Anniversary of the Winter-Seminar

Brand

Sölden

Heiligenblut

Klosters

Zuoz

Wolkenstein

Winter- Seminar

Brand

40

Klosters

Sölden

Zuoz

Heiligenblut

Wolkenstein

Winter- Seminar

Why master clans are fitter than master sequences

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Austria
and
The Santa Fe Institute, Santa Fe, New Mexico, USA

50. Winterseminar

Klosters, 10.– 16.01.2015

Web-Page for further information:

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

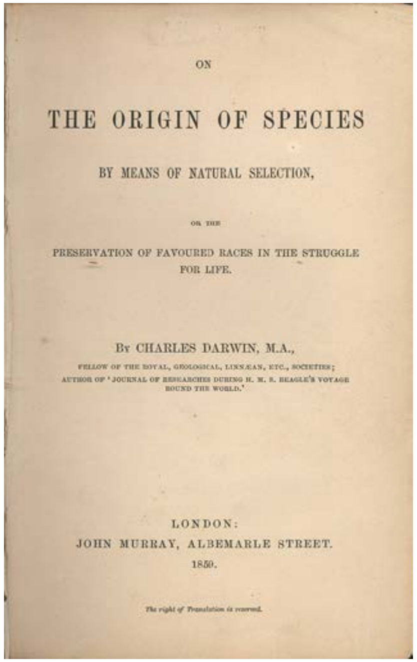
Santa Fe Institute Preprint Series # 12-06-006

1. Prologue
2. Landscapes and evolutionary dynamics
3. Realistic landscapes and mutant clans
4. Mutation flow analysis
5. Neutrality in evolutionary dynamics
6. Concluding remarks and perspectives

1. **Prologue**
2. Landscapes and evolutionary dynamics
3. Realistic landscapes and mutant clans
4. Mutation flow analysis
5. Neutrality in evolutionary dynamics
6. Concluding remarks and perspectives

Charles Darwin,
1809 - 1882

Charles Darwin,
1809 - 1882



ON

THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

OR THE

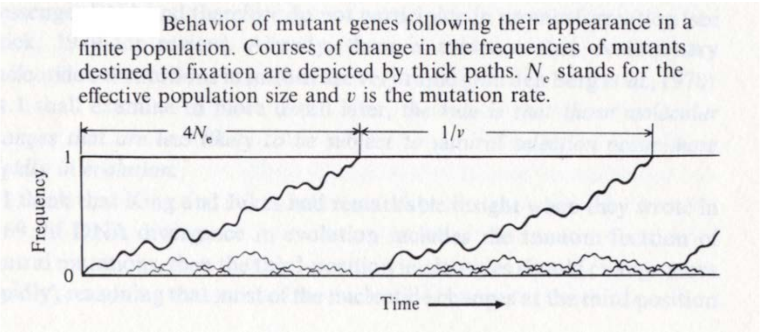
PRESERVATION OF FAVOURED RACES IN THE STRUGGLE
FOR LIFE.

By CHARLES DARWIN, M.A.,
FELLOW OF THE ROYAL, GEOLOGICAL, LINNÆAN, 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.

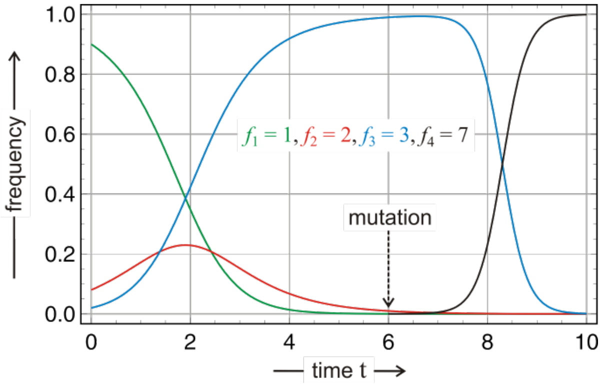
Charles Darwin,
1809 - 1882



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.

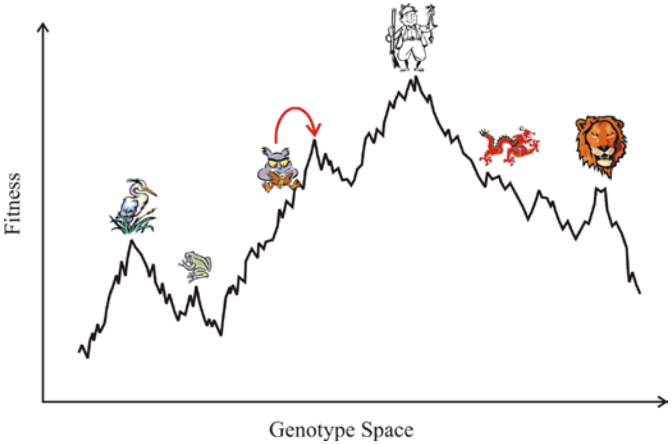
survival of the survivor

Sewall Wright,
1889 - 1988

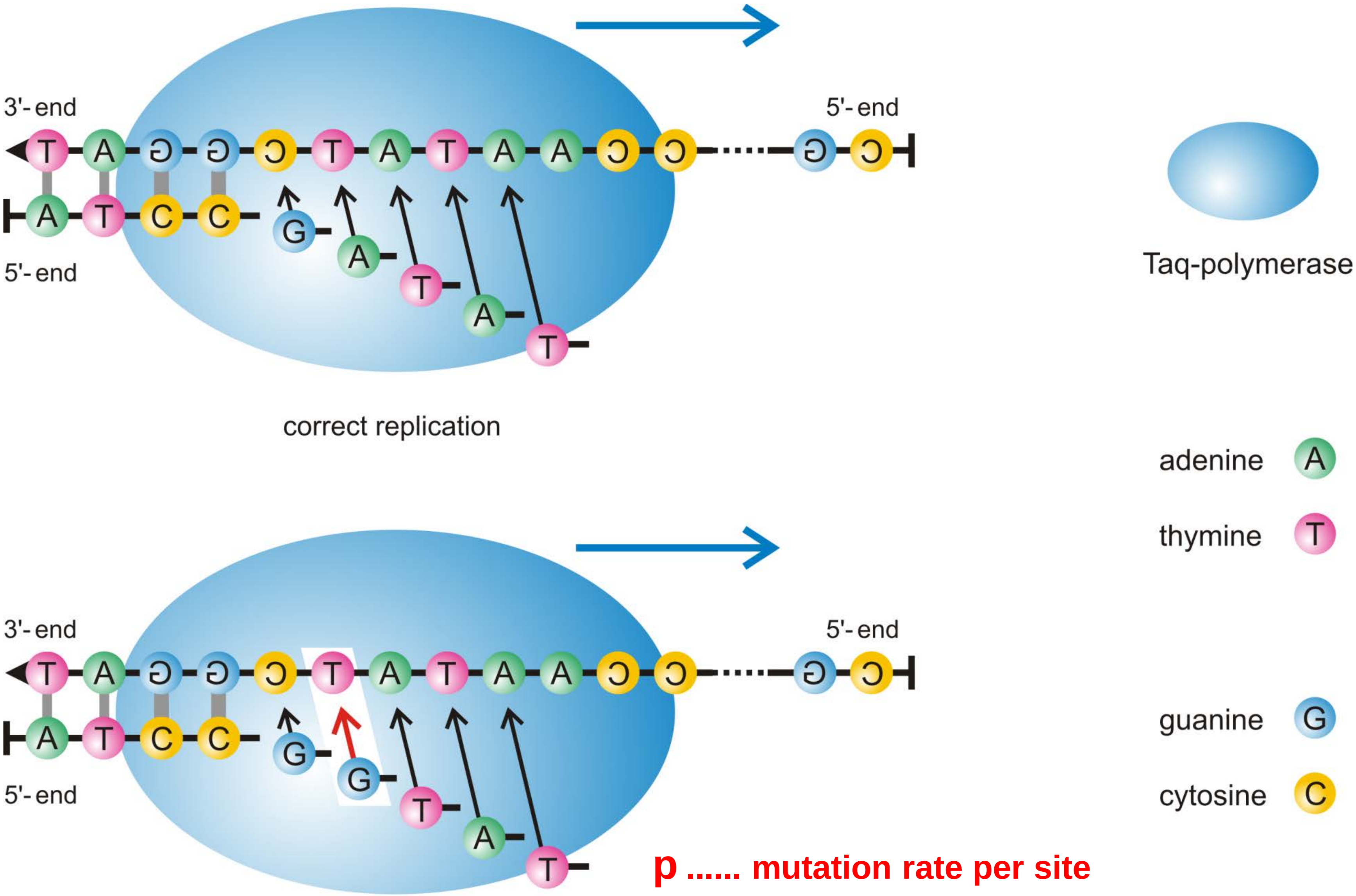


survival of the fittest

Motoo Kimura,
1924 - 1994



fitness landscape



correct replication

mutation

p ...... mutation rate per site and replication

DNA replication and mutation

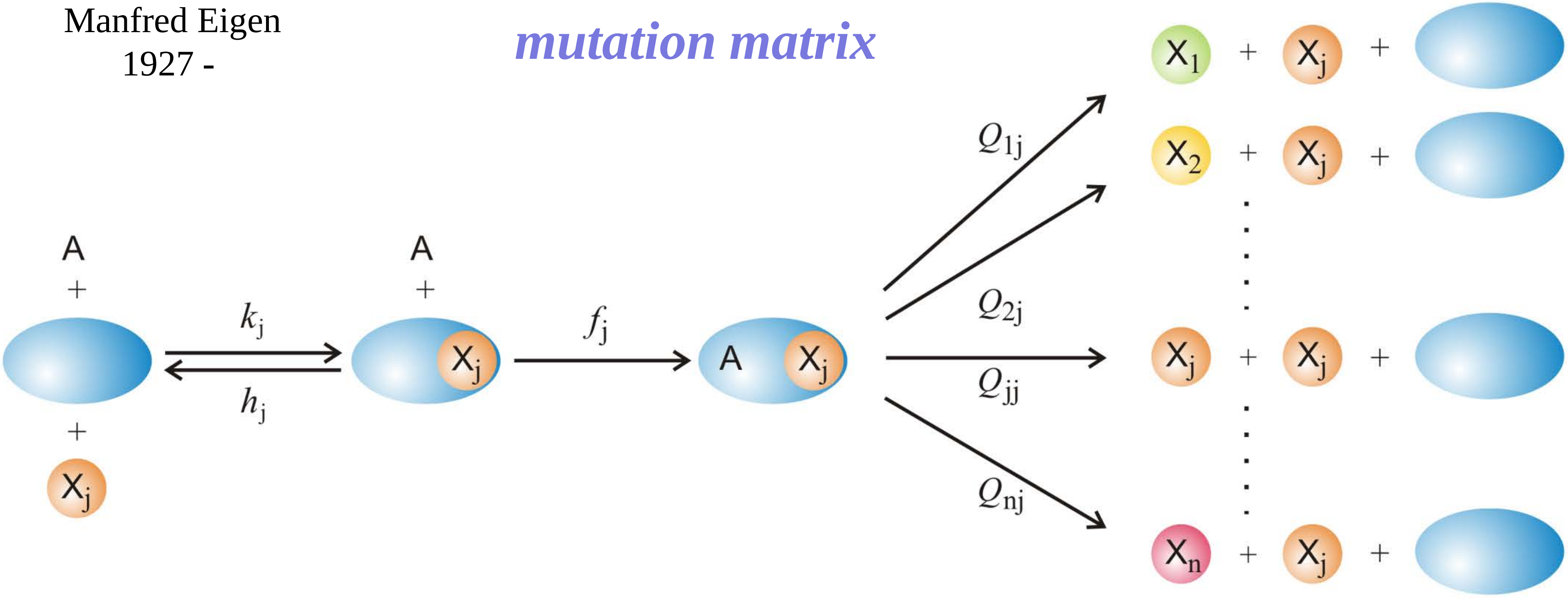
Manfred Eigen
1927 -

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

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

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

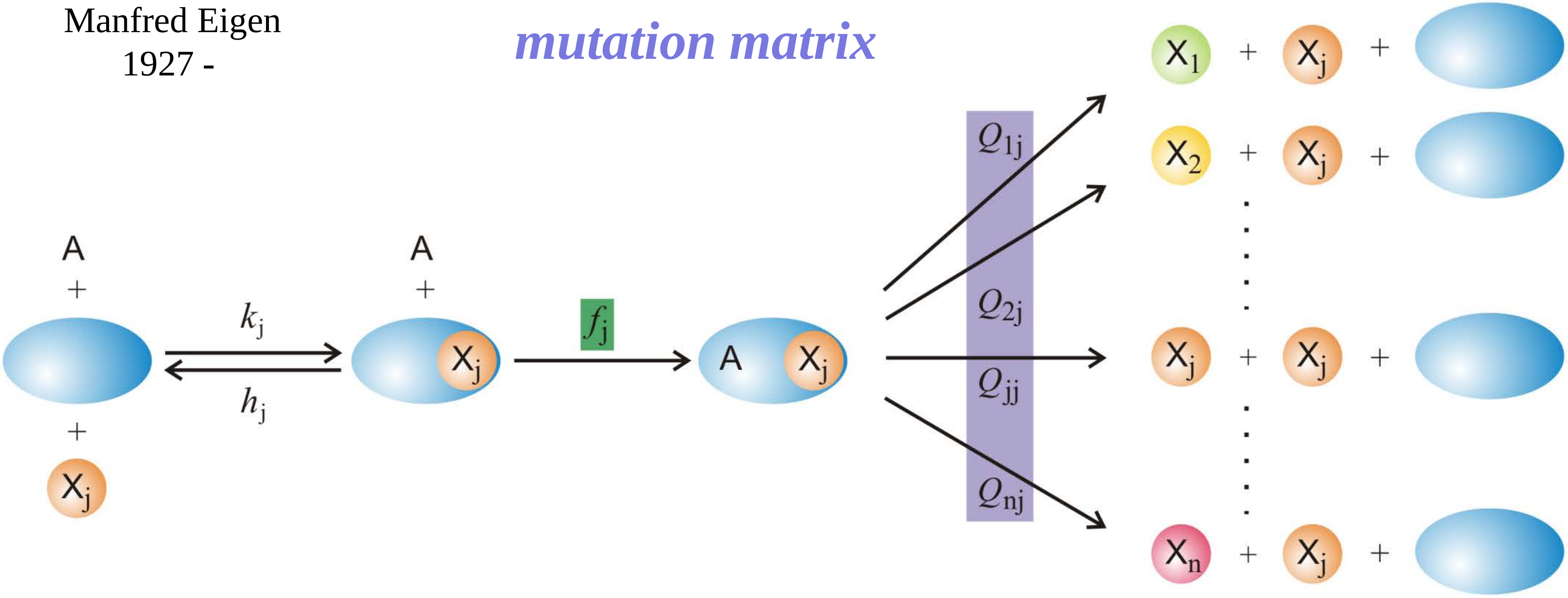
Manfred Eigen
1927 -

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = \sum\nolimits_{i=1}^{n} W_{ji}\, x_i - x_j\, \Phi\,; \quad j = 1, 2, \ldots, n$$

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

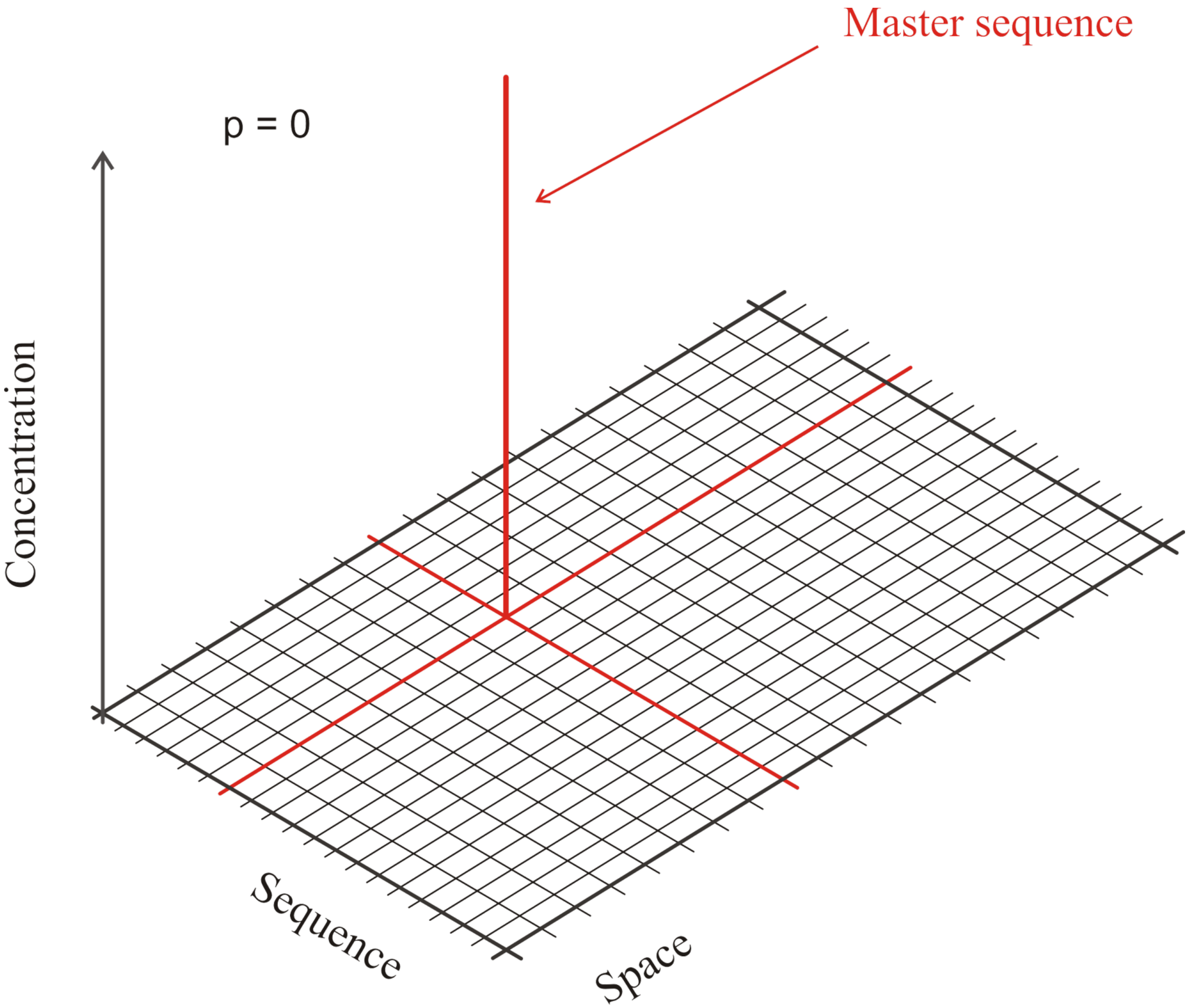
fitness landscape

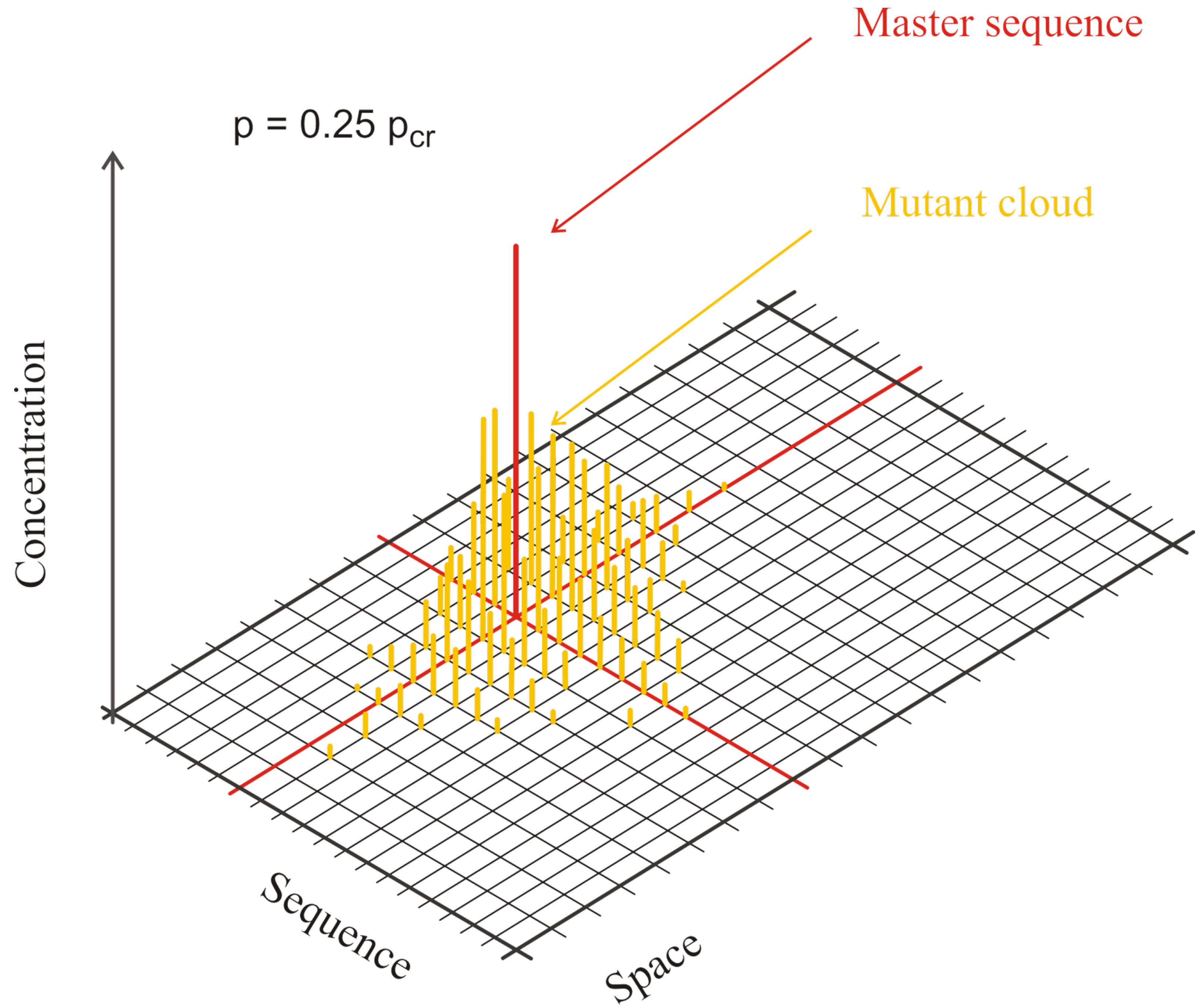
mutation matrix

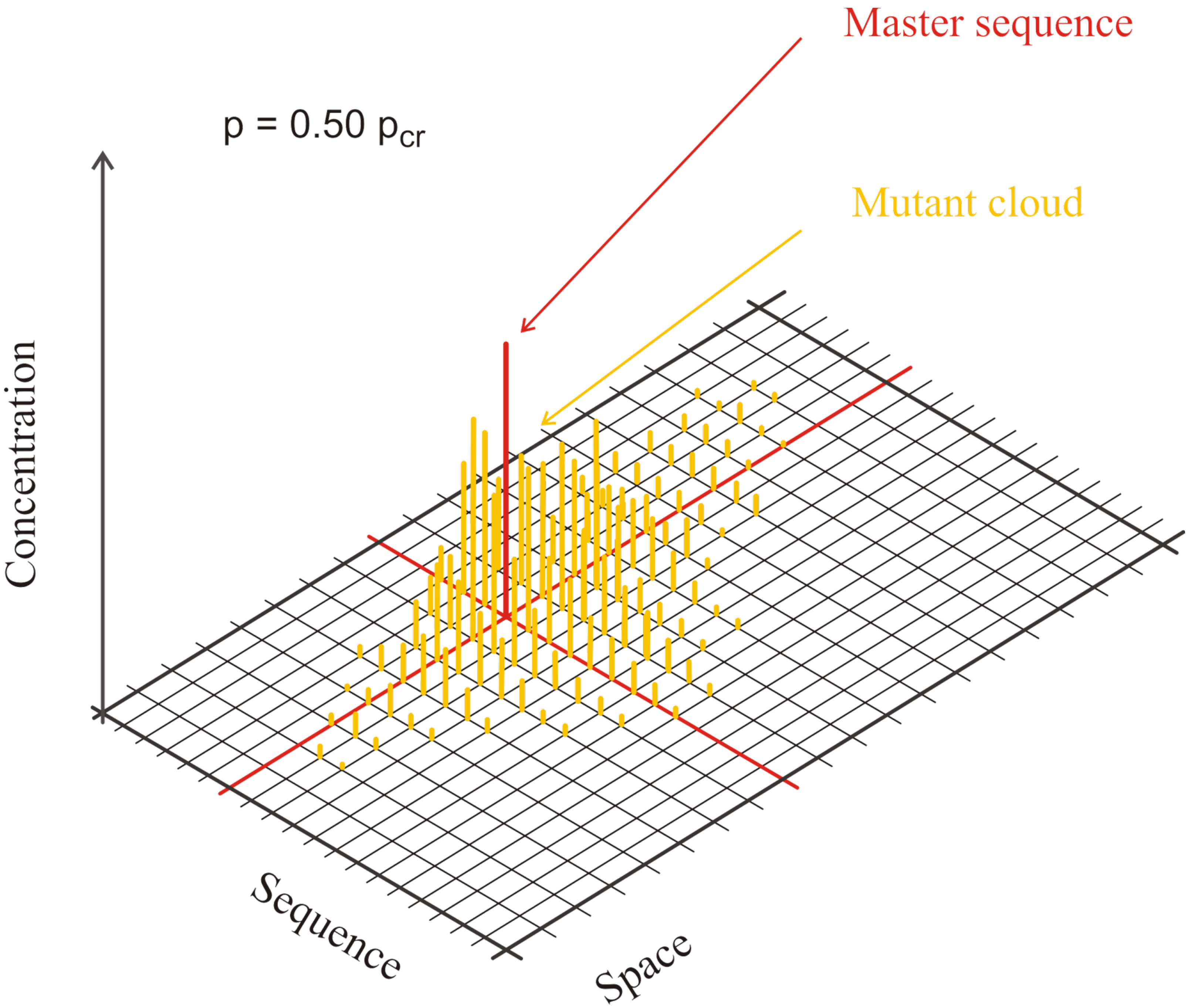


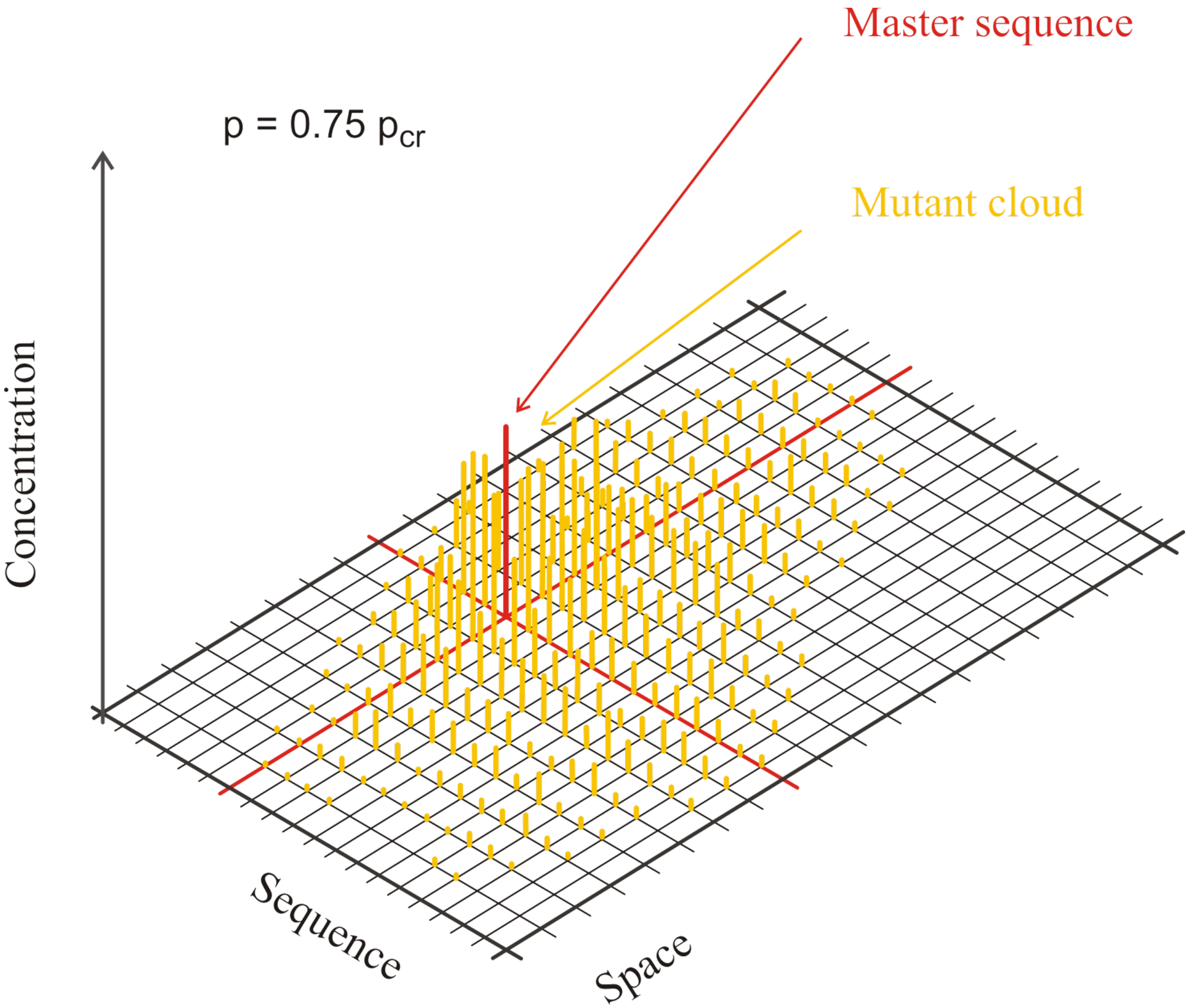
Mutation and (correct) replication as parallel chemical reactions

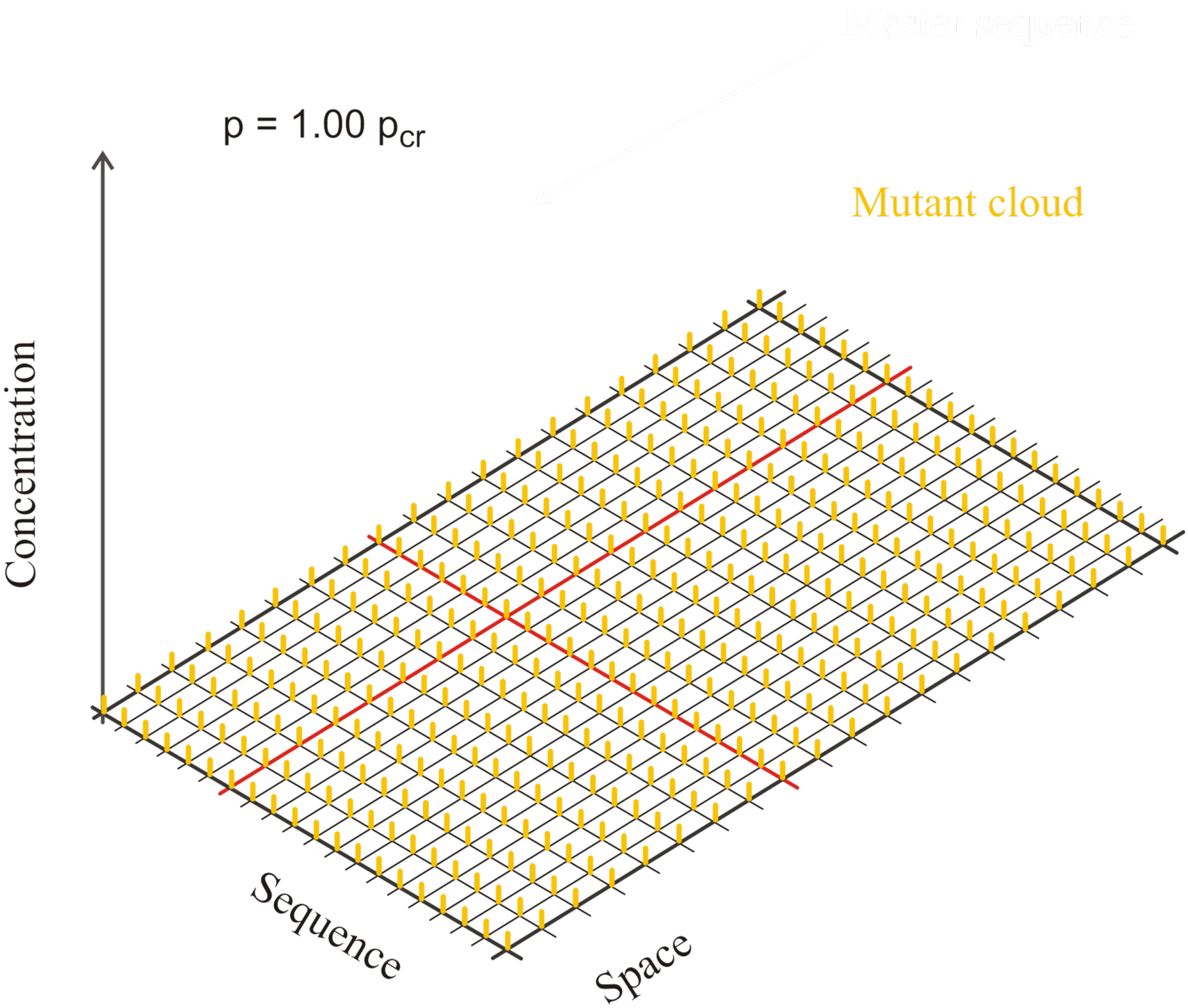
M. Eigen. 1971. *Naturwissenschaften* 58:465,
M. Eigen & P. Schuster.1977-78. *Naturwissenschaften* 64:541, 65:7 und 65:341


Master sequence
p = 0
Concentration
Sequence
Space

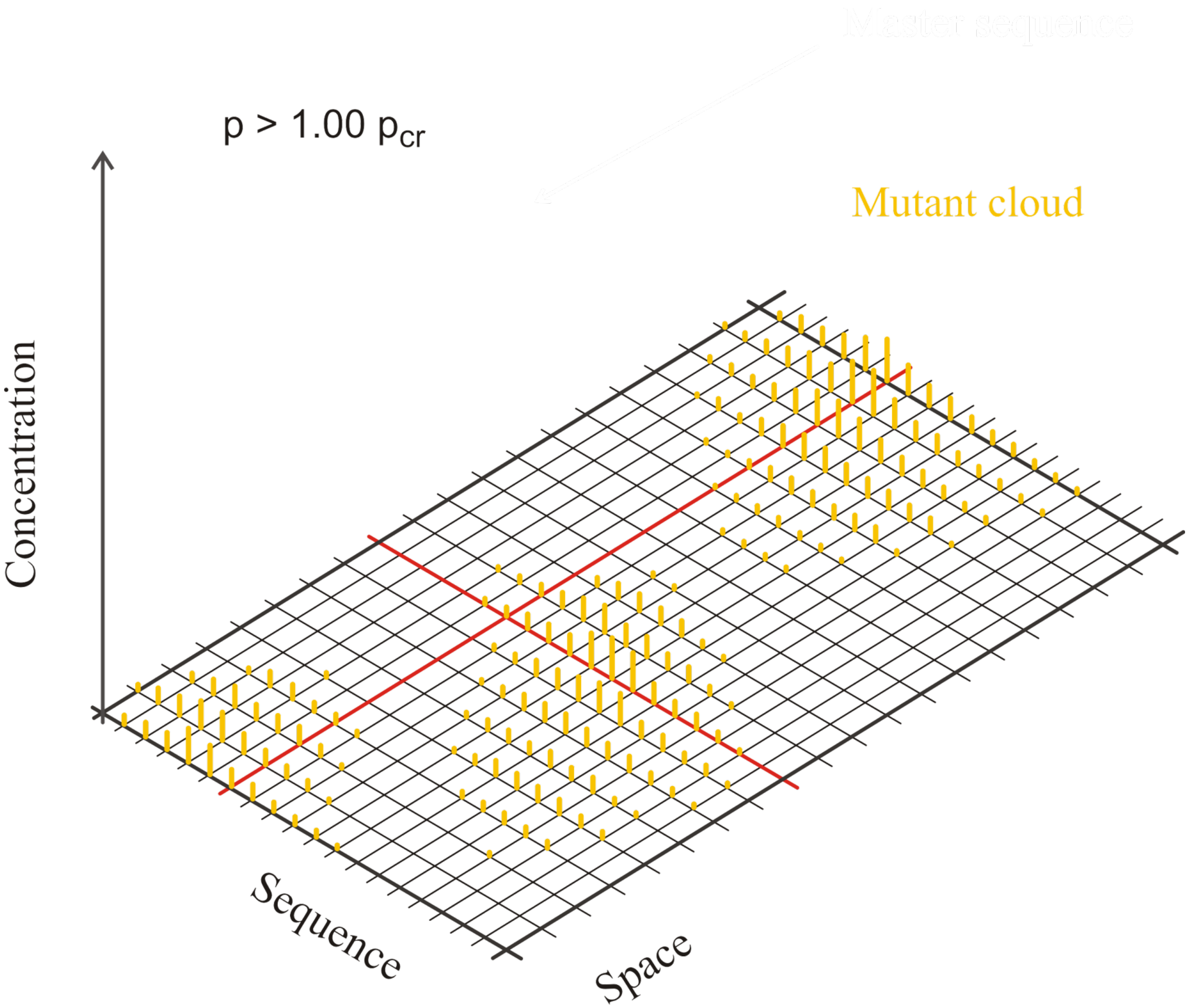

Master sequence
p = 0.25 p_{cr}
Mutant cloud
Concentration
Sequence
Space

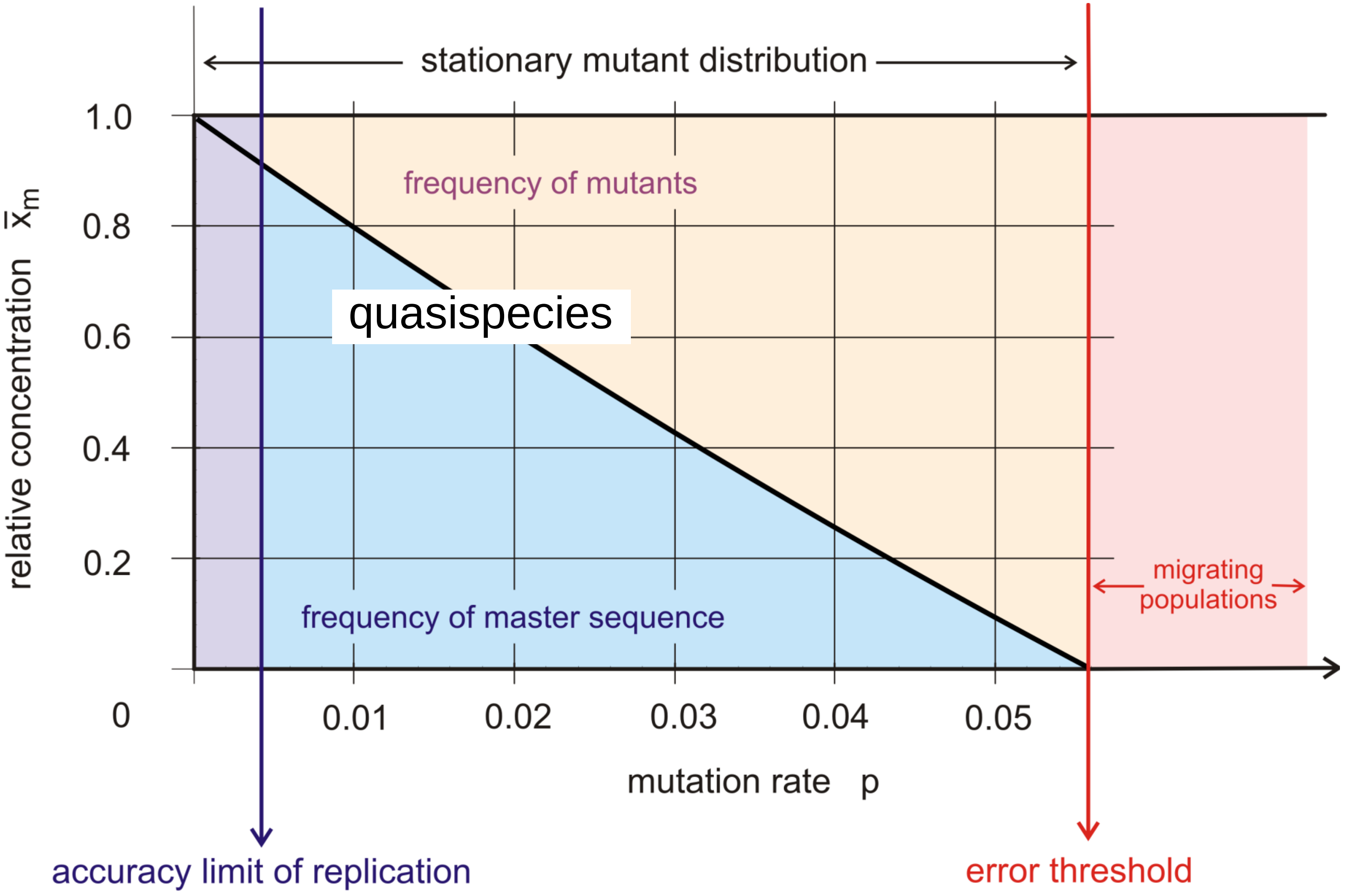

Master sequence
Mutant cloud
$p = 0.50\ p_{cr}$
Concentration
Sequence
Space


Master sequence
Mutant cloud
p = 0.75 p_{cr}
Concentration
Sequence
Space

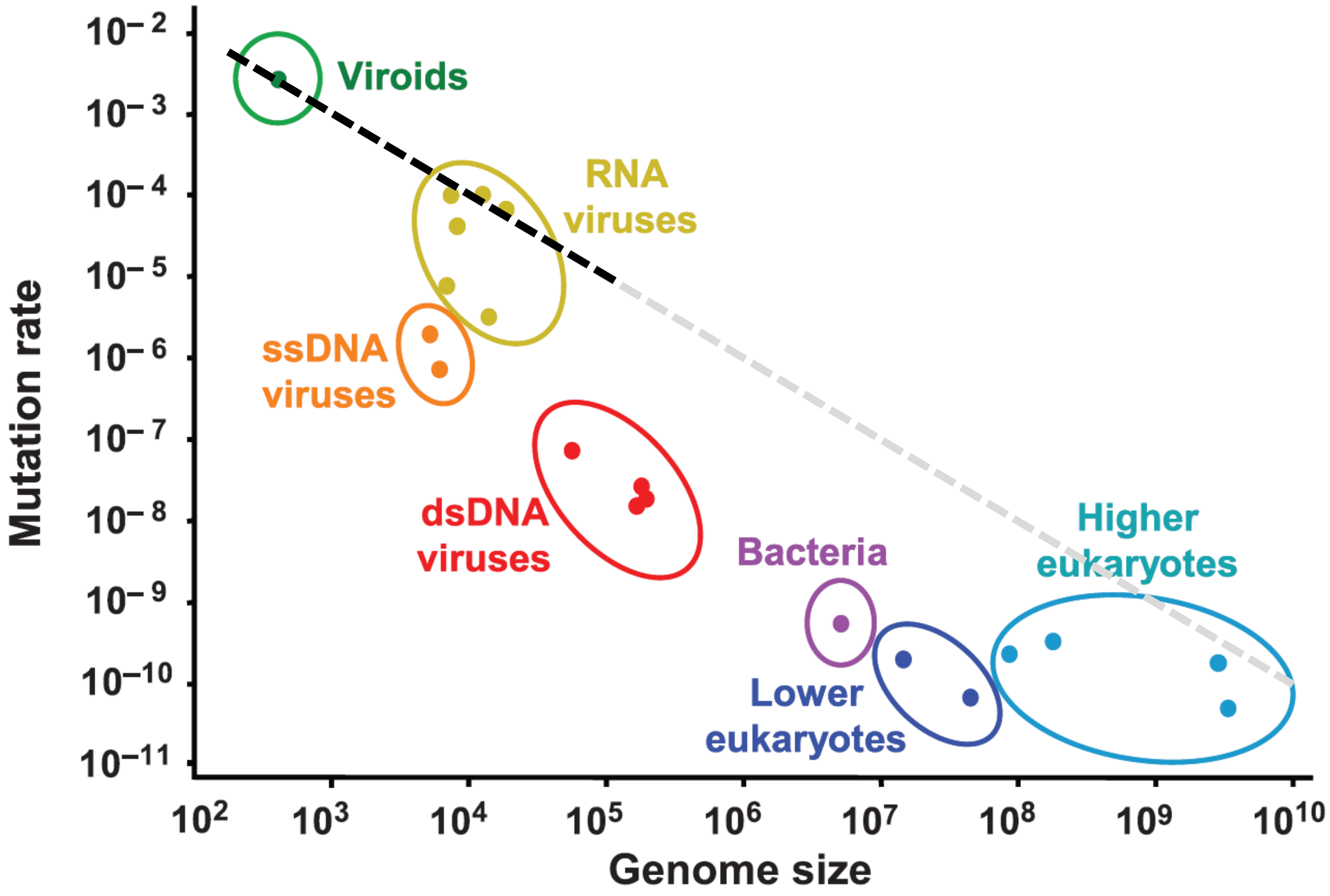

Master sequence
p = 1.00 p_cr
Mutant cloud
Concentration
Sequence
Space

$p > 1.00\ p_{cr}$


Master sequence
Mutant cloud
Concentration
Sequence
Space



The error threshold in replication and mutation



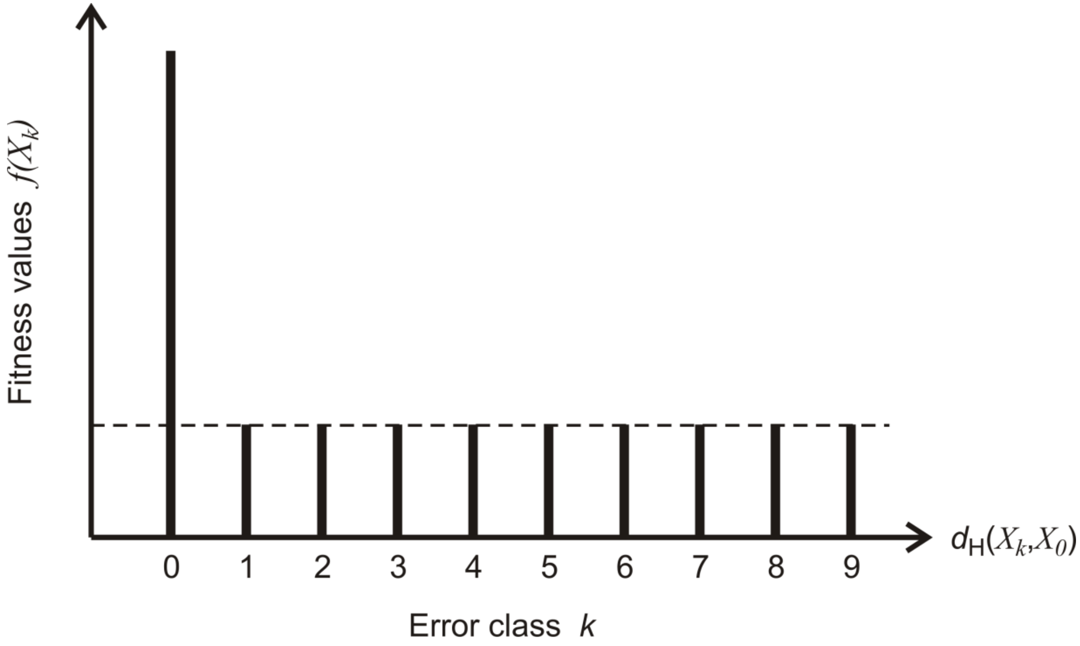
Selma Gago, Santiago F. Elena, Ricardo Flores, Rafael Sanjuán. 2009. Extremely high mutation rate of a hammerhead viroid. Science 323:1308.

Mutation rate and genome size

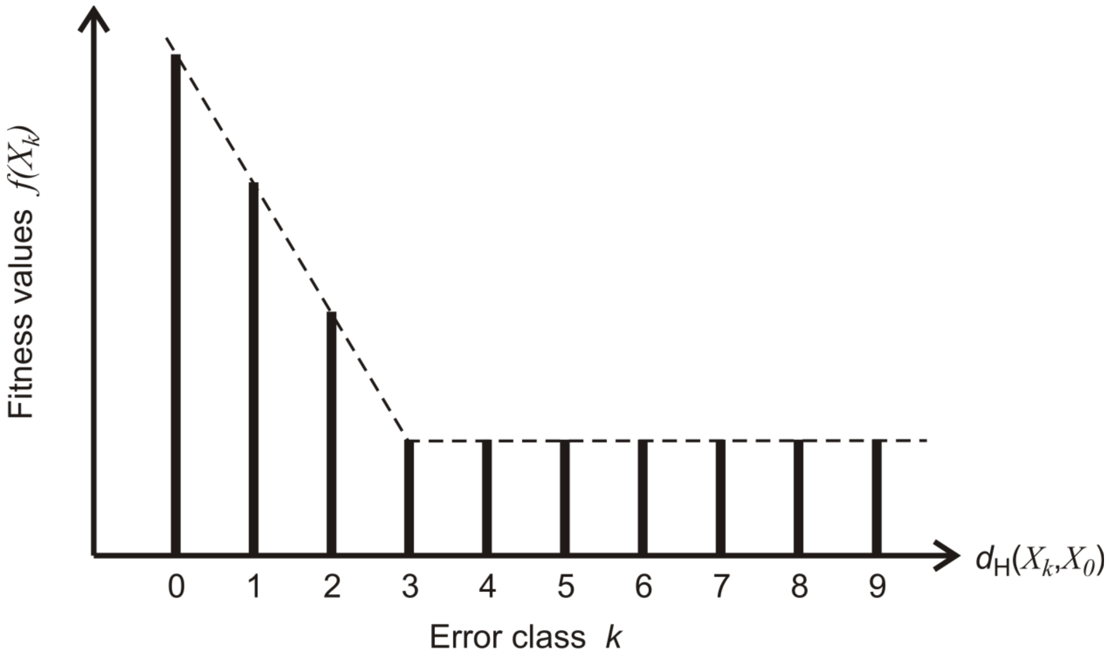
1. Prologue
2. **Landscapes and evolutionary dynamics**
3. Realistic landscapes and mutant clans
4. Mutation flow analysis
5. Neutrality in evolutionary dynamics
6. Concluding remarks and perspectives

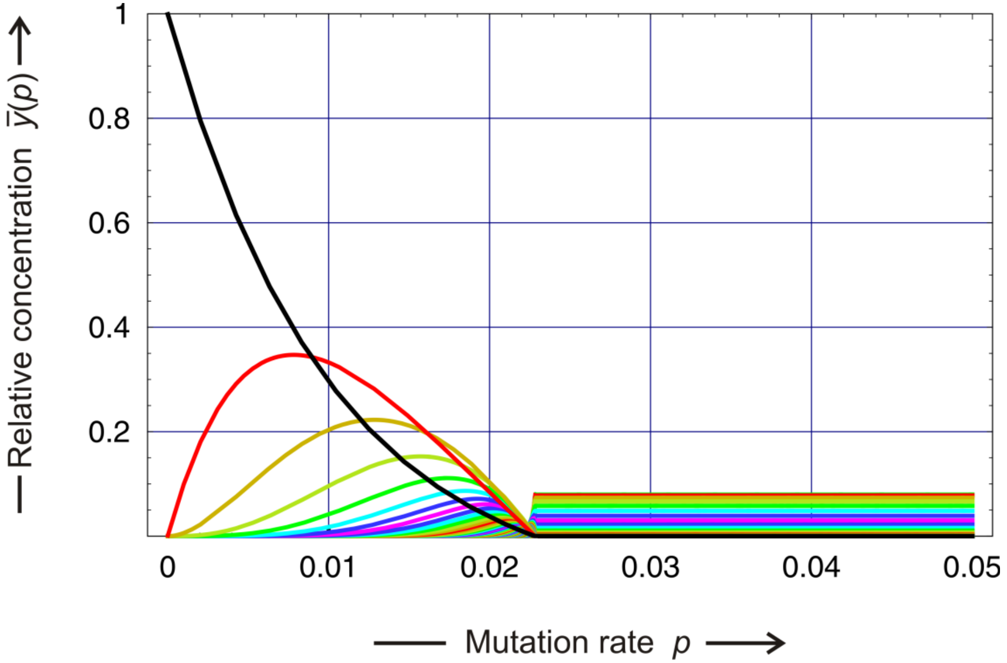
Model fitness landscapes I

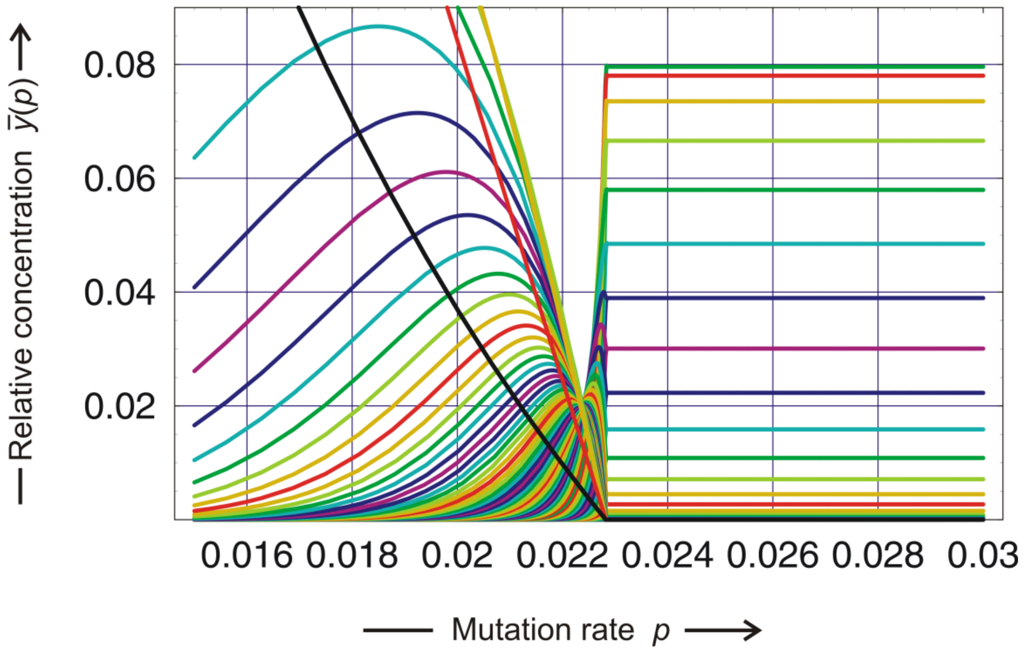
single peak landscape



step linear landscape



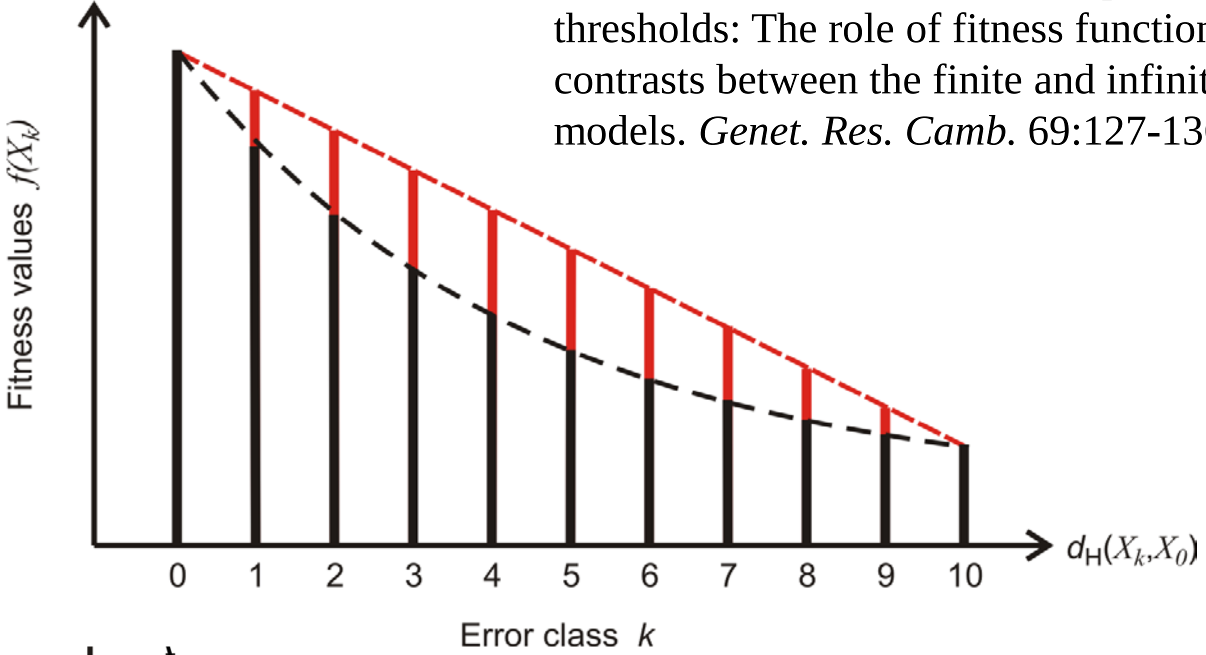




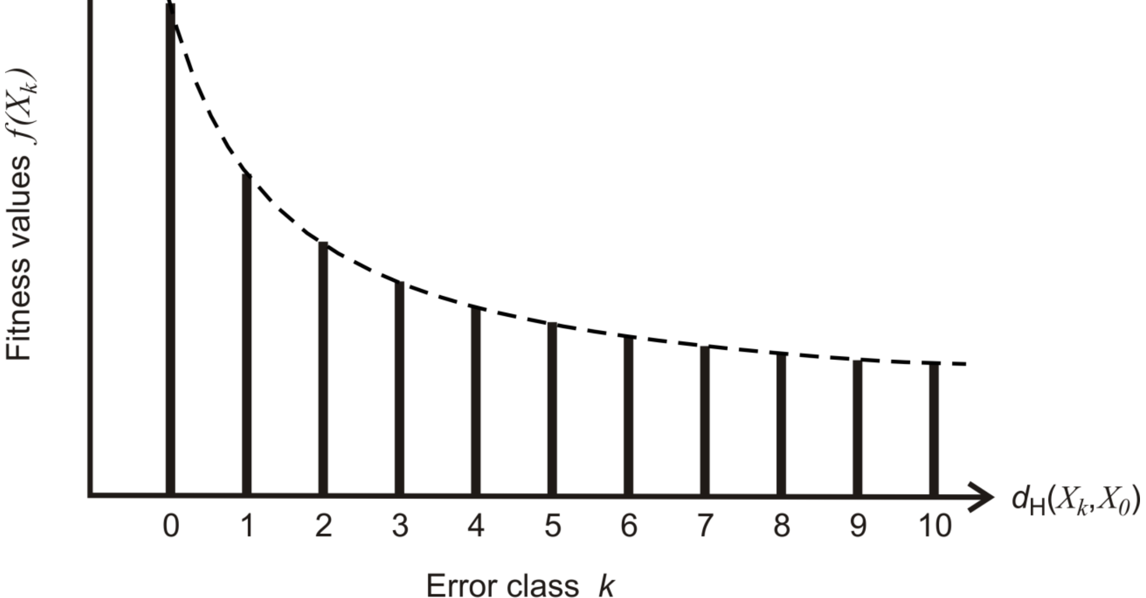
Error threshold on the single peak landscape

Thomas Wiehe. 1997. Model dependency of error thresholds: The role of fitness functions and contrasts between the finite and infinite sites models. *Genet. Res. Camb.* 69:127-136

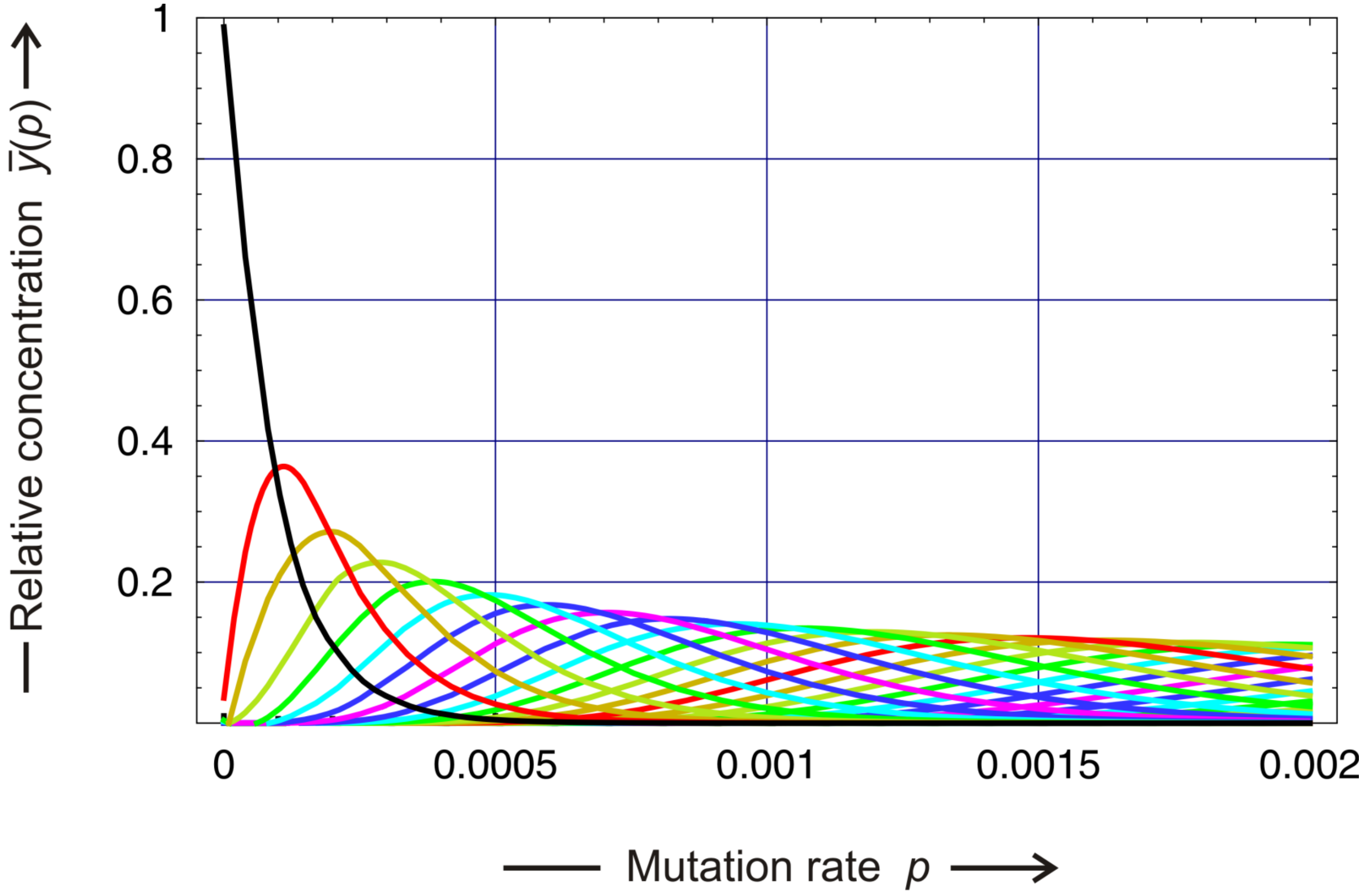
linear and multiplicative



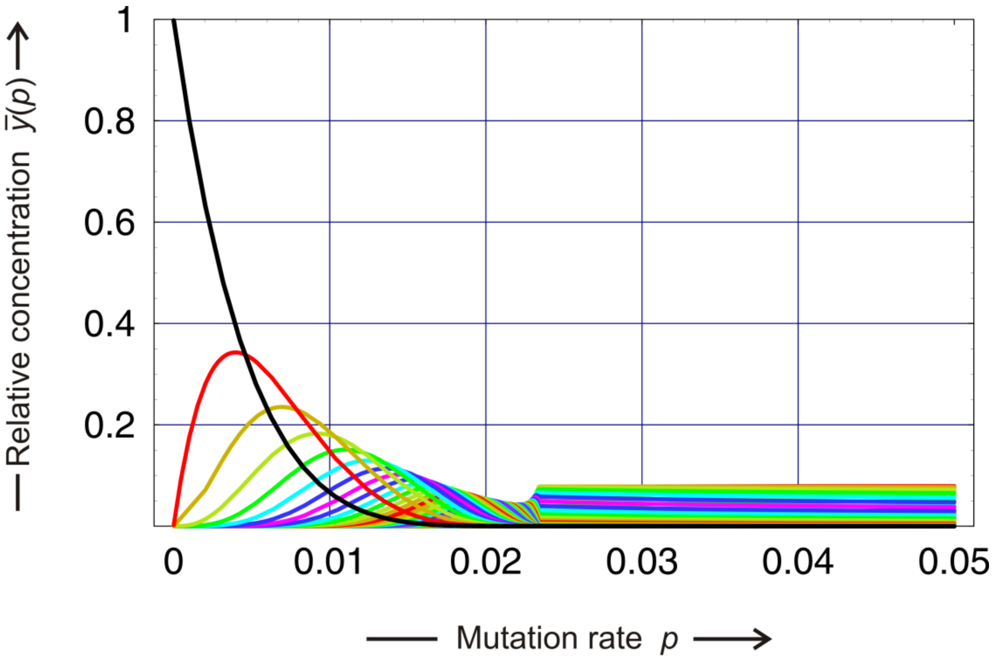
hyperbolic

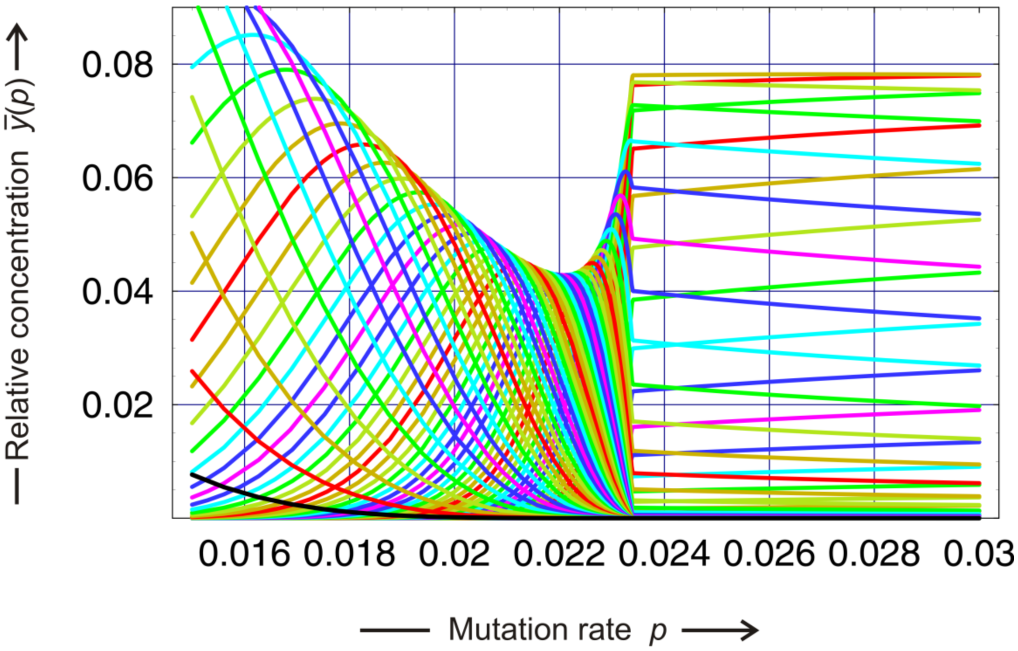


Model fitness landscapes II



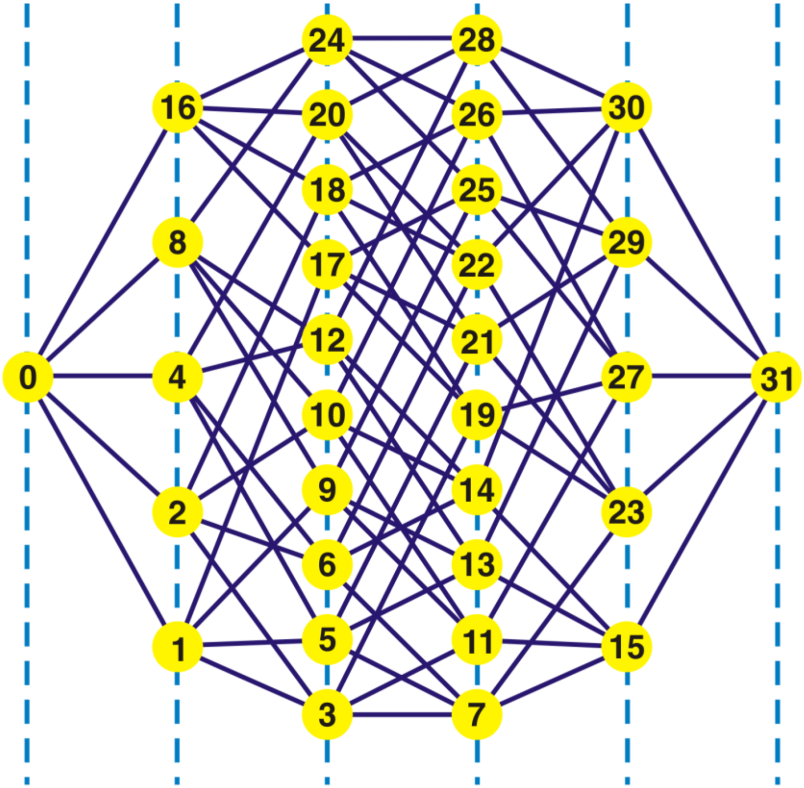
The linear fitness landscape shows no error threshold





Error threshold on the
hyperbolic landscape

1. Prologue
2. Landscapes and evolutionary dynamics
3. **Realistic landscapes and mutant clans**
4. Mutation flow analysis
5. Neutrality in evolutionary dynamics
6. Concluding remarks and perspectives



Mutant classes	0	1	2	3	4	5
# sequences	1	5	10	10	5	1

Binary sequences are encoded by their decimal equivalents:

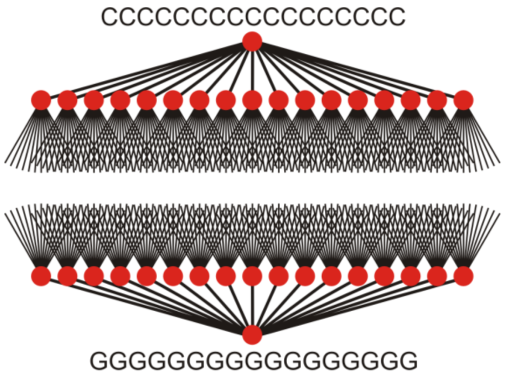
C = 0 and G = 1, for example,

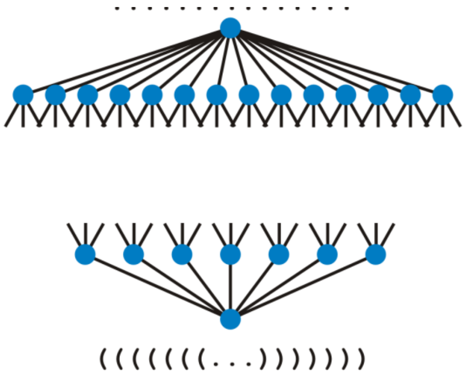
"0" ≡ 00000 = CCCCC,

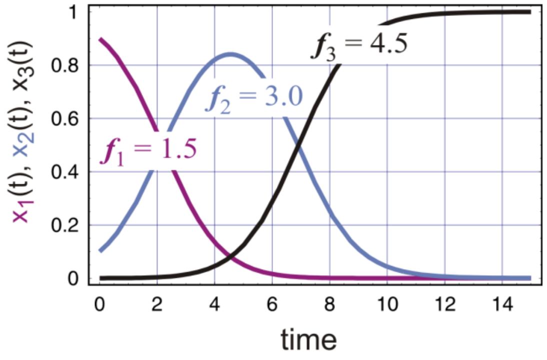
"14" ≡ 01110 = CGGGC,

"29" ≡ 11101 = GGGCG, etc.

$\mathcal{Q}_5$: the space of binary sequences of chain lenght $l = 5$







sequence space

shape space

parameter space

$\mathcal{Q}$

$\mathcal{S}$

$\mathbb{R}_+$

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

$\Psi: (\mathcal{S}, d_S) \Rightarrow \mathbb{R}_+$

$$X \longrightarrow S = \Phi(X) \longrightarrow f = \Psi(S)$$

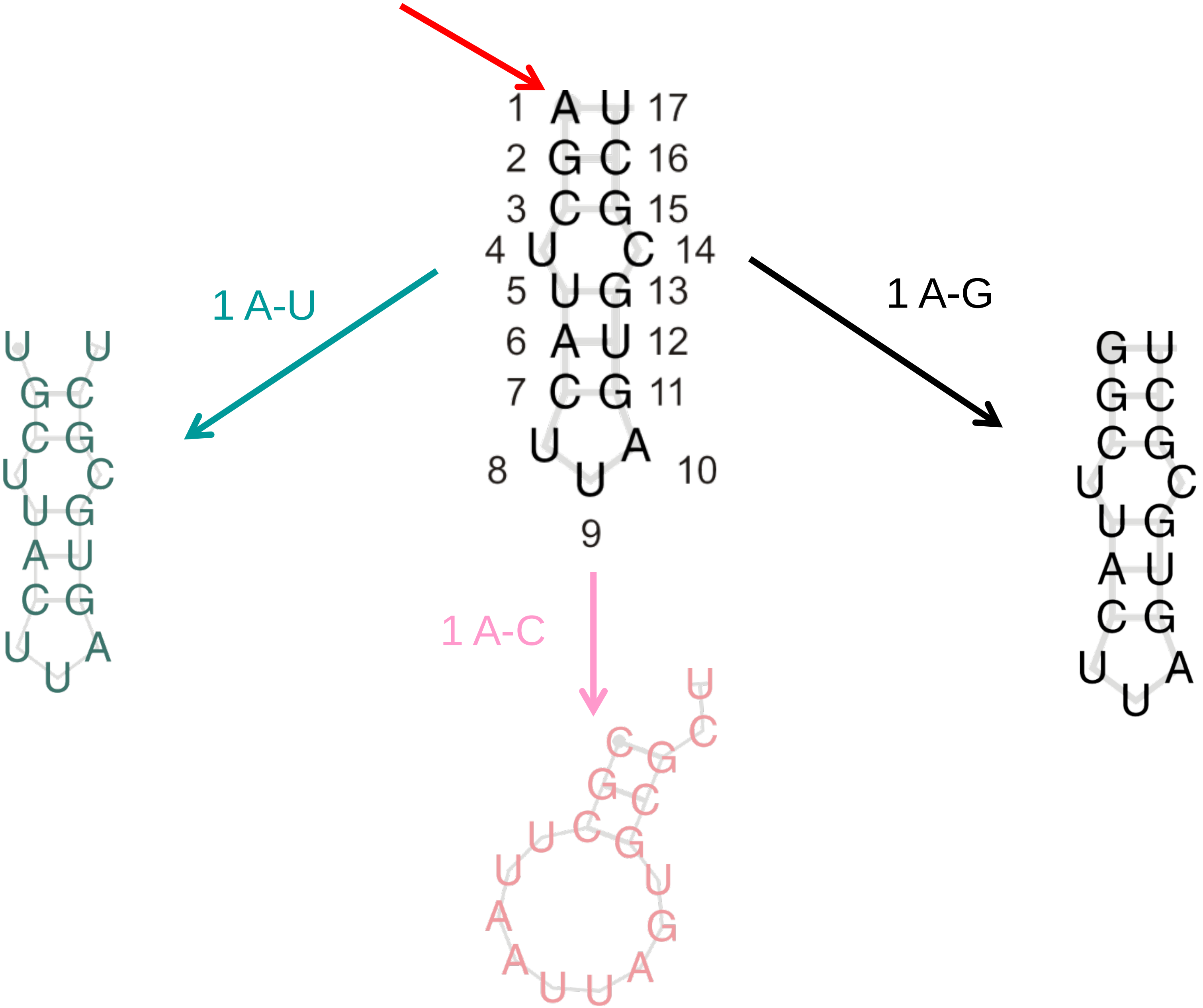
sequence

structure

function

genotype ⟶ **phenotype** ⟶ **selection**

Evolution as a global phenomenon in genotype space

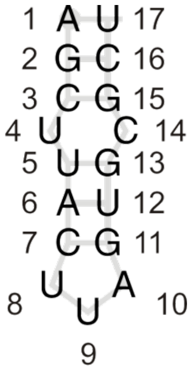


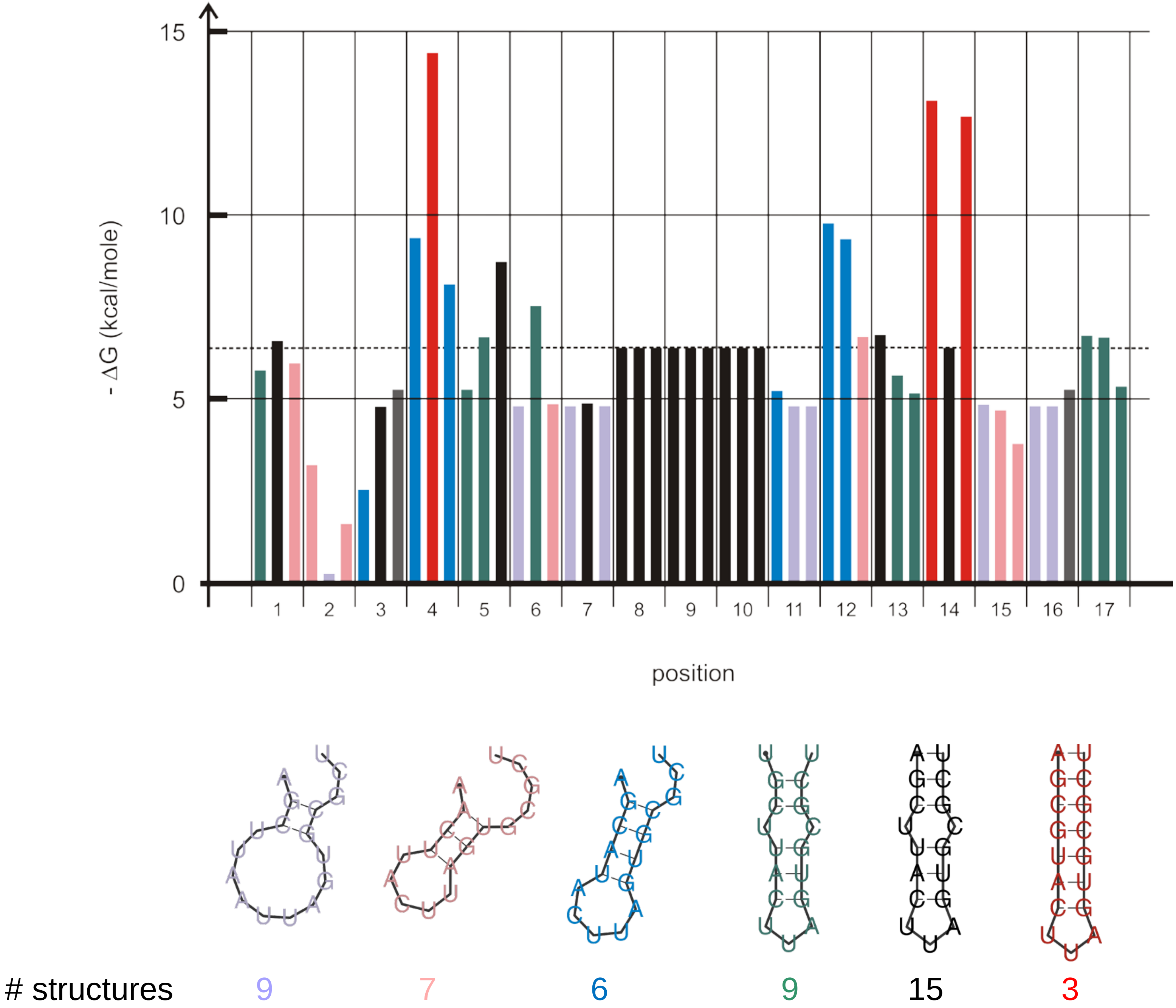
1 A-U

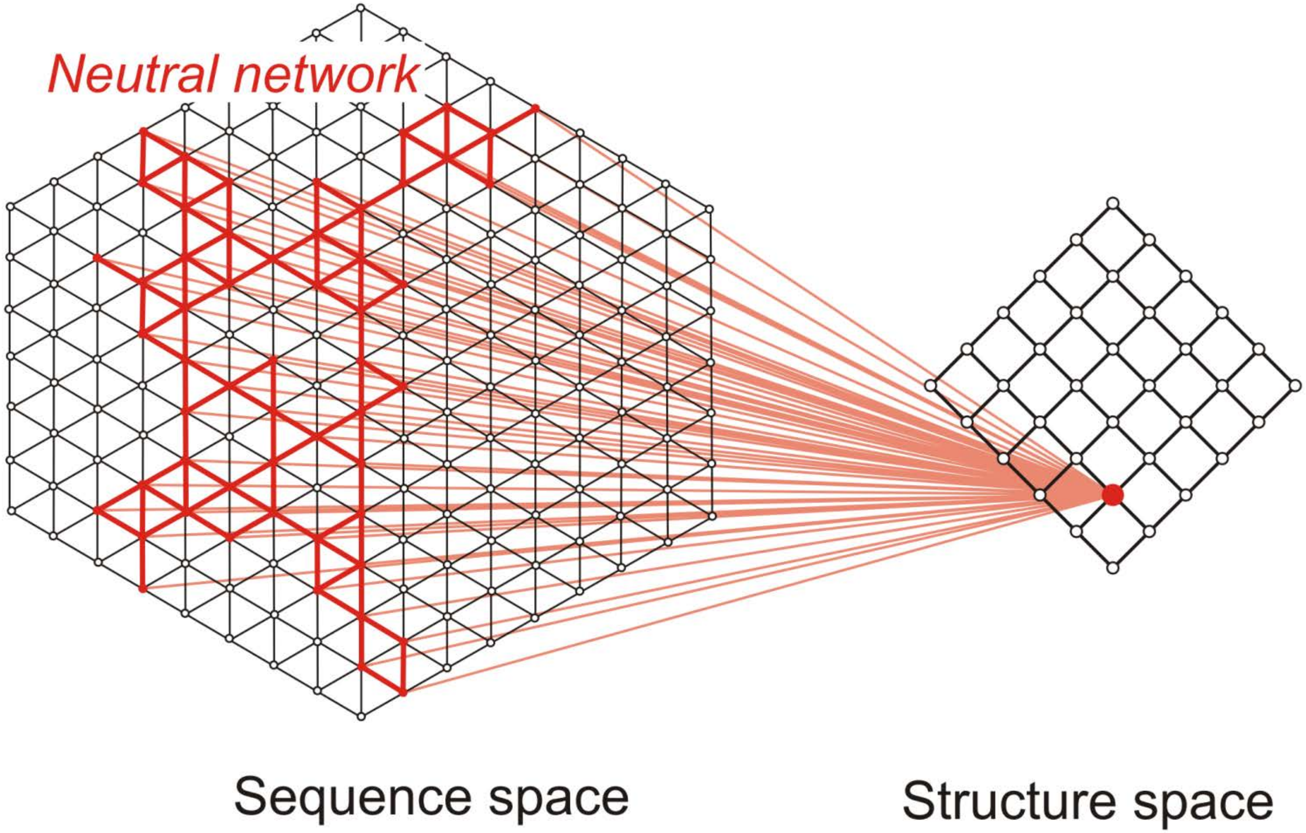
1 A-G

1 A-C

AGCUUAACUUAGUCGCU


1 A U 17
2 G C 16
3 C G 15
4 U C 14
5 U G 13
6 A U 12
7 C G 11
8 U A 10
U
9


- ΔG (kcal/mole)
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
position
structures
9
7
6
9
15
3



many genotypes ⇒ one phenotype

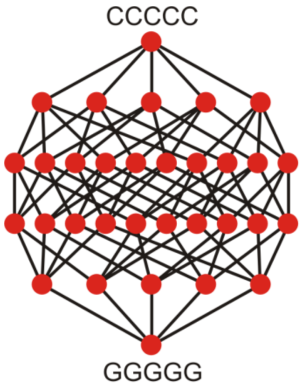
Fitness landscapes became experimentally accessible!

Protein landscapes: Yuuki Hayashi, Takuyo Aita, Hitoshi Toyota, Yuzuru Husimi, Itaru Urabe, Tetsuya Yomo. 2006. Experimental rugged fitness landscape in protein sequence space. *PLoS One* 1:e96.

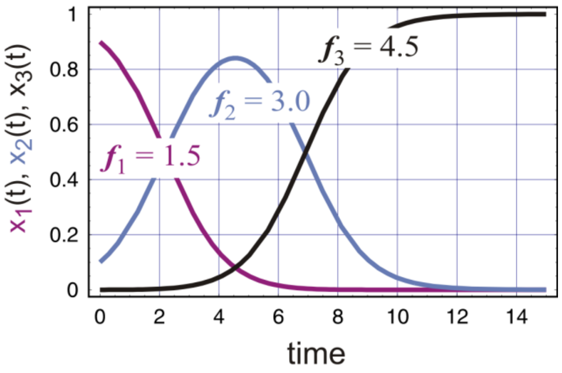
RNA landscapes: Sven Klussman, Ed. 2005. The aptamer handbook. Wiley-VCh, Weinheim (Bergstraße), DE.
Jason N. Pitt, Adrian Ferré-D'Amaré. 2010. Rapid construction of empirical RNA fitness landscapes. *Science* 330:376-379.

RNA viruses: Esteban Domingo, Colin R. Parrish, John J. Holland, Eds. 2007. Origin and evolution of viruses. Second edition. Elesvier, San Diego, CA.

Retroviruses: Roger D. Kouyos, Gabriel E. Leventhal, Trevor Hinkley, Mojgan Haddad, Jeannette M. Whitcomb, Christos J. Petropoulos, Sebastian Bonhoeffer. 2012. Exploring the complexity of the HIV-I fitness landscape. *PLoS Genetics* 8:e1002551



sequence space



parameter space

$$S \longrightarrow f = \Psi(Y)$$

sequence

function

The simplified model

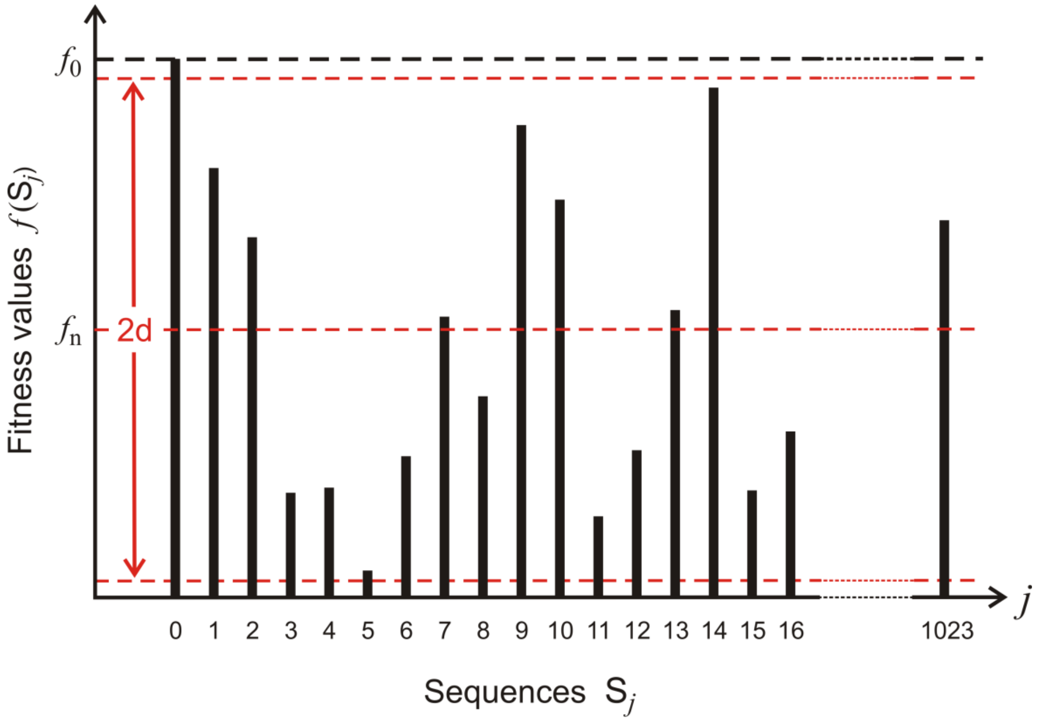
$$f(S_j) = f_n + 2d\,(f_0 - f_n)\left(\eta_j^{(s)} - 0.5\right)$$

$$j = 1, 2, \ldots, N\,;\; j \neq m$$

$\eta \ldots$ random number

$s \ldots$ seeds

„realistic" landscape



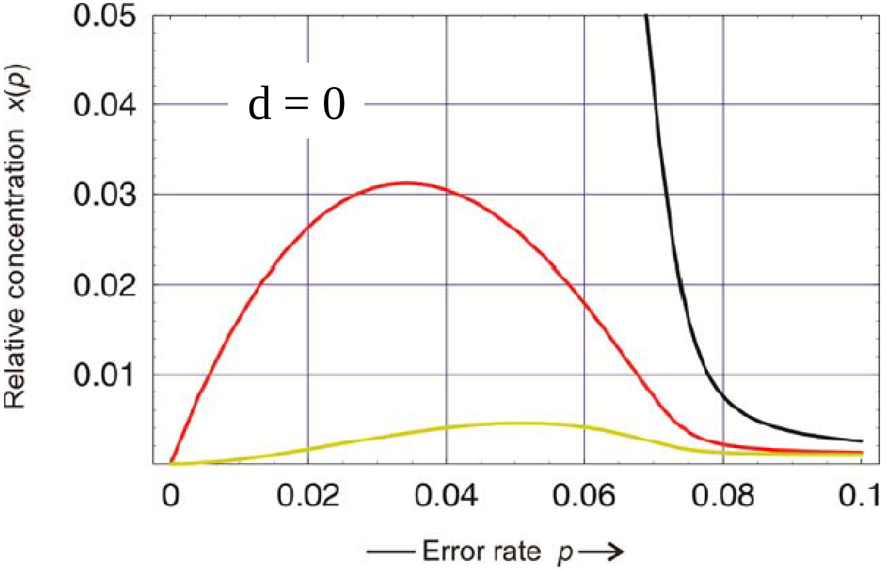
"experimental computer biology":

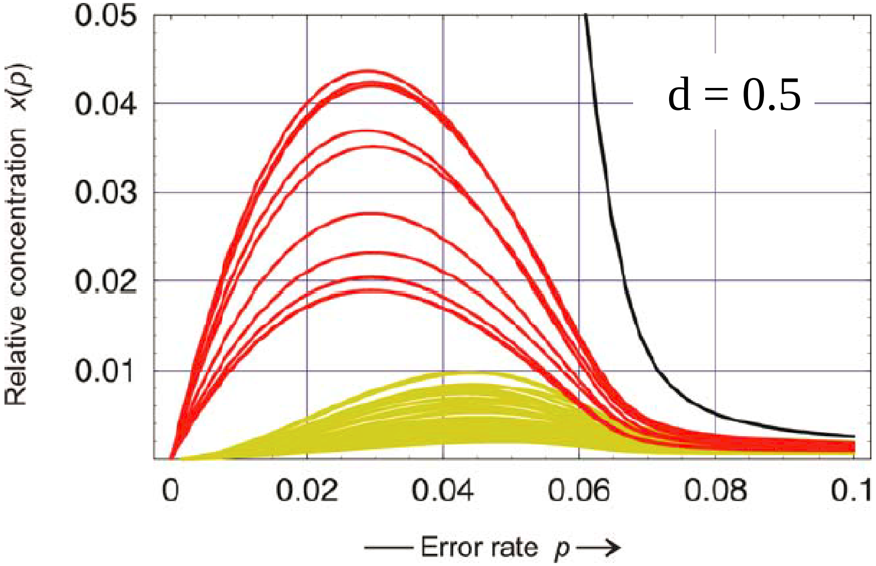
(i) choose seeds, e.g., s ∈ {000, … , 999},

(ii) compute landscape, $f(S_j)$, $j = 1, \ldots, N$,

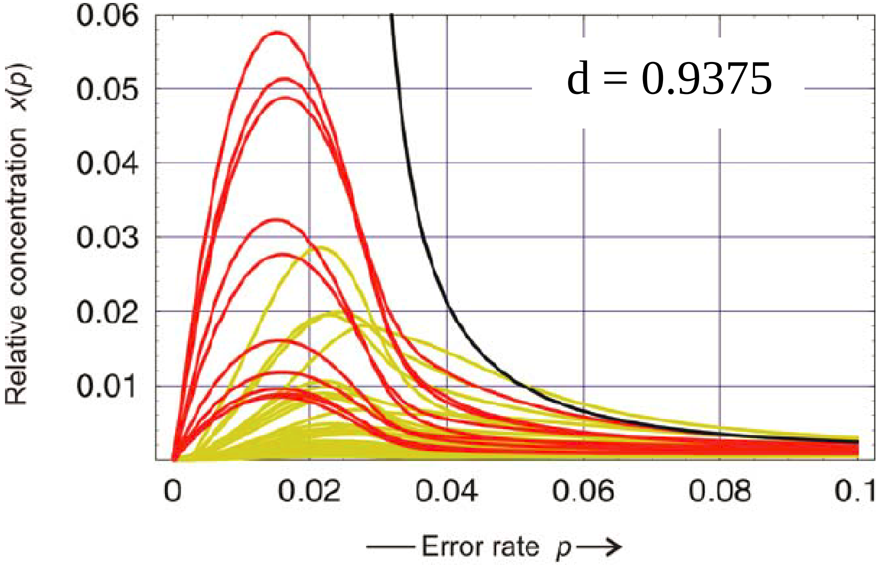
(iii) compute and analyze quasispecies, $\Upsilon(p,d)$

Rugged fitness landscapes over individual binary sequences with n = 10



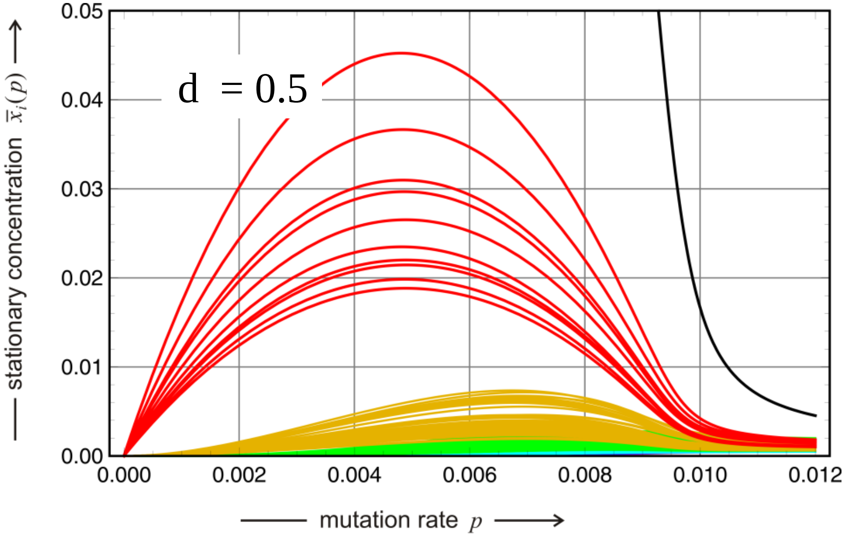


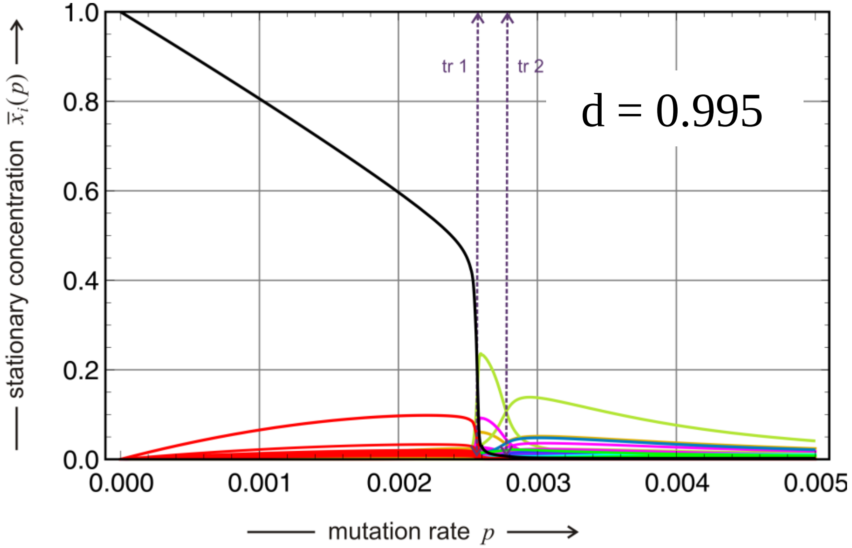
Quasispecies with increasing random scatter *d*



Error threshold: Individual sequences

n = 10, σ = 2, s = 491 and d = 0, 0.5, 0.9375



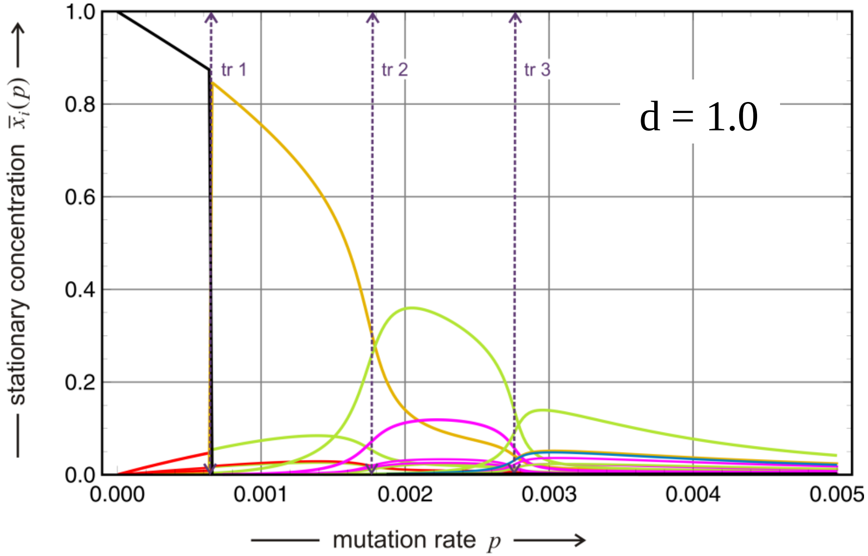


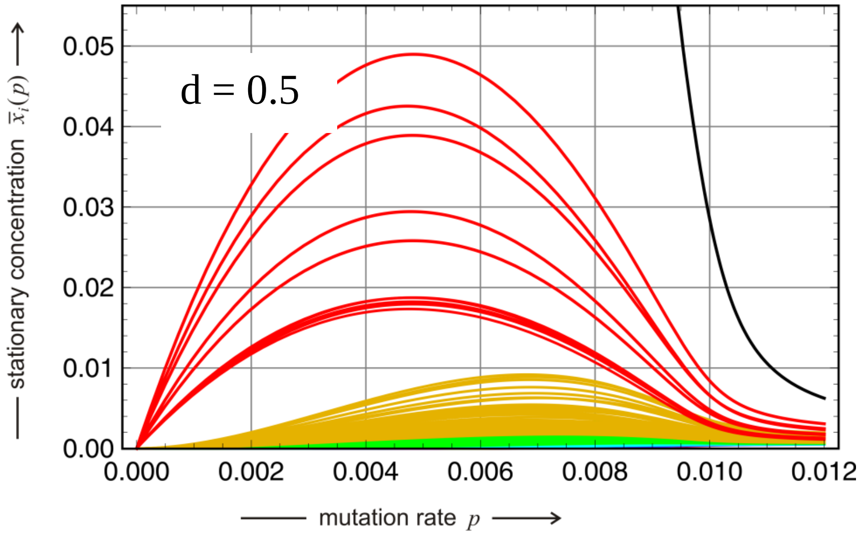
Choice of random scatter:

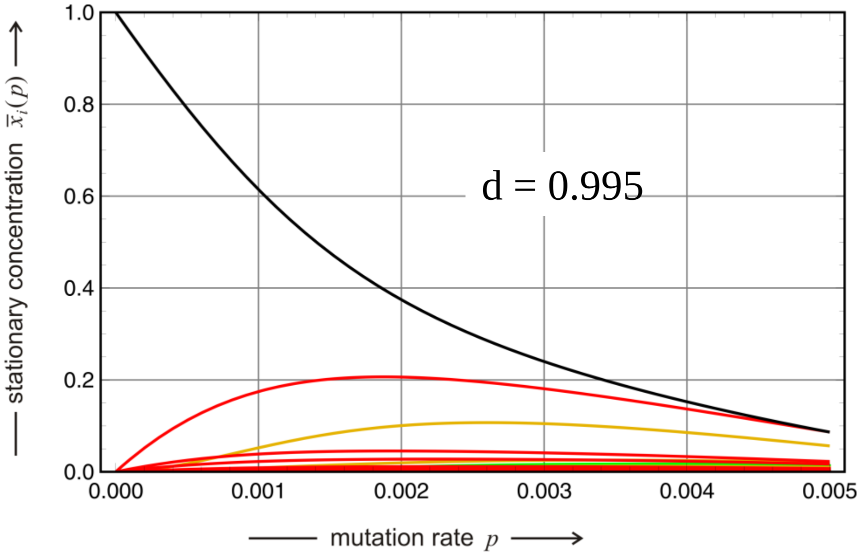
s = 637

Error threshold on ‚realistic‘ landscapes

n = 10, f_0 = 1.1, f_n = 1.0, s = 637

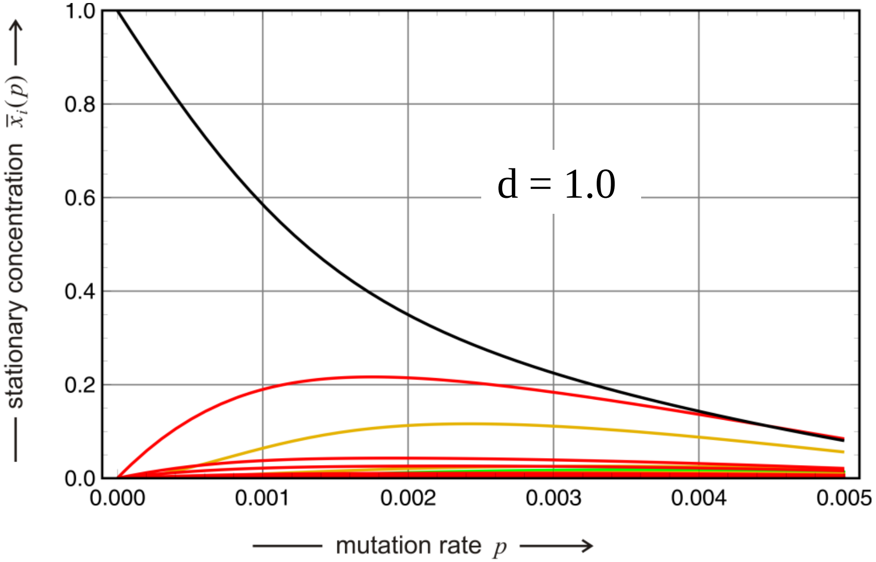






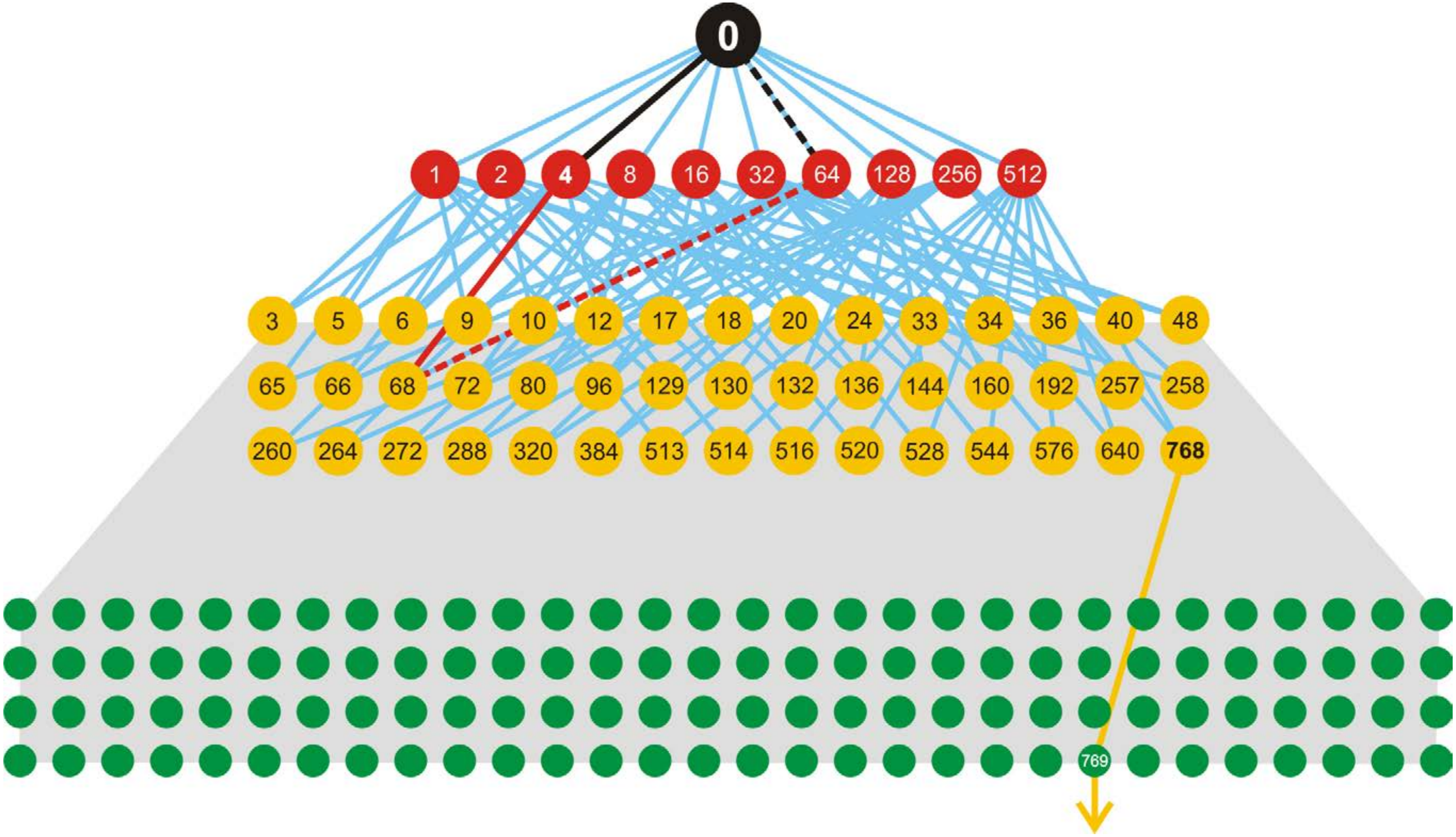
Choice of random scatter:

s = 919

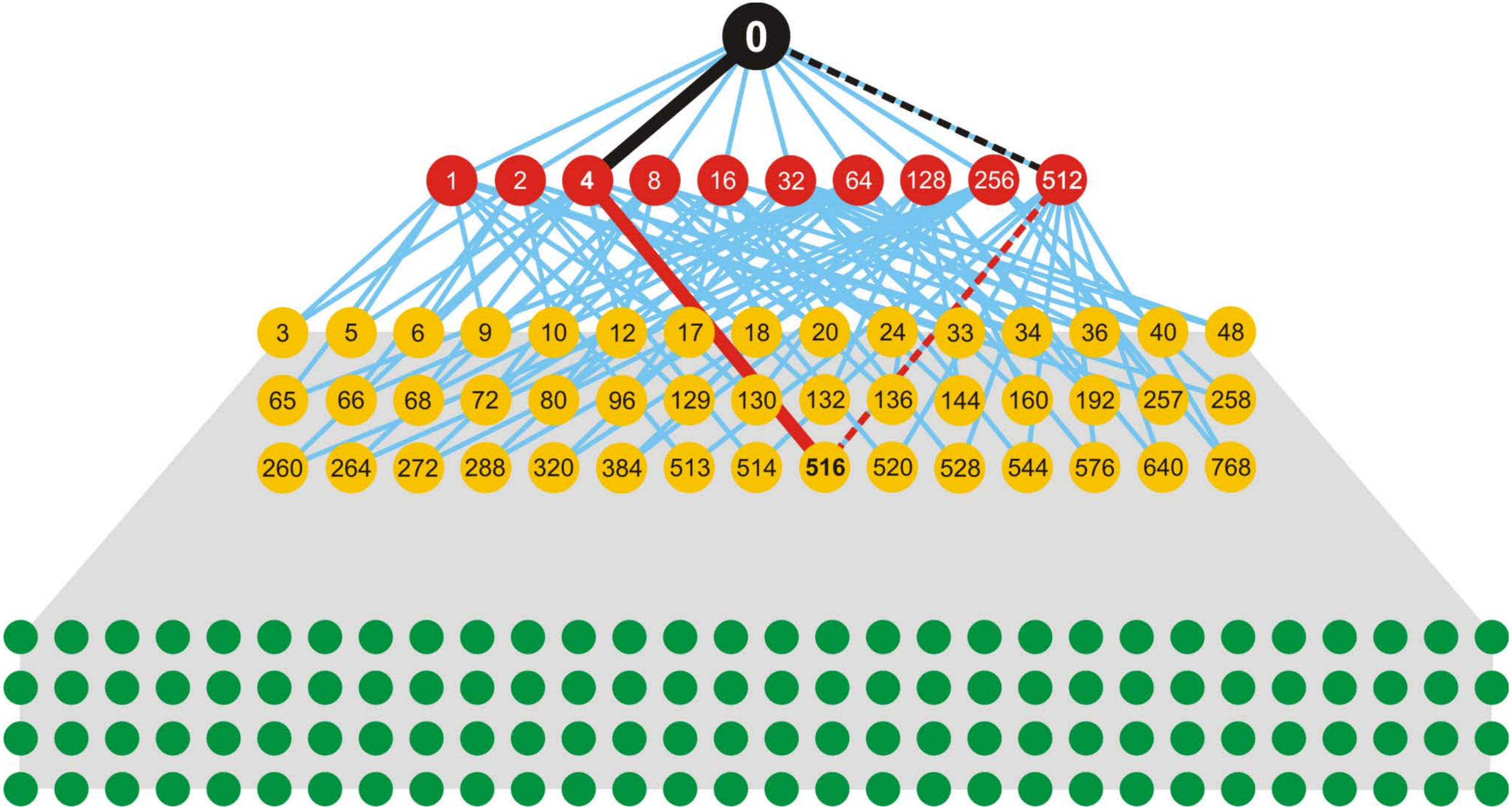


Error threshold on ‚realistic‘ landscapes

n = 10, $f_0 = 1.1$, $f_n = 1.0$, s = 919



Determination of the dominant mutation flow: $d = 1$, $s = 613$



Determination of the dominant mutation flow: $d = 1$, $s = 919$

Predictions of the strong quasispecies concept

1. A strong quasispecies is dominated by a **clan** of mutationally coupled closely related sequences.

2. A four-membered clan consists of the **master sequence** being the fittest sequence, its **fittest one error mutant**, the **fittest two-error mutant** that **has to lie in the one-error neighborhood of the fittest one-error mutant**, and the **fourth sequence** completing the mutationally coupled quartet.

3. A strong quasispecies is stable against changes in the mutation rate and hence provides an evolutionary advantage over conventional quasispecies.

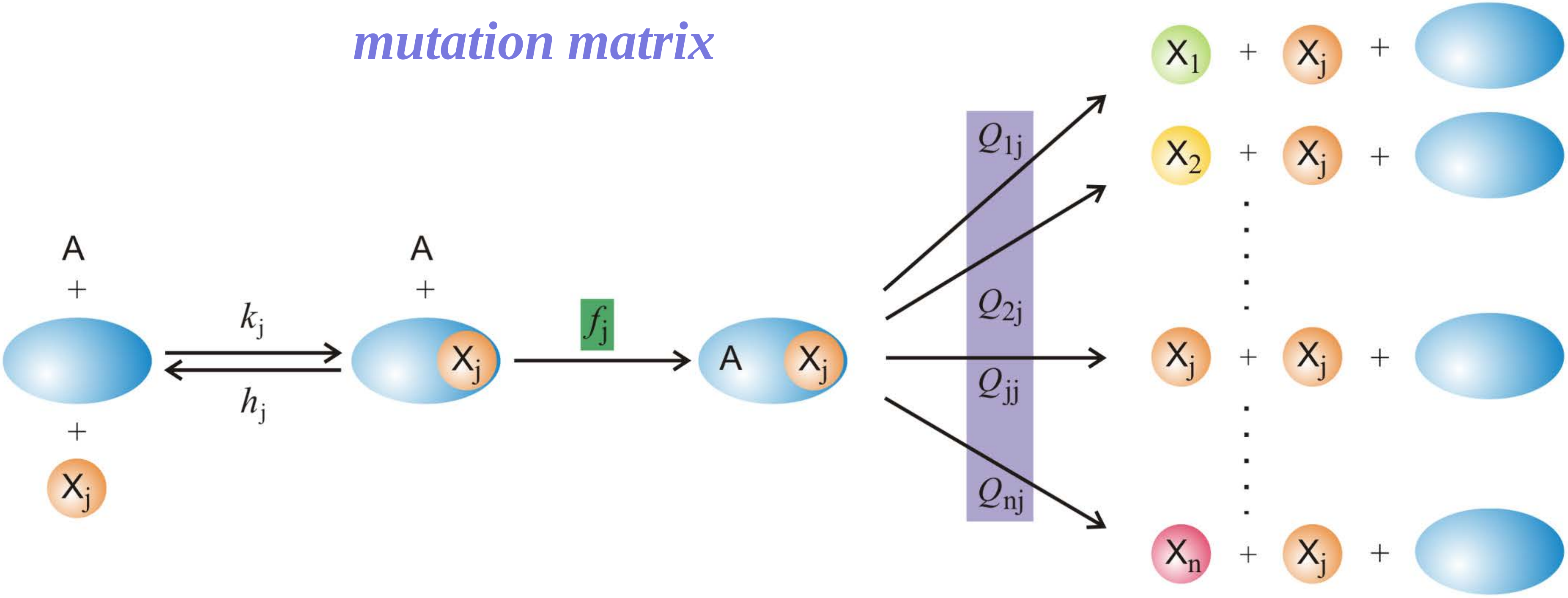
1. Prologue
2. Landscapes and evolutionary dynamics
3. Realistic landscapes and mutant clans
4. **Mutation flow analysis**
5. Neutrality in evolutionary dynamics
6. Concluding remarks and perspectives

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = \sum\nolimits_{i=1}^{n} W_{ji}\, x_i - x_j\, \Phi\,; \quad j = 1, 2, \ldots, n$$

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

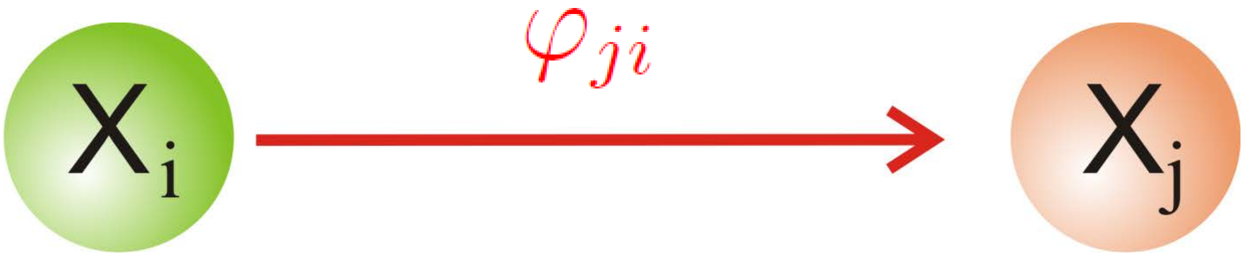
fitness landscape

mutation matrix



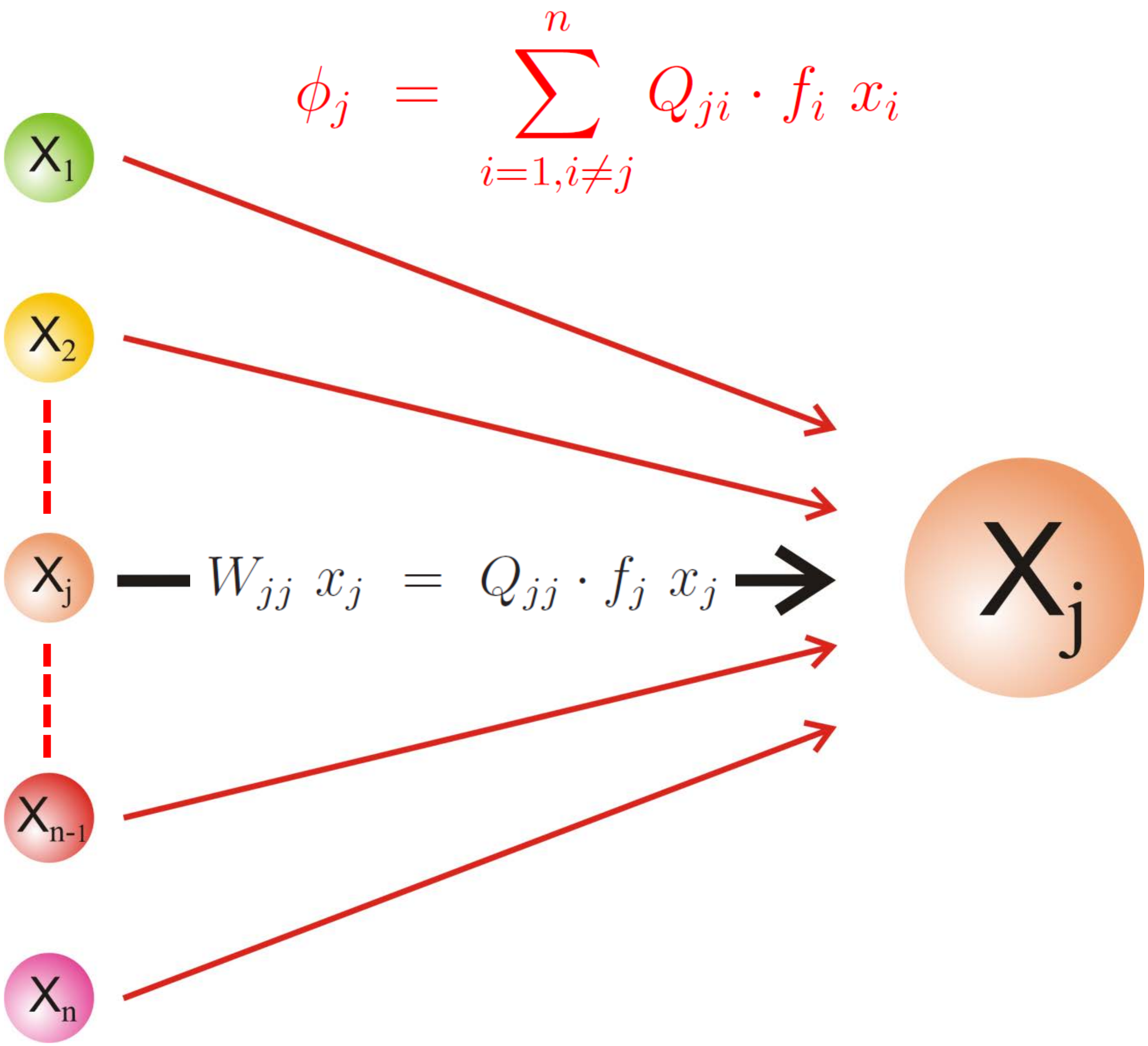
Correct replication and mutation as parallel chemical reactions

$$\varphi_{ji} = Q_{ji}\cdot f_i\,x_i\,;\; Q_{ji} \approx (1-p)^{l-\mathrm{d}^{\mathrm{H}}_{ji}}\cdot p^{\mathrm{d}^{\mathrm{H}}_{ji}}$$



$$\phi_j = \sum_{i=1,i\neq j}^{n} \varphi_{ji} = \sum_{i=1,i\neq j}^{n} Q_{ji}\cdot f_i\,x_i$$

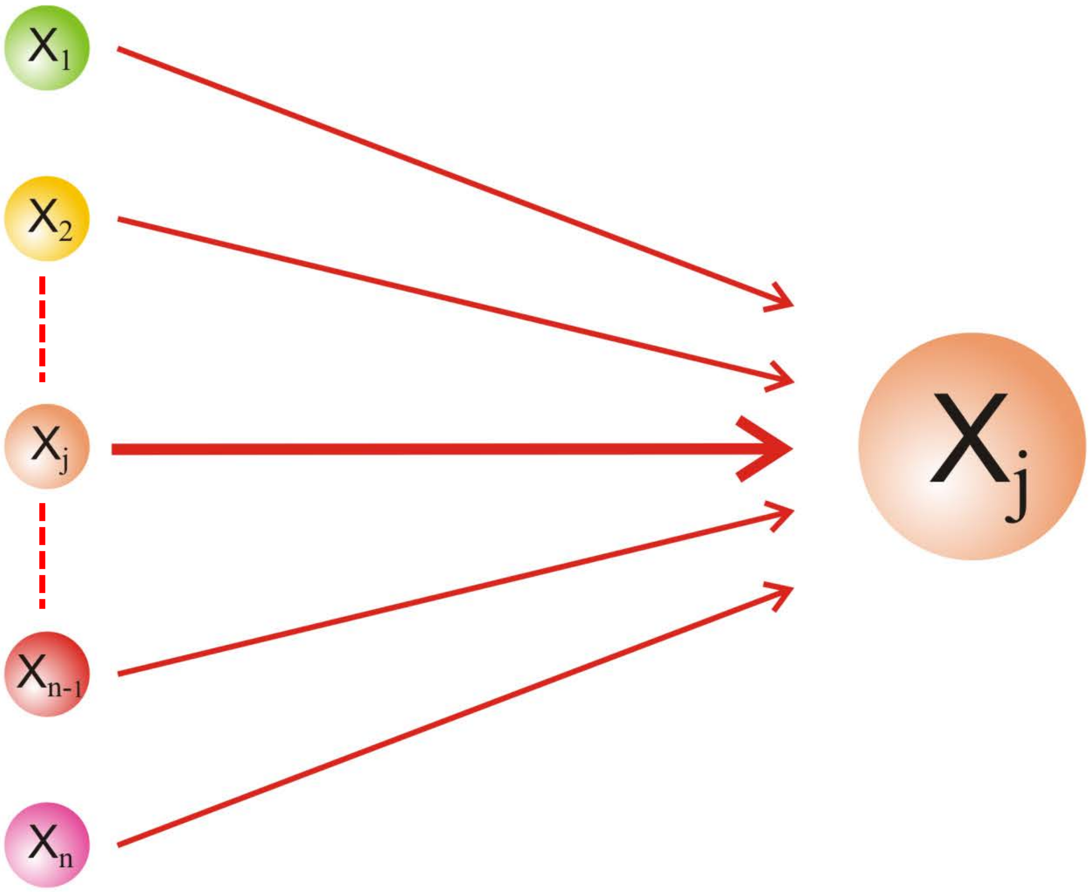
Mutation flow component and mutation flow



$$\phi_j = \sum_{i=1, i\neq j}^{n} Q_{ji} \cdot f_i \, x_i$$

$$W_{jj} \, x_j = Q_{jj} \cdot f_j \, x_j$$

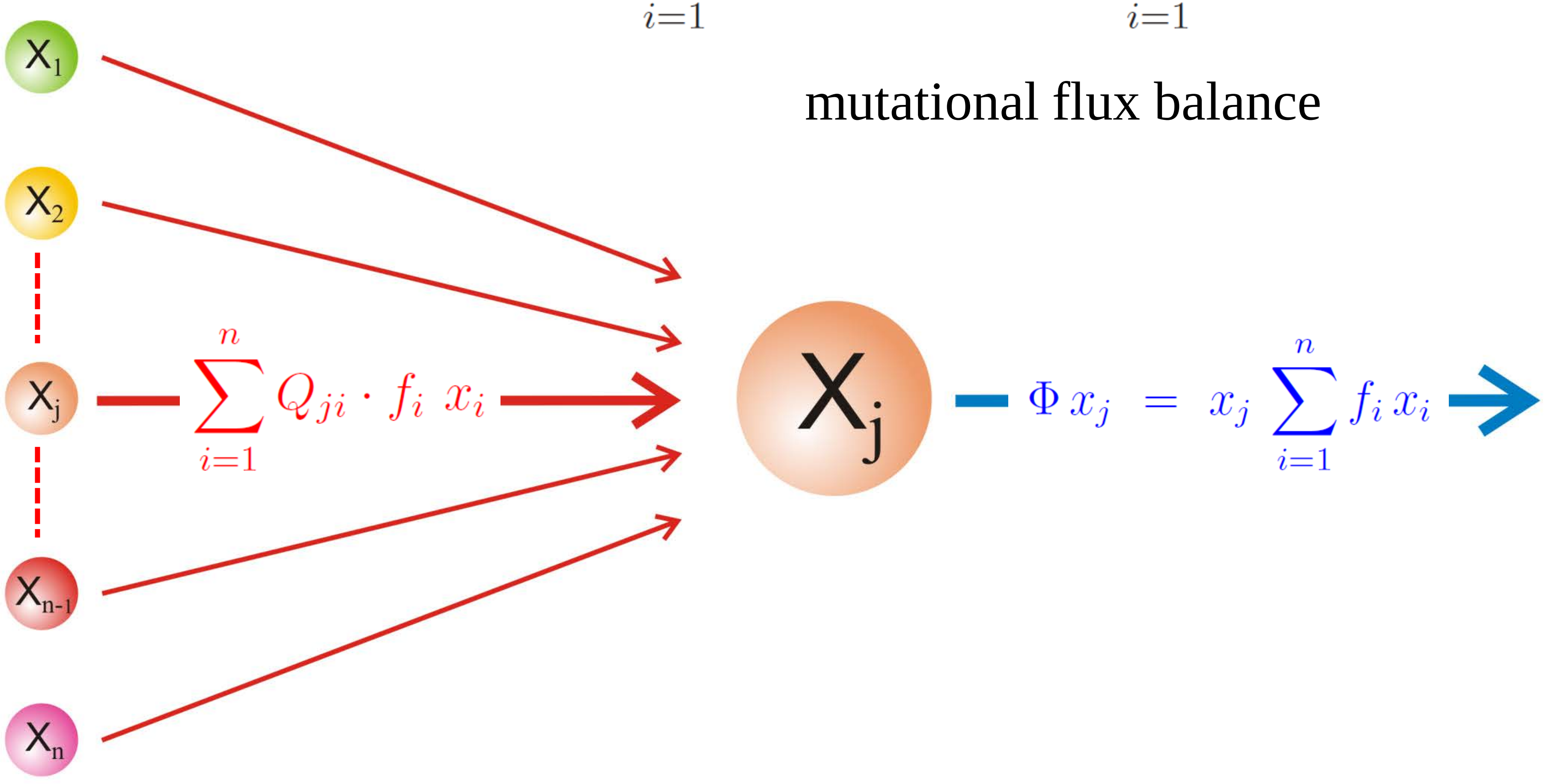
Definition of the mutation flow



Mutational flux balance and quasispecies

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

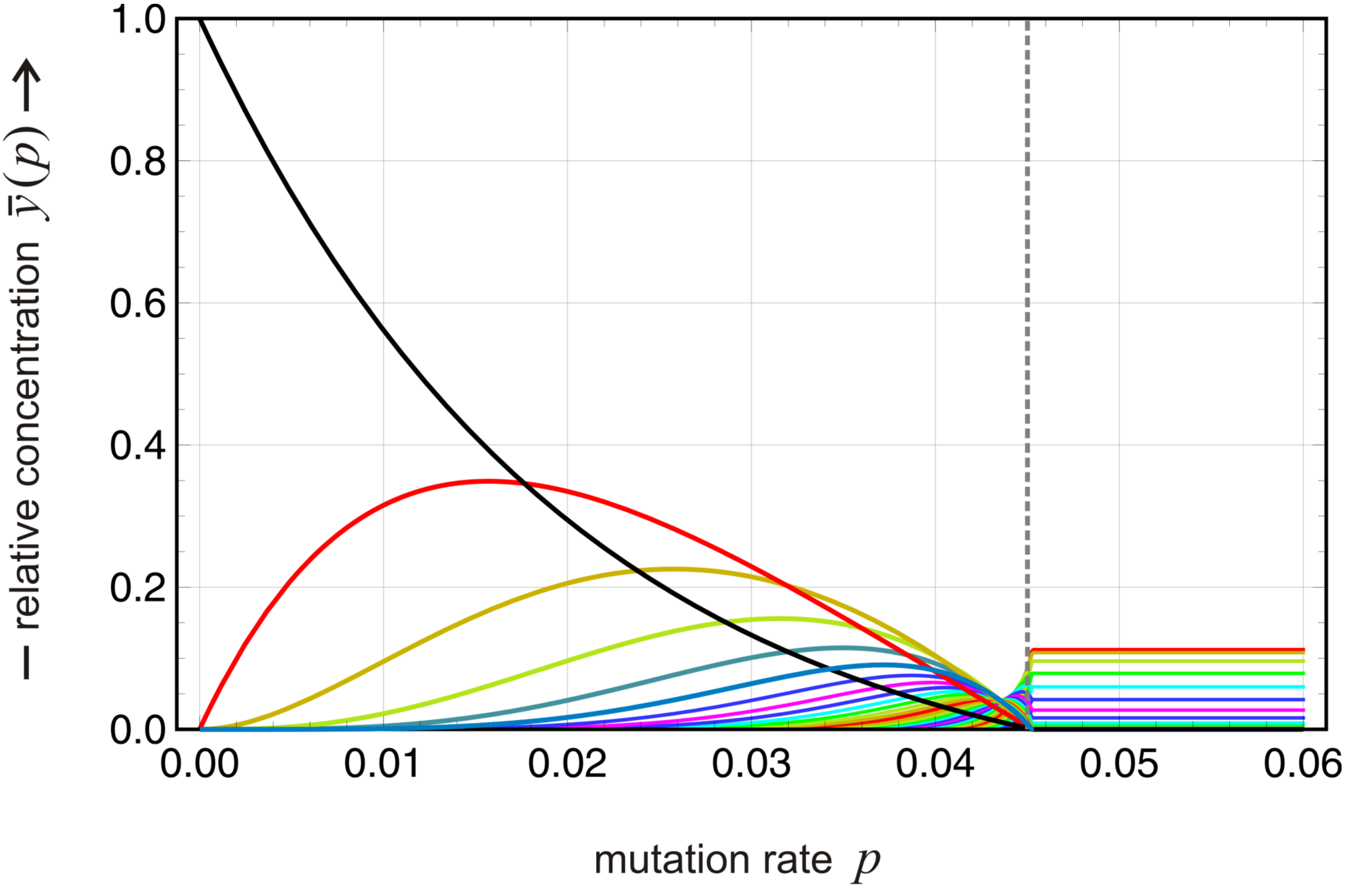
mutational flux balance



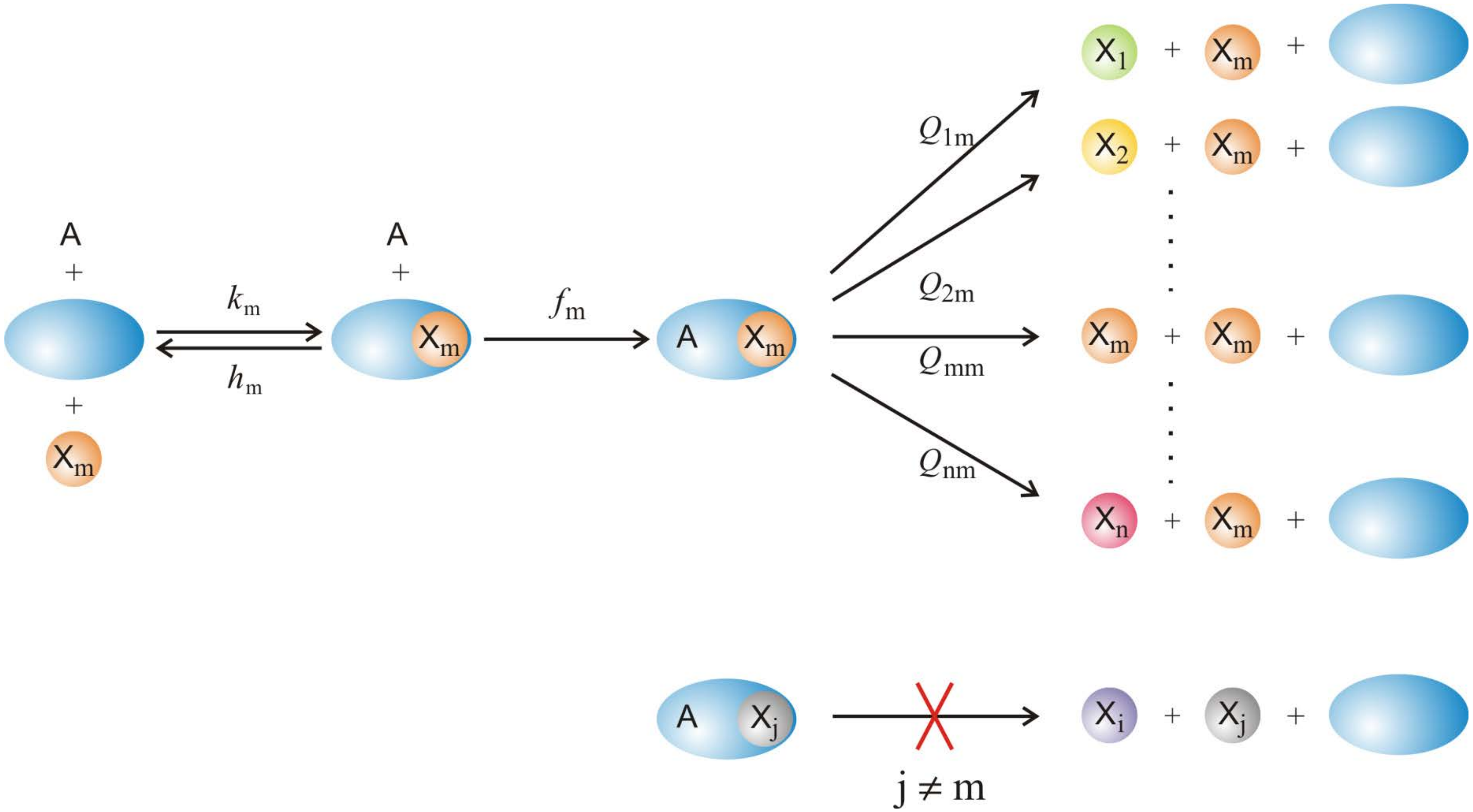
$$\sum_{i=1}^{n} Q_{ji} \cdot f_i \; x_i$$

$$\Phi \, x_j \; = \; x_j \sum_{i=1}^{n} f_i \, x_i$$

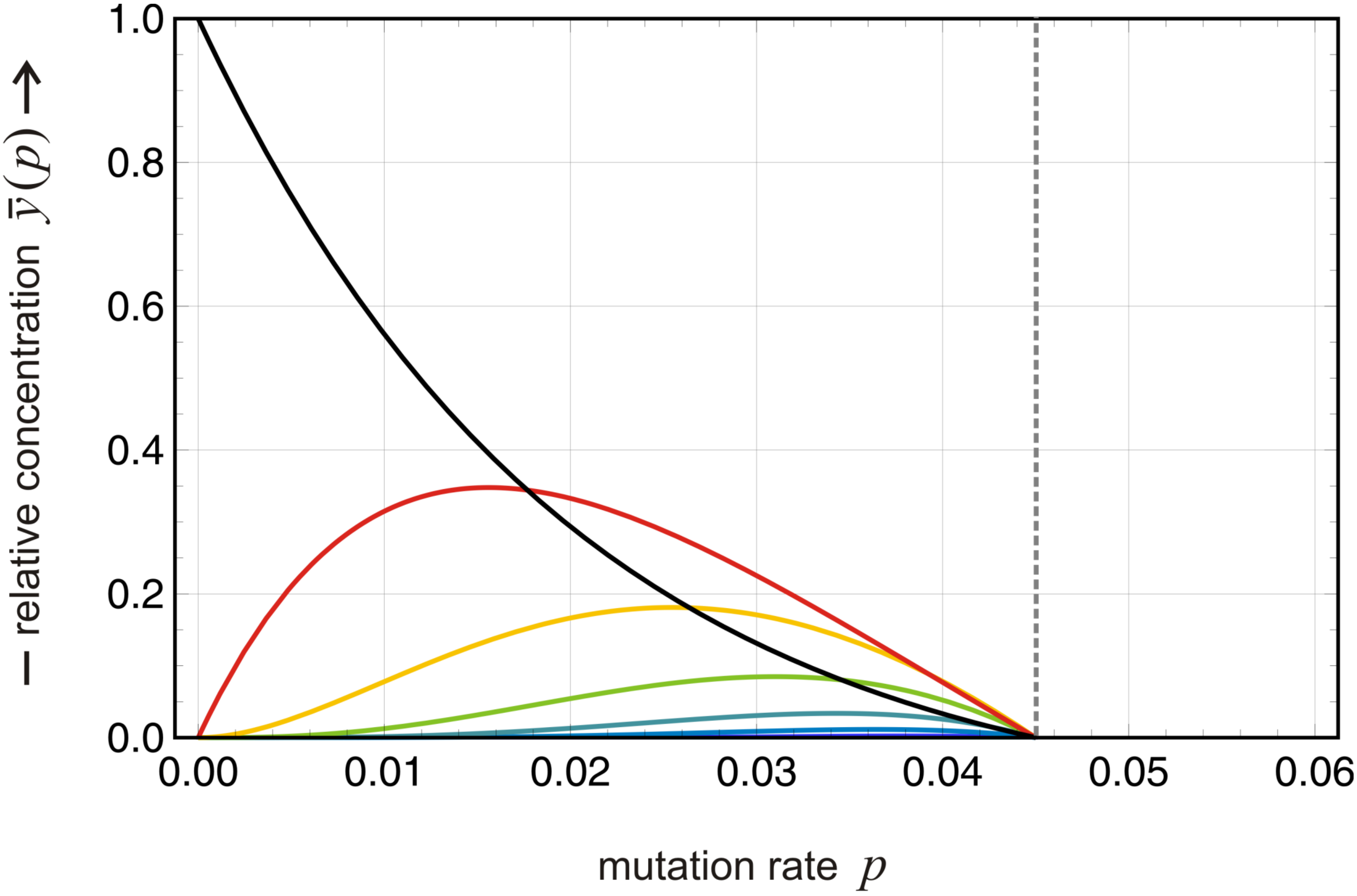
Mutational flux balance and quasispecies



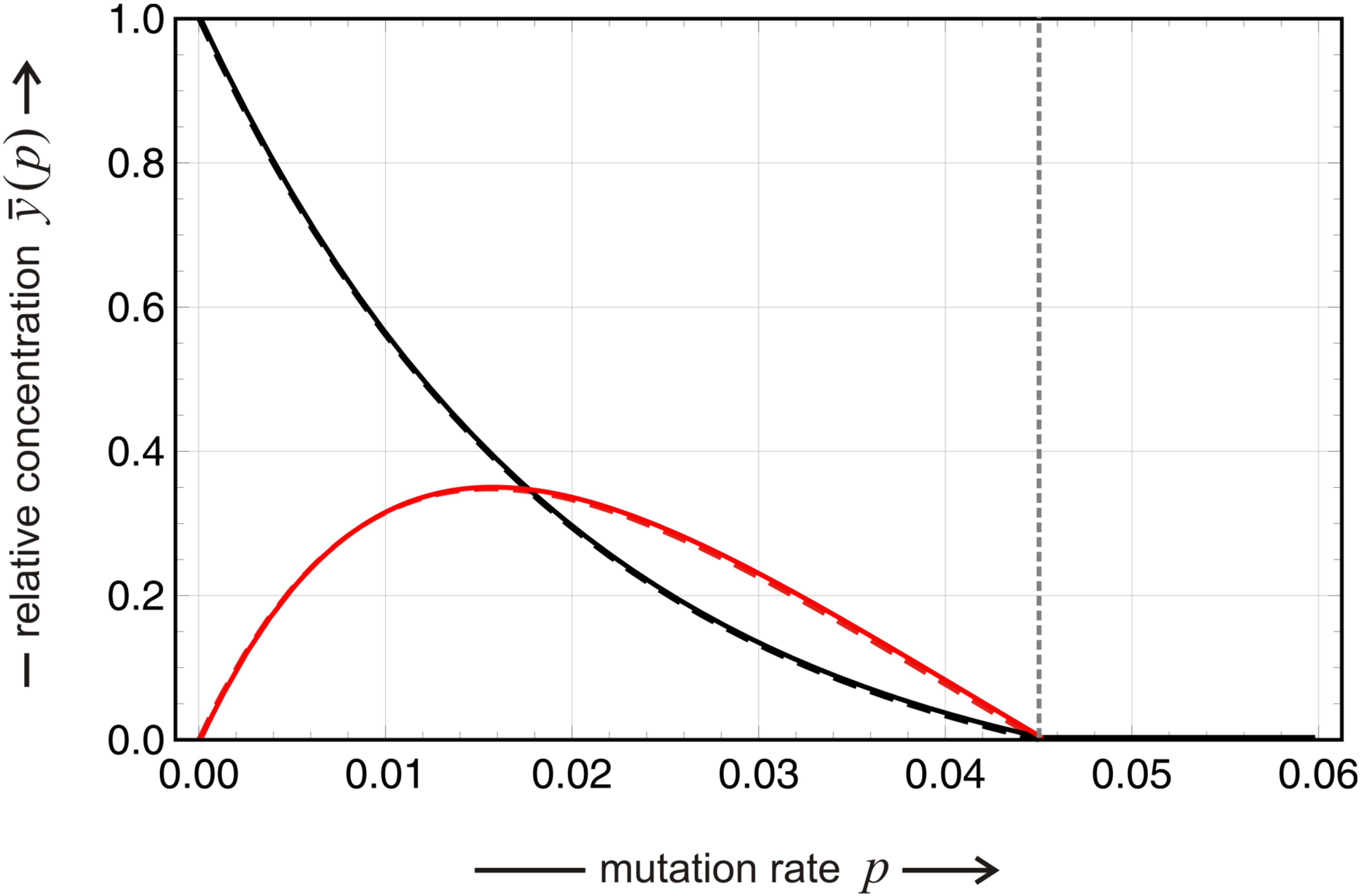
Exact quasispecies: $l = 50, f_0 = 10, f_j = 1 \;\forall\; j \neq 0$



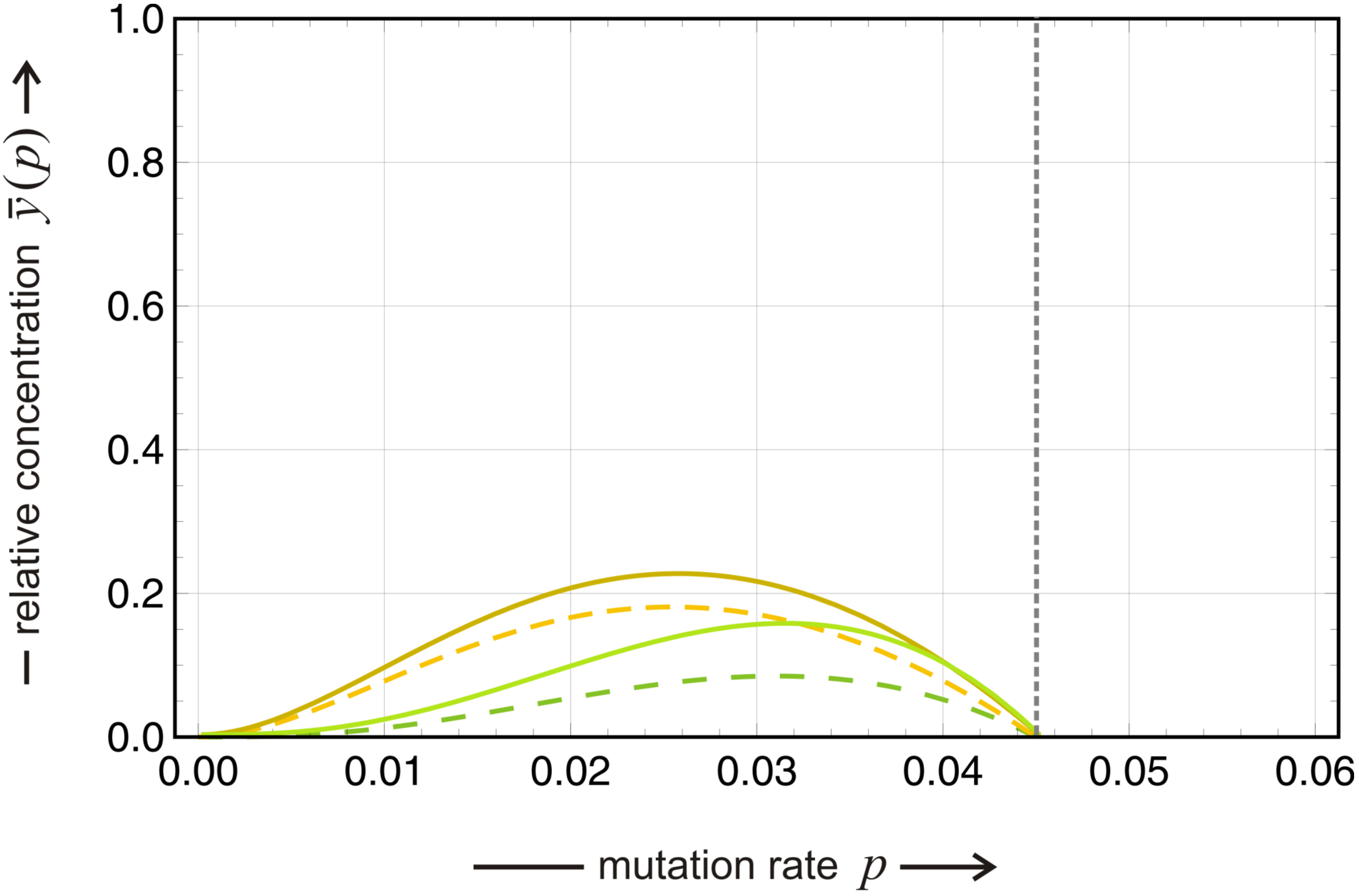
Zero mutation backflow



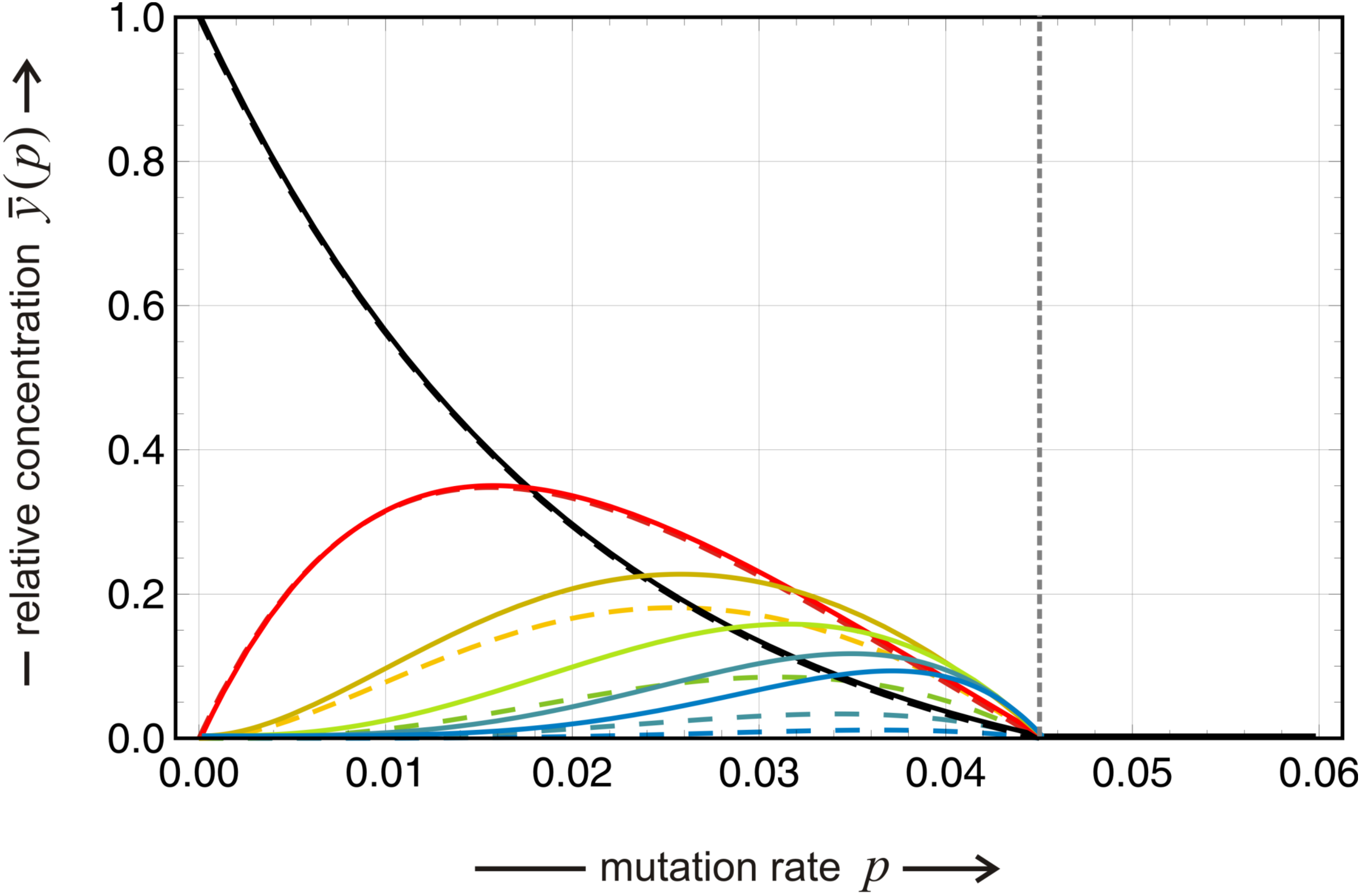
Zero mutation backflow: $l = 50$, $f_0 = 10$, $f_j = 1 \; \forall \; j \neq 0$



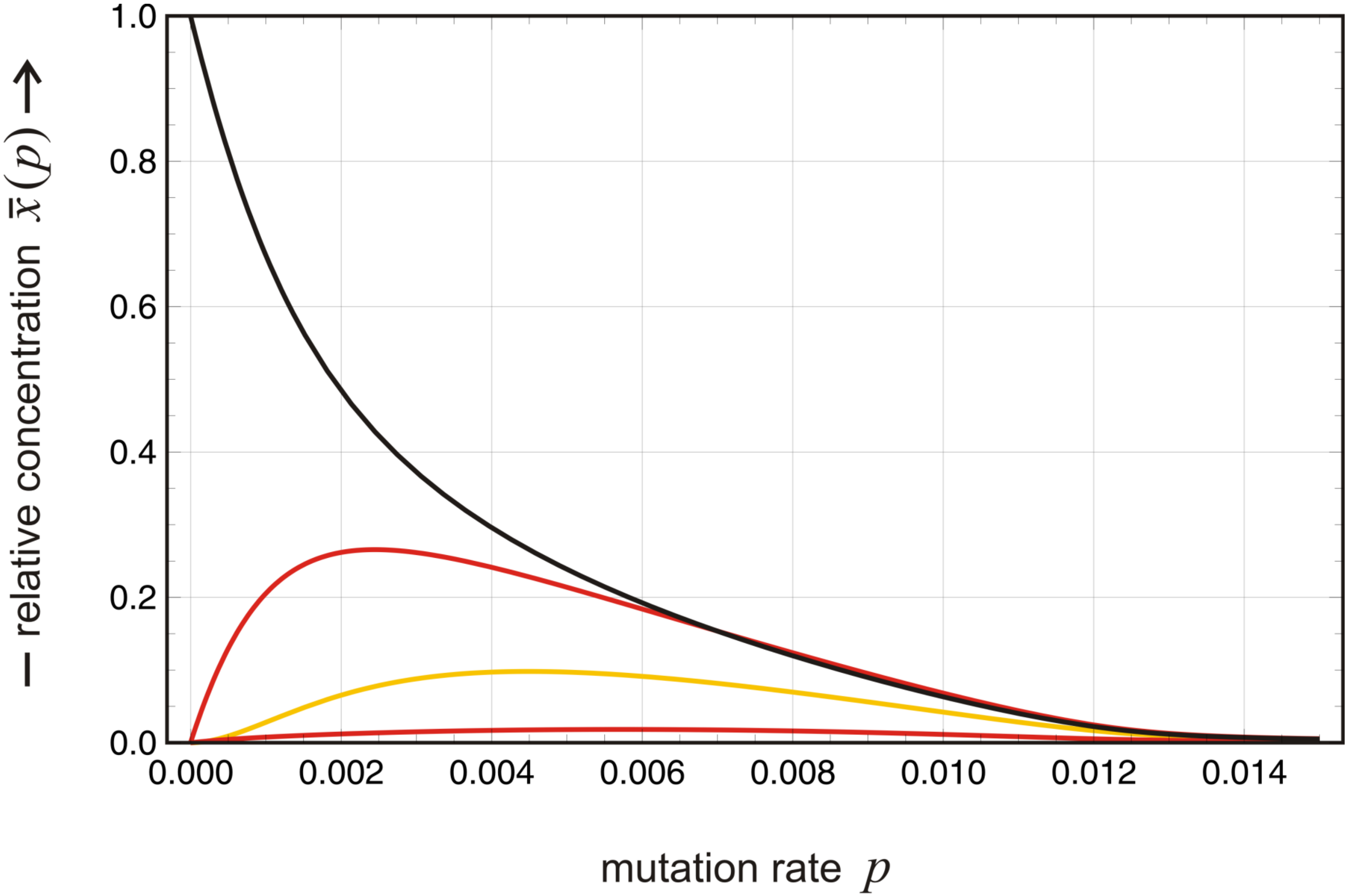
Zero mutation backflow: $l = 50$, $f_0 = 10$, $f_j = 1 \; \forall \; j \neq 0$



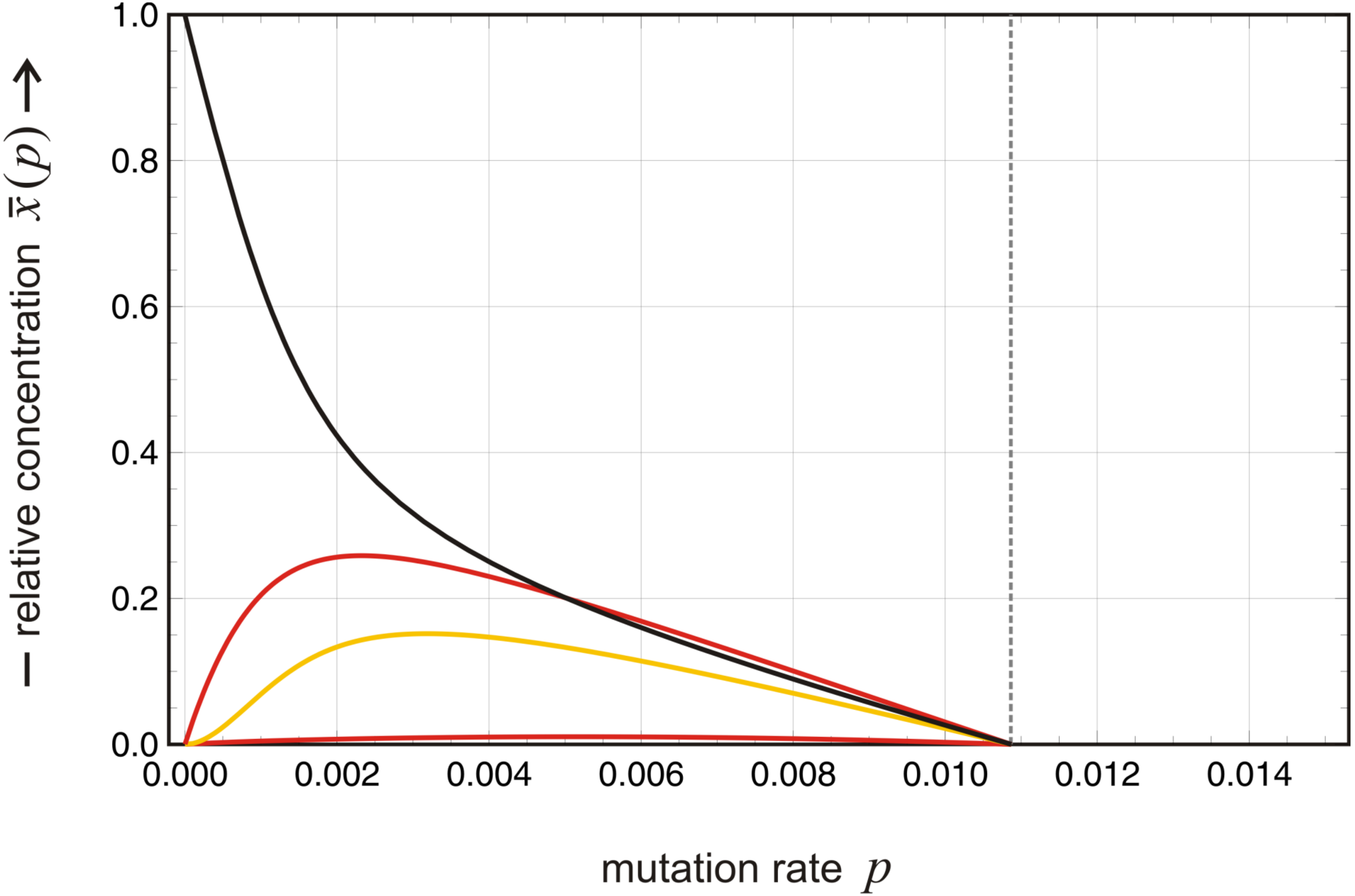
Zero mutation backflow: $l = 50$, $f_0 = 10$, $f_j = 1 \; \forall \; j \neq 0$



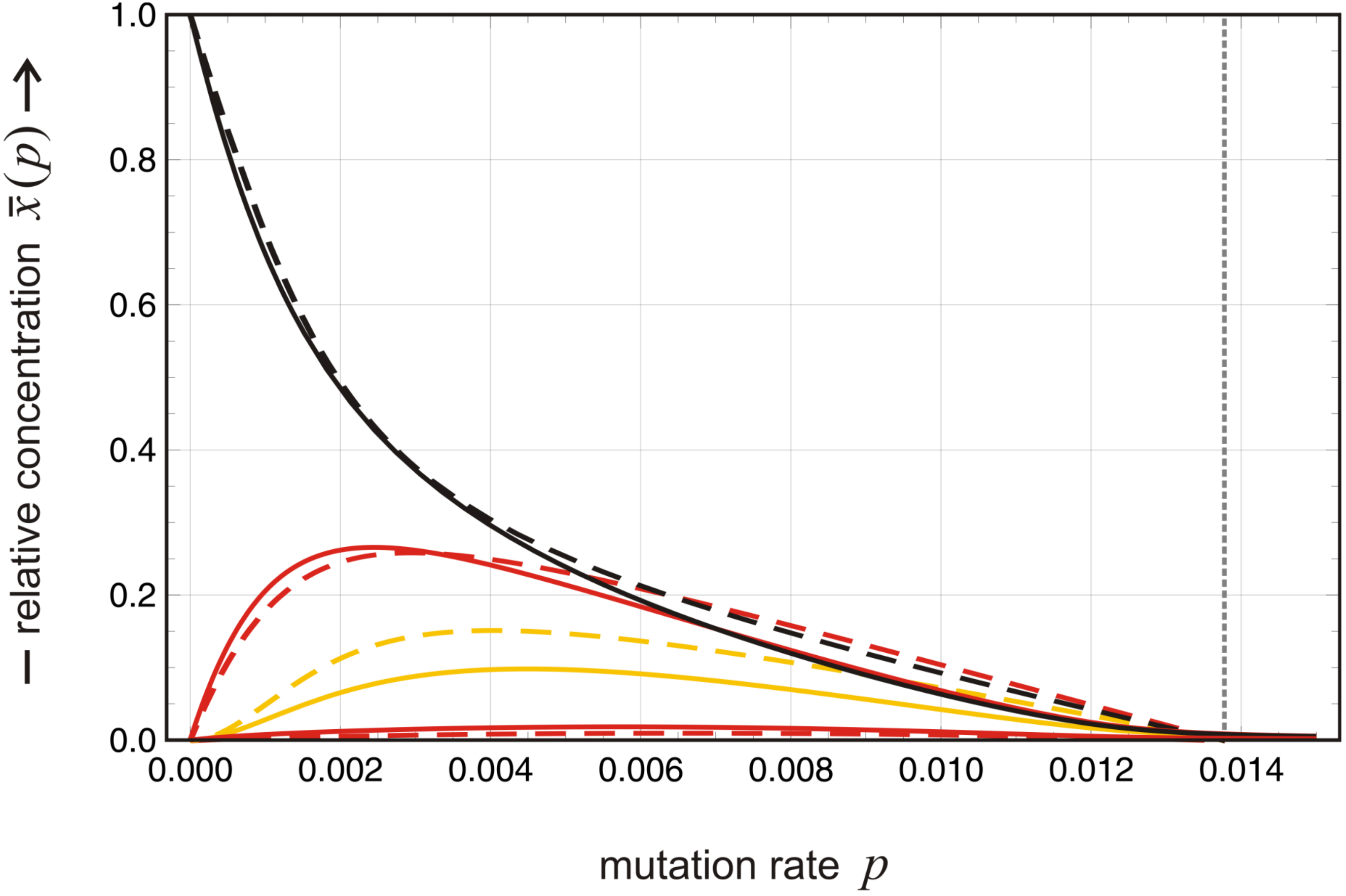
Zero mutation backflow: $l = 50$, $f_0 = 10$, $f_j = 1 \; \forall \; j \neq 0$



Strong quasispecies: $l = 10$, $f_0 = 1.1$, $f_n = 1.0$, d = 1.0, s = 919



Zero mutational backflow: $l = 10$, $f_0 = 1.1$, $f_n = 1.0$, $f_4 = 1.09659$, $f_{516} = 1.09703$



Strong quasispecies: comparison

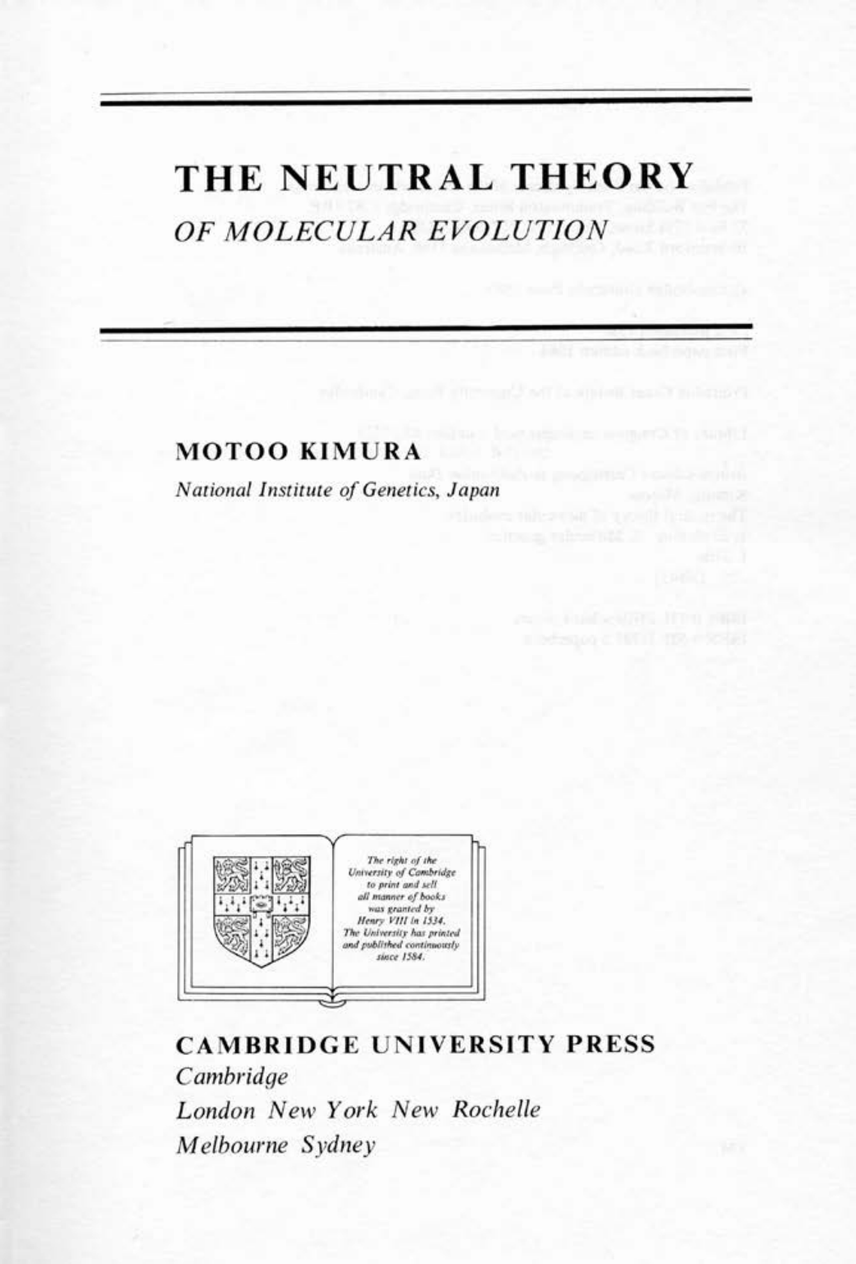
1. Prologue
2. Landscapes and evolutionary dynamics
3. Realistic landscapes and mutant clans
4. Mutation flow analysis
5. **Neutrality in evolutionary dynamics**
6. Concluding remarks and perspectives

Motoo Kimura, 1924 - 1994

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.



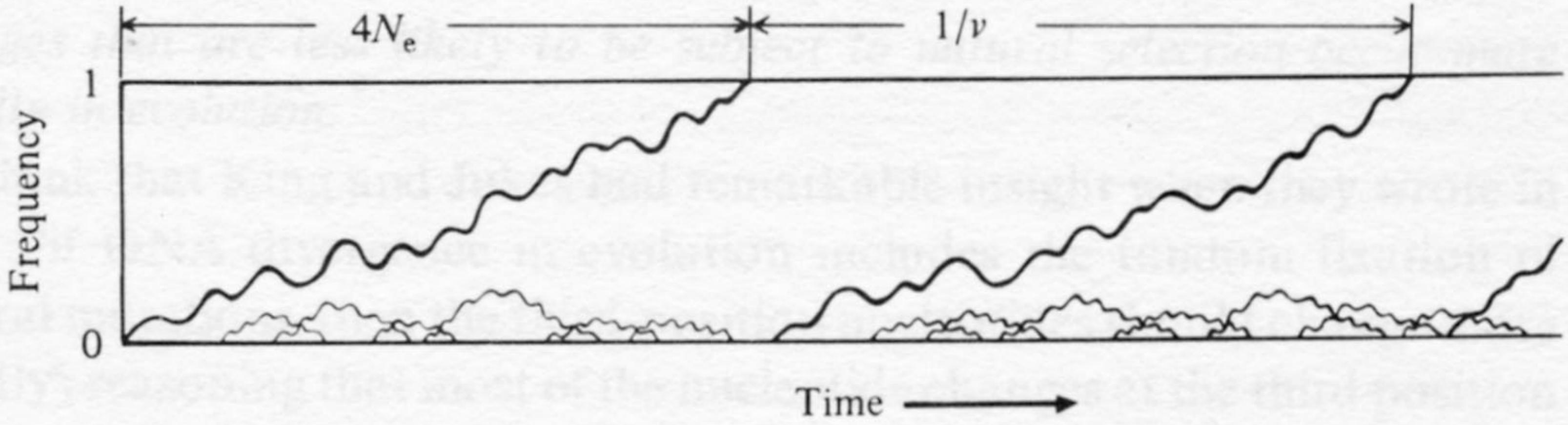
THE NEUTRAL THEORY
OF MOLECULAR EVOLUTION

MOTOO KIMURA
National Institute of Genetics, Japan

The right of the
University of Cambridge
to print and sell
all manner of books
was granted by
Henry VIII in 1534.
The University has printed
and published continuously
since 1584.

CAMBRIDGE UNIVERSITY PRESS
Cambridge
London New York New Rochelle
Melbourne Sydney

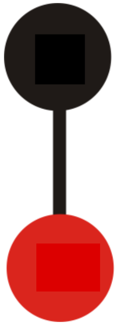
Fig. 3.1. Behavior of mutant genes following their appearance in a finite population. Courses of change in the frequencies of mutants destined to fixation are depicted by thick paths. N_e stands for the effective population size and v is the mutation rate.



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.

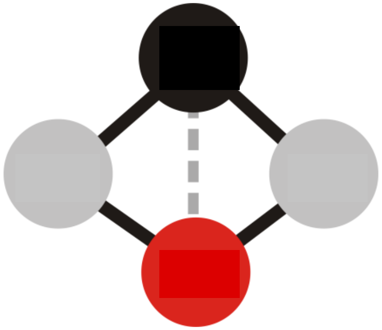
Fixation leads to selection of a single variant in the sense of „survival of the survivor“.

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



$d_H = 1$

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



$d_H = 2$

$$\lim_{p\to 0} x_1(p) = \alpha/(1+\alpha)$$

$$\lim_{p\to 0} x_2(p) = 1/(1+\alpha)$$

$d_H \geq 3$

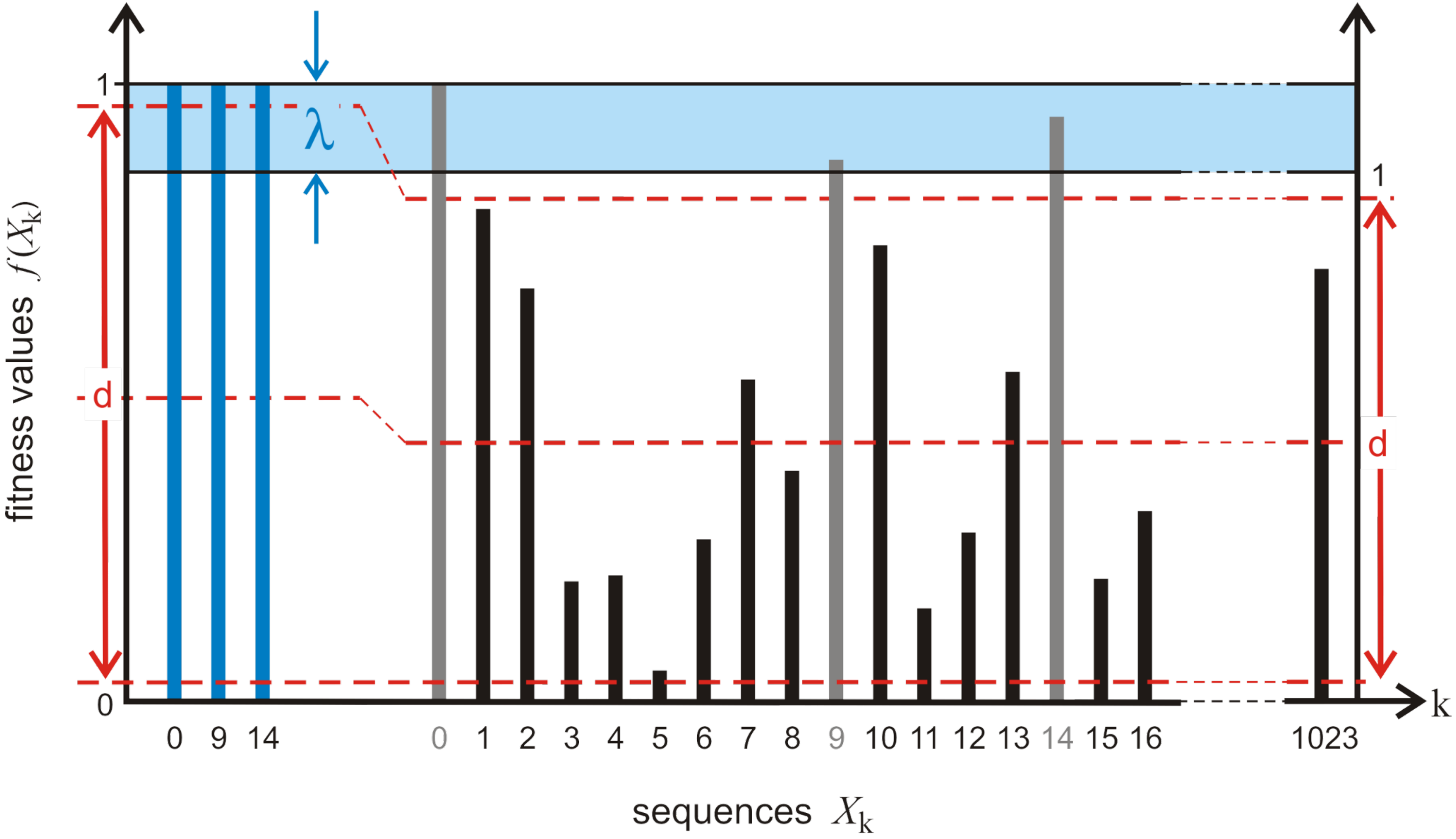
$$\lim_{p\to 0} x_1(p) = 1, \lim_{p\to 0} x_2(p) = 0 \quad \text{or}$$

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

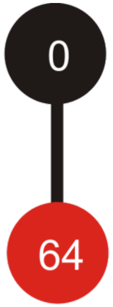
Survival of the survivor or random fixation in the sense of **Motoo Kimura**

Pairs of neutral sequences in replication networks

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

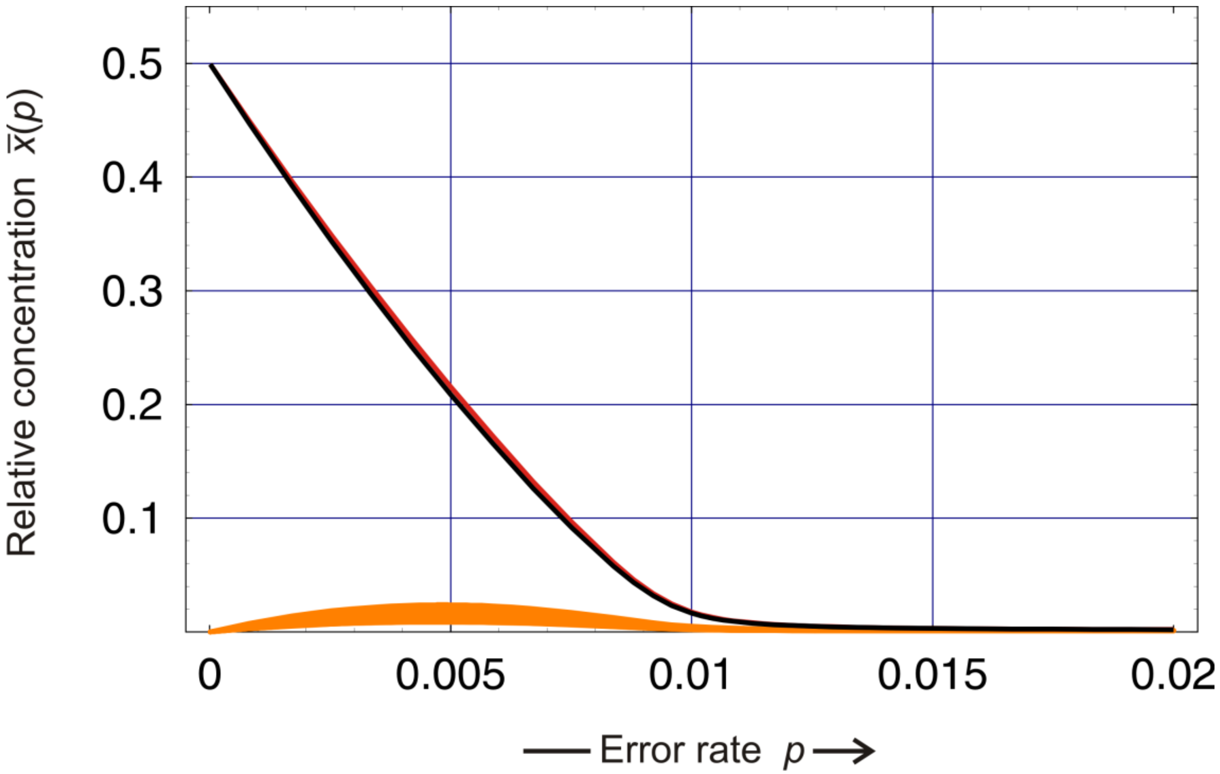


A fitness landscape including neutrality



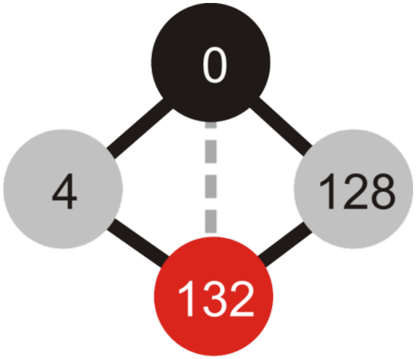
neutral network

$\lambda = 0.01$, s = 367



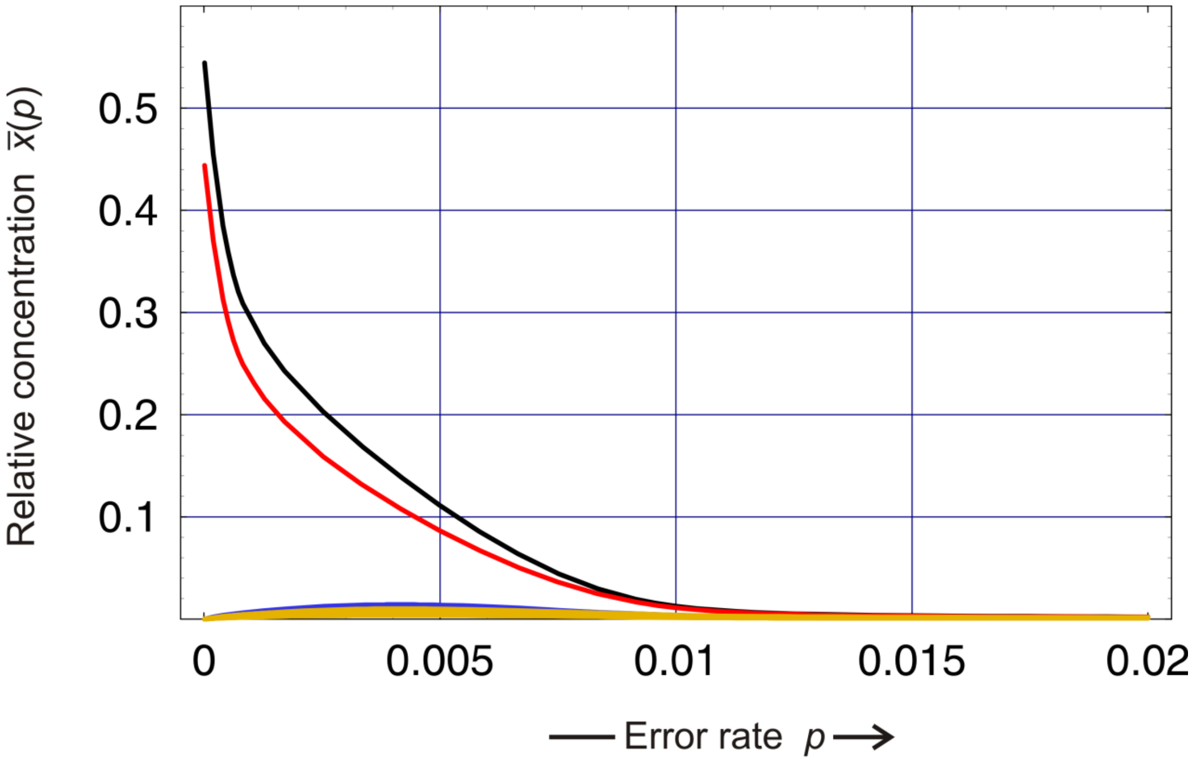
Neutral network: Individual sequences

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



neutral network

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



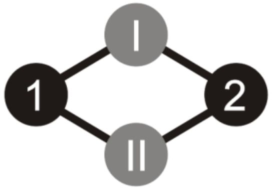
Neutral network: Individual sequences

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

······· ACAUGCGAA ······· master sequence 1
······· AUAUACGAA ·······
······· ACAUGCGCA ·······
······· GCAUACGAA ·······
······· ACAUGCUAA ·······
······· ACAUGCGAG ·······
······· ACACGCGAA ·······
······· ACGUACGAA ·······
······· ACAUAGGAA ·······
······· ACAUACGAA ······· master sequence 2

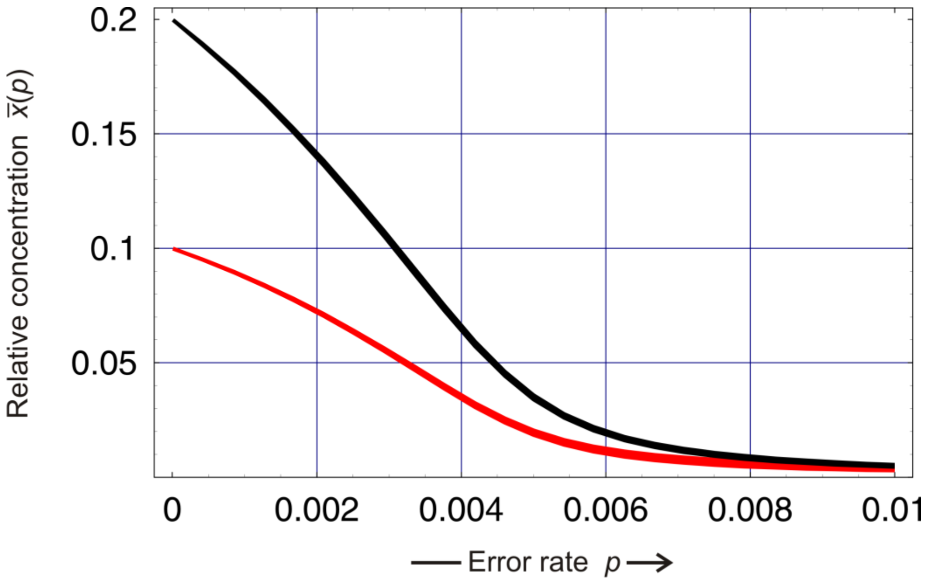
······· ACAU$\genfrac{}{}{0pt}{}{G}{A}$CGAA ······· consensus sequence

······· ACAGUCAGAA ······· master sequence 1
······· ACAGUCCGAA ······· intermediate I
······· AUAAUCCGAA ·······
······· ACAGUCAGCA ·······
······· GCAGUCAGAA ·······
······· ACAGUCAUAA ·······
······· ACAGUCAGAG ·······
······· ACAACCCGAA ·······
······· ACGGUCAGAA ·······
······· ACAGUGAGAA ·······
······· ACAAUCAGAA ······· intermediate II
······· ACAAUCCGAA ······· master sequence 2



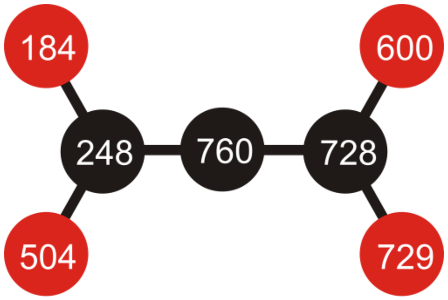
······· ACA$\genfrac{}{}{0pt}{}{G}{A}UC\genfrac{}{}{0pt}{}{A}{C}$GAA ······· consensus sequence

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

λ = 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: λ = 0.10, s = 229

1. Prologue
2. Landscapes and evolutionary dynamics
3. Realistic landscapes and mutant clans
4. Mutation flow analysis
5. Neutrality in evolutionary dynamics
6. **Concluding remarks and perspectives**

The quasispecies and the error threshold concepts

1. Realistic landscapes combine two seemingly conflicting features: (i) **ruggedness** and (ii) **neutrality**.

2. The ostensible contradiction between ruggedness and neutrality is resolved by the **sequence-structure relation** of **biopolymers**.

3. Rugged landscapes and all landscapes with sufficient "steepness" sustain **error thresholds** in the sense that an almost uniform distribution of variants is approached already far away from the point of random replication defined by equal rates for correct replication and mutation.

4. Provided certain requirements on the fitness values of one and two error mutants of the master sequence are fulfilled, the populations form especially **stable quasispecies**.

5. In stable quasispecies or in the corresponding neutral clusters **clans of sequences** replace the single survivors – deterministic or random.

Perspectives of molecular evolution

1. Populations with high and low mutation rates are described within the same model based on the quasispecies concept.

2. Accurate predictions on in vitro evolution and virus evolution can be made wherever fitness parameters are available.

3. The modeling approach can be extended in qualitative terms to other prokaryotic and eukaryotic populations provided enough data are available.

4. The mechanism of reproduction can be extended to more complex mechanisms like sexual reproduction and reproduction including epigenetic effects.

Coworkers

Universität Wien

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

Peter Stadler, Universität Leipzig, DE

Walter Fontana, Harvard Medical School, MA

Christian Reidys, University of Southern Denmark, Odense, DK

Thomas Wiehe, Universität Köln, DE

Martin Nowak, Harvard University, MA

Stefan Bonhoeffer, ETH Zürich, CH

Christian Forst, Southwestern Medcial Center, University of Texas, Dallas, TX

Erich Bornberg-Bauer, Münster, DE

Acknowledgement of support

Universität Wien

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

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

Santa Fe Institute Preprint Series # 12-06-006