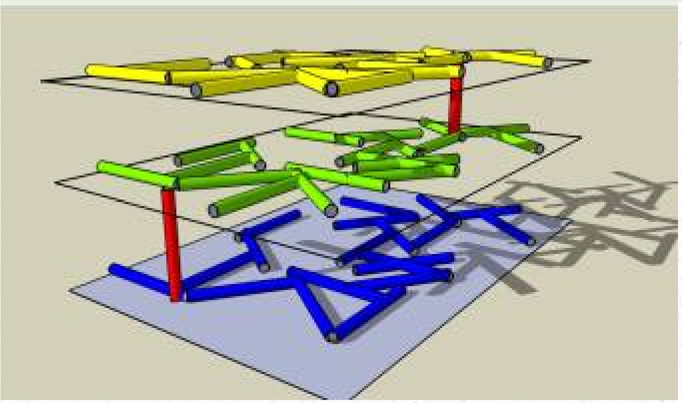
# *Neutral evolution:*
## *adaptation behind the curtain*



**José A. Cuesta & Susanna Manrubia**

GISC – UC3M CAB

gisc

Universidad Carlos III de Madrid

# Evolution:
# The current paradigm

# Building blocks

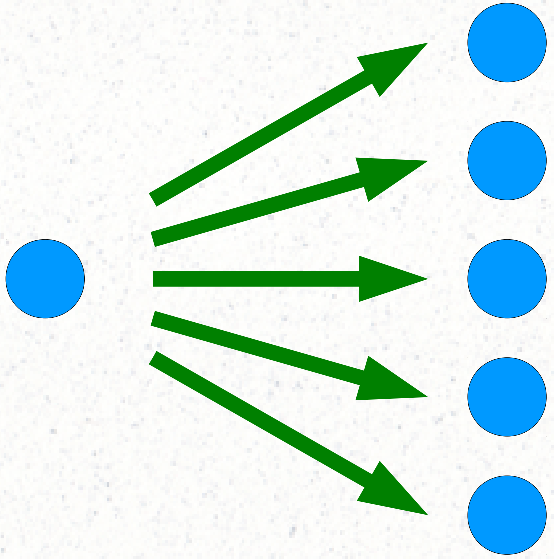
# *Building blocks*

**Replication**

**Selection**

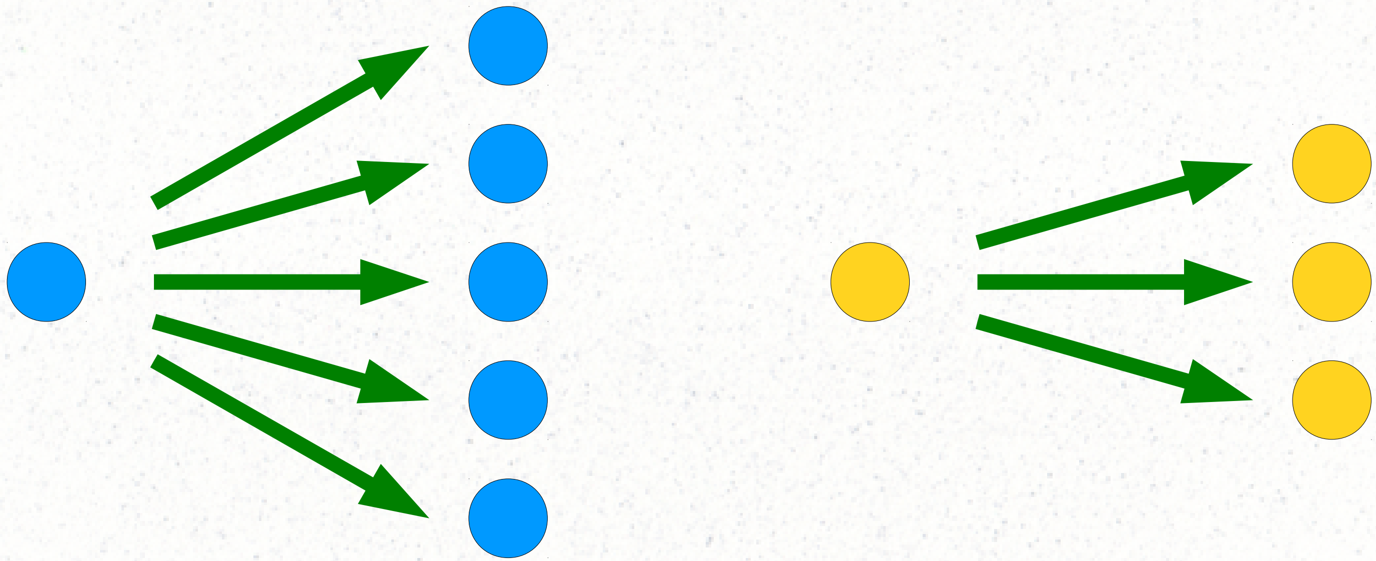
**Mutation**

# Replication



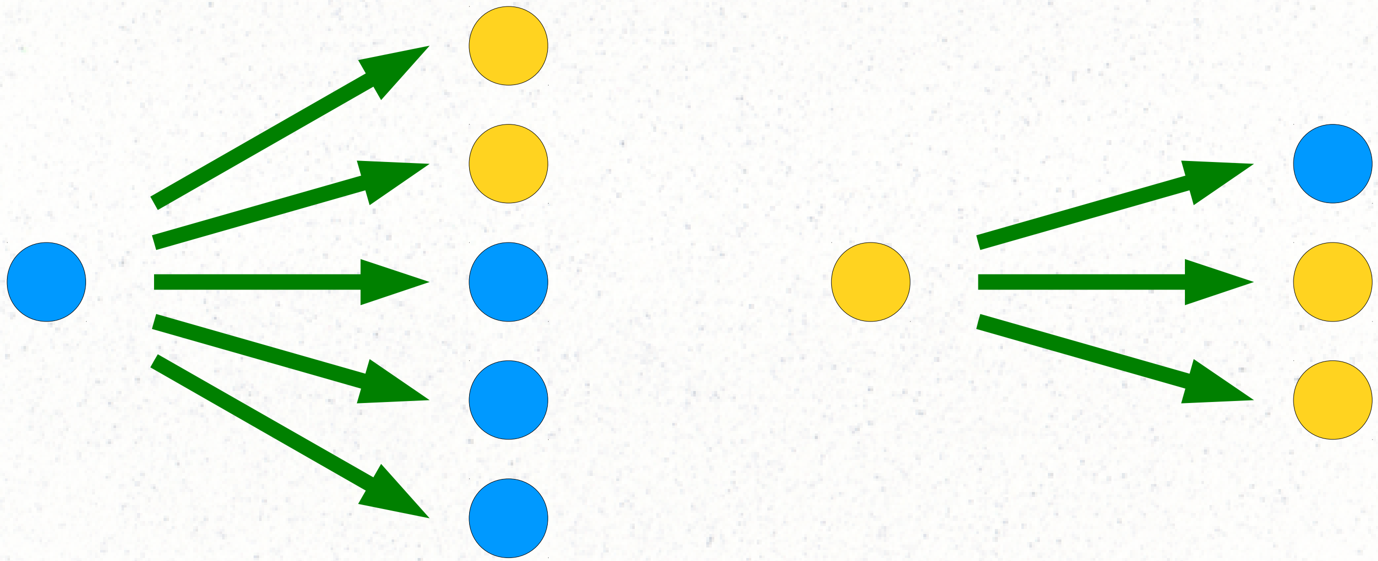
$$P\{X=k\}=p(k) \qquad k=0,1,2\ldots$$

# Selection



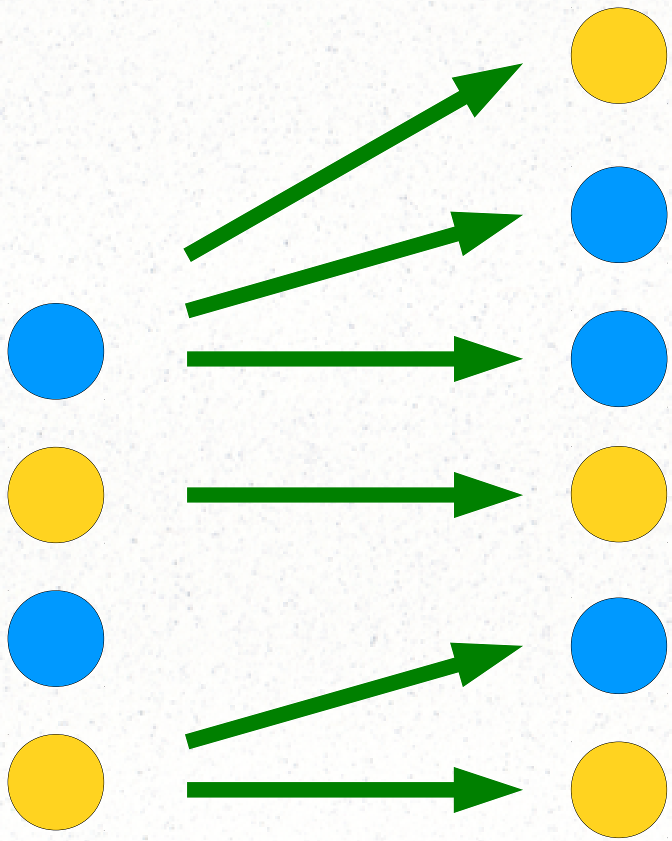
$$P_i\{X=k\}=p_i(k) \qquad k=0,1,2\ldots$$

# Mutation



$$P_i\{X=k\}=p_i(k) \qquad k_i=0,1,2\ldots$$

# Branching process



$Z^{(n)}$ $Z^{(n+1)}$

# Mutation-selection matrix

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

$W \equiv RQ$ mutation-selection matrix

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

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

$$Q\boldsymbol{u}^T = \boldsymbol{u}^T \qquad \boldsymbol{u} = (1, \ldots, 1)$$

# Evolution equation for the mean

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

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

**quasi-species equation**

# Some properties

1. Exponential growth

$$\boldsymbol{N}^{(n)} \sim \lambda_{max}^{n} \boldsymbol{v}_{max} \qquad \boldsymbol{v}_{max} \mathrm{W} = \lambda_{max} \boldsymbol{v}_{max}$$

2. Survival of the fittest

$$\mathrm{W} = \mathrm{R} \qquad \lambda_{max} = r_i \qquad \boldsymbol{v}_{max} = \boldsymbol{e}_i \qquad r_i > r_j$$

3. Mutations are the source of variability

# Steady state

quasi-species equation in frequencies

$$\boldsymbol{x}^{(n+1)} = \frac{\boldsymbol{x}^{(n)}\mathrm{W}}{\boldsymbol{x}^{(n)}\mathrm{W}\,\boldsymbol{u}^T} \qquad \boldsymbol{x}^{(n)}\boldsymbol{u}^T = 1$$

$$\boldsymbol{u} = (1,\ldots,1)$$

steady state

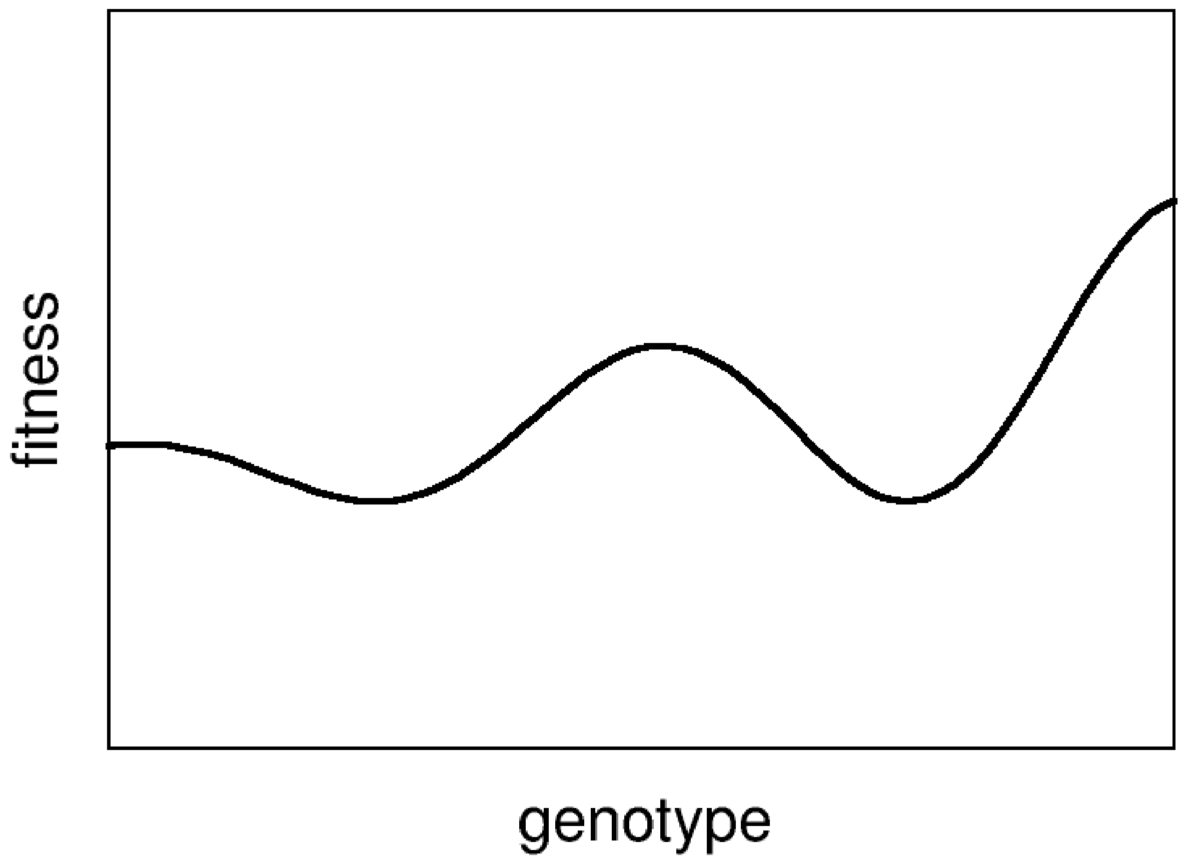
$$\boldsymbol{x}\mathrm{W} = \phi\,\boldsymbol{x}$$

$$\phi = \boldsymbol{x}\mathrm{W}\,\boldsymbol{u}^T = \boldsymbol{x}\mathrm{R}\,\boldsymbol{u}^T = \sum_i r_i x_i$$

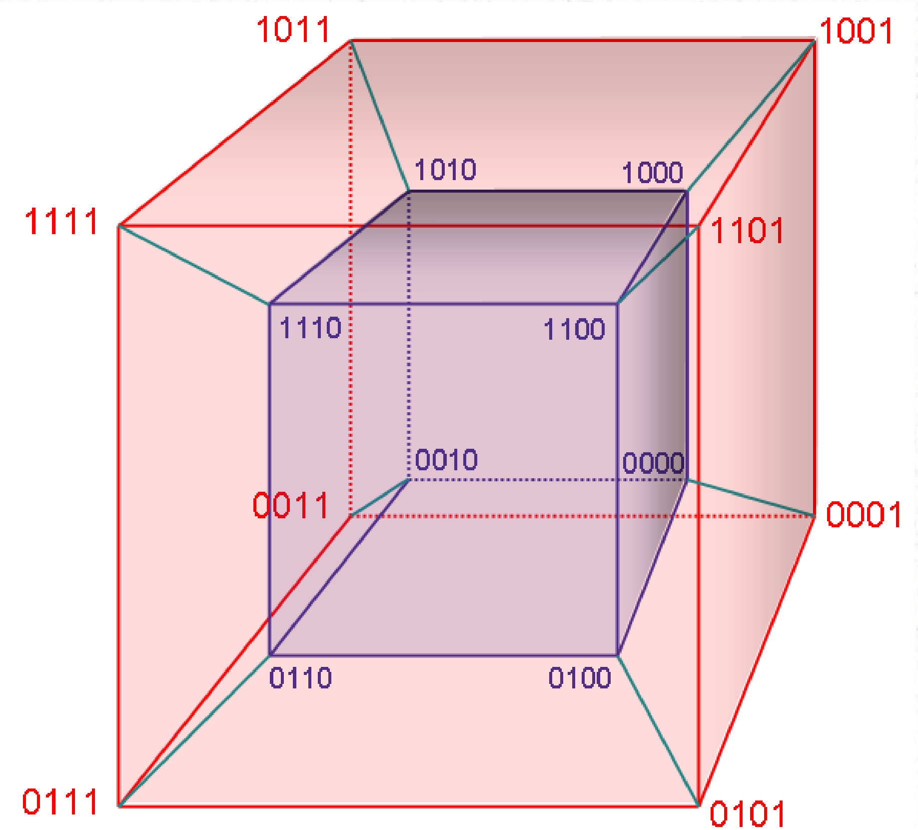
average fitness

# Fitness landscapes

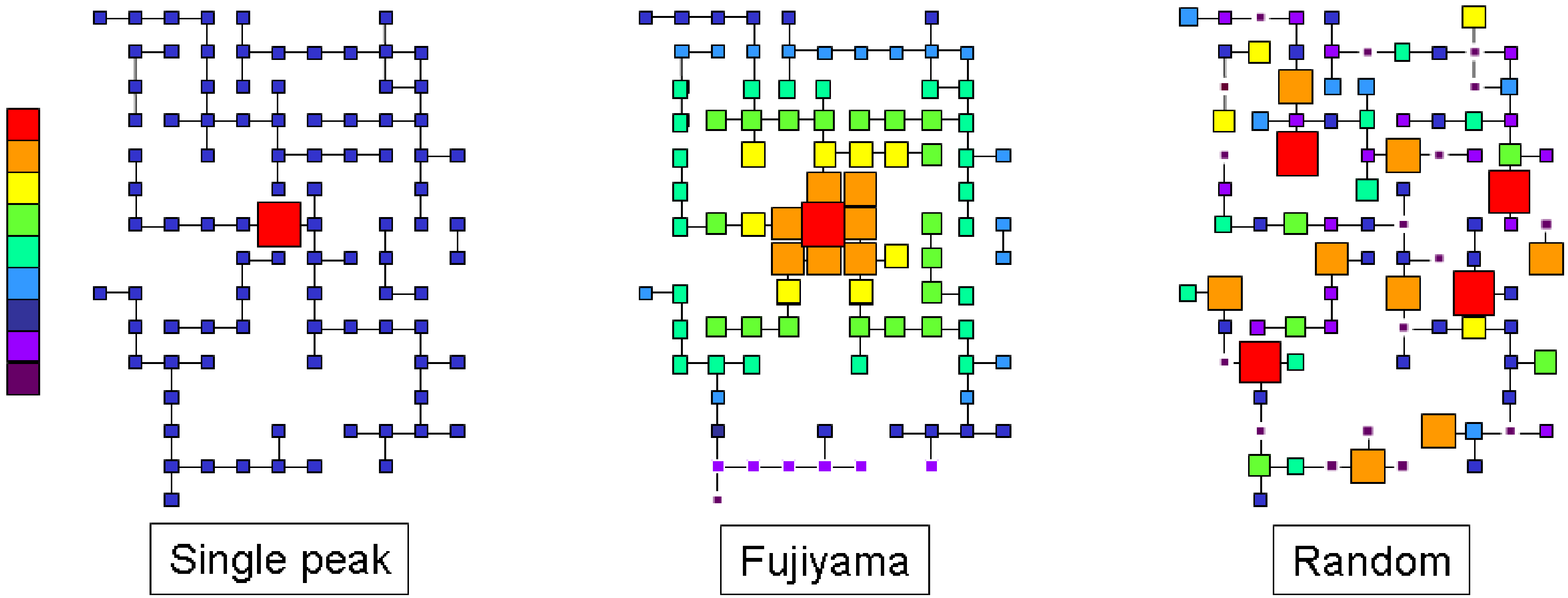
# Fitness landscapes

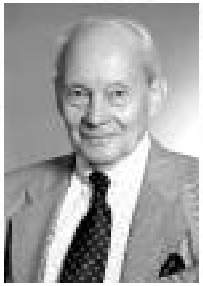


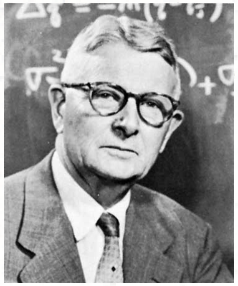
# Genome space



# Metaphors





# *Quasi-species*

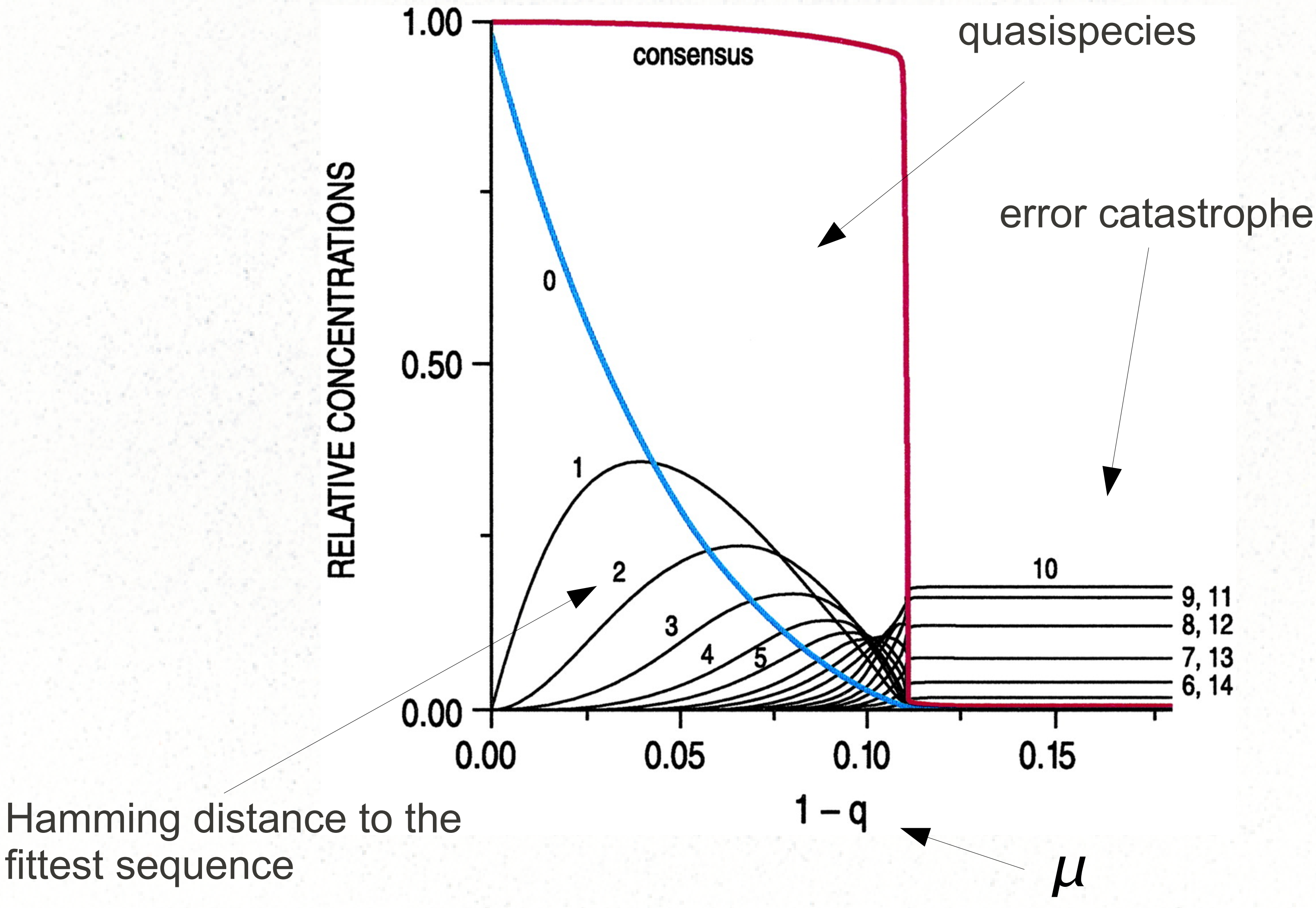
ATTTGGAAATGCCGCAATTTACGGGA
ACTTGCAAATTCCGCAAATTTCGGGG
AGTTGGAACTTCCGCAATTCTCGGGA
ACTTGGACATTCCGATATTCTCGGGA
GGTTGGAAATACCCCAATTTTCGGGA
ACTTTGAAATTCCGCAACGGTCGGGA
ACATGGAAATTCCGCAATTTTCGGGA

---

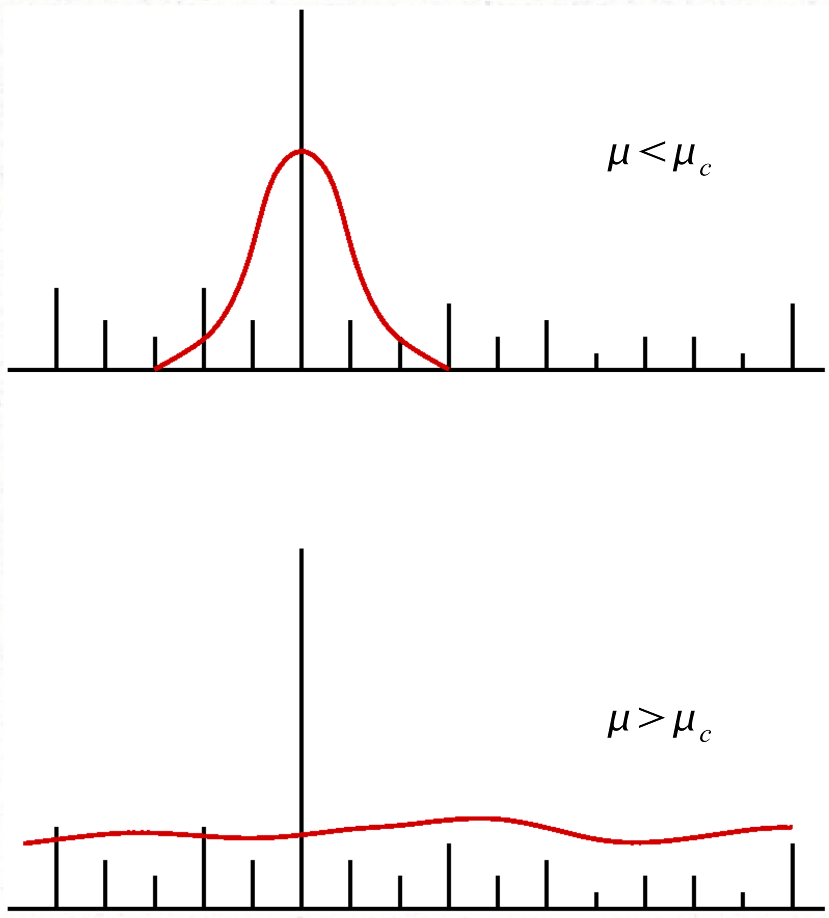
ACTTGGAAATTCCGCAATTTTCGGGA

consensus sequence

# Error catastrophe

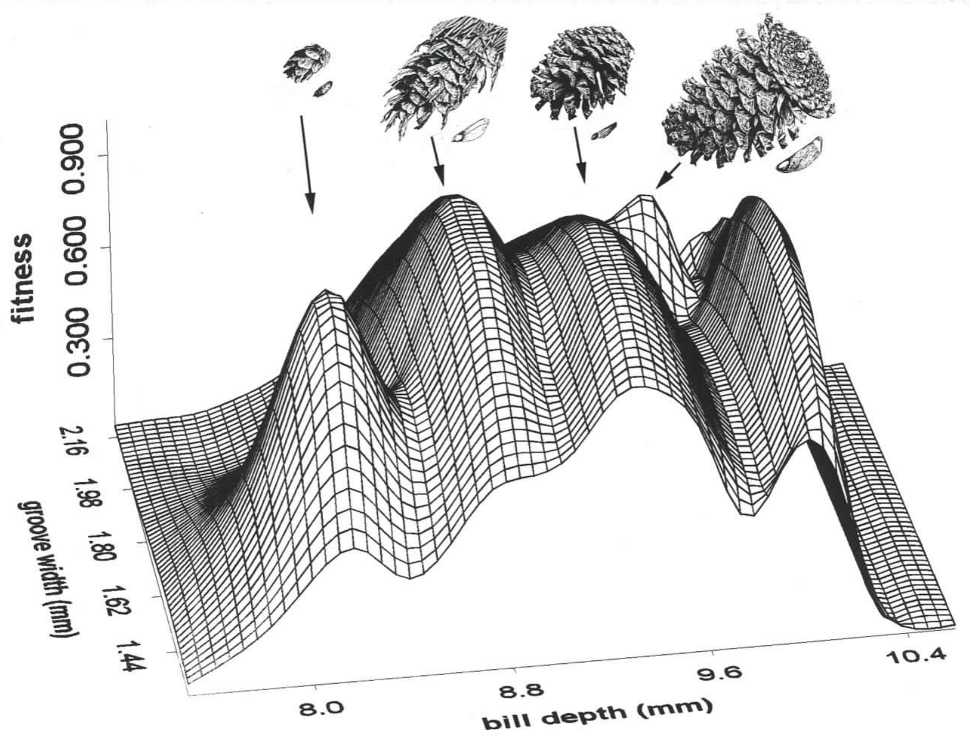


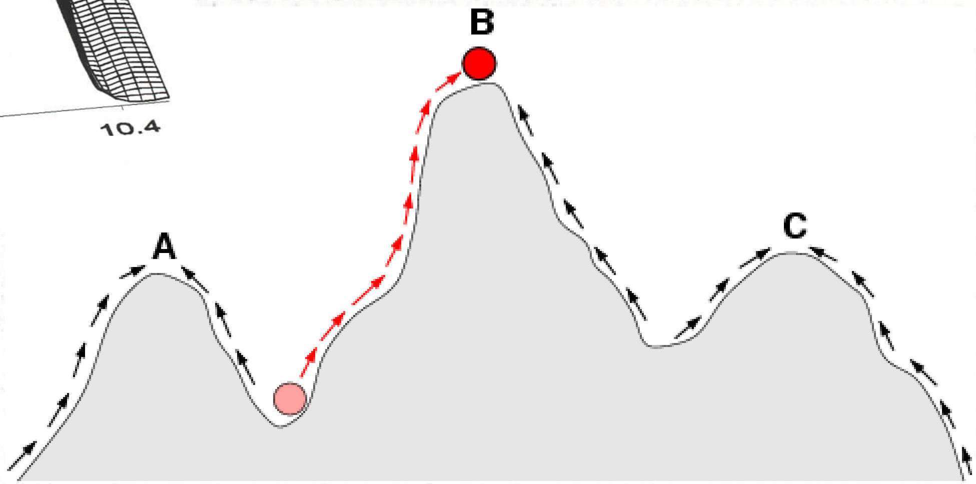
# Error catastrophe




Eigen, *Naturwissenschaften* (1971)

# Speciation in rugged landscapes

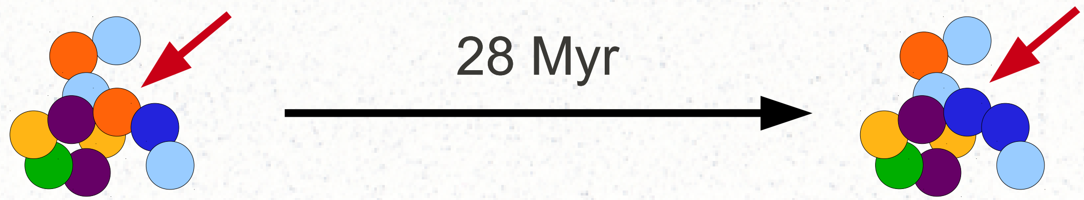


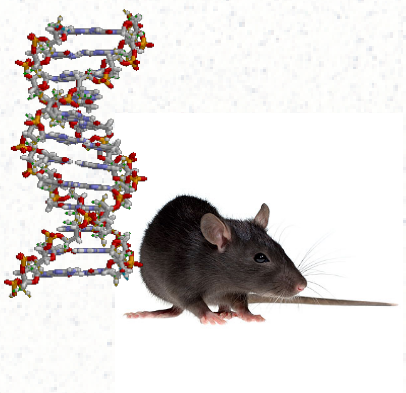
# Evolution:
# The new paradigm

# Neutral evolution

# *Evidence*




100 aa

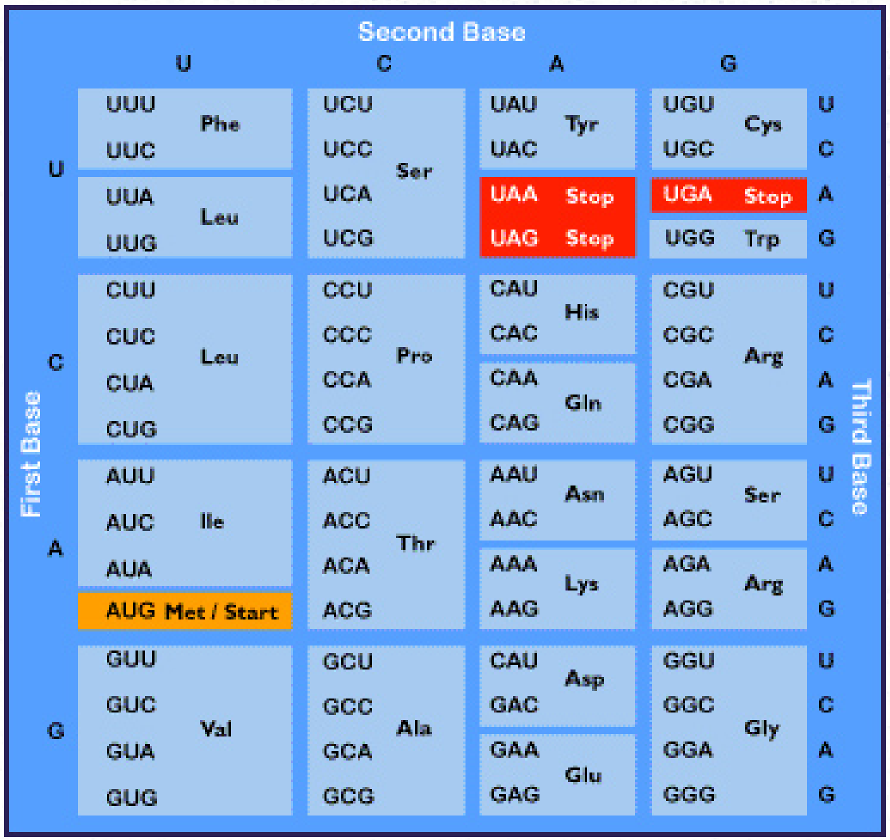


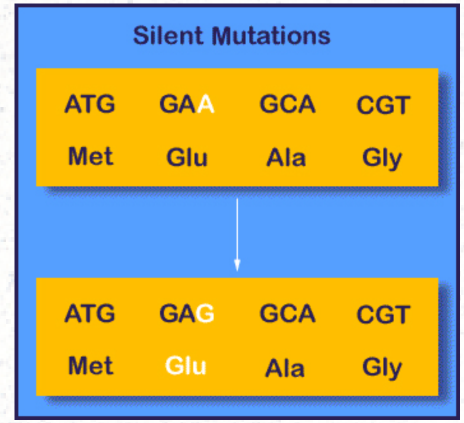
4000 Mbp (80% redundancy)

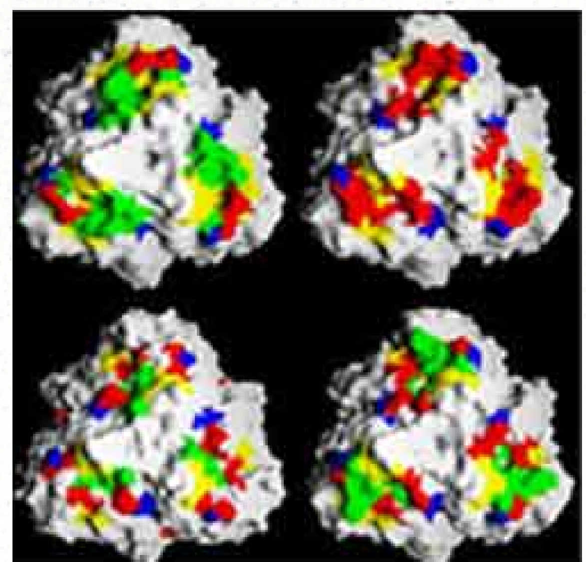
1 aa substitution / genome every 2 yr !!

Kimura, *Nature* (1968)

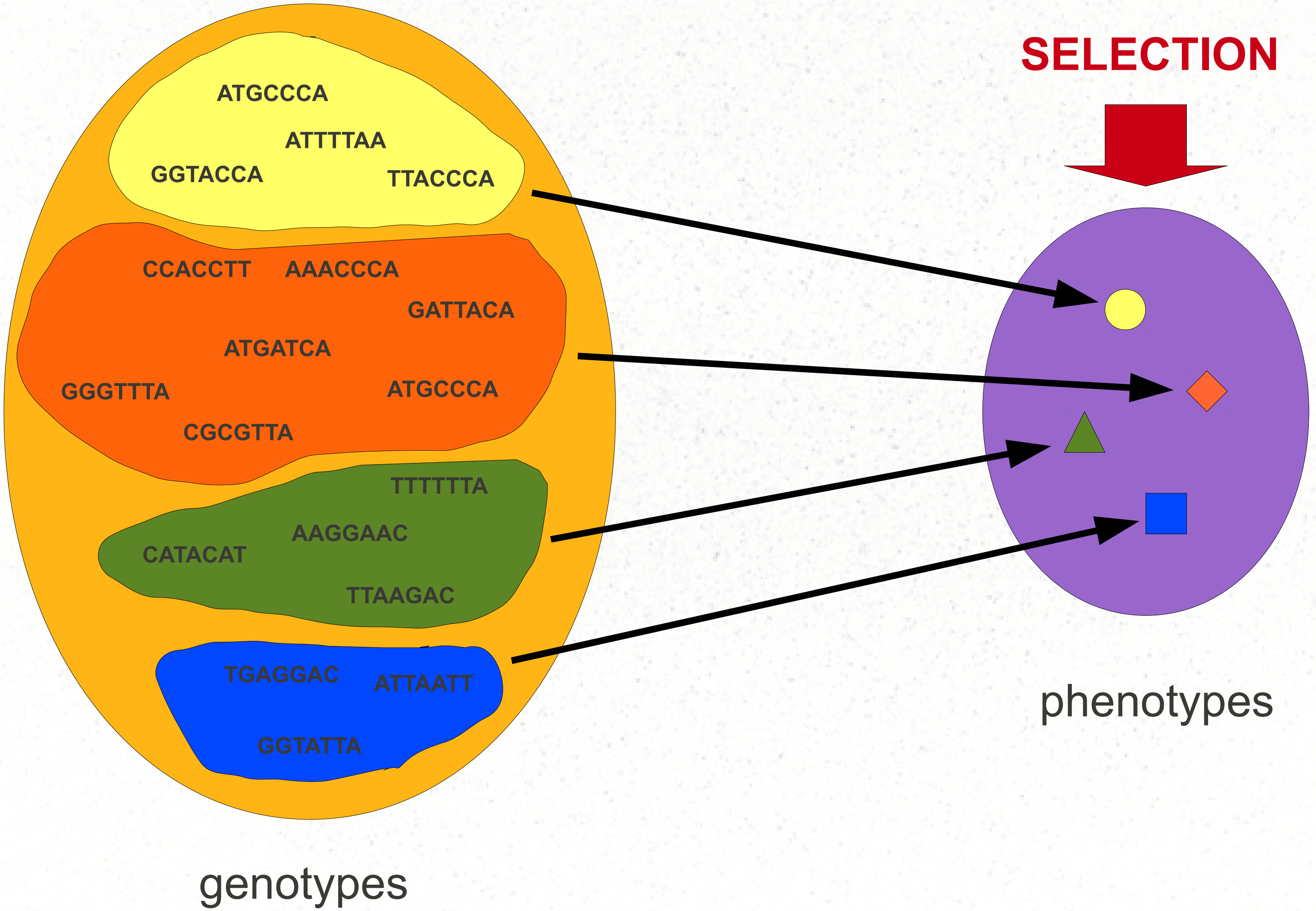
# High redundancy



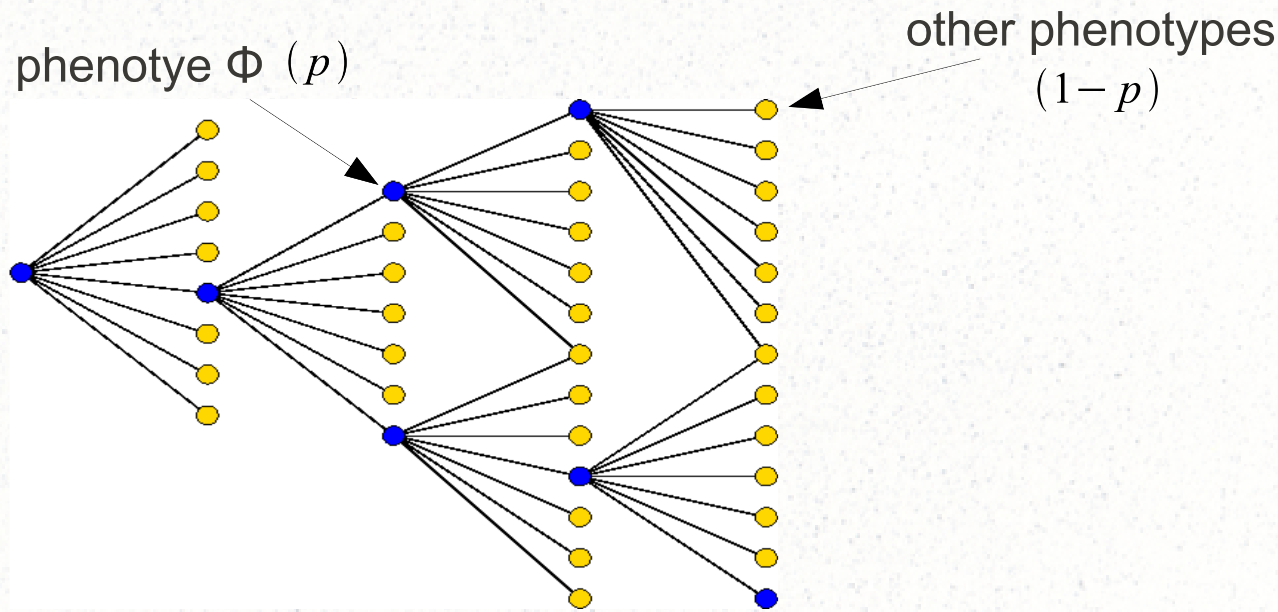



# Genotype-phenotype mapping



# Distribution of phenotypes

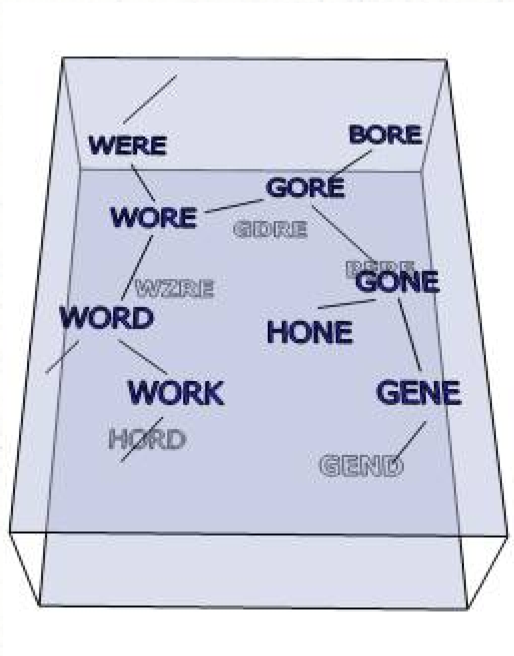
$$D \gg 1$$



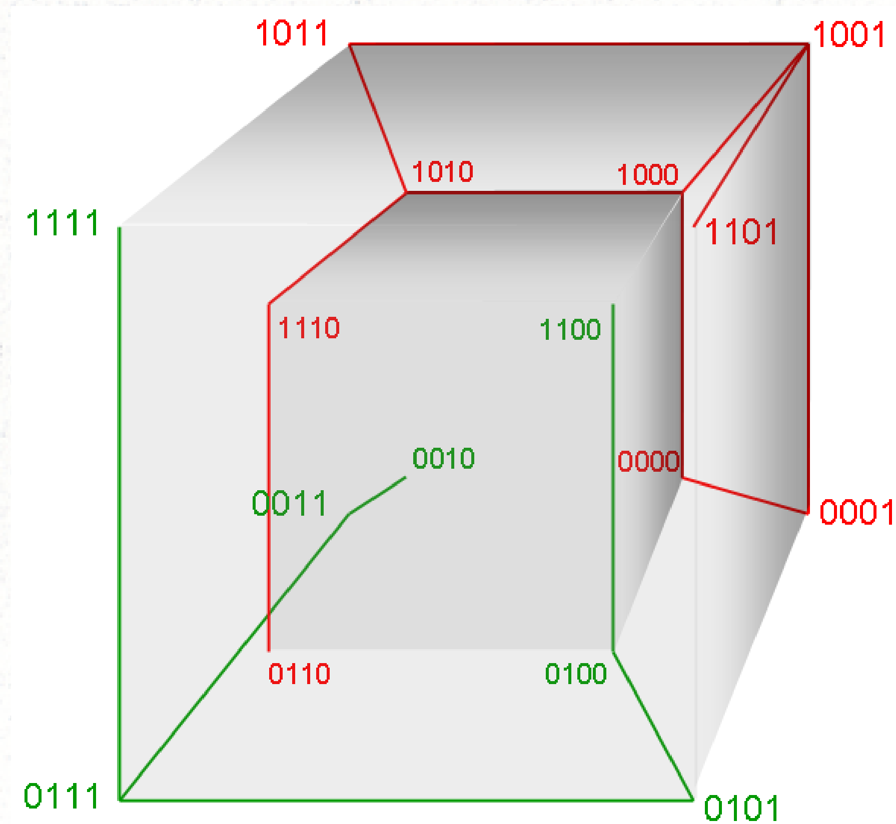

$$E\{k\} = (D-1)\,p_c = 1 \quad \Leftrightarrow \quad p_c = \frac{1}{D-1} \approx \frac{1}{D}$$

Maynard Smith, *Nature* (1970)

# *Neutral networks*




Maynard Smith, *Nature* (1970)



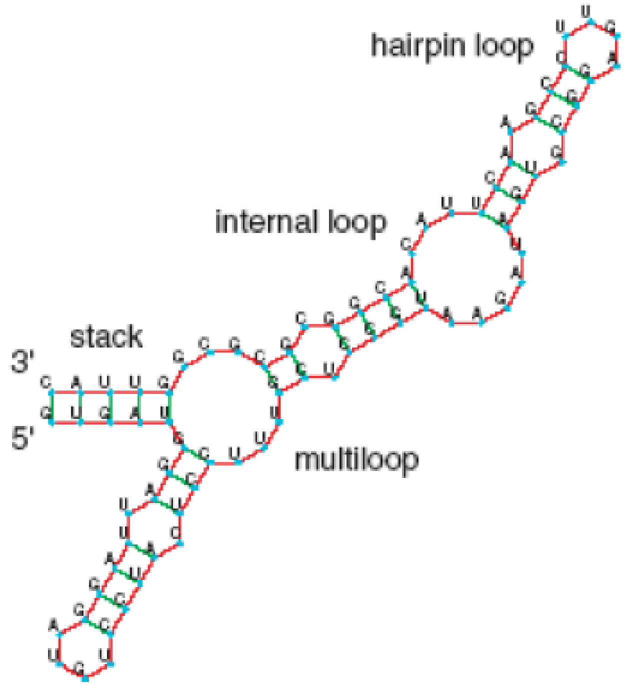
# Working example: RNA

# RNA secondary structures

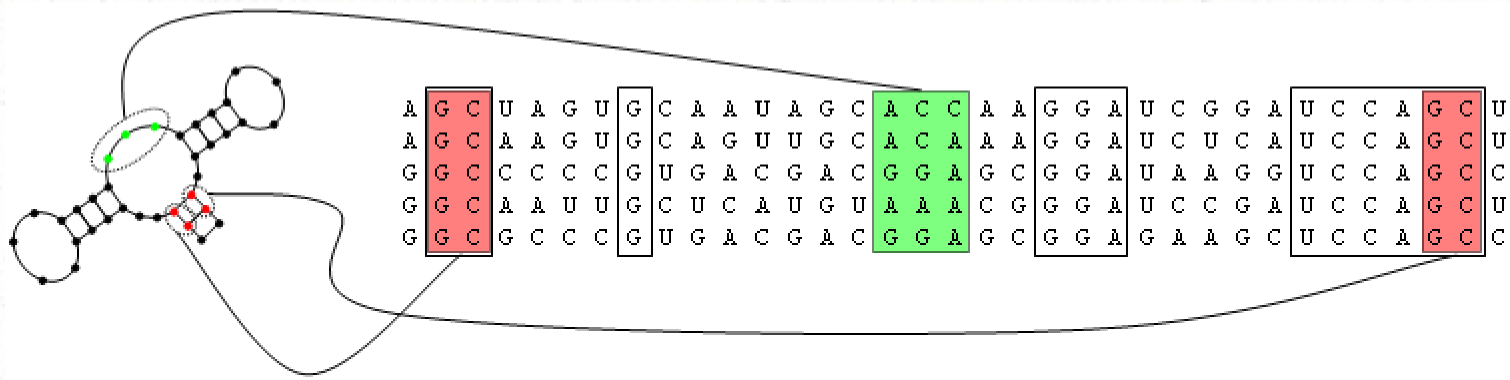
G≡C – 3 Kcal/mol

A=U – 2 Kcal/mol

G–U – 1 Kcal/mol



# RNA secondary structures




mean number of sequences of length *n* folding into the same secondary structure:

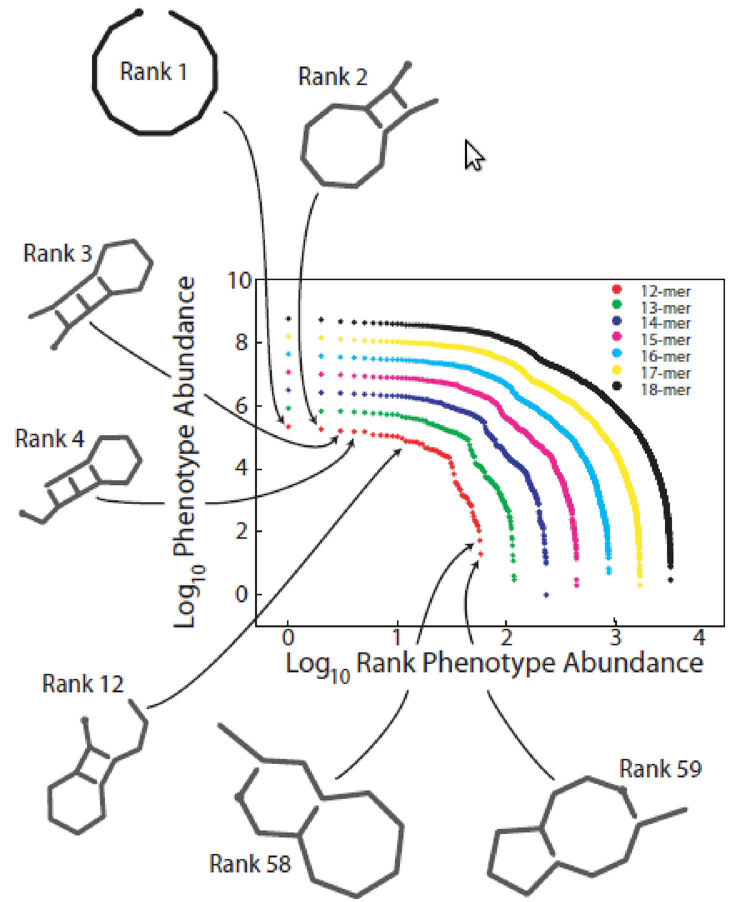
$$\sim 0.6735\, n^{3/2} (2.1635)^n$$

Schuster et al, *Proc. Roy. Soc. London B* (1994)

# RNA phenotype landscape

$p_c$=2.8%

| | Rank | Abundance | Nc | Dhamm | Dspl |
|---|---|---|---|---|---|
| 85.4% | 1 | 14325304 | nd | nd | nd |
| 1.3% | 2 | 218567 | 16 | 11 | 36 |
| 1.1% | 3 | 183791 | 10 | 12 | 32 |
| 1.0% | 4 | 162009 | 26 | 12 | 39 |
| | 5 | 152393 | 9 | 12 | 25 |
| | 6 | 151705 | 15 | 12 | 29 |
| | 7 | 122332 | 8 | 12 | 32 |
| | 8 | 117213 | 21 | 12 | 35 |
| | 9 | 113896 | 8 | 12 | 33 |
| | 10 | 110509 | 8 | 11 | 32 |
| | 11 | 105538 | 8 | 12 | 47 |
| | 12 | 93866 | 7 | 8 | 24 |
| | . . . . | | | | |
| | 38 | 2260 | 12 | 9 | 18 |
| | 39 | 2208 | 1 | 6 | 12 |
| | 40 | 1525 | 16 | 9 | 21 |
| | 41 | 1379 | 15 | 7 | 14 |
| | 42 | 1368 | 2 | 5 | 10 |
| | 43 | 1299 | 22 | 8 | 16 |
| | 44 | 1188 | 34 | 7 | 12 |
| | 45 | 1139 | 23 | 8 | 18 |
| | 46 | 860 | 3 | 7 | 13 |
| | 47 | 800 | 3 | 6 | 15 |
| | 48 | 713 | 3 | 7 | 17 |
| | 49 | 665 | 15 | 8 | 15 |
| | 50 | 411 | 11 | 5 | 8 |
| | 51 | 314 | 3 | 4 | 6 |
| | 52 | 240 | 3 | 4 | 6 |
| | 53 | 220 | 4 | 4 | 8 |
| | 54 | 197 | 5 | 3 | 6 |
| | 55 | 165 | 4 | 6 | 10 |
| | 56 | 153 | 4 | 6 | 10 |
| | 57 | 109 | 6 | 6 | 12 |
| | 58 | 54 | 1 | 4 | 6 |
| | 59 | 20 | 1 | 4 | 8 |

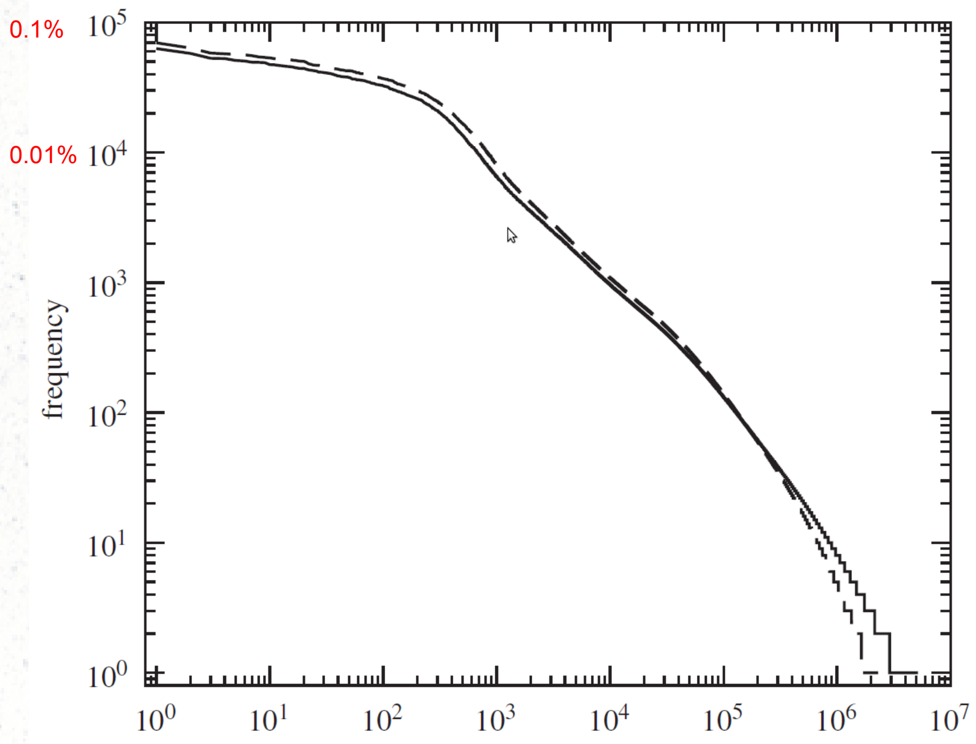



Cowperthwait et al, *PloS Comp. Biol.* (2008)

# *RNA phenotype landscape*

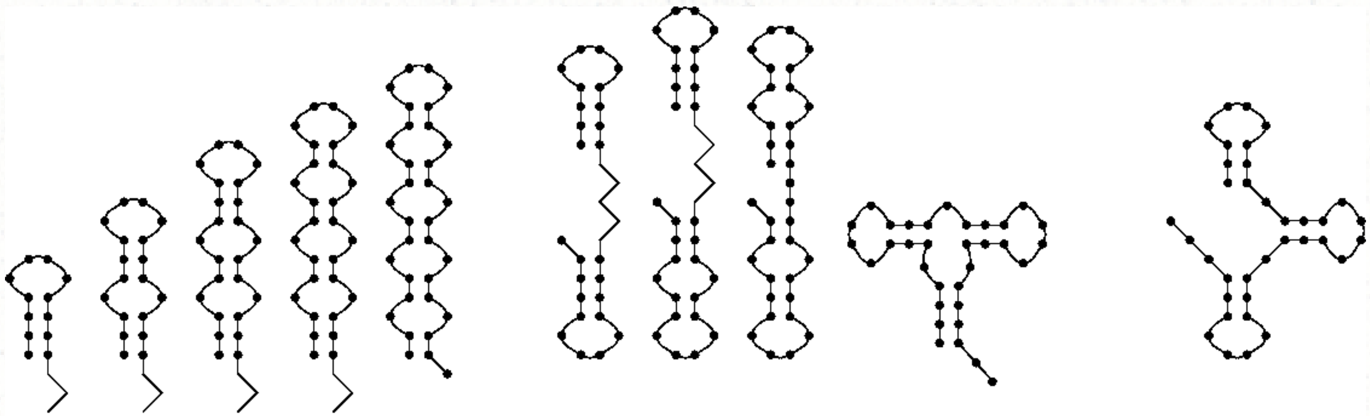
folding $10^8$ 35-mers

$p_c$=1.0%




Stich, Briones & Manrubia, *J. Theor. Biol.* (2008)

# *RNA phenotype landscape*

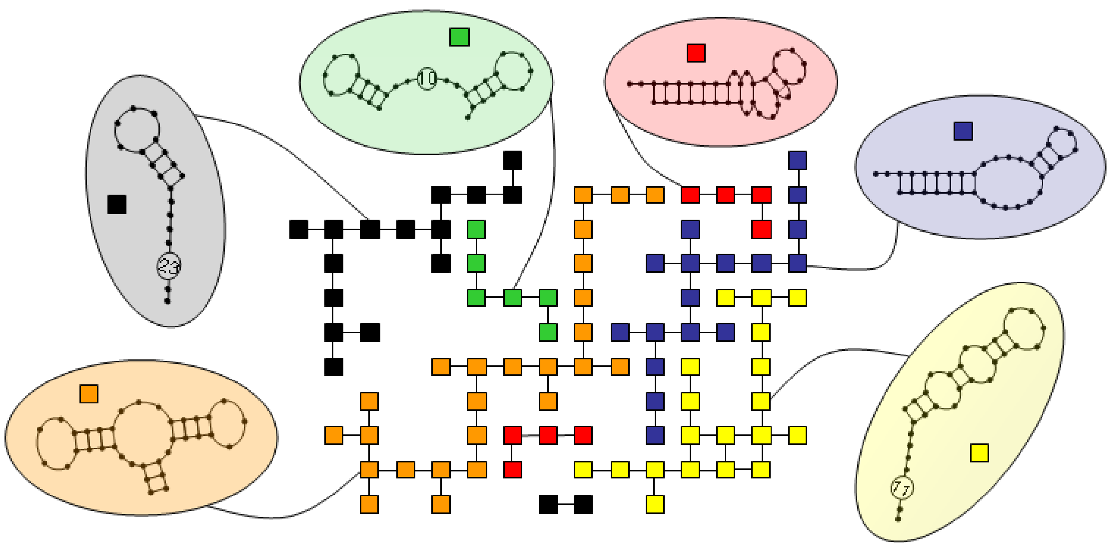


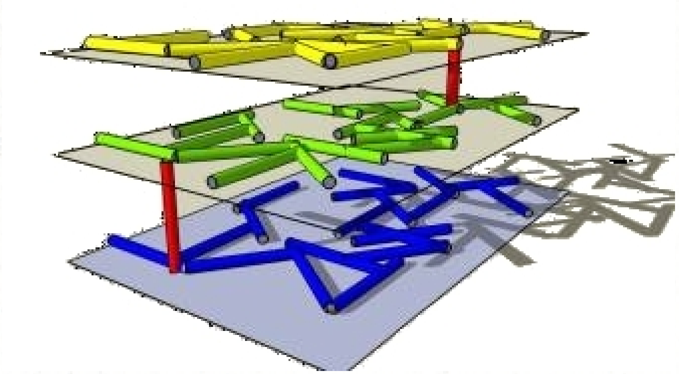

| Name | No. seq. | Ratio seq. (%) | No. struc. | Ratio struc. (%) | Seq./struc. |
|---|---|---|---|---|---|
| Open | 2 133 048 | 2.1330 | 1 | ⩽0.0001 | 2133048.0 |
| SL | 20 054 055 | 20.0541 | 2330 | 0.0451 | 8606.9 |
| HP | 37 359 257 | 37.3593 | 182 569 | 3.5359 | 204.6 |
| HP2 | 22 580 884 | 22.5809 | 1 624 464 | 31.4616 | 13.9 |
| HP3 | 4 583 268 | 4.5833 | 1 771 143 | 34.3024 | 2.6 |
| DSL | 7 782 386 | 7.7824 | 82 554 | 1.5989 | 94.3 |
| DSL2 | 4 295 840 | 4.2958 | 699 668 | 13.5507 | 6.1 |
| DSL3 | 427 878 | 0.4279 | 299 045 | 5.7917 | 1.4 |
| HH | 433 103 | 0.4331 | 203 886 | 3.9487 | 2.1 |
| TSL | 23 495 | 0.0235 | 16 652 | 0.3225 | 1.4 |

Stich, Briones & Manrubia, *J. Theor. Biol.* (2008)

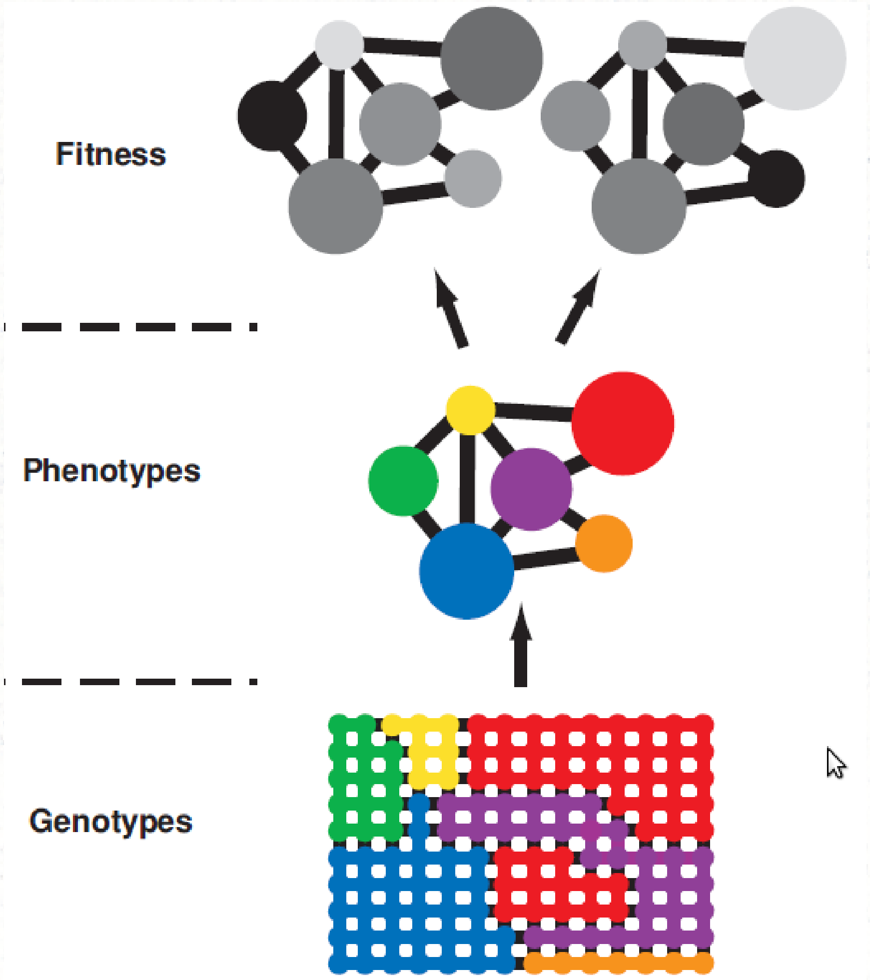
# Landscapes revisited

# Distribution of phenotypes

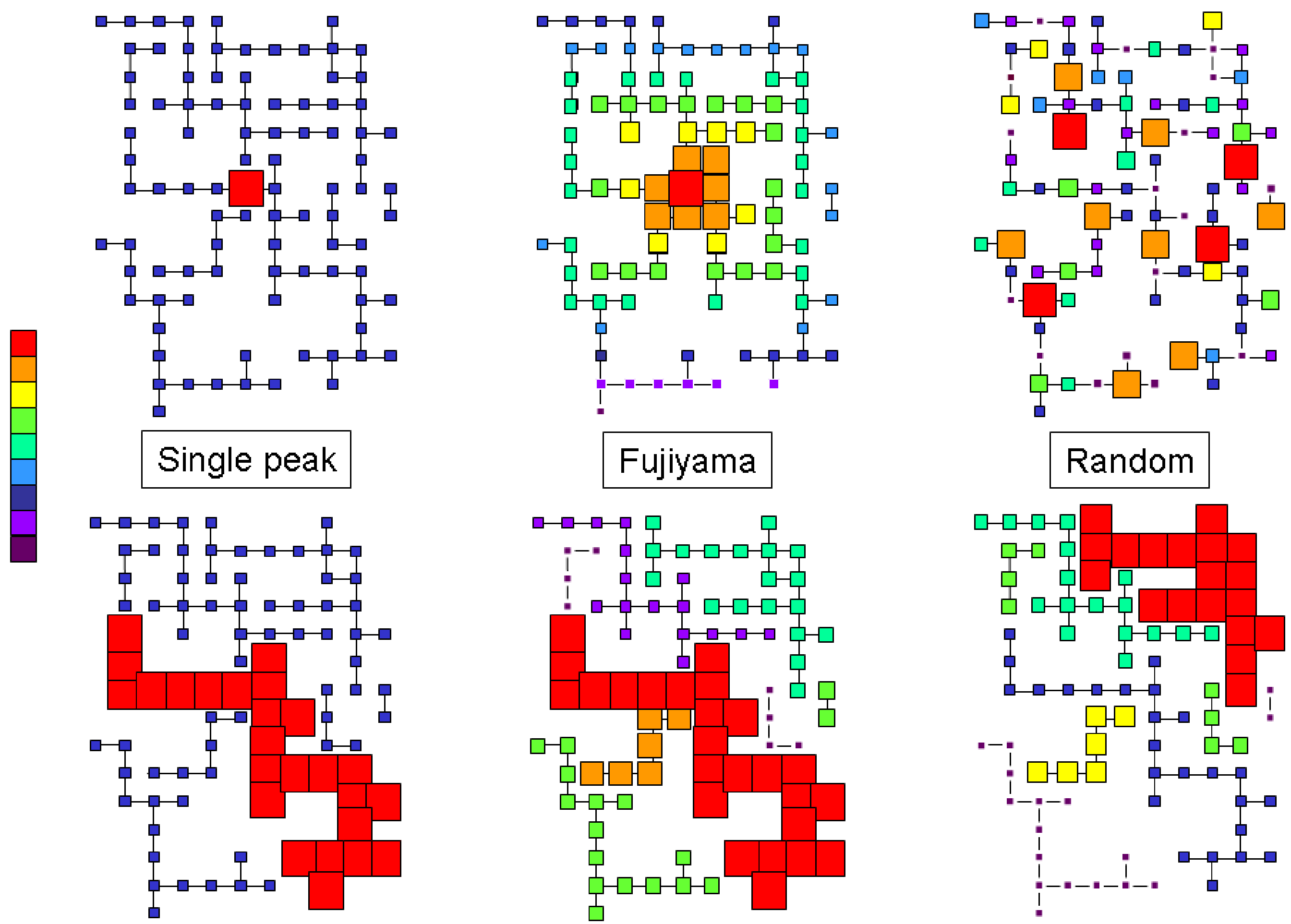


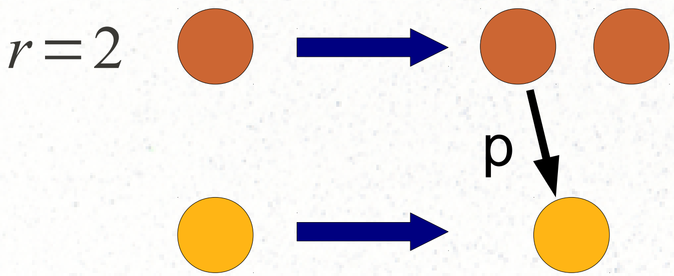
# Phenotype landscapes



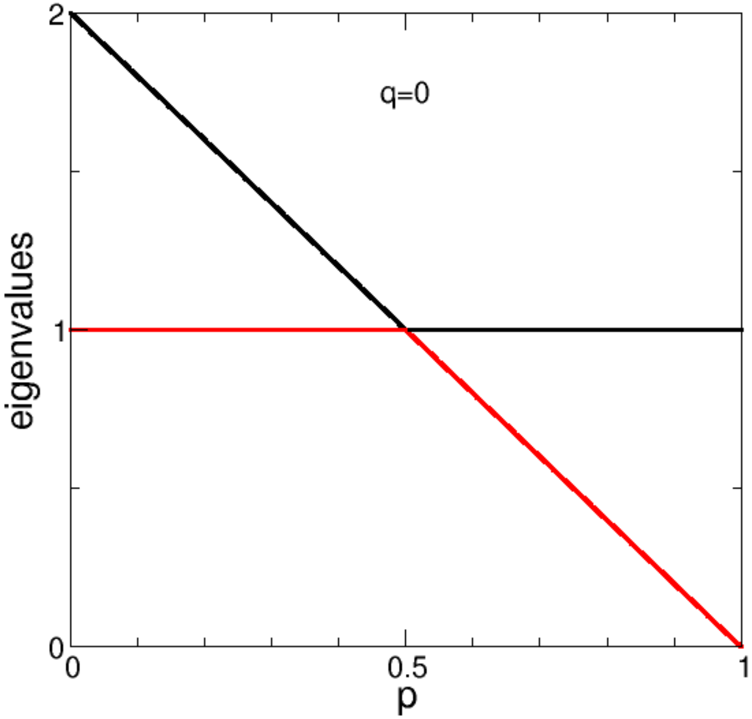
# New metaphors

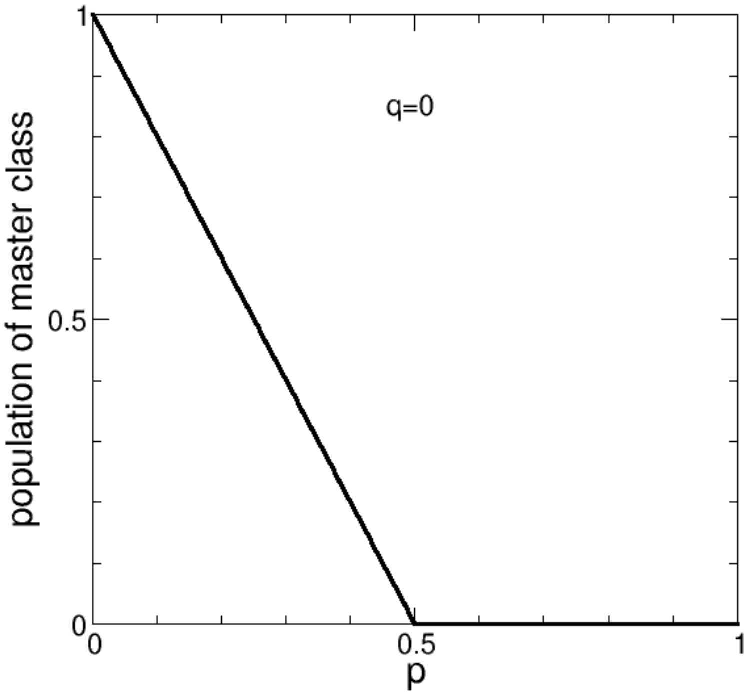


# Error catastrophe?

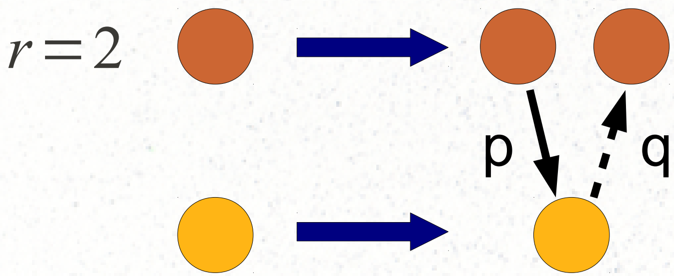



$$W = \begin{pmatrix} r(1-p) & r\,p \\ 0 & 1 \end{pmatrix}$$

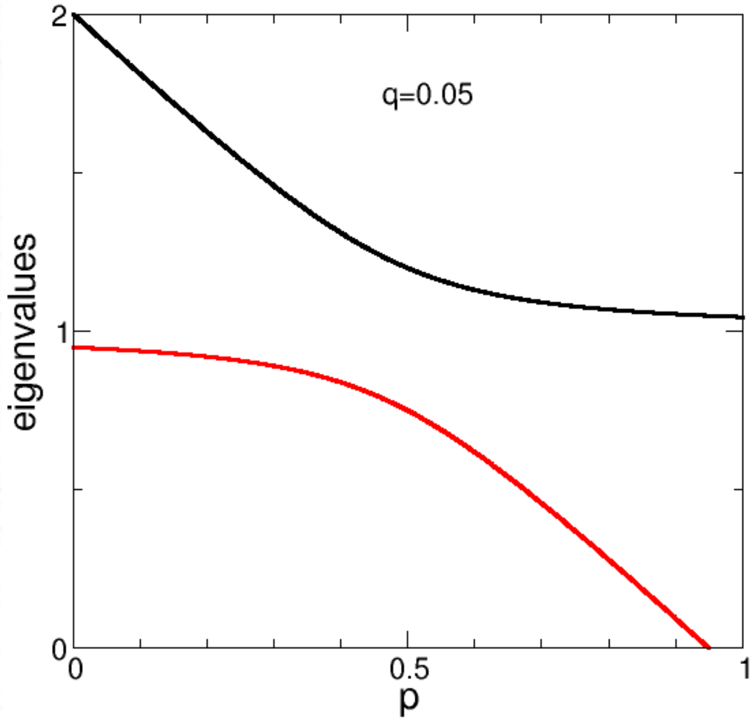


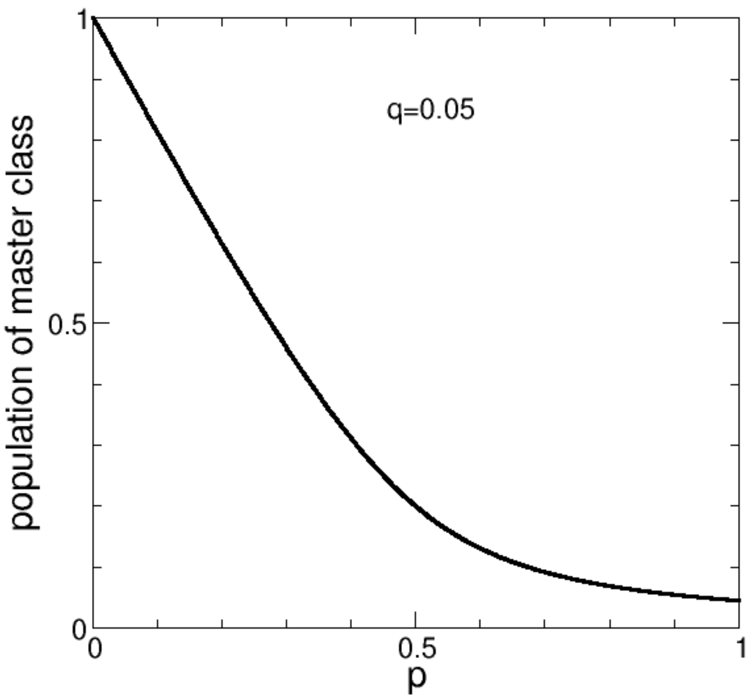



# Error catastrophe?

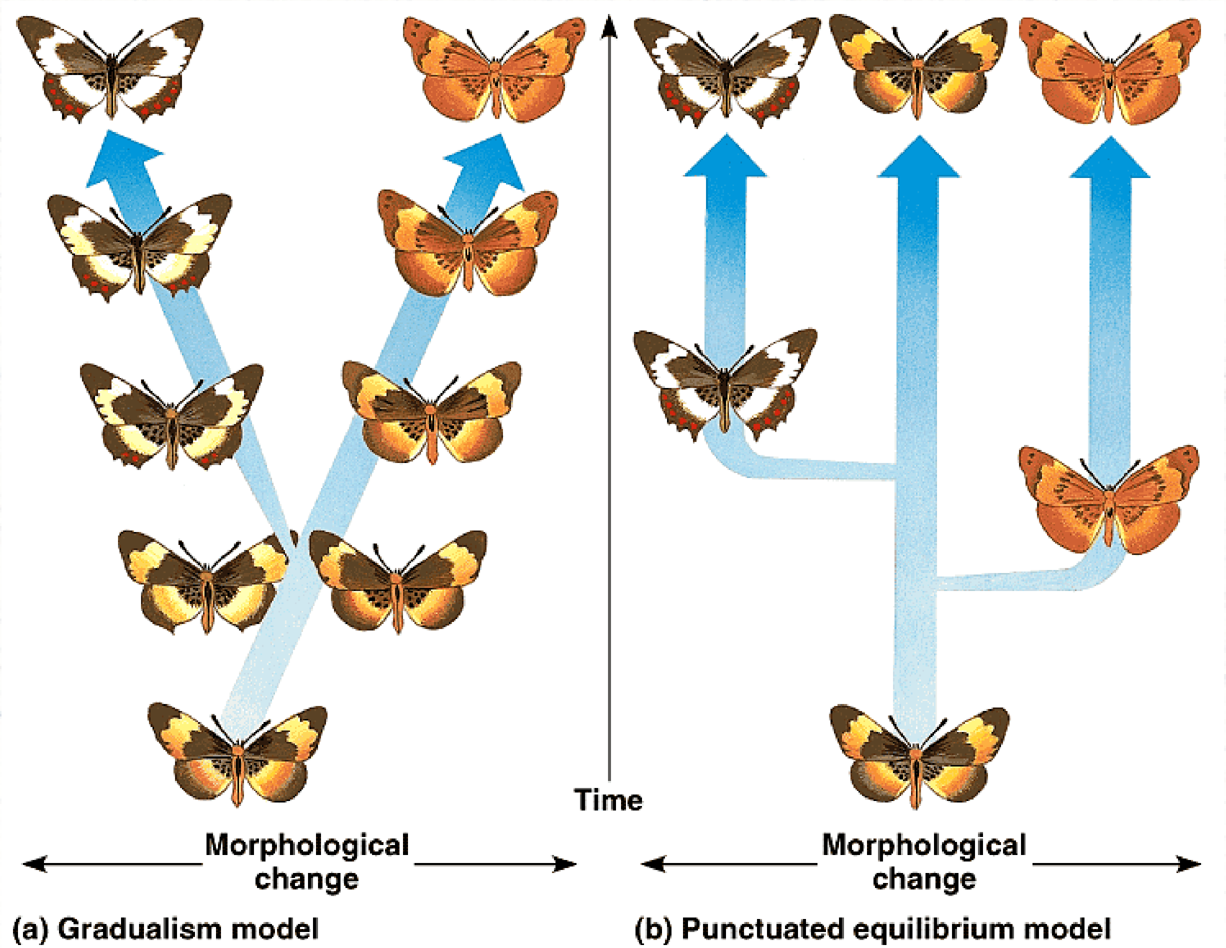



$$W=\begin{pmatrix} r(1-p) & r\,p \\ q & 1-q \end{pmatrix}$$



# Punctuated equilibrium



# Conclusions

- Current paradigm: population genetics
  - genotype-based
  - gradual evolution
  - problems to explain speciation
- New paradigm: neutral evolution
  - most evolution is neutral: great redundancy
  - many genotypes map to the same phenotype
  - a few phenotypes dominate the genotype space
  - small changes in genotype may induce big changes in phenotype
  - evolutionary preference for abundant phenotypes
    (abundance - replicability $\leftrightarrow$ entropy - energy)
  - punctuated equilibrium (at least at molecular level)