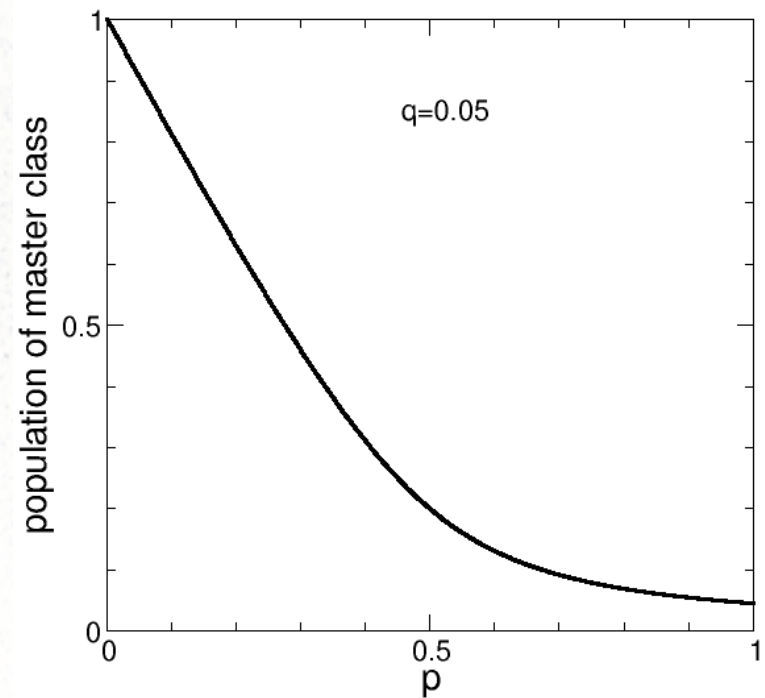
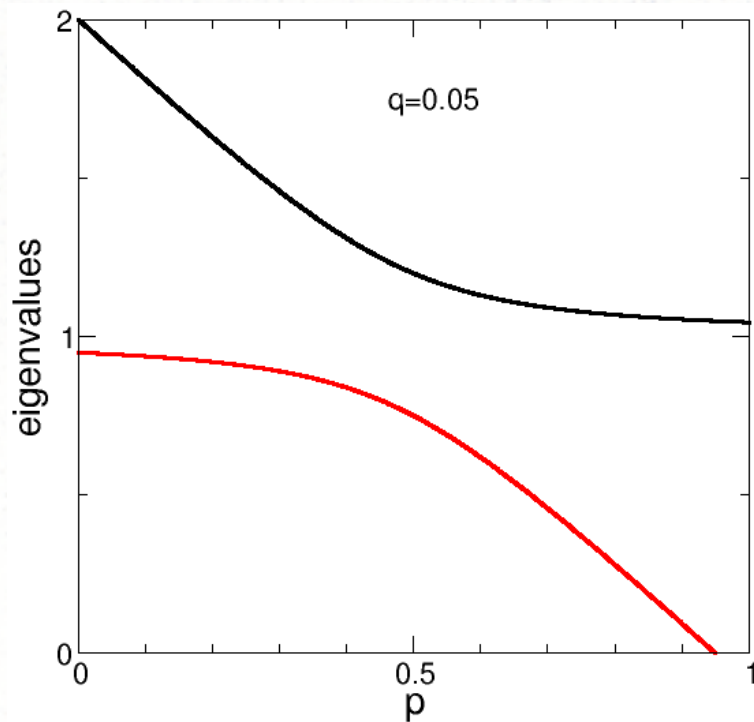
$$W = \begin{pmatrix} r(1-p) & r p \\ 0 & 1 \end{pmatrix}$$



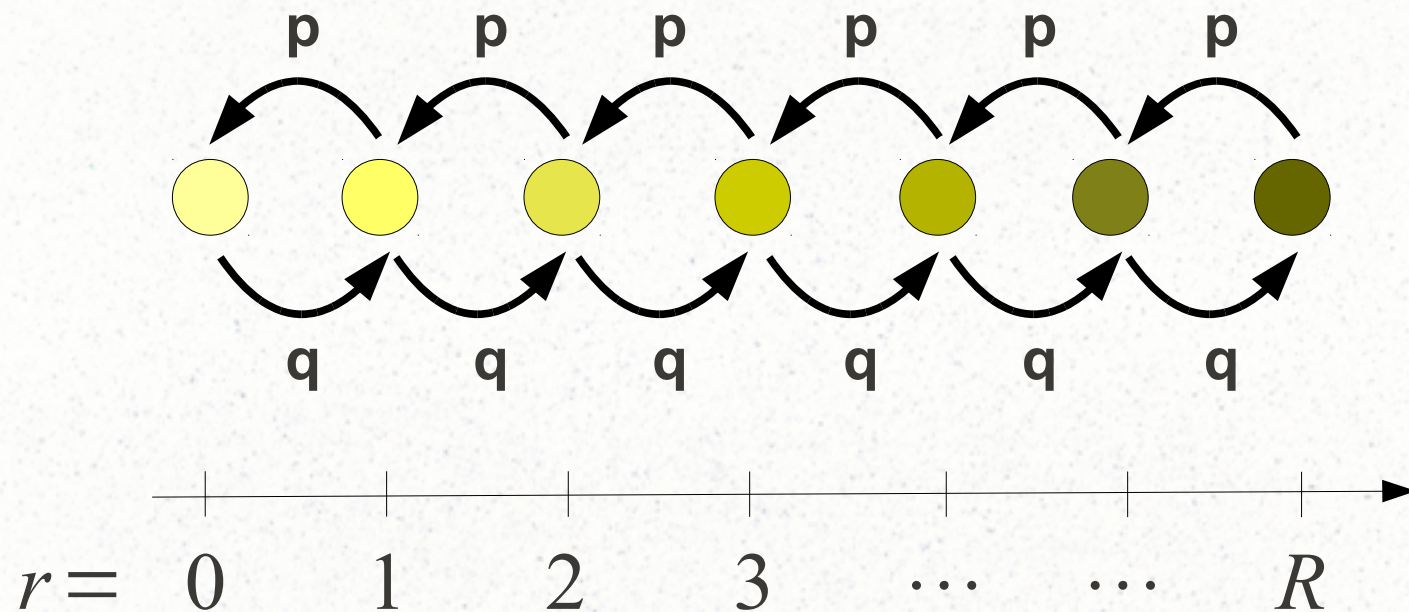
Error catastrophe?



$$W = \begin{pmatrix} r(1-p) & r p \\ q & 1-q \end{pmatrix}$$



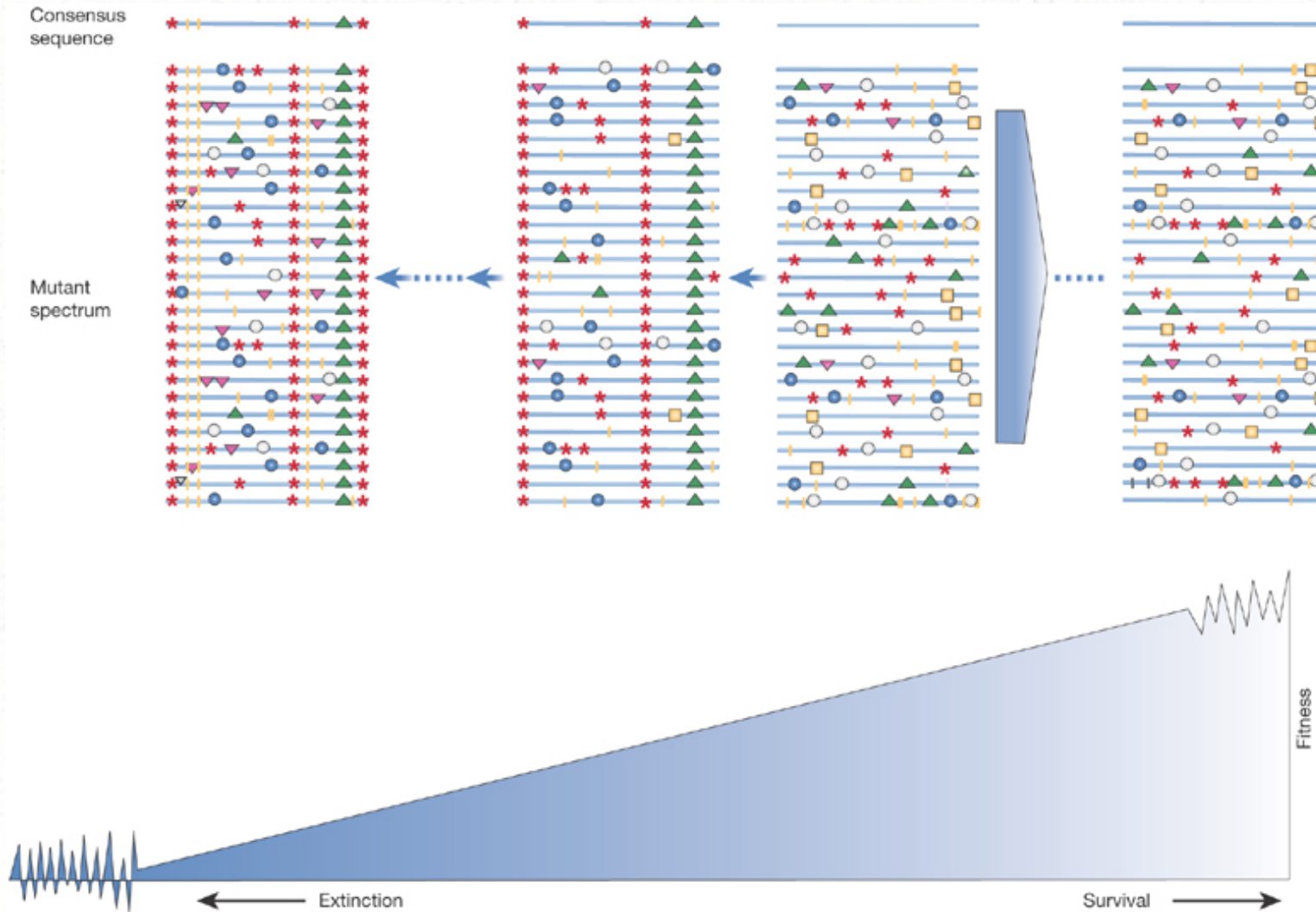
A simple generalization



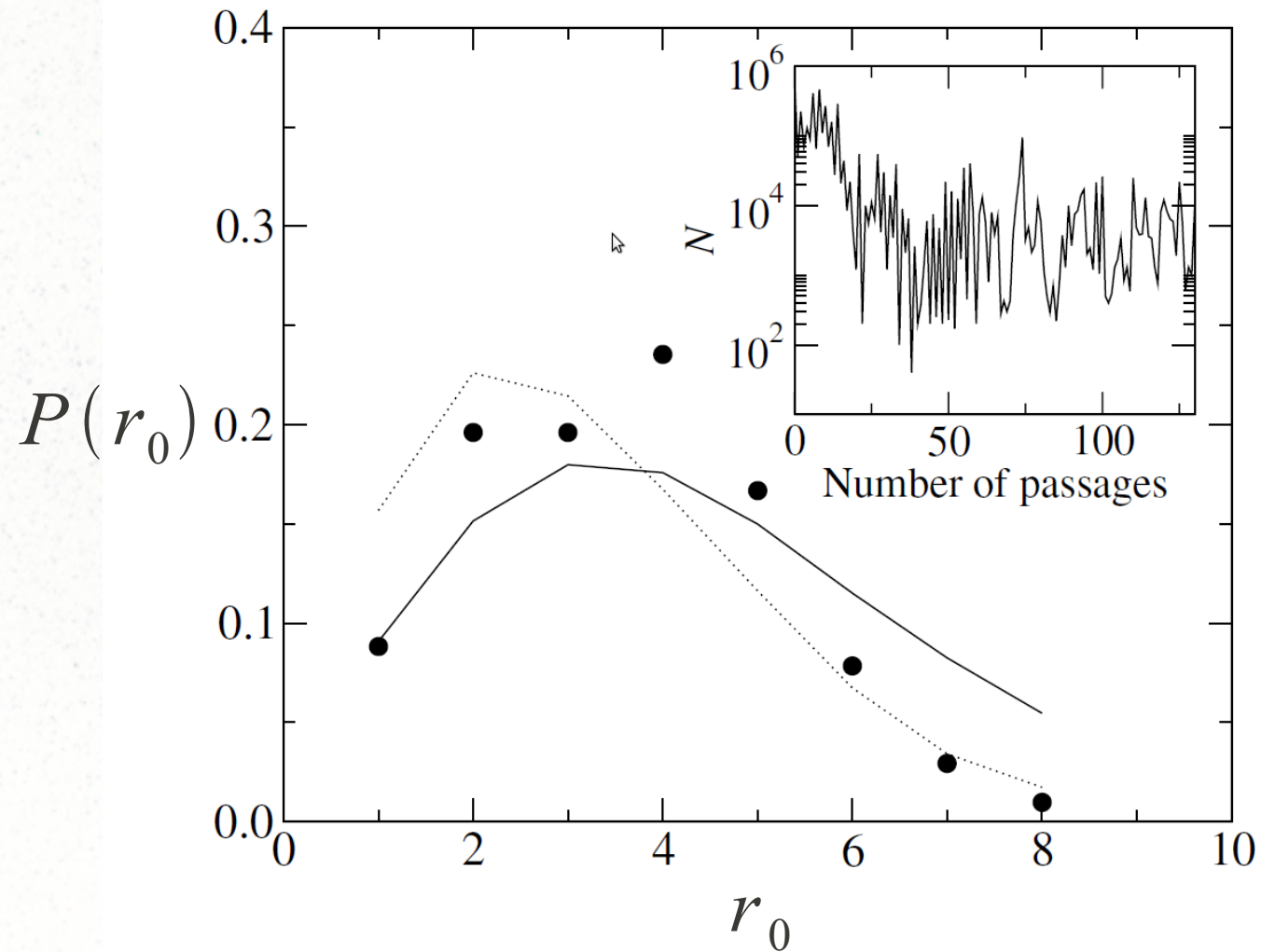
$$N_r^{(n+1)} = q(r-1)N_{r-1}^{(n)} + (1-p-q)rN_r^{(n)} + p(r+1)N_{r+1}^{(n)}$$

$$N_R^{(n+1)} = q(R-1)N_{R-1}^{(n)} + (1-p)RN_R^{(n)} \quad 0 < r < R$$

Muller's ratchet



Actual behavior



Conclusions

- **REPLICATION + SELECTION + MUTATION = EVOLUTION**
- Variability yields **quasi-species**
- Evolution of sequences is well described by the **quasi-species equation**
- Assuming an optimal sequence and neglecting back mutations leads to the **error catastrophe**
- There is high **redundancy** in biology (a huge number of **genotypes** produce the same **phenotype**) → no optimal sequence
- **Redundancy** generates **compensatory mutations** eliminating the **error catastrophe**
- Simple models with **compensatory mutations** capture experimental behavior in **viral evolution**