Bridging from Chemistry to the Life Sciences

Evolution seen with the Glasses of a Physicist

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Austria

and

The Santa Fe Institute, Santa Fe, New Mexico, USA

Manfred Eigen Award Lecture

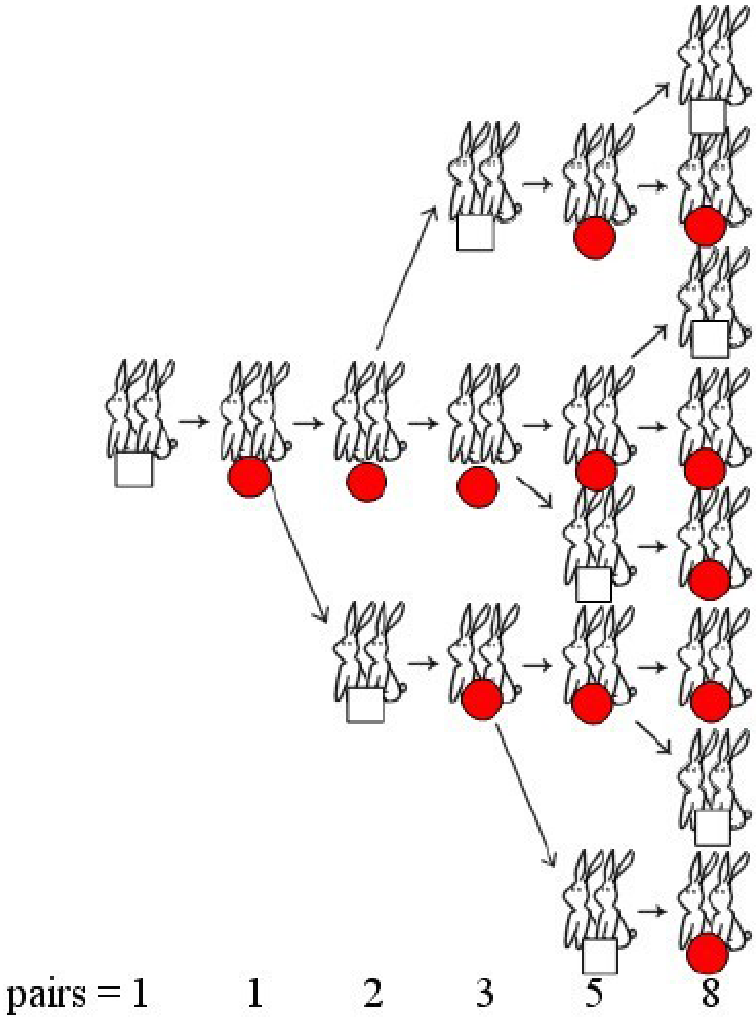
Göttingen, 09.05.2018

Web-Page for further information:

http://www.tbi.univie.ac.at/~pks

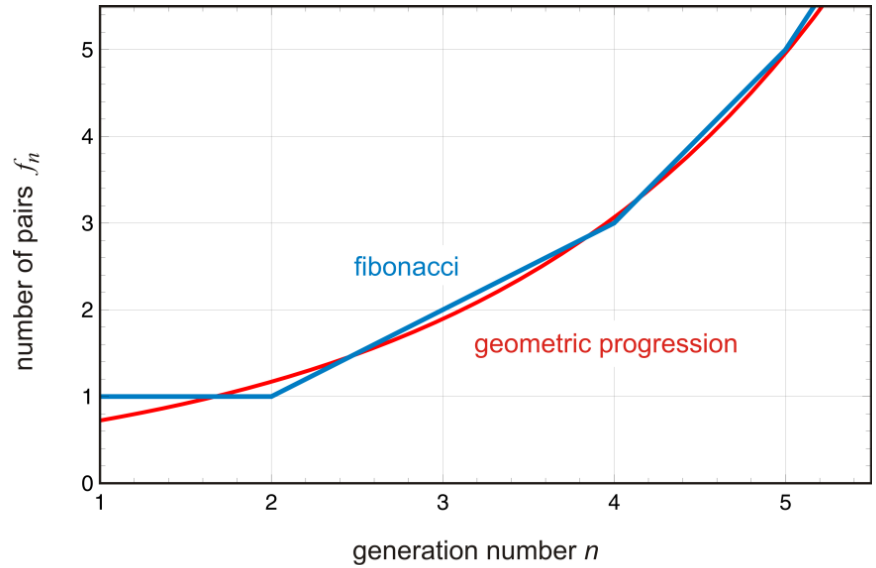
1. Mathematical concepts before Darwin
2. Theory of molecular evolution
3. Evolution in realistically small populations
4. Fitness landscapes and evolution
5. Evolution and present day molecular genetics

1. **Mathematical concepts before Darwin**
2. Theory of molecular evolution
3. Evolution in realistically small populations
4. Fitness landscapes and evolution
5. Evolution and present day molecular genetics



Leonardo da Pisa
„Fibonacci“
~1180 – ~1240

Fibonacci series: 1,1,2,3,5,8,13,21,…



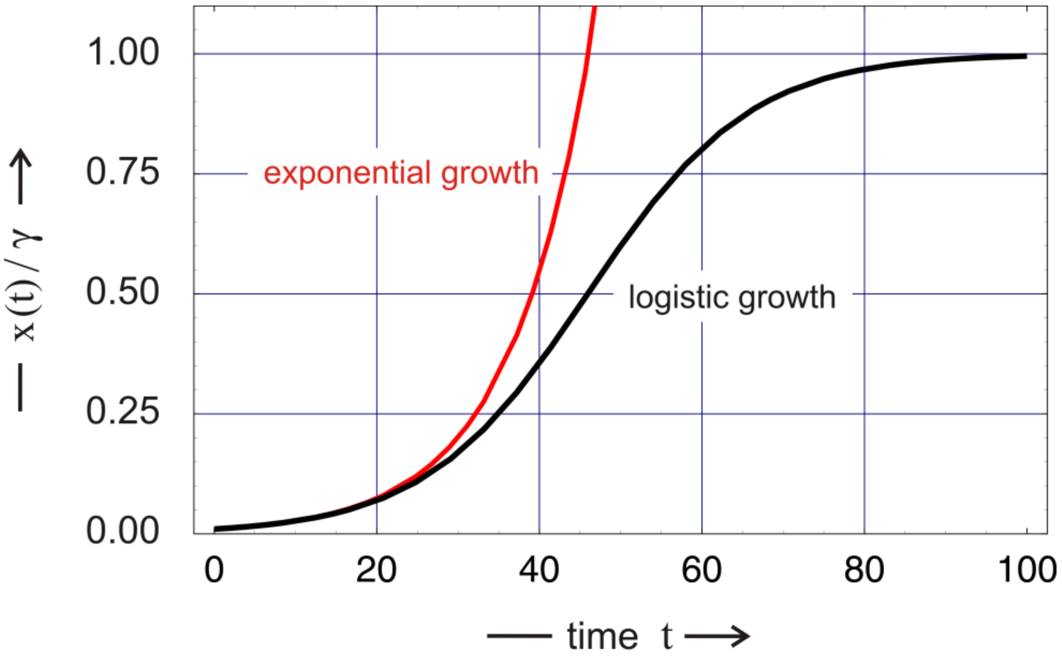
geometric progression

exponential function

Thomas Robert Malthus,
1766 – 1834

Leonhard Euler,
1717 – 1783

Pierre-François Verhulst,
1804-1849

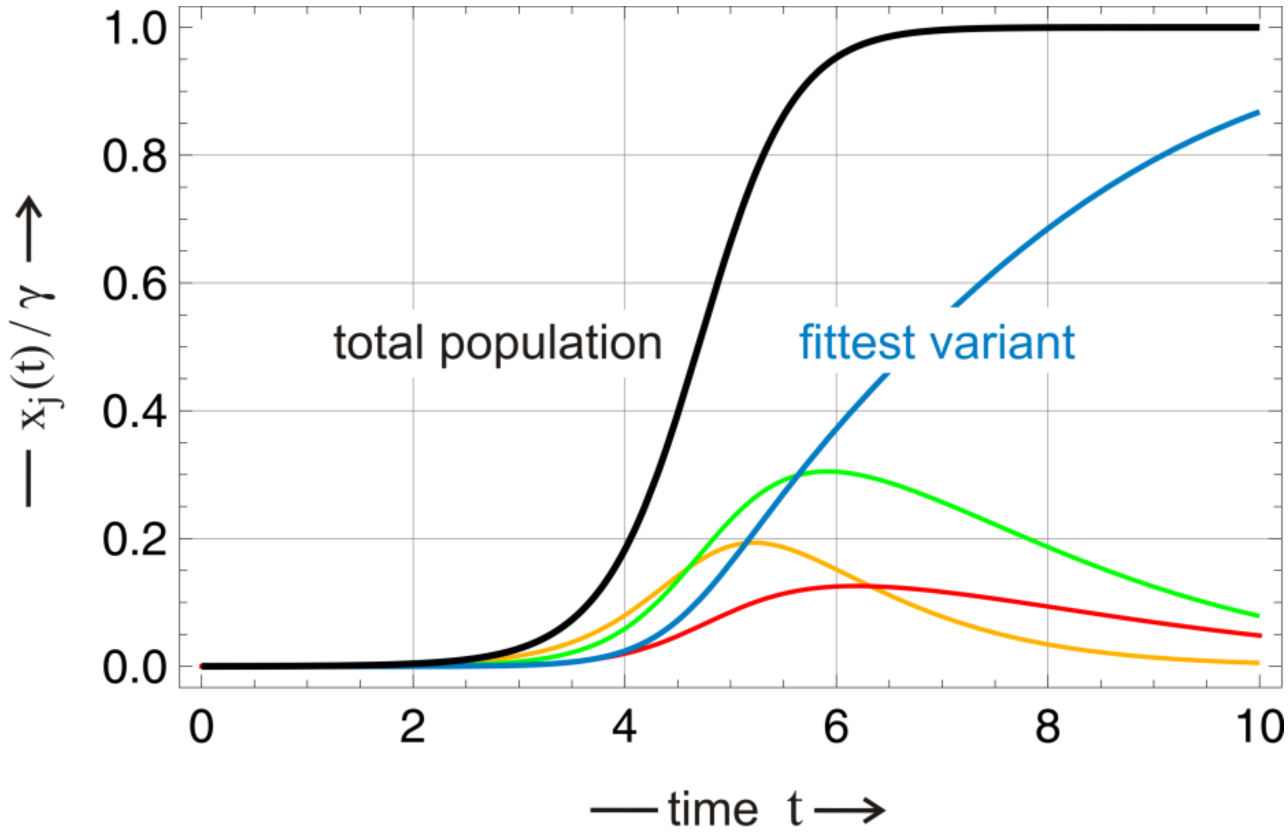


population: $\Pi = \{X\}$

the consequence of finite resources

$$\frac{dx}{dt} = f\,x\left(1 - \frac{x}{\gamma}\right) \Rightarrow x(t) = \frac{x_0\,\gamma}{x_0 + (\gamma - x_0)\exp(-ft)}$$

the logistic equation: Verhulst 1838



fitness values: $f_1 = 2.80$, $f_2 = 2.35$, $f_3 = 2.25$, and $f_4 = 1.75$

population: $\Pi = \{X_1, X_2, X_3, X_4\}$

logistic growth leads to natural selection in heterogeneous populations

$$A + X \longrightarrow 2X$$

autocatalysis

$$\frac{dx}{dt} = f\,x \quad \Rightarrow \quad x(t) = x(0)\,\exp(f\,t)$$

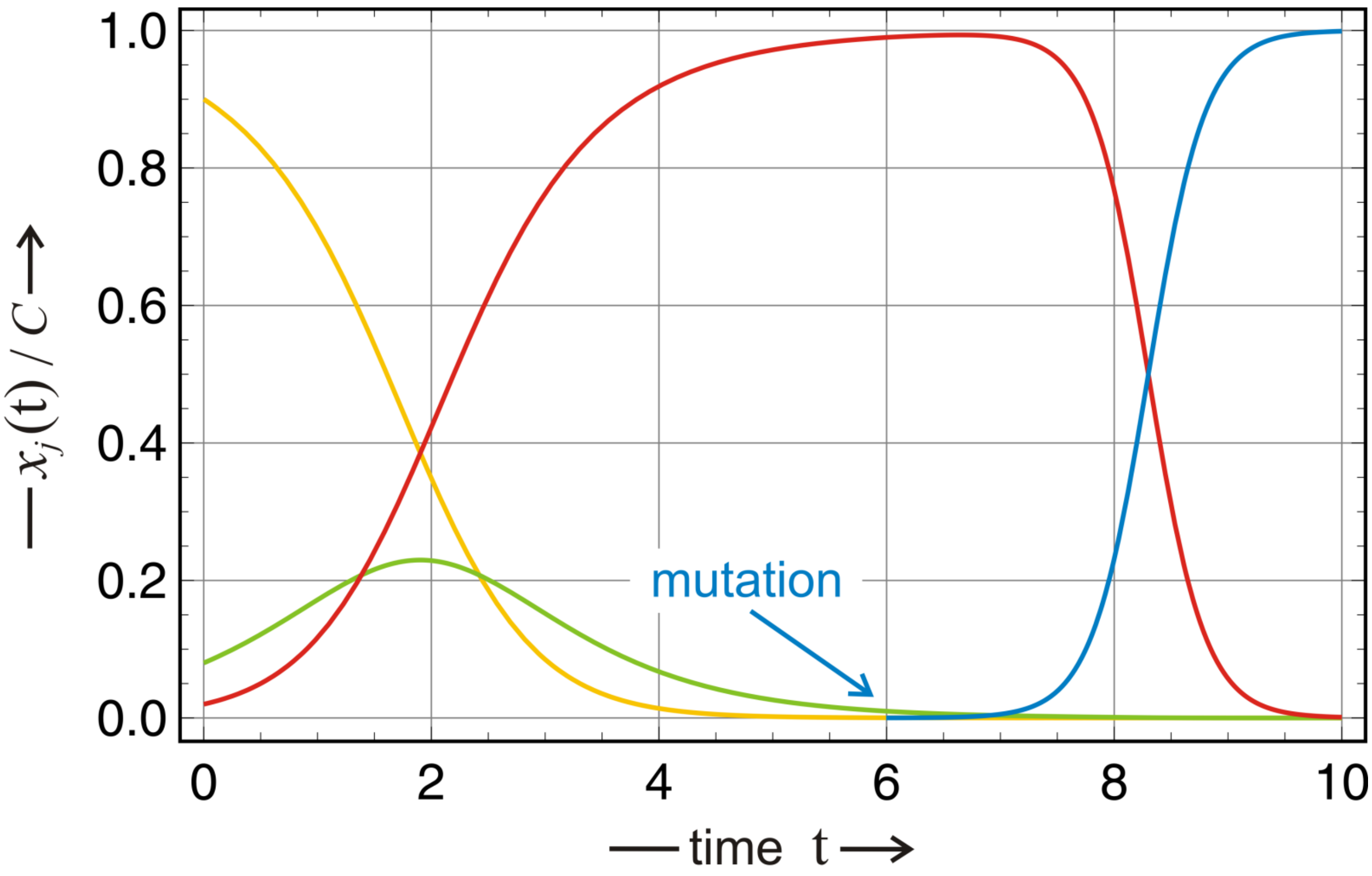
$$A + X_k \longrightarrow 2X_k \;\;;\;\; k = 1, 2, \ldots, n$$

competition

$$\frac{dx_k}{dt} = f_k\,x_k\,;\ \mathrm{k} = 1, 2, \ldots, \mathrm{n}\,;\ \sum_{k=1}^{n} x_k = 1$$

$$x_k(t) = \frac{x_k(0)\,\exp(f_k\,t)}{\sum_{j=1}^{n} x_j(0)\,\exp(f_j\,t)} \quad \Rightarrow \quad \text{selection of the fittest}$$

The chemistry and the mathematics of reproduction



$f_1 = 1\ ,\ f_2 = 2\ ,\ f_3 = 3\ ,\ f_4 = 7$

before the development of molecular biology mutation was treated as a "deus ex machina"

Charles Darwin's five ideas:

(i) evolution has happened and species change,

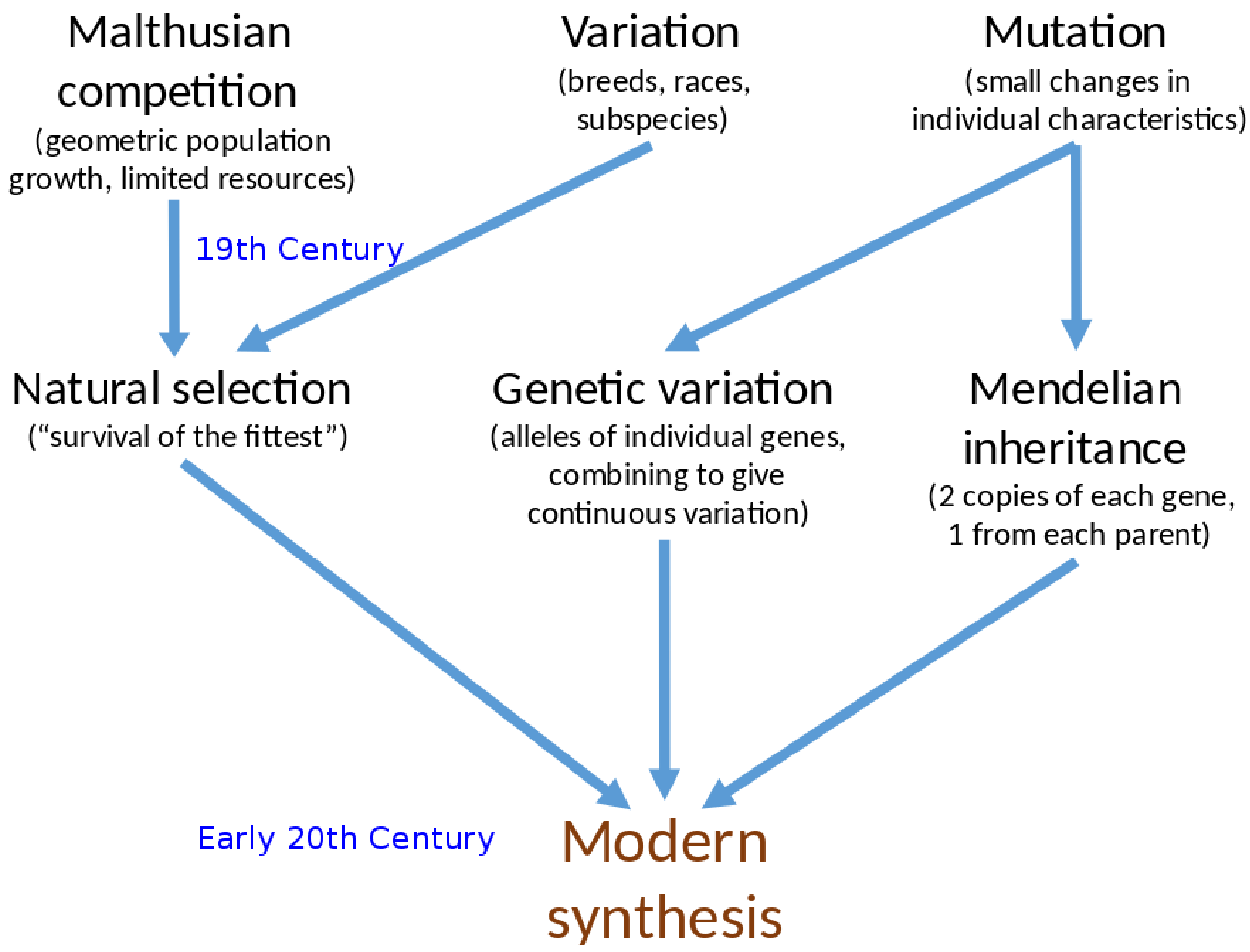
(ii) multiplication of species led to biological diversity,

(iii) all life had a common ancestor,

(iv) all change happened gradually, and

(v) natural selection.

Ernst Mayr, Ulrich Kutschera

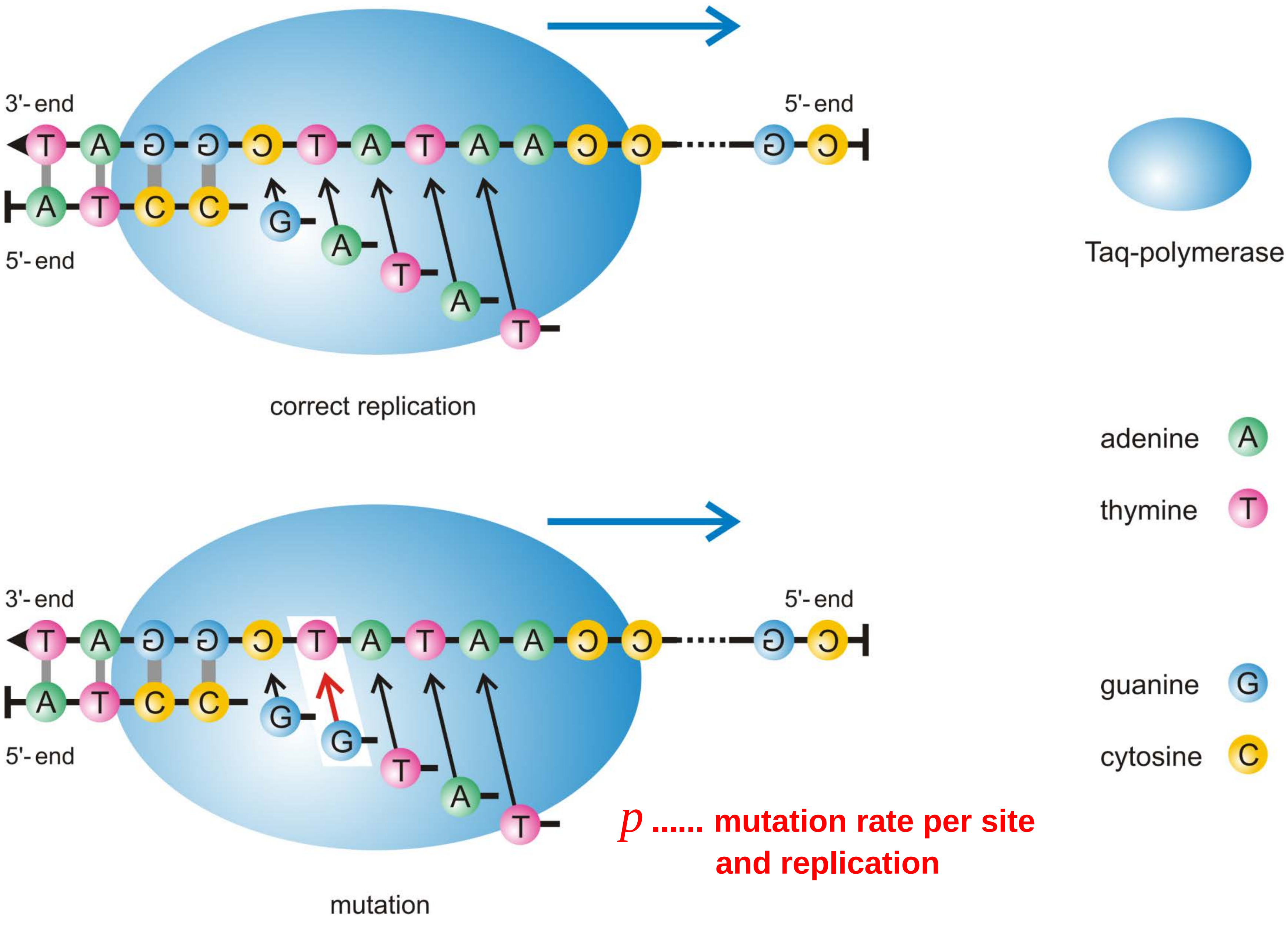

Malthusian competition
(geometric population growth, limited resources)
Variation
(breeds, races, subspecies)
Mutation
(small changes in individual characteristics)
19th Century
Natural selection
("survival of the fittest")
Genetic variation
(alleles of individual genes, combining to give continuous variation)
Mendelian inheritance
(2 copies of each gene, 1 from each parent)
Early 20th Century
Modern synthesis

1. Mathematical concepts before Darwin

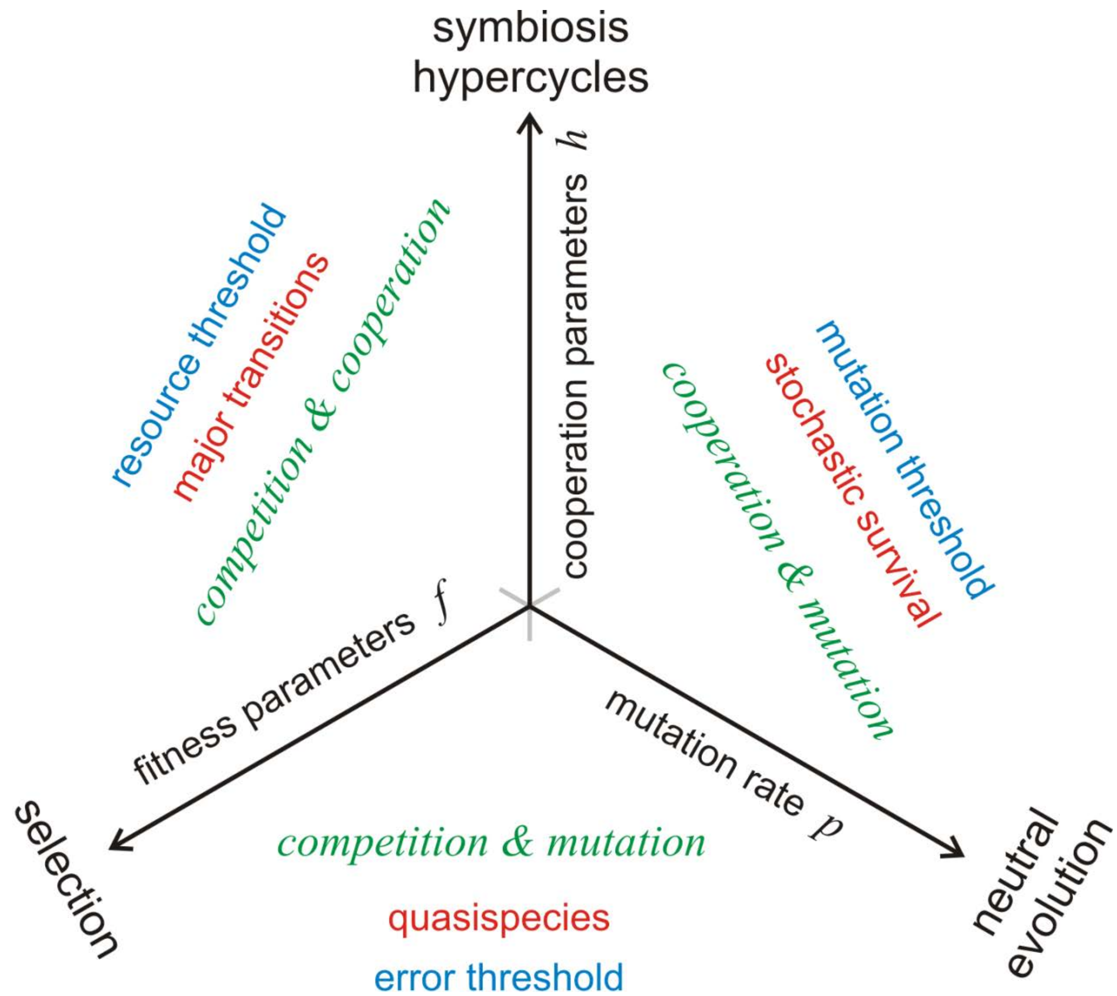
2. **Theory of molecular evolution**

3. Evolution in realistically small populations

4. Evolution and present day molecular genetics



DNA replication and mutation



evolution dependent on three basic parameters

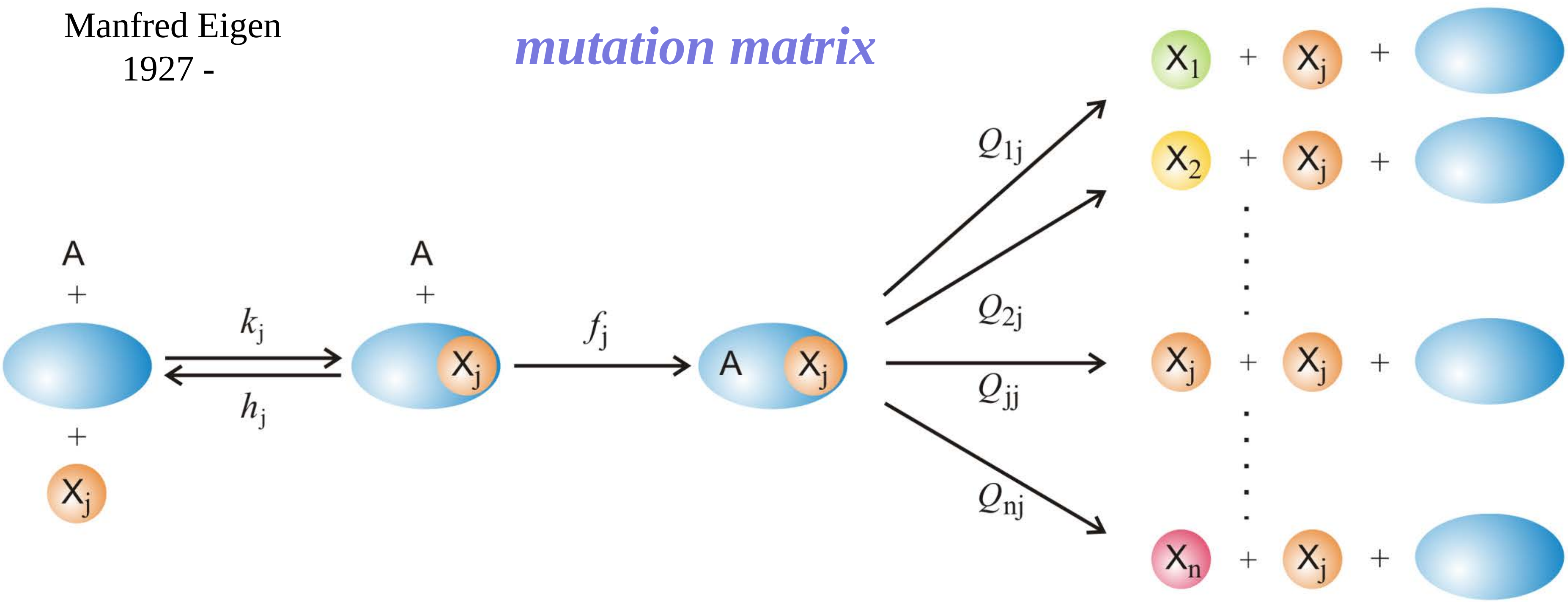
Manfred Eigen
1927 -

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} W_{ji}\, x_i - x_j\, \Phi\,; \quad j = 1, 2, \ldots, n$$

$$W_{ji} = Q_{ji} \cdot f_i\,, \quad \sum_{i=1}^{n} x_i = 1, \quad \Phi = \sum_{i=1}^{n} f_i\, x_i$$

fitness landscape

mutation matrix



Mutation and (correct) replication as parallel chemical reactions

M. Eigen. 1971. *Naturwissenschaften* 58:465,
M. Eigen & P. Schuster.1977-78. *Naturwissenschaften* 64:541, 65:7 und 65:341

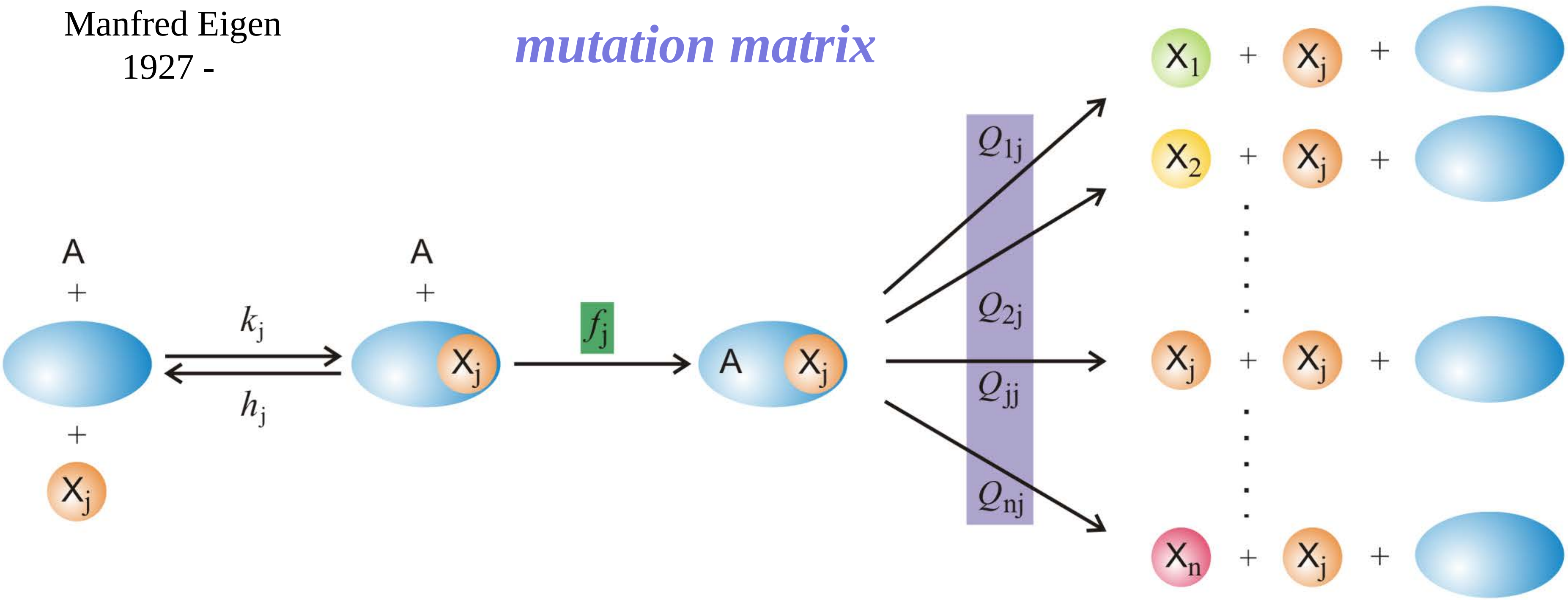
Manfred Eigen
1927 -

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} W_{ji}\, x_i - x_j\, \Phi\,; \quad j = 1, 2, \ldots, n$$

$$W_{ji} = Q_{ji} \cdot f_i\,, \quad \sum_{i=1}^{n} x_i = 1, \quad \Phi = \sum_{i=1}^{n} f_i\, x_i$$

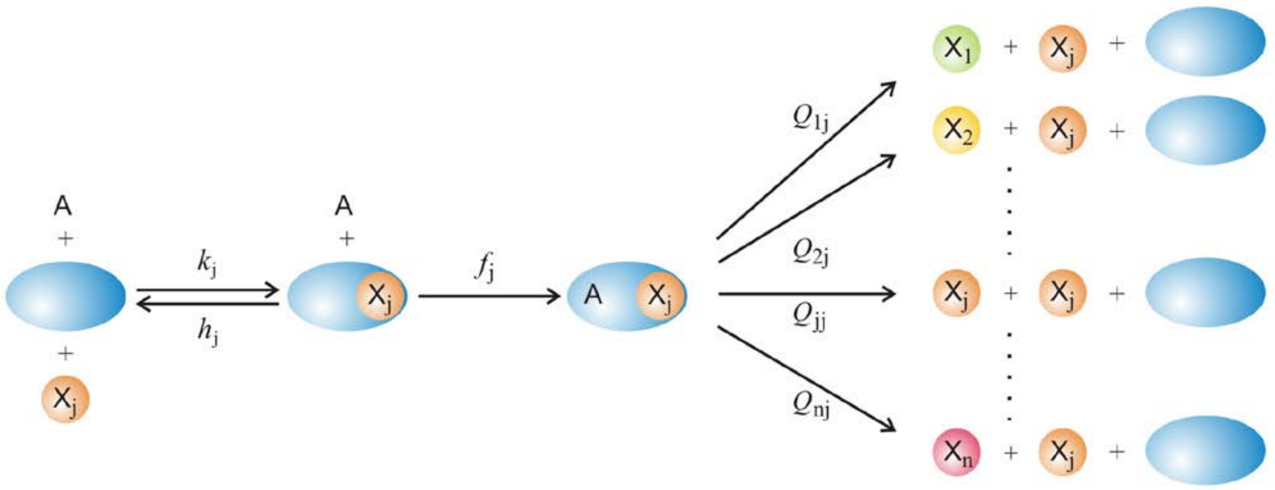
fitness landscape

mutation matrix

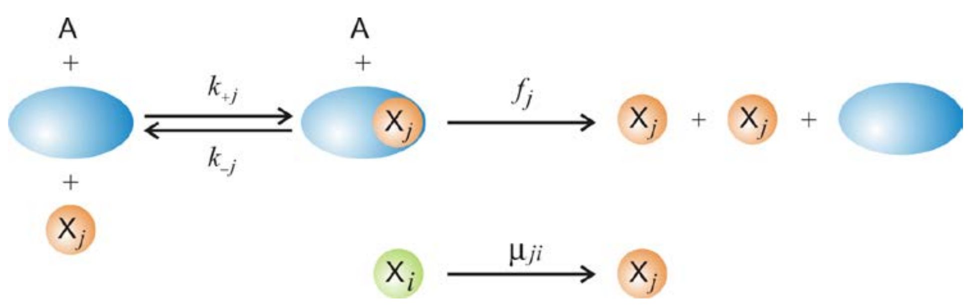


Mutation and (correct) replication as parallel chemical reactions

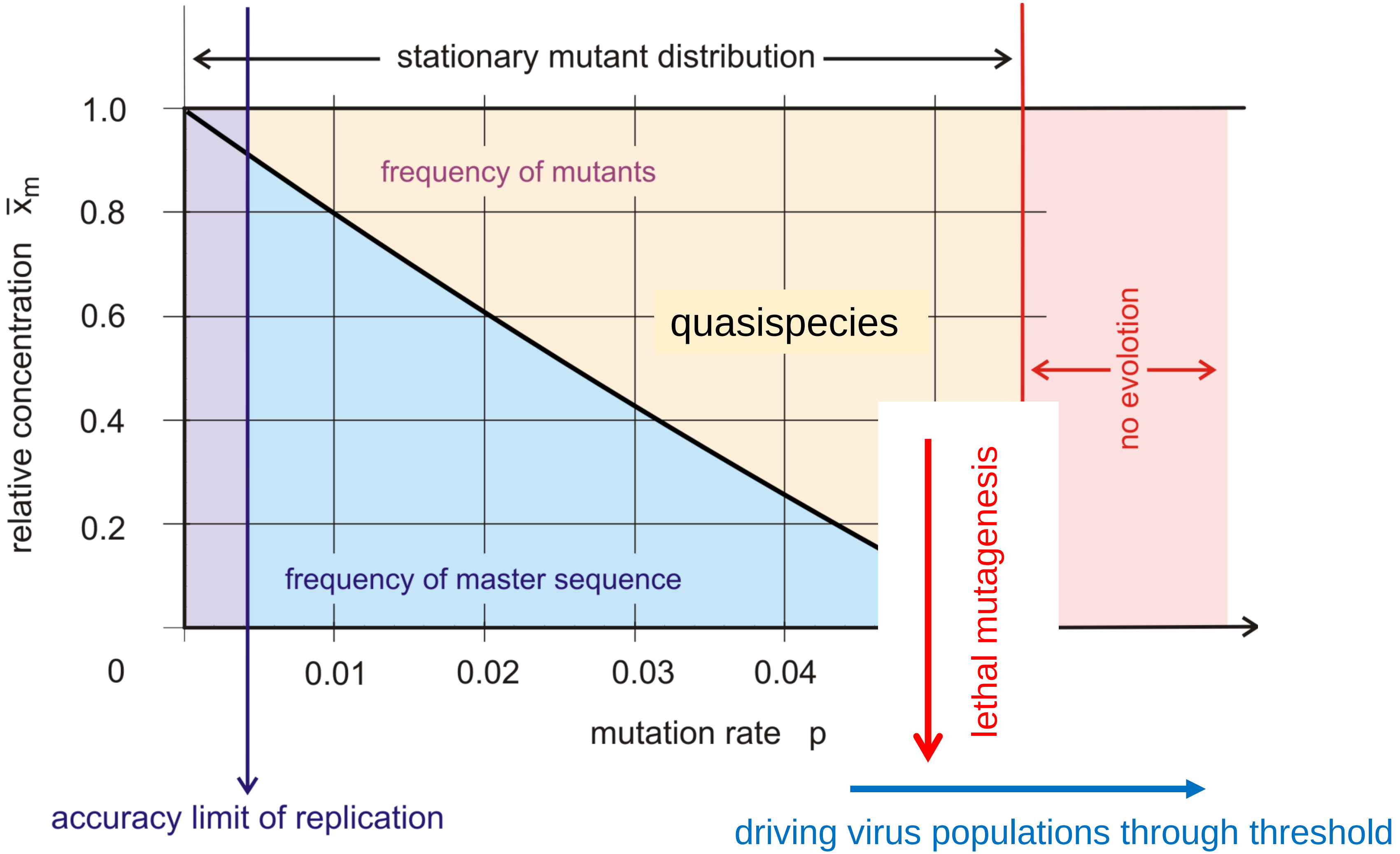
M. Eigen. 1971. *Naturwissenschaften* 58:465,
M. Eigen & P. Schuster.1977-78. *Naturwissenschaften* 64:541, 65:7 und 65:341



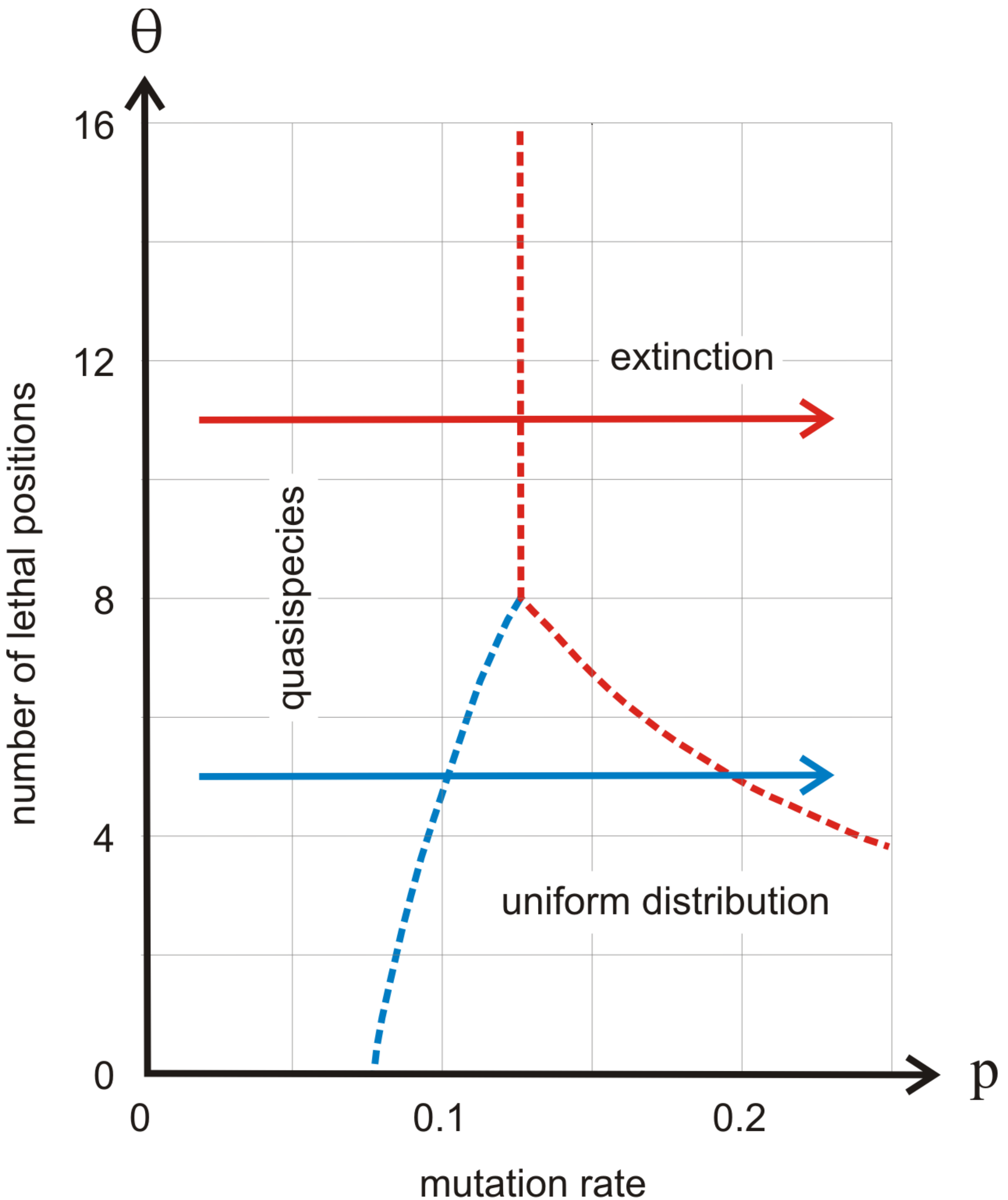
Manfred Eigen. 1971. *Naturwissenschaften* 58:465



James F.Crow, Motoo Kimura. An introduction to population genetics theory.
Harper & Row, New York, 1970, p.265

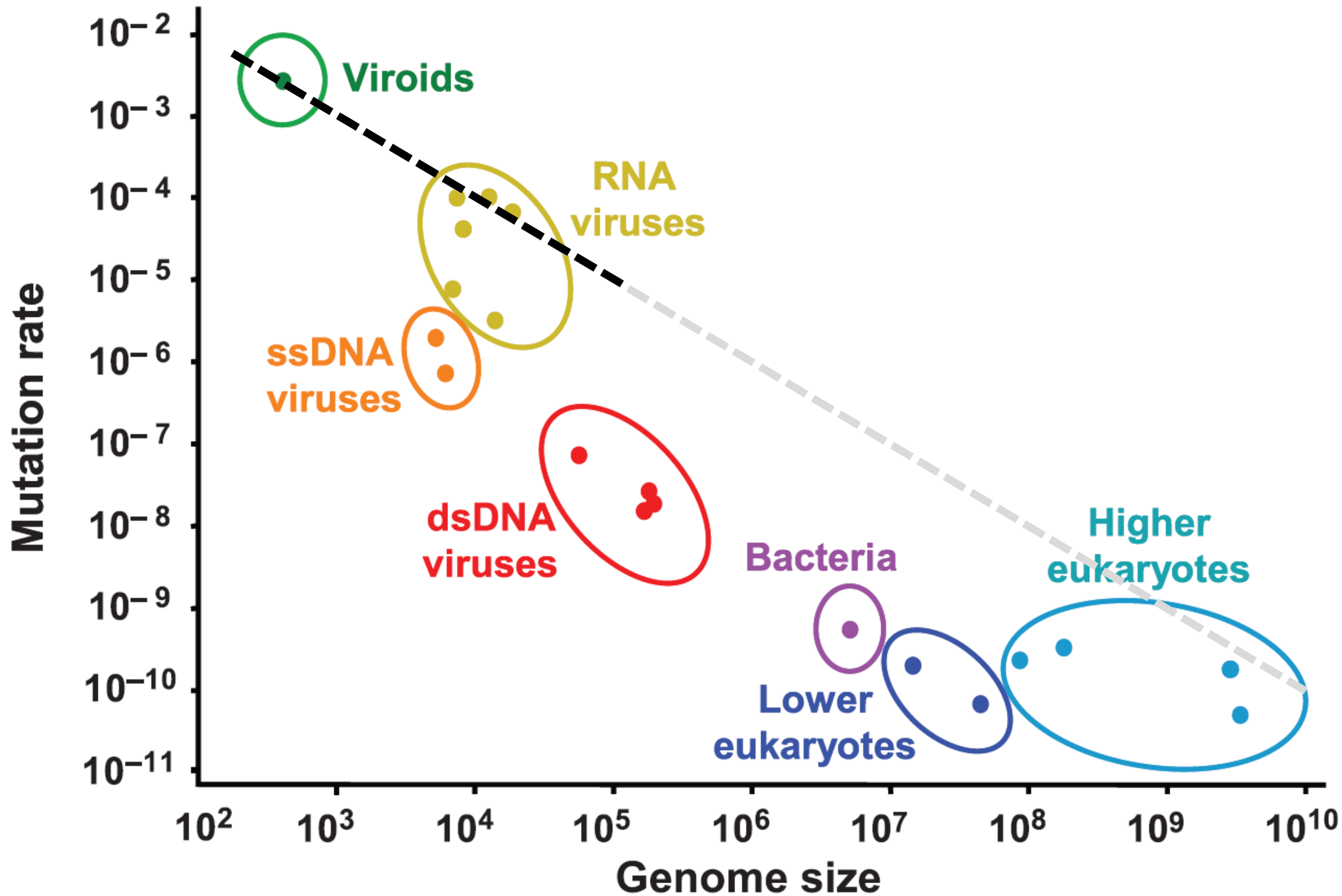


the error threshold in replication



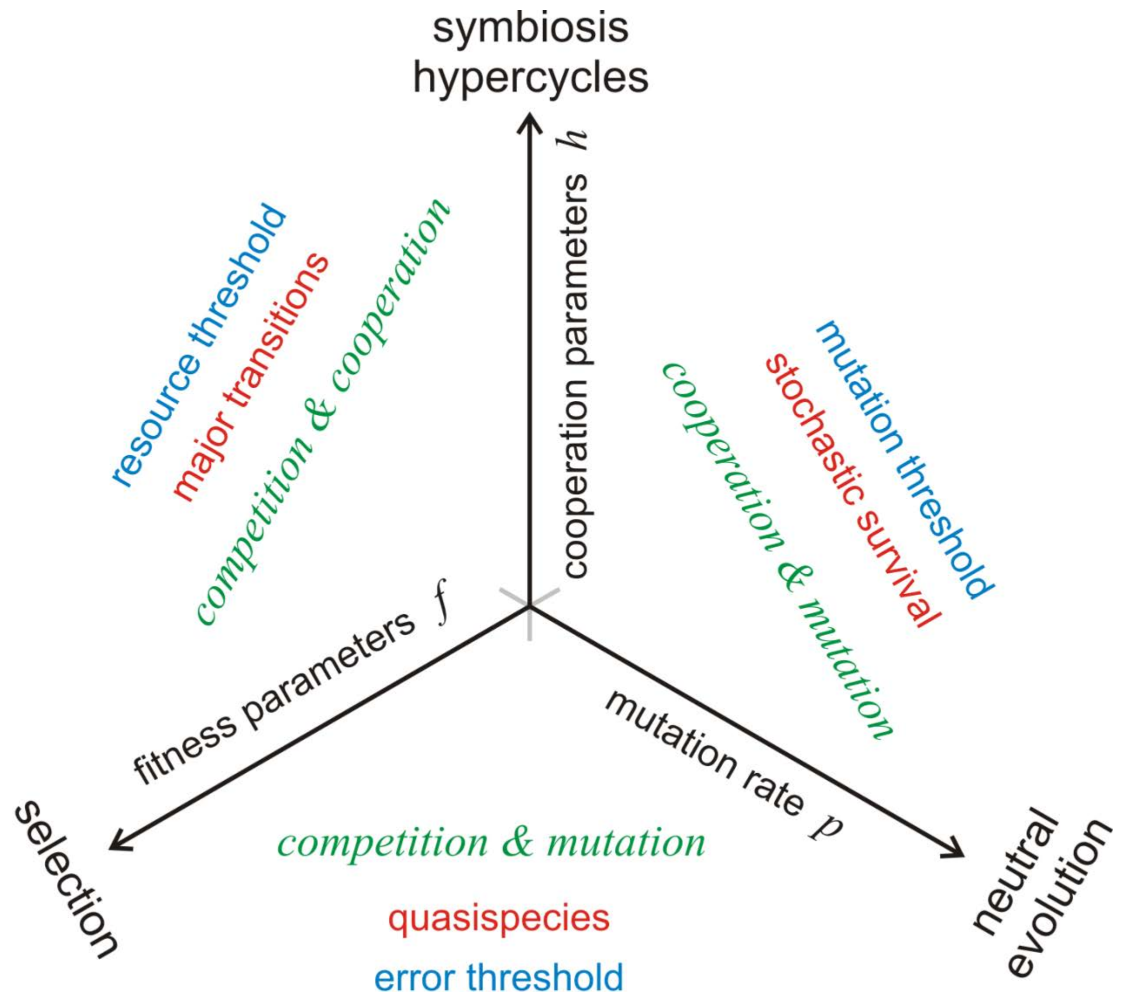
error threshold and
lethal mutagenesis

Héctor Tejero, Aeturo Marín,
Francisco Montero. 2010.
J.Theor.Biol. 262:733-741.



Selma Gago, Santiago F. Elena, Ricardo Flores, Rafael Sanjuán. 2009, Extremely high mutation rate of a hammerhead viroid. Science 323:1308.

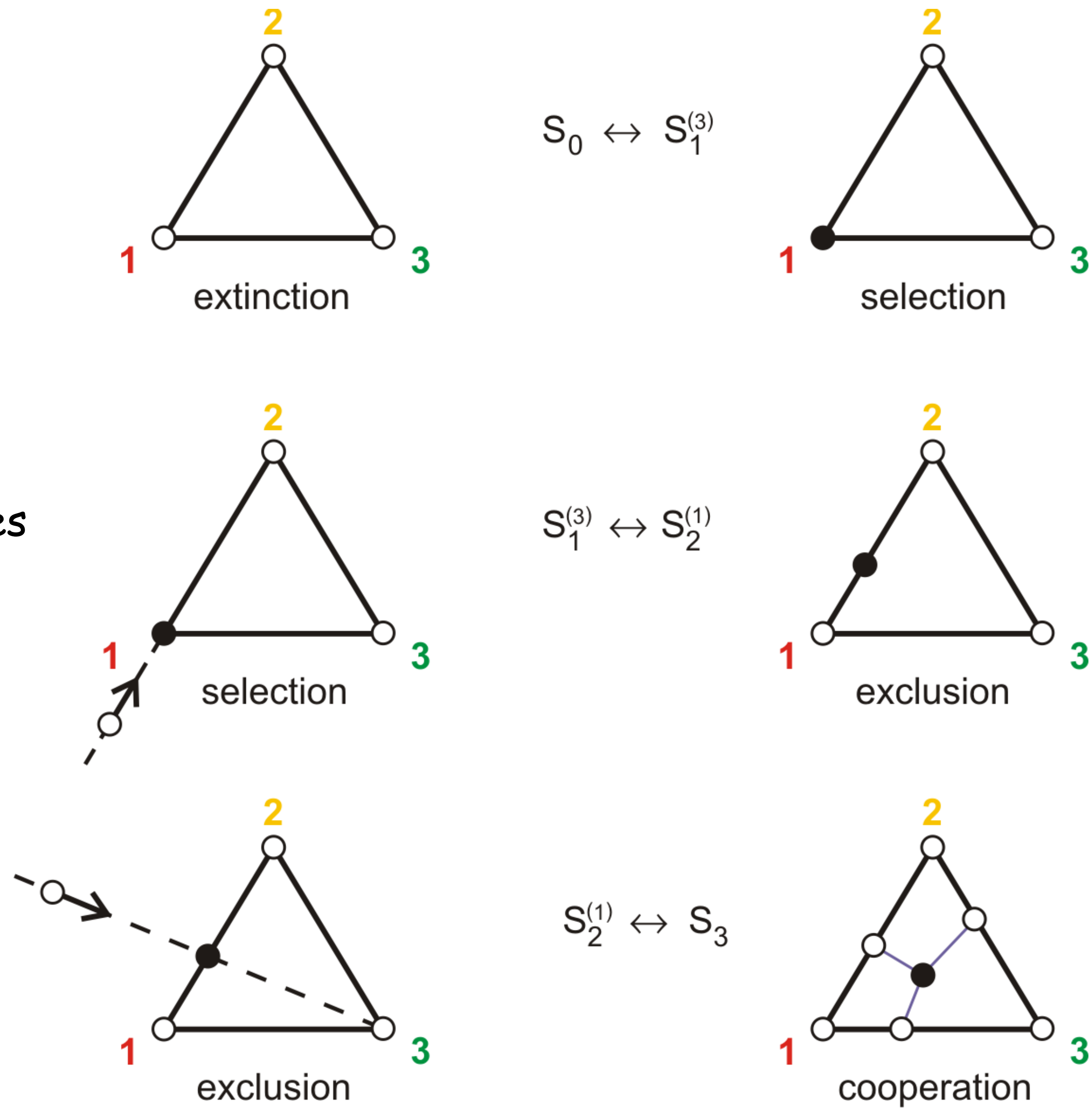
Mutation rate and genome size



evolution dependent on three basic parameters

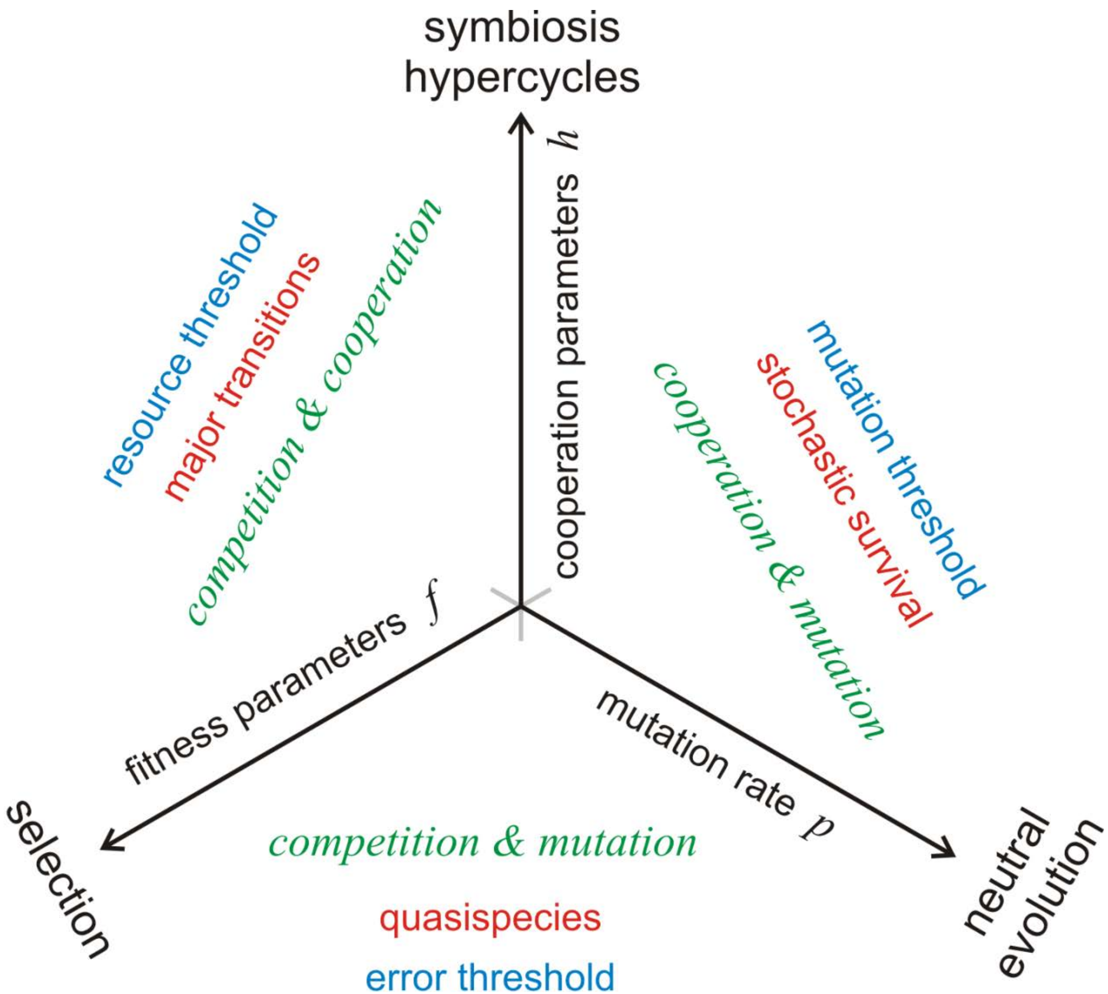
driving force:

availability of resources



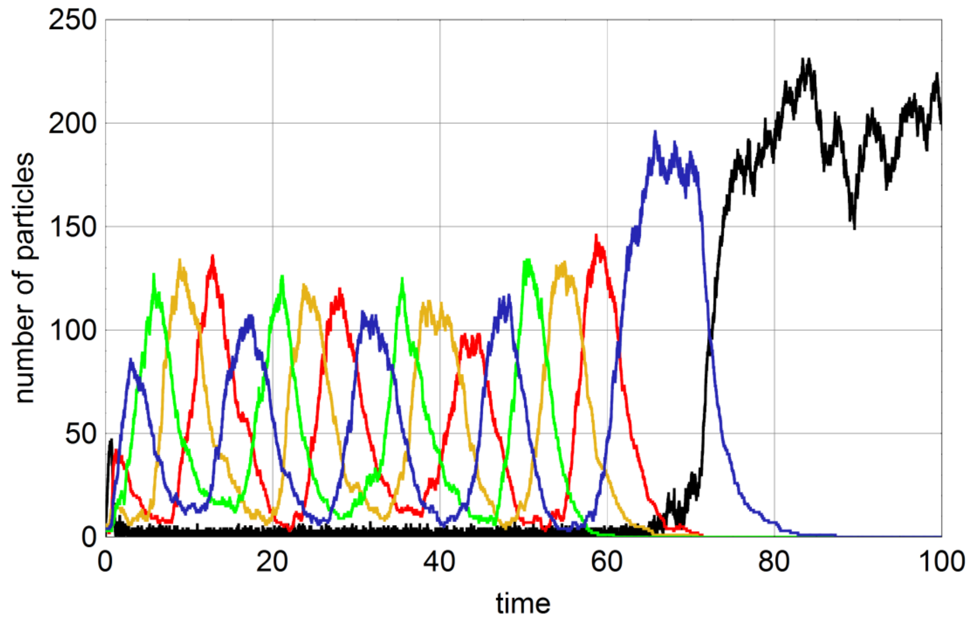
sequence of states

$$S_0 \leftrightarrow S_1^{(1)} \leftrightarrow S_2^{(3)} \leftrightarrow S_3$$

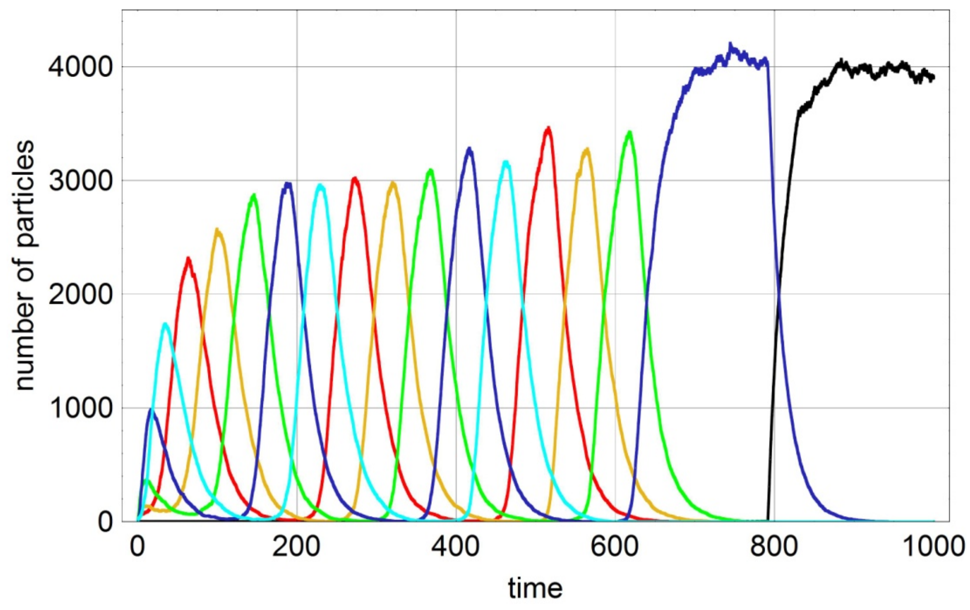


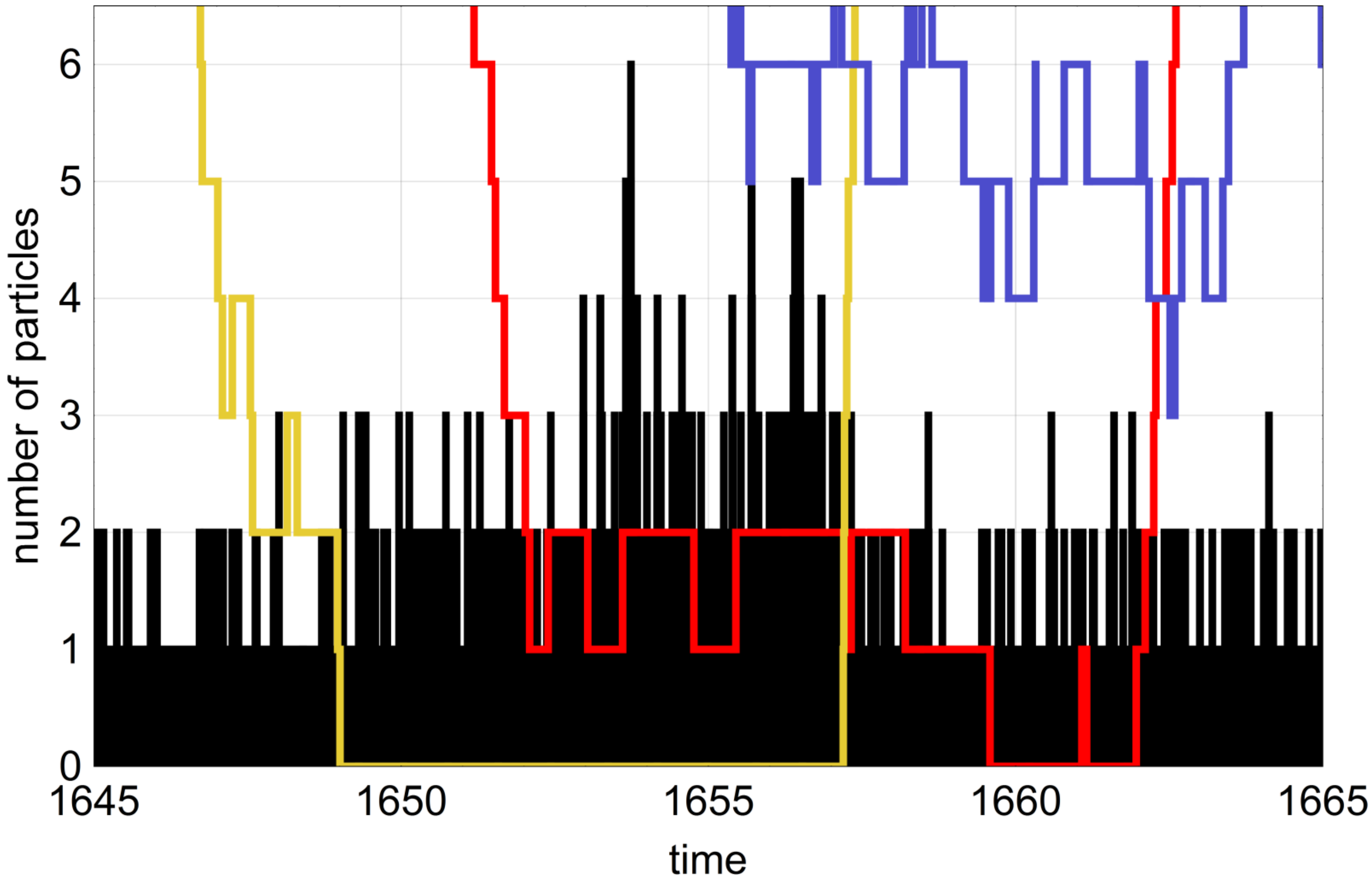
evolution dependent on three basic parameters

stochastic hypercycles with $n = 4$

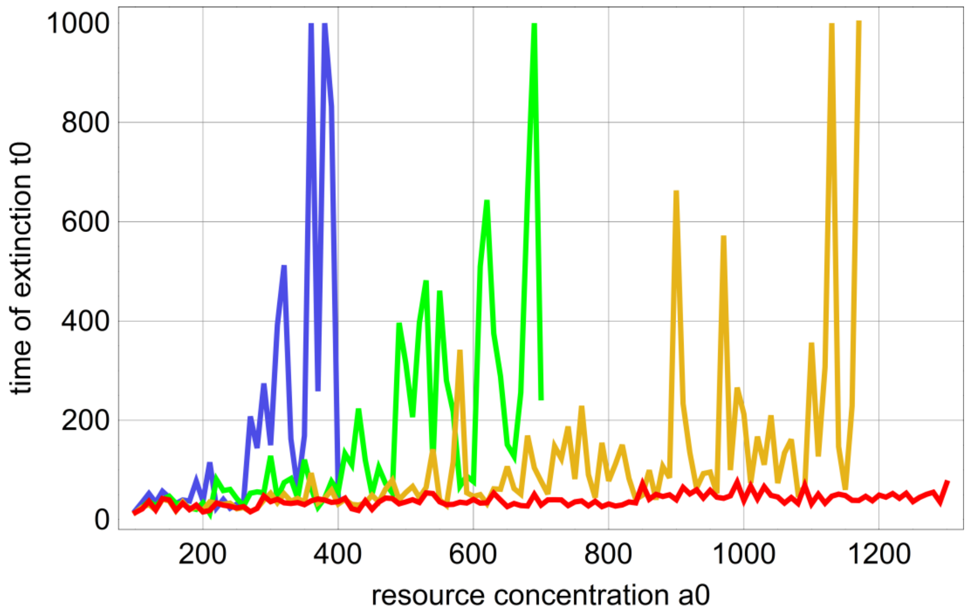


stochastic hypercycles with $n = 5$





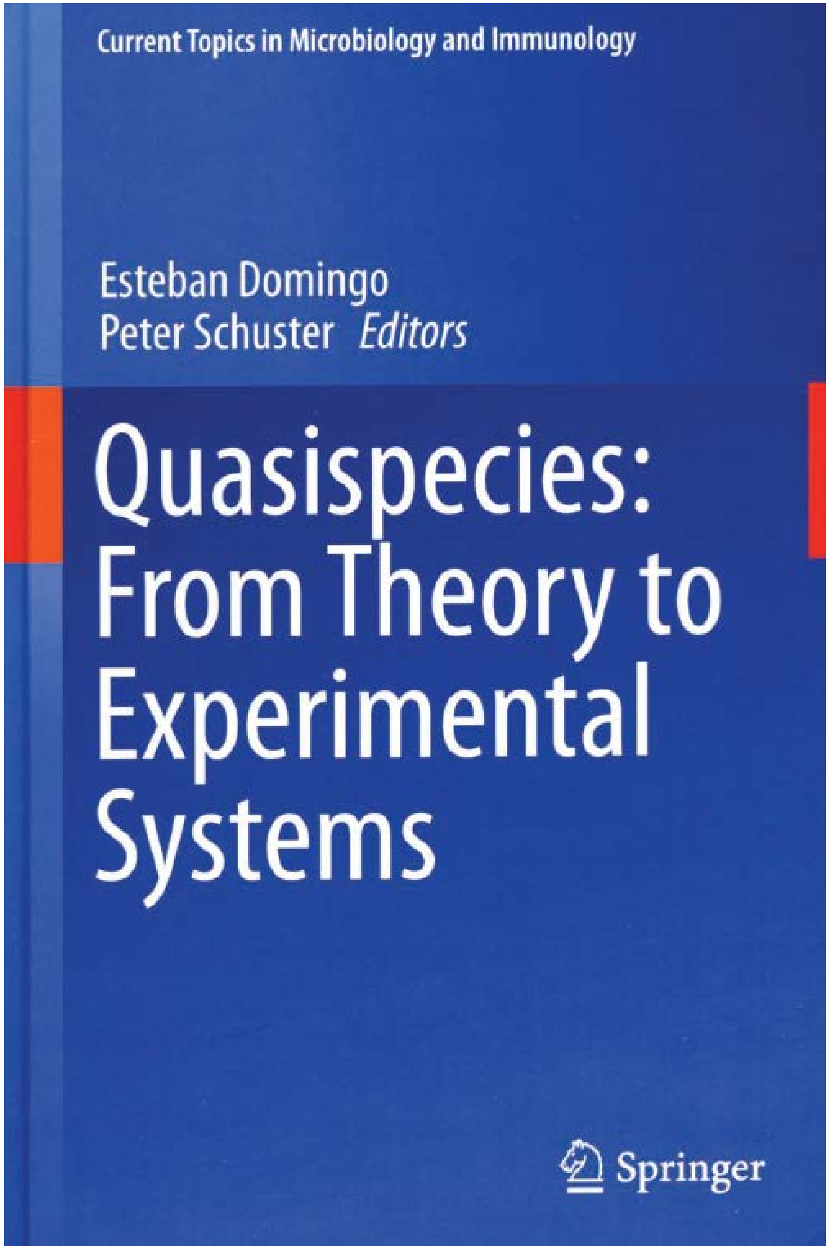
oscillatory hypercycles: simulation for n=4, enlargement

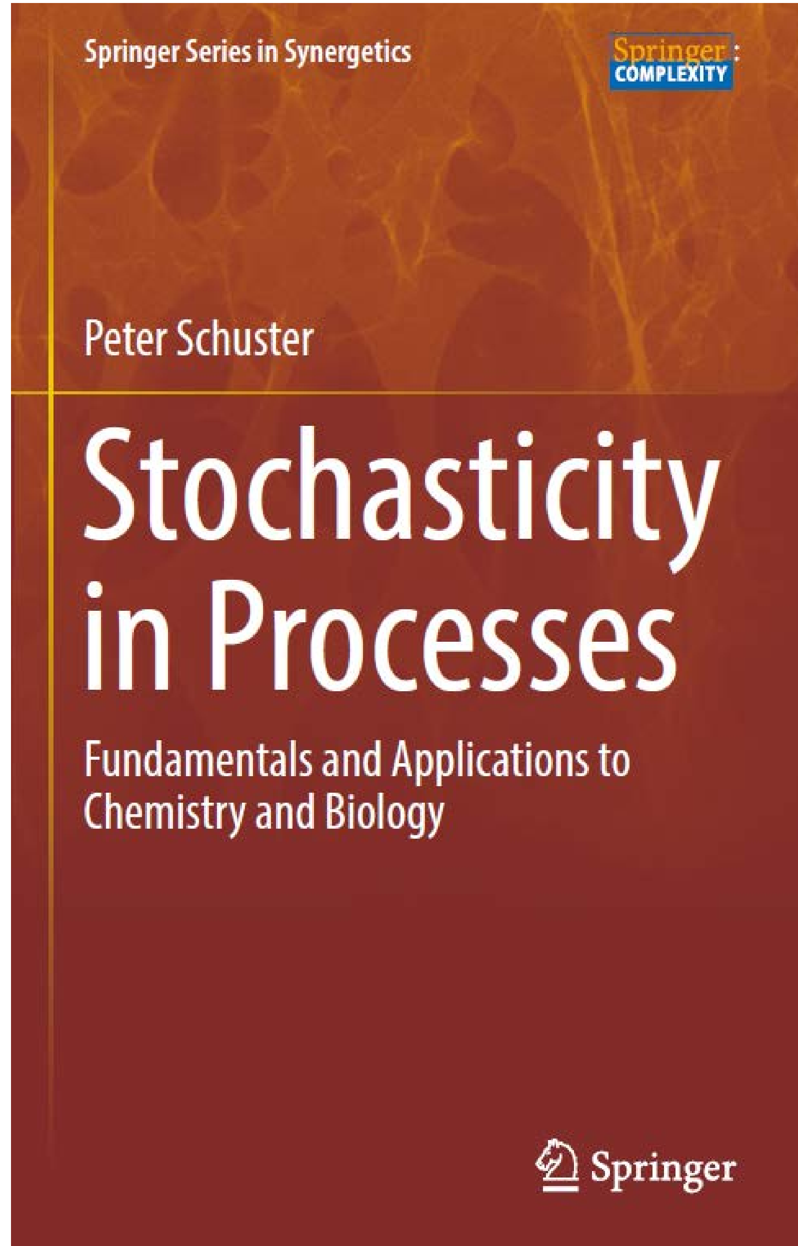


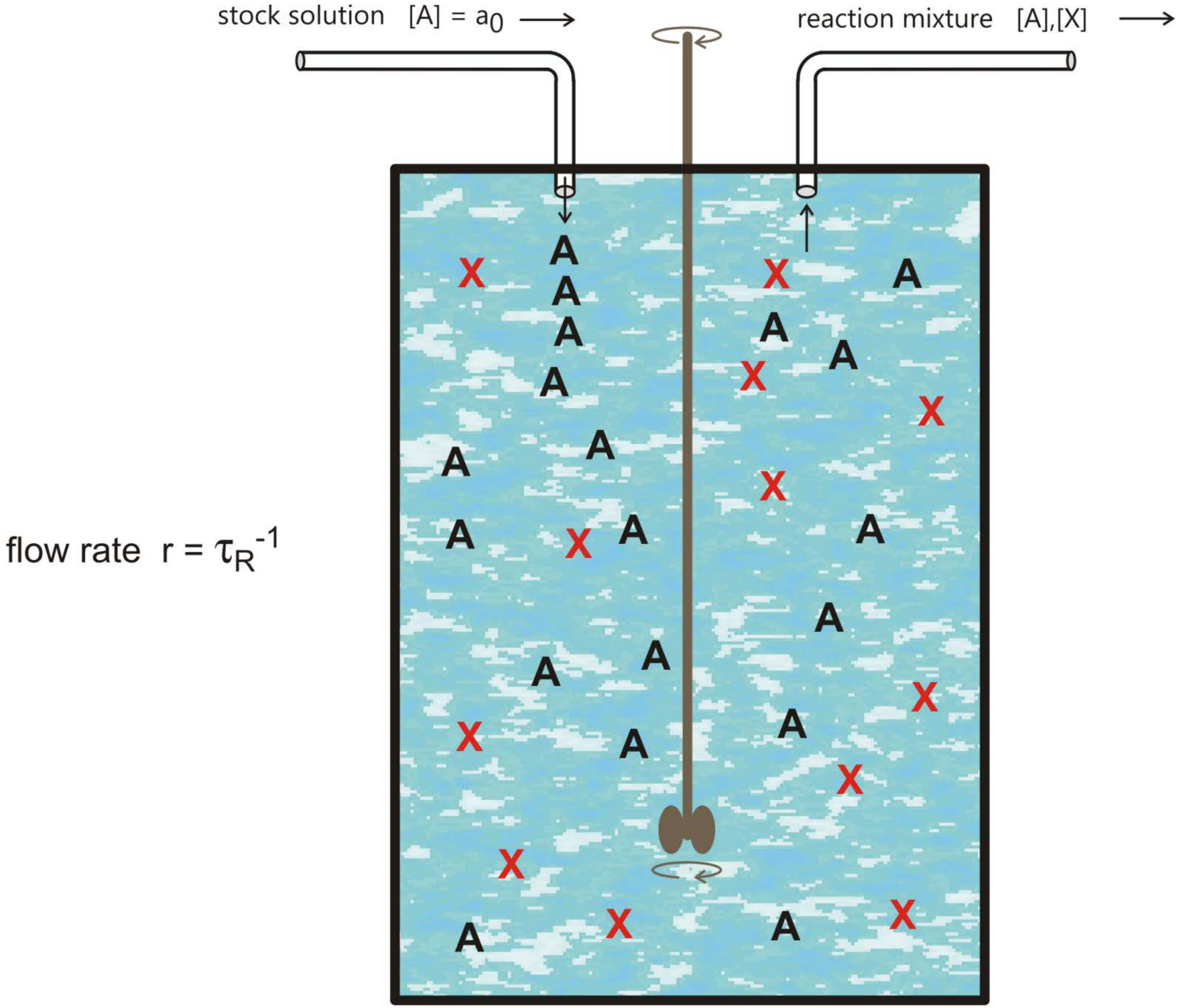
mutation rate: p = **0.0000**, p=**0.0005**, p = **0.0010** and p = **0.0020**

oscillatory hypercycles: simulation for n = 5, ‚pentagram‘

1. Mathematical concepts before Darwin
2. Theory of molecular evolution
3. **Evolution in realistically small populations**
4. Fitness landscapes and evolution
5. Evolution and present day molecular genetics


Current Topics in Microbiology and Immunology
Esteban Domingo
Peter Schuster Editors
Quasispecies: From Theory to Experimental Systems
Springer


Springer Series in Synergetics
Springer Complexity
Peter Schuster
Stochasticity in Processes
Fundamentals and Applications to Chemistry and Biology
Springer



the continuously fed stirred tank reactor (CFSTR)

$$
\begin{aligned}
* \quad &\xrightarrow{\;a_0\, r\;} \quad \mathbf{A} \ , \\
\mathbf{A} + \mathbf{X}_i \quad &\xrightarrow{\;f_i\;} \quad 2\,\mathbf{X}_i \ ; \quad i = 1, \ldots, n \ , \\
\mathbf{A} \quad &\xrightarrow{\;r\;} \quad \varnothing \ , \quad \text{and} \\
\mathbf{X}_i \quad &\xrightarrow{\;r\;} \quad \varnothing \ ; \ i = 1, \ldots, n \ .
\end{aligned}
$$

selection in the flow reactor

$\Delta f / f$	t_e	Population size N = 100				Population size N = 200			
		$A(t_e)$	$X_1(t_e)$	$X_2(t_e)$	$X_3(t_e)$	$A(t_e)$	$X_1(t_e)$	$X_2(t_e)$	$X_3(t_e)$
0.0	600	**1.5** ± 1.3	**30.5** ± 3.9	**34.2** ± 4.6	**33.4** ± 4.1	**0.5** ± 0.9	**30.6** ± 4.6	**30.9** ± 5.0	**32.0** ± 4.7
0.02	600	**1.8** ± 1.4	**41.8** ± 4.8	**32.9** ± 3.8	**23.4** ± 4.0	**0.6** ± 0.8	**50.4** ± 5.7	**27.7** ± 4.9	**17.3** ± 2.6
0.04	400	**2.4** ± 2.1	**45.4** ± 5.0	**31.3** ± 4.5	**19.9** ± 2.5	**0.7** ± 0.8	**58.3** ± 4.6	**25.6** ± 4.5	**11.0** ± 2.9
0.1	400	**2.1** ± 1.7	**59.8** ± 5.5	**28.0** ± 4.1	**10.0** ± 2.9	**0.4** ± 0.5	**73.9** ± 4.1	**20.6** ± 3.5	**4.8** ± 1.9
0.2	400	**1.9** ± 1.1	**68.3** ± 4.5	**23.1** ± 3.7	**6.7** ± 2.8	**0.5** ± 0.7	**76.6** ± 4.1	**19.3** ± 2.8	**3.6** ± 1.7
0.4	400	**2.3** ± 1.8	**71.7** ± 6.0	**20.8** ± 5.2	**5.2** ± 2.4	**0.9** ± 0.6	**82.0** ± 4.2	**13.8** ± 3.8	**3.3** ± 1.7
1.0	200	**2.7** ± 2.4	**78.4** ± 4.7	**15.8** ± 3.3	**3.1** ± 1.5	**0.9** ± 0.9	**83.6** ± 4.0	**12.6** ± 3.2	**2.9** ± 1.5
1.8	200	**4.3** ± 1.1	**80.8** ± 2.9	**13.6** ± 3.1	**1.3** ± 1.2	**1.5** ± 1.3	**83.8** ± 3.3	**12.7** ± 2.5	**2.0** ± 1.7

n = 3: $\mathbf{X}_1$, $f_1 = f + \Delta f / 2f$; $\mathbf{X}_2$, $f_2 = f$; $\mathbf{X}_3$, $f_3 = f - \Delta f / 2f$; $f = 0.1$

initial particle numbers: $X_1(0) = X_2(0) = X_3(0) = 1$

probability of selection

1. Mathematical concepts before Darwin
2. Theory of molecular evolution
3. Evolution in realistically small populations
4. **Fitness landscapes and evolution**
5. Evolution and present day molecular genetics

Sewall Wright, 1889 - 1988

+ …….. wild type
a .... alternative allele
on locus A
:
:
:
abcde … alternative alleles
on all five loci

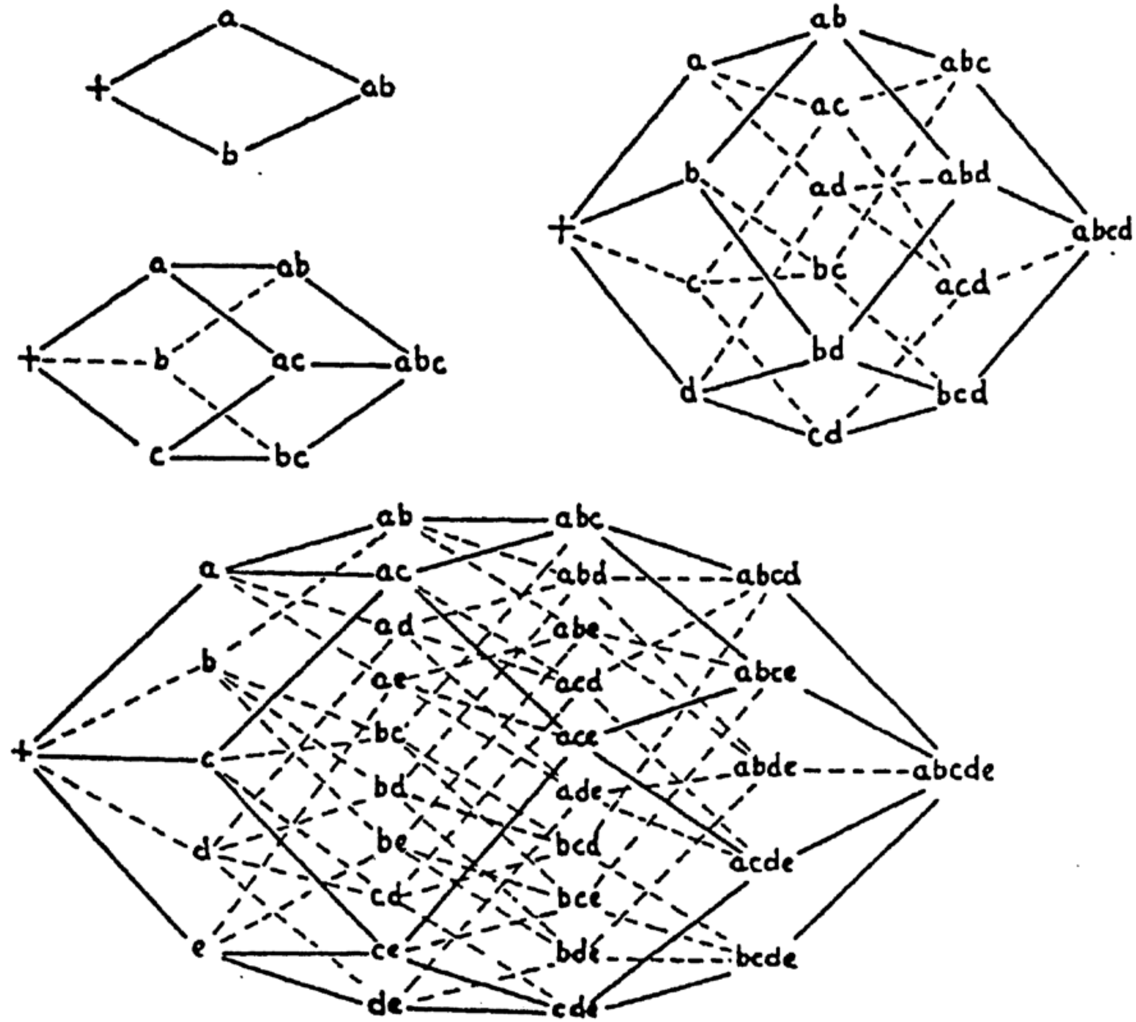


FIG. 1.—The combinations of from 2 to 5 paired allelomorphs.

the multiplicity of gene replacements with two alleles on each locus

Sewall Wright. 1988. Surfaces of selective value revisited. American Naturalist 131:115-123

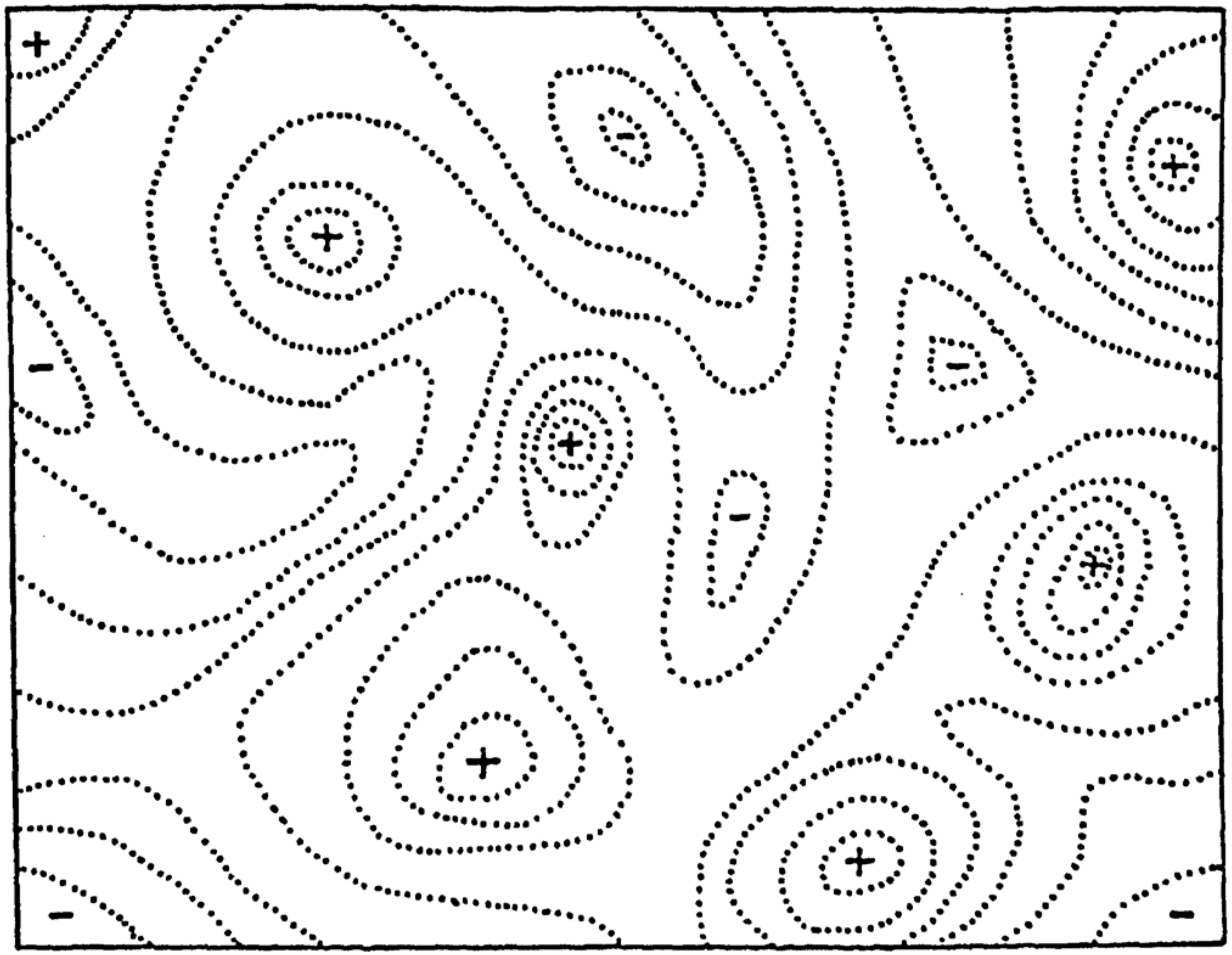
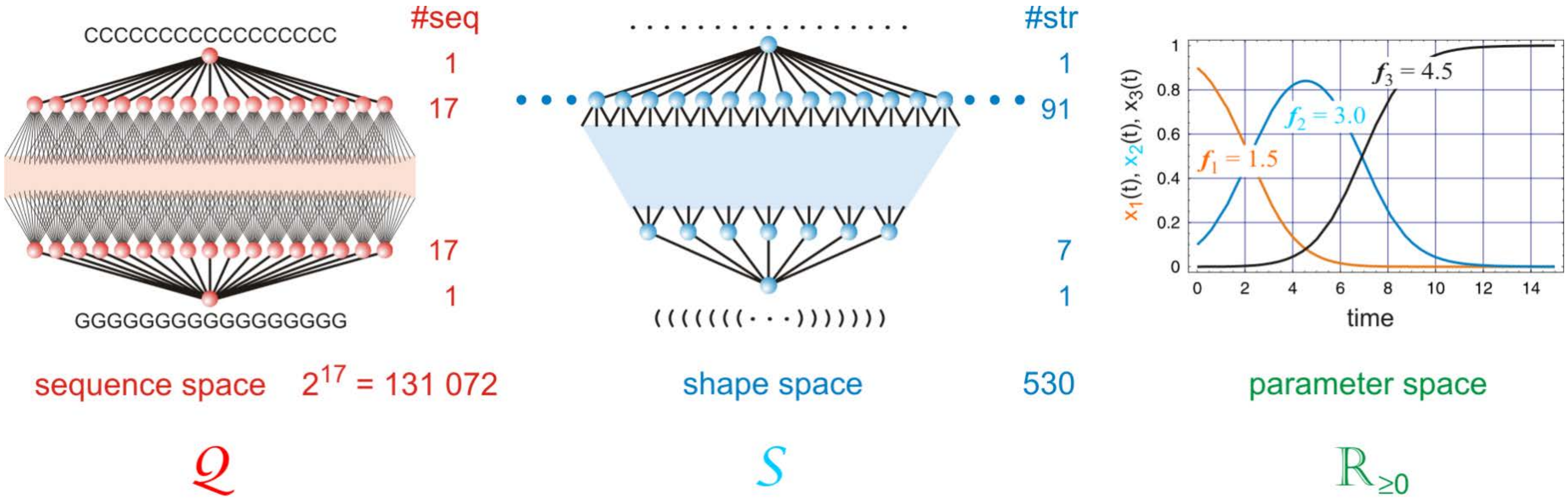


FIG. 2.—Diagrammatic representation of the field of gene combinations in two dimensions instead of many thousands. Dotted lines represent contours with respect to adaptiveness.

Evolution is hill climbing of populations or subpopulations

Sewall Wright. 1988. Surfaces of selective value revisited. American Naturalist 131:115-123

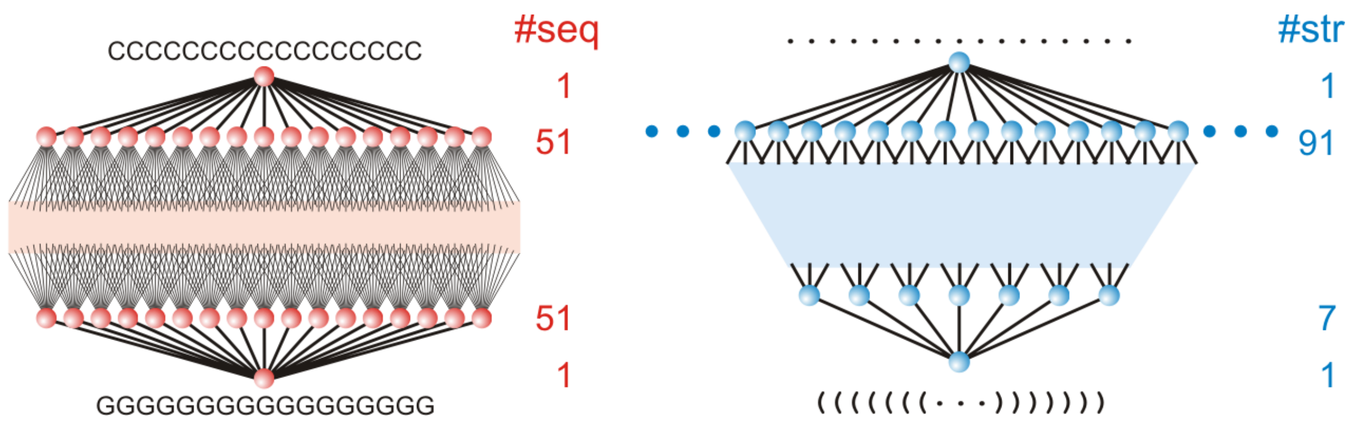


X $\xrightarrow{\Phi: (\mathcal{Q}, d_H) \Rightarrow (\mathcal{S}, d_S)}$ $\mathrm{S} = \Phi(\mathrm{X})$ $\xrightarrow{\Psi: (\mathcal{S}, d_S) \Rightarrow \mathbb{R}_{\geq 0}}$ $f = \Psi(\mathrm{S})$

sequence → structure → function

genotype → phenotype → selection

fitness of RNA secondary structures through evaluation of phenotypes



sequence space $4^{17} = 17\ 179\ 869\ 184$ shape space 530

$\mathcal{Q}$ $\mathcal{S}$

X $\Phi: (\mathcal{Q}, d_H) \Rightarrow (\mathcal{S}, d_S)$ S = Φ(X)

sequence ⟶ structure

genotype ⟶ phenotype

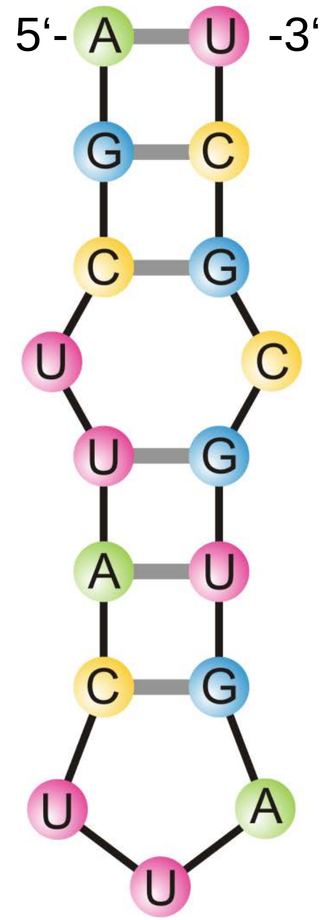
formation of RNA secondary structures as genotype-phenotype mapping

RNA sequence – structure mappings

1. ruggedness and neutrality
2. existence of extended neutral networks
3. shape space covering

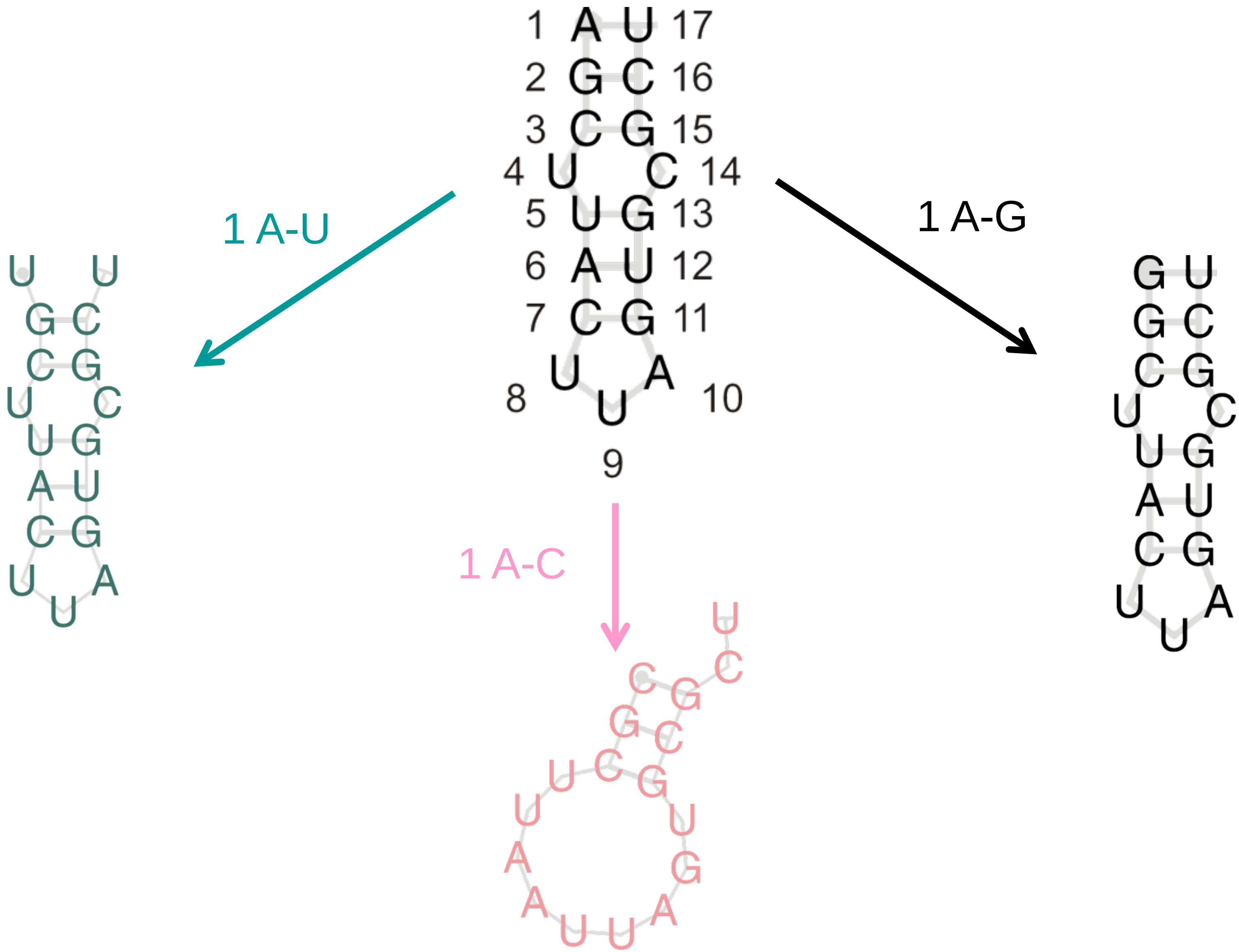
5‘- AGCUUACUUAGUGCGCU-3‘

5‘-(((· (((· · ·))) ·))) -3‘



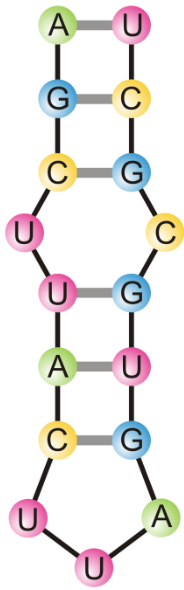
the minimum free energy structure of a small RNA molecule

AGCUUAACUUAGUCGCU

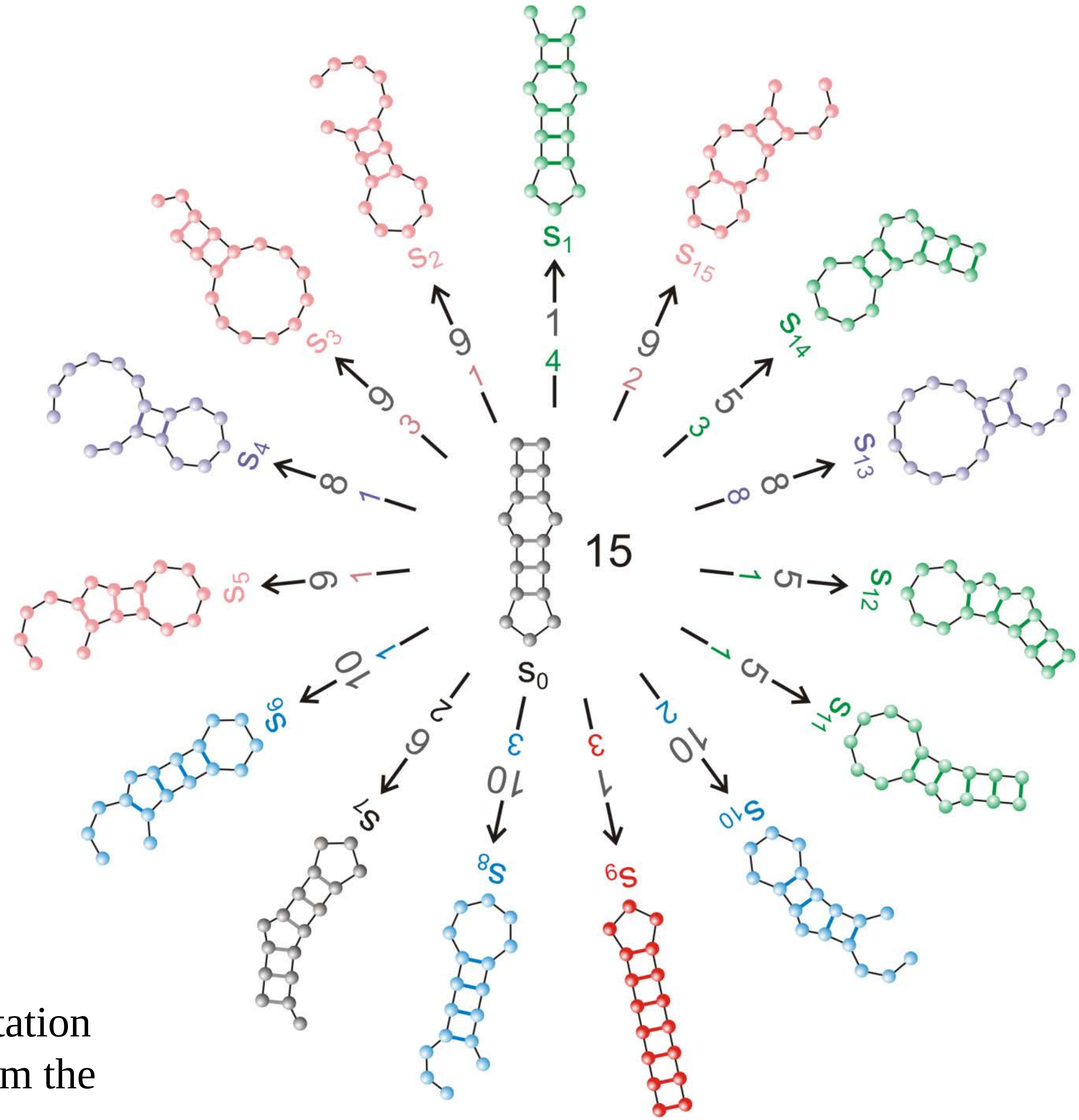


AGCUUACUUAGUGCGCU

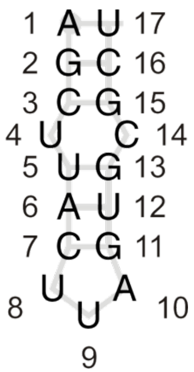
(((• (((• • •))) •)))



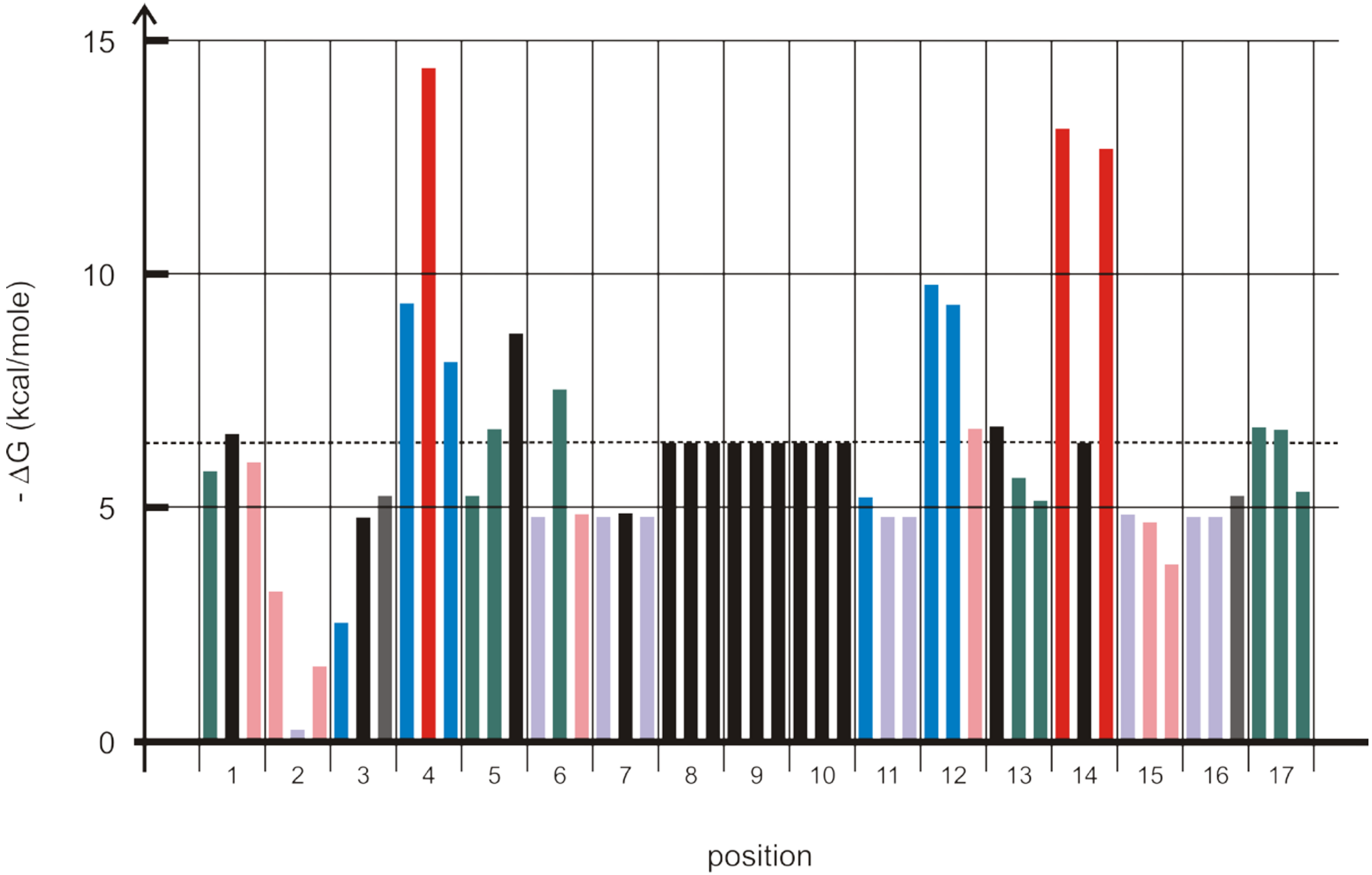
reference

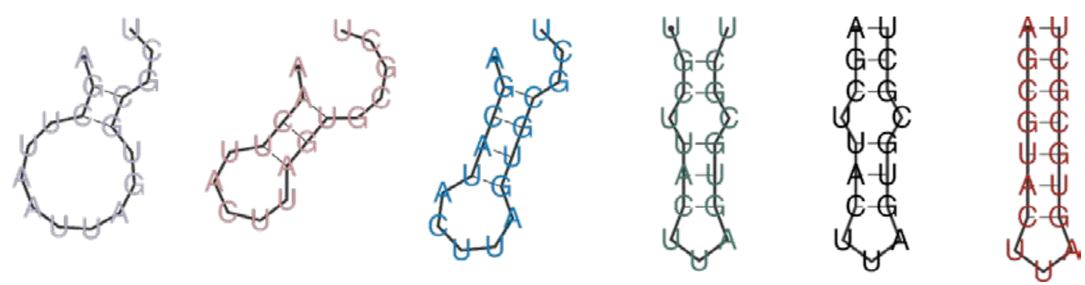


frequencies of 51 point mutation structures and distances from the reference structure

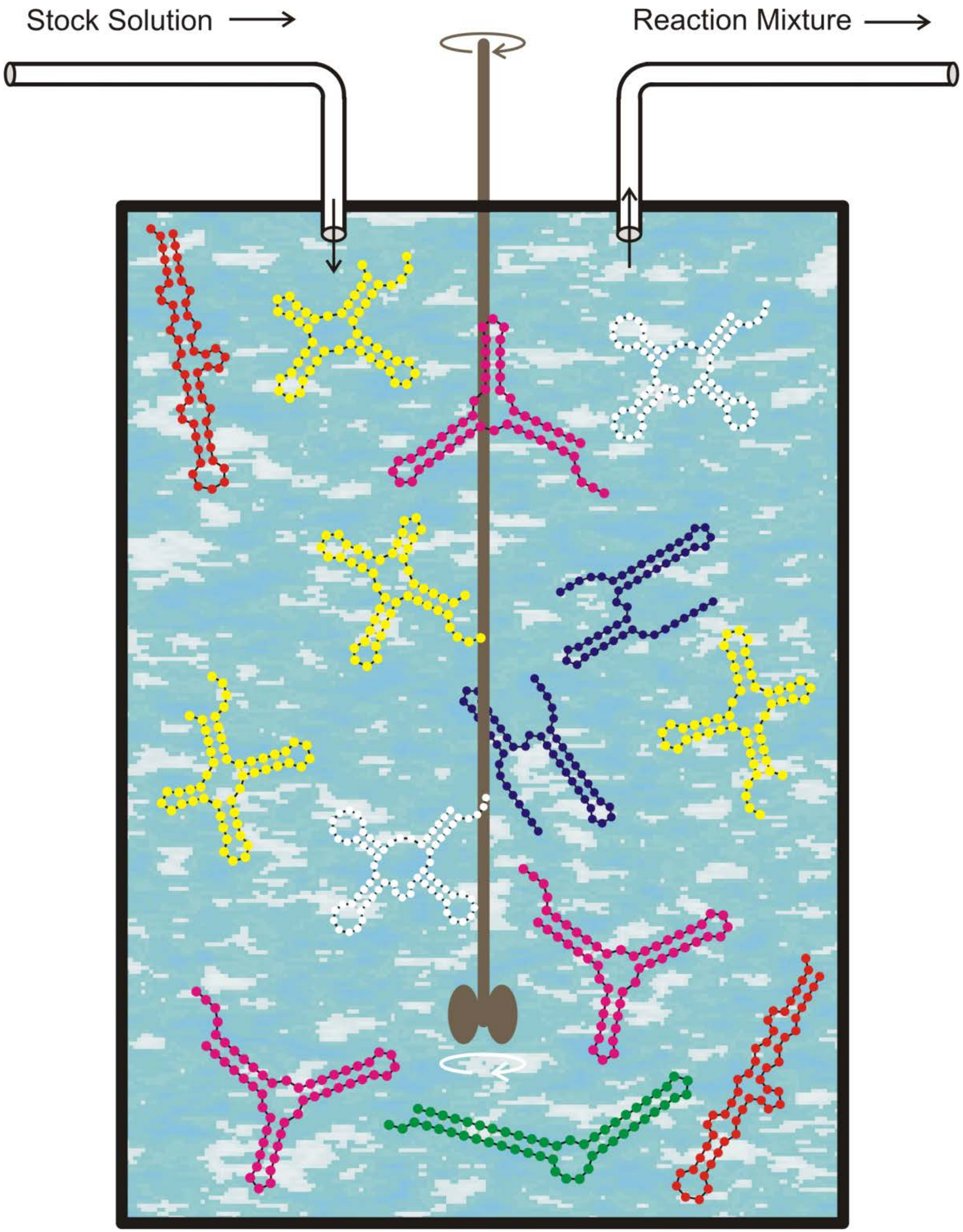


reference





free energy of formation
(ΔG_0) of 51 point mutants
Of the reference sequence



replication rate constant

(fitness):

$$f_k = \gamma \,/\, [\alpha + \Delta d_S^{\,(k)}]$$

$$\Delta d_S^{\,(k)} = d_H(S_k, S_\tau)$$

selection pressure:

The population size,

N = # RNA molecules,

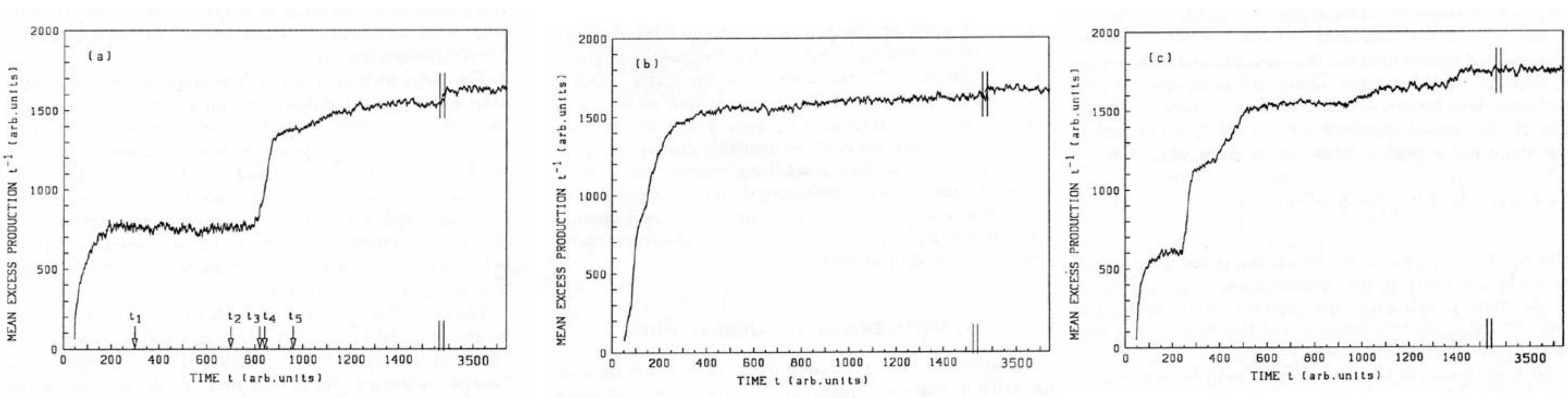
is determined by the flow:

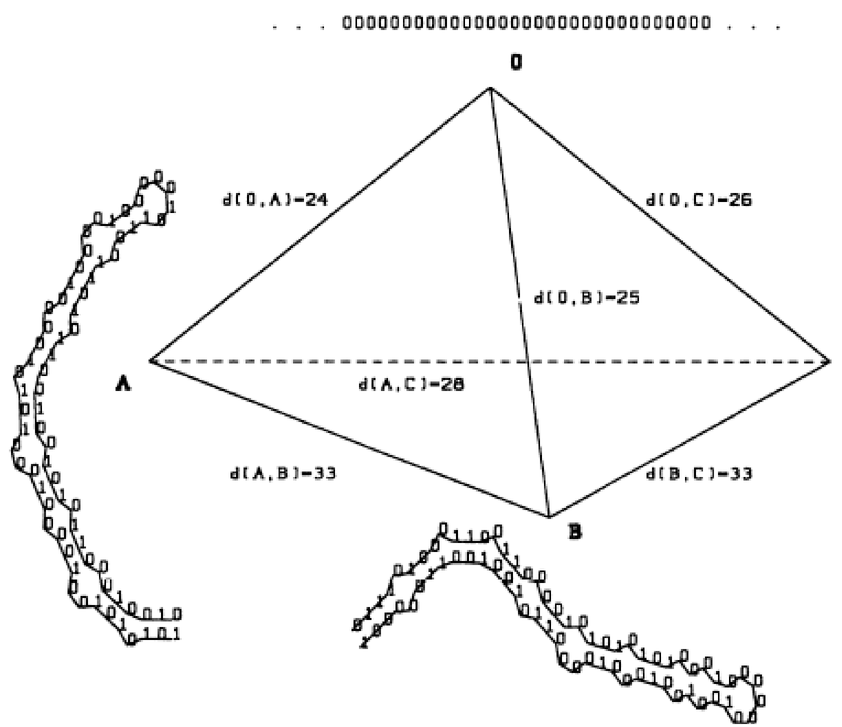
$$N(t) \approx \overline{N} \pm \sqrt{\overline{N}}$$

mutation rate:

p = 0.001 / nucleotide × replication

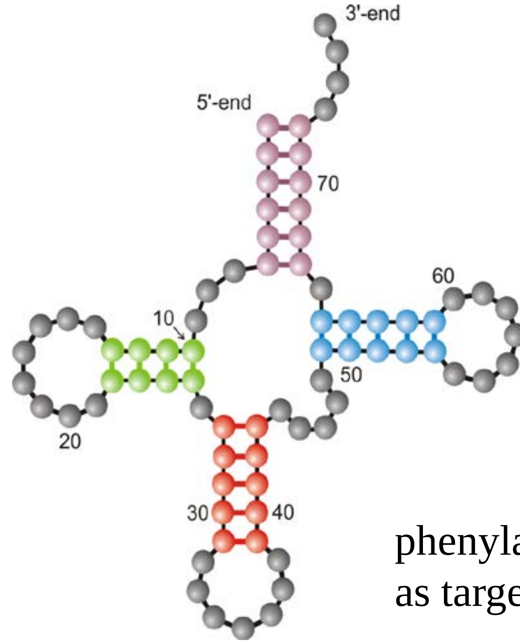
the flow reactor as a device for
studying the evolution of molecules
in vitro and *in silico*.





Walter Fontana, Wolfgang Schnabl, and
Peter Schuster, Phys.Rev.A 40:3301-3321, 1989

structure of
randomly chosen
initial sequence



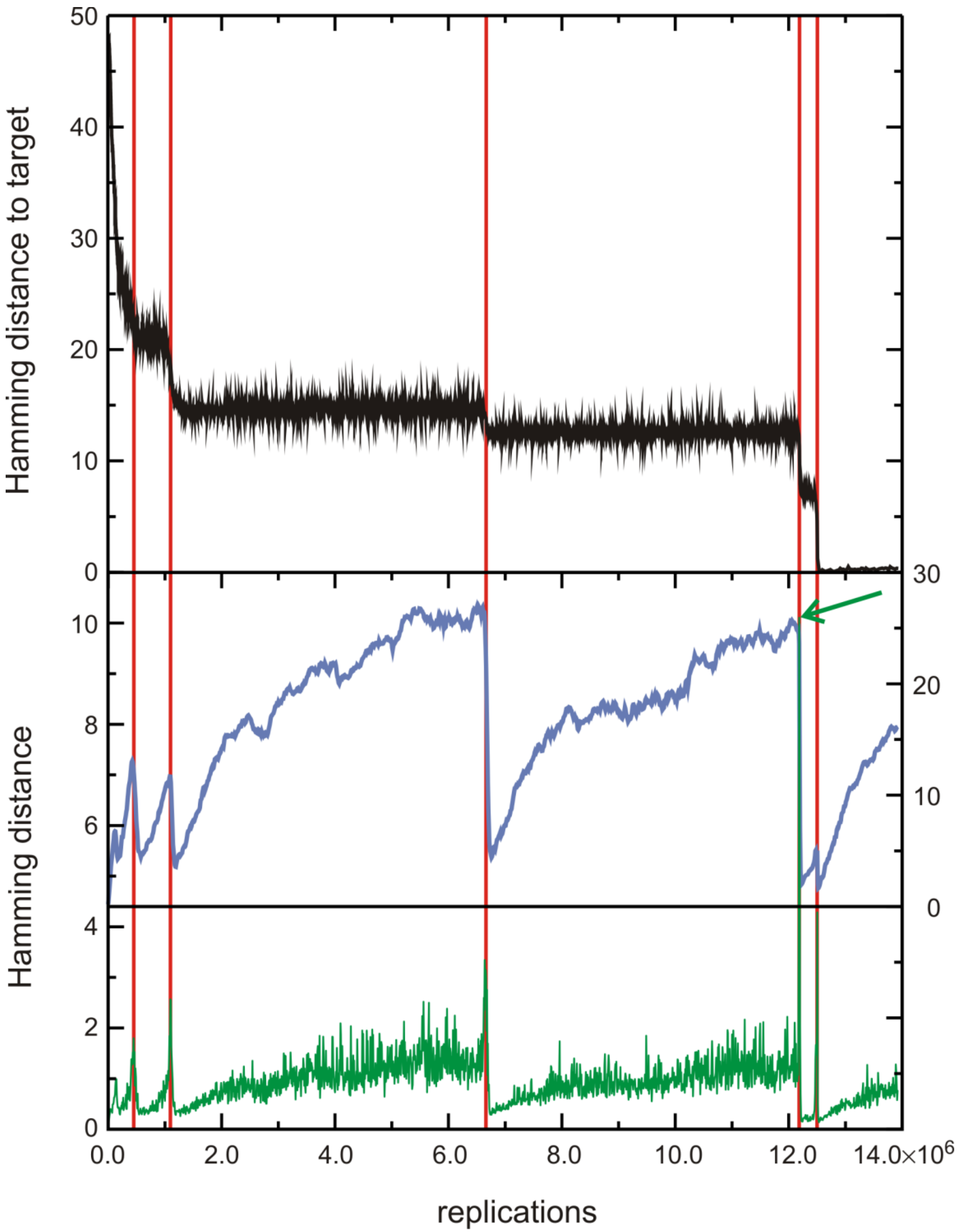
phenylalanyl-tRNA
as target structure

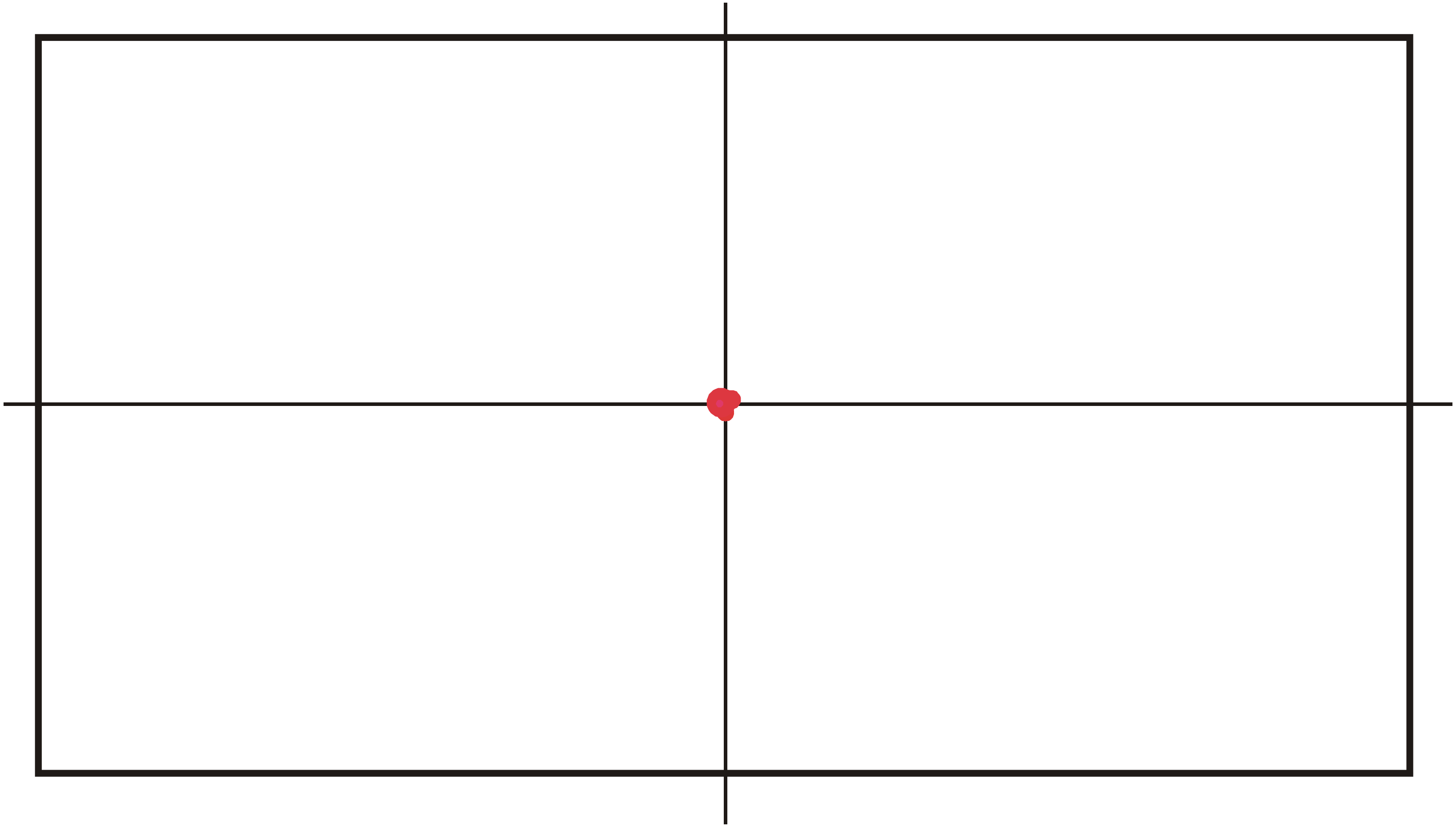
evolution *in silico*.

W. Fontana, P. Schuster, *Science* **280** (1998), 1451-1455 evolutionary trajectory

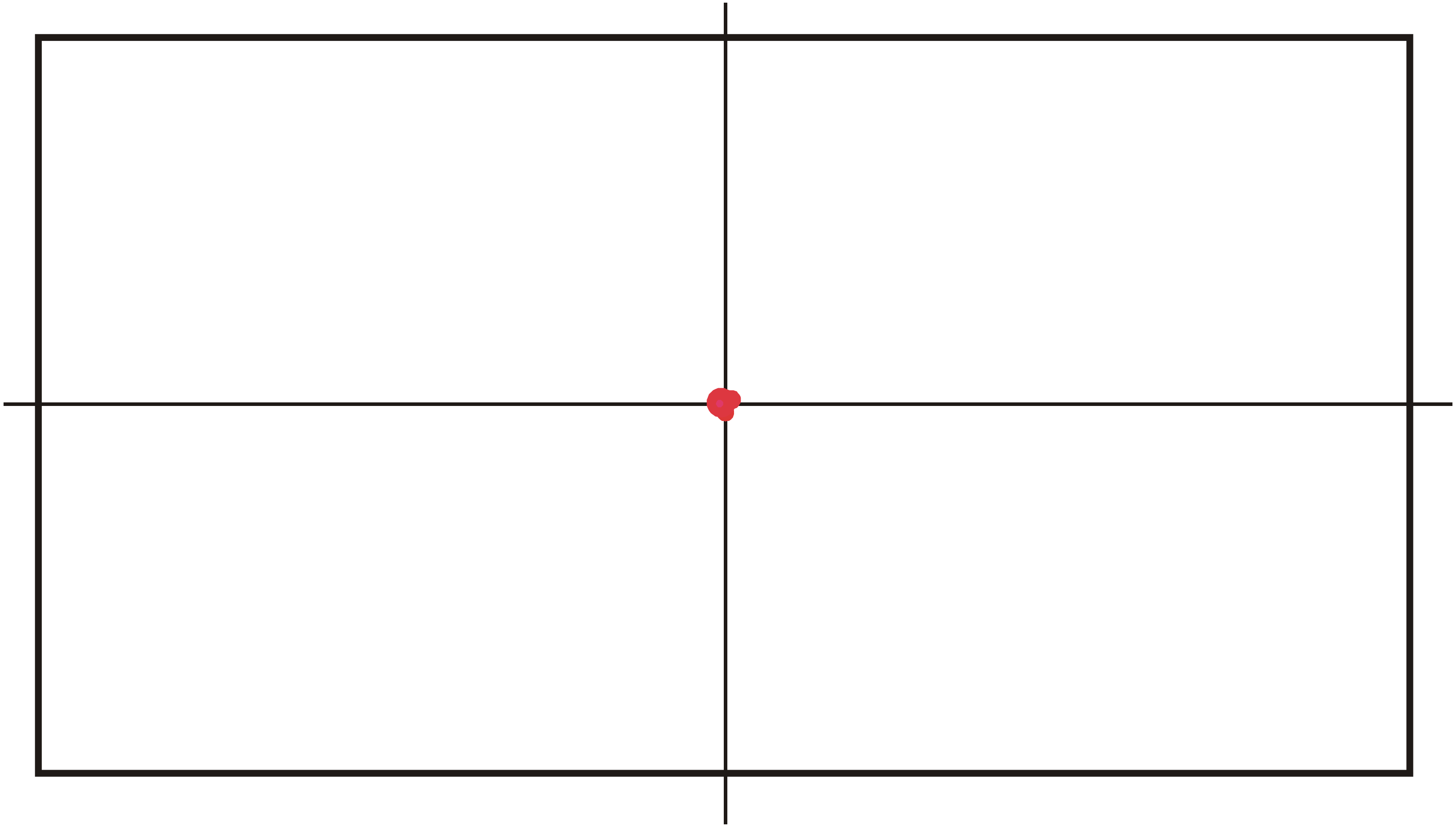
spreading of the population on neutral networks

drift of the population center in sequence space

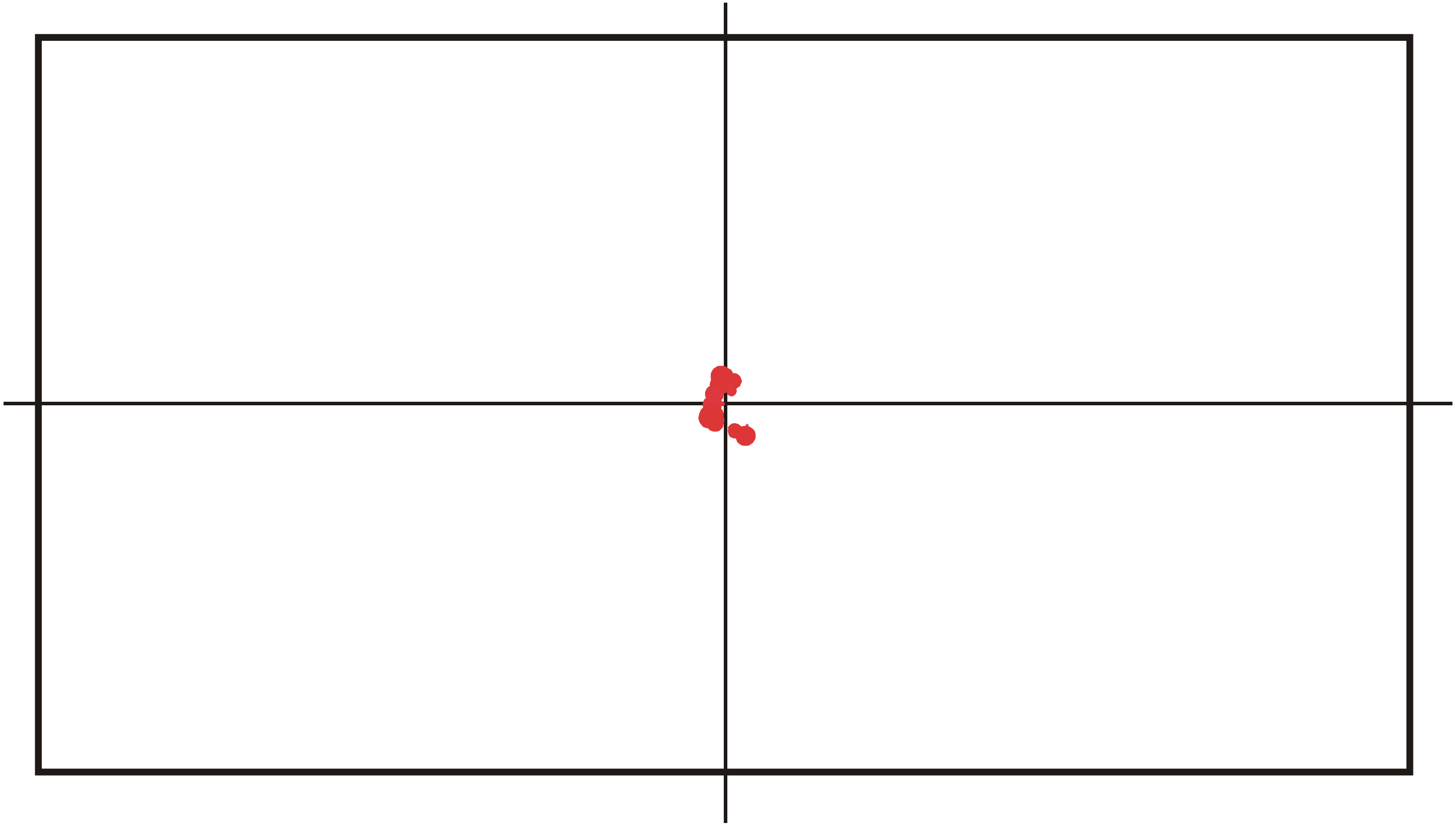




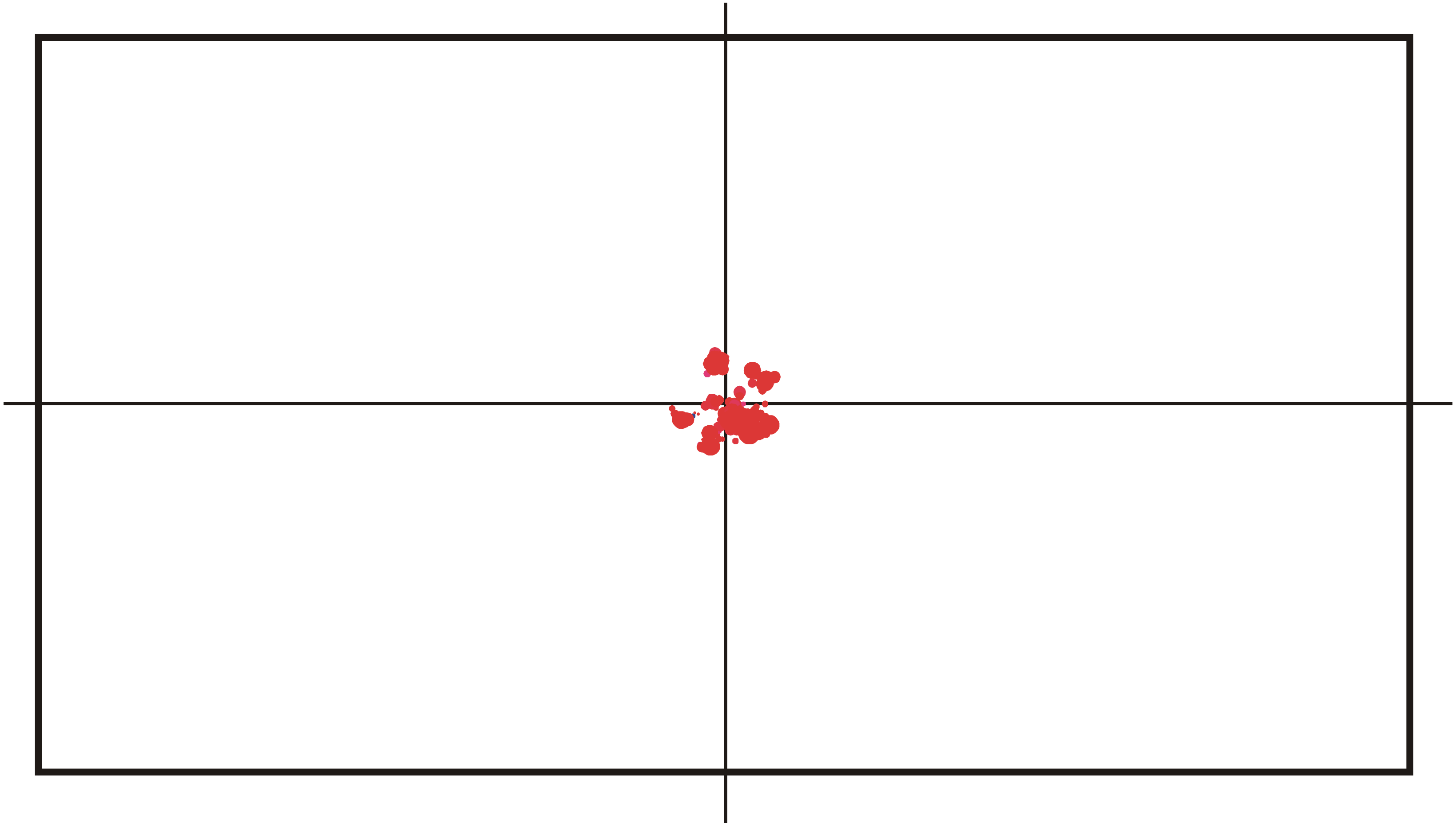
spreading and evolution of a population on a neutral network: t = 150



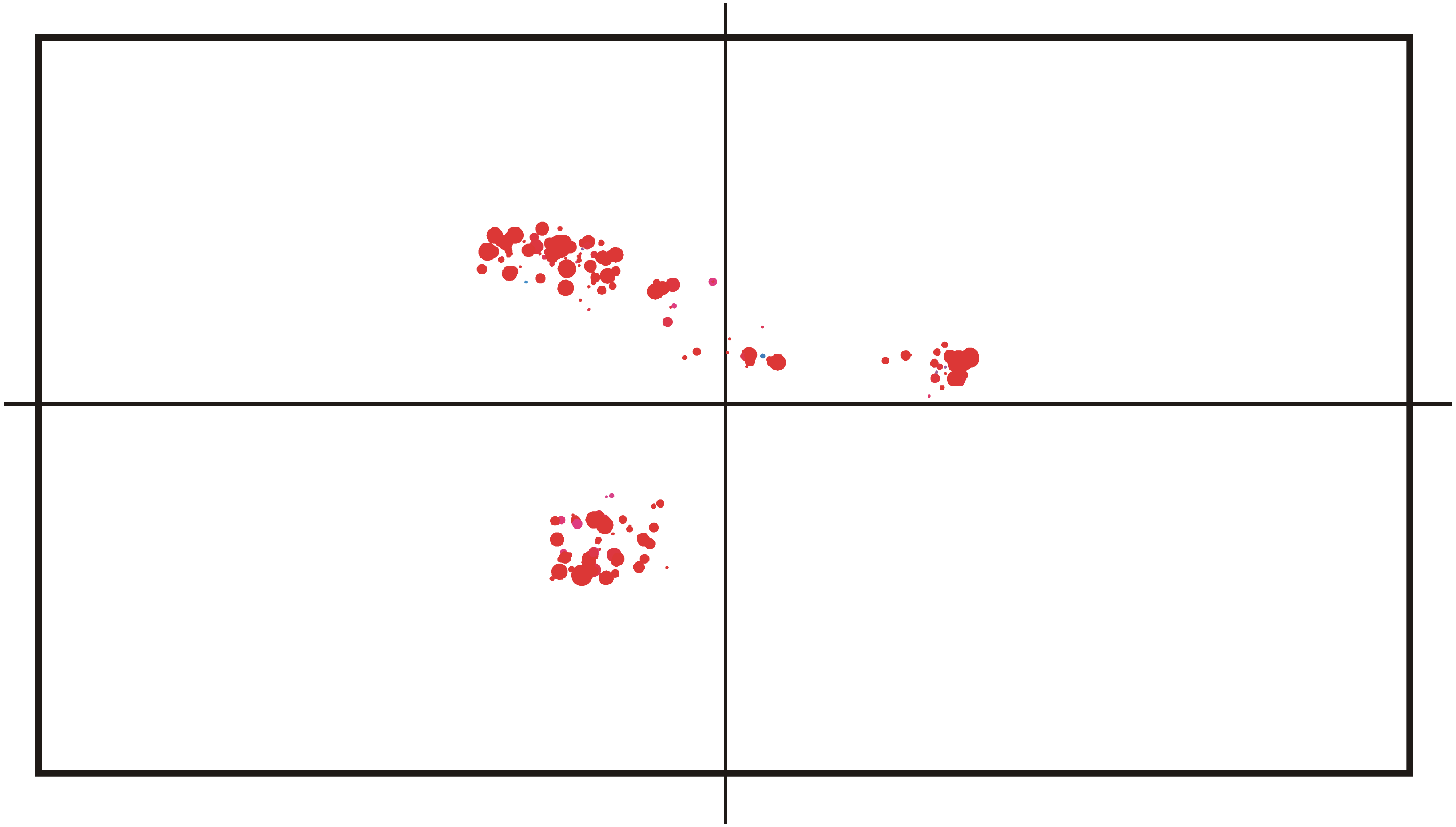
Spreading and evolution of a population on a neutral network: t = 150



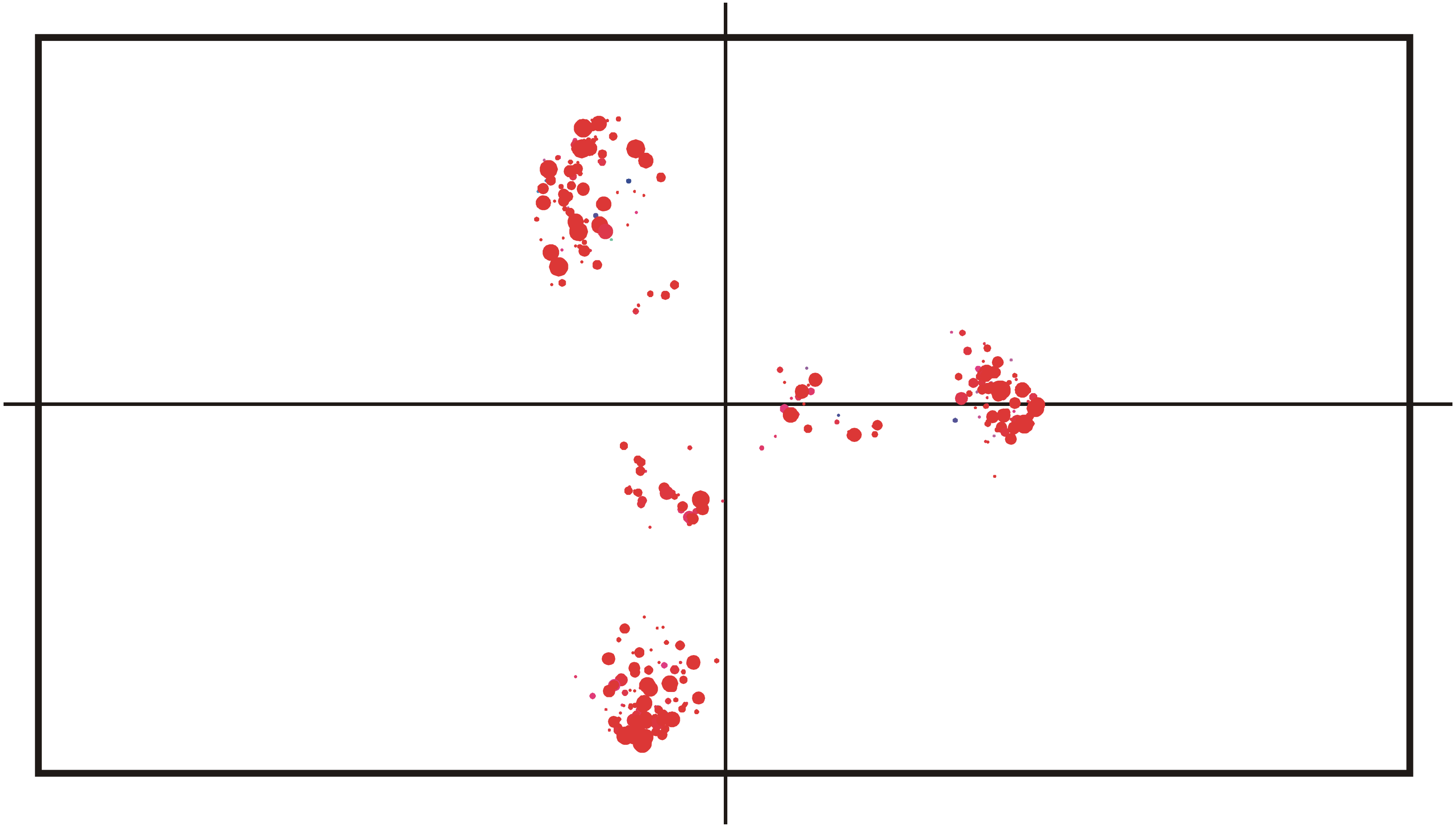
Spreading and evolution of a population on a neutral network: t = 170



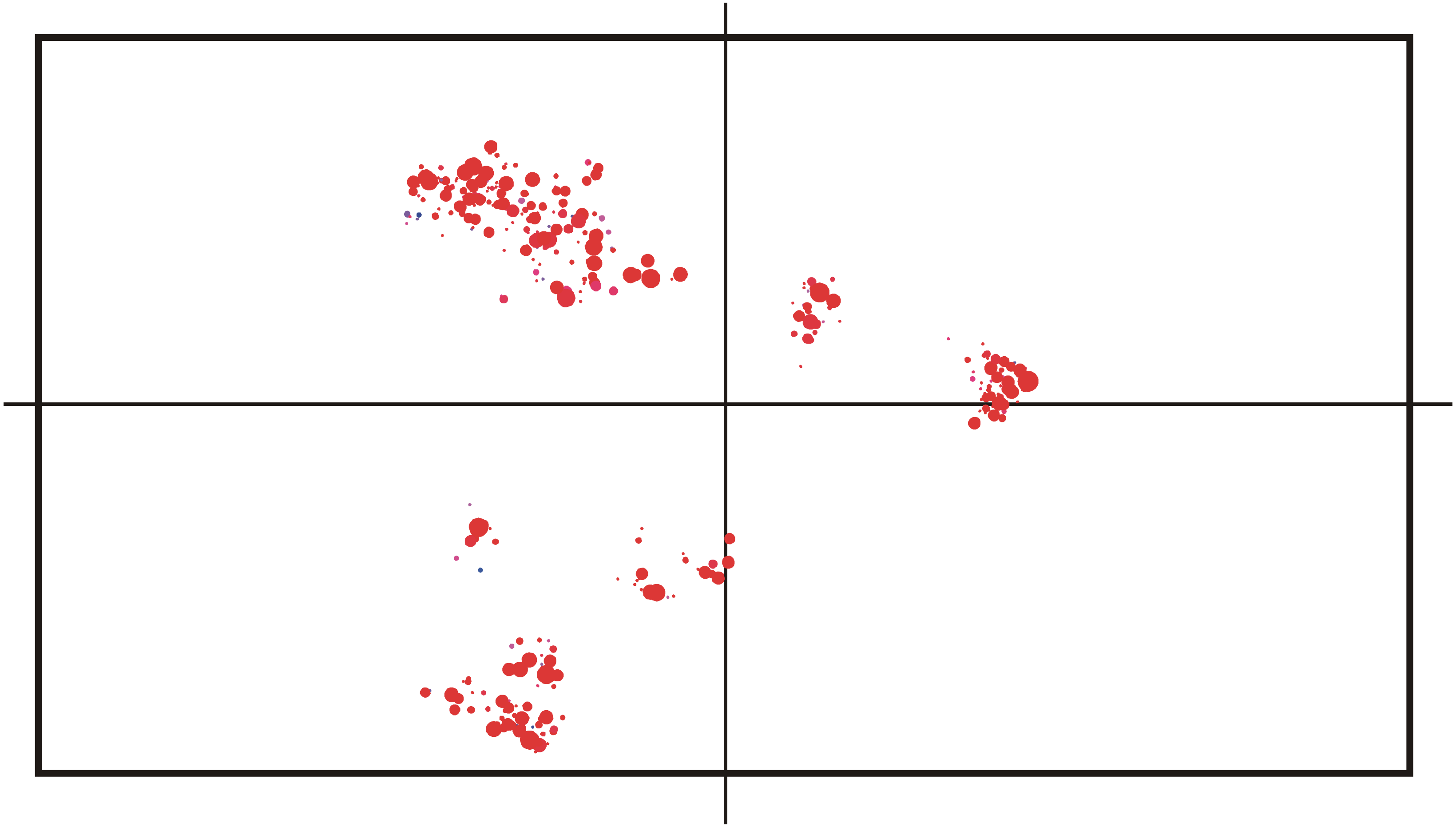
Spreading and evolution of a population on a neutral network: t = 200



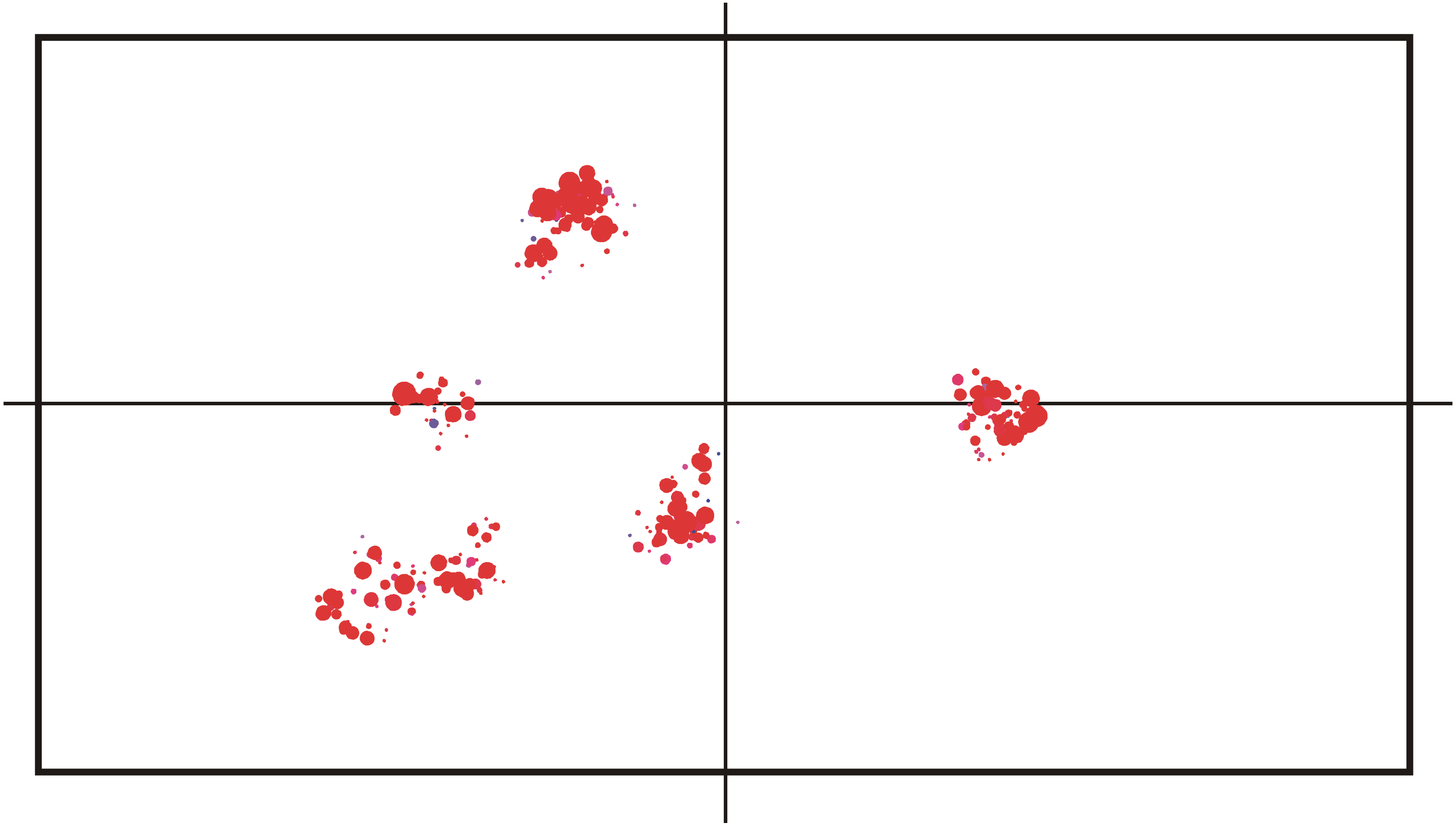
Spreading and evolution of a population on a neutral network: t = 350



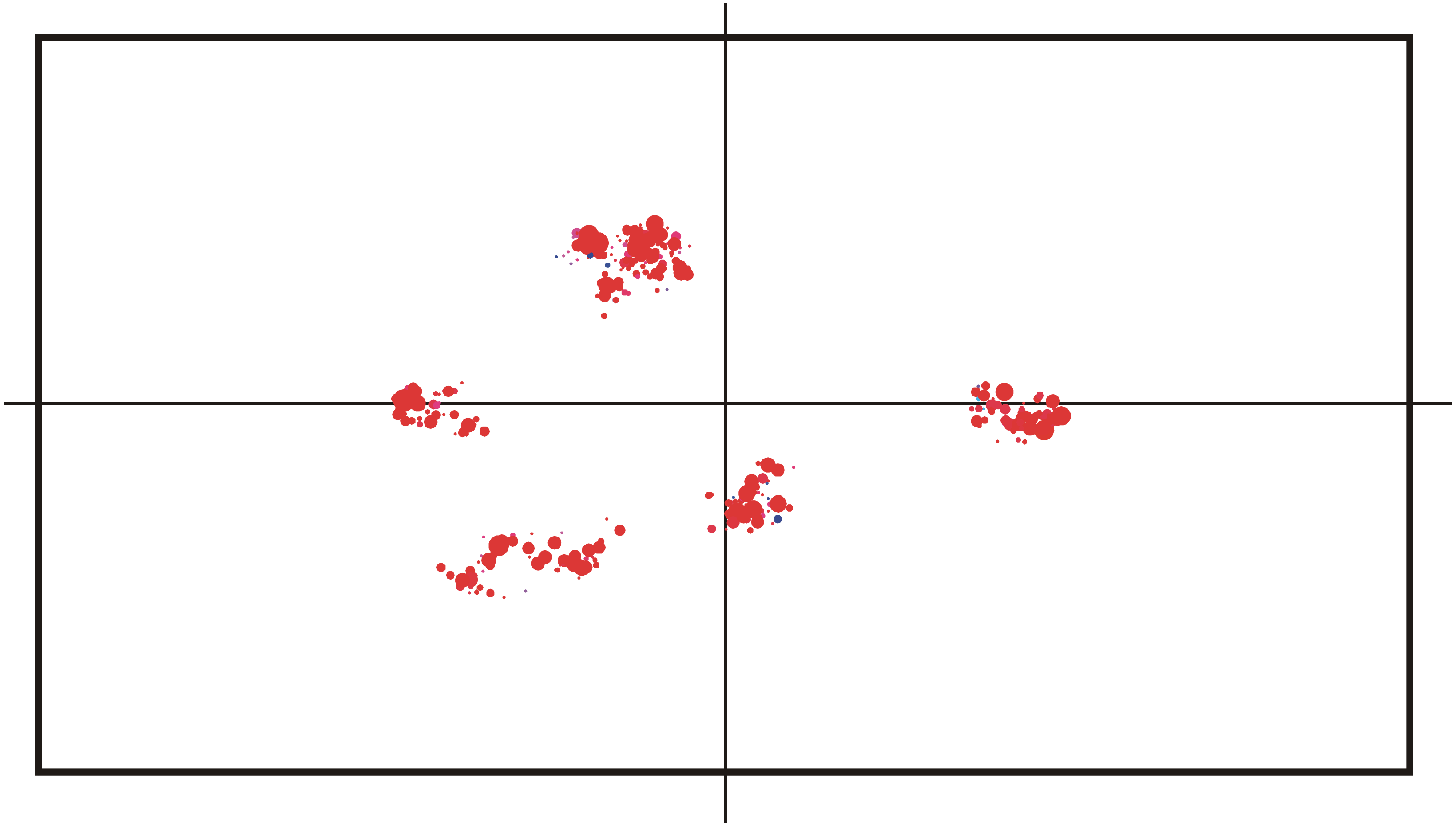
Spreading and evolution of a population on a neutral network: t = 500



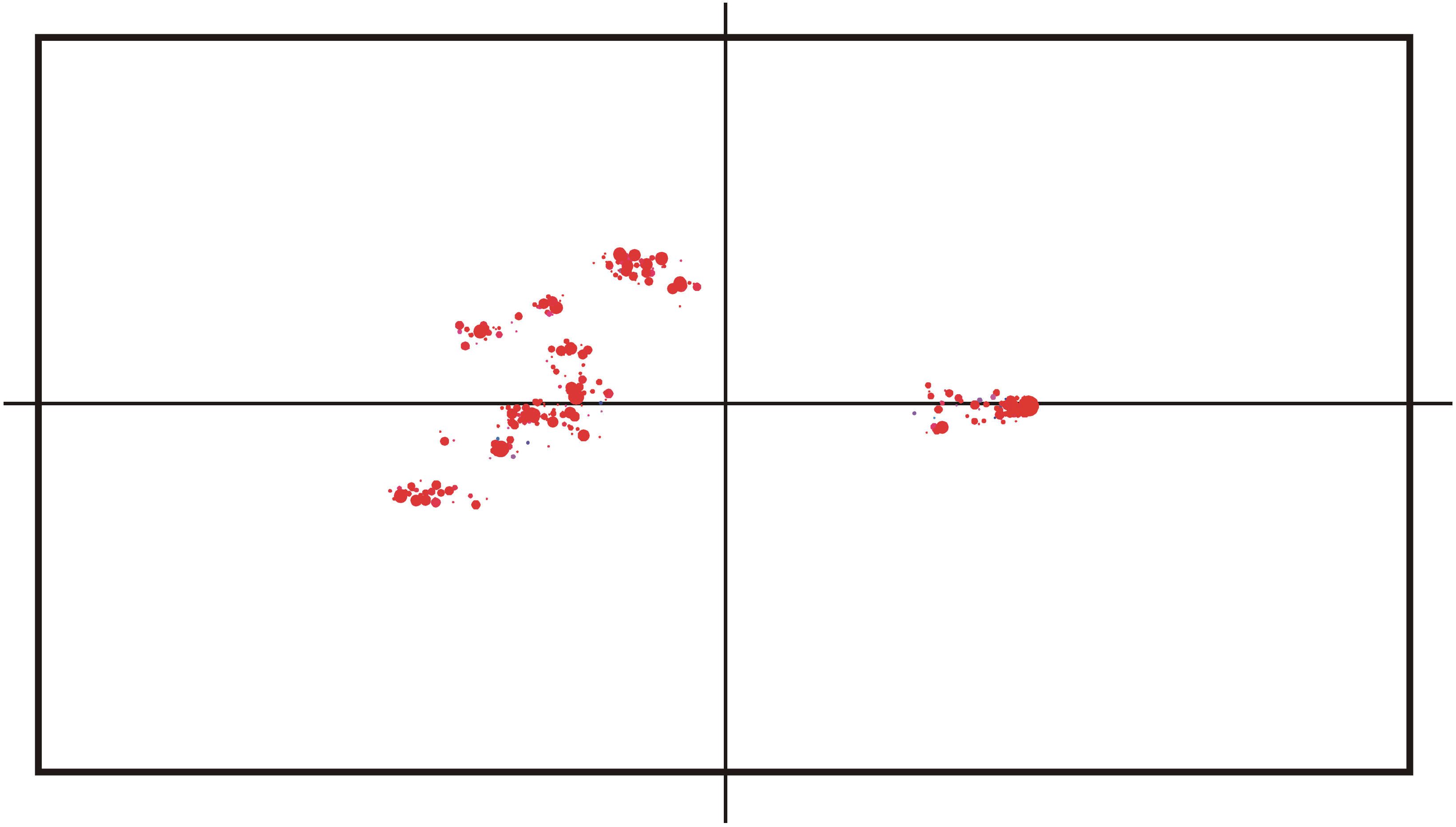
Spreading and evolution of a population on a neutral network: t = 650



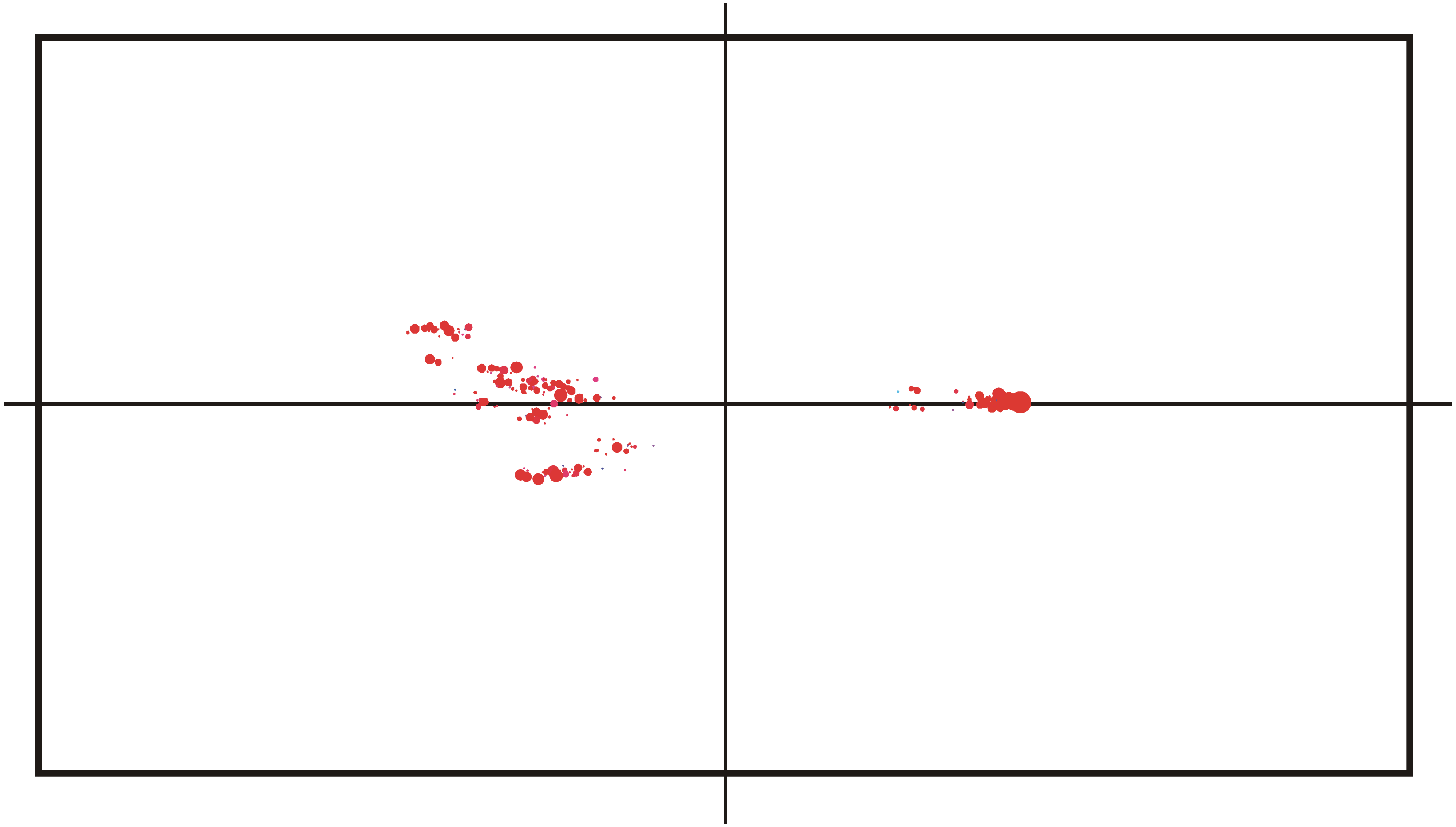
Spreading and evolution of a population on a neutral network: t = 820



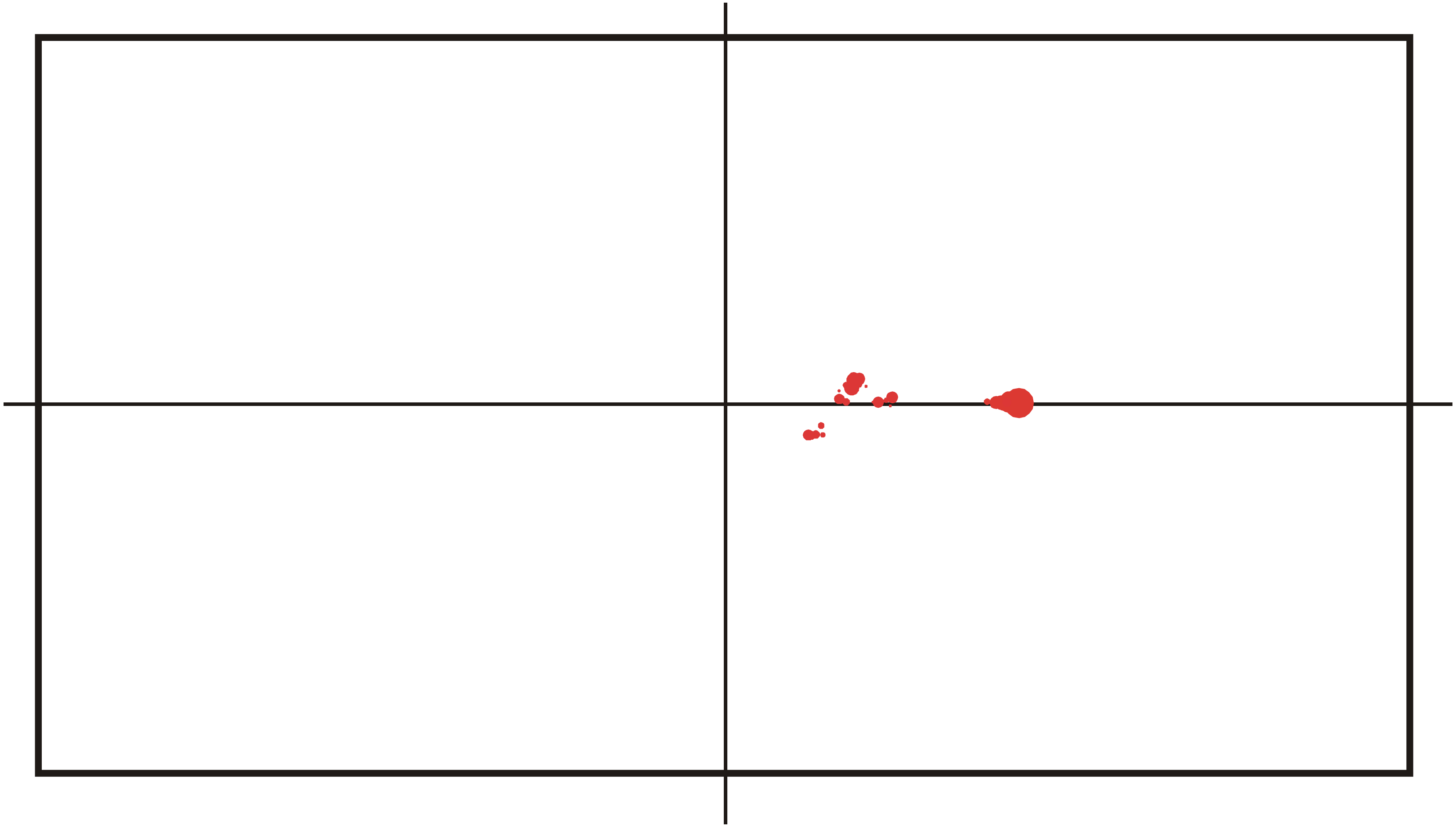
Spreading and evolution of a population on a neutral network: t = 825



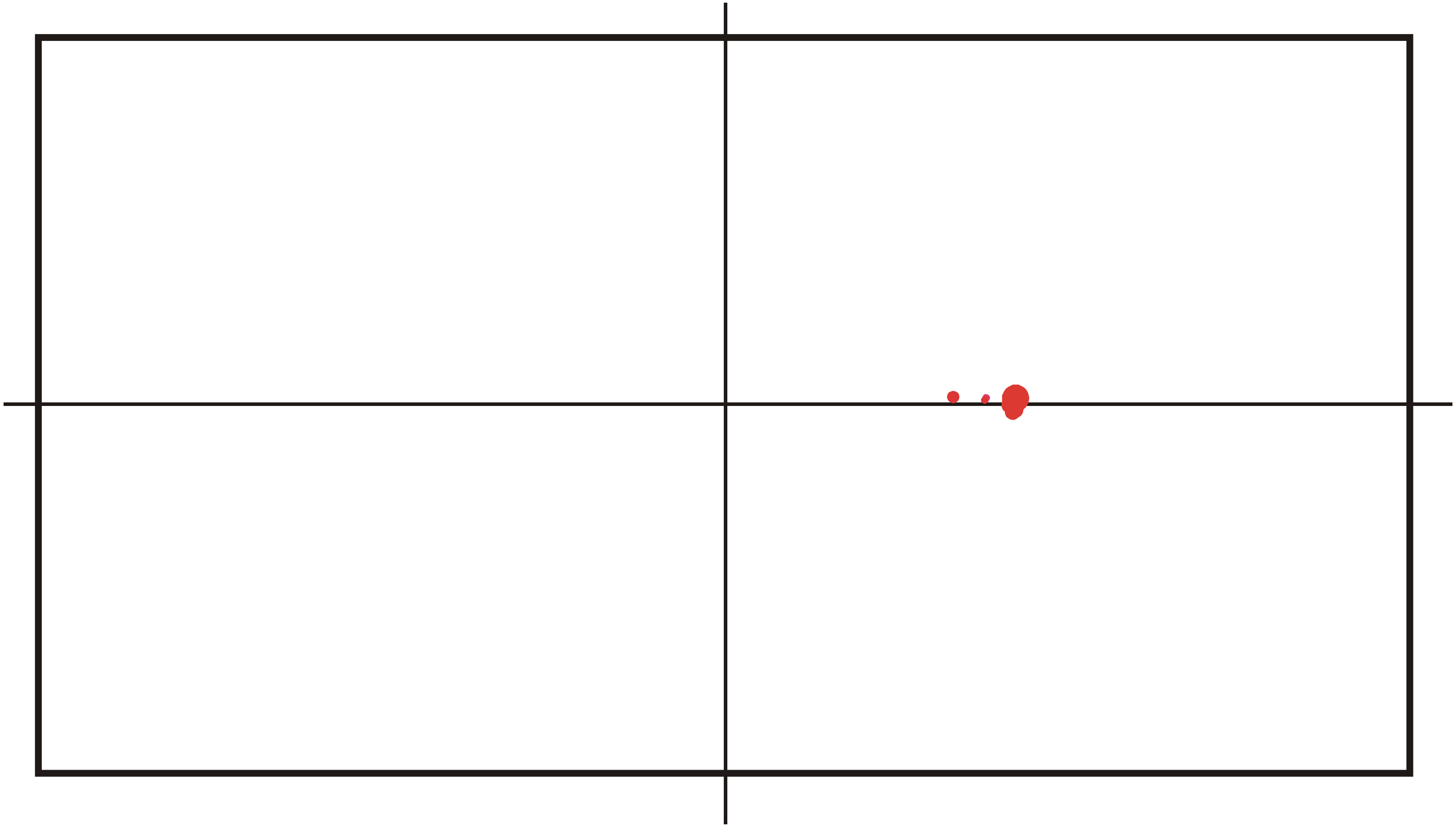
Spreading and evolution of a population on a neutral network: t = 830



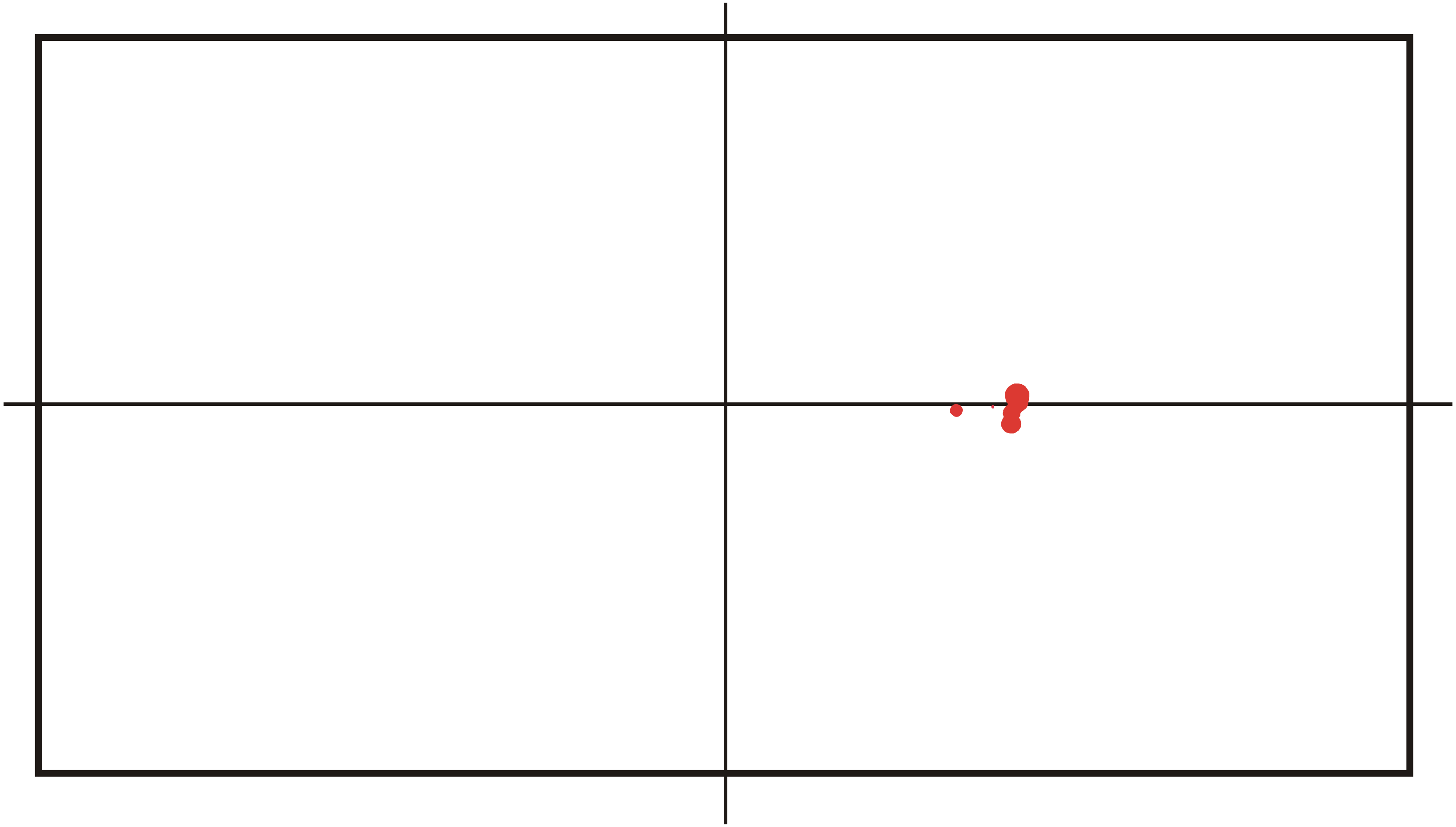
Spreading and evolution of a population on a neutral network: t = 835



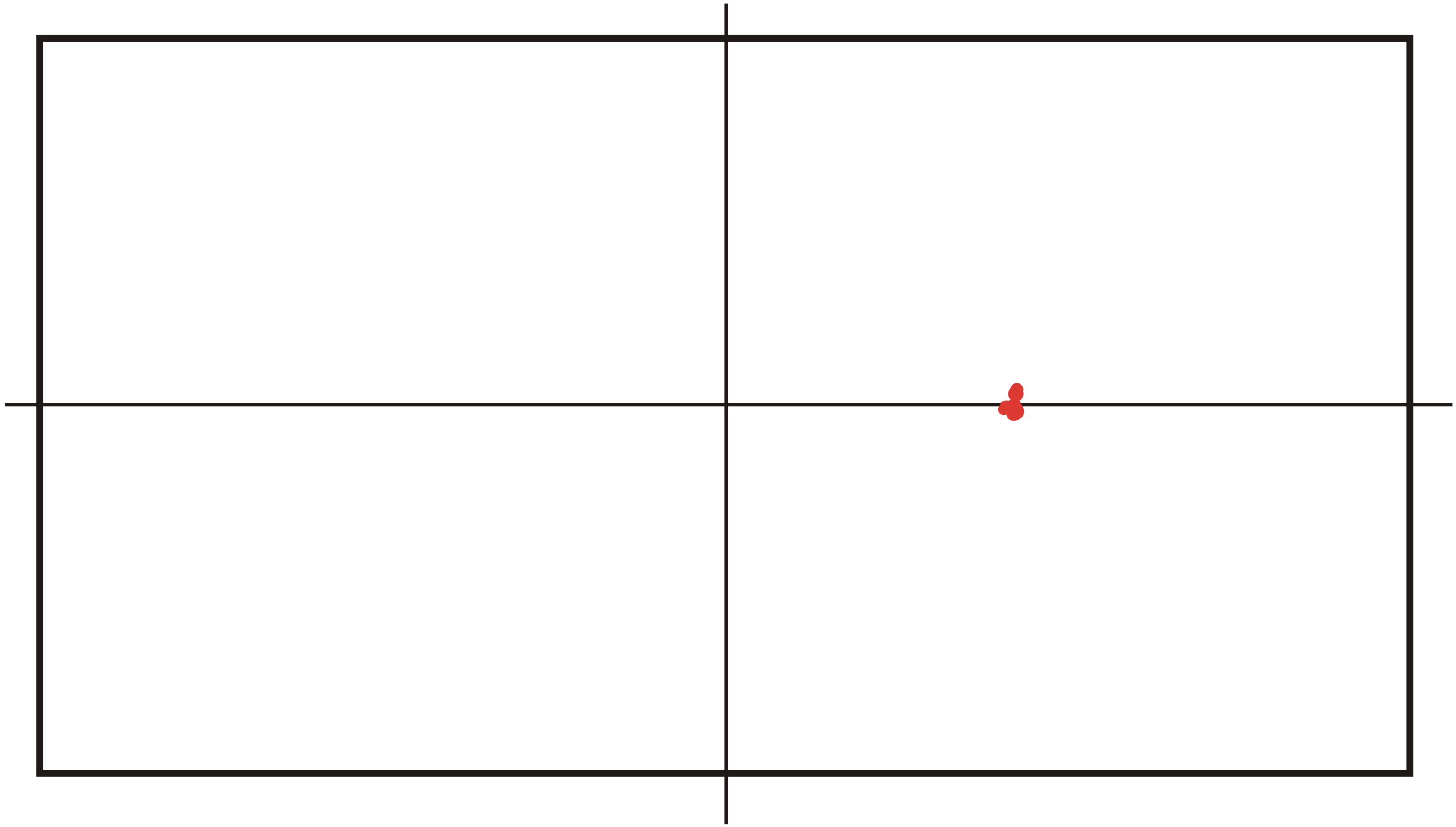
Spreading and evolution of a population on a neutral network: t = 840



Spreading and evolution of a population on a neutral network: t = 845



Spreading and evolution of a population on a neutral network: t = 850



Spreading and evolution of a population on a neutral network: t = 855

scenario	selection	mutation	population size	initial values	selected object	optimum
Darwinian	**strong**	**weak**	**large**	**large**	**single variant**	**yes**
molecular	no matter	strong	large	large	quasispecies	yes
stochastic	no matter	no matter	no matter	small	single variant	no
random drift	weak	strong	small	no matter	drifting clones	no

Darwinian evolution and optimization

scenario	selection	mutation	population size	initial values	selected object	optimum
Darwinian	strong	weak	large	large	single variant	yes
molecular	**no matter**	**strong**	**large**	**large**	**quasispecies**	**yes**
stochastic	no matter	no matter	no matter	small	single variant	no
random drift	weak	strong	small	no matter	drifting clones	no

molecular evolution and optimization

scenario	selection	mutation	population size	initial values	selected object	optimum
Darwinian	strong	weak	large	large	single variant	yes
molecular	no matter	strong	large	large	quasispecies	yes
stochastic	**no matter**	**no matter**	**no matter**	**small**	**single variant**	**no**
random drift	weak	strong	small	no matter	drifting clones	no

stochastic evolution and optimization

scenario	selection	mutation	population size	initial values	selected object	optimum
Darwinian	strong	weak	large	large	single variant	yes
molecular	no matter	strong	large	large	quasispecies	yes
stochastic	no matter	no matter	no matter	small	single variant	no
random drift	**weak**	**strong**	**small**	**no matter**	**drifting clones**	**no**

random drift and optimization

1. Mathematical concepts before Darwin
2. Theory of molecular evolution
3. Evolution in realistically small populations
4. **Evolution and present day molecular genetics**

An epigenetic trait is a stably inheritable phenotype resulting from chages in a chromosome without alterations in the DNA sequence.

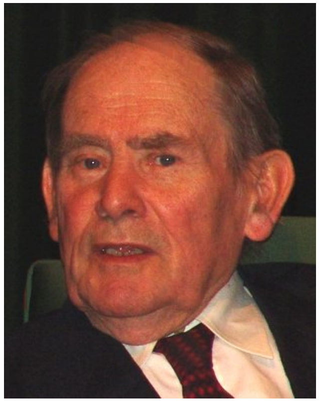
definition of epigenetic trait.

Cold Spring Harbor-Symposium, December 2008

What else is epigenetics than a funny form of enzymology ?

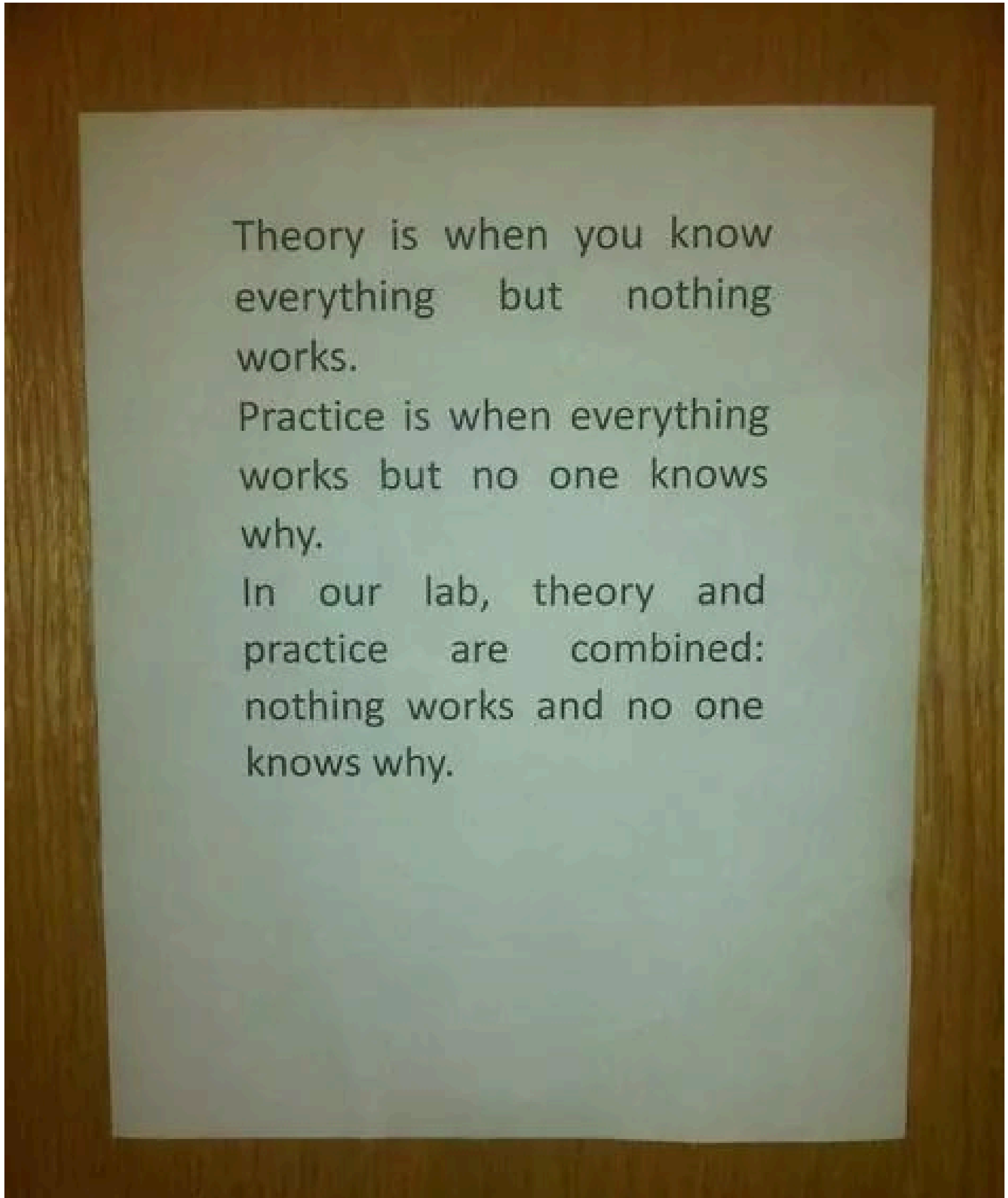
Each protein, after all, comes from some piece of DNA.

Sydney Brenner,
Wageningen, NL, 2010



Theory - mathematics and computation - cannot remove complexity, but it shows what kind of regular behavior can be expected and what experiments have to be done to get a grasp on the irregularities.

Manfred Eigen. In: E. Domingo et al., eds.
Origin and Evolution of Viruses, second edition.
Elsevier – Academic Press, Amsterdam, NL, 2008


Theory is when you know everything but nothing works.
Practice is when everything works but no one knows why.
In our lab, theory and practice are combined: nothing works and no one knows why.

Coworkers

Universität Wien

Peter Stadler, **Bärbel M. Stadler**, Universität Leipzig, GE

Walter Fontana, Harvard Medical School, MA

Martin Nowak, Harvard University, MA

Christian Reidys, Virginia Tech, Blacksburg, VA

Christian V. Forst, Icahn School of Medicine at Mount Sinai, NY

Ivo L.Hofacker, **Christoph Flamm**, Universität Wien, AT

Acknowledgement of support

Universität Wien

Fonds zur Förderung der wissenschaftlichen Forschung (FWF)

Österreichische Akademie der Wissenschaften

Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF)

Jubiläumsfonds der Österreichischen Nationalbank

European Commission

Austrian Genome Research Program – GEN-AU: Bioinformatics Network (BIN)

Siemens AG, Austria

Universität Wien and the Santa Fe Institute

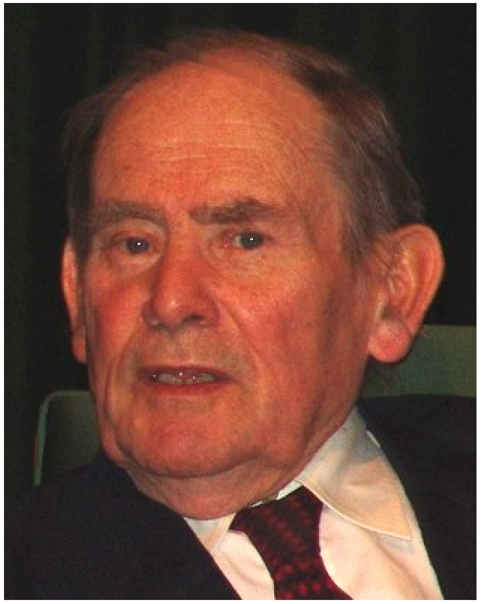
Thank you for your attention!

Web-Page for further information:

http://www.tbi.univie.ac.at/~pks

…. I was taught in the pregenomic era to be a hunter. I learnt how to identify the wild beasts and how to go out, hunt them down and kill them. We are now urged to be gatherers, to collect everything lying around and put it into storehouses.

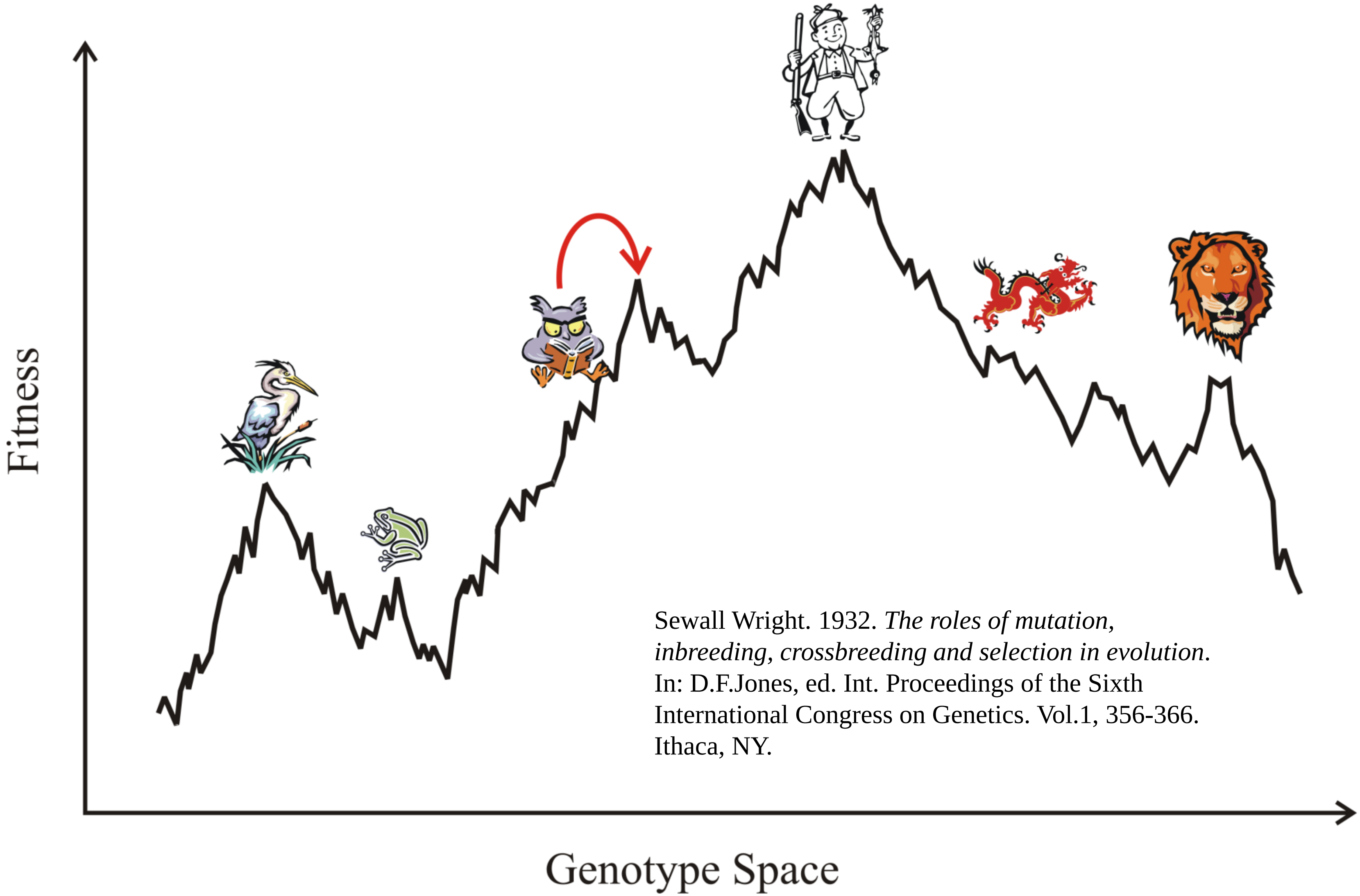
Someday, it is assumed, someone will come and sort through the storehouses, discard all the junk, and keep the rare finds. The only difficulty is how to recognize them.



Sydney Brenner, 1927 -

Sydney Brenner. Hunters and gatherers. *The Scientist* **16**(4): 14, 2002

the „big data“ problem in bioinformatics



Sewall Wrights fitness landscape as metaphor for Darwinian evolution