Evolution at Molecular Resolution

Peter Schuster

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

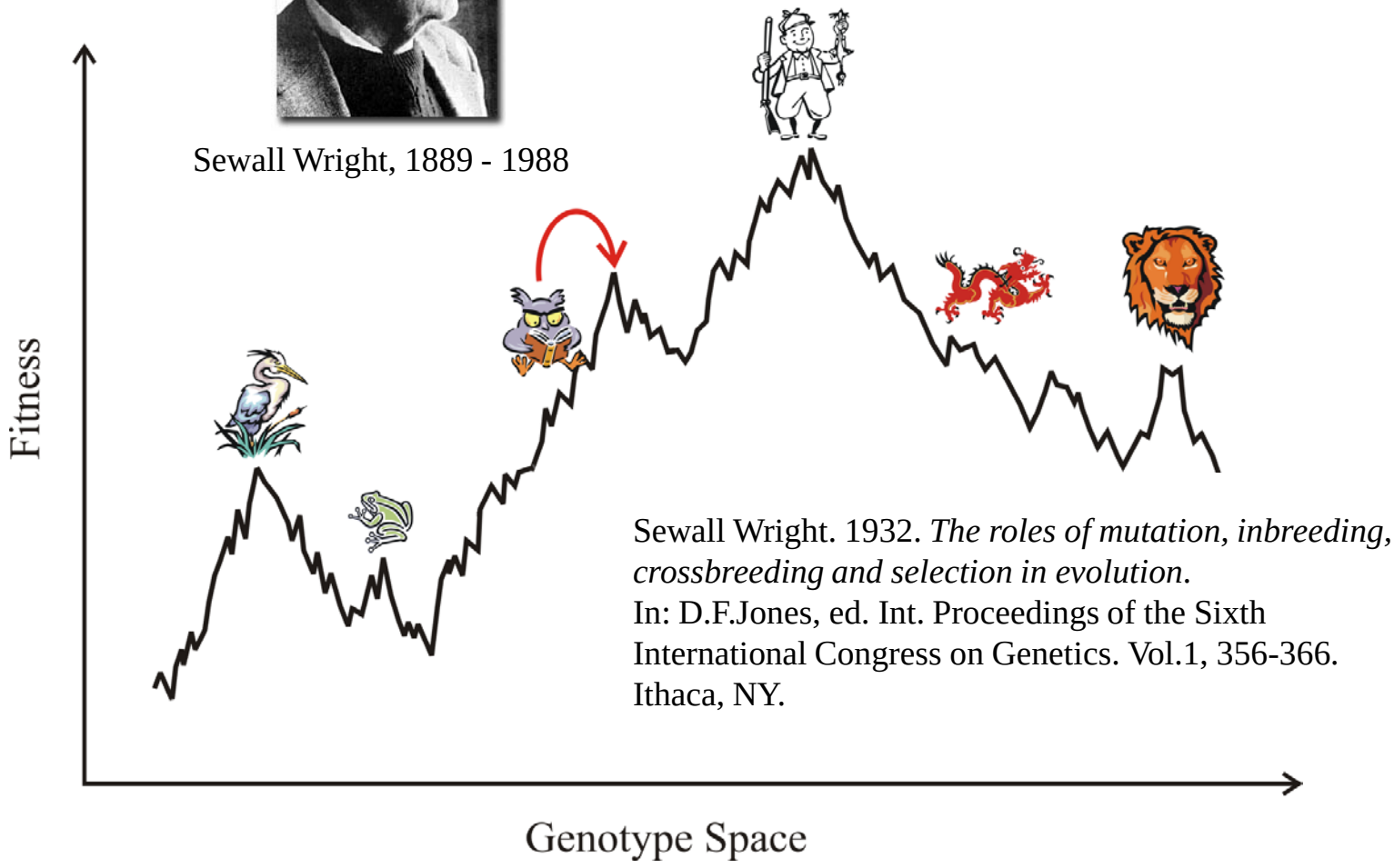
EMBO Members‘ Meeting 2014

Heidelberg, 29.– 31.10.2014

Web-Page for further information:

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

Sewall Wright, 1889 - 1988



Sewall Wright. 1932. *The roles of mutation, inbreeding, crossbreeding and selection in evolution.*
In: D.F.Jones, ed. Int. Proceedings of the Sixth International Congress on Genetics. Vol.1, 356-366. Ithaca, NY.

Sewall Wrights fitness landscape as metaphor for Darwinian evolution

+ …….. wild type
a ..... alternative allele
on locus A

:

:

:

abcde … alternative alleles
on all five loci

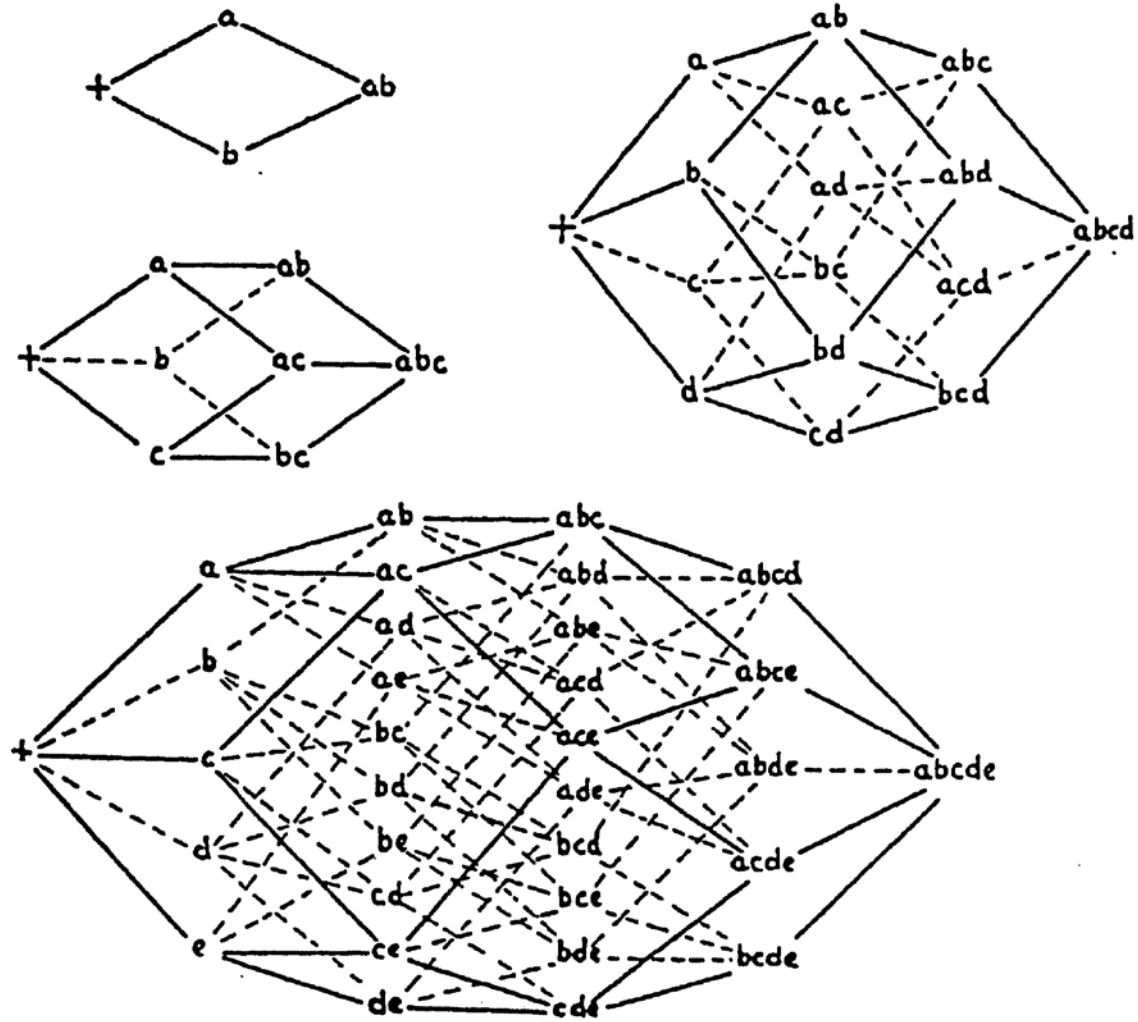
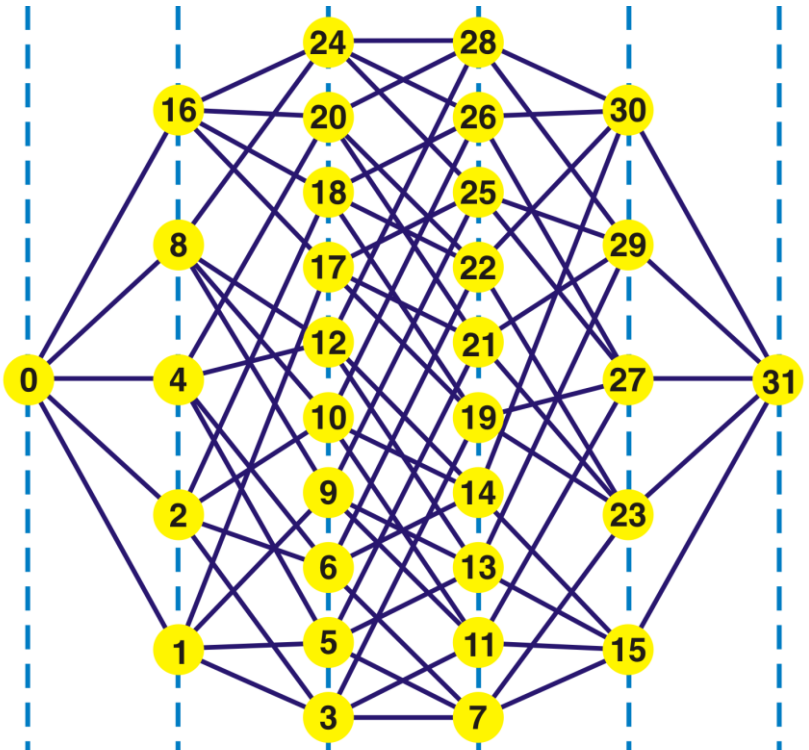


FIG. 1.—The combinations of from 2 to 5 paired allelomorphs.

The multiplicity of gene replacements with two alleles on each locus

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



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 $l = 5$

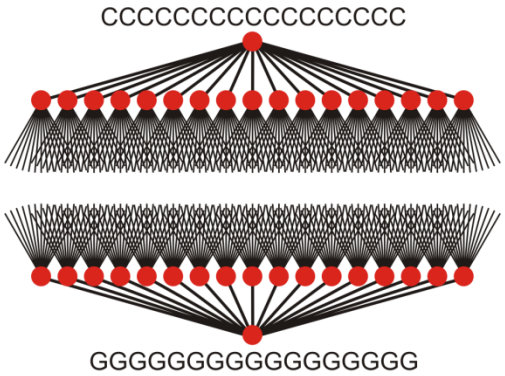
Fitness landscapes became experimentally accessible!

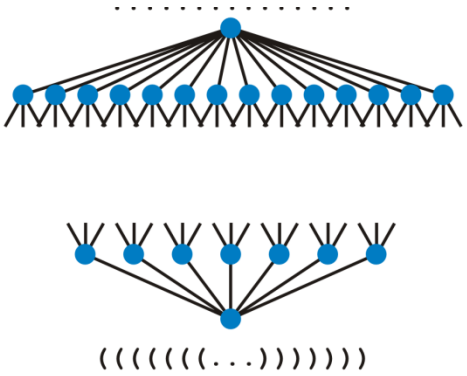
Protein landscapes: Yuuki Hayashi, Takuyo Aita, Hitoshi Toyota, Yuzuru Husimi, Itaru Urabe, Tetsuya Yomo. 2006. Experimental rugged fitness landscape in protein sequence space. *PLoS One* 1:e96.

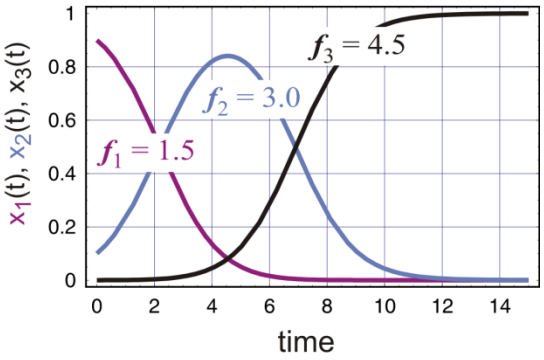
RNA landscapes: Sven Klussman, Ed. 2005. The aptamer handbook. Wiley-VCh, Weinheim (Bergstraße), DE.
Jason N. Pitt, Adrian Ferré-D'Amaré. 2010. Rapid construction of empirical RNA fitness landscapes. *Science* 330:376-379.

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 | shape space | parameter space

$\mathcal{Q}$ | S | $\mathbb{R}_+$

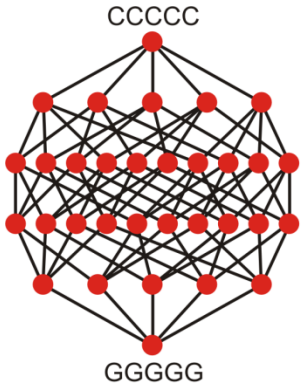
$\Phi: (\mathcal{Q}, d_H) \Rightarrow (S, d_S)$ $\qquad$ $\Psi: (S, d_S) \Rightarrow \mathbb{R}_+$

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

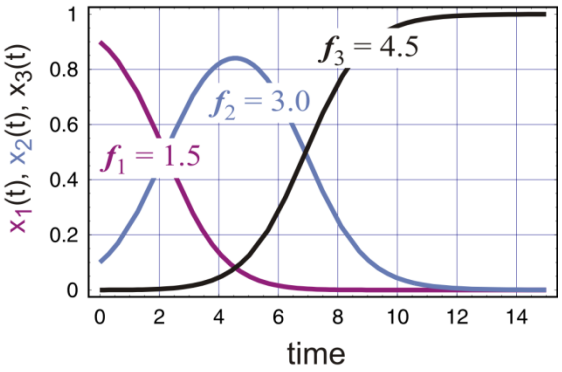
sequence | structure | function

genotype $\longrightarrow$ **phenotype** $\longrightarrow$ **selection**

Evolution as a global phenomenon in genotype space



sequence space



parameter space

S

sequence

$f = \Psi(Y)$

function

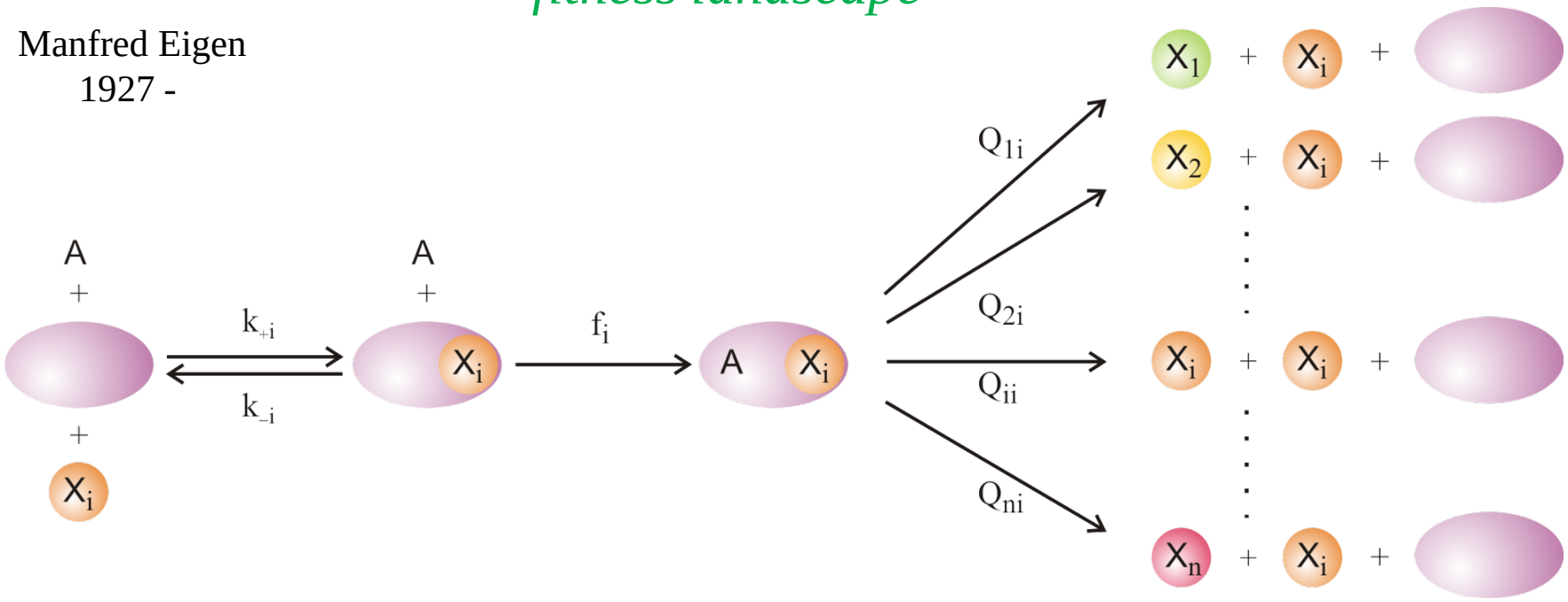
The simplified model

Manfred Eigen
1927 -

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

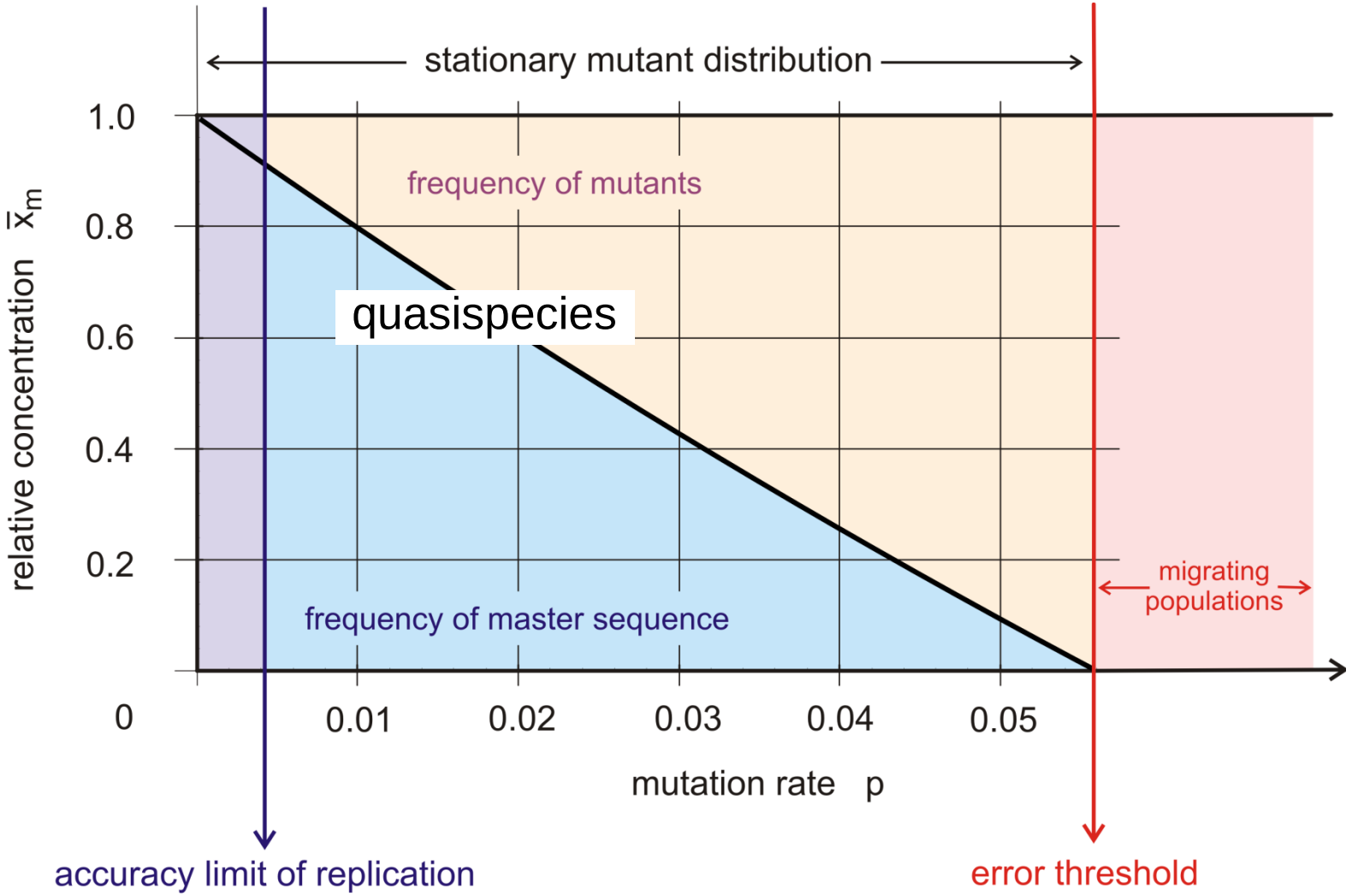
$$W_{ji} = Q_{ji} \cdot f_i\,, \quad \sum_{i=1}^{n} x_i = 1, \quad \Phi = \sum_{i=1}^{n} f_i\, x_i$$

fitness landscape

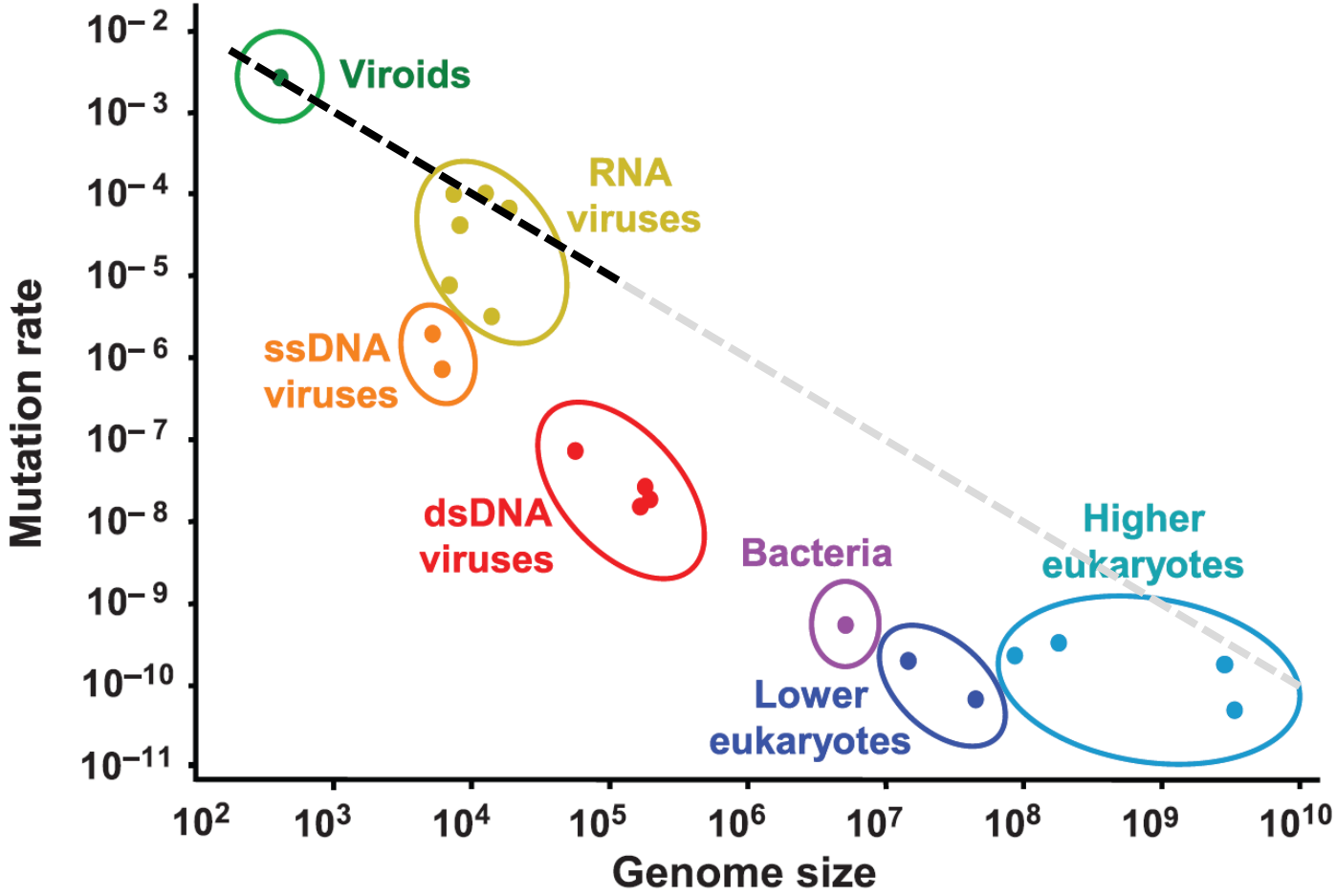


Mutation and (correct) replication as parallel chemical reactions

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

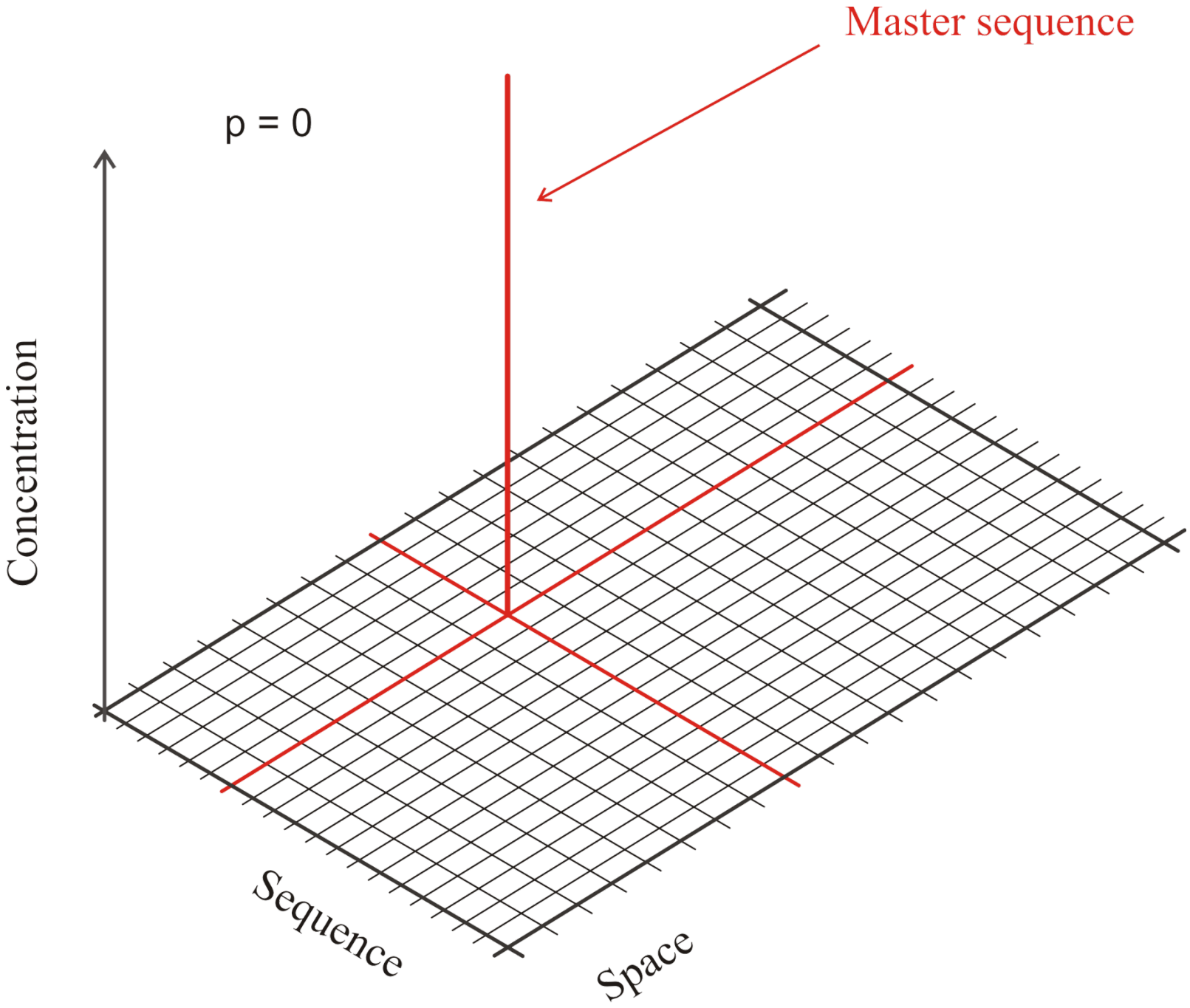


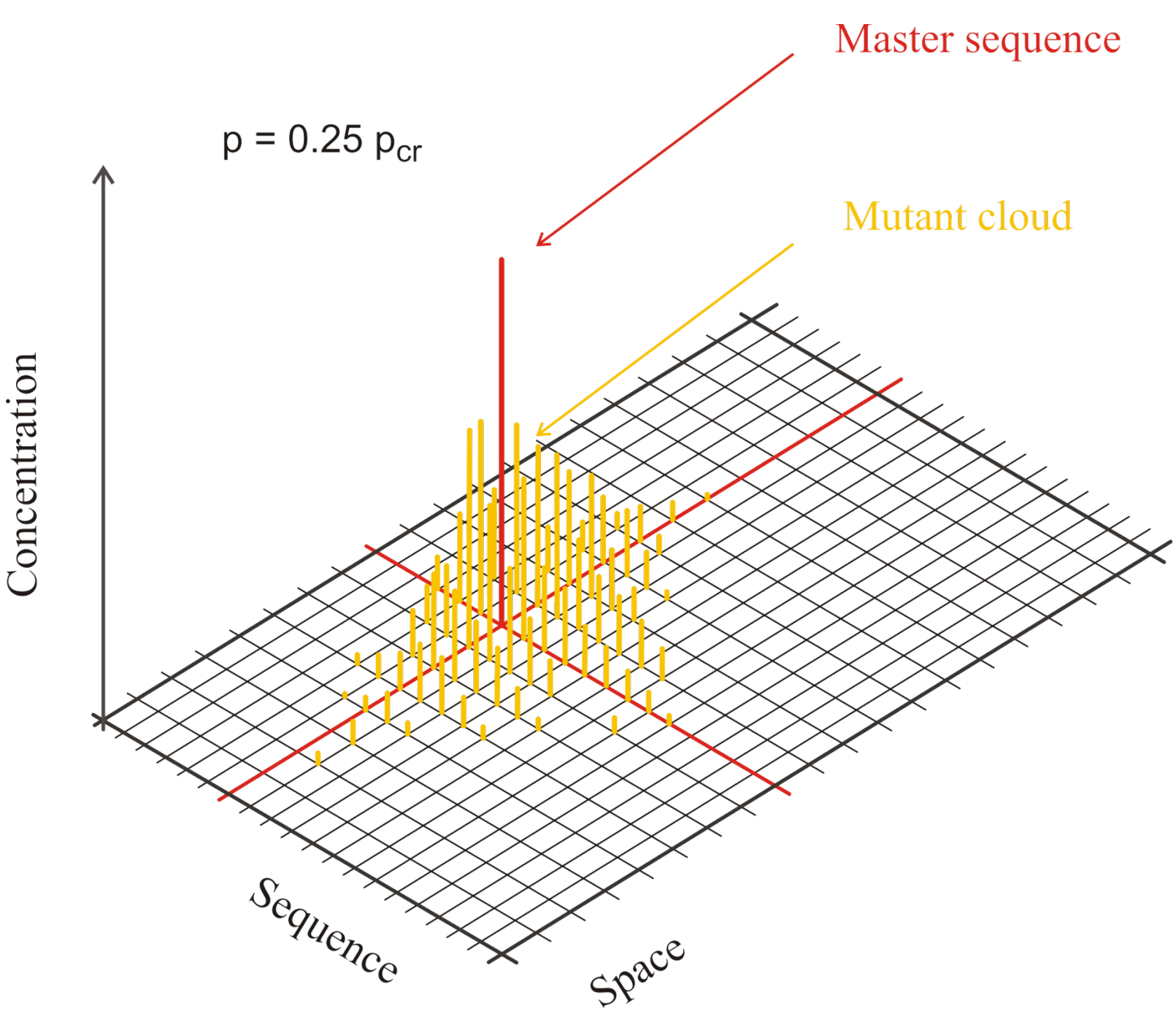
The error threshold in replication and mutation

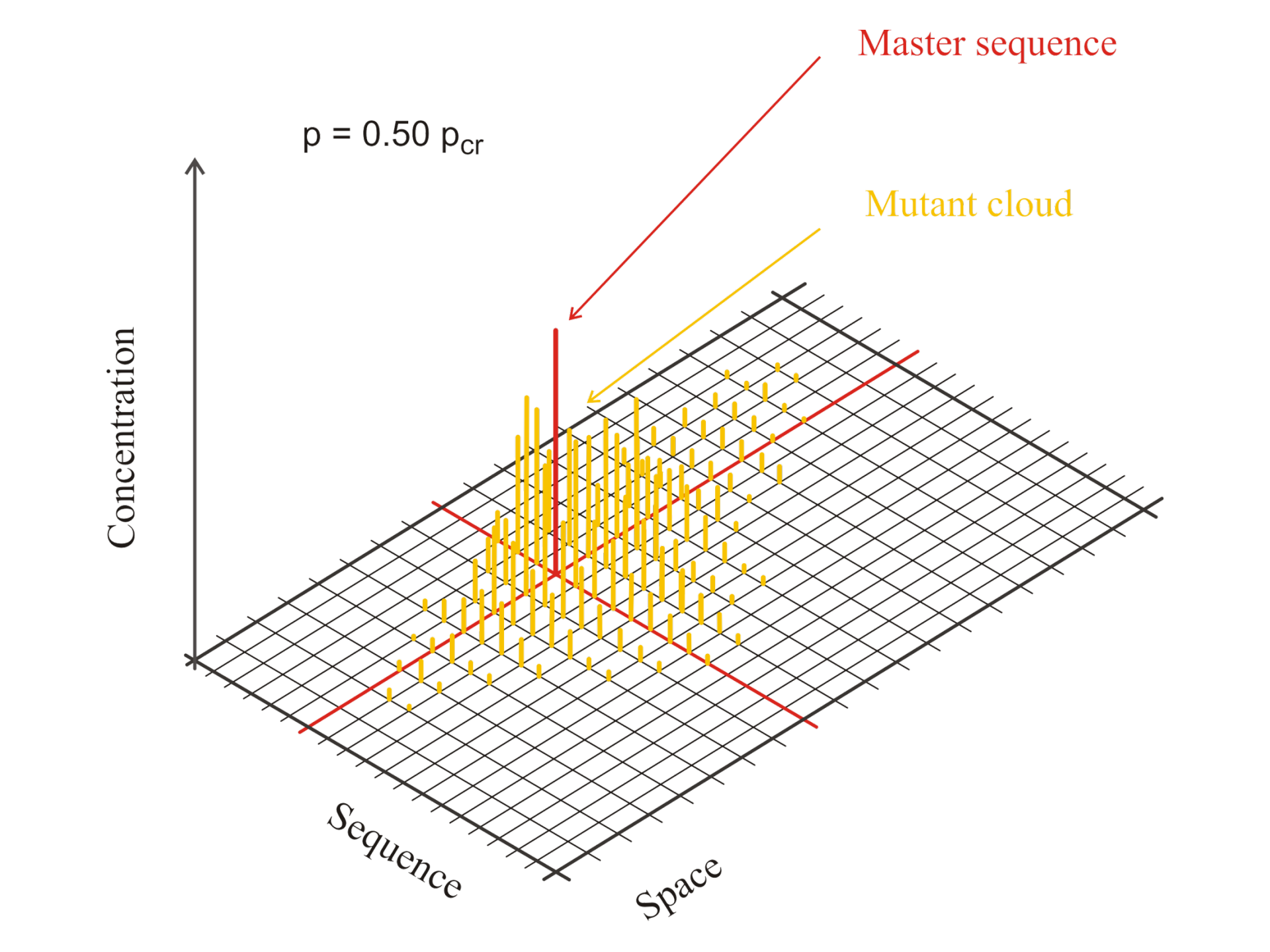


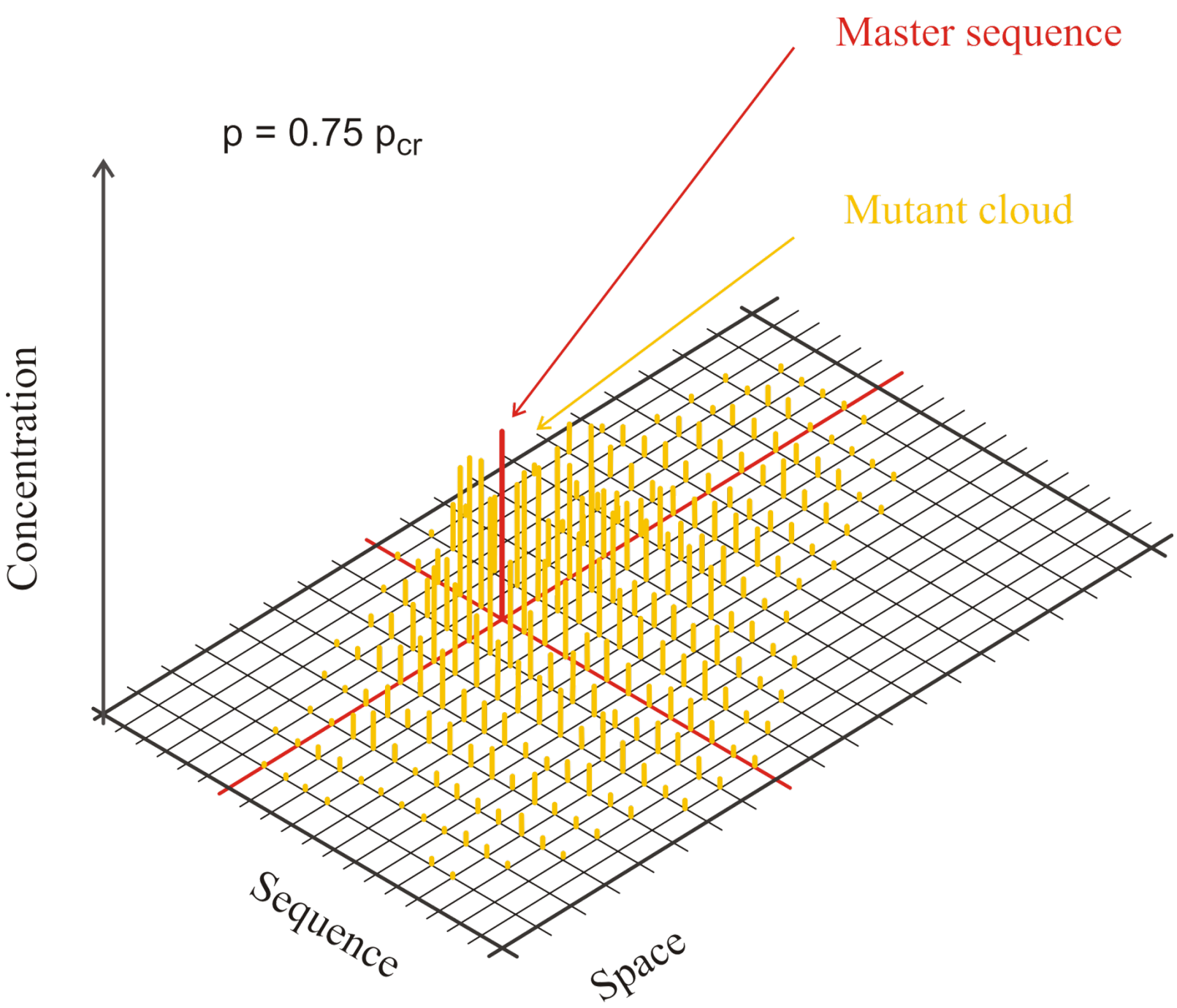
Selma Gago, Santiago F. Elena, Ricardo Flores, Rafael Sanjuán. 2009. Extremely high mutation rate of a hammerhead viroid. Science 323:1308.

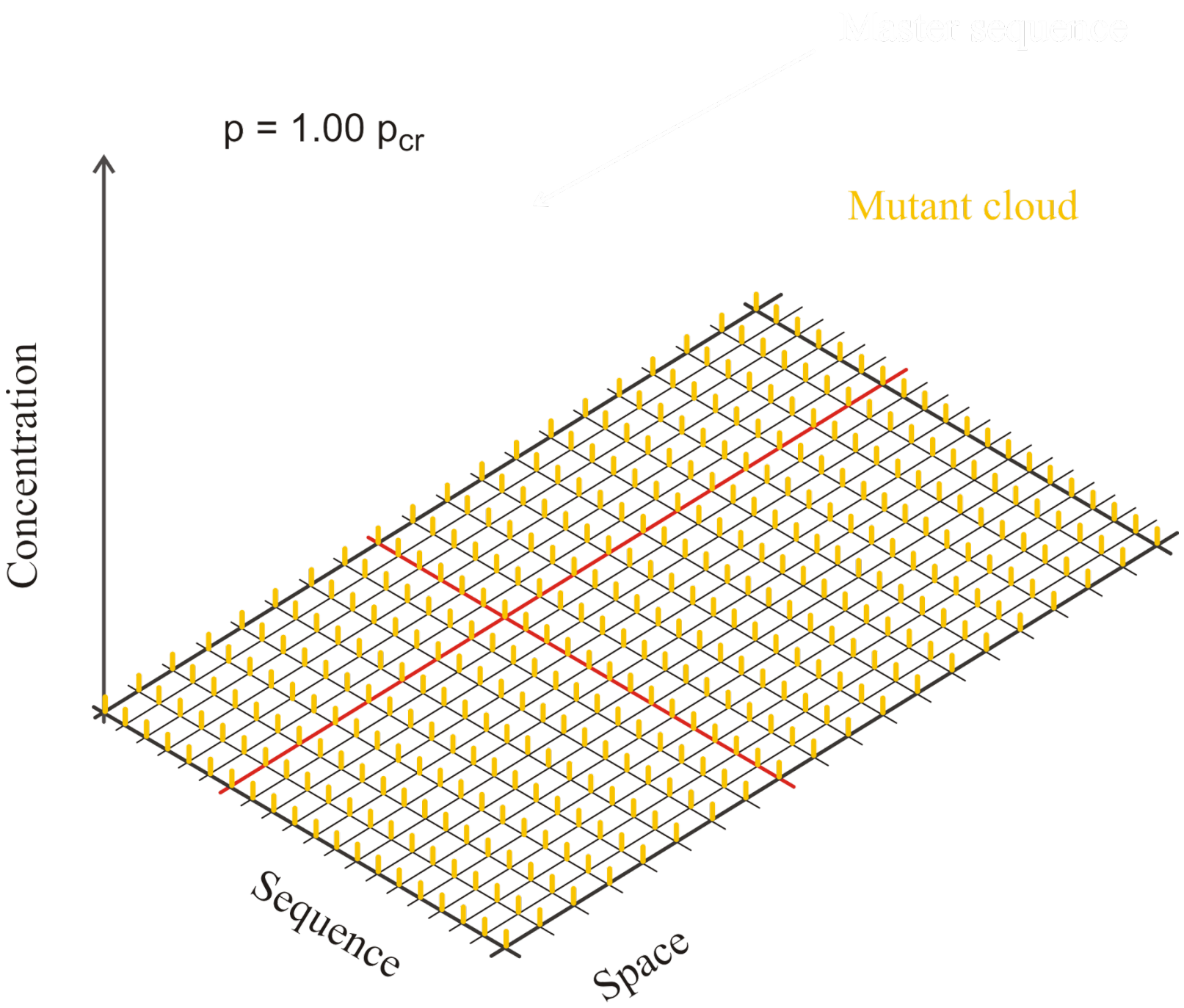
Mutation rate and genome size

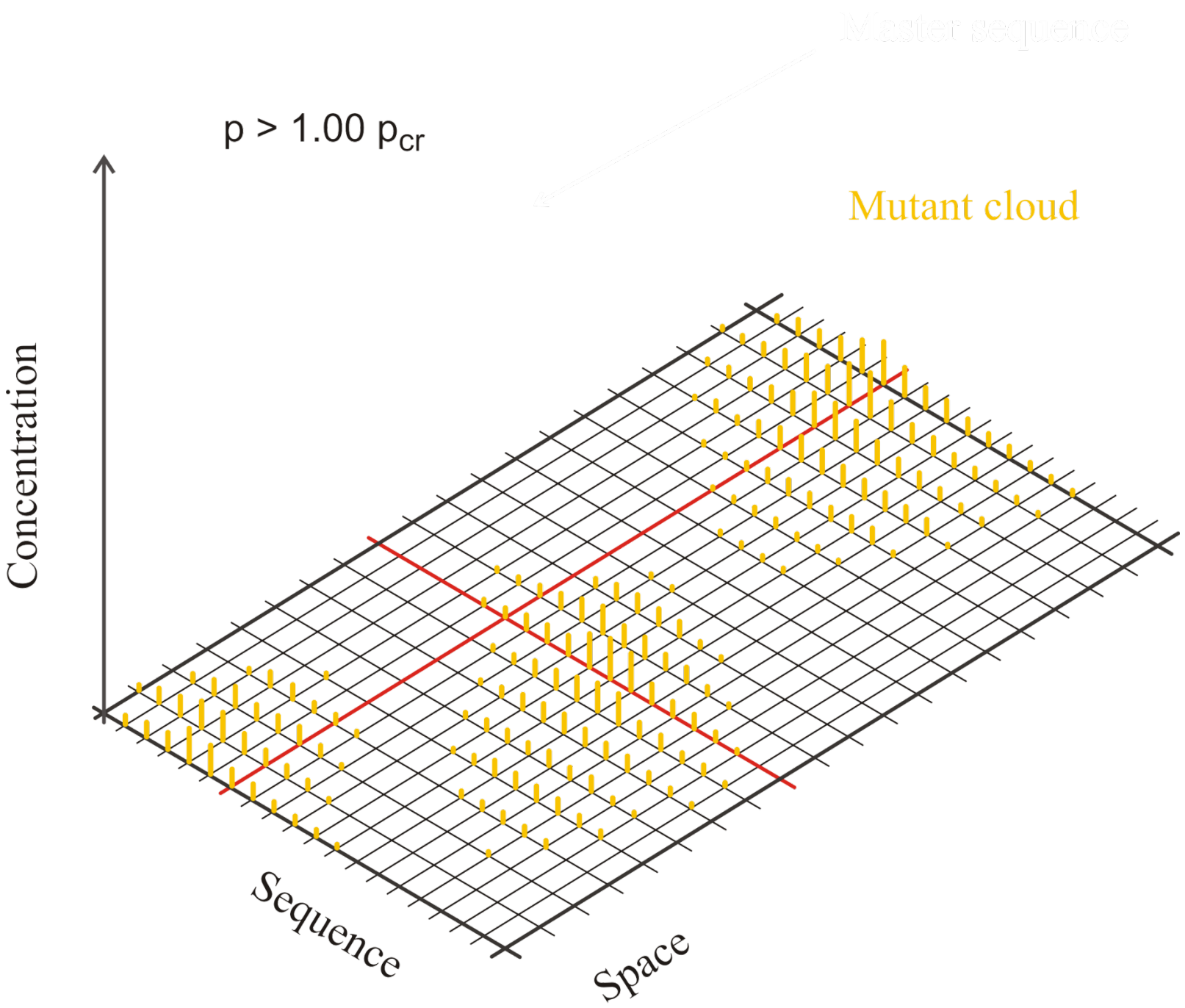

Master sequence
p = 0
Concentration
Sequence
Space


Master sequence
p = 0.25 p_cr
Mutant cloud
Concentration
Sequence
Space

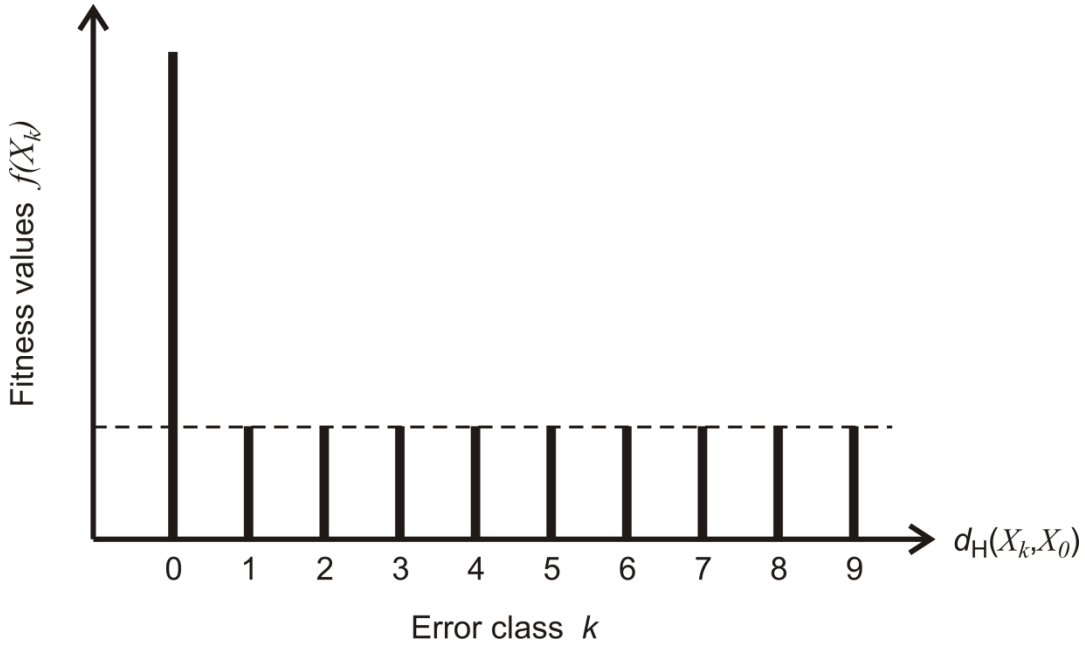

Master sequence
p = 0.50 p_cr
Mutant cloud
Concentration
Sequence
Space


Master sequence
p = 0.75 p_cr
Mutant cloud
Concentration
Sequence
Space


Master sequence
p = 1.00 p_cr
Mutant cloud
Concentration
Sequence
Space


Master sequence
p > 1.00 p_cr
Mutant cloud
Concentration
Sequence
Space

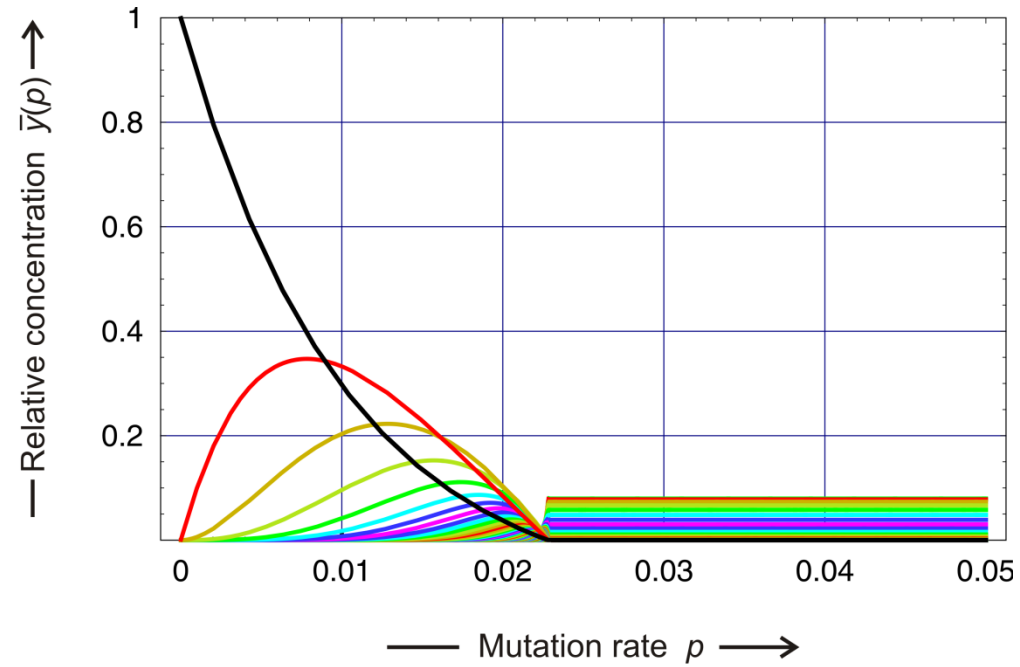
single peak landscape

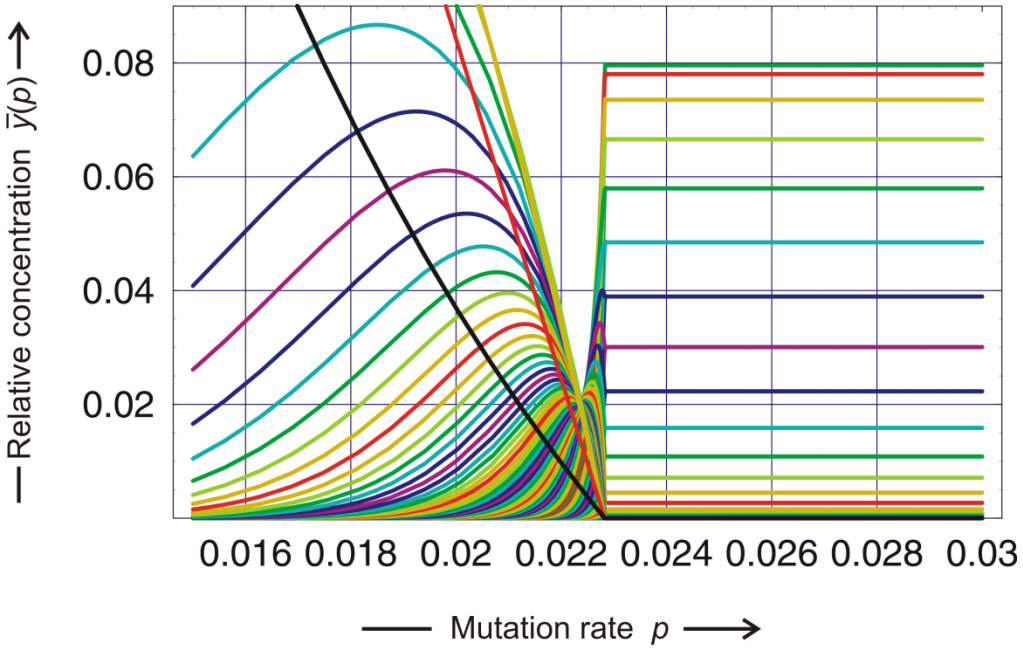


step linear landscape

Model fitness landscapes I

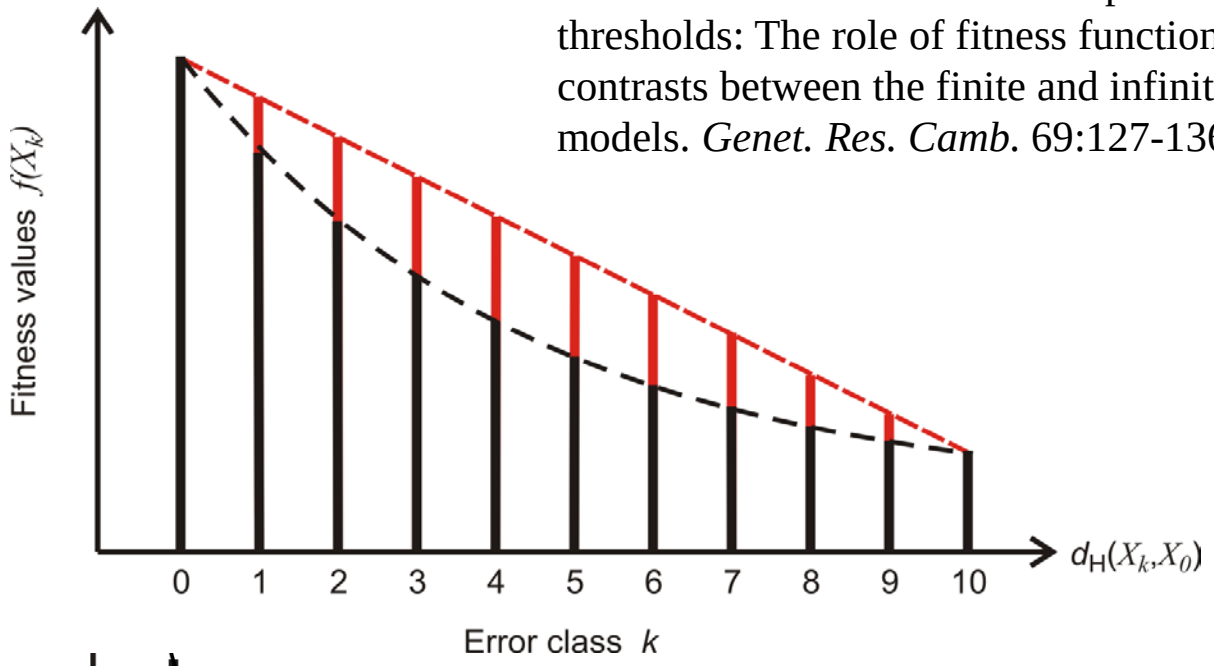
Error threshold on the single peak landscape



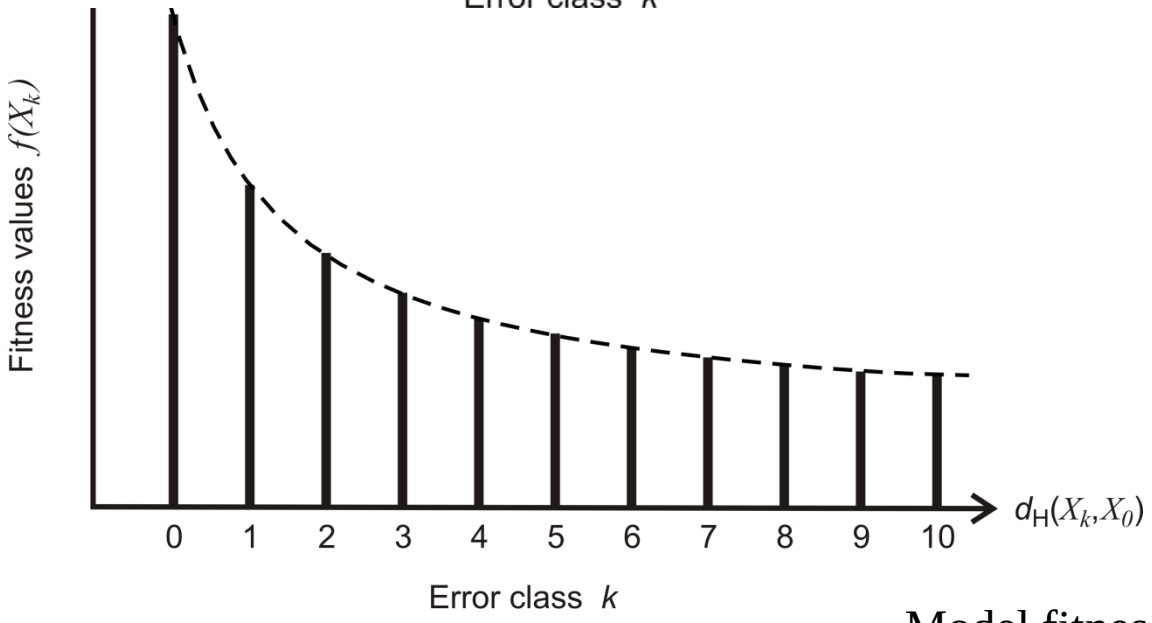


Thomas Wiehe. 1997. Model dependency of error thresholds: The role of fitness functions and contrasts between the finite and infinite sites models. *Genet. Res. Camb.* 69:127-136

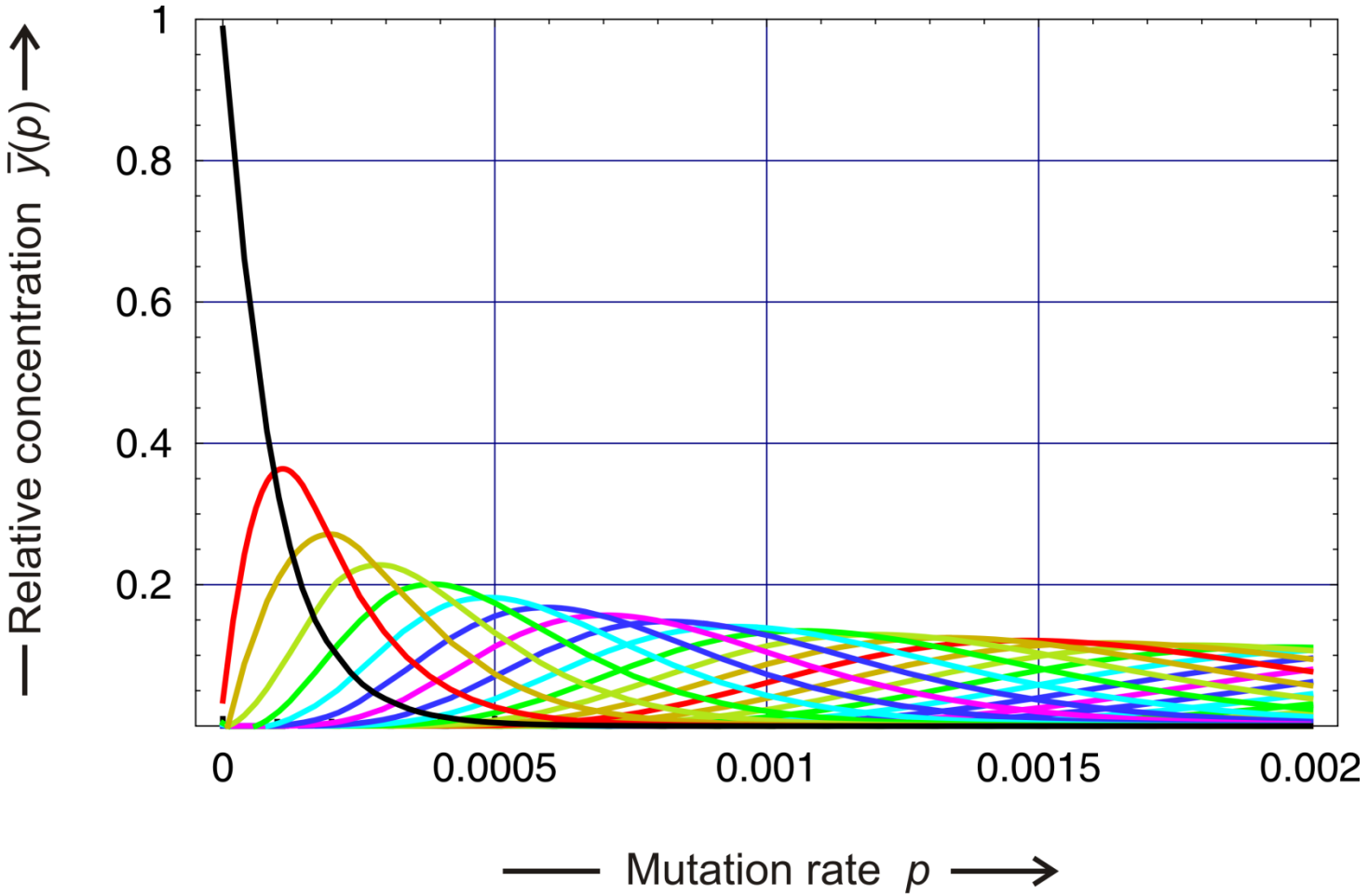
linear and multiplicative



hyperbolic

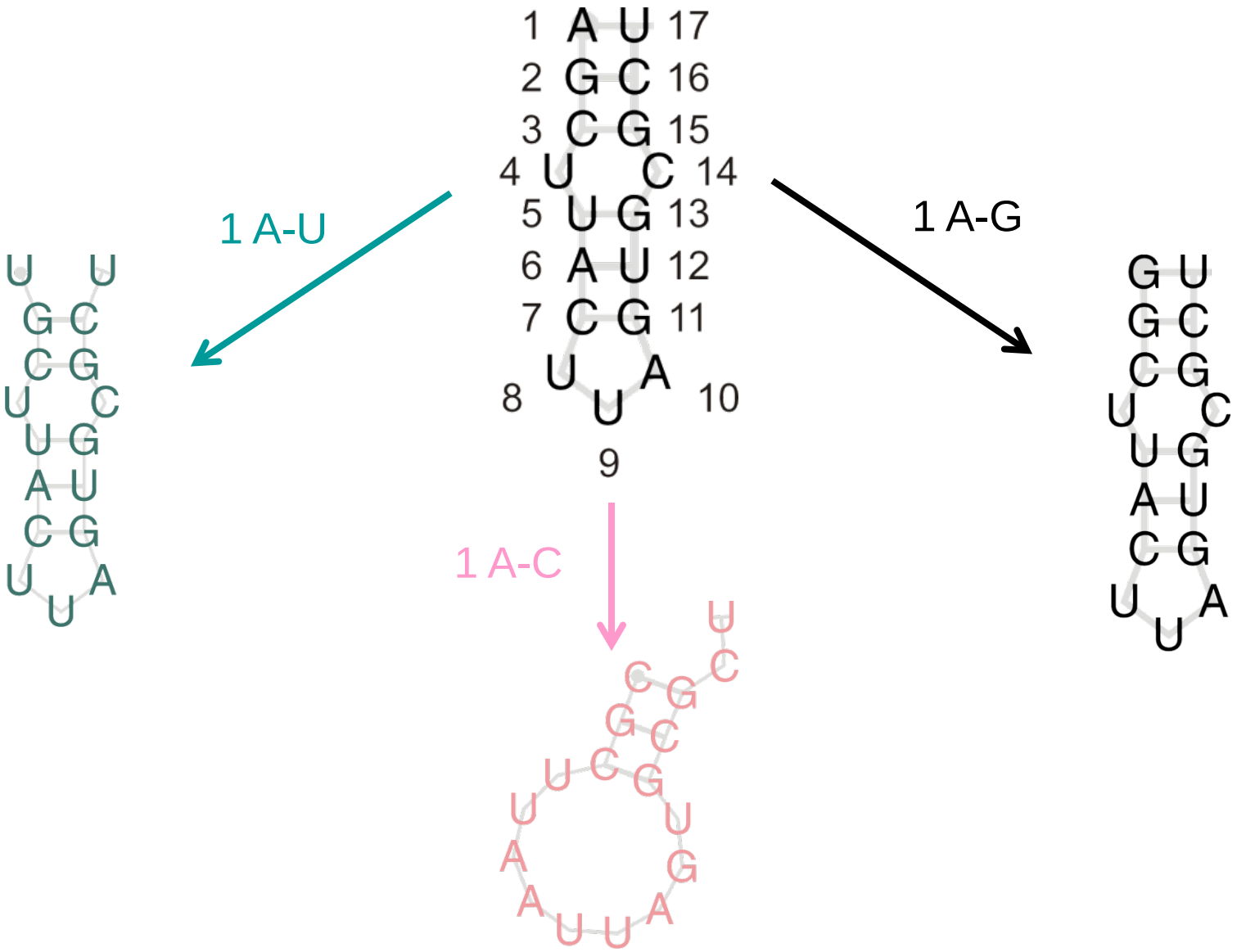


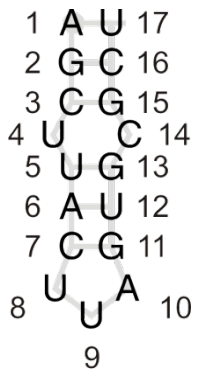
Model fitness landscapes II

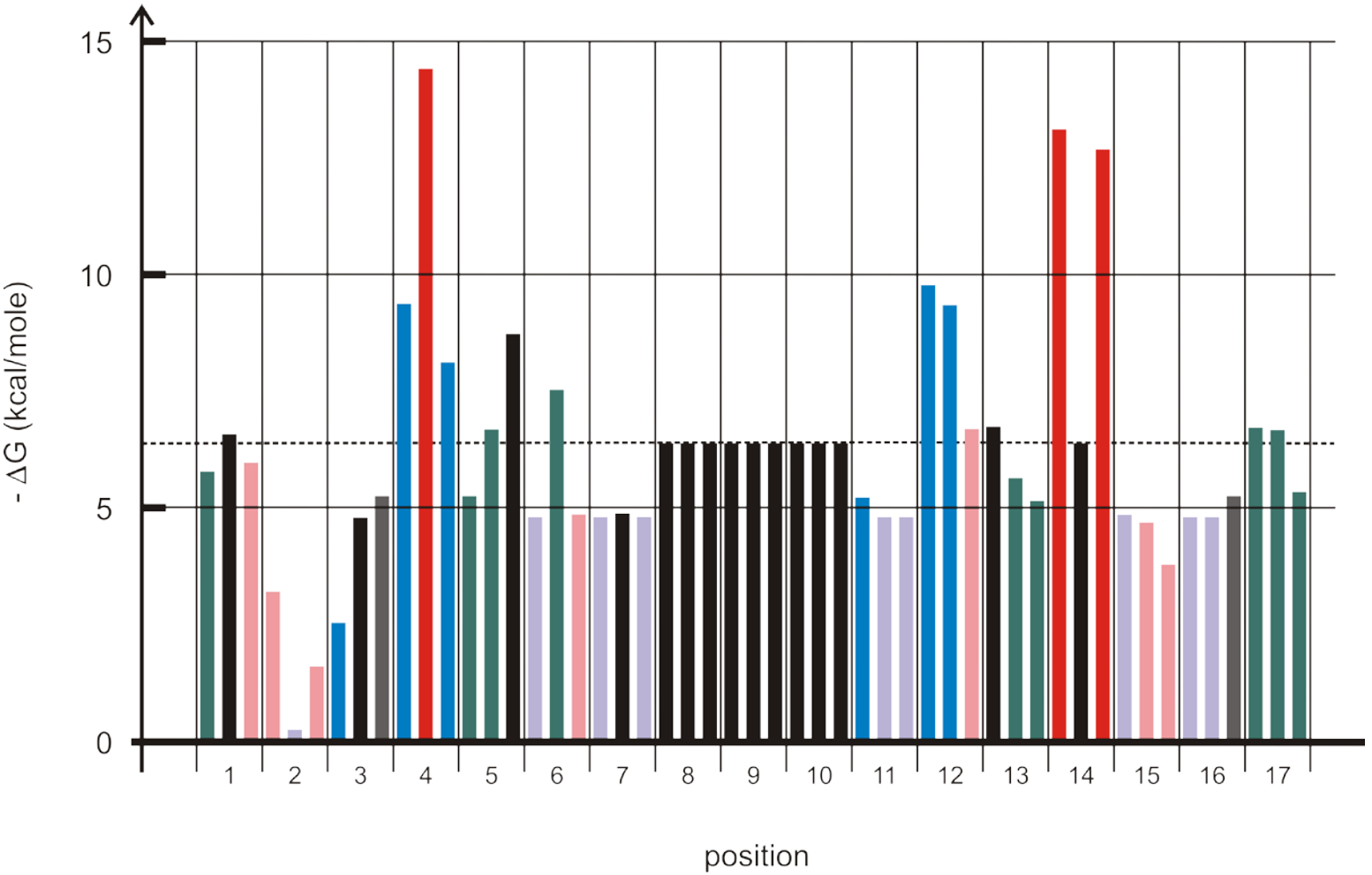


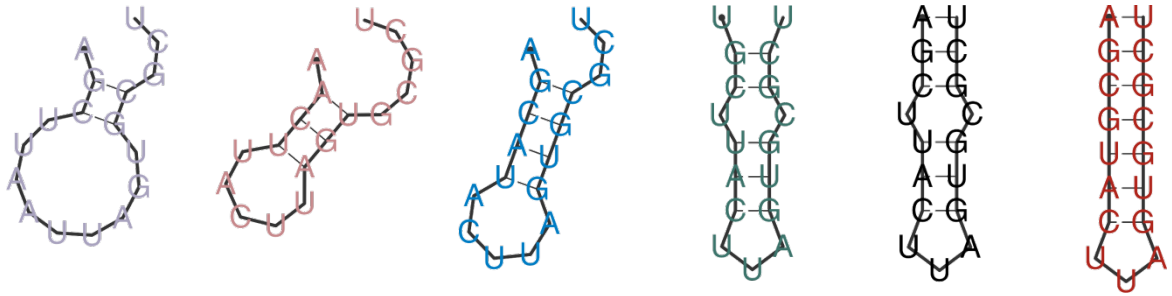
The linear fitness landscape shows no error threshold

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
U A
8 U 10
9


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

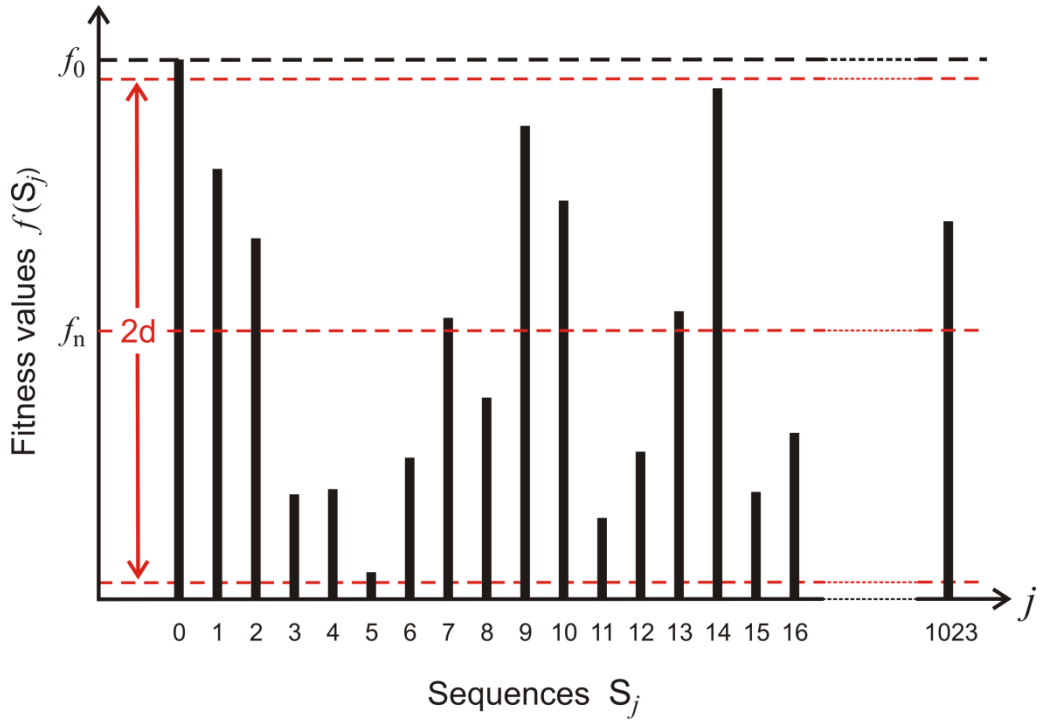


$$f(S_j) = f_n + 2d\,(f_0 - f_n)\left(\eta_j^{(s)} - 0.5\right)$$

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

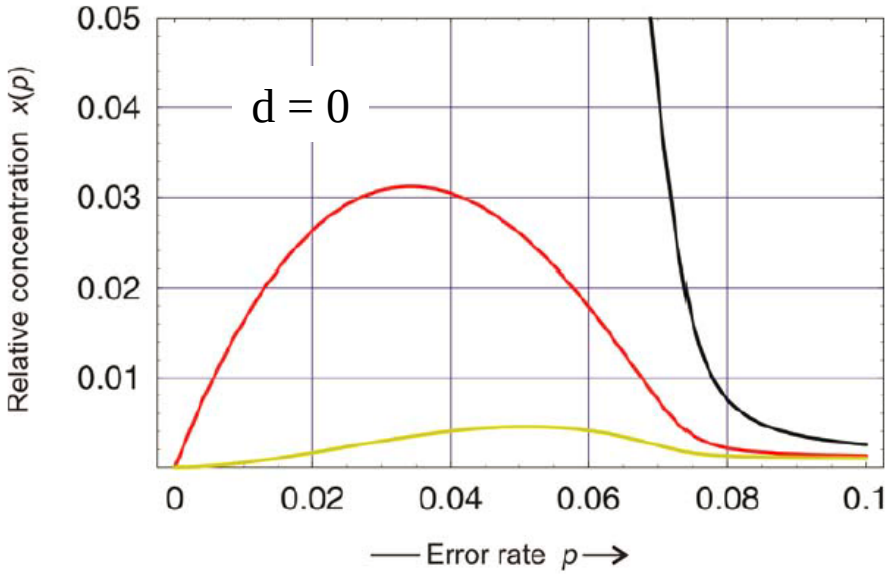
η … random number

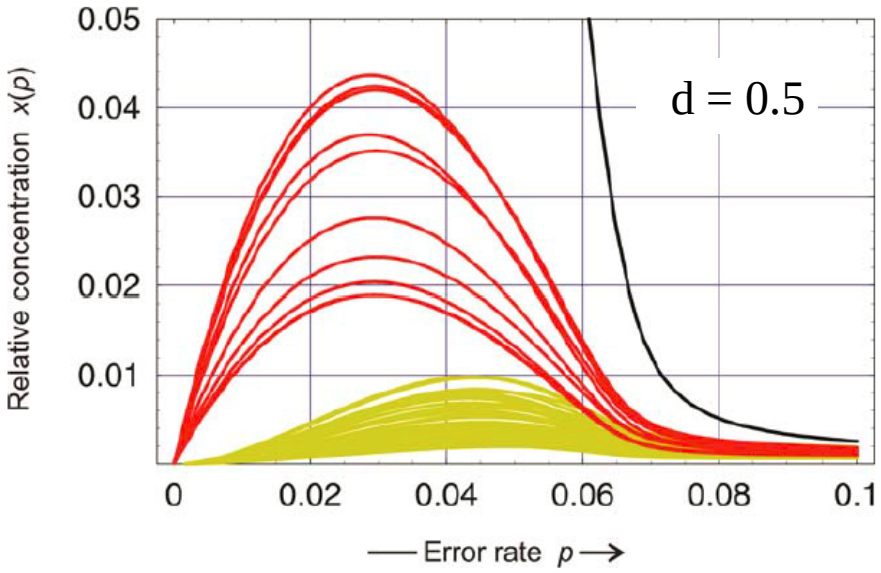
s … seeds

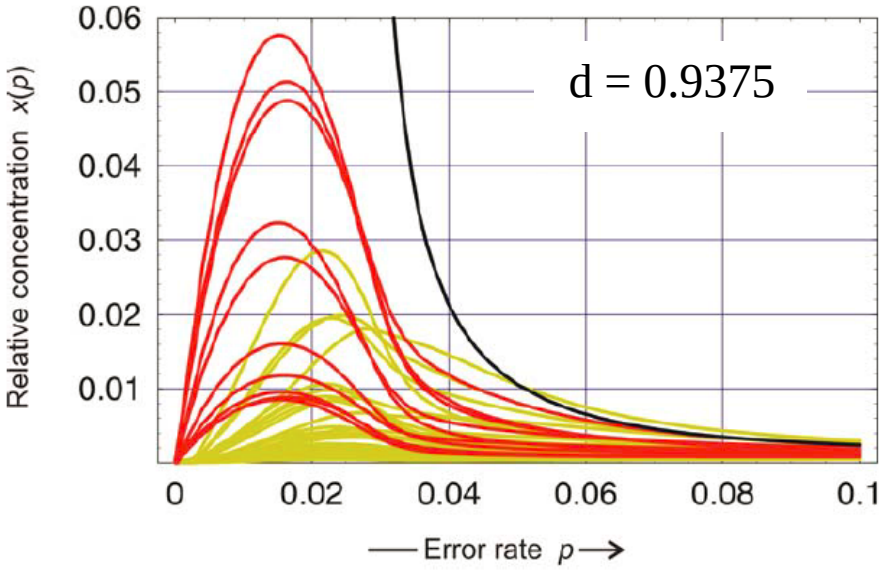


„realistic“ landscape

Rugged fitness landscapes over individual binary sequences with n = 10



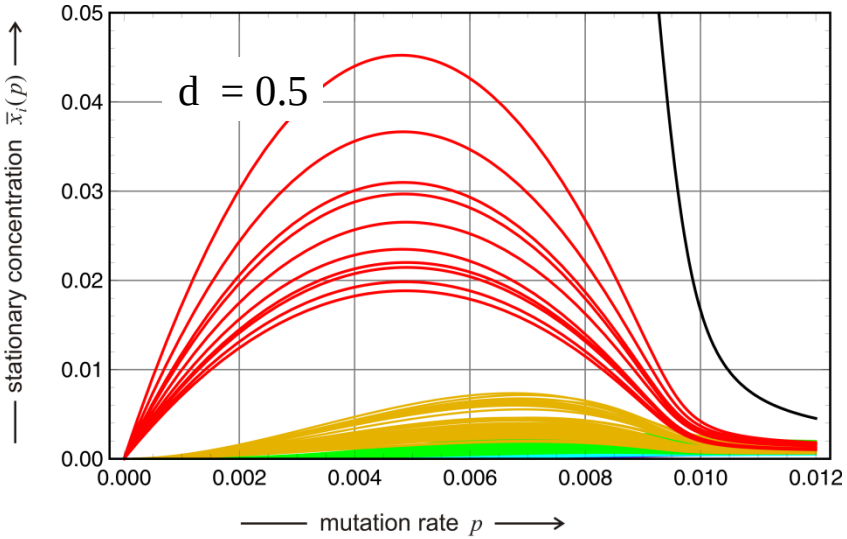


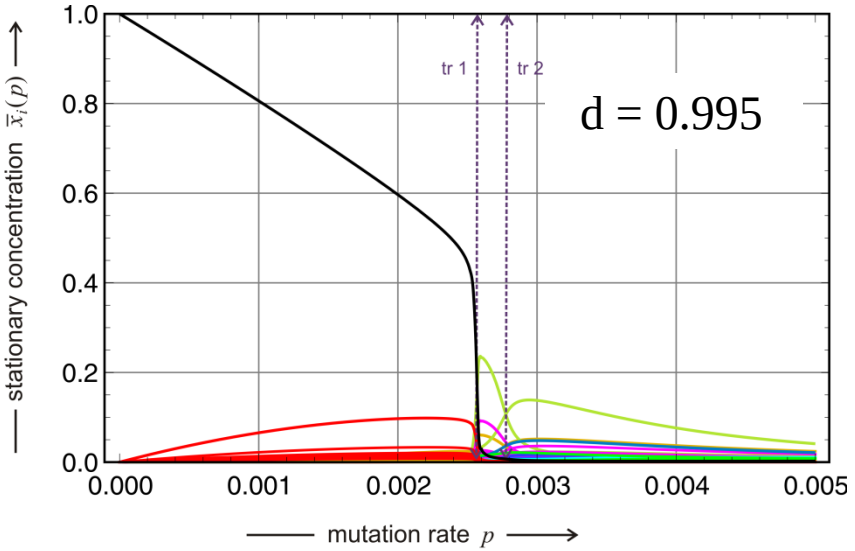


Quasispecies with increasing random scatter *d*

Error threshold: Individual sequences

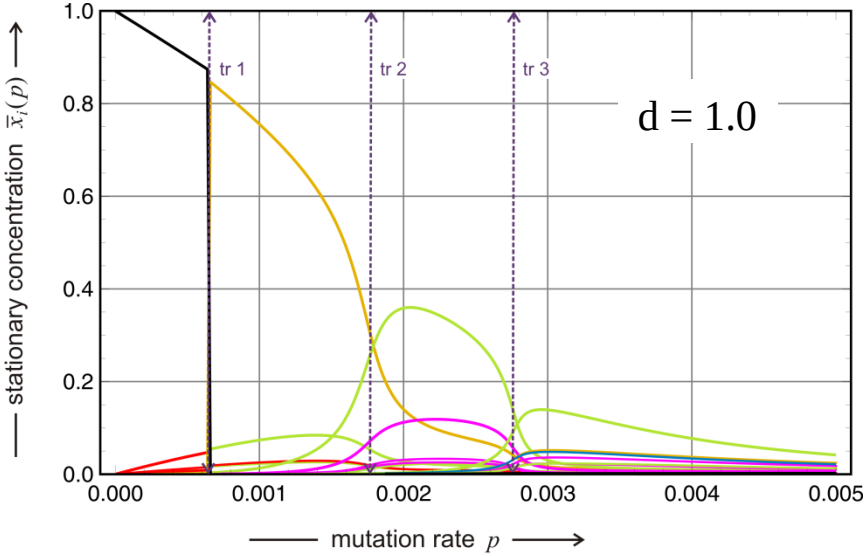
n = 10, σ = 2, s = 491 and d = 0, 0.5, 0.9375





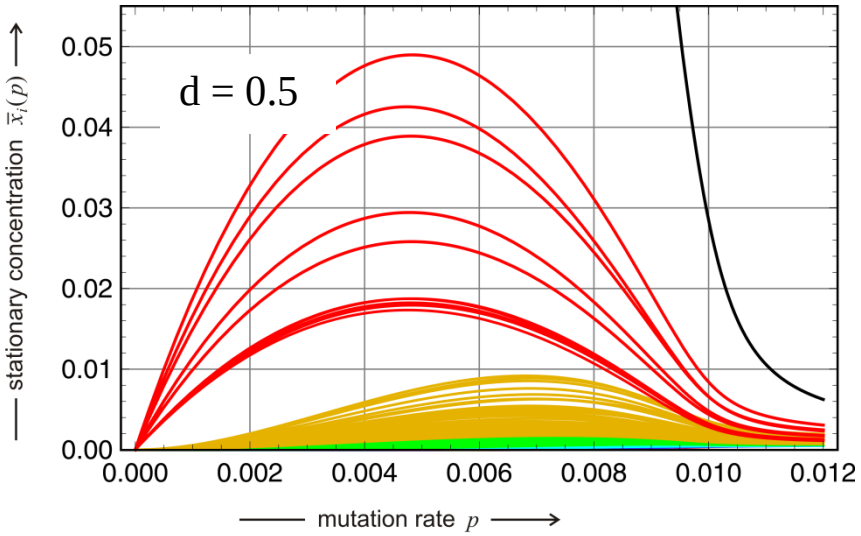
Choice of random scatter:

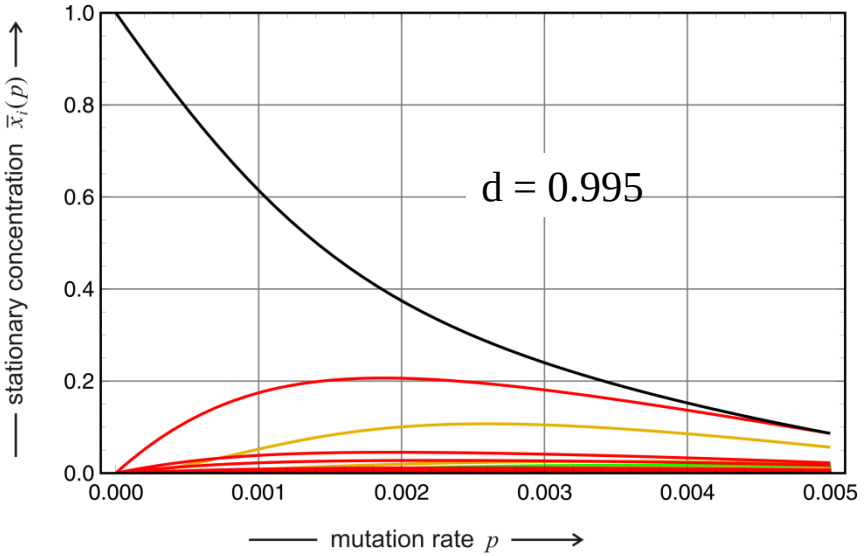
s = 637



Error threshold on ‚realistic' landscapes

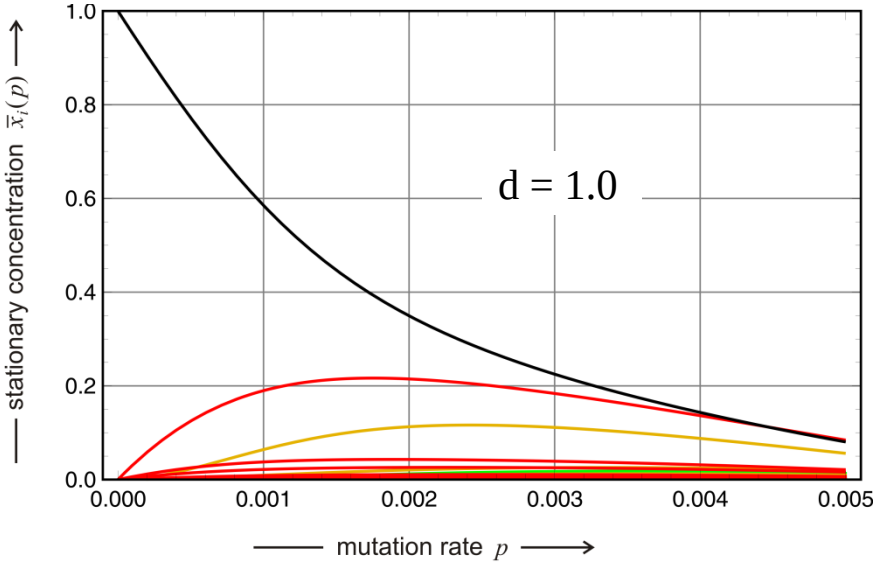
n = 10, $f_0 = 1.1$, $f_n = 1.0$, s = 637





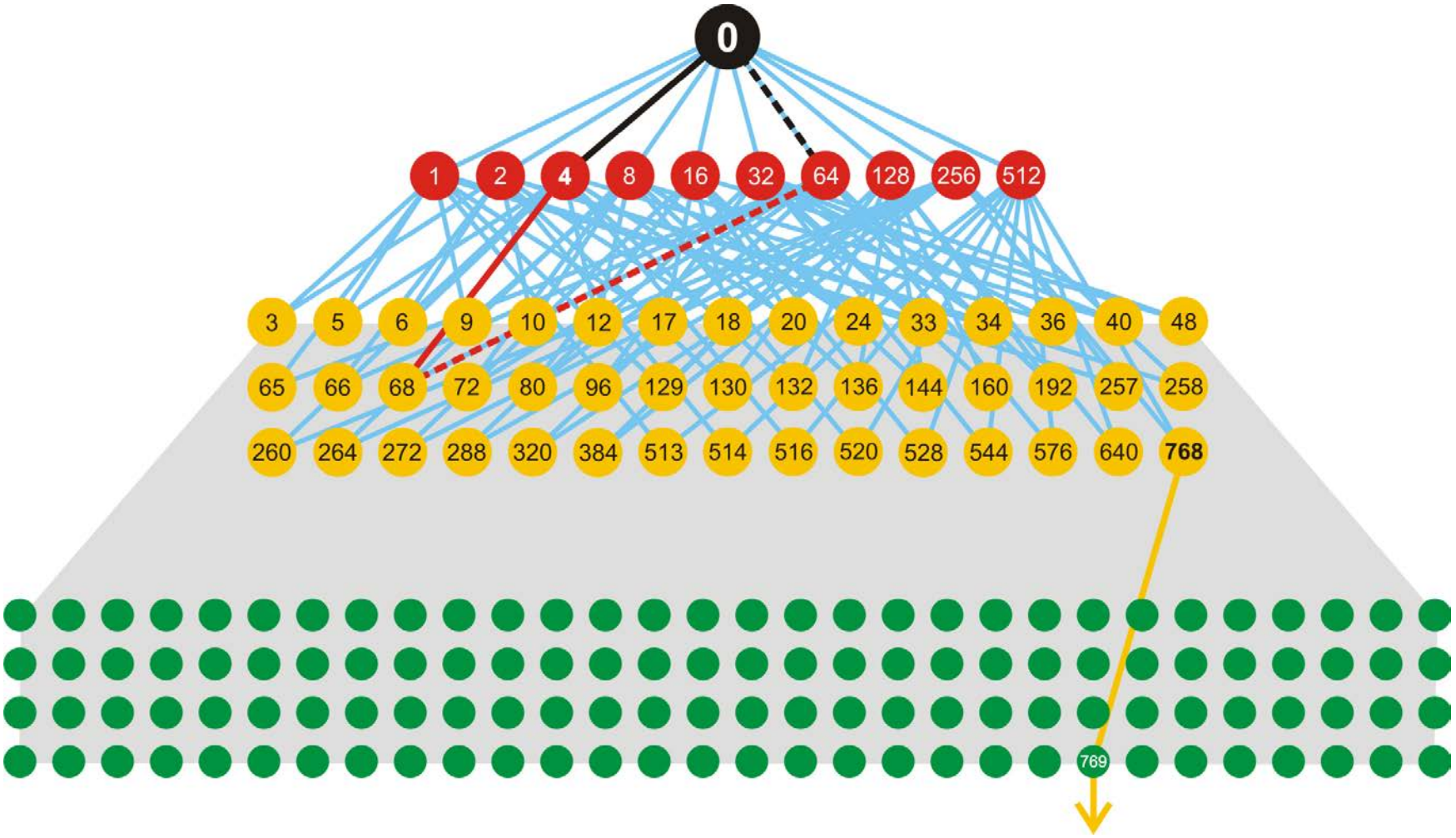
Choice of random scatter:

s = 919

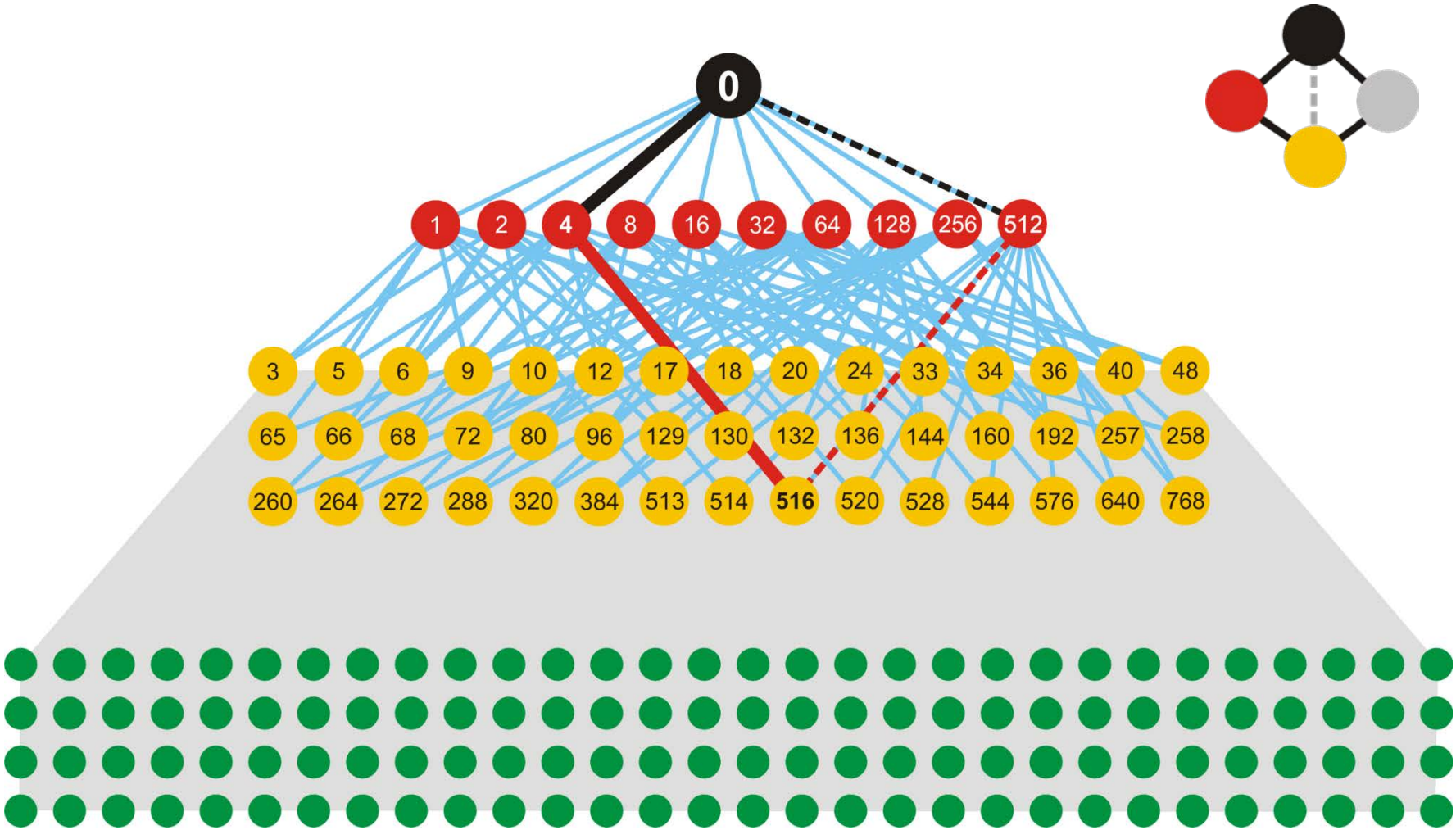


Error threshold on ‚realistic' landscapes

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



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



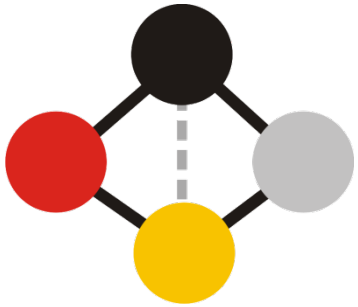
Determination of the dominant mutation flow: $d = 1$, $s = 919$

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. **Strong quasispecies** reproduce more efficiently, are **stable to mutation**, and should be **favored by evolution**.

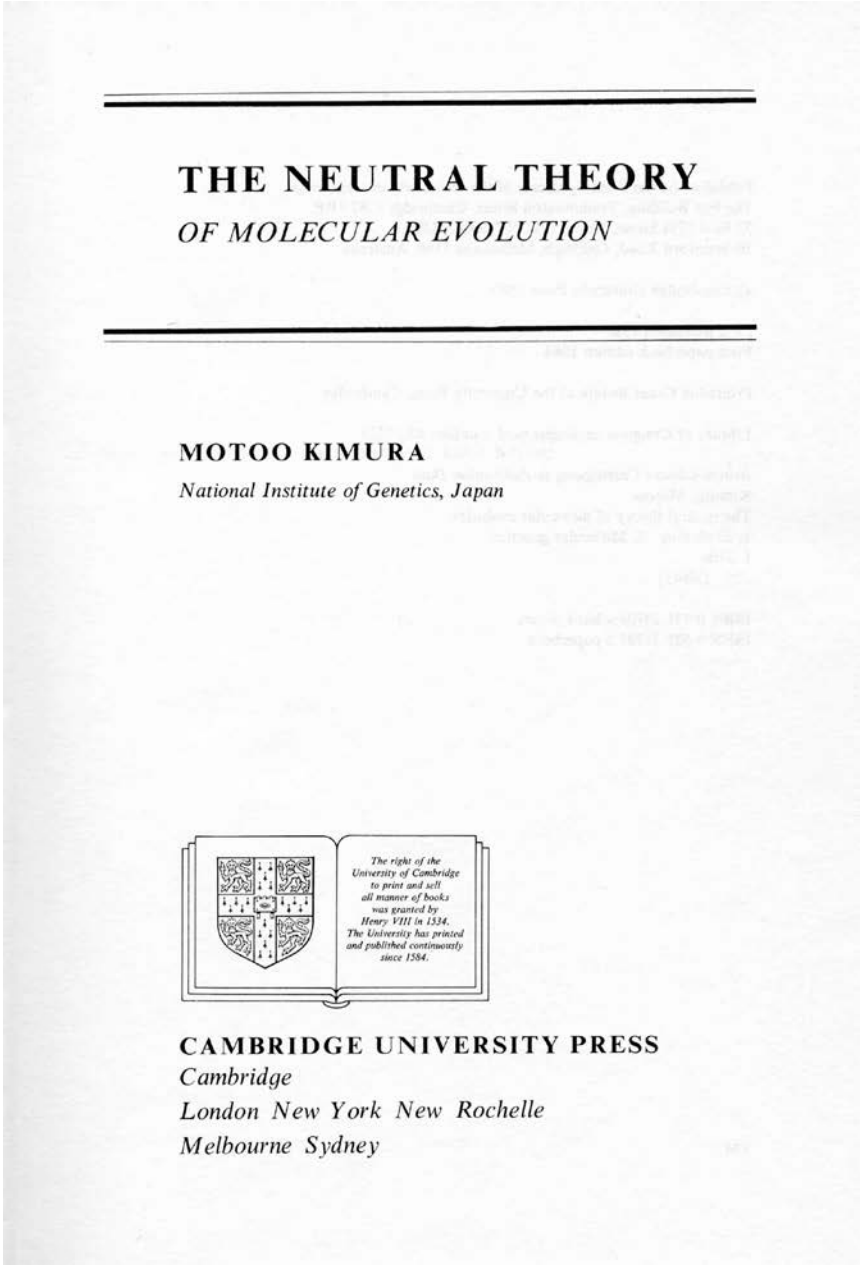


Motoo Kimura, 1924 - 1994

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

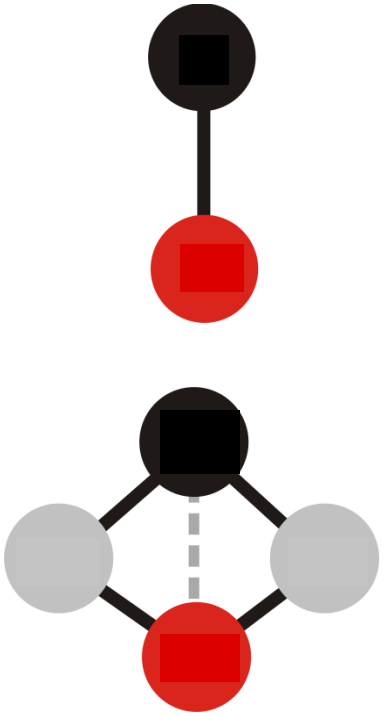
OF MOLECULAR EVOLUTION

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



$\mathbf{d_H = 1}$

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

$\mathbf{d_H = 2}$

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

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

$\mathbf{d_H \geq 3}$

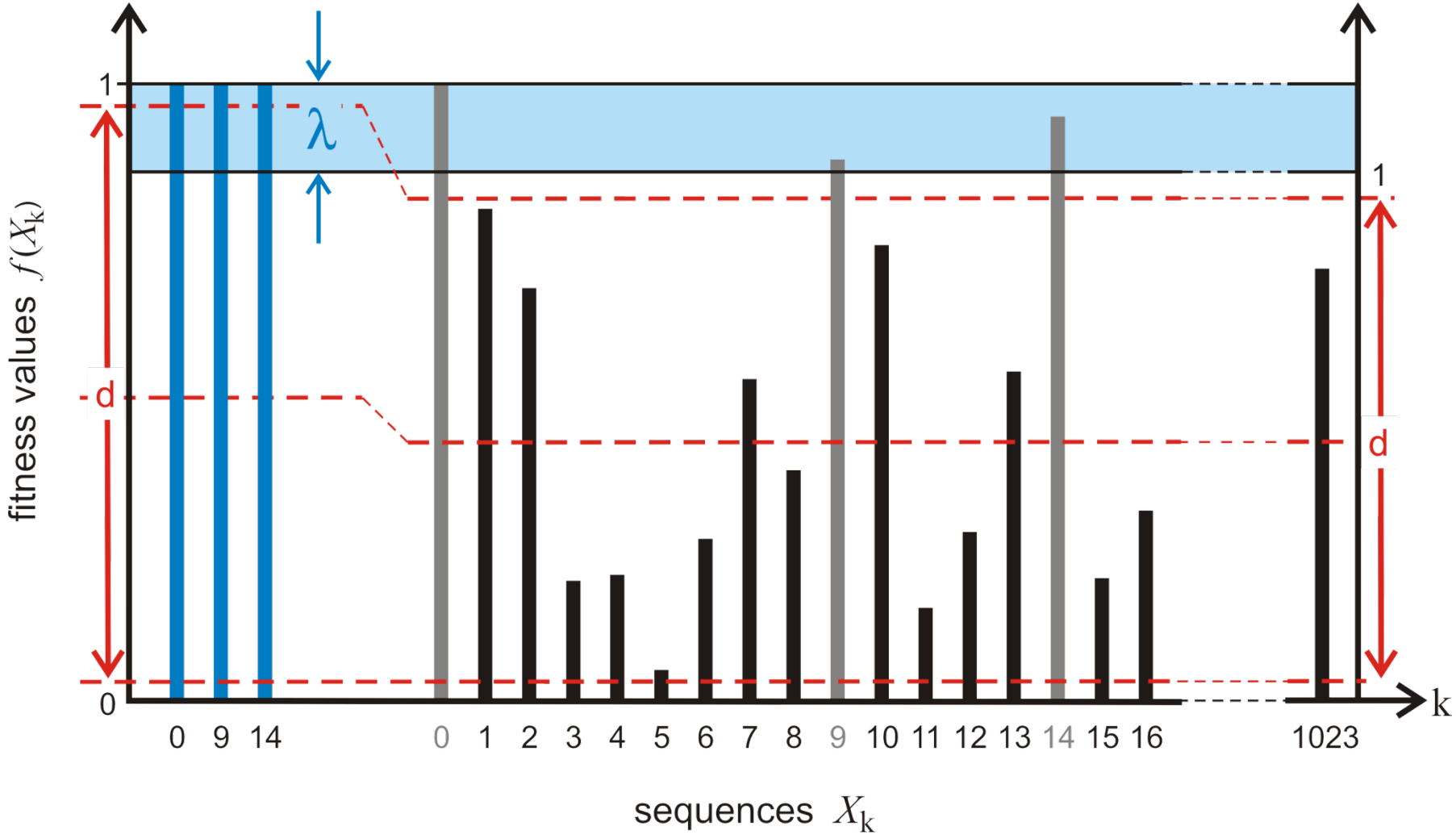
$$\lim_{p\to 0} x_1(p) = 1, \lim_{p\to 0} x_2(p) = 0 \quad \text{or}$$

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

Pairs of neutral sequences in replication networks

Random fixation in the sense of Motoo Kimura

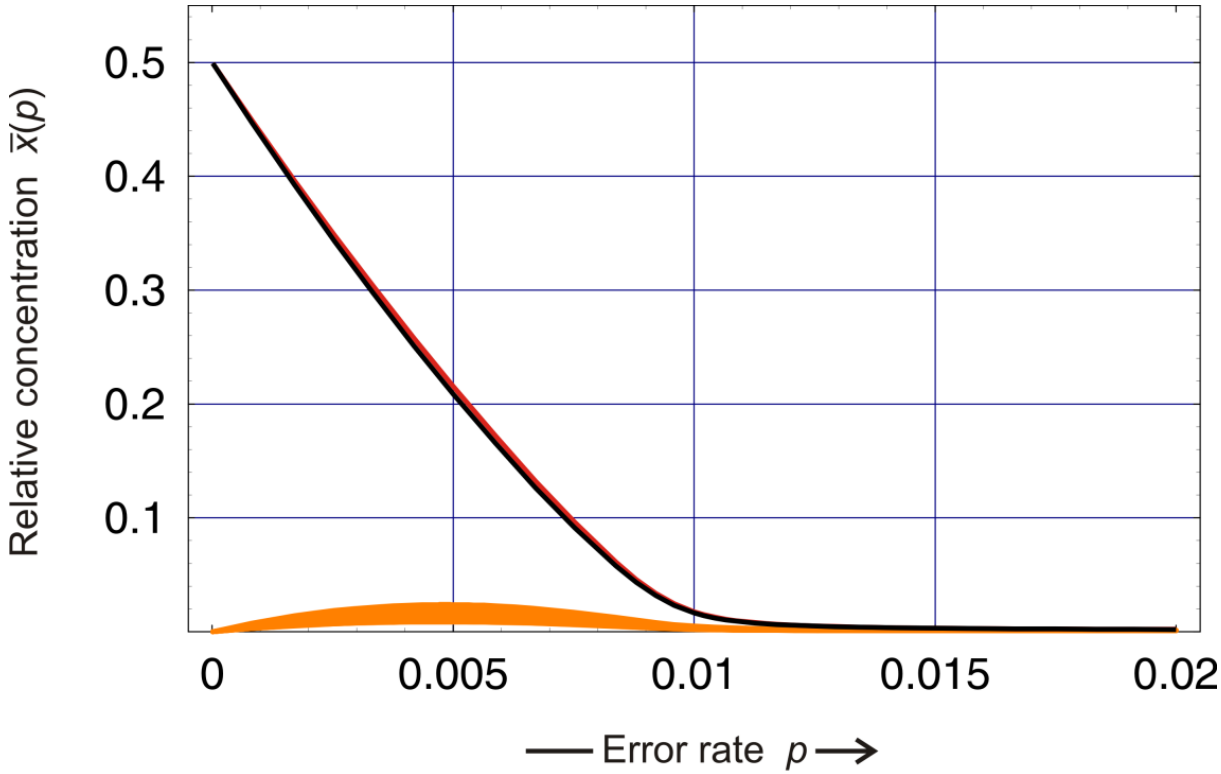
P. Schuster, J. Swetina. 1988. Bull. Math. Biol. 50:635-650



A fitness landscape including neutrality

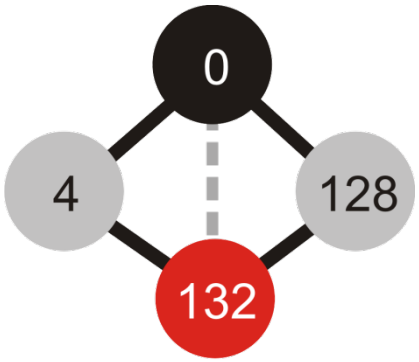
neutral network

$\lambda = 0.01,\ s = 367$



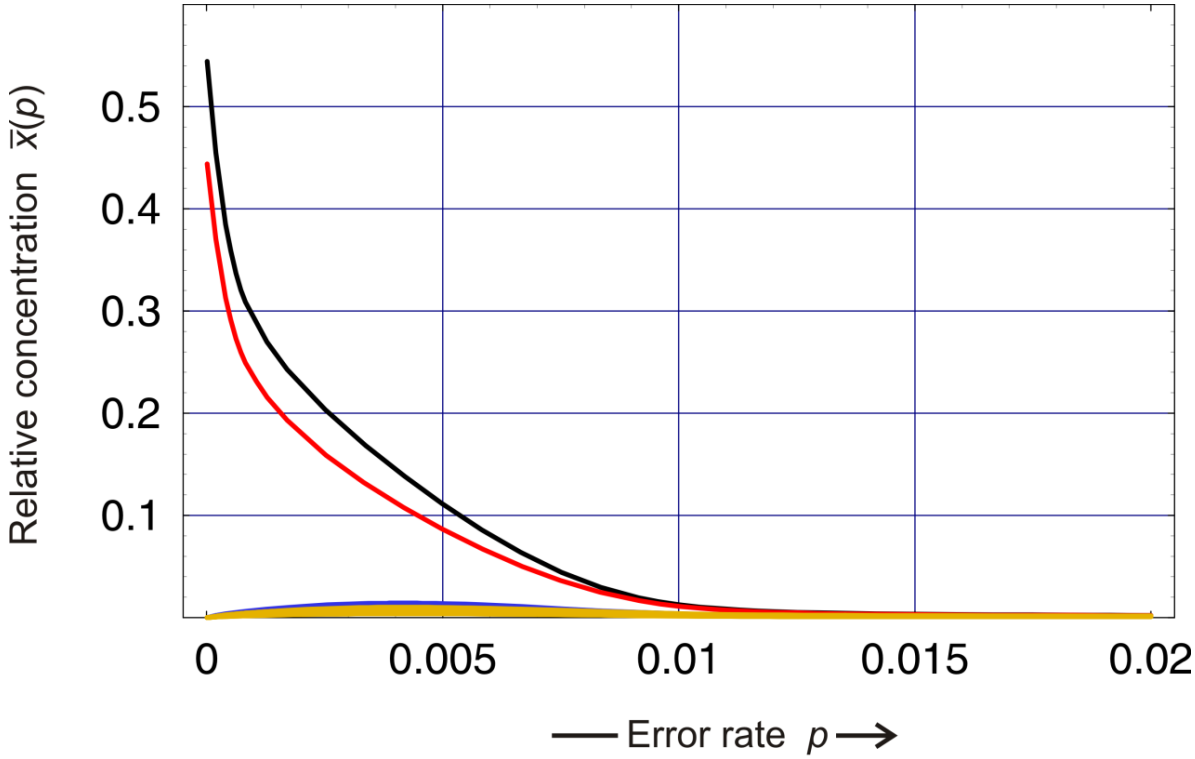
Neutral network: Individual sequences

n = 10, σ = 1.1, d = 1.0



neutral network

$\lambda = 0.01,\ s = 877$



Neutral network: Individual sequences

n = 10, σ = 1.1, d = 1.0

```
..... ACAUGCGAA .....
..... AUAUACGAA .....
..... ACAUGCGCA .....
..... GCAUACGAA .....
..... ACAUGCUAA .....
..... ACAUGCGAG .....
..... ACACGCGAA .....
..... ACGUACGAA .....
..... ACAUAGGAA .....
..... ACAUACGAA .....
```

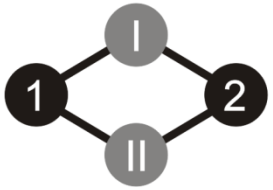
master sequence 1

master sequence 2

..... ACAU$^{G}_{A}$CGAA ..... consensus sequence

```
..... ACAGUCAGAA .....
..... ACAGUCCGAA .....
..... AUAAUCCGAA .....
..... ACAGUCAGCA .....
..... GCAGUCAGAA .....
..... ACAGUCAUAA .....
..... ACAGUCAGAG .....
..... ACAACCCGAA .....
..... ACGGUCAGAA .....
..... ACAGUGAGAA .....
..... ACAAUCAGAA .....
..... ACAAUCCGAA .....
```

master sequence 1
intermediate I



intermediate II
master sequence 2

..... ACA$^{G}_{A}$UC$^{A}_{C}$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.

Conclusions

1. Realistic **fitness landscapes** sustain **error thresholds**.

2. **Quasispecies** may be centered around **clans of sequences** with high fitness, which provide **evolutionary stability** against increasing **mutation rates**.

3. Pairs of **neutral sequences** with **Hamming distances one** or **two** form clans and are **not subjected** to Kimura's **random selection**.

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

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

Universität Wien

Thank you for your attention!

Web-Page for further information:

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

Preprint:

Santa Fe Institute Working Paper: # 12-06-12