

What we can learn from simple models of evolution

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Austria

and

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



Seminar Lecture

Haifa, 03.03.2013

Web-Page for further information:

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

1. Gradualism and punctualism
2. Contingency in evolution experiments
3. Neutrality and its consequences
4. In silico-evolution of RNA structures

1. Gradualism and punctualism

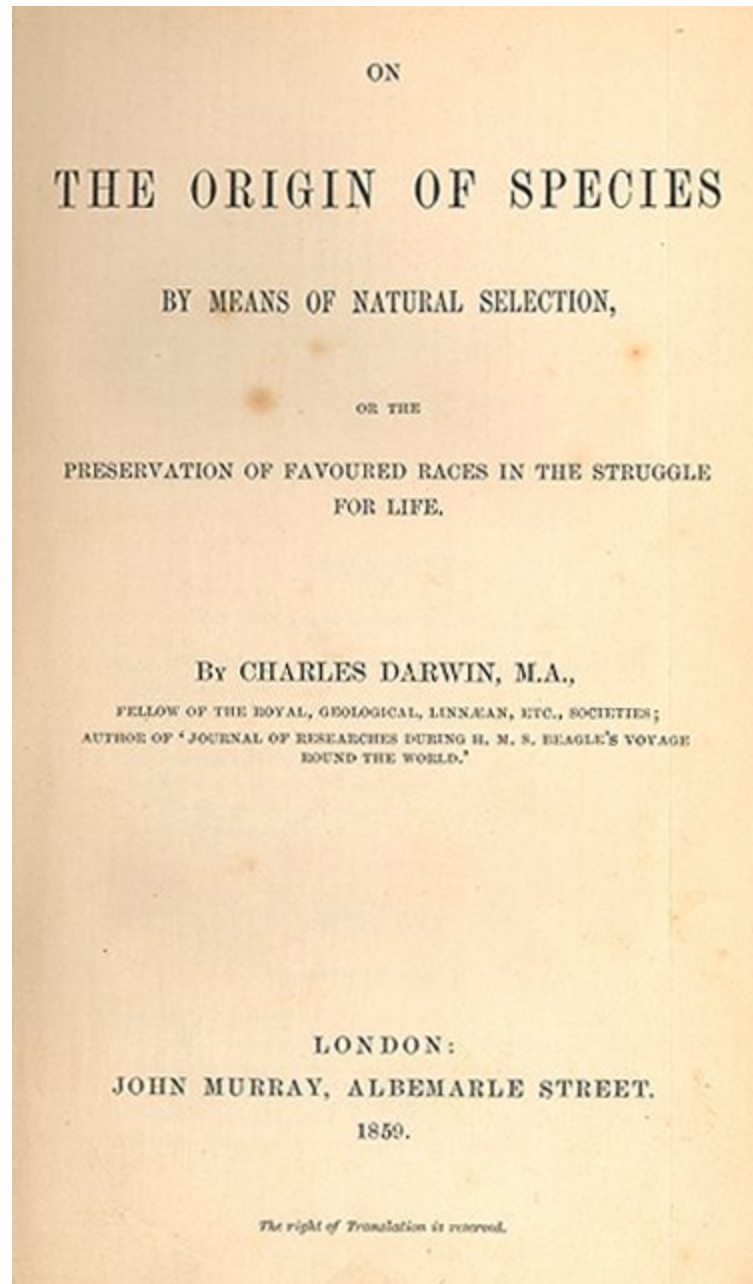
2. Contingency in evolution experiments

3. Neutrality and its consequences

4. In silico-evolution of RNA structures



Charles Darwin, 1809 - 1882



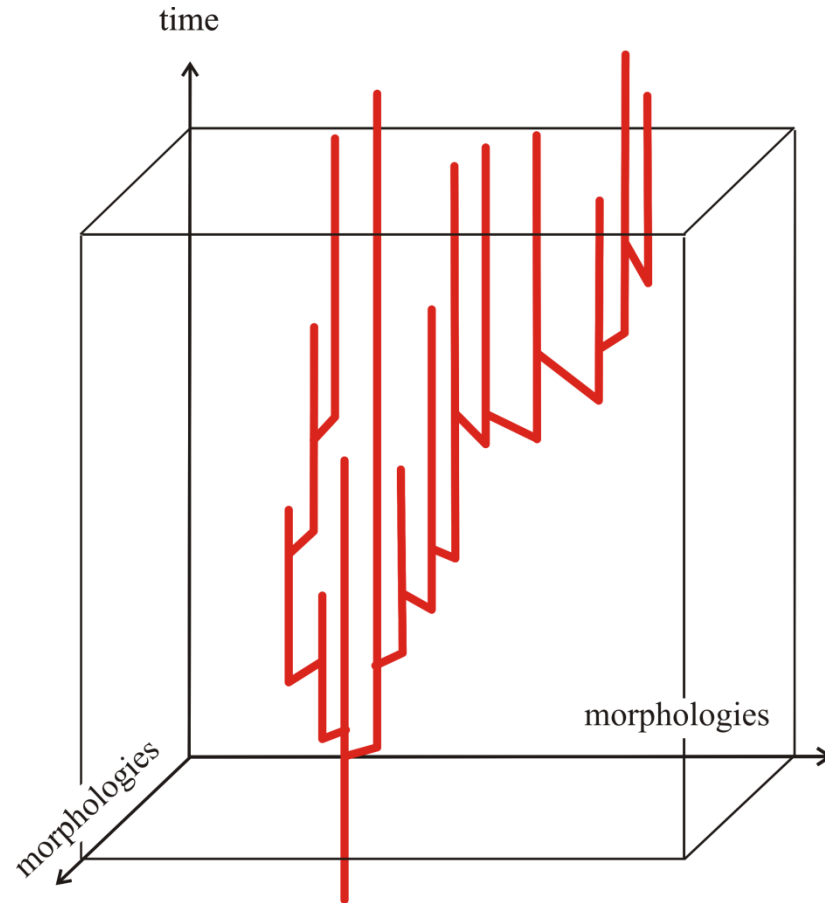
The five concepts of Darwin's theory of evolution from the
„Origin of Species“, 23.11.1859

Ernst Mayr. 1991. One long argument. Harvard University Press.

1. **evolution** - *the fact as such*
2. **common descent** - *all organisms have a common ancestor*
3. **multiplication of species** - *the formation of new species from existing ones*
4. **gradualism** - *all changes happen in (very) small steps*
5. **natural selection** - *adaptation to the environment as a result of the fact that only few individuals can master the competition for limited resources*



Stephen J. Gould,
1941 - 2002



Niles Eldredge, 1943 -

The concept of punctuated equilibrium

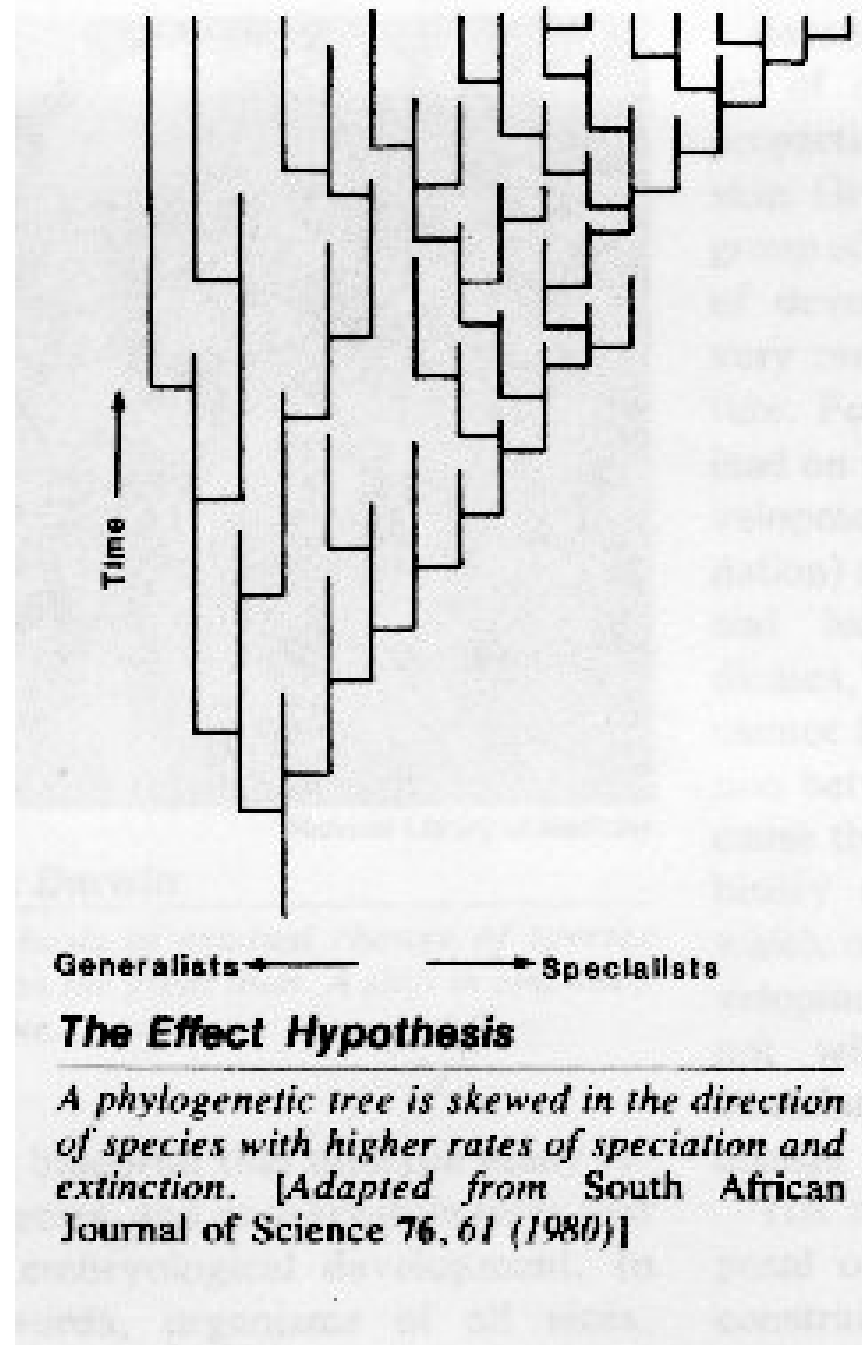


Gradualism versus punctualism in butterfly species formation



Elisabeth Vrba, 1943 -

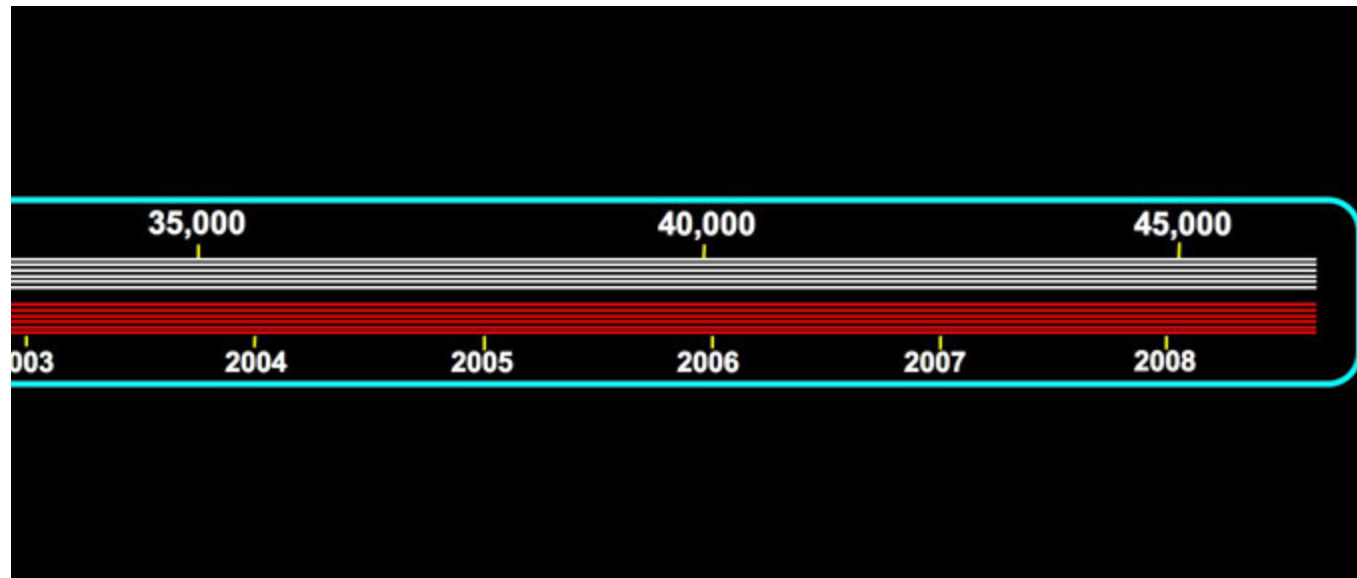
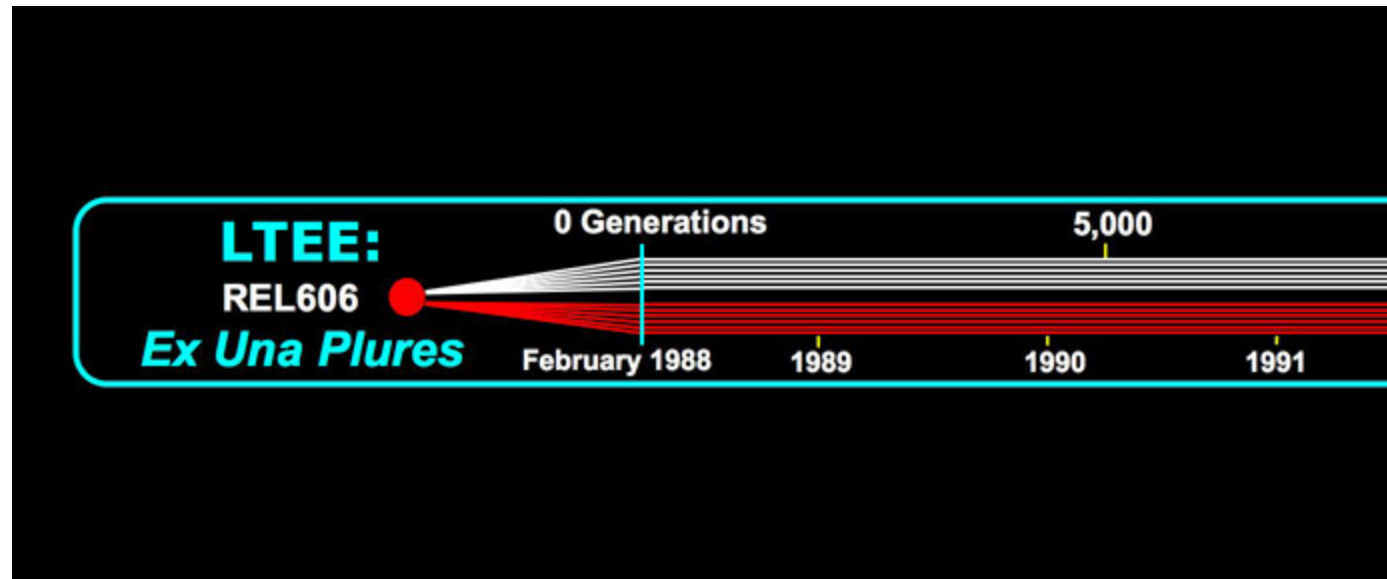
A speciation model based on
punctuated equilibrium



1. Gradualism and punctualism
2. **Contingency in evolution experiments**
3. Neutrality and its consequences
4. In silico-evolution of RNA structures

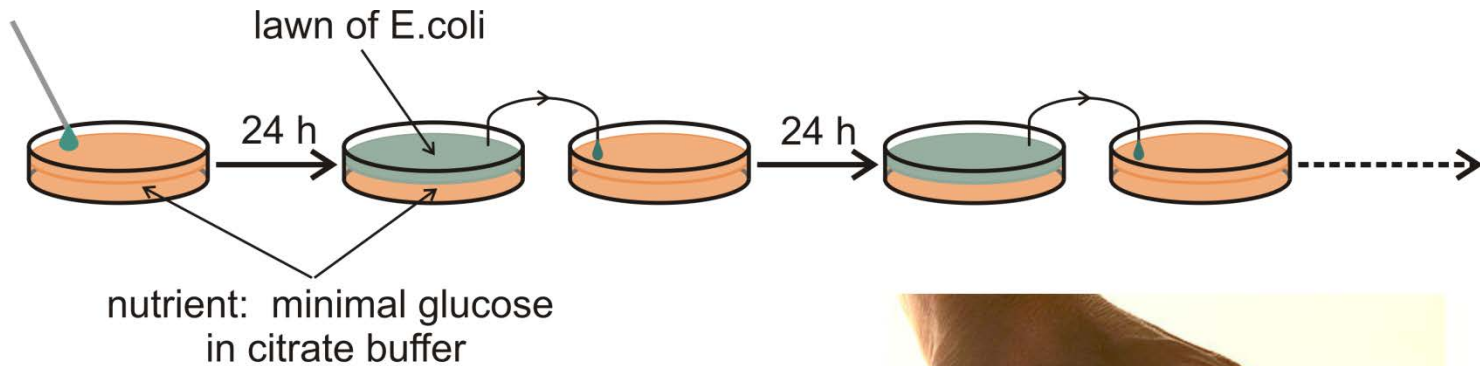


Richard Lenski, 1956 -



Bacterial evolution under controlled conditions: A twenty years experiment.

Richard Lenski, University of Michigan, East Lansing



medium supports $\approx 5 \times 10^8$ bacteria

1 day ≈ 6.67 generations

1 month ≈ 200 generations

1 year ≈ 2400 generations

Serial transfer of bacterial cultures in Petri dishes



Bacterial evolution under controlled conditions: A twenty years experiment.

Richard Lenski, University of Michigan, East Lansing



The twelve populations of Richard Lenski's long time evolution experiment

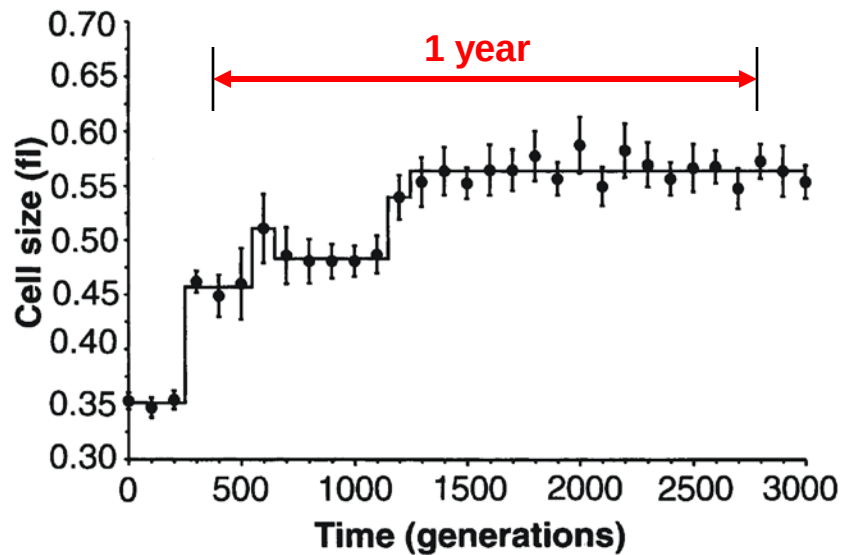


Fig. 1. Change in average cell size (1 fl = 10^{-15} L) in a population of *E. coli* during 3000 generations of experimental evolution. Each point is the mean of 10 replicate assays (22). Error bars indicate 95% confidence intervals. The solid line shows the best fit of a step-function model to these data (Table 1).

Epochal evolution of bacteria in serial transfer experiments under constant conditions

S. F. Elena, V. S. Cooper, R. E. Lenski. *Punctuated evolution caused by selection of rare beneficial mutants*. Science **272** (1996), 1802-1804

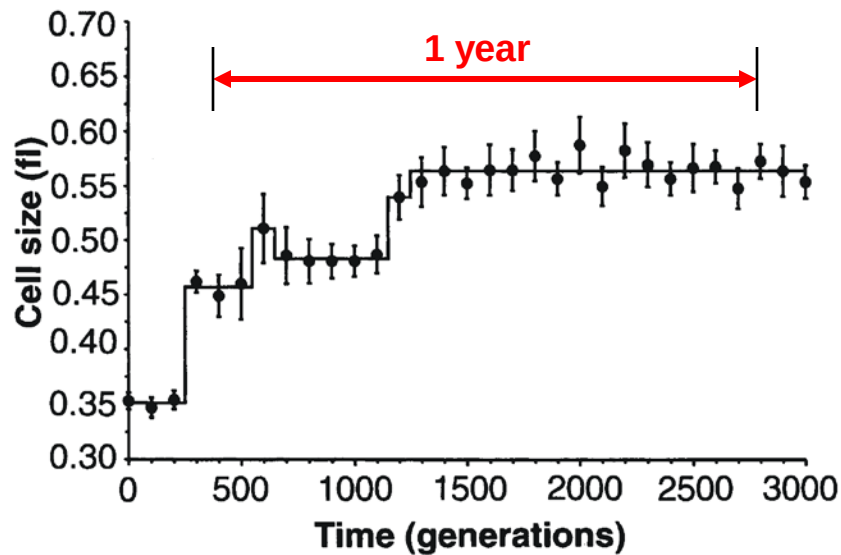


Fig. 1. Change in average cell size (1 fl = 10^{-15} L) in a population of *E. coli* during 3000 generations of experimental evolution. Each point is the mean of 10 replicate assays (22). Error bars indicate 95% confidence intervals. The solid line shows the best fit of a step-function model to these data (Table 1).

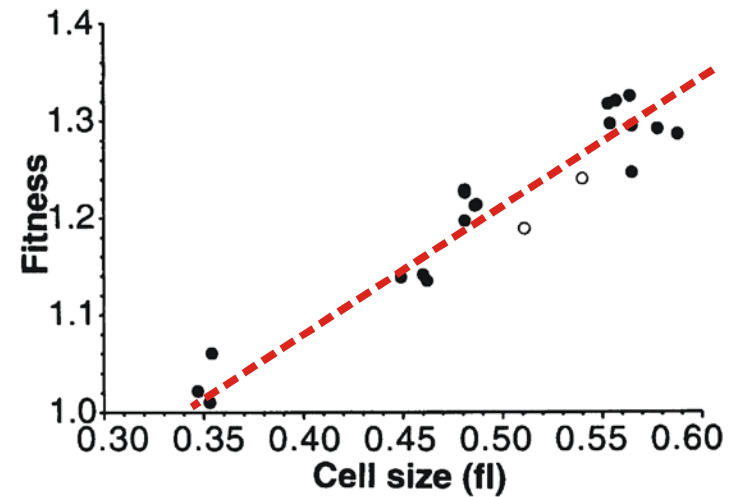


Fig. 2. Correlation between average cell size and mean fitness, each measured at 100-generation intervals for 2000 generations. Fitness is expressed relative to the ancestral genotype and was obtained from competition experiments between derived and ancestral cells (6, 7). The open symbols indicate the only two samples assigned to different steps by the cell size and fitness data.

Epochal evolution of bacteria in serial transfer experiments under constant conditions

S. F. Elena, V. S. Cooper, R. E. Lenski. *Punctuated evolution caused by selection of rare beneficial mutants.* Science **272** (1996), 1802-1804



The twelve populations of Richard Lenski's long time evolution experiment
Enhanced turbidity in population A-3

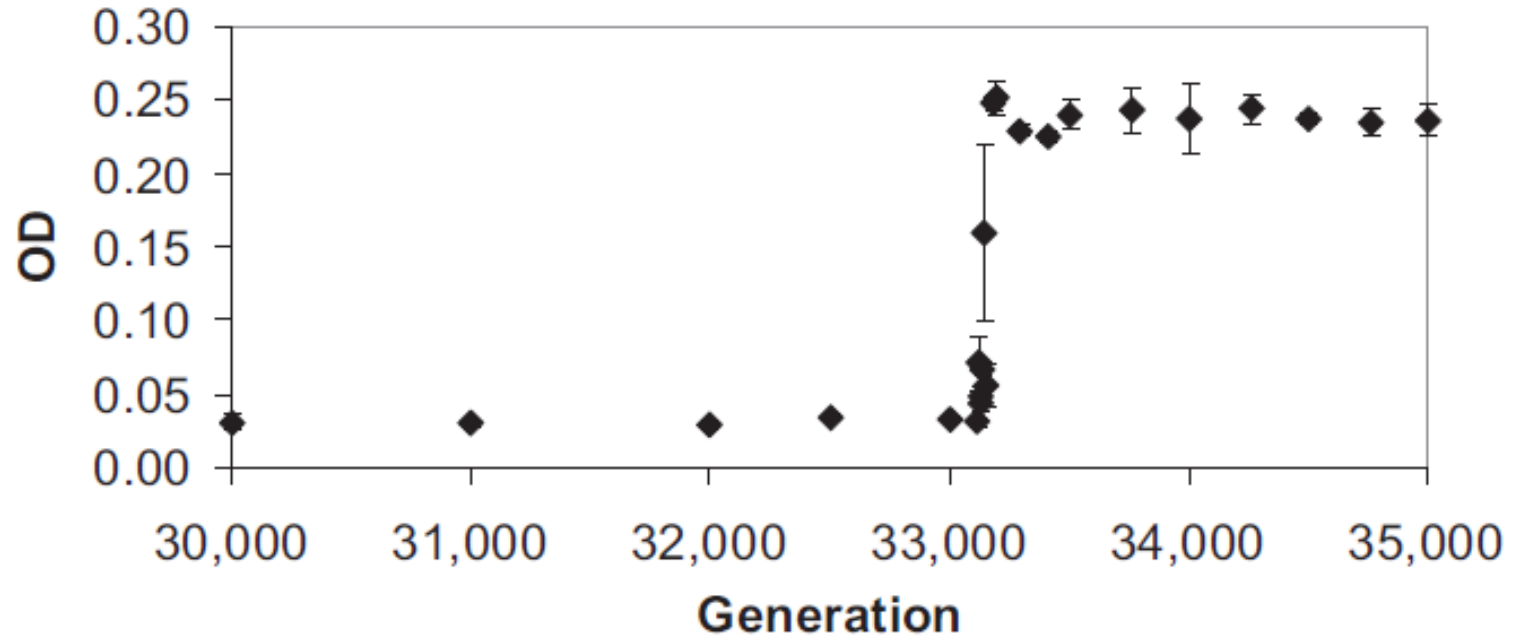


Fig. 1. Population expansion during evolution of the Cit⁺ phenotype. Samples frozen at various times in the history of population Ara-3 were revived, and three DM25 cultures were established for each generation. Optical density (OD) at 420 nm was measured for each culture at 24 h. Error bars show the range of three values measured for each generation.

Innovation by mutation in long time evolution of *Escherichia coli* in constant environment

Z.D. Blount, C.Z. Borland, R.E. Lenski. 2008. *Proc.Natl.Acad.Sci.USA* 105:7899-7906

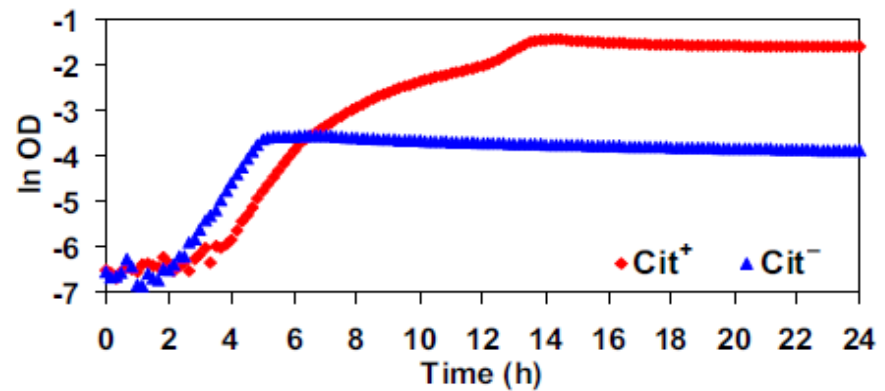


Fig. 2. Growth of *Cit⁻* (blue triangles) and *Cit⁺* (red diamonds) cells in DM25 medium. Each trajectory shows the average OD for eight replicate mixtures of three clones, all from generation 33,000 of population Ara-3.

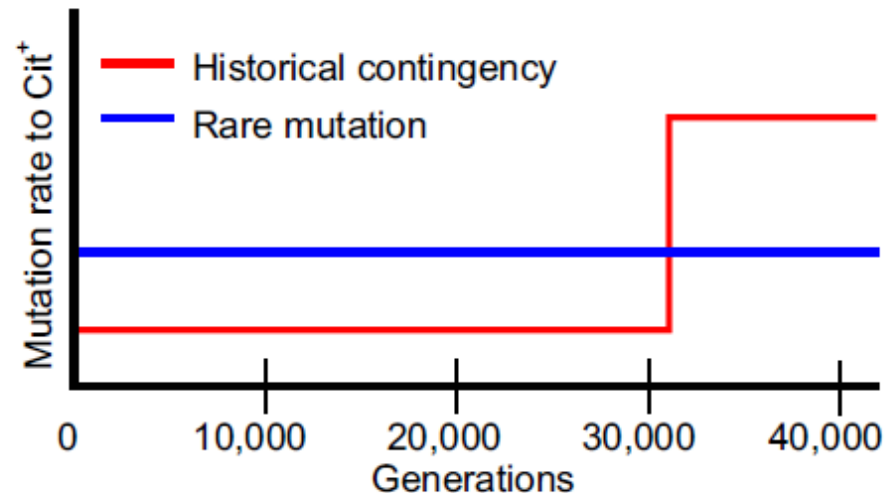


Fig. 3. Alternative hypotheses for the origin of the *Cit⁺* function. According to the rare-mutation hypothesis, the probability of mutation from *Cit⁻* to *Cit⁺* was low but constant over time. Under the historical-contingency hypothesis, the probability of this transition increased when a mutation arose that produced a genetic background with a higher mutation rate to *Cit⁺*.

Innovation by mutation in long time evolution of *Escherichia coli* in constant environment

Z.D. Blount, C.Z. Borland, R.E. Lenski. 2008.
Proc.Natl.Acad.Sci.USA
105:7899-7906

Table 1. Summary of replay experiments

Generation	First experiment		Second experiment		Third experiment	
	Replicates	Independent Cit ⁺ mutants	Replicates	Independent Cit ⁺ mutants	Replicates	Independent Cit ⁺ mutants
Ancestor	6	0	10	0	200	0
5,000	—	—	—	—	200	0
10,000	6	0	30	0	200	0
15,000	—	—	—	—	200	0
20,000	6	0	30	0	200	2
25,000	6	0	30	0	200	0
27,000	—	—	—	—	200	2
27,500	6	0	30	0	—	—
28,000	—	—	—	—	200	0
29,000	6	0	30	0	200	0
30,000	6	0	30	0	200	0
30,500	6	1	30	0	—	—
31,000	6	0	30	0	200	1
31,500	6	1	30	0	200	1
32,000	6	0	30	4	200	2
32,500	6	2	30	1	200	0
Totals	72	4	340	5	2,800	8

Contingency of E. coli evolution experiments

1. Gradualism and punctualism
2. Contingency in long-time evolution
3. **Neutrality and its consequences**
4. In silico-evolution of RNA structures

What is neutrality ?

Selective neutrality =
= several genotypes having the **same fitness**.

Structural neutrality =
= several genotypes forming molecules with
the **same structure**.



ON
THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

OR THE
PRESERVATION OF FAVOURED RACES IN THE STRUGGLE
FOR LIFE.

By CHARLES DARWIN, M.A.,

FELLOW OF THE ROYAL, GEOLOGICAL, LINNEAN, ETC., SOCIETIES;
AUTHOR OF 'JOURNAL OF RESEARCHES DURING H. M. S. BEAGLE'S VOYAGE
ROUND THE WORLD.'

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1859.

The right of Translation is reserved.

This preservation of favourable individual differences and variations, and the destruction of those which are injurious, I have called Natural Selection, or the Survival of the Fittest. Variations neither useful nor injurious would not be affected by natural selection, and would be left either a fluctuating element, as perhaps we see in certain polymorphic species, or would ultimately become fixed, owing to the nature of the organism and the nature of the conditions.

Charles Darwin. *The Origin of Species*. Sixth edition. John Murray. London: 1872



Motoo Kimura's population genetics of neutral evolution.

Evolutionary rate at the molecular level.
Nature **217**: 624-626, 1955.

The Neutral Theory of Molecular Evolution.
 Cambridge University Press. Cambridge,
 UK, 1983.

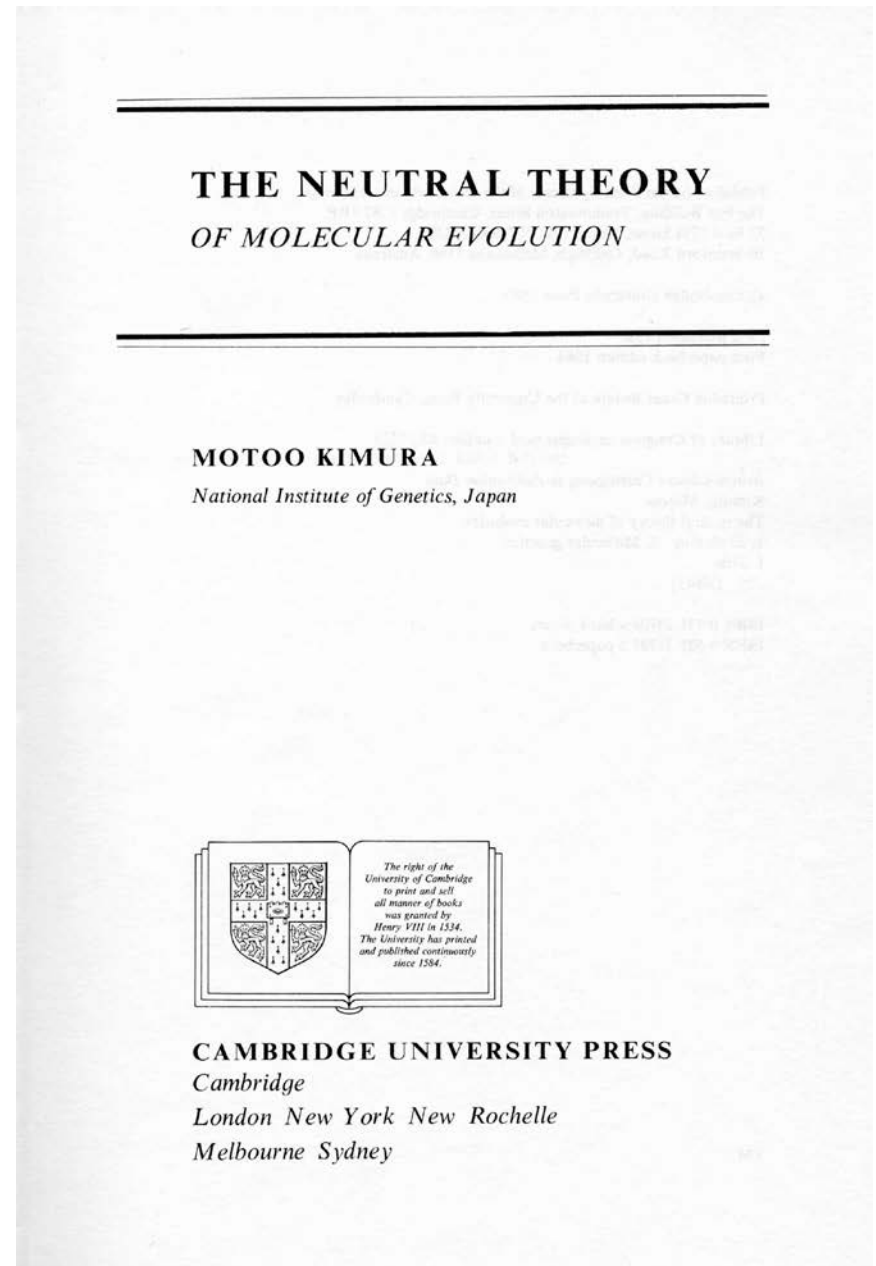
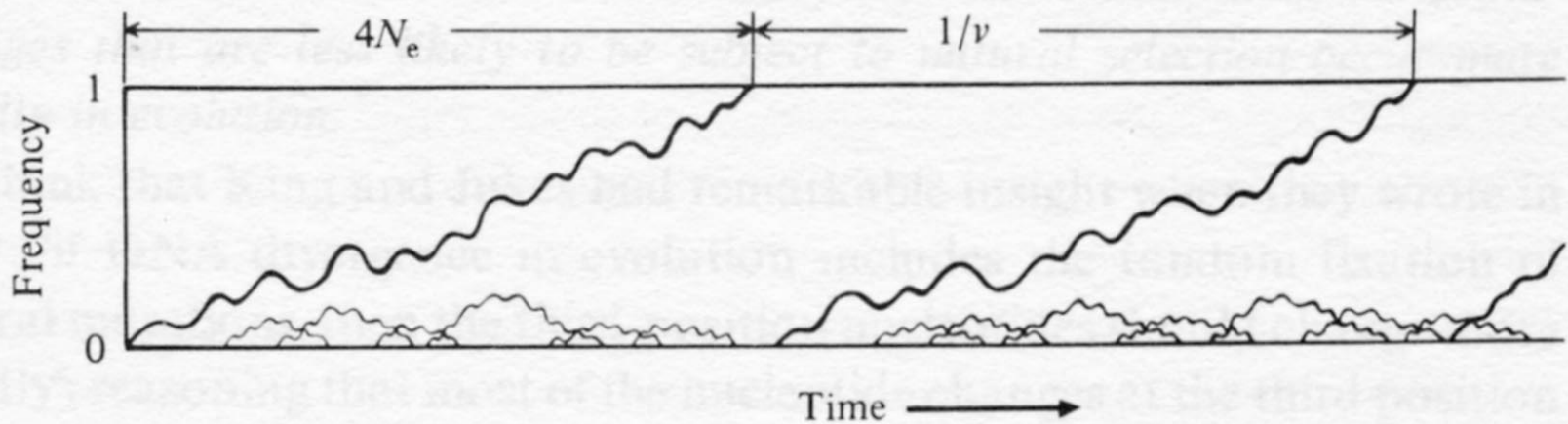


Fig. 3.1. Behavior of mutant genes following their appearance in a finite population. Courses of change in the frequencies of mutants destined to fixation are depicted by thick paths. N_e stands for the effective population size and v is the mutation rate.



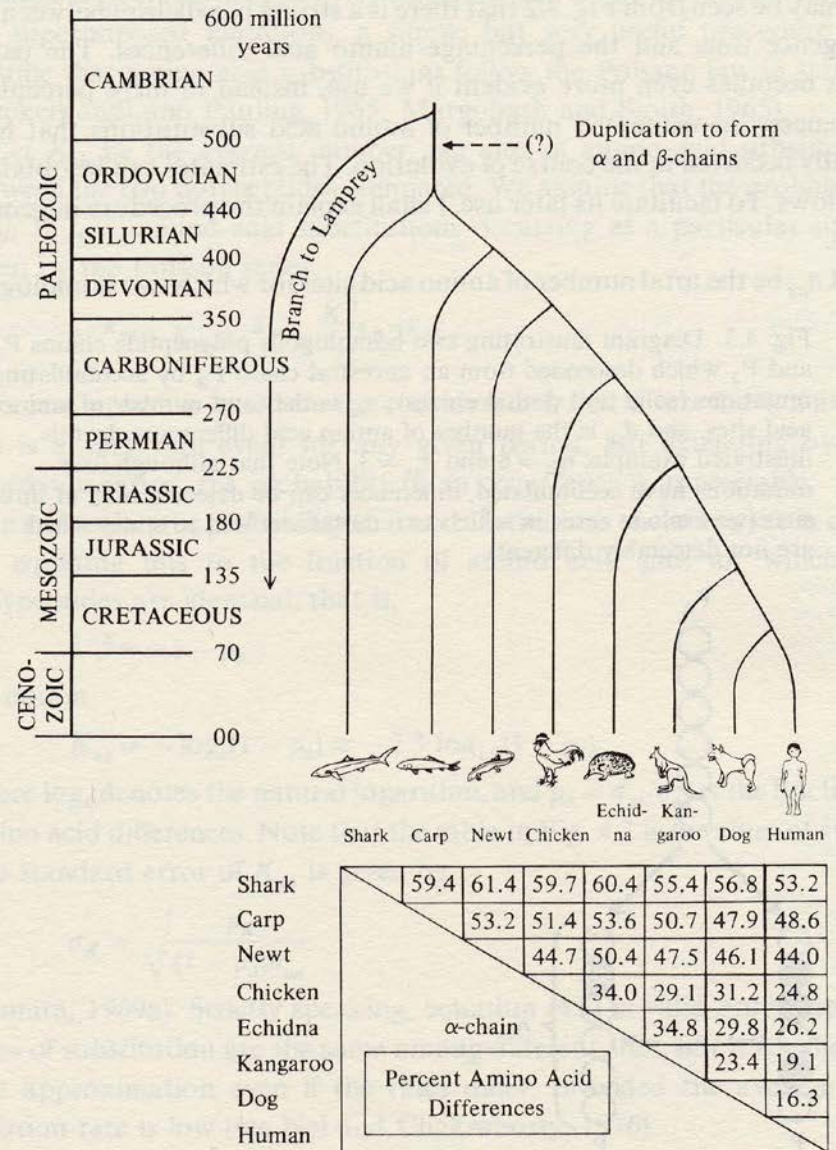
The average time of replacement of a dominant genotype in a population is the reciprocal mutation rate, $1/v$, and therefore independent of population size.

Fixation of mutants in neutral evolution (Motoo Kimura, 1955)

The molecular clock of evolution

Motoo Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press. Cambridge, UK, 1983.

Fig. 4.2. Percentage amino acid differences when the α hemoglobin chains are compared among eight vertebrates together with their phylogenetic relationship and the times of divergence.

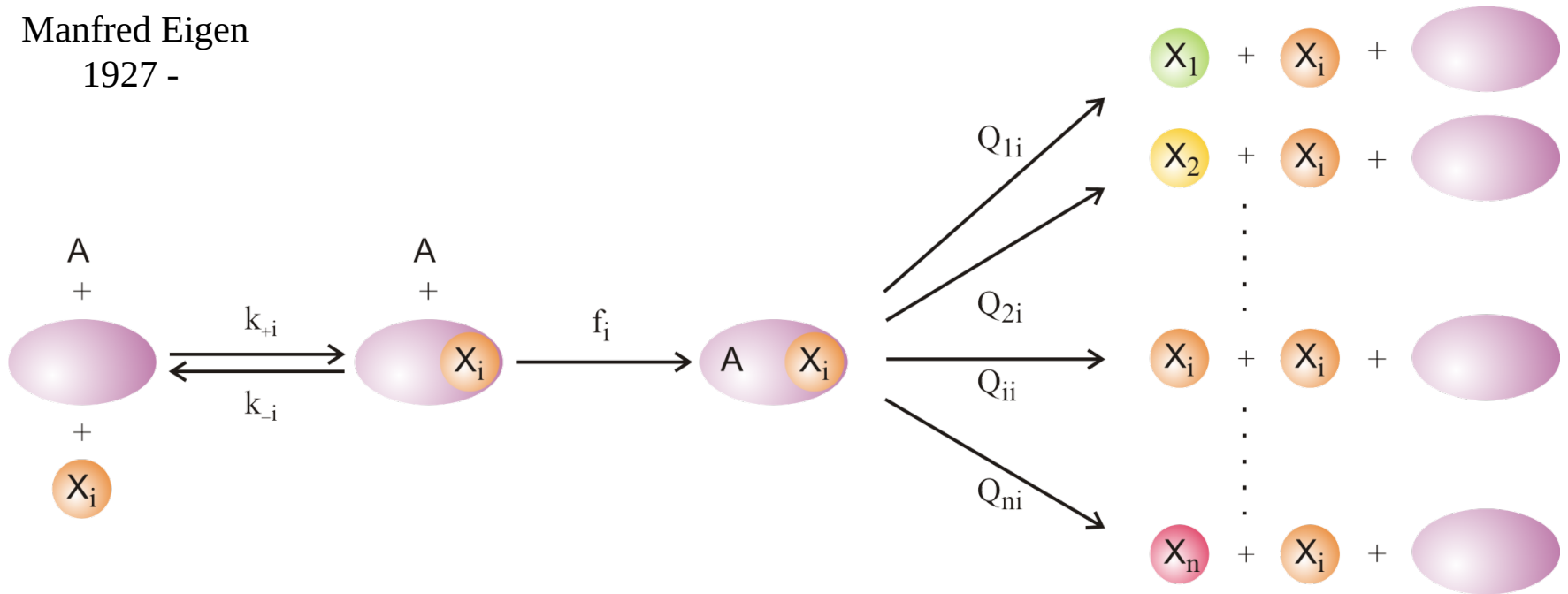




Manfred Eigen
1927 -

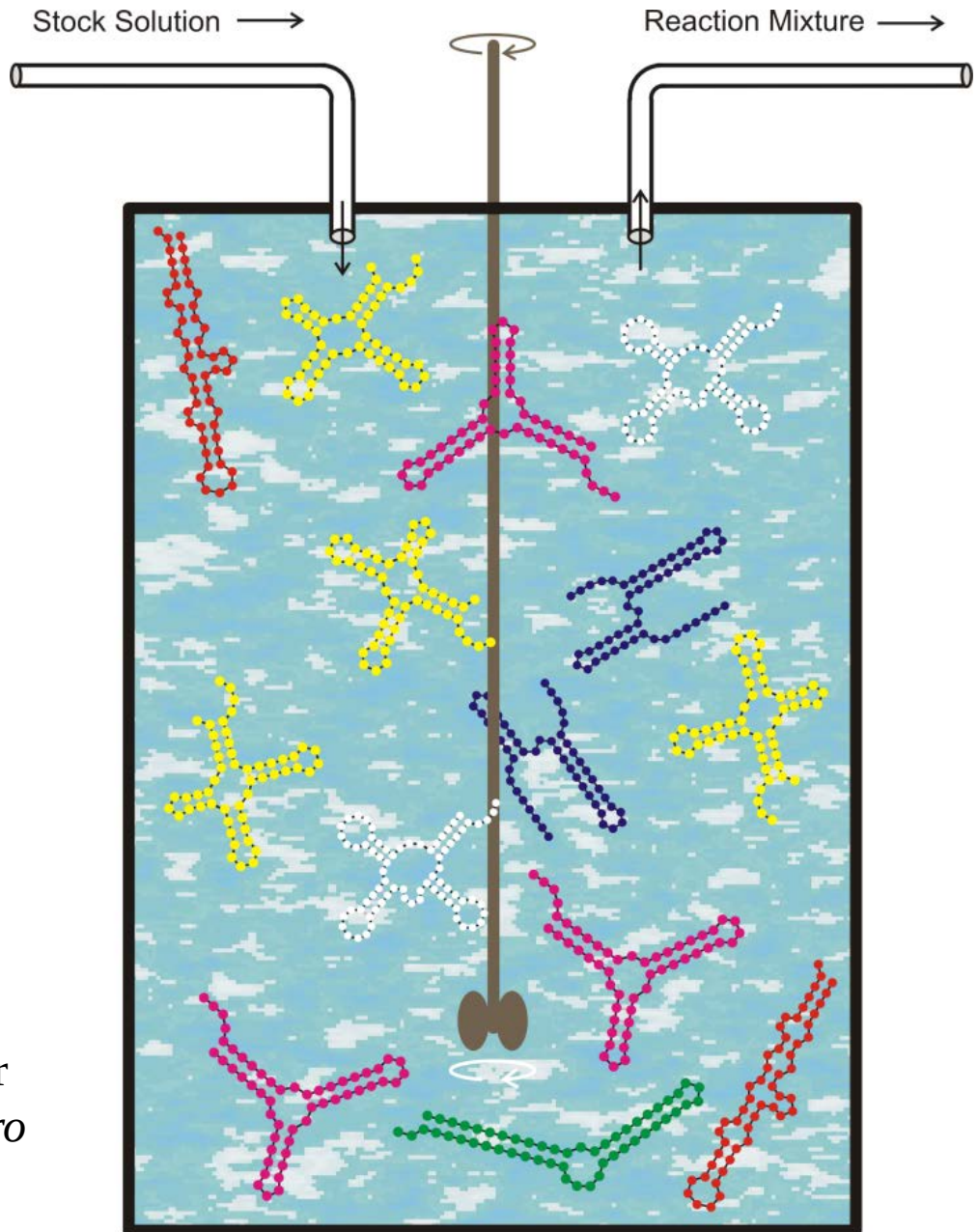
$$\frac{dx_j}{dt} = \sum_{i=1}^n W_{ji} x_i - x_j \Phi ; \quad j = 1, 2, \dots, n$$

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



Mutation and (correct) replication as parallel chemical reactions

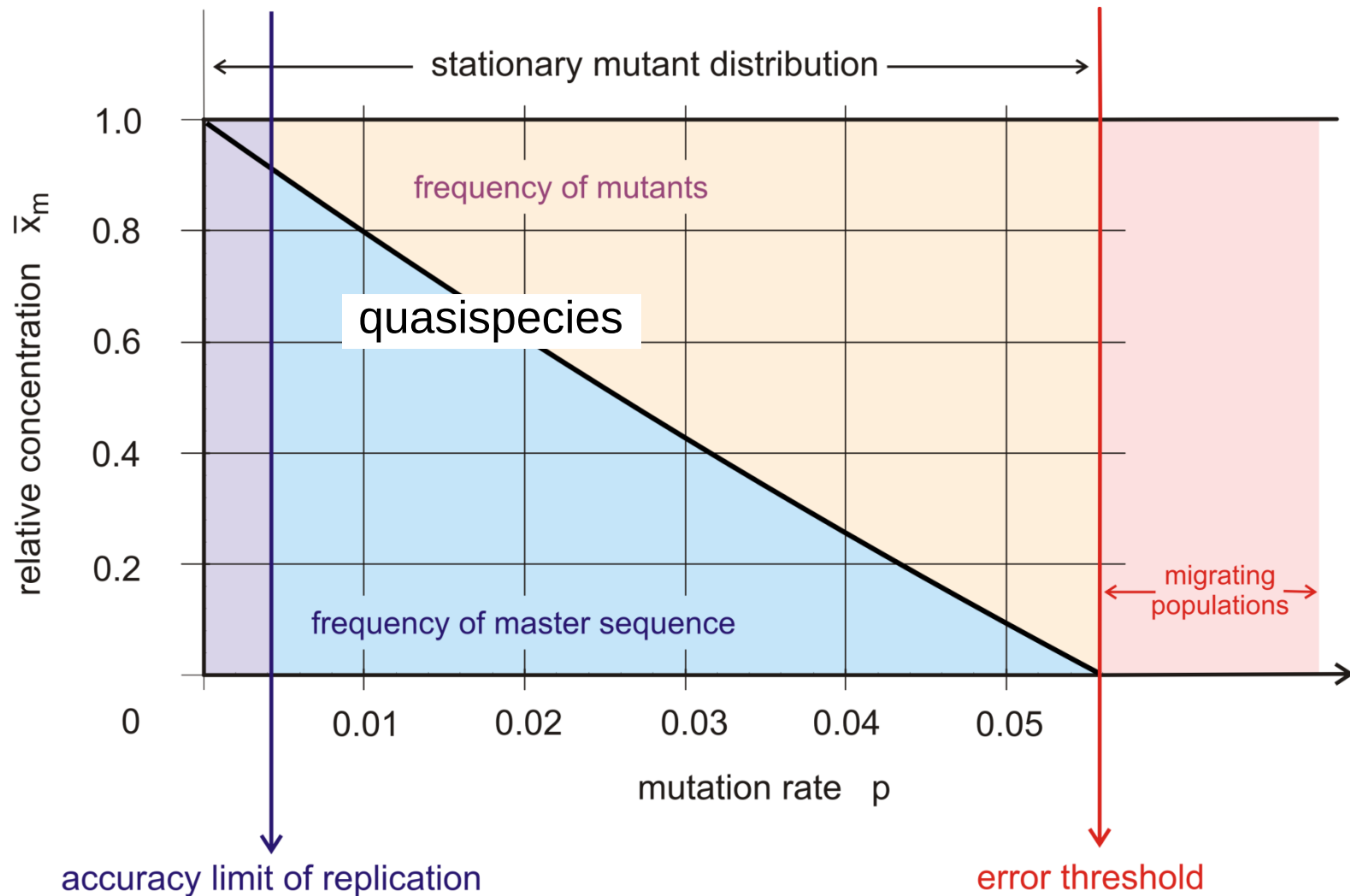
M. Eigen. 1971. *Naturwissenschaften* 58:465,
M. Eigen & P. Schuster. 1977. *Naturwissenschaften* 64:541, 65:7 und 65:341



Stock solution:

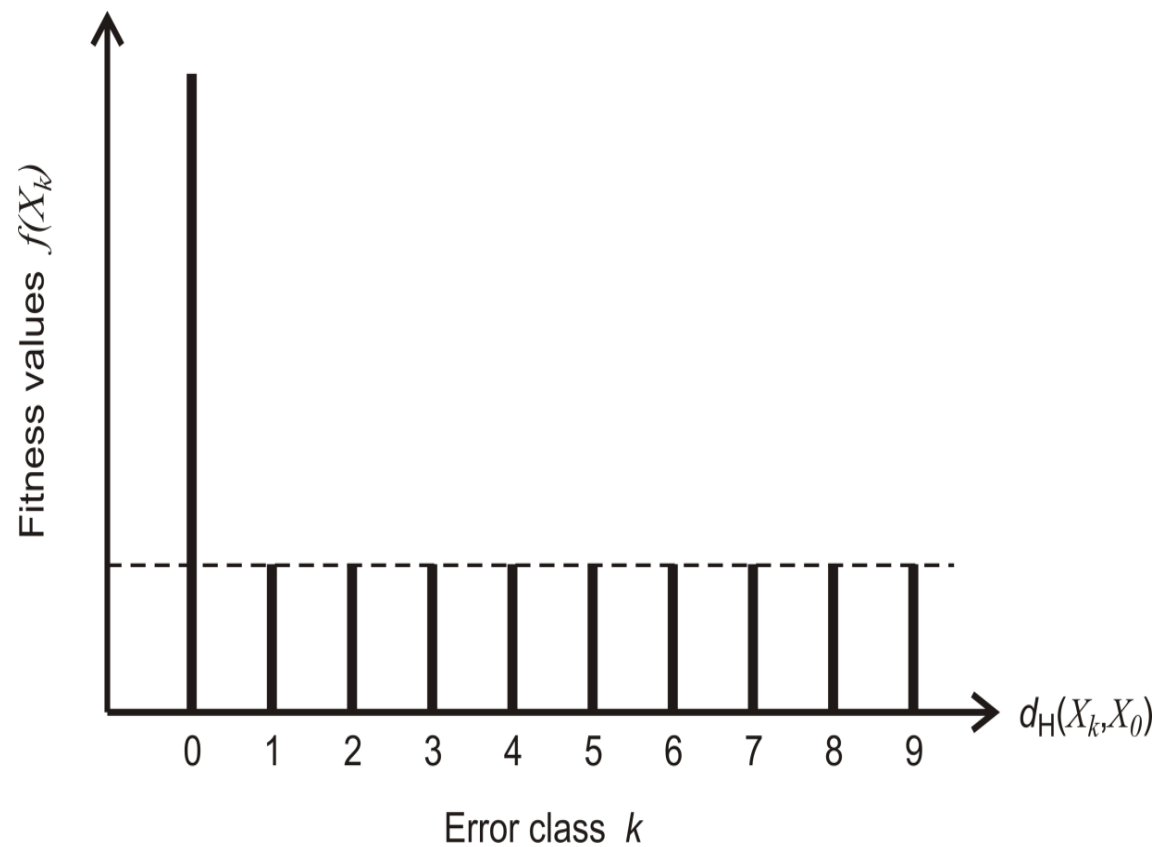
activated monomers, **ATP, CTP, GTP, UTP**;
a replicase, an enzyme that performs
complementary replication;
buffer solution

The continuously stirred tank reactor
(CSTR) as a tool for studies on *in vitro*
evolution and computer simulation.



The error threshold in replication and mutation

single peak landscape



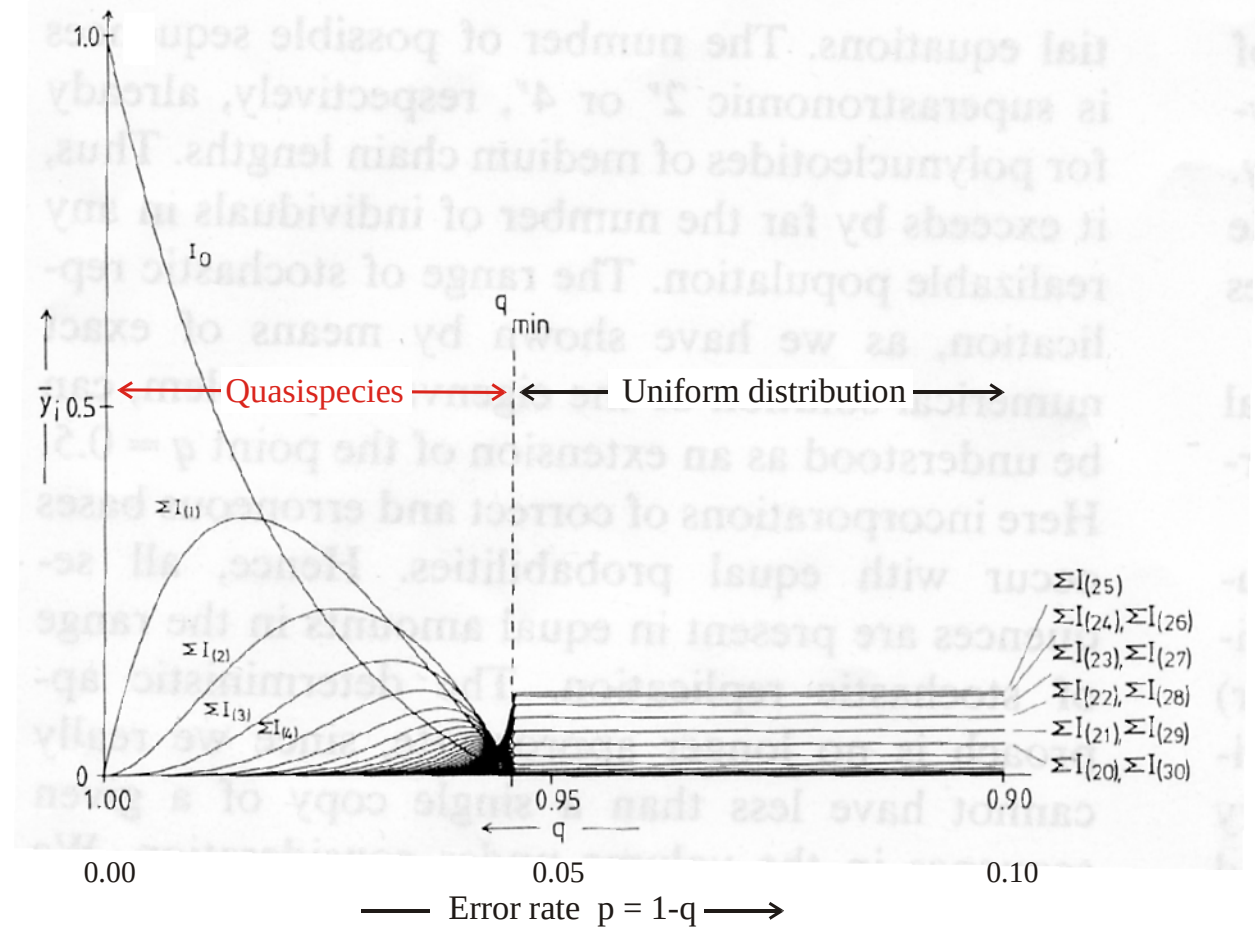
A model fitness landscape that was accessible to computation in the nineteen eighties

SELF-REPLICATION WITH ERRORS

A MODEL FOR POLYNUCLEOTIDE REPLICATION **

Jörg SWETINA and Peter SCHUSTER *

Institut für Theoretische Chemie und Strahlenchemie der Universität, Währingerstraße 17, A-1090 Wien, Austria



Stationary population or
quasispecies as a function
of the mutation or error
rate p

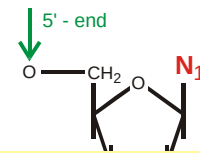
Realistic fitness landscapes

1. **Ruggedness**: nearby lying genotypes may develop into very different phenotypes

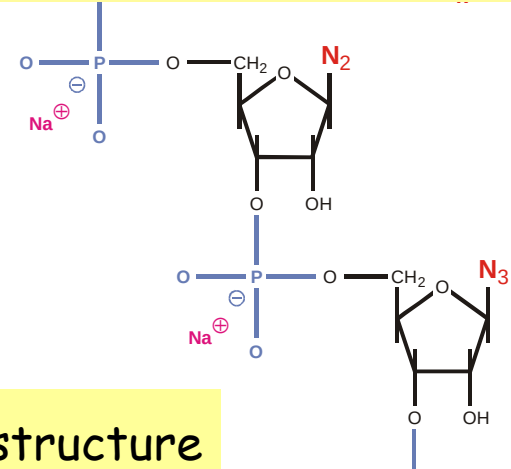
2. **Neutrality**: many different genotypes give rise to phenotypes with identical selection behavior

3. **Combinatorial explosion**: the number of possible genomes is prohibitive for systematic searches

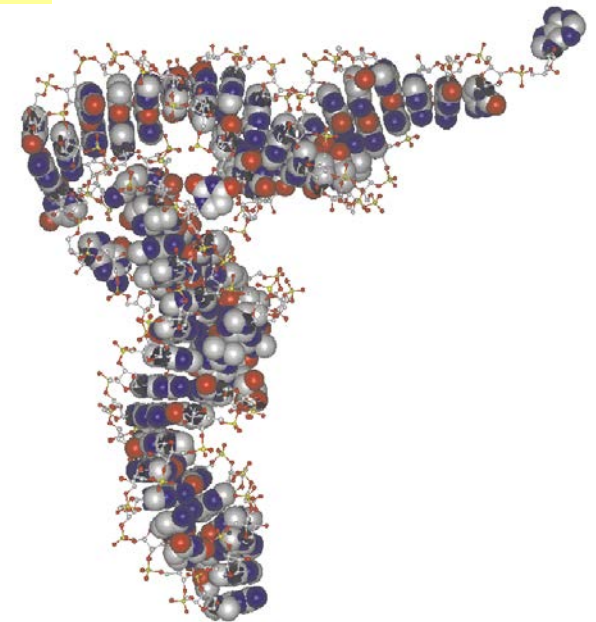
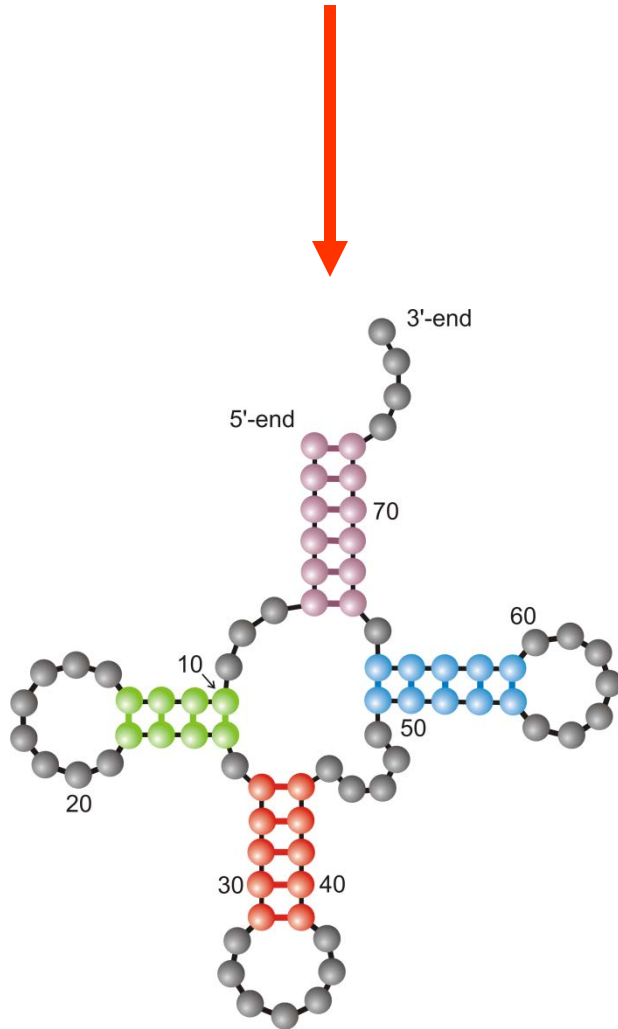
and hence, any successful and applicable theory of molecular evolution must be able to predict evolutionary dynamics from a small or at least in practice measurable number of fitness values.

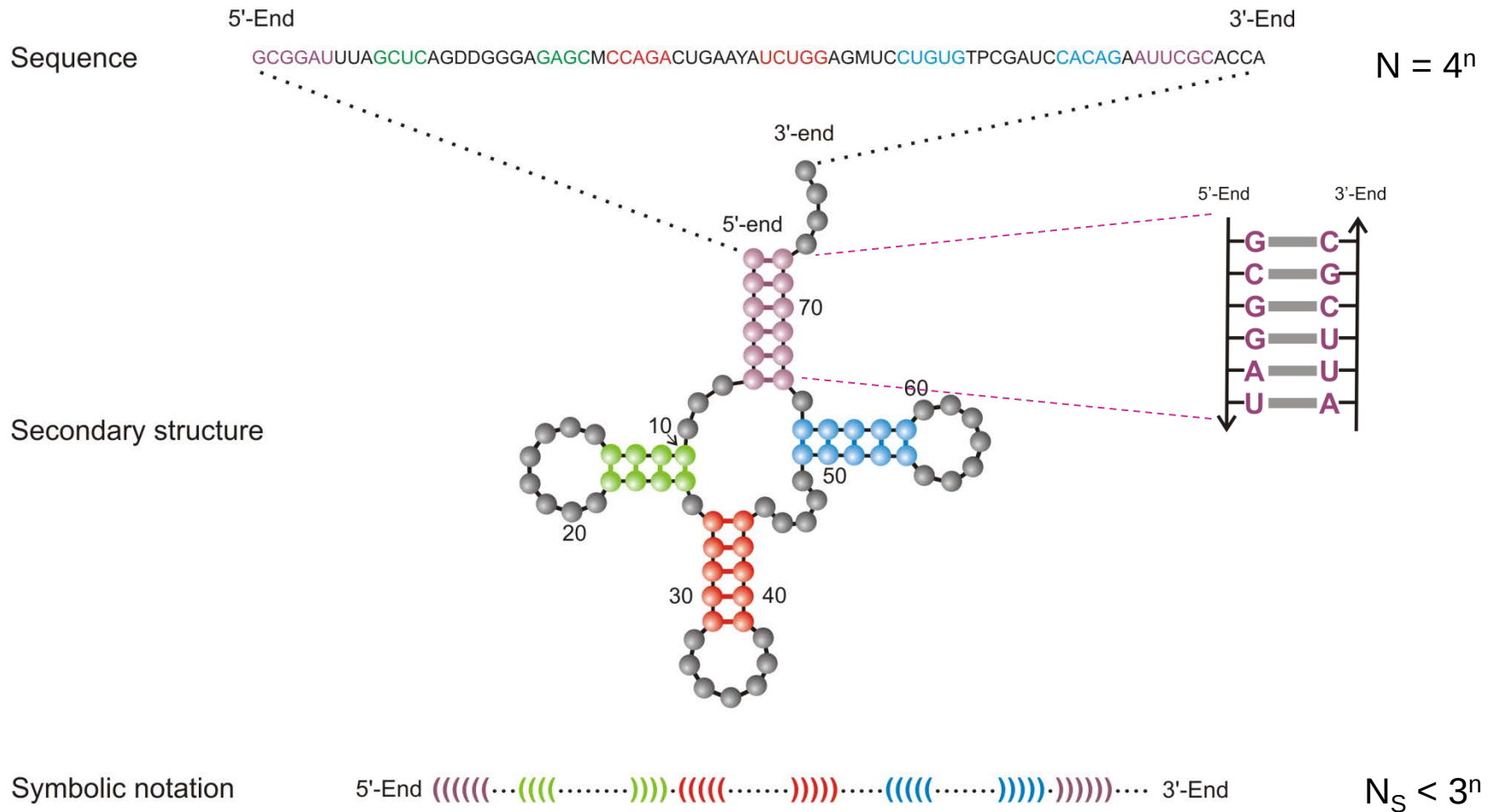


5'-end **GCGGAUUUAGCUC**AGUUGGGAGAG**CGCCAGACUGAAGAUCUGG**AGGUC**CUGUGUUCGAUCCACAGAAUUCGCACCA** 3'-end



Definition of RNA structure

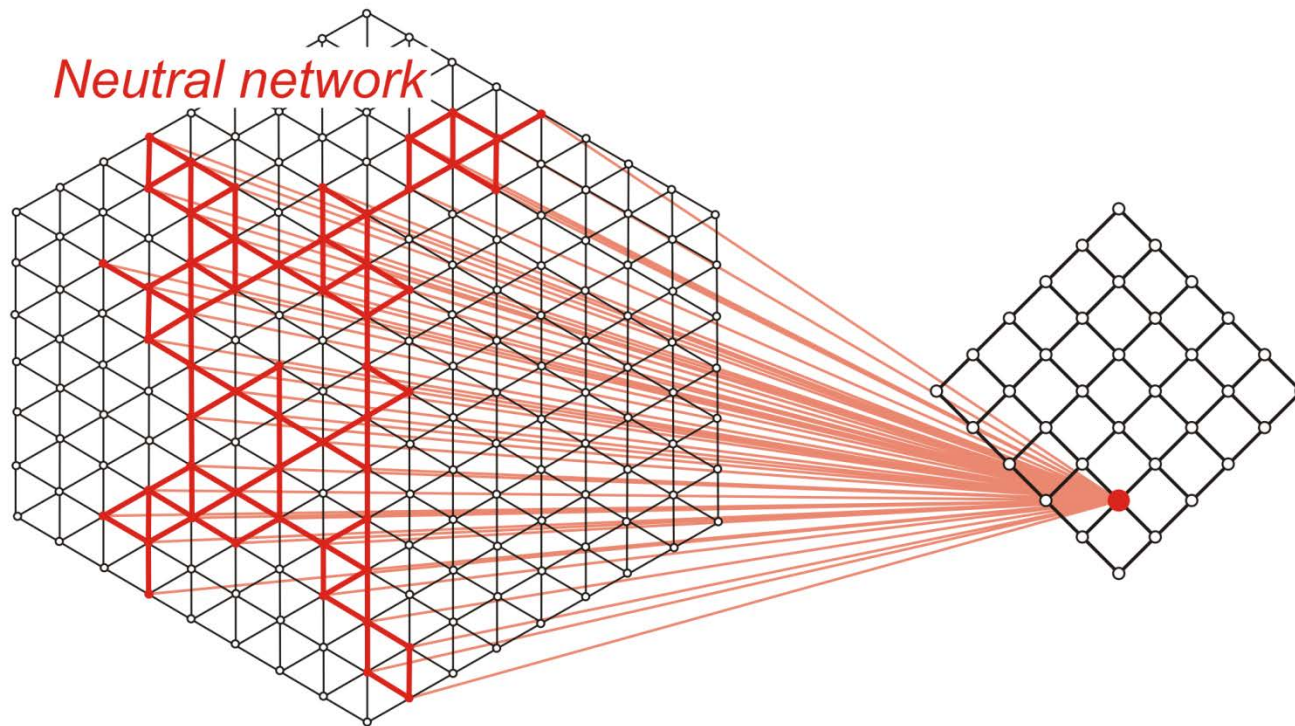




Criterion: Minimum free energy (mfe)

Rules: $_ (_) _ \in \{AU, CG, GC, GU, UA, UG\}$

A symbolic notation of RNA secondary structure that is equivalent to the conventional graphs



Sequence space

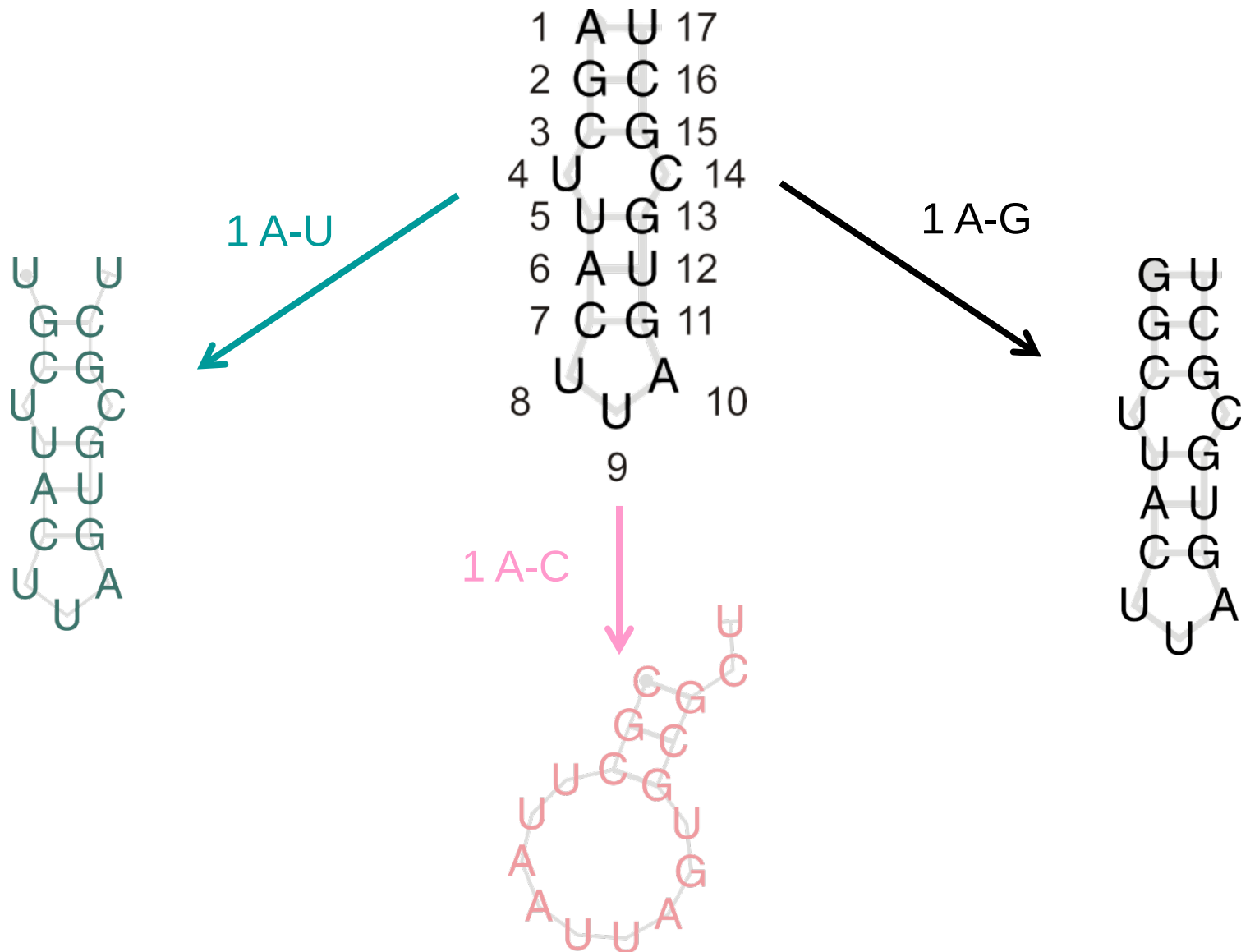
Structure space

many genotypes

$\Rightarrow$

one phenotype

AGCUUAACUUAGUCGCU



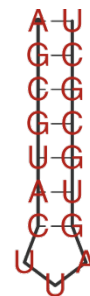
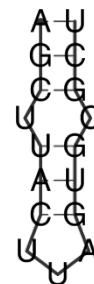
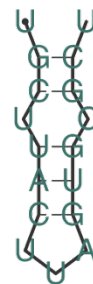
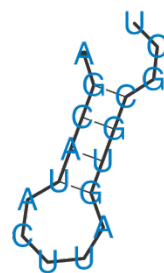
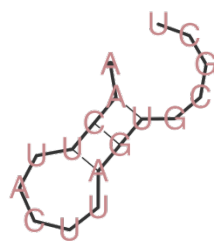
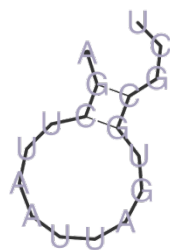
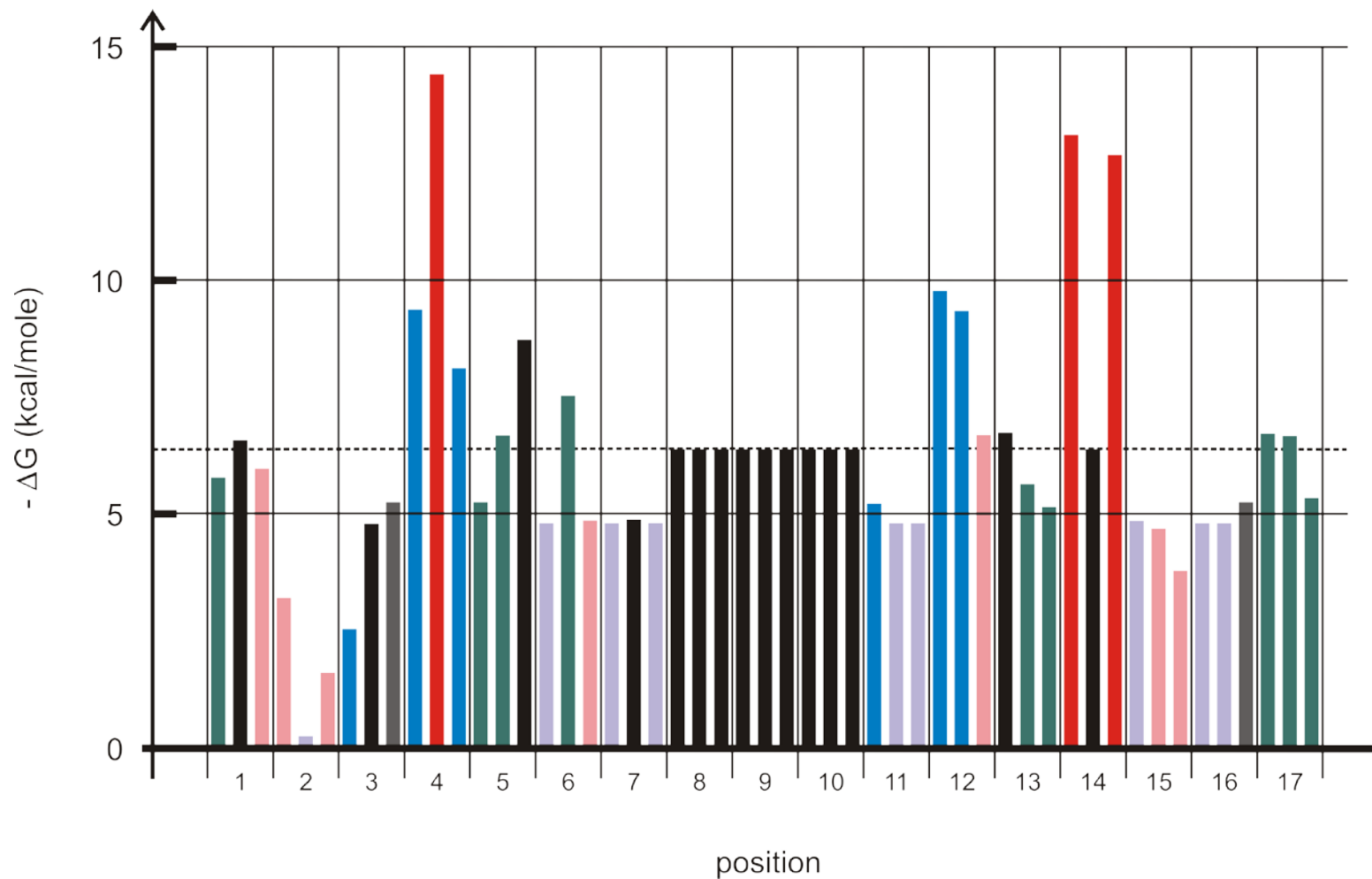
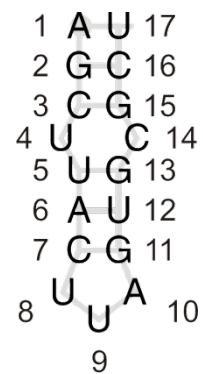
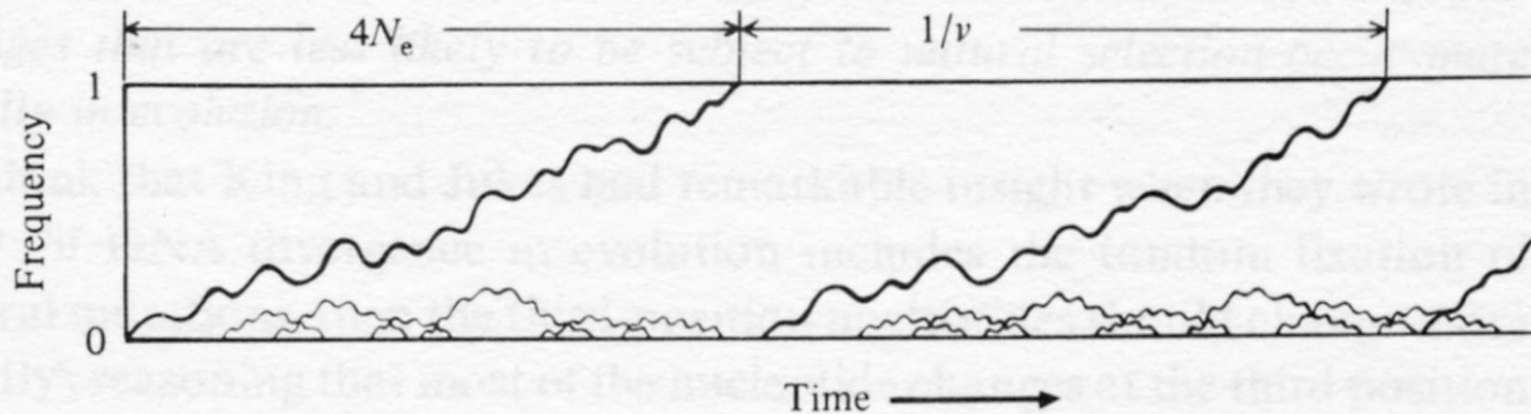


Fig. 3.1. Behavior of mutant genes following their appearance in a finite population. Courses of change in the frequencies of mutants destined to fixation are depicted by thick paths. N_e stands for the effective population size and v is the mutation rate.



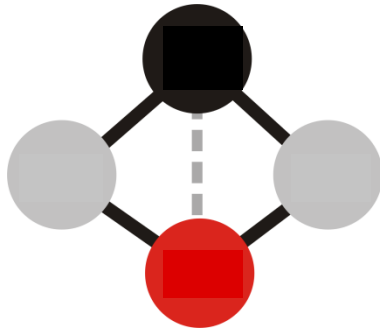
Motoo Kimura

Is the Kimura scenario correct for frequent mutations?



$$d_H = 1$$

$$\lim_{p \rightarrow 0} x_1(p) = x_2(p) = 0.5$$



$$d_H = 2$$

$$\lim_{p \rightarrow 0} x_1(p) = a$$

$$\lim_{p \rightarrow 0} x_2(p) = 1 - a$$

$$d_H \geq 3$$

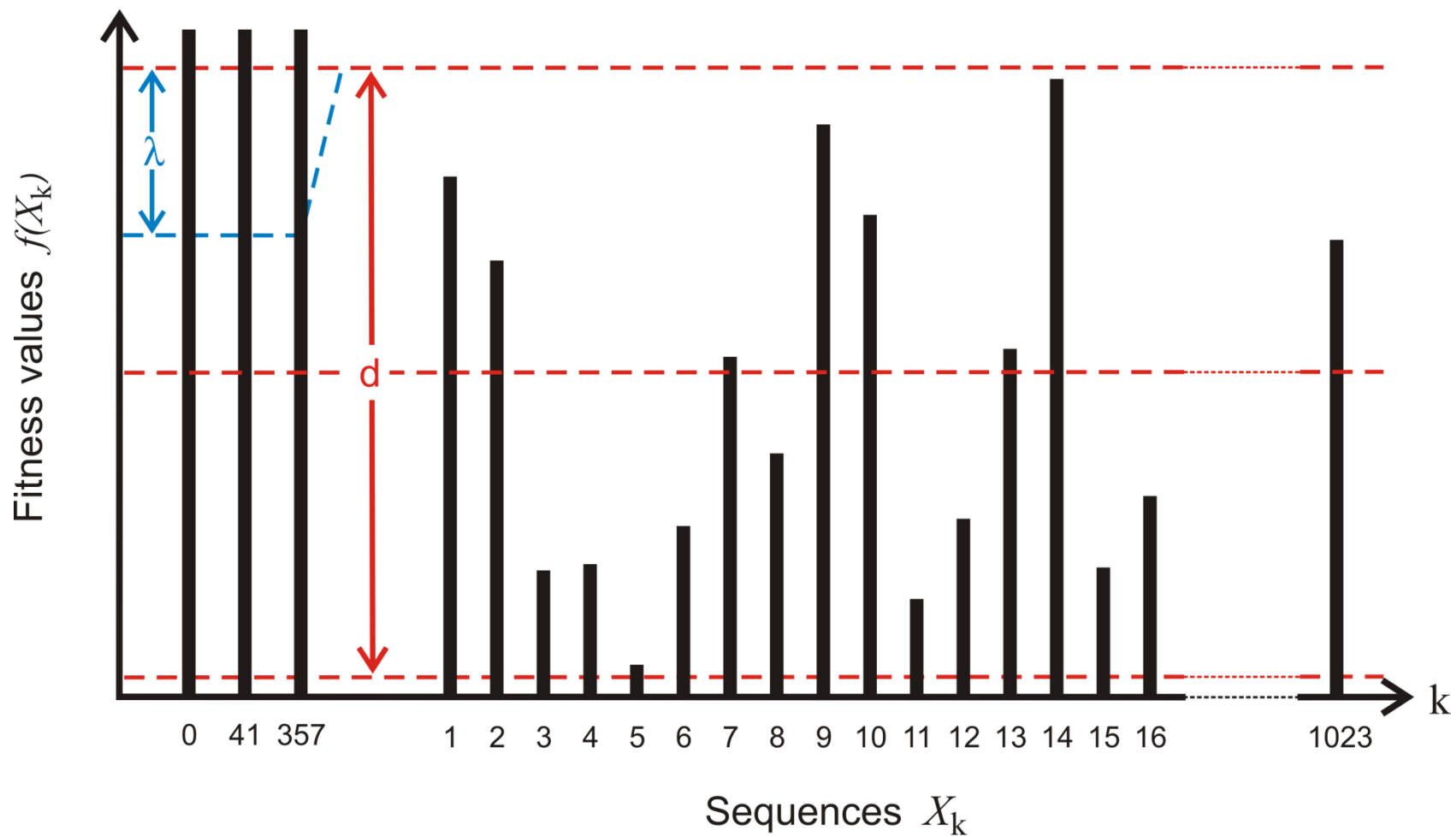
$$\lim_{p \rightarrow 0} x_1(p) = 1, \lim_{p \rightarrow 0} x_2(p) = 0 \quad \text{or}$$

$$\lim_{p \rightarrow 0} x_1(p) = 0, \lim_{p \rightarrow 0} x_2(p) = 1$$

Pairs of neutral sequences in replication networks

Random fixation in the
sense of Motoo Kimura

P. Schuster, J. Swetina. 1988. Bull. Math. Biol. 50:635-650

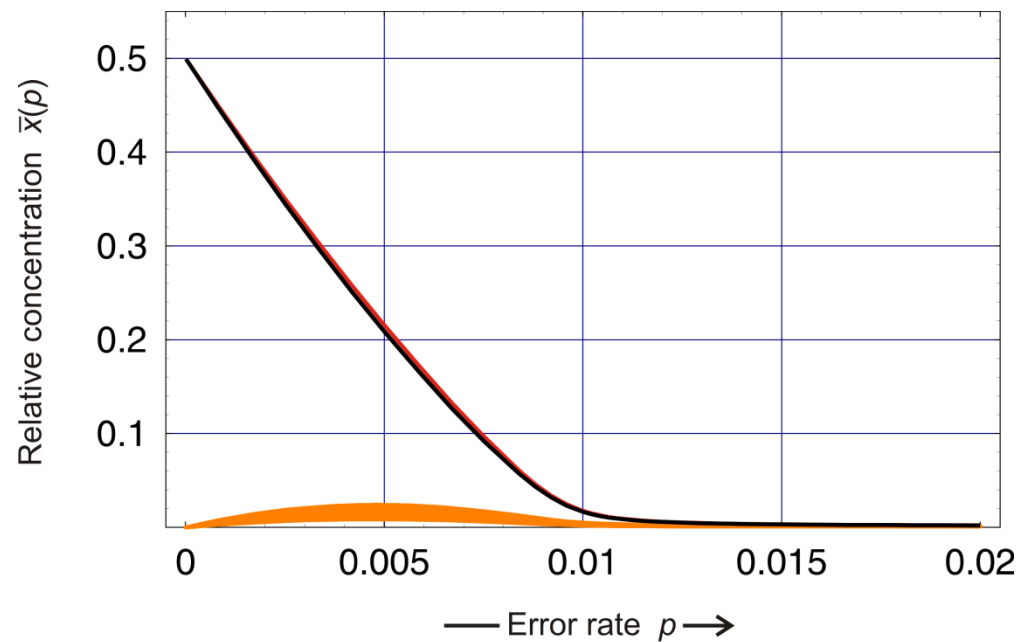


A fitness landscape including neutrality



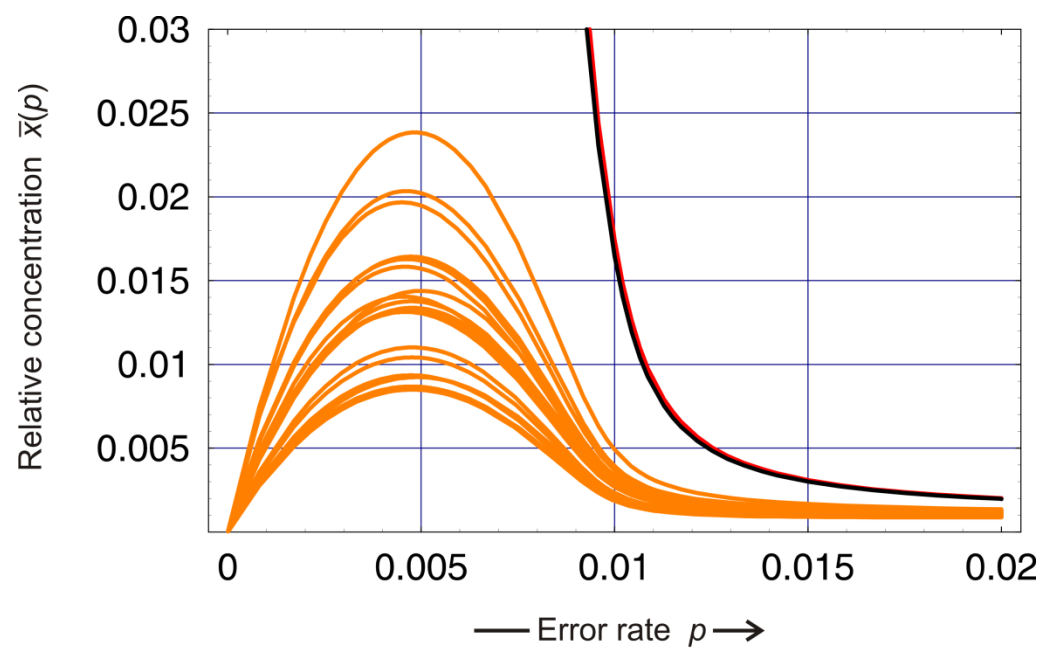
Neutral network

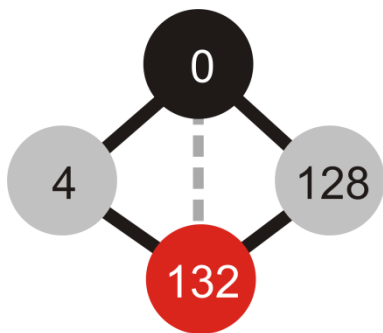
$\lambda = 0.01$, $s = 367$



Neutral network: Individual sequences

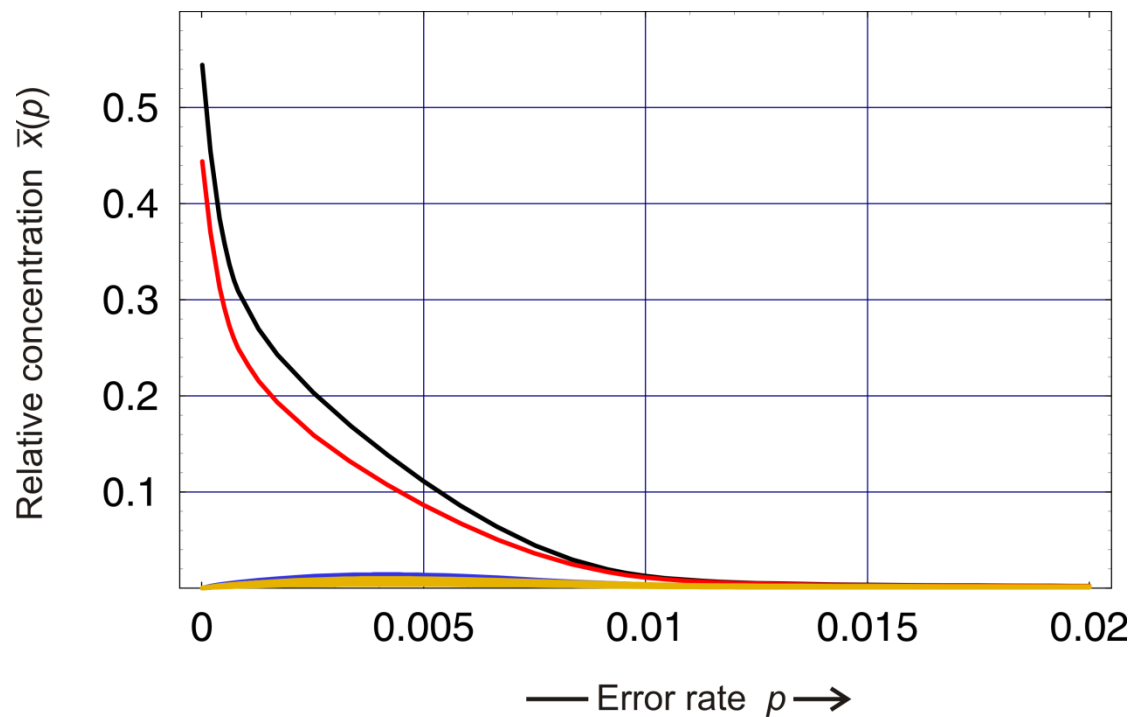
$n = 10$, $\sigma = 1.1$, $d = 1.0$





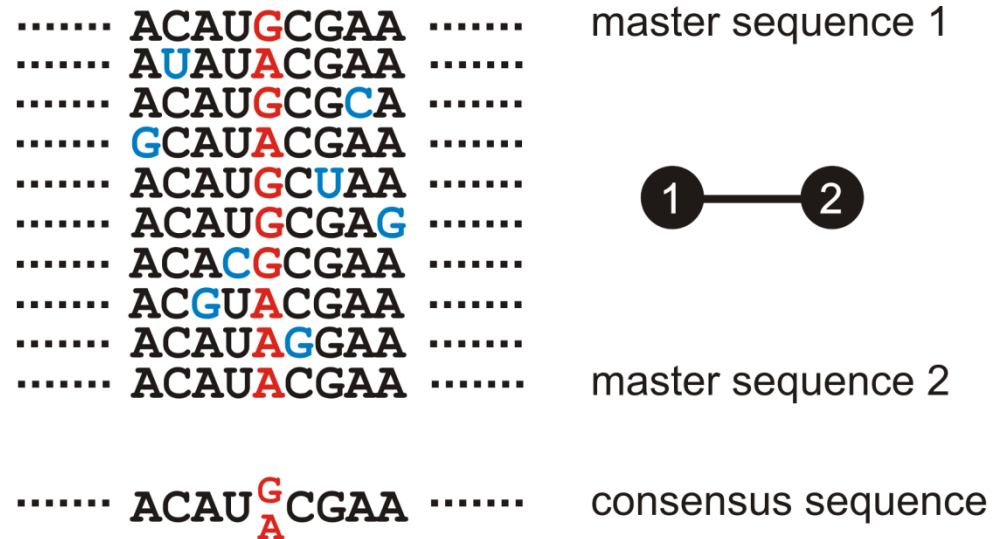
Neutral network

$\lambda = 0.01$, $s = 877$

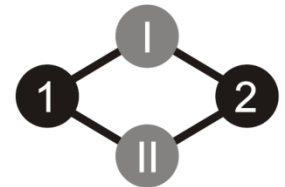
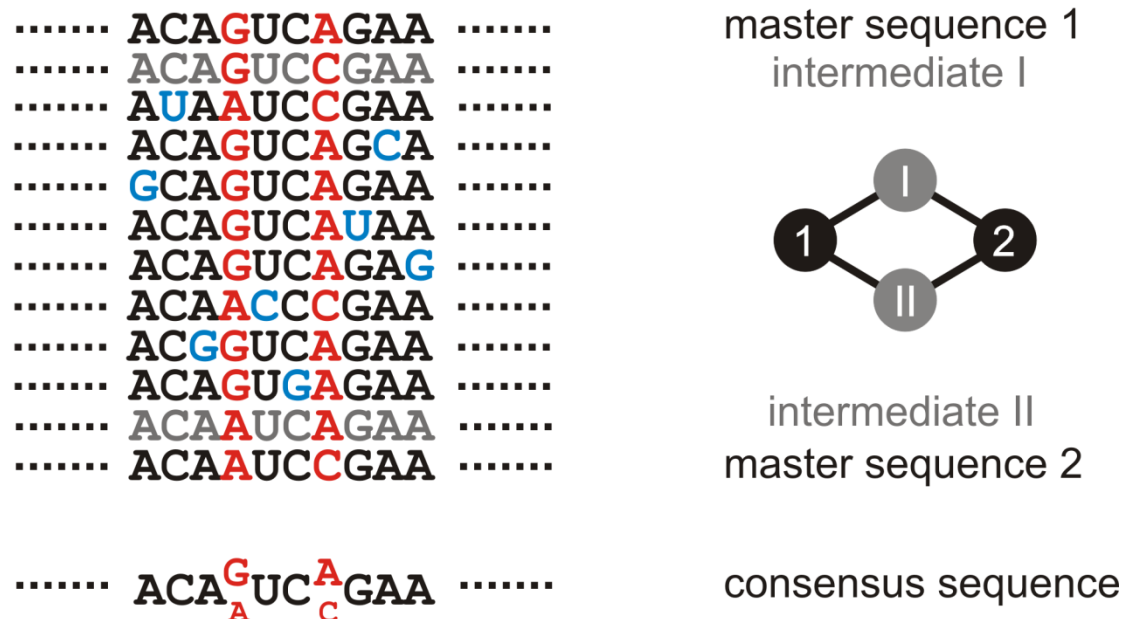


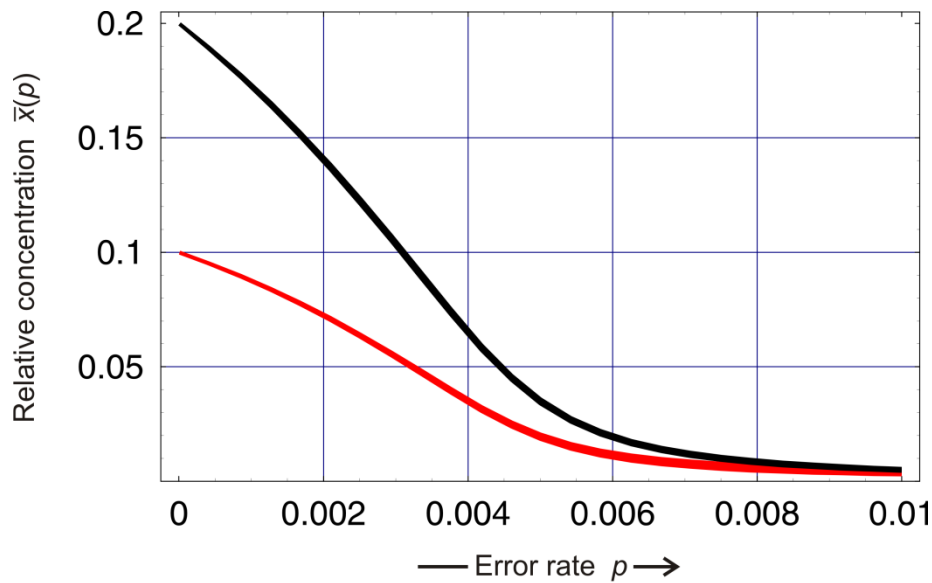
Neutral network: Individual sequences

$n = 10$, $\sigma = 1.1$, $d = 1.0$



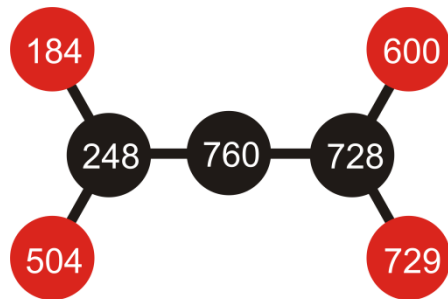
Consensus sequences of a
quasispecies of two strongly
coupled sequences of
Hamming distance
 $d_H(X_i, X_j) = 1$ and 2.





Perturbation matrix W

$$W = \begin{pmatrix} f & 0 & \varepsilon & 0 & 0 & 0 & 0 \\ 0 & f & \varepsilon & 0 & 0 & 0 & 0 \\ \varepsilon & \varepsilon & f & \varepsilon & 0 & 0 & 0 \\ 0 & 0 & \varepsilon & f & \varepsilon & 0 & 0 \\ 0 & 0 & 0 & \varepsilon & f & \varepsilon & \varepsilon \\ 0 & 0 & 0 & 0 & \varepsilon & f & 0 \\ 0 & 0 & 0 & 0 & \varepsilon & 0 & f \end{pmatrix}$$



Neutral network

$$\lambda = 0.10, s = 229$$

Adjacency matrix

Largest eigenvector of W

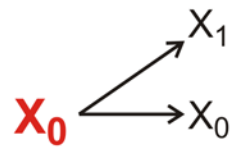
$$\xi_0 = (0.1, 0.1, 0.2, 0.2, 0.2, 0.1, 0.1) .$$

Neutral networks with increasing λ : $\lambda = 0.10, s = 229$

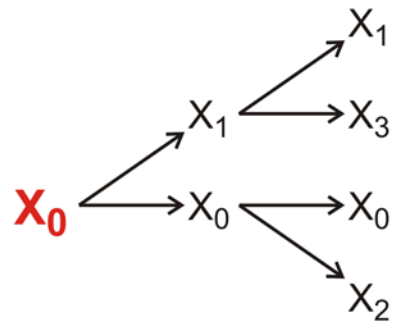
1. Gradualism and punctualism
2. Contingency in evolution experiments
3. Neutrality and its consequences
4. **In silico-evolution of RNA structures**

X_0

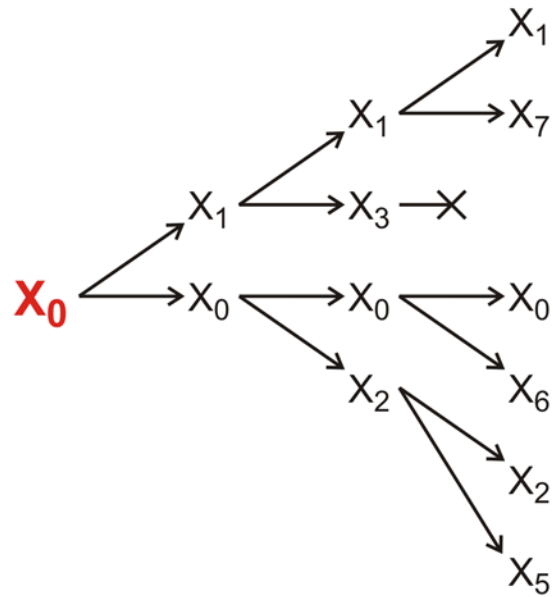
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



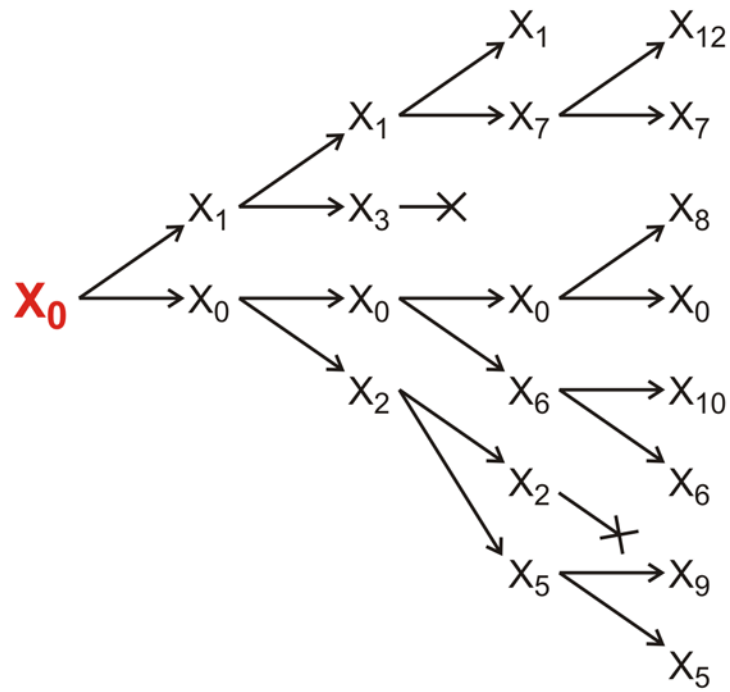
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



Evolution of RNA molecules as a Markow process and its analysis by means of the relay series

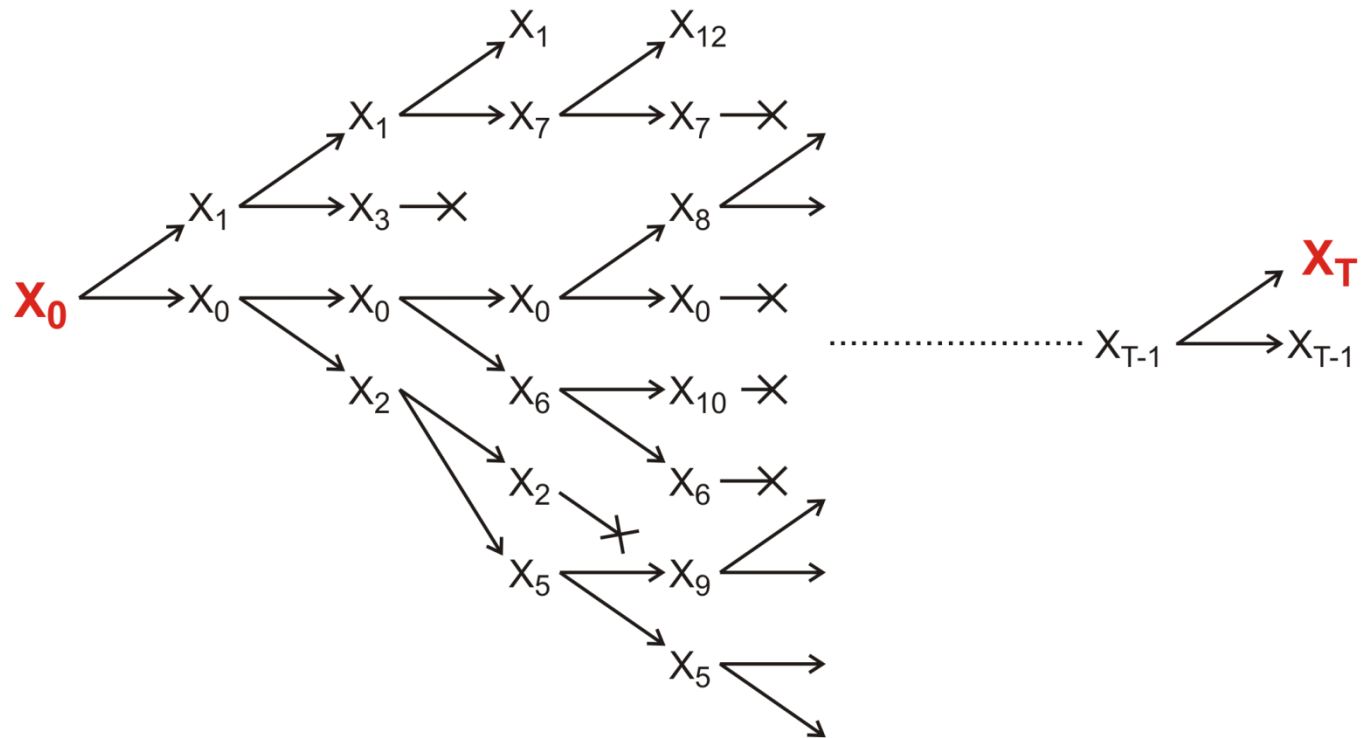


Evolution of RNA molecules as a Markow process and its analysis by means of the relay series

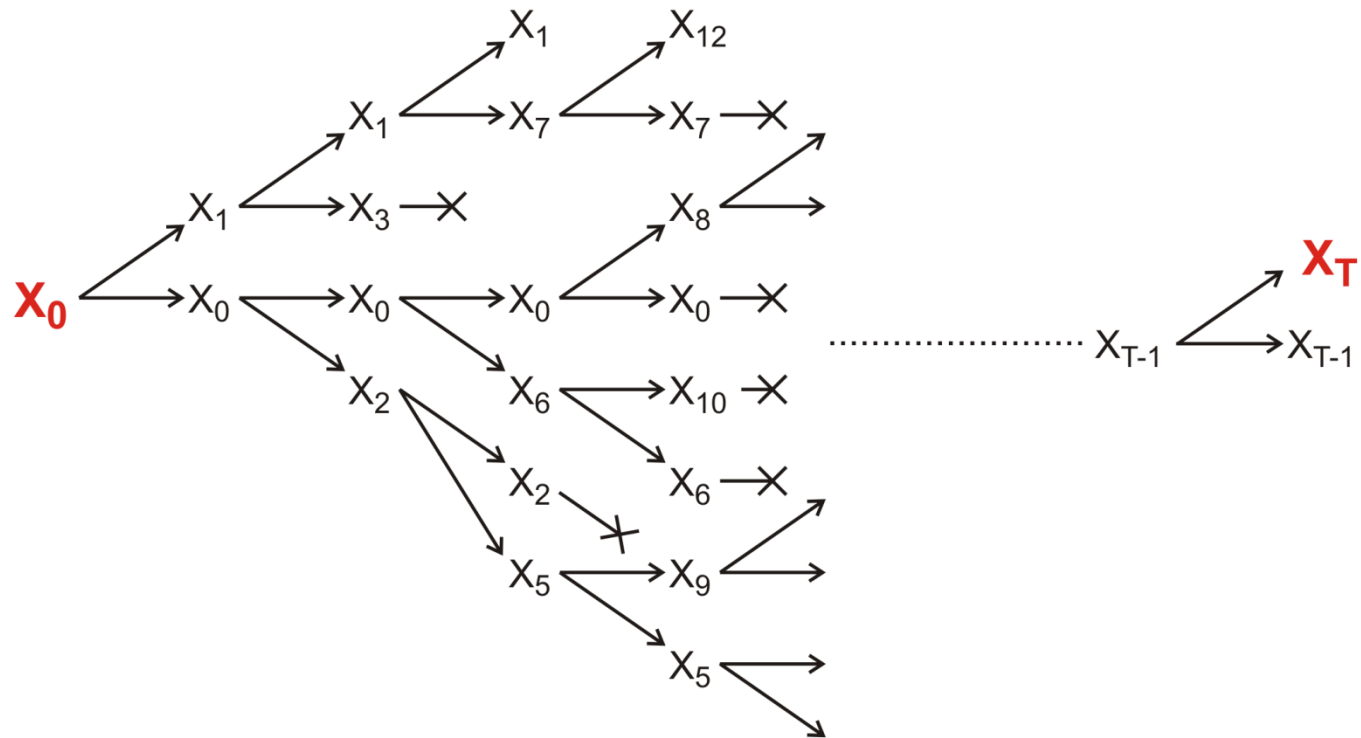


Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



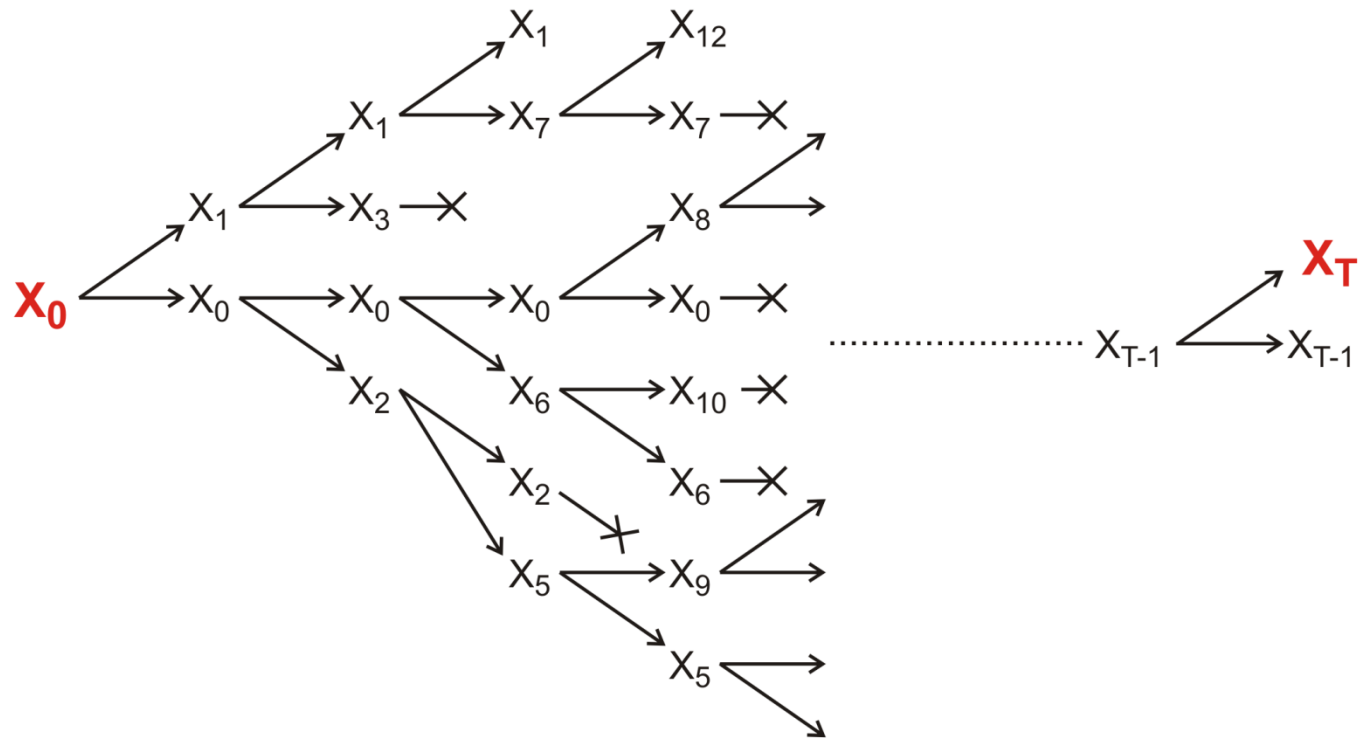


Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



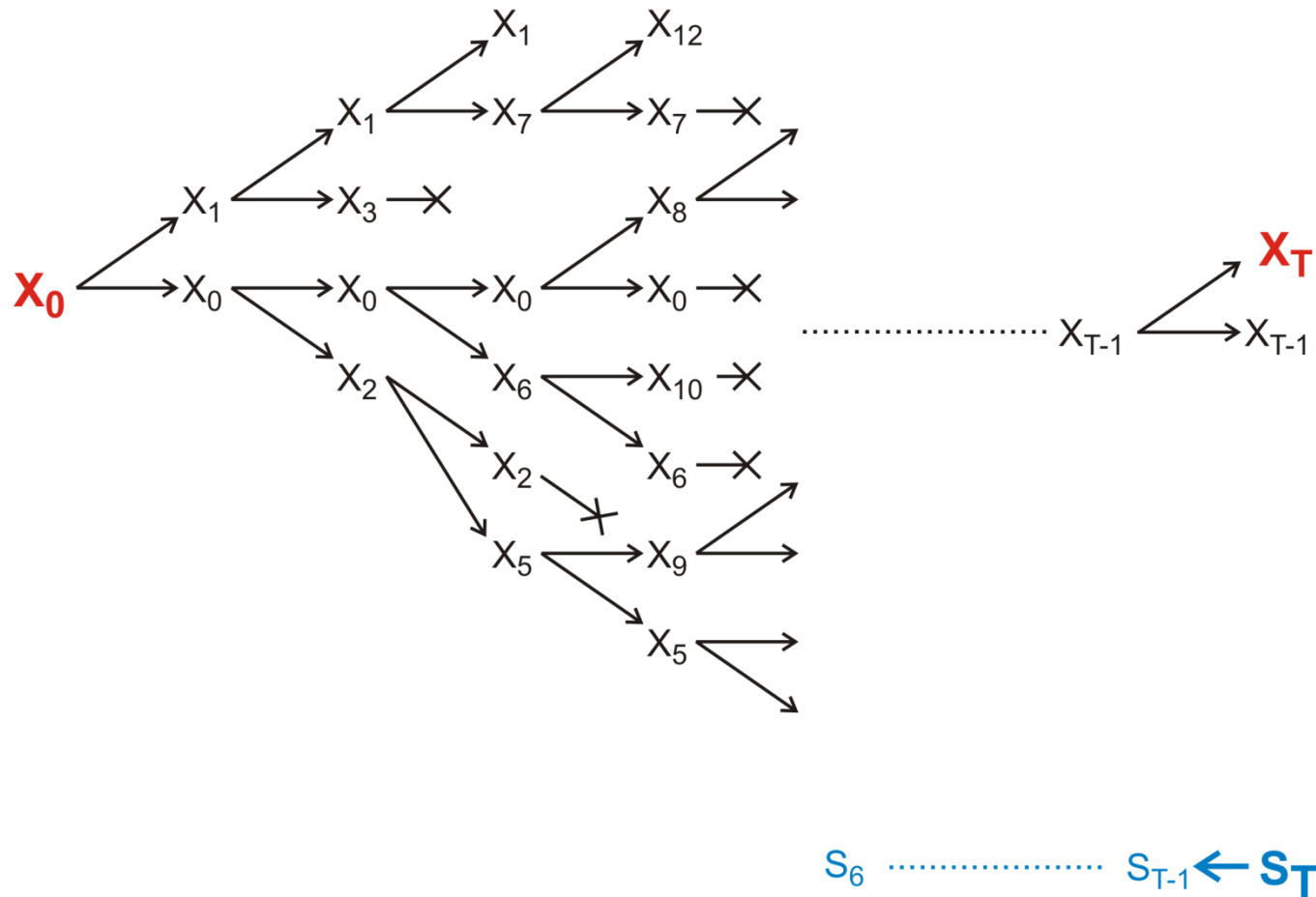
S_T

Evolution of RNA molecules as a Markow process and its analysis by means of the relay series

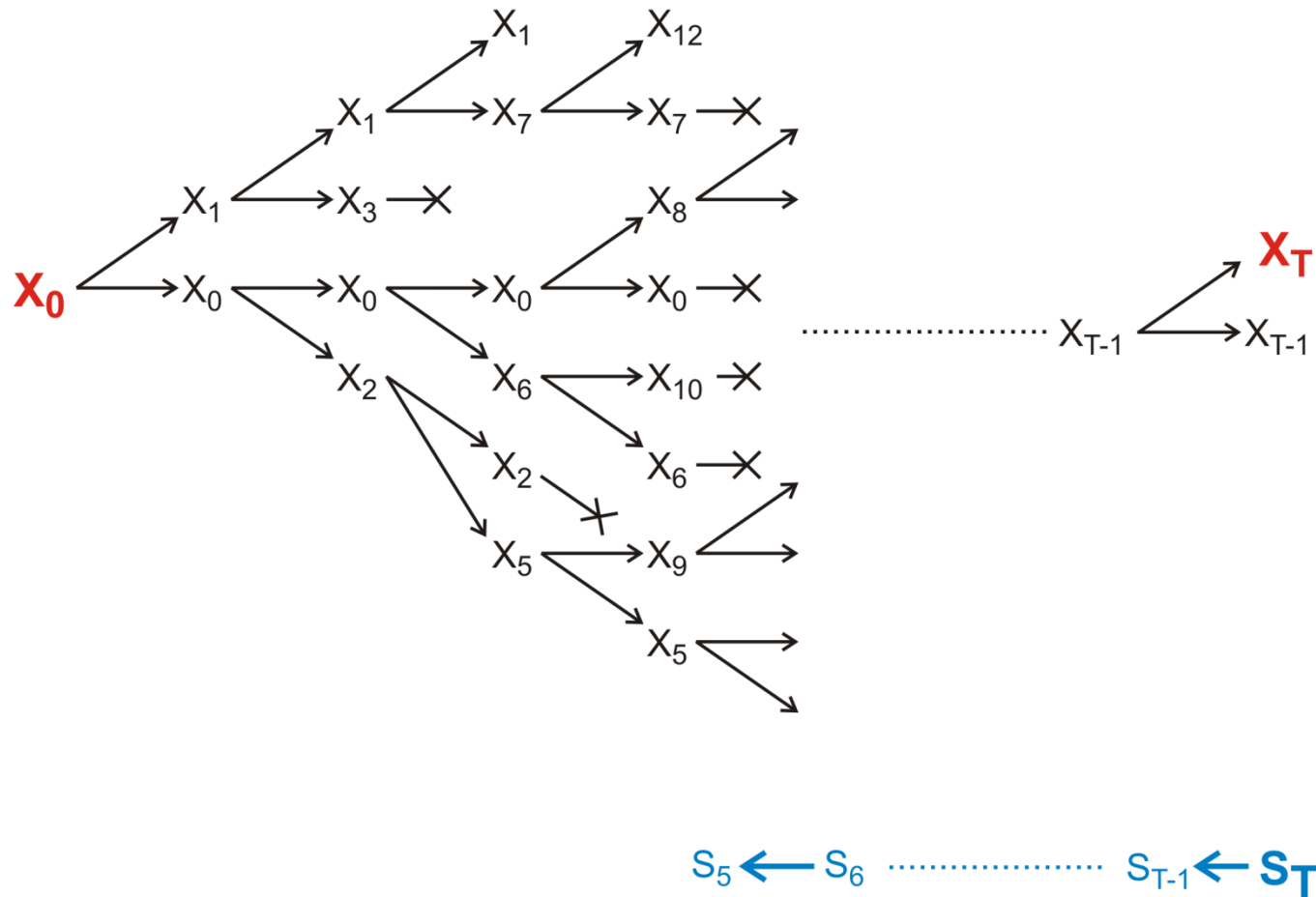


$S_{T-1} \leftarrow S_T$

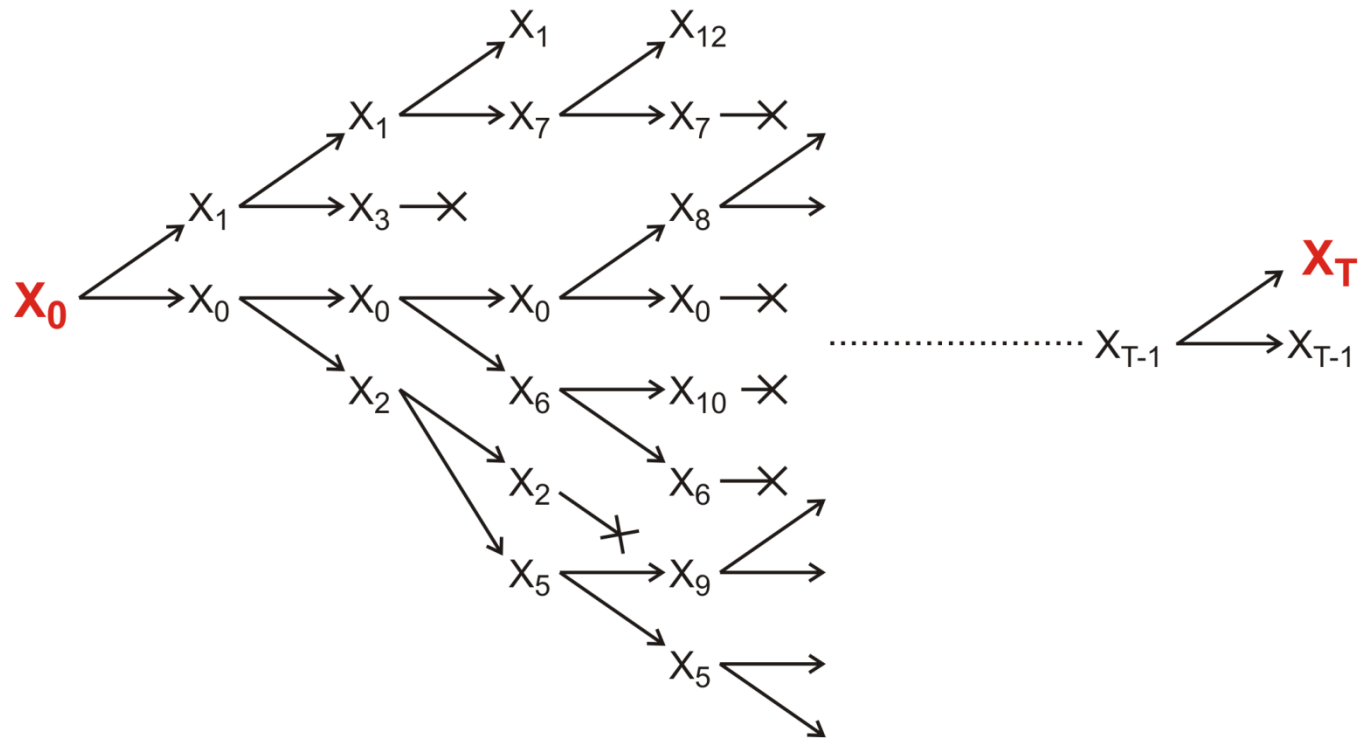
Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



$$S_0 \leftarrow S_1 \leftarrow S_2 \leftarrow S_3 \leftarrow S_4 \leftarrow S_5 \leftarrow S_6 \dots S_{T-1} \leftarrow S_T$$

Evolution of RNA molecules as a Markow process and its analysis by means of the relay series



Evolution of RNA molecules as a Markow process and its analysis by means of the relay series

Computer simulation of RNA optimization

Walter Fontana and Peter Schuster,
Biophysical Chemistry 26:123-147, 1987

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

PHYSICAL REVIEW A

VOLUME 40, NUMBER 6

SEPTEMBER 15, 1989

Physical aspects of evolutionary optimization and adaptation

Walter Fontana, Wolfgang Schnabl, and Peter Schuster*

Institut für Theoretische Chemie der Universität Wien, Währingerstrasse 17, A-1090 Wien, Austria

(Received 2 February 1989; revised manuscript received 5 May 1989)

A model of an objective function based on polynucleotide folding is used to investigate the dynamics of evolutionary adaptation in finite populations. Binary sequences are optimized with respect to their kinetic properties through a stochastic process involving mutation and selection. The objective function consists in a mapping from the set of all binary strings with given length into a set of two-dimensional structures. These structures then encode the kinetic properties, expressed in terms of parameters of reaction probability distributions. The objective function obtained thereby represents a realistic example of a highly "rugged landscape." Ensembles of molecular strings adapting to this landscape are studied by tracing their escape path from local optima and by applying multivariate analysis. Effects of small population numbers in the tail of the sequence distribution are discussed quantitatively. Close upper bounds to the number of distinct values produced by our objective function are given. The distribution of values is explored by means of simulated annealing and reveals a random scatter in the locations of optima in the space of all sequences. The genetic optimization protocol is applied to the "traveling salesman" problem.

Biophysical Chemistry 26 (1987) 123-147
Elsevier

BPC 01133

123

A computer model of evolutionary optimization

Walter Fontana and Peter Schuster

Institut für theoretische Chemie und Strahlenchemie der Universität Wien, Währingerstraße 17, A-1090 Wien, Austria

Accepted 27 February 1987

Molecular evolution; Optimization; Polyribonucleotide folding; Quasi-species; Selective value; Stochastic reaction kinetics

Molecular evolution is viewed as a typical combinatorial optimization problem. We analyse a chemical reaction model which considers RNA replication including correct copying and point mutations together with hydrolytic degradation and the dilution flux of a flow reactor. The corresponding stochastic reaction network is implemented on a computer in order to investigate some basic features of evolutionary optimization dynamics. Characteristic features of real molecular systems are mimicked by folding binary sequences into unknotted two-dimensional structures. Selective values are derived from these molecular 'phenotypes' by an evaluation procedure which assigns numerical values to different elements of the secondary structure. The fitness function obtained thereby contains nontrivial long-range interactions which are typical for real systems. The fitness landscape also reveals quite involved and bizarre local topologies which we consider also representative of polynucleotide replication in actually occurring systems. Optimization operates on an ensemble of sequences via mutation and natural selection. The strategy observed in the simulation experiments is fairly general and resembles closely a heuristic widely applied in operations research areas. Despite the relative smallness of the system - we study 2000 molecules of chain length $\nu = 70$ in a typical simulation experiment - features typical for the evolution of real populations are observed as there are error thresholds for replication, evolutionary steps and quasistationary sequence distributions. The relative importance of selectively neutral or almost neutral variants is discussed quantitatively. Four characteristic ensemble properties, entropy of the distribution, ensemble correlation, mean Hamming distance and diversity of the population, are computed and checked for their sensitivity in recording major optimization events during the simulation.

1. Molecular evolution and optimization

Conventional population genetics treats mutation as an external stochastic source. Moreover, mutations are considered as very rare events. In the absence of genetic recombination populations of haploid organisms are expected to be usually homogeneous. Experimental evidence on viral and bacterial populations is available now and it contradicts these expectations. Mutations appear much more frequently than was originally assumed.

Dedicated to Professor Manfred Eigen on the occasion of his 60th birthday.

Correspondence address: P. Schuster, Institut für theoretische Chemie und Strahlenchemie der Universität Wien, Währingerstraße 17, A-1090 Wien, Austria.

The molecular approach considers error-free replication and mutation as parallel reactions within the same mechanism. Detailed information on the molecular mechanisms of polynucleotide replication provides direct insight into the nature of mutations and their role in evolution. Several classes of mutations are properly distinguished: point mutations, deletions and insertions. Point mutations are of special importance: they represent the most frequent mutations and are easily incorporated into theoretical models of molecular evolution. This does not mean, however, that the other classes of mutations are not important in evolution. To give an example: there is a general belief that insertions leading to gene duplication played a major role in the development of present day enzyme families.

The first theoretical model of molecular evolu-

- random individuals. The primer pair used for genomic DNA amplification is 5'-TCTCCCTGGATTCT-CATTTA-3' (forward) and 5'-TCTTTGTCTCTGT-TCCACC-3' (reverse). Reactions were performed in 25 μ l using 1 unit of Taq DNA polymerase with each primer at 0.4 μ M; 200 μ M each dATP, dTTP, dGTP, and dCTP; and PCR buffer [10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂] in a cycle condition of 94°C for 1 min and then 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s followed by 72°C for 6 min. PCR products were purified (Qiagen), digested with Xmn I, and separated in a 2% agarose gel.
32. A nonsense mutation may affect mRNA stability and result in degradation of the transcript [L. Maquat, *Am. J. Hum. Genet.* **59**, 279 (1996)].
33. Data not shown; a dot blot with poly (A)⁺ RNA from 50 human tissues (The Human RNA Master Blot, 7770-1, Clontech Laboratories) was hybridized with a probe from exons 29 to 47 of MYO15 using the same condition as Northern blot analysis (13).
34. Smith-Magenis syndrome (SMS) is due to deletions of 17p11.2 of various sizes, the smallest of which includes MYO15 and perhaps 20 other genes [6]; K.-S. Chen, L. Potocki, J. R. Lupski, *MDD Res. Rev.* **2**, 122 (1996). MYO15 expression is easily detected in the pituitary gland (data not shown). Haploinsufficiency for MYO15 may explain a portion of the SMS

phenotype such as short stature. Moreover, a few SMS patients have sensorineural hearing loss, possibly because of a point mutation in MYO15 in trans to the SMS 17p11.2 deletion.

35. R. A. Fridell, data not shown.
36. K. B. Avraham *et al.*, *Nature Genet.* **11**, 369 (1995); X.-Z. Liu *et al.*, *ibid.* **17**, 268 (1997); F. Gibson *et al.*, *Nature* **374**, 62 (1995); D. Wei *et al.*, *ibid.*, p. 60.
37. RNA was extracted from cochlea (membranous labyrinth) obtained from human fetuses at 18 to 22 weeks of development in accordance with guidelines established by the Human Research Committee at the Brigham and Women's Hospital. Only samples without evidence of degradation were pooled for poly (A)⁺ selection over oligo(dT) columns. First-strand cDNA was prepared using an Advantage RT-PCR kit (Clontech Laboratories). A portion of the first-strand cDNA (4%) was amplified by PCR with Advantage cDNA polymerase mix (Clontech Laboratories) using human MYO15-specific oligonucleotide primers (forward, 5'-GCATGACCTGCCGGTAAT-GGG-3'; reverse, 5'-CTCAGGGCTTCTGCTAGT-GCTCGGCTGGC-3'). Cycling conditions were 40 s at 94°C; 40 s at 66°C (3 cycles); 60°C (5 cycles); and 55°C (25 cycles); and 1 min at 68°C. PCR products were visualized by ethidium bromide staining after fractionation in a 1% agarose gel. A 688-bp PCR

product is expected from amplification of the human MYO15 cDNA. Amplification of human genomic DNA with this primer pair would result in a 2903-bp fragment.

38. We are grateful to the people of Bengala, Bali, and the two families from India. We thank J. R. Lupski and K.-S. Chen for providing the human chromosome 17 cosmid library. For technical and computational assistance, we thank N. Dietrich, M. Ferguson, A. Gupta, E. Sorbello, F. Torzadeh, C. Varner, M. Walker, G. Bouffard, and S. Beckstrom-Sternberg (National Institutes of Health Intramural Sequencing Center). We thank J. T. Hinnant, I. N. Arhya, and S. Winata for assistance in Bali, and T. Barber, S. Sullivan, E. Green, D. Drayna, and J. Battey for helpful comments on this manuscript. Supported by the National Institute on Deafness and Other Communication Disorders (NIDCD) (Z01 DC 00035-01 and Z01 DC 00038-01 to T.B.F. and E.R.W. and R01 DC 03402 to C.C.M.), the National Institute of Child Health and Human Development (R01 HD30428 to S.A.C.), and a National Science Foundation Graduate Research Fellowship to F.J.P. This paper is dedicated to J. B. Snow Jr. on his retirement as the Director of the NIDCD.

9 March 1998; accepted 17 April 1998

Continuity in Evolution: On the Nature of Transitions

Walter Fontana and Peter Schuster

To distinguish continuous from discontinuous evolutionary change, a relation of nearness between phenotypes is needed. Such a relation is based on the probability of one phenotype being accessible from another through changes in the genotype. This nearness relation is exemplified by calculating the shape neighborhood of a transfer RNA secondary structure and provides a characterization of discontinuous shape transformations in RNA. The simulation of replicating and mutating RNA populations under selection shows that sudden adaptive progress coincides mostly, but not always, with discontinuous shape transformations. The nature of these transformations illuminates the key role of neutral genetic drift in their realization.

A much-debated issue in evolutionary biology concerns the extent to which the history of life has proceeded gradually or has been punctuated by discontinuous transitions at the level of phenotypes (1). Our goal is to make the notion of a discontinuous transition more precise and to understand how it arises in a model of evolutionary adaptation.

We focus on the narrow domain of RNA secondary structure, which is currently the simplest computationally tractable, yet realistic phenotype (2). This choice enables the definition and exploration of concepts that may prove useful in a wider context. RNA secondary structures represent a coarse level of analysis compared with the three-dimensional structure at atomic resolution. Yet, secondary structures are empir-

ically well defined and obtain their biophysical and biochemical importance from being a scaffold for the tertiary structure. For the sake of brevity, we shall refer to secondary structures as "shapes." RNA combines in a single molecule both genotype (replicable sequence) and phenotype (selectable shape), making it ideally suited for in vitro evolution experiments (3, 4).

To generate evolutionary histories, we used a stochastic continuous time model of an RNA population replicating and mutating in a capacity-constrained flow reactor under selection (5, 6). In the laboratory, a goal might be to find an RNA aptamer binding specifically to a molecule (4). Although in the experiment the evolutionary end product was unknown, we thought of its shape as being specified implicitly by the imposed selection criterion. Because our intent is to study evolutionary histories rather than end products, we defined a target shape in advance and assumed the replication rate of a sequence to be a function of

the similarity between its shape and the target. An actual situation may involve more than one best shape, but this does not affect our conclusions.

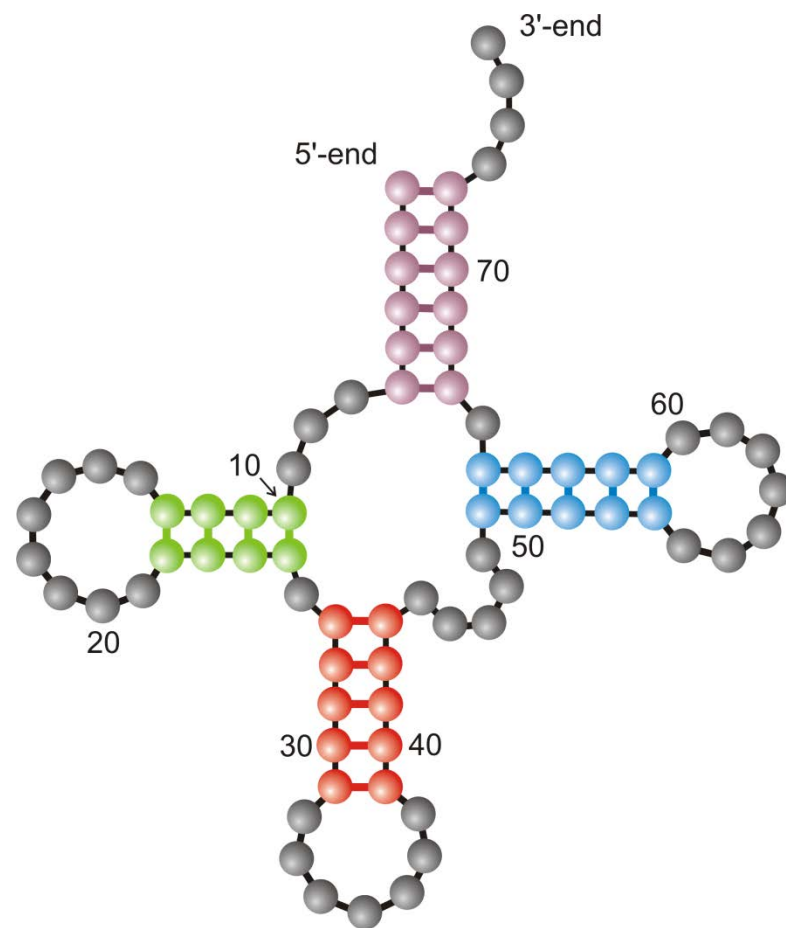
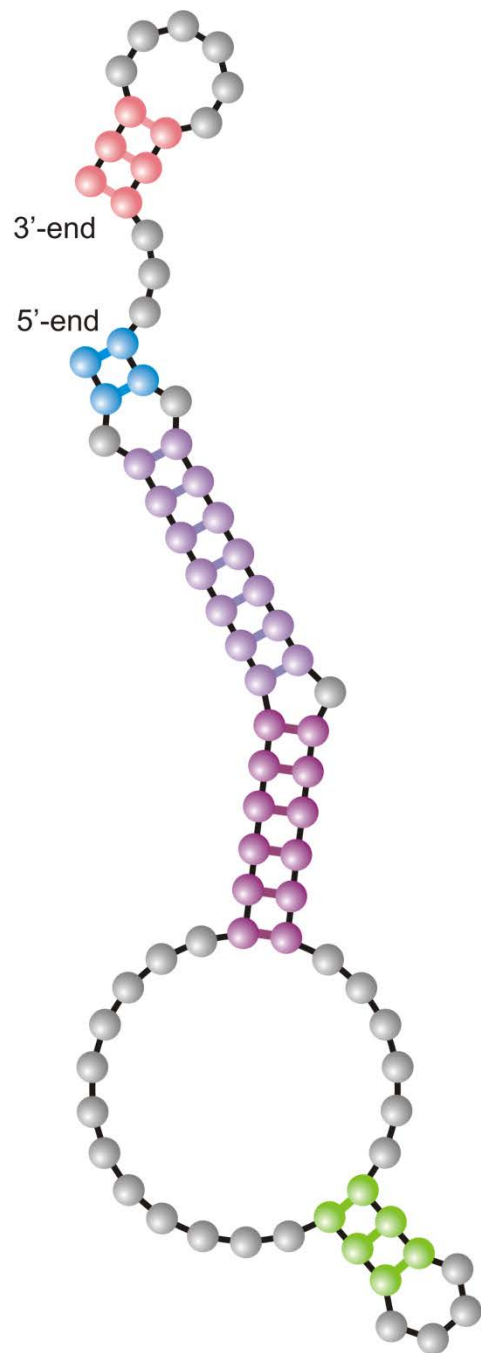
An instance representing in its qualitative features all the simulations we performed is shown in Fig. 1A. Starting with identical sequences folding into a random shape, the simulation was stopped when the population became dominated by the target, here a canonical tRNA shape. The black curve traces the average distance to the target (inversely related to fitness) in the population against time. Aside from a short initial phase, the entire history is dominated by steps, that is, flat periods of no apparent adaptive progress, interrupted by sudden approaches toward the target structure (7). However, the dominant shapes in the population not only change at these marked events but undergo several fitness-neutral transformations during the periods of no apparent progress. Although discontinuities in the fitness trace are evident, it is entirely unclear when and on the basis of what the series of successive phenotypes itself can be called continuous or discontinuous.

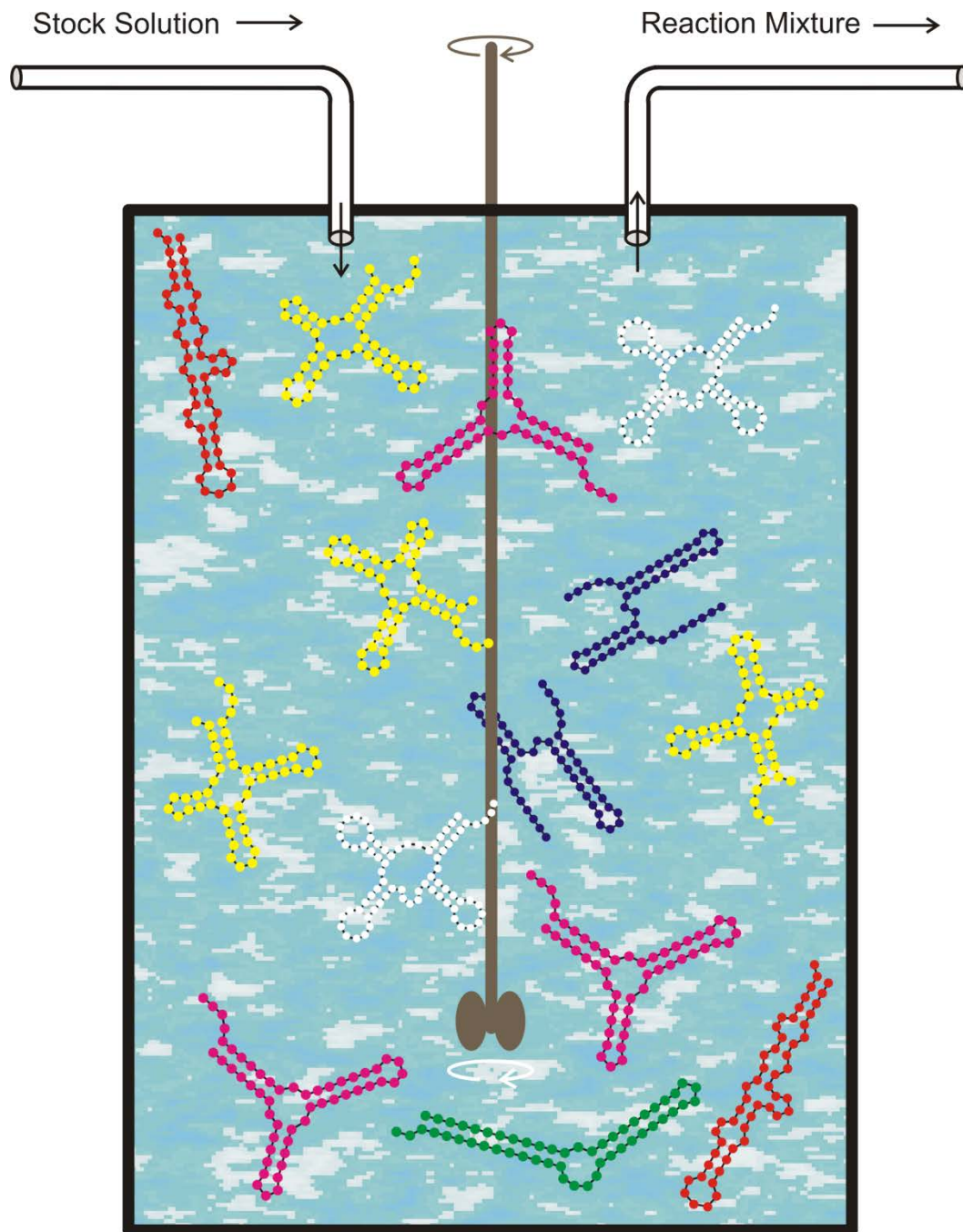
A set of entities is organized into a (topological) space by assigning to each entity a system of neighborhoods. In the present case, there are two kinds of entities: sequences and shapes, which are related by a thermodynamic folding procedure. The set of possible sequences (of fixed length) is naturally organized into a space because point mutations induce a canonical neighborhood. The neighborhood of a sequence consists of all its one-error mutants. The problem is how to organize the set of possible shapes into a space. The issue arises because, in contrast to sequences, there are

Evolution *in silico*

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

Institut für Theoretische Chemie, Universität Wien, Währingerstrasse 17, A-1090 Wien, Austria, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA, and International Institute for Applied Systems Analysis (IIASA), A-2361 Laxenburg, Austria.





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 flux:

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

Mutation rate:

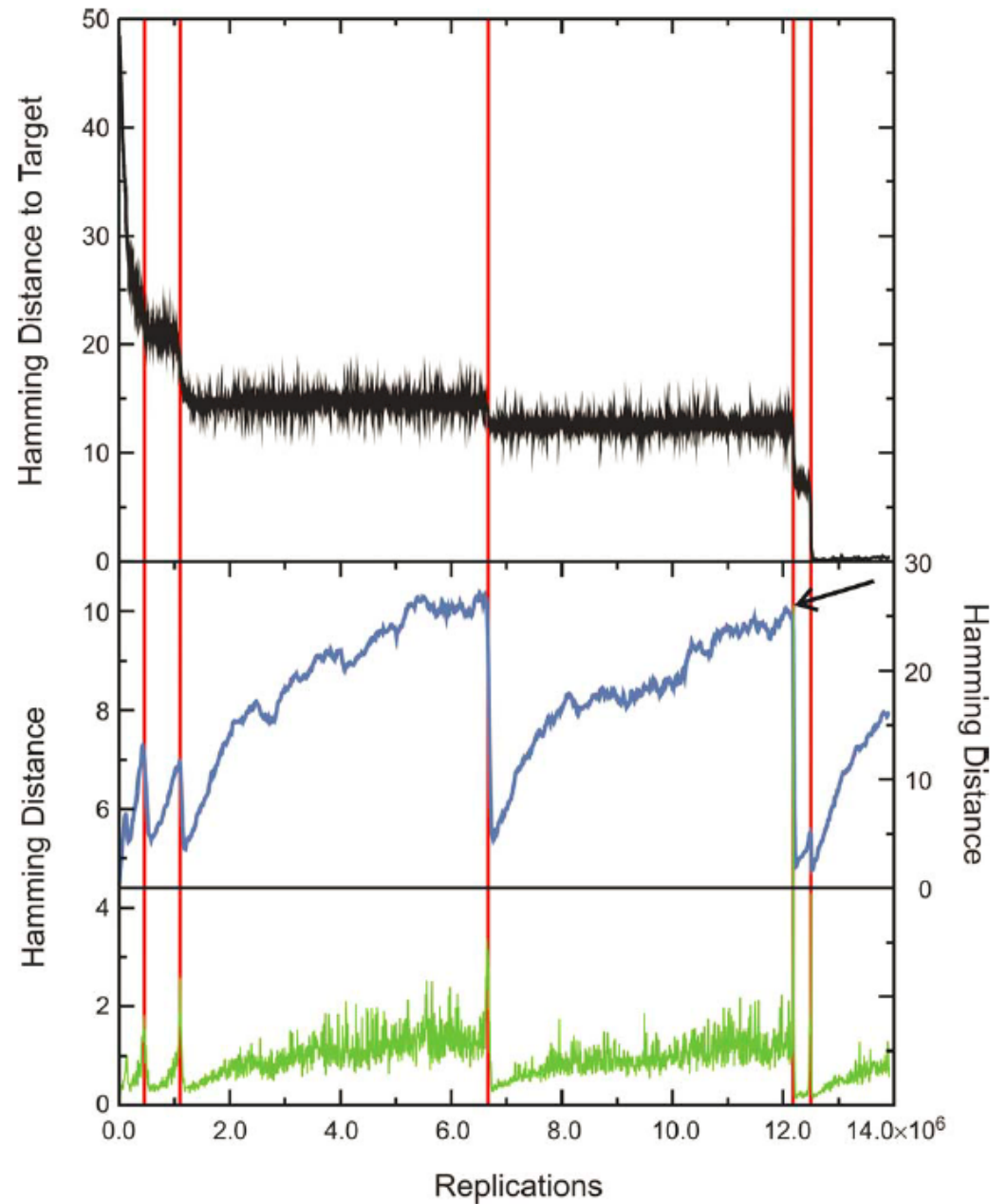
$$p = 0.001 / \text{Nucleotide} \times \text{Replication}$$

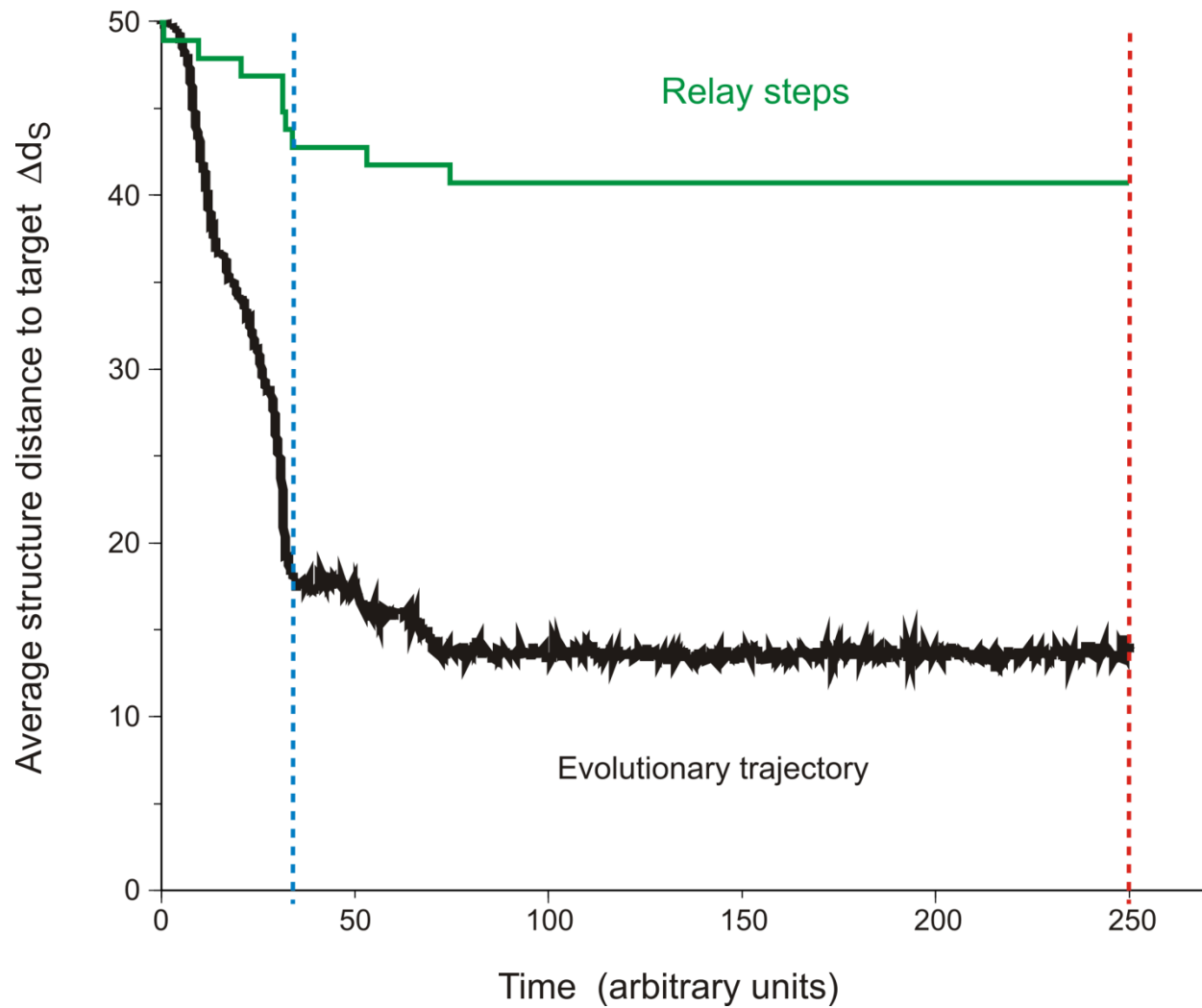
The flow reactor as a device for studying the evolution of molecules
in vitro and *in silico*.

Evolutionary trajectory

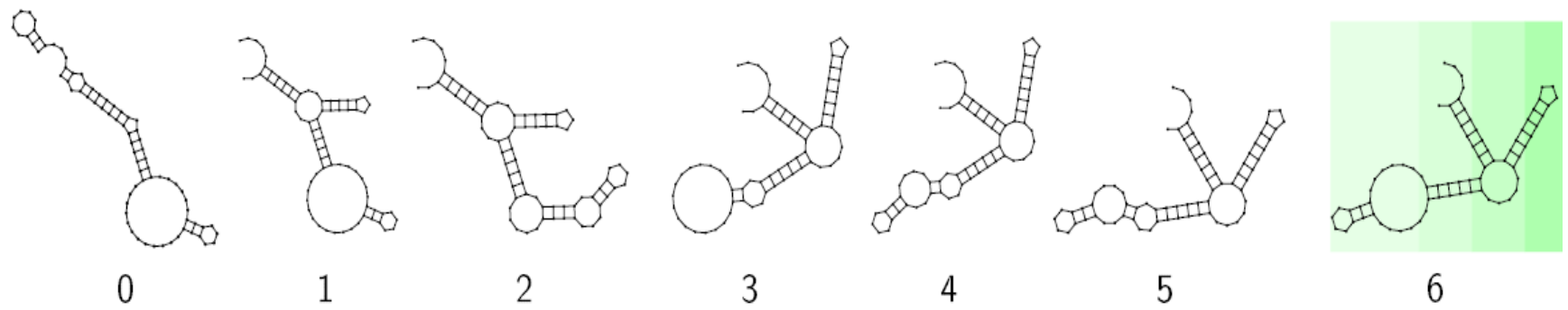
Spreading of the population
on neutral networks

Drift of the population center
in sequence space





First adaptive phase in RNA structure optimization



First adaptive phase in RNA structure optimization: RNA structures

```

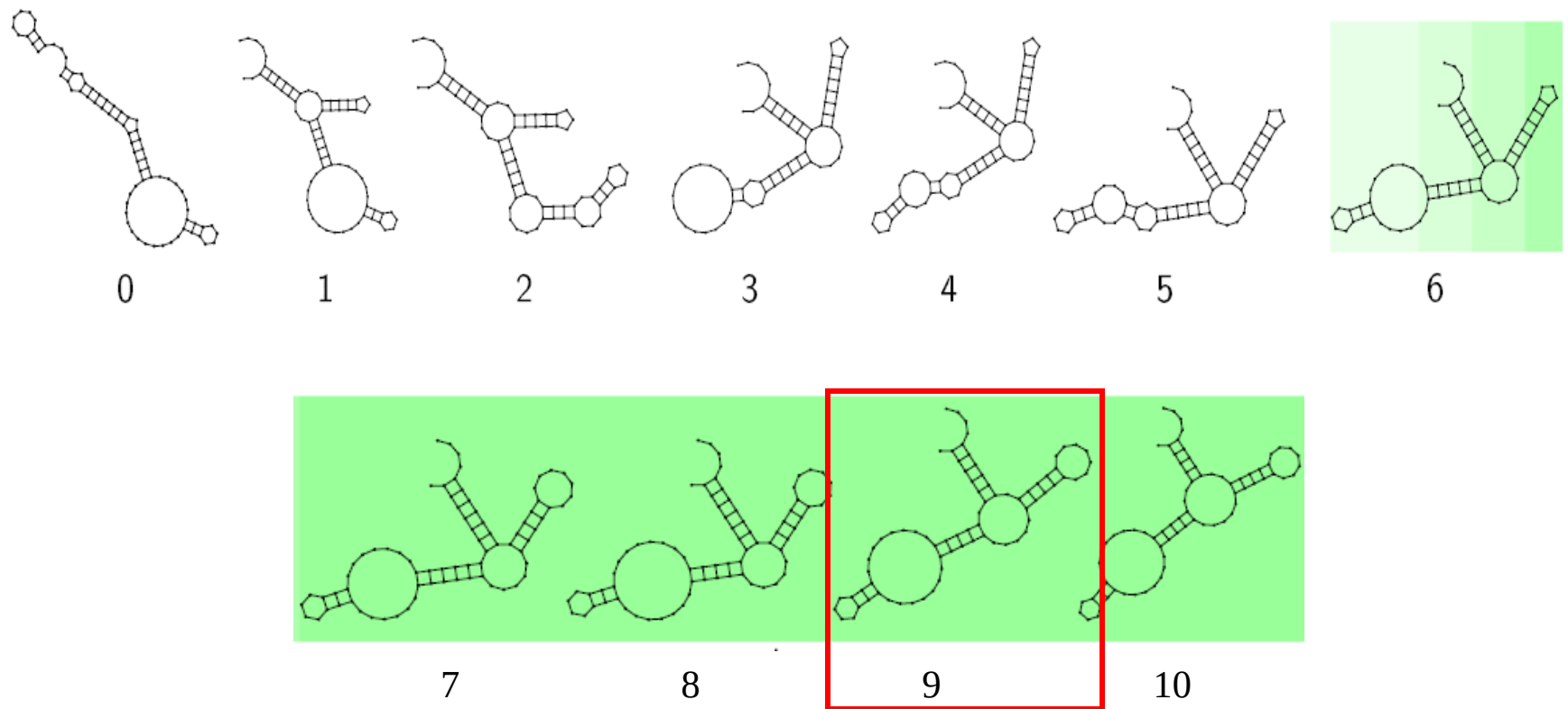
entry  GUUAUGGGCGAUGAGGAGUAGUGUUUAAACCAAUCGGUCAAUGAUCUCGUGUGCCCAUUGCAUAUCCCGUACAGGA
0      ((.((((((((((((((((.....(((.....))).....)))))).)))))).)...(((.....)))
exit   GUUAUGGGCGAUGAGGAGUAGUGUUUAAACCAAUCGGUCAAUGAUCUCGUGUGCCCAUUGCAUAUCCCGUACAGGA
entry  GUUAUGGGCGAUGAGGAGUAGUGUUUAAACCAAUCGGUCAAUGAUCUCGUGUGCCCAUUGCAUAUCCCGUACAGGA
1      ..((((((..((((((((.....(((.....))).....))))))((((((...)))))).))))))....
exit   GUUAUGGGCGAUGAGGAGUAGUGUUUAAACCAAUCGGUCAAUGAUCUCGUGUGCGCAUUGCAUAUCCCGUACAGGA
entry  GUUAUGGGCGAUGAGGAGUAGUGUUUAAACCAAUCGGUCAAUCAUCUCGUGUGCGCAUUGCAUAUCCCGUACAGGA
2      ..((((((..((((((((.....(((.....))).....))))))((((((...)))))).))))))....
exit   GUUAUGGGCGAUGAGGAGUAGUGUUUAAACCAAUCGGUCAACGAUCUCGUGUGCGCAUUGCAUAUCCCGUACAGGA
entry  GUUAUGGGCGUUGAGGAGUAGUGUUUAAACCAAUCGGUCAACGAUCUCGUGUGCGCAUUGCAUAUCCCGUACAGGA
3      ..((((((((((((((..(.....)))))).))))))....((((((...)))))))))....
exit   GUUAUGGGCGUUGAGGAGUAGUGUUUAAACCAAUCGGUCAACGAUCUCGUGUGCGCAUUGCAUAUCCCGUACAGGA
entry  GUUAUGGGCGUUGAGGAGUAGGGUUUAAACCAAUCGGUCAACGAUCUCGUGUGCGCAUUGCAUAUCCCGUACAGGA
4      ..((((((((((((((..(.....(((.....)))..))..))))))....((((((...)))))))))....
exit   GUUAUGGGCGUUGAGGAGUAGGGUUUAAACCAAUCGGUCAACGAUCUCGUGUGCGCAUUGCAUAUCCCGUACAGGA
entry  GGUAUGGGCGUUGAGGAGUAGGGUUUAAACCAAUCGGUCAACGAUCUCGUGUGCGCAUUGCAUAUCCCGUACAGGA
5      .((((((((((((((((..(.....(((.....)))..))..))))))....((((((...)))))))))....
exit   GGUAUGGGCGUUGAGGAGUAGGGUUUAAACCAAUCGGUCAACGAUCUCGUGUGCGCAUUGCAUAUCCCGUACAGGA
entry  GGUAUGGGCGUUGAGUAGUAGGGUUUAAACCAAUCGGUCAACGAUCUCGUGUGCGCAUUGCAUAUCCCGUACAGGA
6      .((((((((((((((((.....(((.....))).....))))))....((((((...)))))))))....
exit   GGUAUGGGCGUUGAGUAGUAGGGUUUAAACCAAUCGGUCAACGAUCUCGUGUGCGCAUUCAUAUCCCGUACAGGA
entry  GGUAUGGGCGUUGAGUAGUAGGGUUUAAACCAAUCGGUCAACGAUCUCGUGUGCGCAUUCAUAUCCCGUACAGGA
7      .((((((((((((((((.....(((.....))).....))))))....((((((...)))))))))....
exit   GGUAUGGGCGUUGAAUAGUAGGGUUUAAACCAAUCGGUCAACGAUCUCGUGUGCGCAUUUCAUAUCCCGUACAGA
entry  GGUAUGGGCGUUGAAUAGUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUCCCGUACAGAA

```

Transition inducing point mutations
change the molecular structure

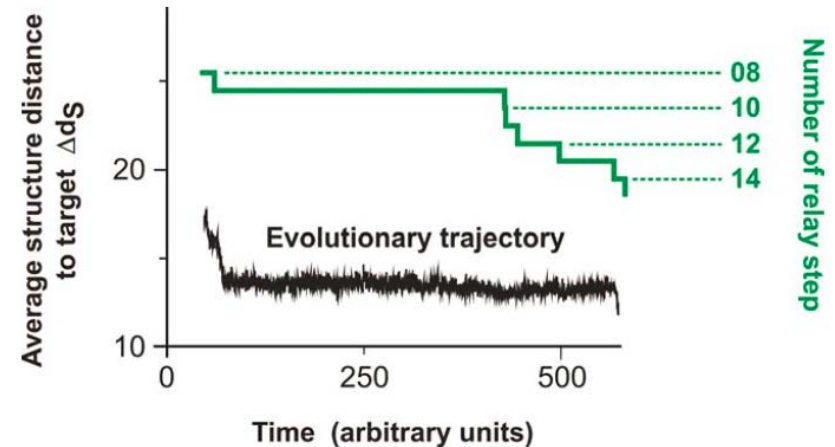
Neutral point mutations leave the
molecular structure unchanged

First adaptive phase in RNA structure optimization: RNA sequences



First adaptive and quasistationary phase in RNA structure optimization

28 neutral point mutations during a long quasi-stationary epoch

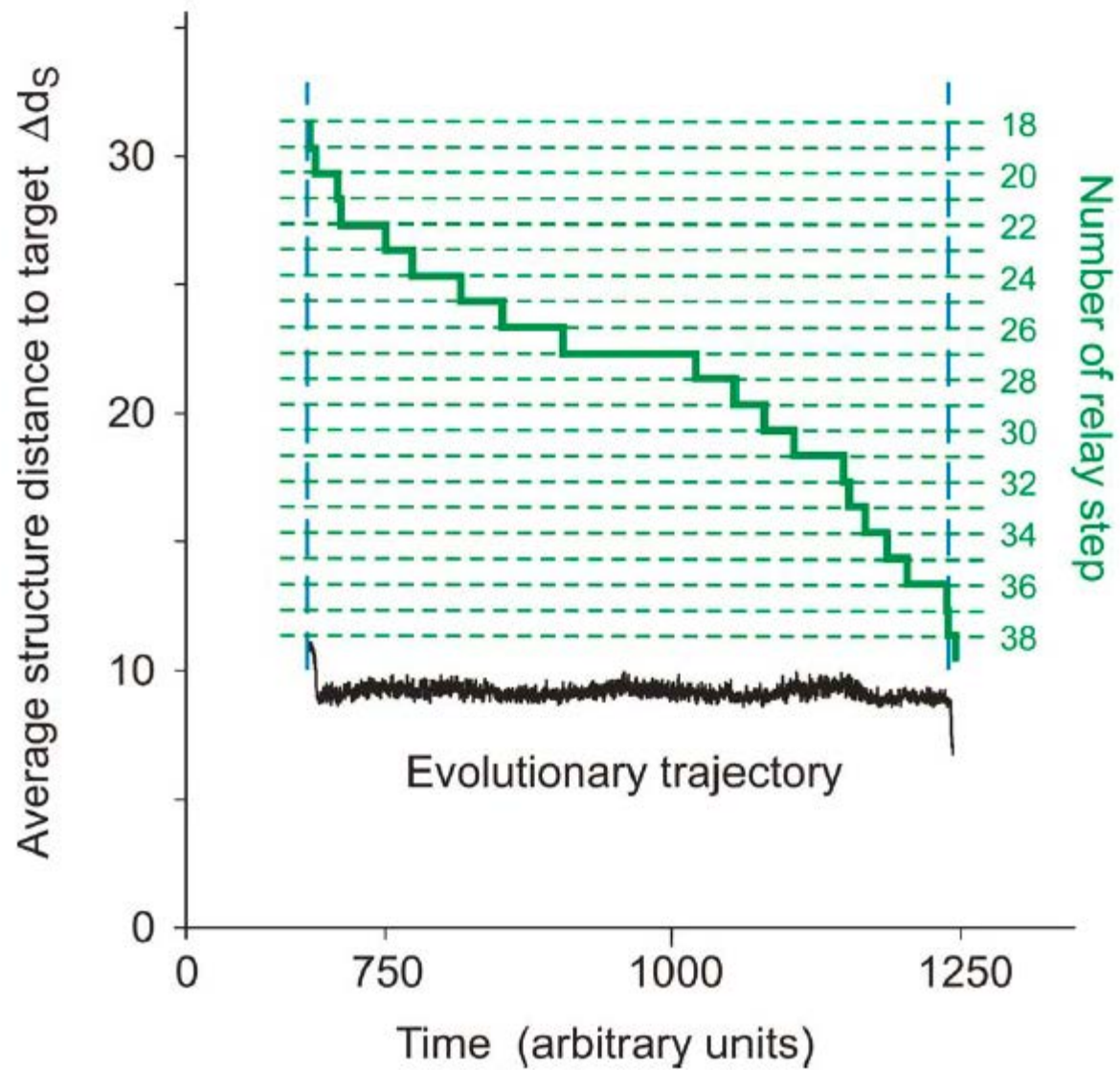


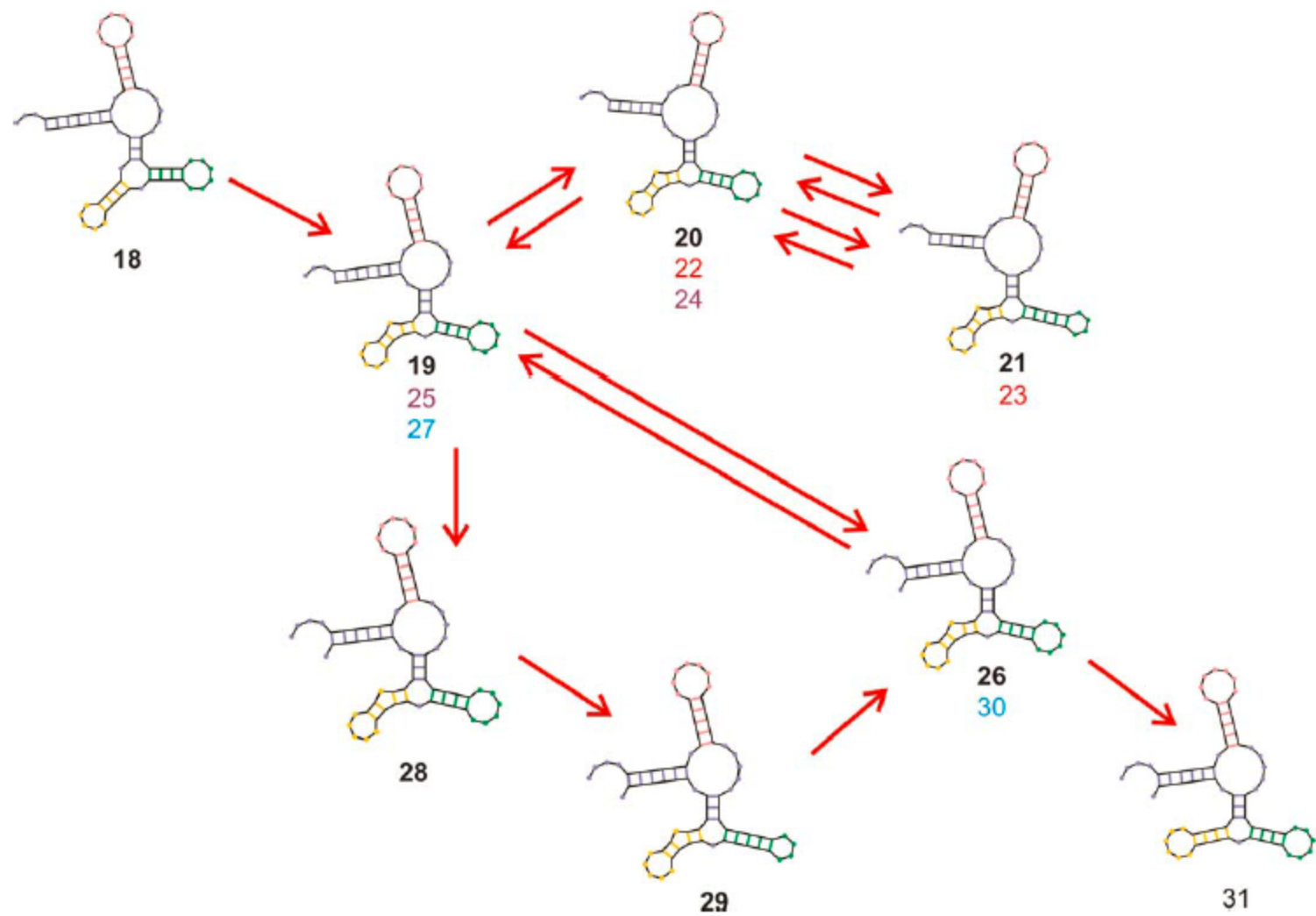
entry	GGUAUGGGCGUUGAAUAGUAGGGUUUAAACCAAUCGG	CAACGAUCUCGUGUGCGCAUUUCAUAUCCCGUACAGAA
8	.((((((((((((((((.....((((.....)))).....)))))).....((((.....))))))))).....	
exit	GGUAUGGGCGUUGAAUA	UAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUCCUACAGAA
entry	GGUAUGGGCGUUGAAUAAUAGGGUUUAAACCAAUCGGCCAACGAUCUCGUGUGCGCAUUUCAUAUA	CCAUACAGAA
9	.(((((((.((((((((((((((((.....((((.....)))).....)))))).....((((.....))))))))).....	
exit	UGGAUGGACGUUGAAUAA	CAAGGUAUCGACCAAACAACCAACGAGUAAAGUGUGUACGCCCCACACACCGUCCCAAG
entry	UGGAUGGACGUUGAAUAA	CAAGGUAUCGACCAAACAACCAACGAGUAAAGUGUGUACGCCCCACACACCGUCCCAAG
10	.(((((((.((((((((((((((((.....((((.....)))).....)))))).....((((.....))))))))).....	
exit	UGGAUGGACGUUGAAUAA	CAAGGUAUCGACCAAACAACCAACGAGUAAAGUGUGUACGCCCCACACAGCGUCCCAAG

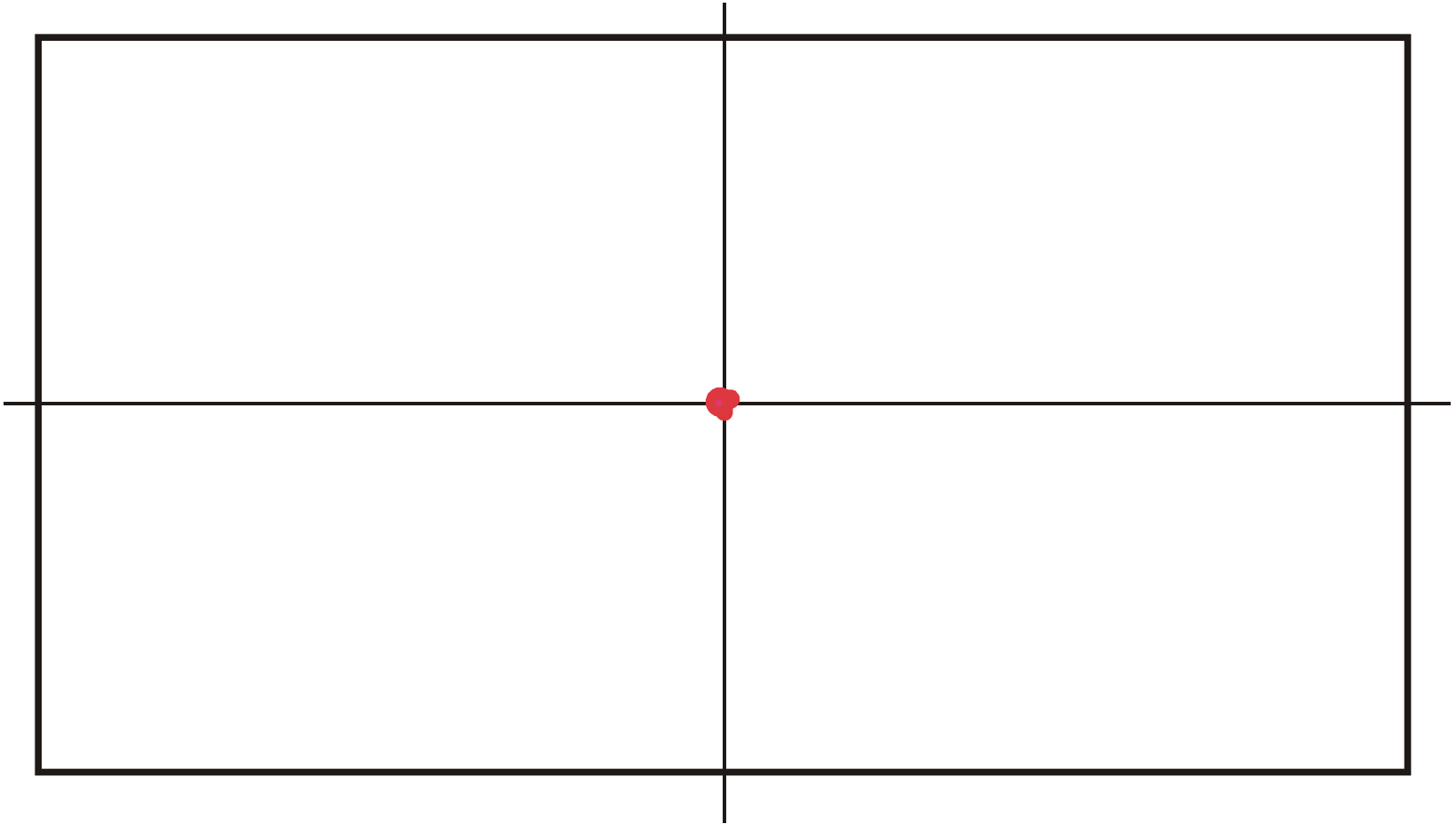
Transition inducing point mutations change the molecular structure

Neutral point mutations leave the molecular structure unchanged

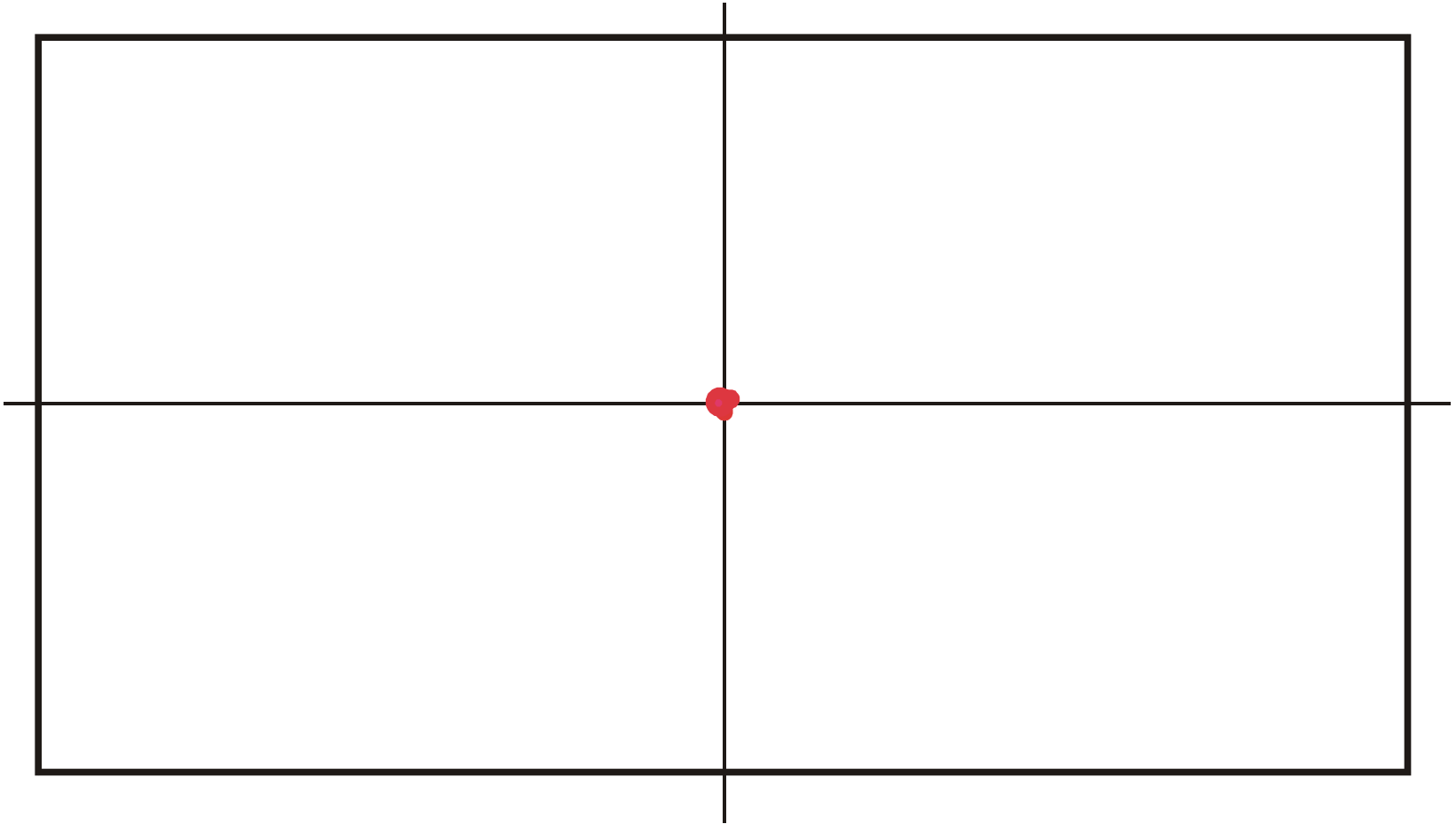
Neutral genotype evolution during phenotypic stasis



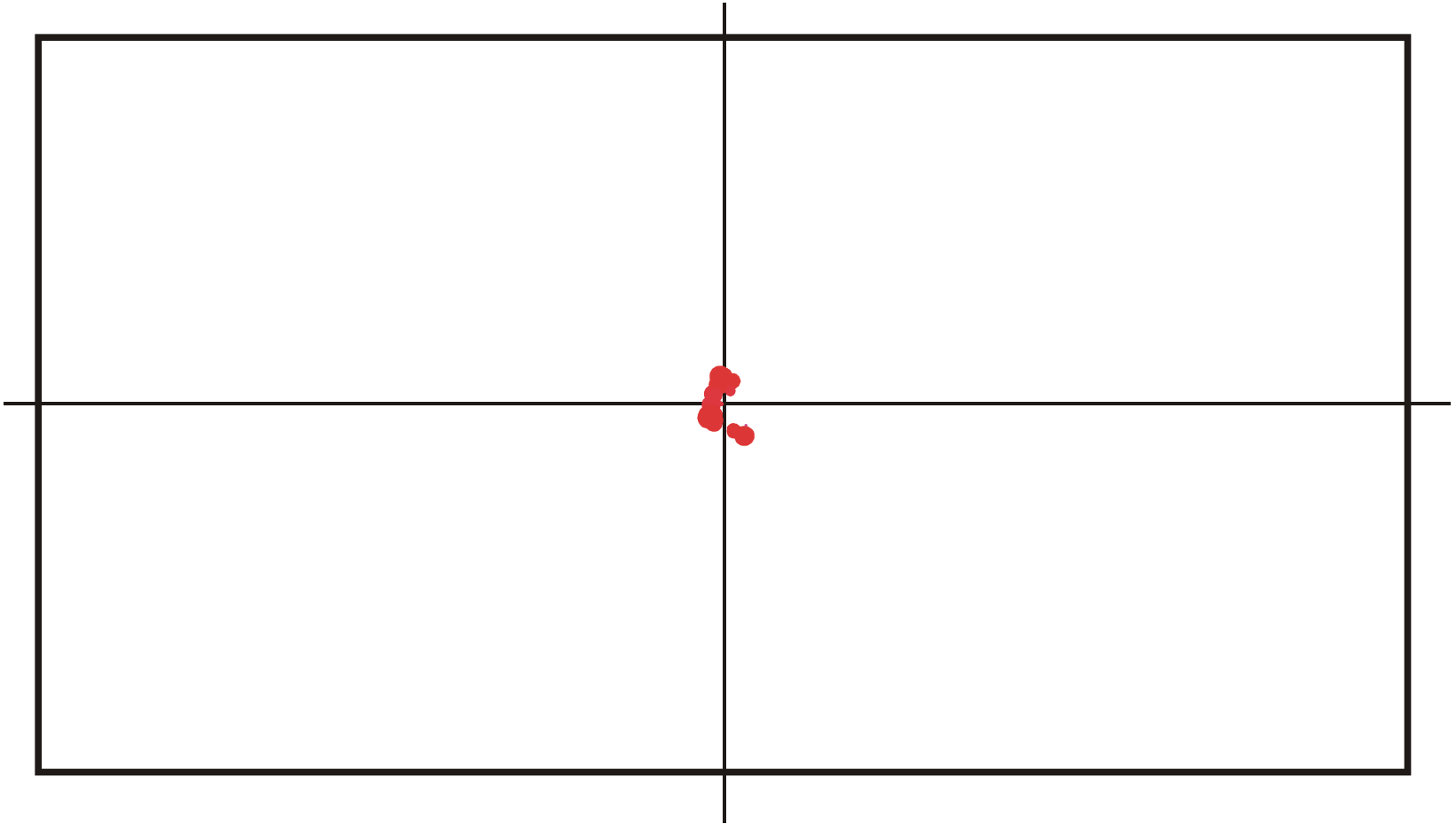




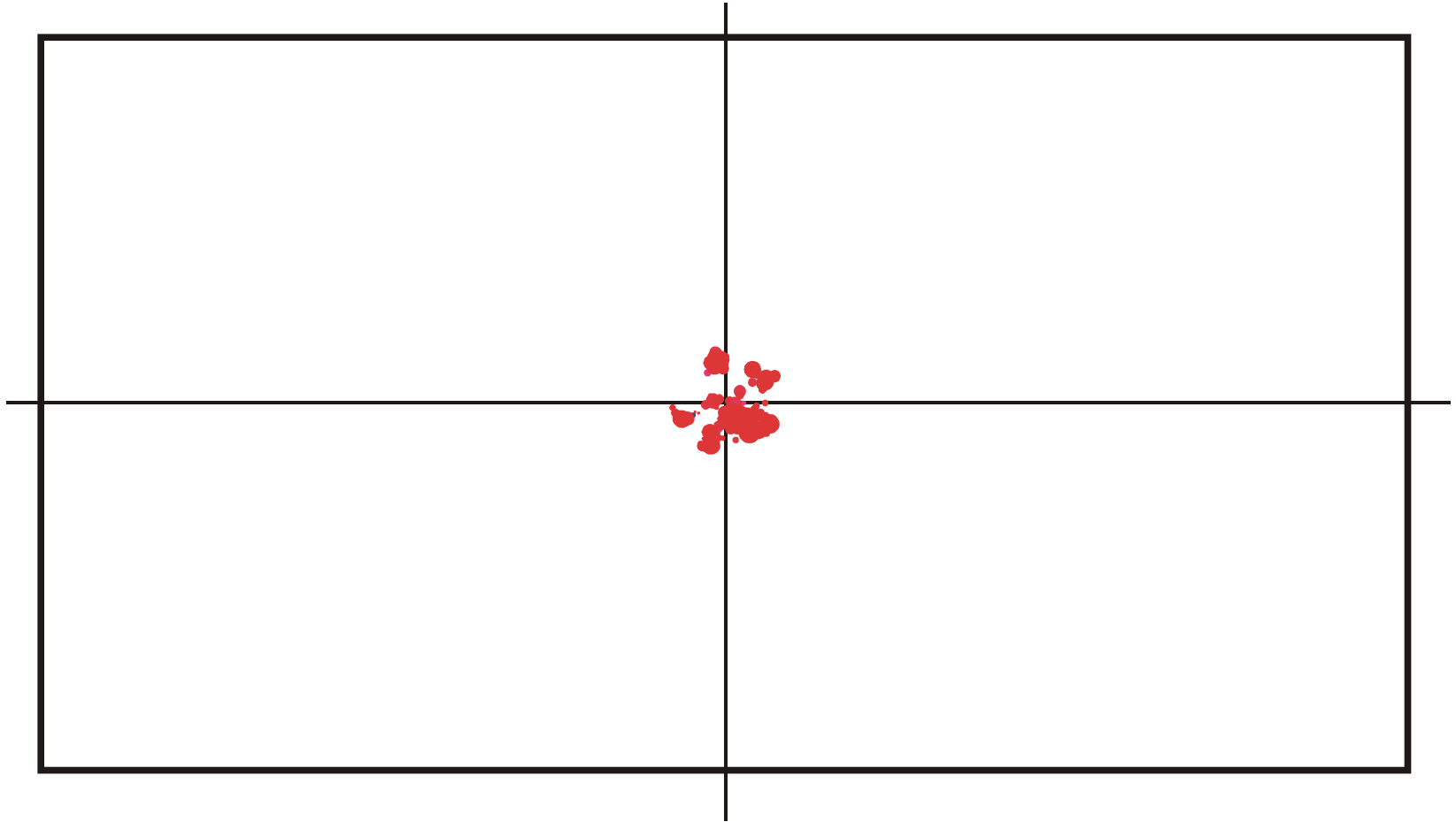
Spreading and evolution of a population on a neutral network: $t = 150$



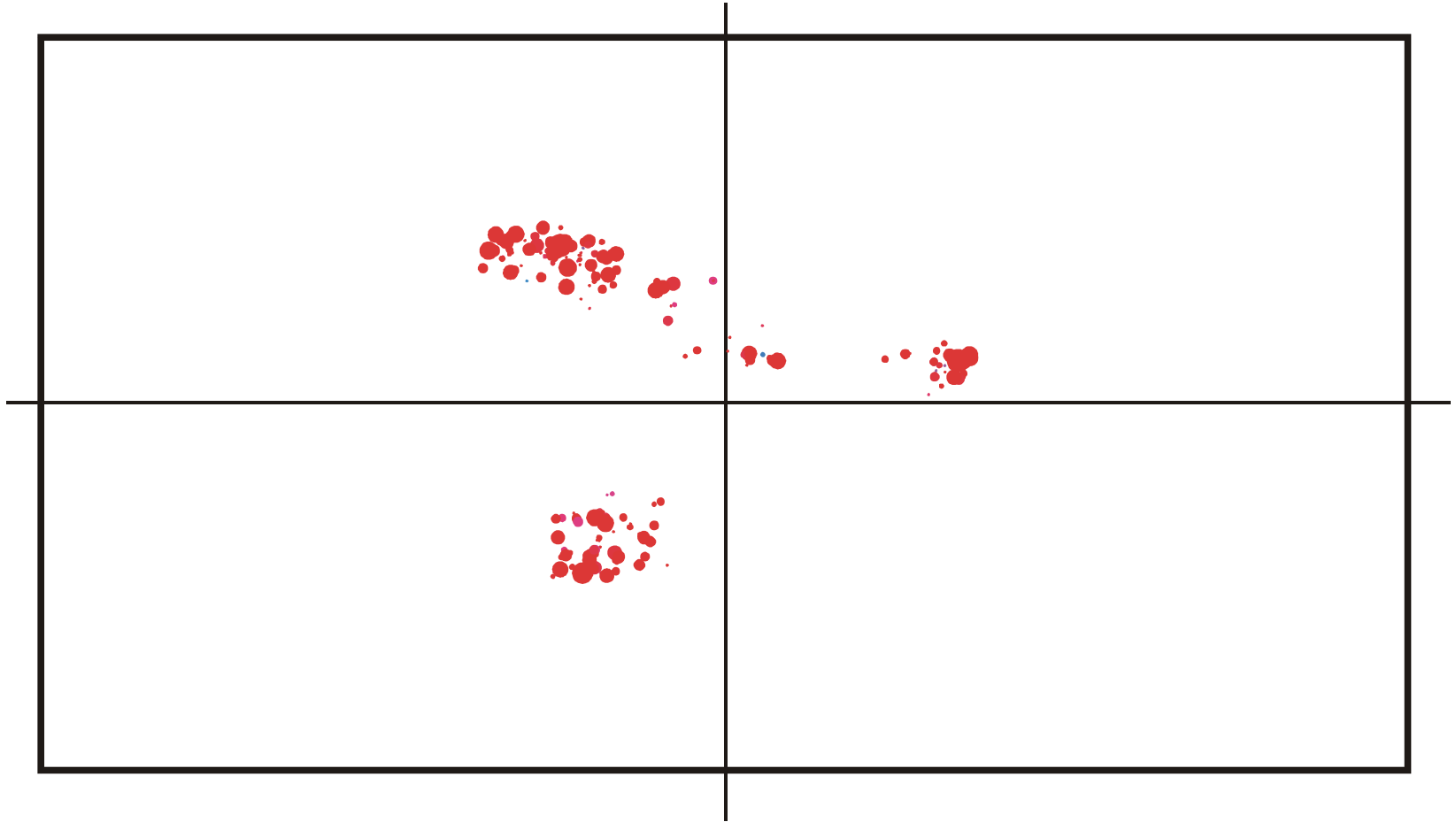
Spreading and evolution of a population on a neutral network: $t = 150$



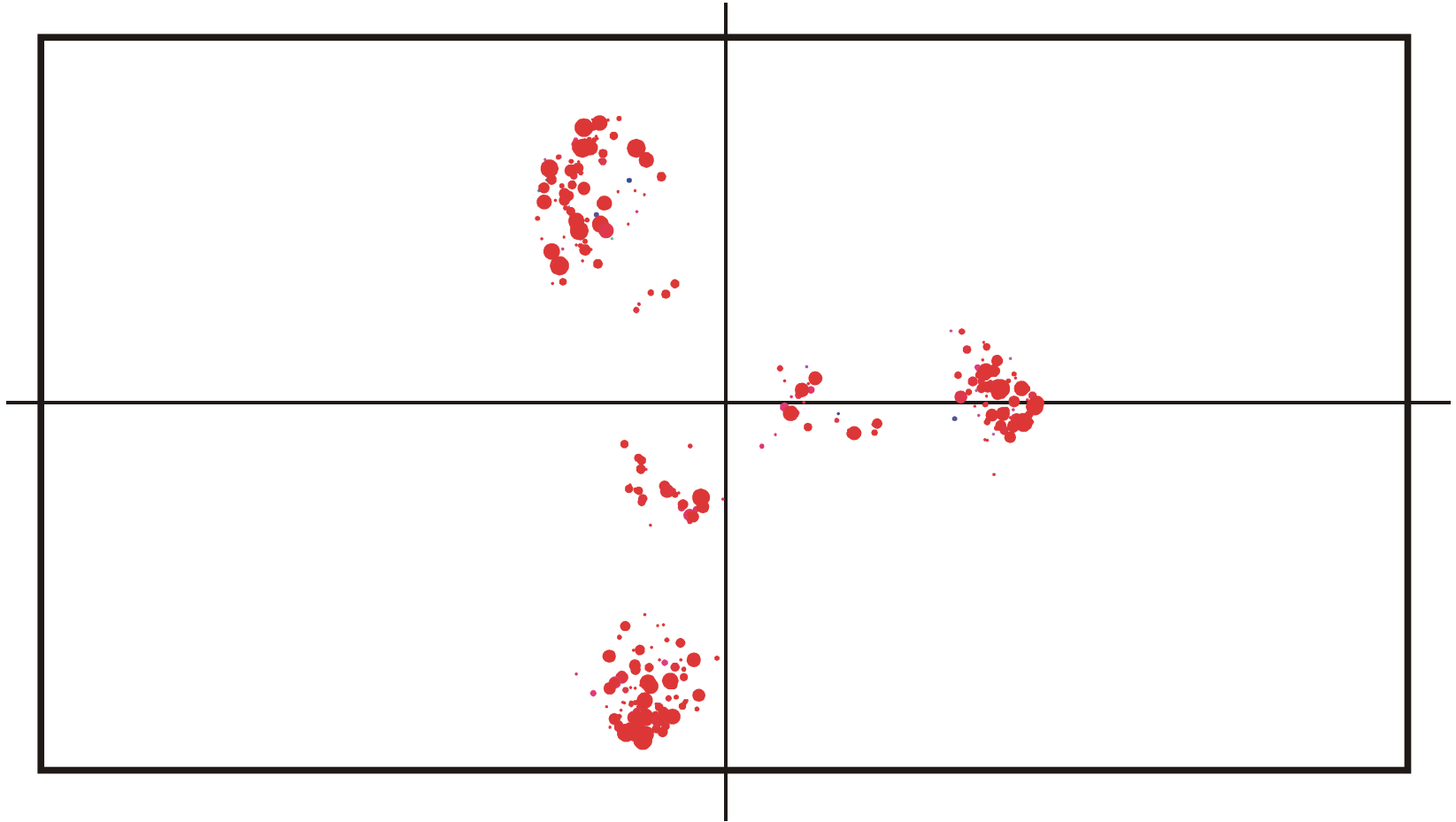
Spreading and evolution of a population on a neutral network : $t = 170$



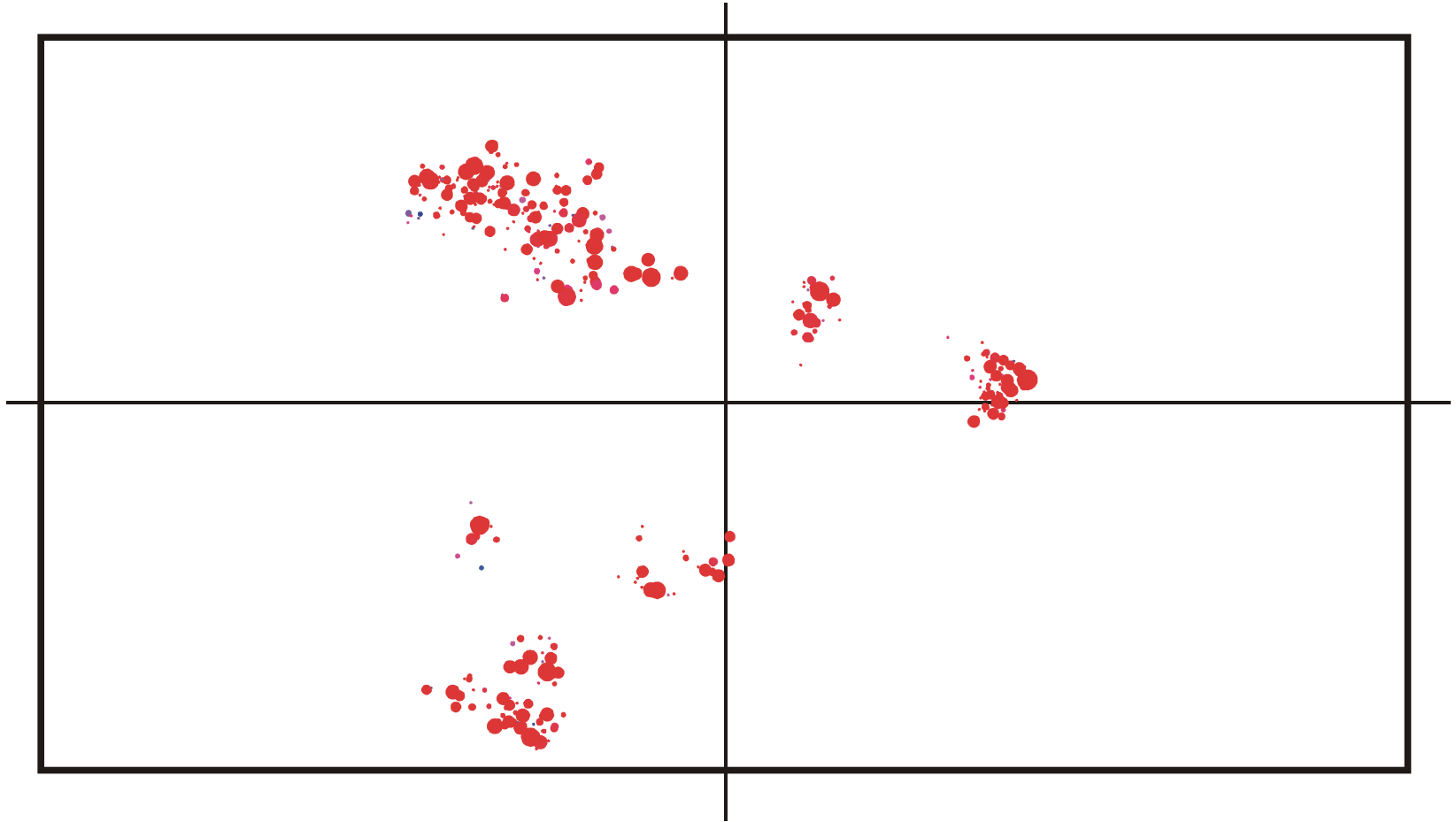
Spreading and evolution of a population on a neutral network : $t = 200$



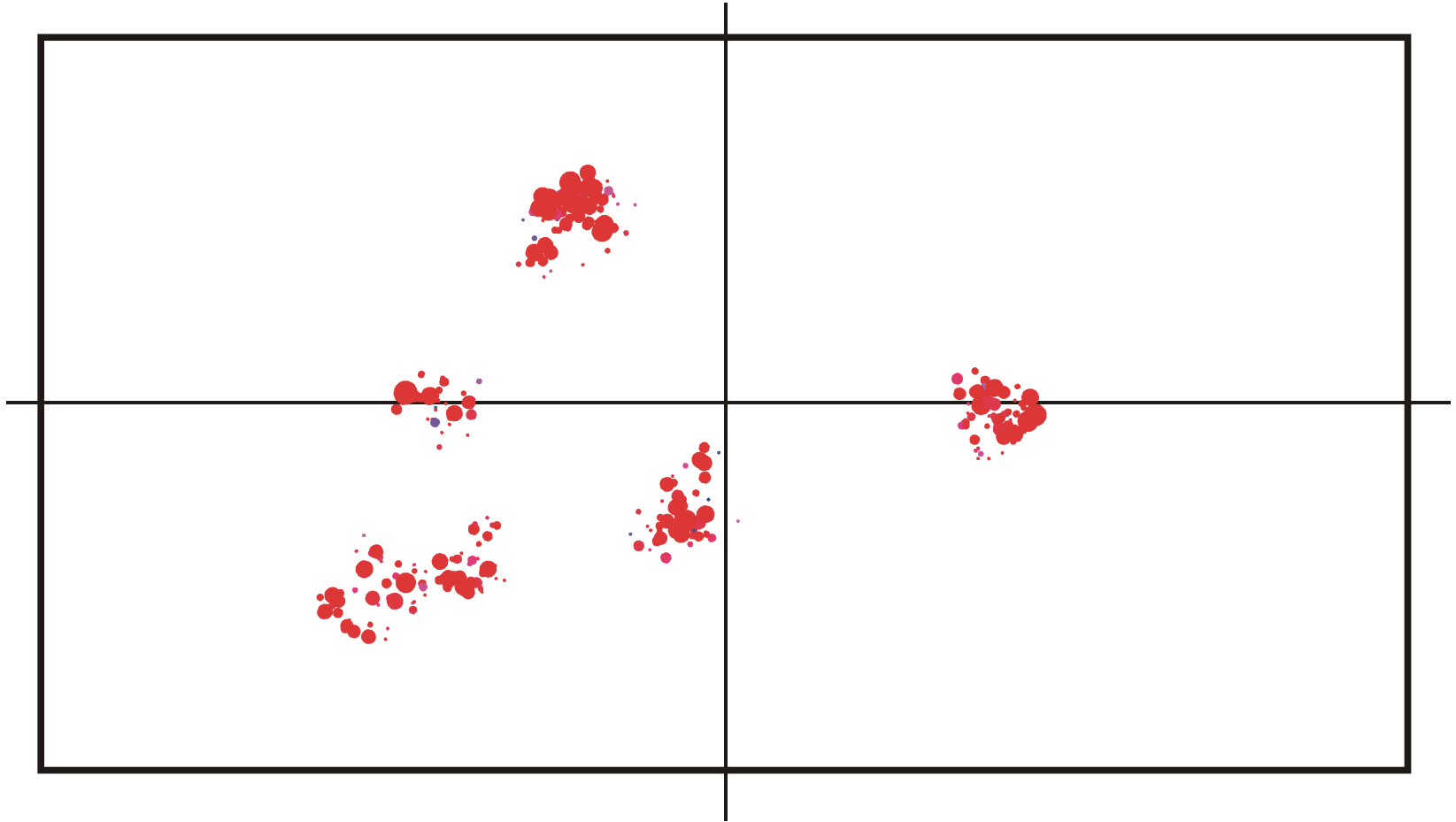
Spreading and evolution of a population on a neutral network : $t = 350$



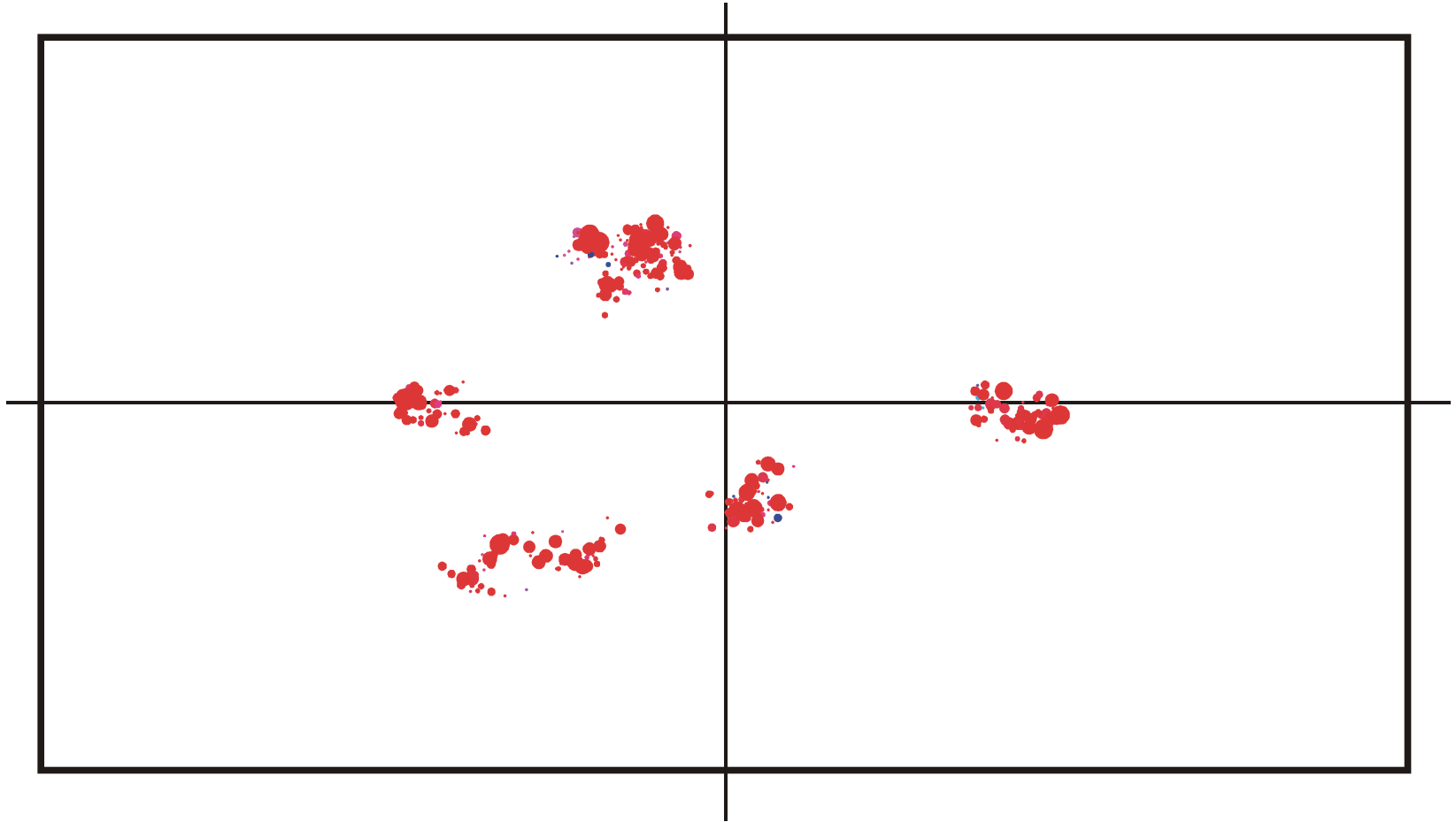
Spreading and evolution of a population on a neutral network : $t = 500$



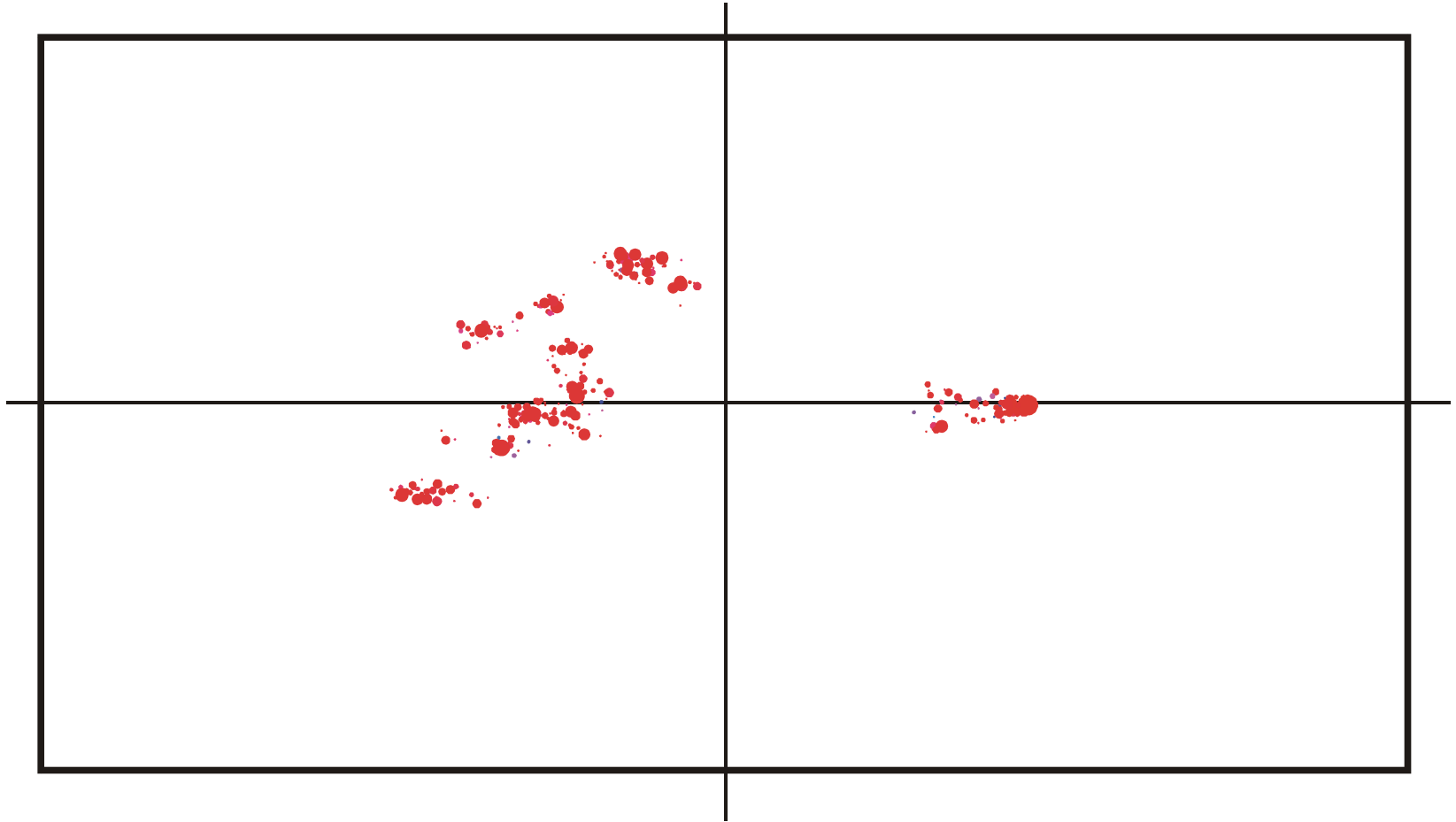
Spreading and evolution of a population on a neutral network : $t = 650$



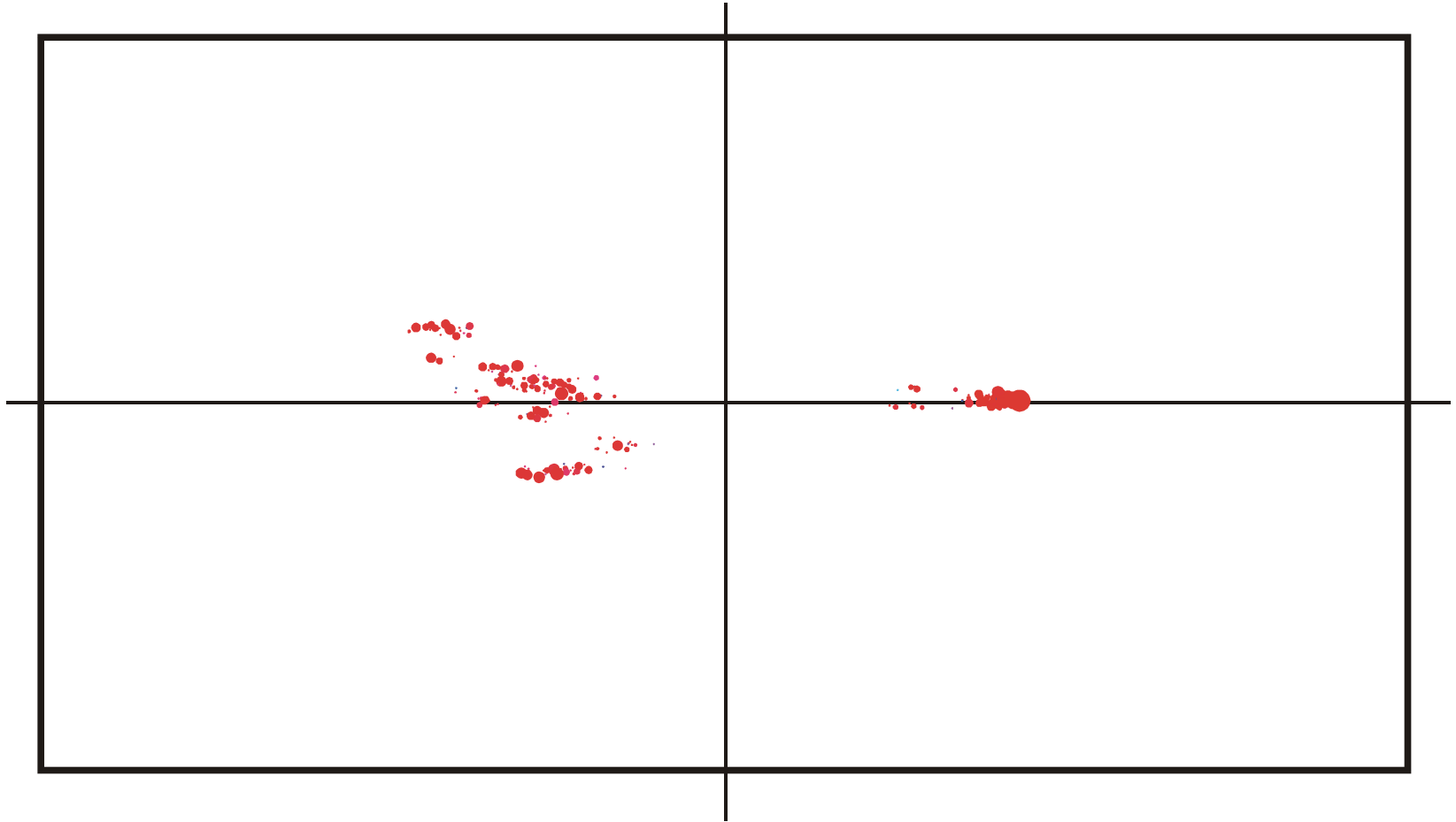
Spreading and evolution of a population on a neutral network : $t = 820$



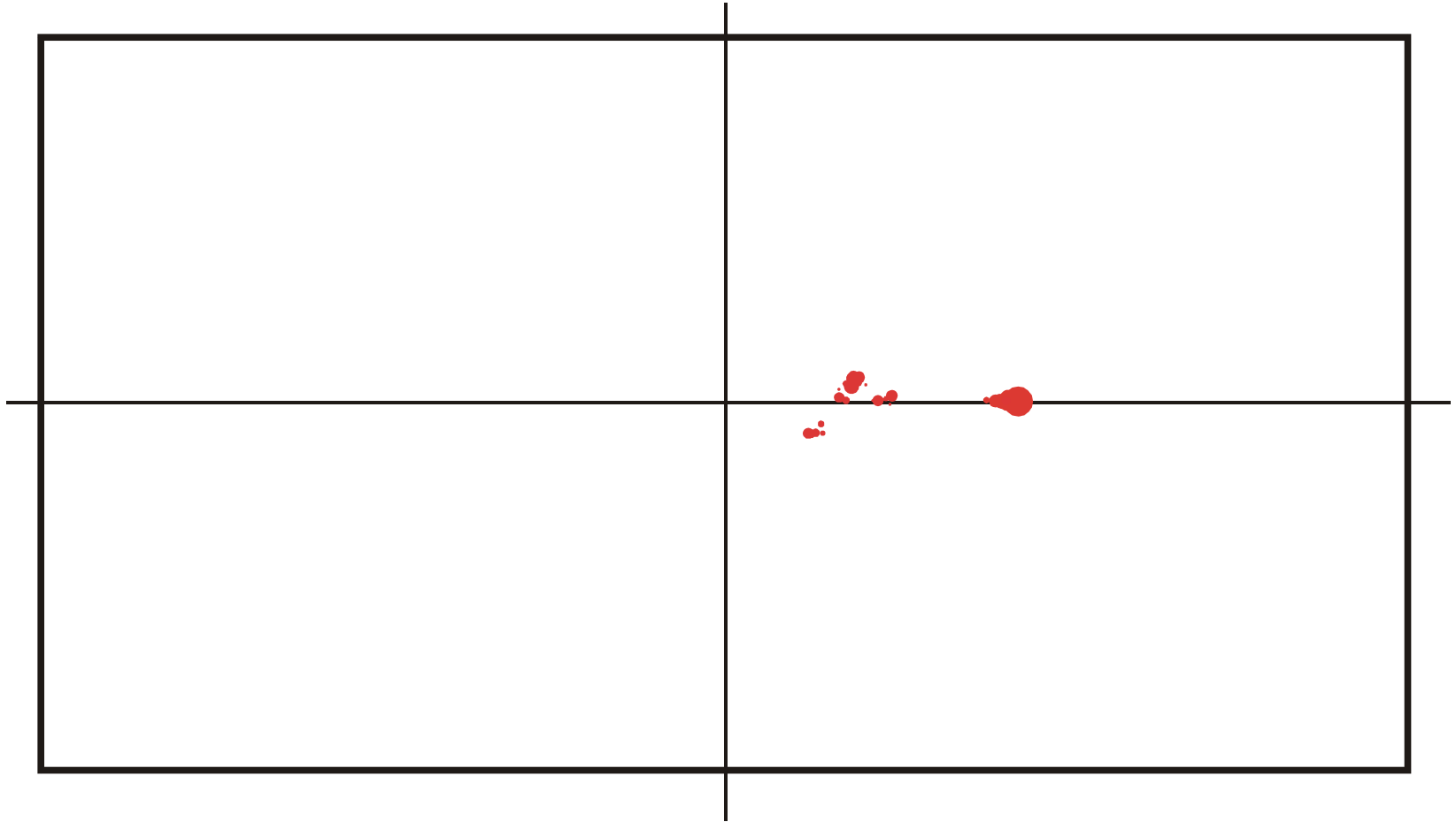
Spreading and evolution of a population on a neutral network : $t = 825$



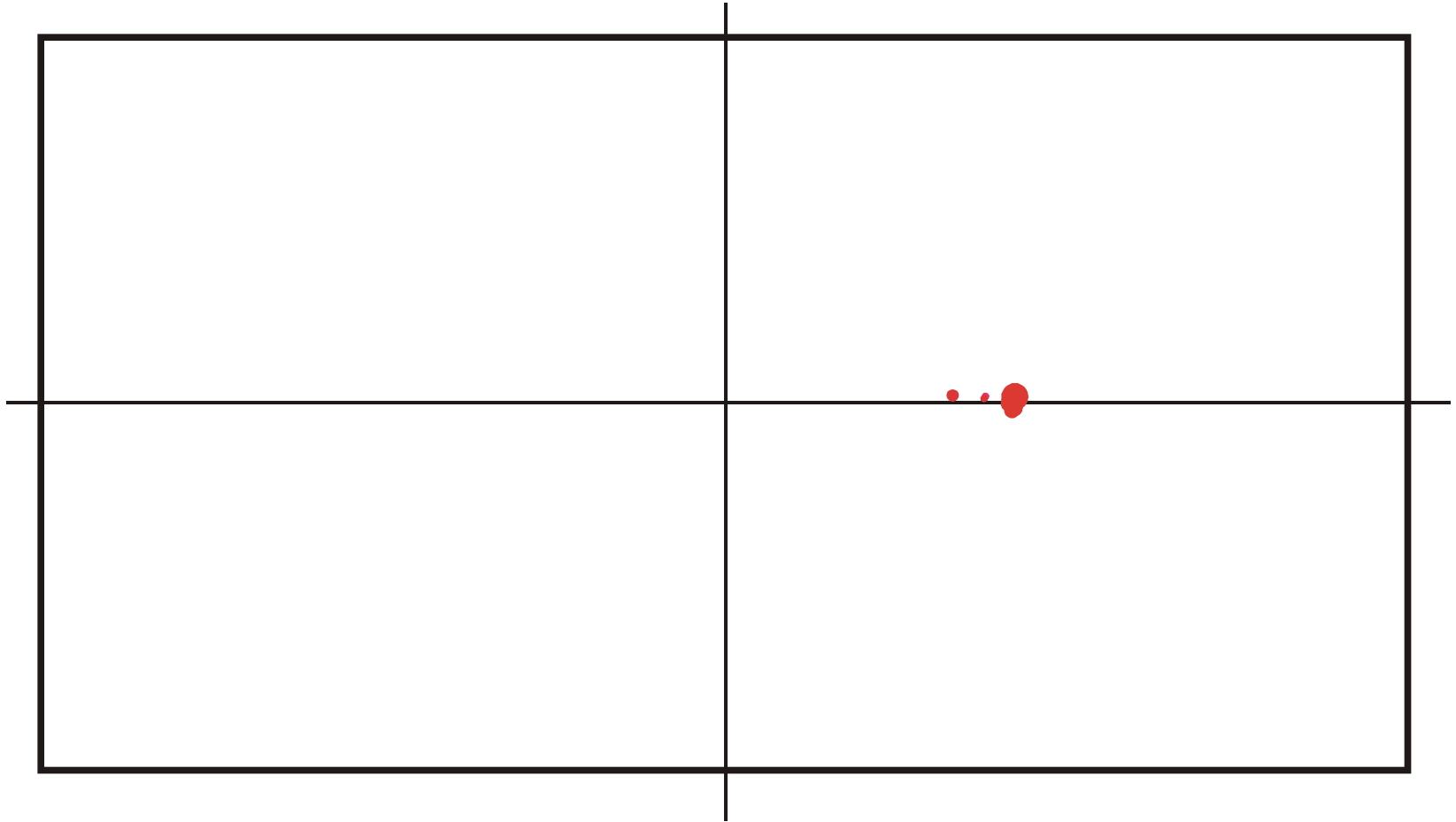
Spreading and evolution of a population on a neutral network : $t = 830$



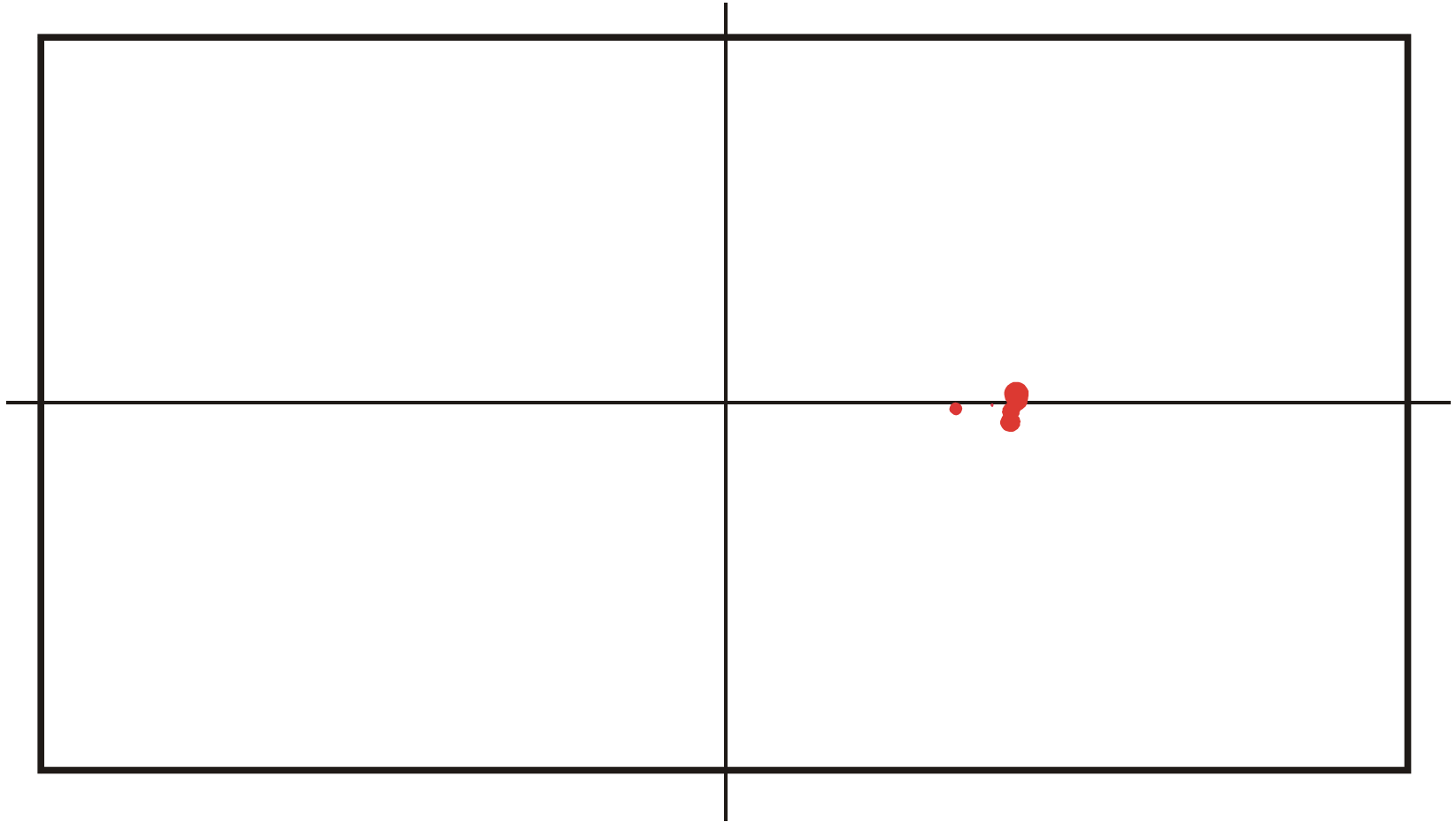
Spreading and evolution of a population on a neutral network : $t = 835$



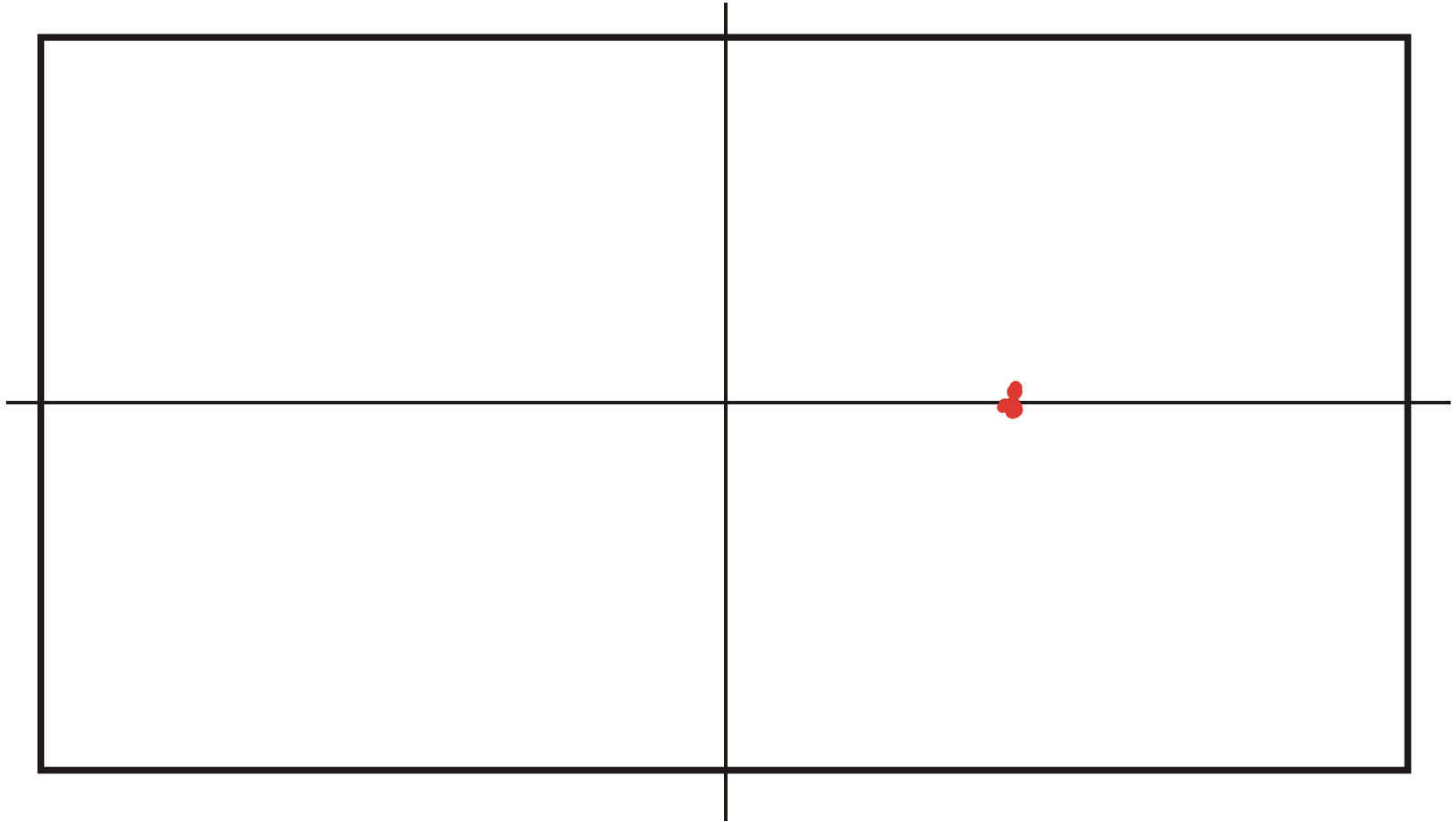
Spreading and evolution of a population on a neutral network : $t = 840$



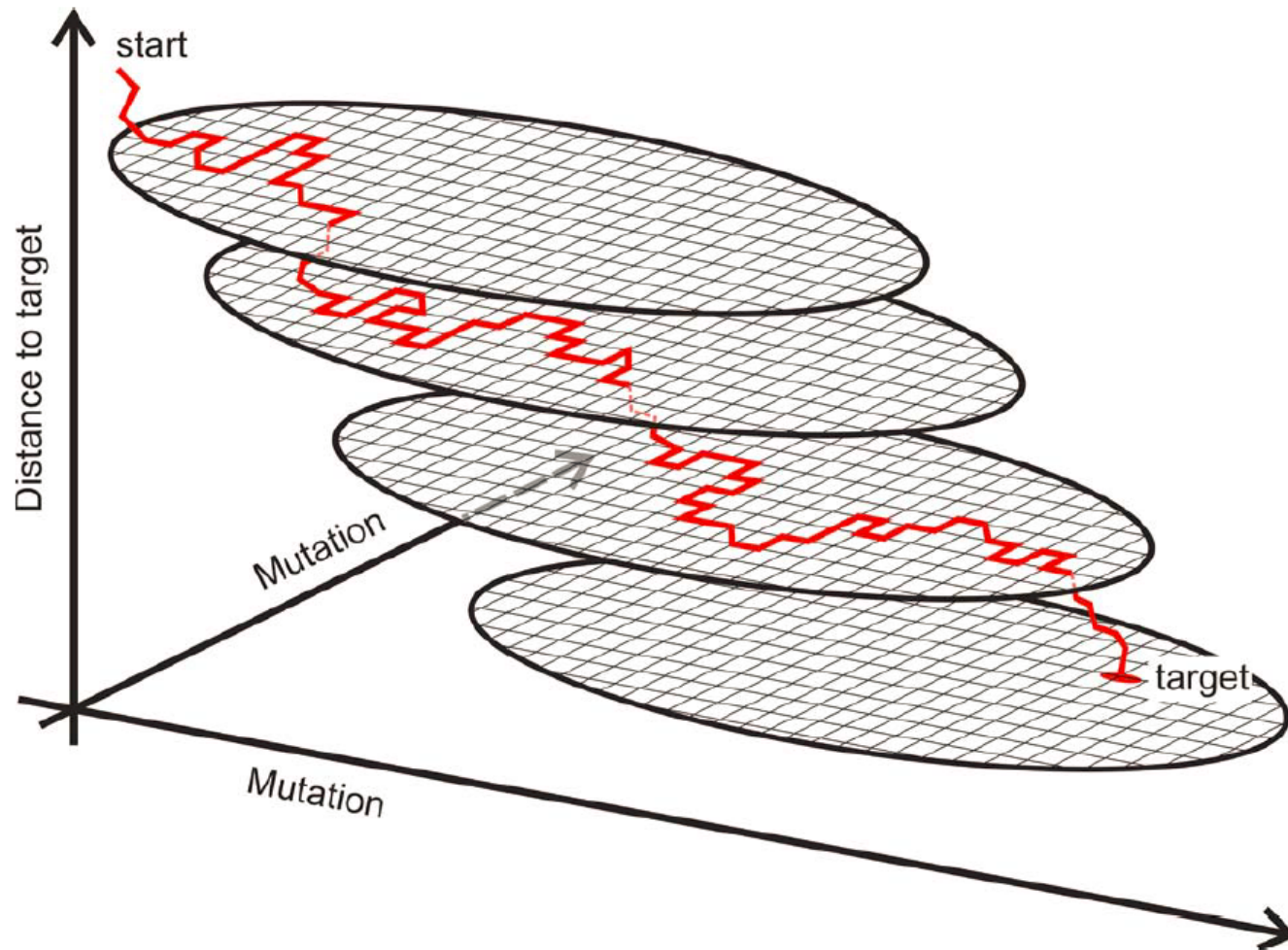
Spreading and evolution of a population on a neutral network : $t = 845$



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



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



A sketch of optimization on neutral networks

Neutrality explains both
punctuation and contingency

Coworkers

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

Paul E. Phillipson, University of Colorado at Boulder, CO

Heinz Engl, Philipp Kügler, James Lu, Stefan Müller, RICAM Linz, AT

Jord Nagel, Kees Pleij, Universiteit Leiden, NL

Walter Fontana, Harvard Medical School, MA

Martin Nowak, Harvard University, MA

Christian Reidys, Nankai University, Tien Tsin, China

Christian Forst, Los Alamos National Laboratory, NM

Thomas Wiehe, Ulrike Göbel, Walter Grüner, Stefan Kopp, Jaqueline Weber,
Institut für Molekulare Biotechnologie, Jena, GE

Ivo L.Hofacker, Christoph Flamm, Andreas Svrček-Seiler, Universität Wien, AT

Kurt Grünberger, Michael Kospach, Andreas Wernitznig, Stefanie Widder,
Stefan Wuchty, Jan Cupal, Stefan Bernhart, Lukas Endler, Ulrike Langhammer,
Rainer Machne, Ulrike Mückstein, Erich Bornberg-Bauer,
Universität Wien, AT



Universität Wien

Acknowledgement of support

Fonds zur Förderung der wissenschaftlichen Forschung (FWF)
Projects No. 09942, 10578, 11065, 13093
13887, and 14898



Universität Wien

Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF)
Project No. Mat05

Jubiläumsfonds der Österreichischen Nationalbank
Project No. Nat-7813

European Commission: Contracts No. 98-0189, 12835 (NEST)

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

Österreichische Akademie der Wissenschaften

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>

