Bioinformatics and the molecular connection to biology

Peter Schuster

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

10 Years of Bioinformatics in Leipzig

Leipzig, 20.09.2012

Web-Page for further information:

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

Prologue

There will never be a Newton of the blade of grass, because human science will never be able to explain how a living being can originate from inanimate matter.

Immanuel Kant, 1790

Three interpretations of Kant‘s „*Newton of the blade of grass*“:

(i) Life science will never be explainable by methods based on physics and chemistry.

(ii) Origin of life questions are outside science.

(iii) Application of mathematics to biology leads nowhere.

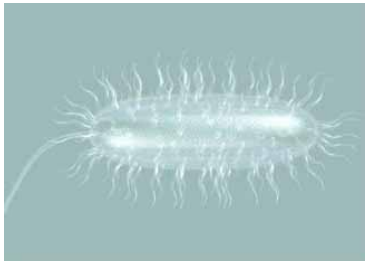
I maintain only that in every special doctrine of nature only so much science can be found as there is mathematics in it.

Three historical examples of using mathematics in biology

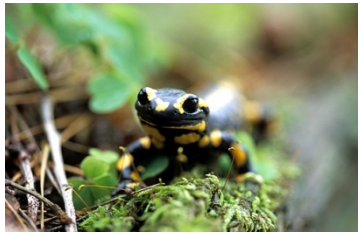
1. The case that did not happen – Charles Darwin
2. Blind insight or correct guess – Gregor Mendel
3. Nature has chosen a less elegant way – Alan Turing

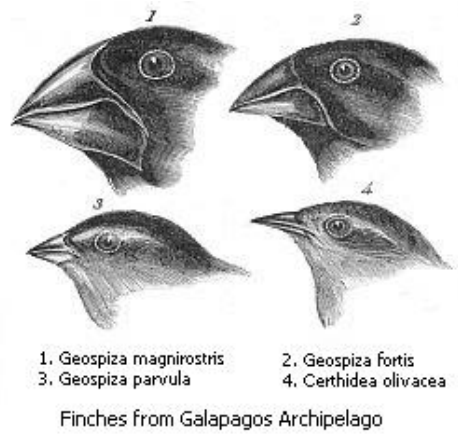
Charles Darwin, 1809 - 1882

Voyage on HMS Beagle, 1831 - 1836

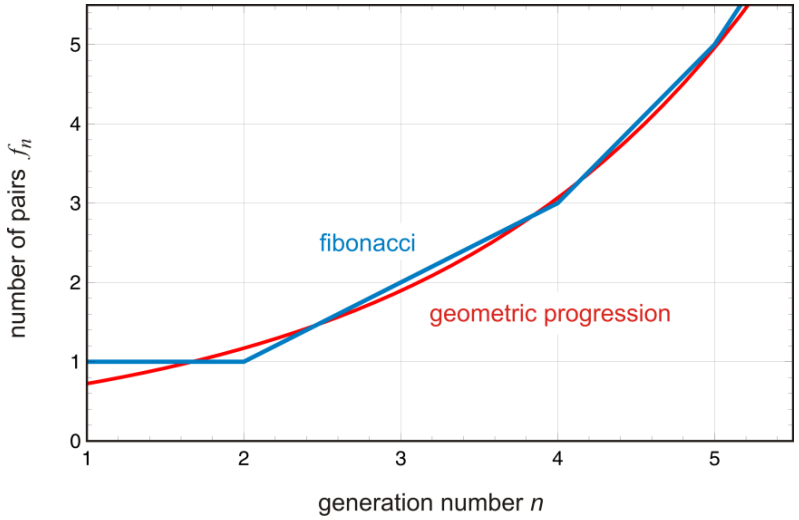


Phenotypes





Finches from Galapagos Archipelago



Fibonacci series

Leonardo da Pisa
„Fibonacci“
~1180 – ~1240

geometric progression

Thomas Robert Malthus,
1766 – 1834

exponential function

Leonhard Euler,
1717 – 1783

autocatalysis

$$\frac{dx}{dt} = f\, x(1-x) \quad \Rightarrow \quad x(t) = x(0)\exp(f\,t)$$

competition

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

$$x_k(t) = x_k(0)\ \exp(f_k\, t)$$

The chemistry and the mathematics of reproduction

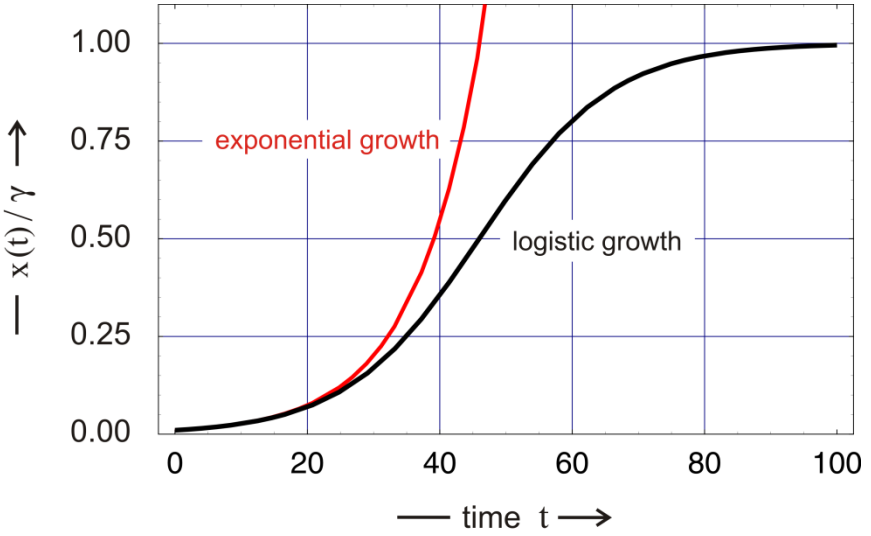
Pierre-François Verhulst,
1804-1849

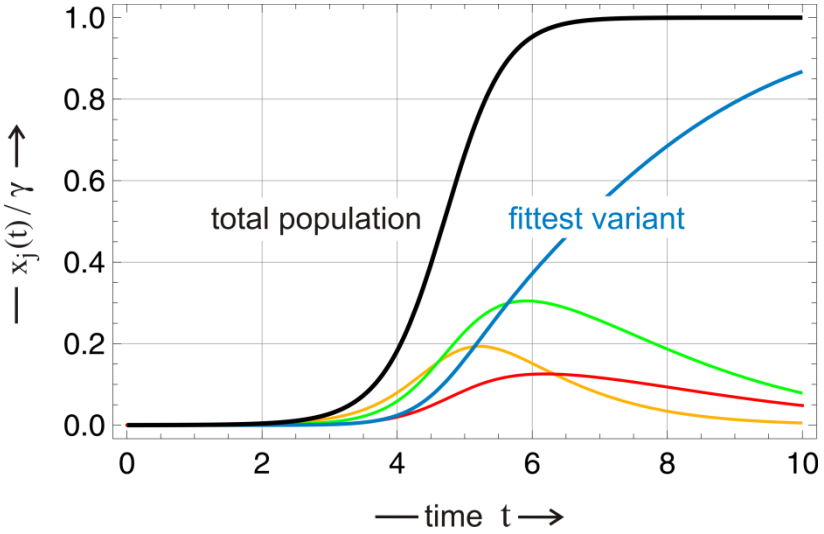
$$\frac{dx}{dt} = f\,x\left(1-\frac{x}{\gamma}\right)$$

$$x(t) = \frac{x_0\,\gamma}{x_0 + (\gamma + x_0)\exp(-ft)}$$

the consequence of finite resources

The logistic equation, 1828





fitness values:

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

Three necessary conditions for Darwinian evolution are:

1. **Multiplication,**
2. **Variation**, and
3. **Selection**.

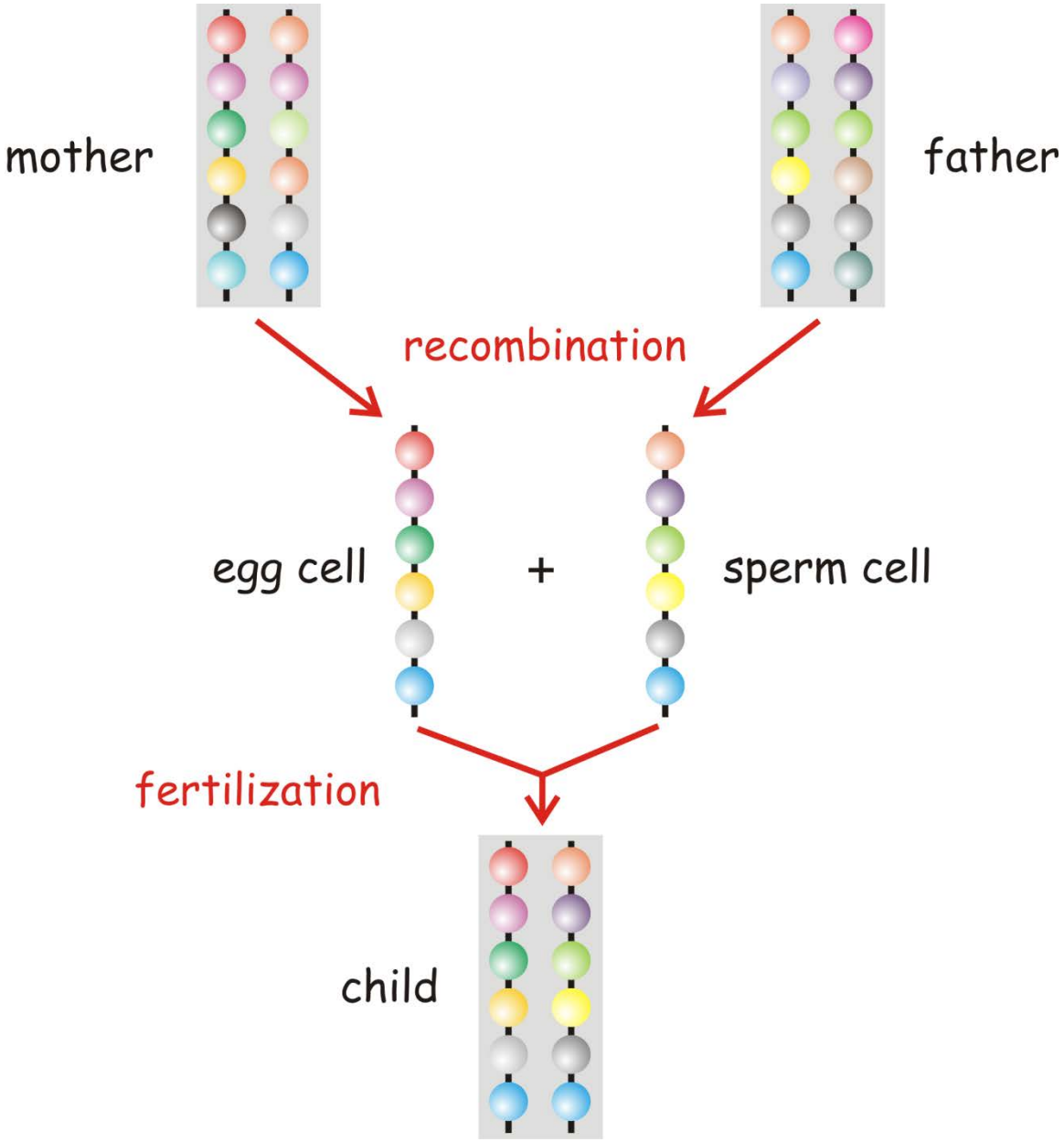
Multiplication is common to all forms of evolving life.

Variation occurs through mutation and recombination.

Selection is a trivial consequence of the finiteness of resources.

One important property of the Darwinian scenario is that variations in the form of mutations or recombination events occur **uncorrelated** with their effects on the selection process.

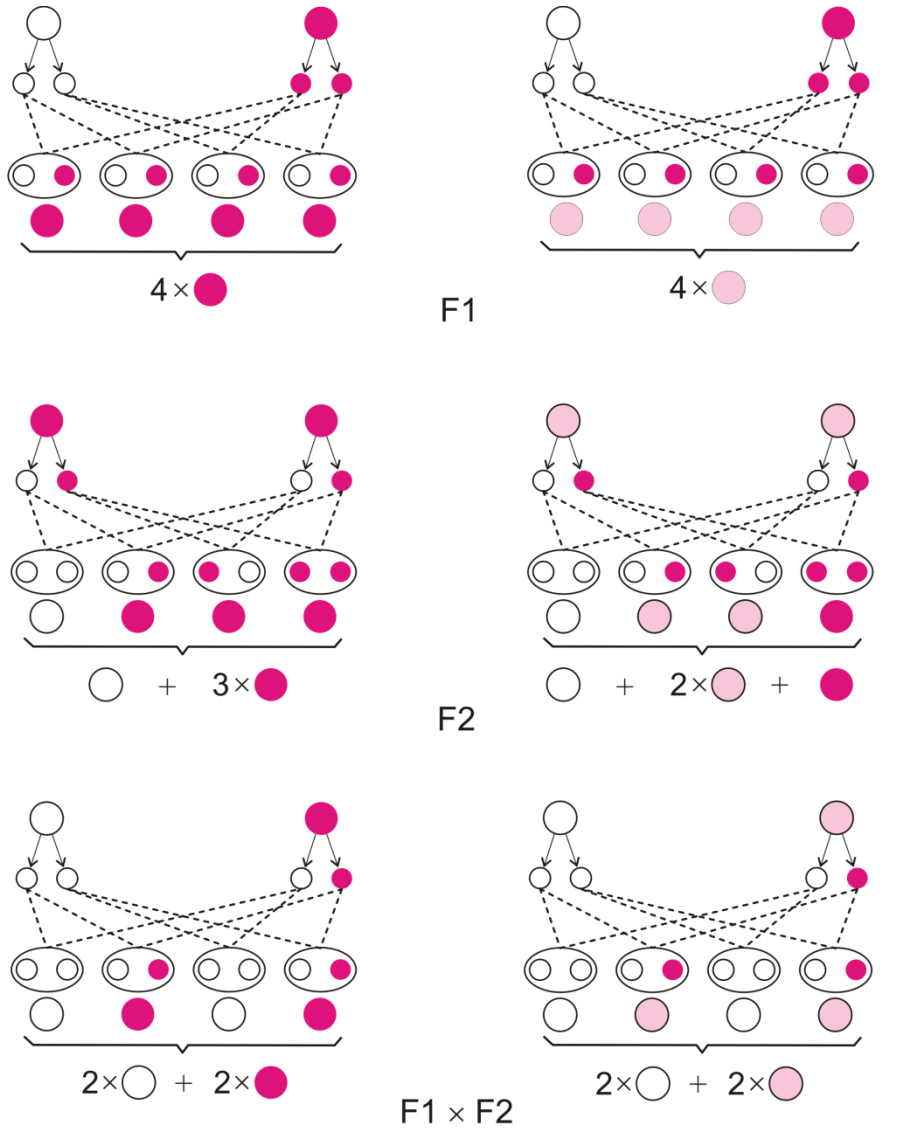
Gregor Mendel
1822 - 1884



Recombination in Mendelian genetics

Mendelian genetics

The 1:3 rule



dominant/recessive pair of alleles

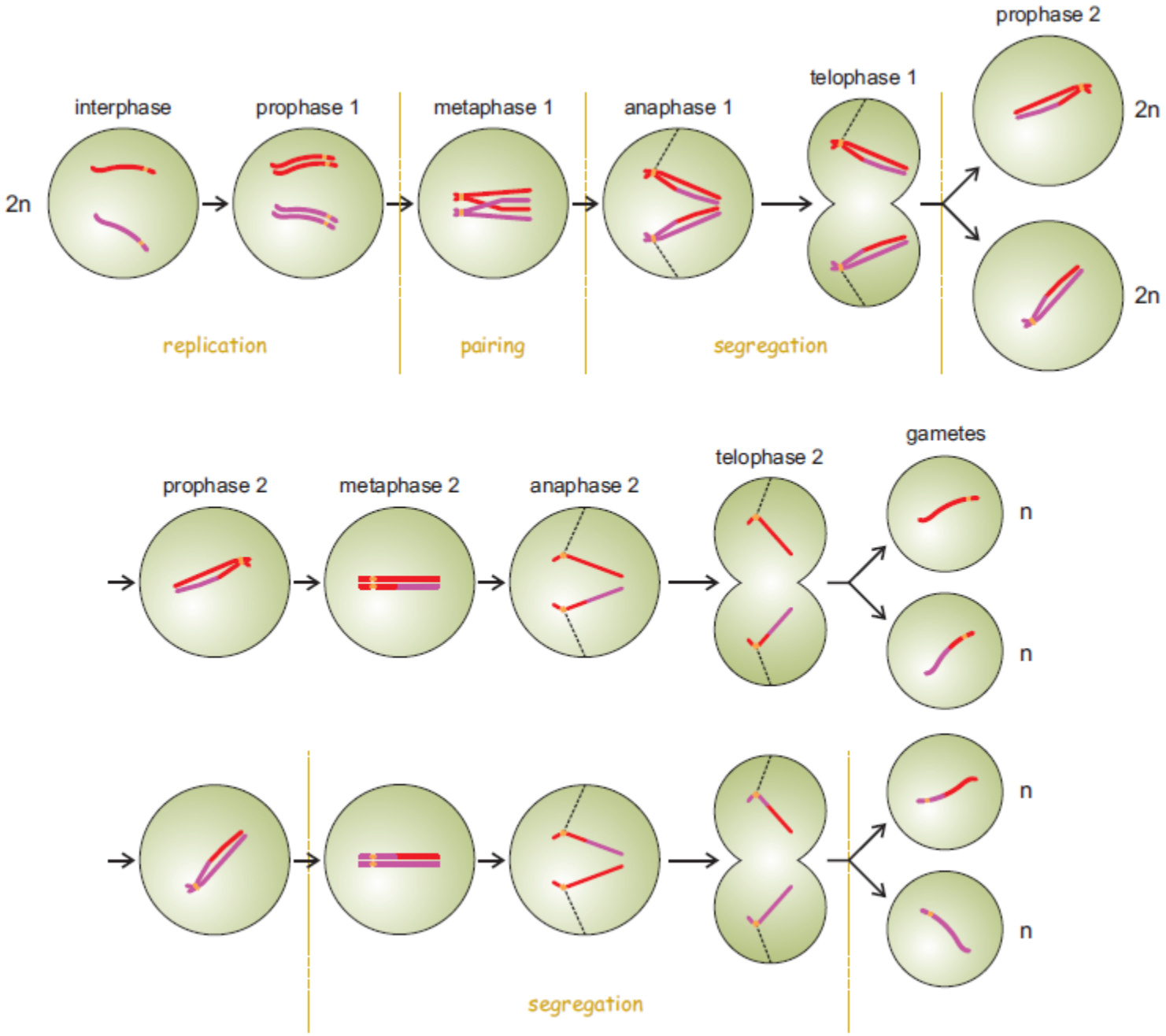
dominance

intermediate pair of alleles

semi-dominance

Char.	Parental phenotype	F1	F2	F2 ratio
1	round × wrinkled seeds	all round	5174 / 1859	2.96
2	yellow × green seeds	all yellow	6022 / 2001	3.01
3	purple × white petals	all purple	705 / 244	3.15
4	inflated × pinched pods	all inflated	882 / 299	2.95
5	green × yellow pods	all green	428 / 152	2.82
6	axial × terminal flowers	all axial	651 / 207	3.14
7	long × short stems	all axial	787 / 277	2.84

The results of the individual experiments Gregor Mendel did with the garden pea *pisum sativum*.


interphase
prophase 1
metaphase 1
anaphase 1
telophase 1
prophase 2
2n
2n
2n
replication
pairing
segregation
prophase 2
metaphase 2
anaphase 2
telophase 2
gametes
n
n
n
n
segregation

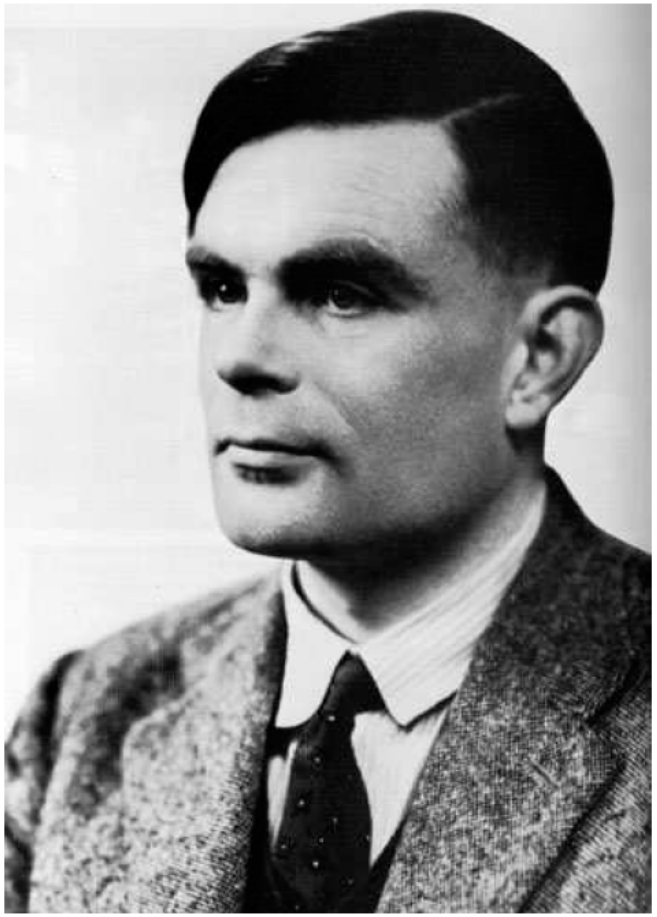
$$\frac{\partial u}{\partial t} = D_u \nabla^2 u + f(u,v)$$

$$\frac{\partial v}{\partial t} = D_v \nabla^2 v + g(u,v)$$

$$u = u(x,y,z,t) \quad \text{and} \quad v = v(x,y,z,t)$$

Change in local concentration =

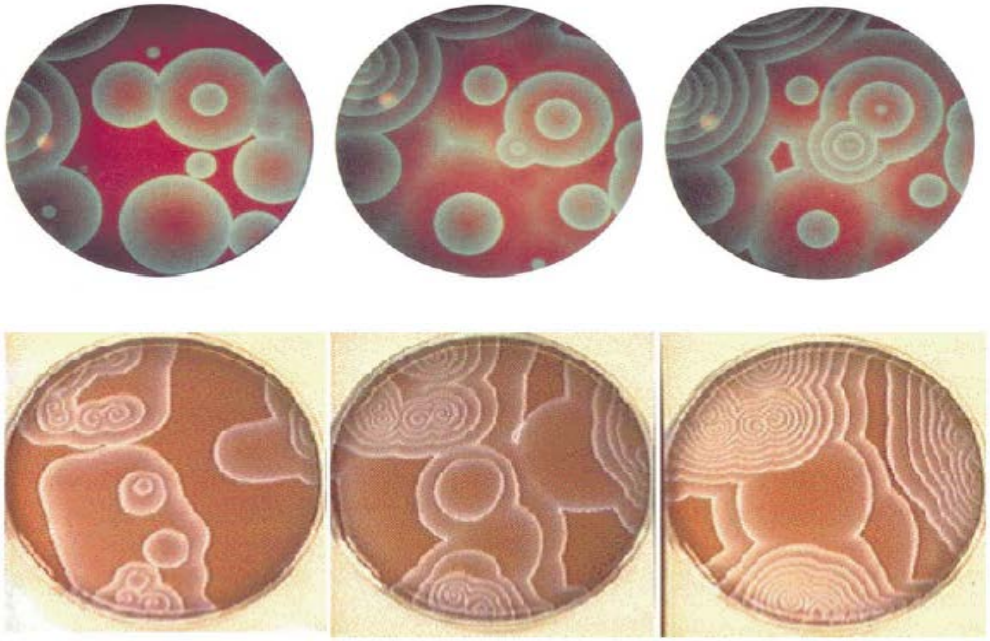
= diffusion + chemical reaction



Alan M. Turing, 1912-1954

A.M. Turing. 1952. The chemical basis of morphogenesis. *Phil.Trans.Roy.Soc.*London B **237**:37-72.

Liesegang rings 1895



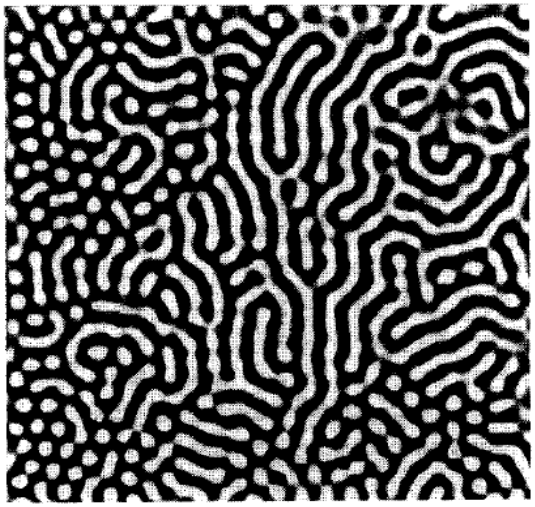
Belousov-Zhabotinskii reaction 1959

Pattern formation through chemical self-organisation:

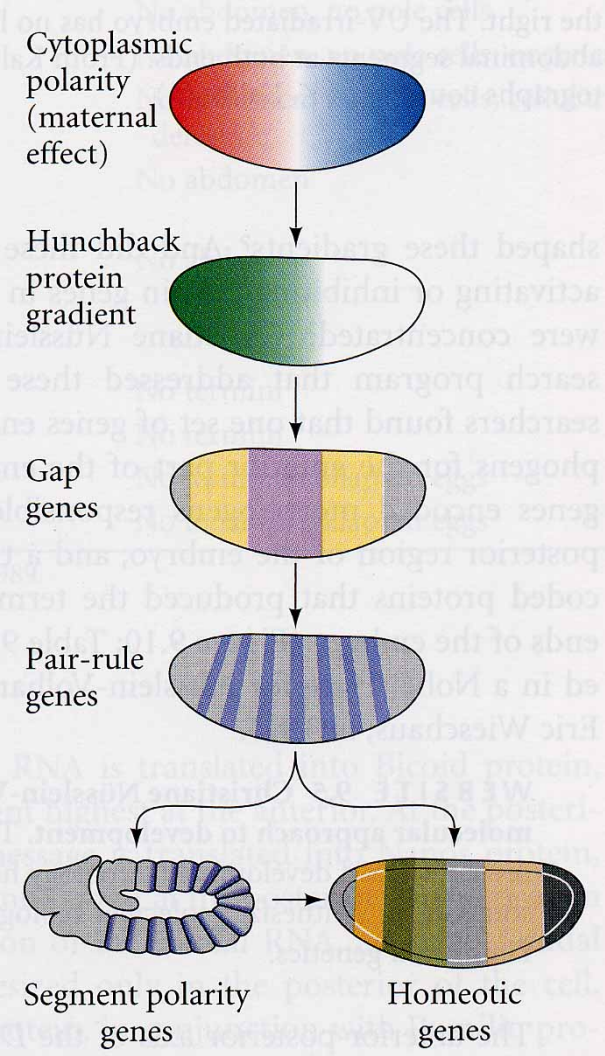
Liesegang rings through crystallisation from supersaturated solutions,

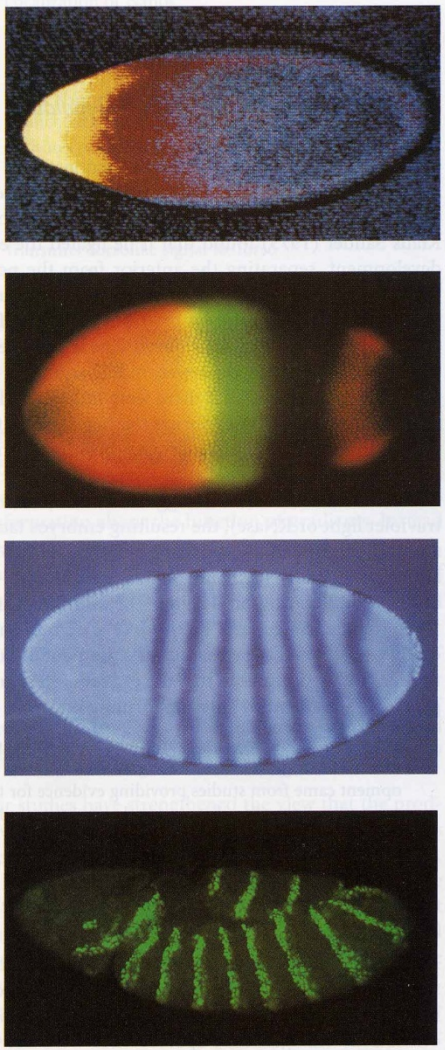
space-time-pattern in the Belousov-Zhabotinskii reaction,

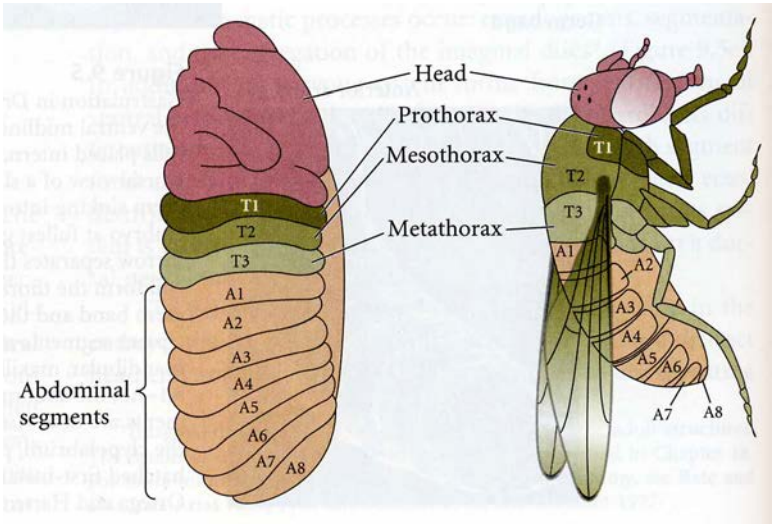
and stationary Turing pattern,



Turing pattern:
Boissonade, De Kepper 1990







Development of the fruit fly *drosophila melanogaster*: Genetics, experiment, and imago

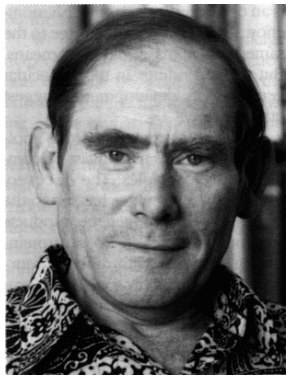
Philip K. Maini, 1959 -

More recently, detailed experimental work on *Drosophila* has shown that the pattern forming process is not, in fact, *via* reaction diffusion, but due to a cascade of gene switching, where certain gene proteins are expressed and, in turn, influence subsequent gene expression patterns. **Therefore, although reaction diffusion theory provides a very elegant mechanism for segmentation nature has chosen a much less elegant way of doing it!**

Philip K. Maini, Kevin J. Painter, and Helene Nguyen Phong Chau. 1997.
Spatial Pattern Formation in Chemical and Biological Systems
J.Chem.Soc., Faraday Transactions **93**:3601-3610.

Unfortunately, **theoretical biology** has a bad name because of its past. Physicists were concerned with questions such as whether biological systems are compatible with the second law of thermodynamics and whether the could be explained by quantum mechanics. Some even expected biology to reveal the presence of new laws of physics. There have also been attempts to seek general mathematical theories of development and of the brain: The application of catastrophe theory is but one example. Even though alternatives have been suggested, such as **computational biology**, **biological systems theory** and **integrative biology**, I have decided to forget and forgive the past and call it

theoretical biology.

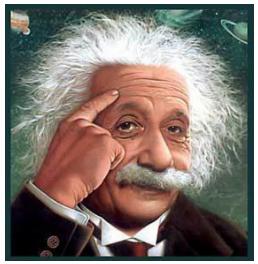


Sydney Brenner, 1999

Theoretical biology in the third millenium.
Phil.Trans.Roy.Soc.London B 354:1963-1965

Biological evolution of higher organisms is an exceedingly complex process not because the mechanism of selection is complex but because cellular metabolism and control of organismic functions is highly sophisticated.

The Darwinian mechanism of selection does neither require organisms nor cells for its operation.



Make things as simple as possible,
but not simpler.

Albert Einstein, 1950 (?)

Pluralitas non est ponenda
sine neccesitate.
Ockham's razor.

William of Ockham, c.1288 – c.1348

Sir William Hamilton, 1852

Replicating molecules

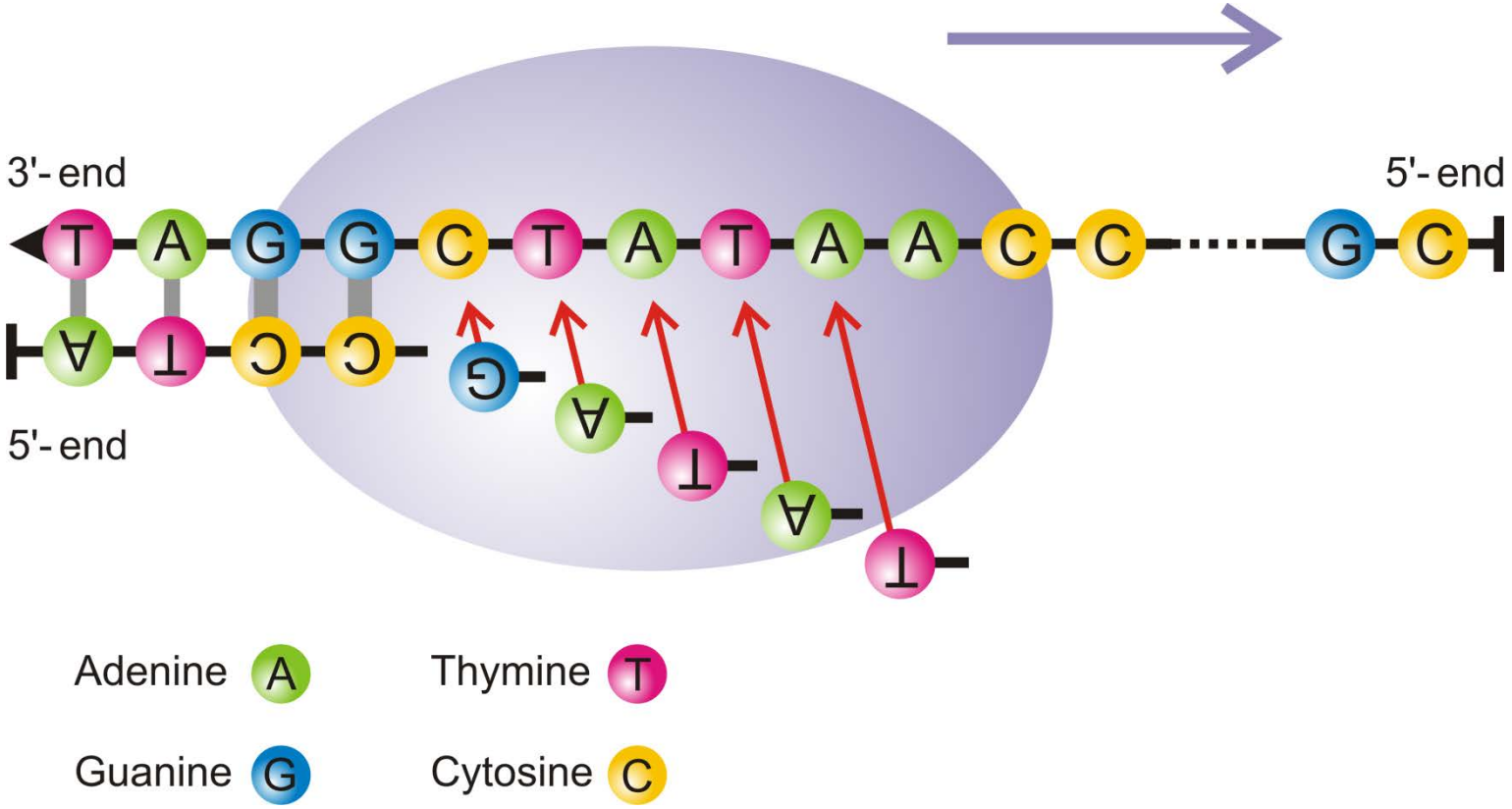
Charles Darwin, 1809-1882

Three necessary conditions for Darwinian evolution are:

1. **Multiplication,**
2. **Variation**, and
3. **Selection**.

All three conditions are fulfilled not only by cellular organisms but also by **nucleic acid molecules** – **DNA** or **RNA** – **in** suitable **cell-free experimental assays**:

Darwinian evolution in the test tube



The replication of DNA by Thermophilus aquaticus polymerase (PCR)

Accuracy of replication: $Q = q_1 \cdot q_2 \cdot q_3 \cdot q_4 \cdot \ldots$

The logics of DNA (or RNA) replication

Evolution in the test tube:

G.F. Joyce, *Angew.Chem.Int.Ed.* **46** (2007), 6420-6436

DOI: 10.1002/anie.200701369

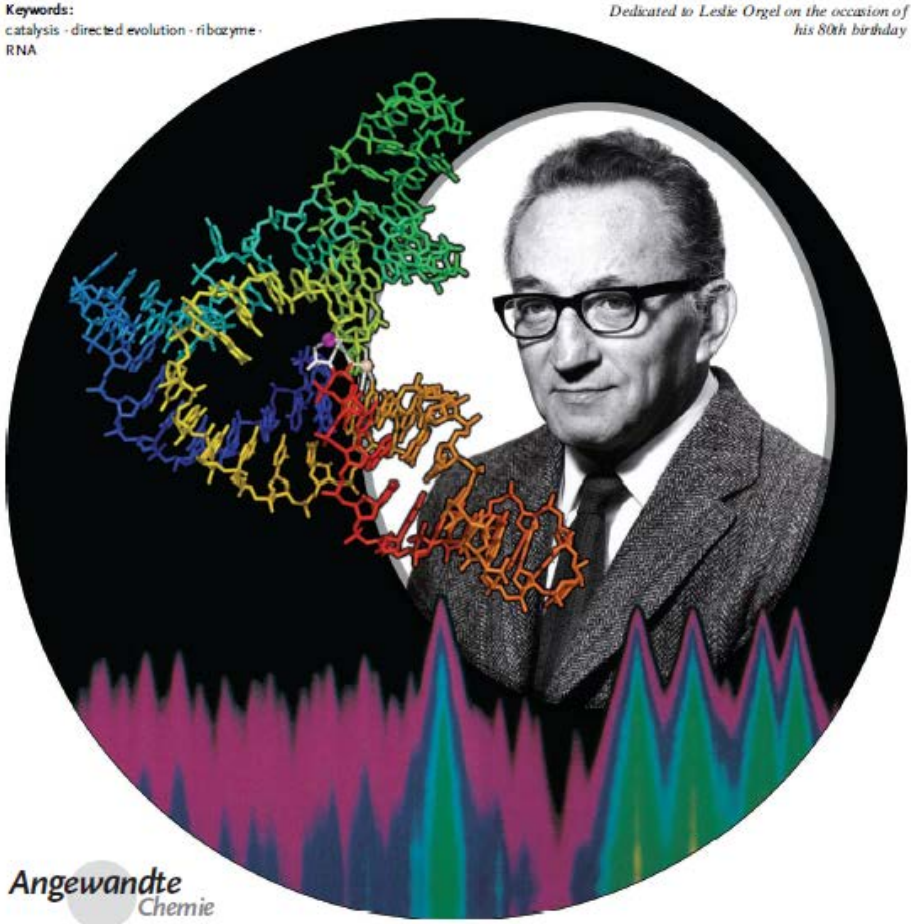
Molecular Evolution

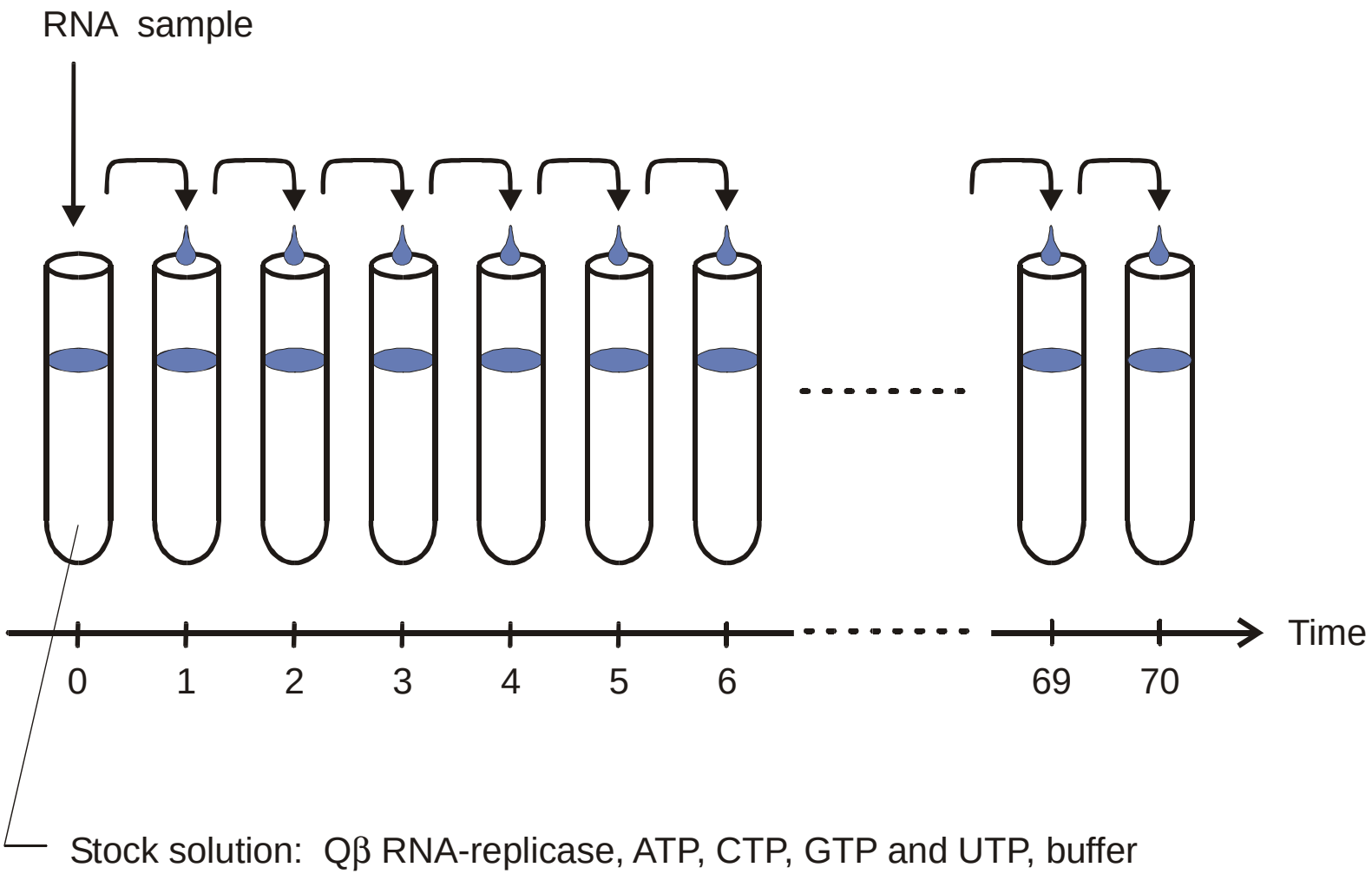
Forty Years of In Vitro Evolution**

*Gerald F. Joyce**

Keywords:
catalysis · directed evolution · ribozyme · RNA

Dedicated to Leslie Orgel on the occasion of his 80th birthday





Application of serial transfer technique to evolution of RNA in the test tube

Reproduction of the original figure of the serial transfer experiment with Qβ RNA

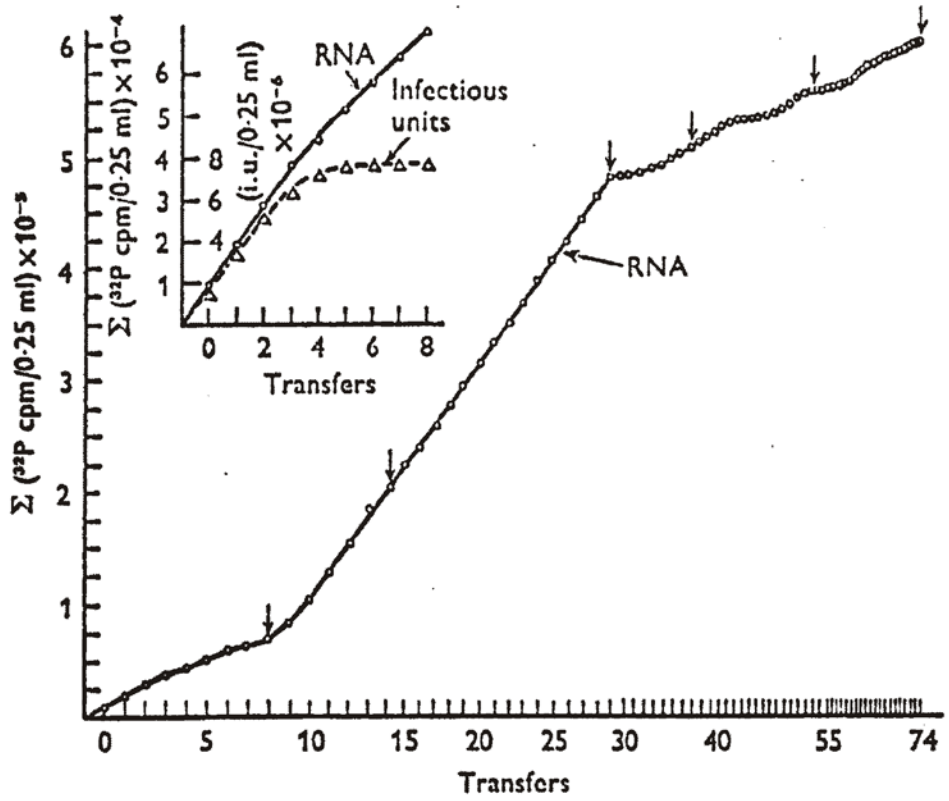
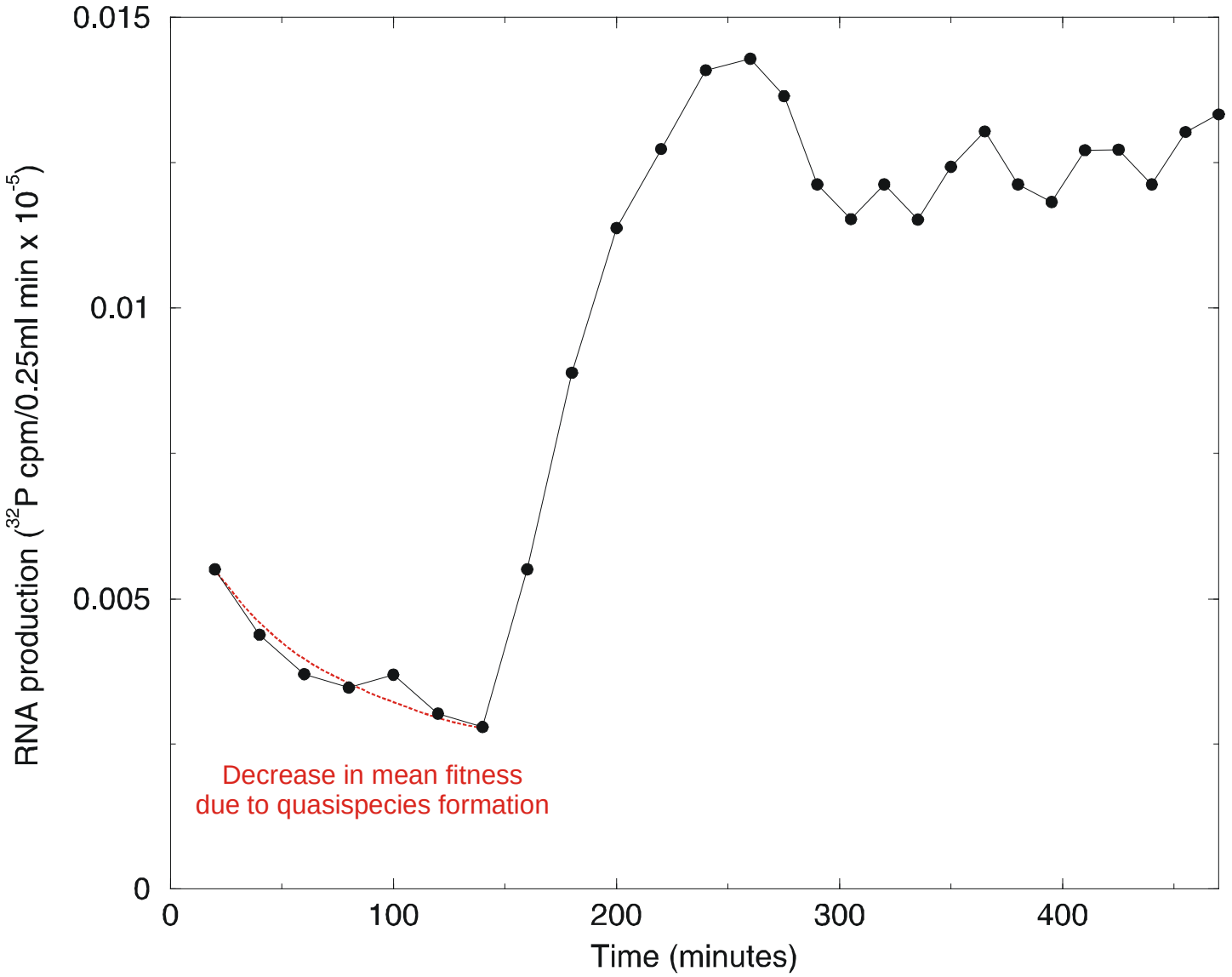


Fig. 9. Serial transfer experiment. Each 0·25 ml standard reaction mixture contained 40 μg of Qβ replicase and ^{32}P-UTP. The first reaction (0 transfer) was initiated by the addition of 0·2 μg ts-1 (temperature-sensitive RNA) and incubated at 35 °C for 20 min, whereupon 0·02 ml was drawn for counting and 0·02 ml was used to prime the second reaction (first transfer), and so on. After the first 13 reactions, the incubation periods were reduced to 15 min (transfers 14–29). Transfers 30–38 were incubated for 10 min. Transfers 39–52 were incubated for 7 min, and transfers 53–74 were incubated for 5 min. The arrows above certain transfers (0, 8, 14, 29, 37, 53, and 73) indicate where 0·001–0·1 ml of product was removed and used to prime reactions for sedimentation analysis on sucrose. The inset examines both infectious and total RNA. The results show that biologically competent RNA ceases to appear after the 4th transfer (Mills *et al.* 1967).

D.R.Mills, R,L,Peterson, S.Spiegelman, *An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule*. Proc.Natl.Acad.Sci.USA **58** (1967), 217-224

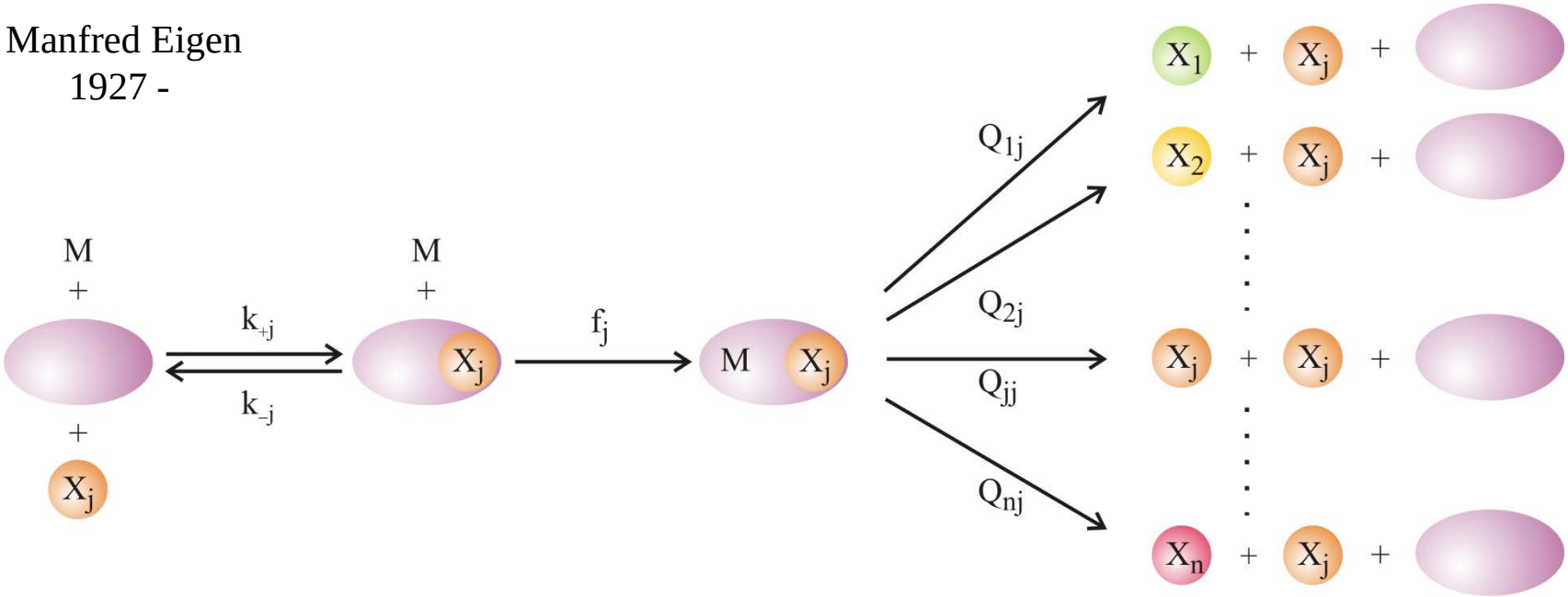


The increase in RNA production rate during a serial transfer experiment

Manfred Eigen
1927 -

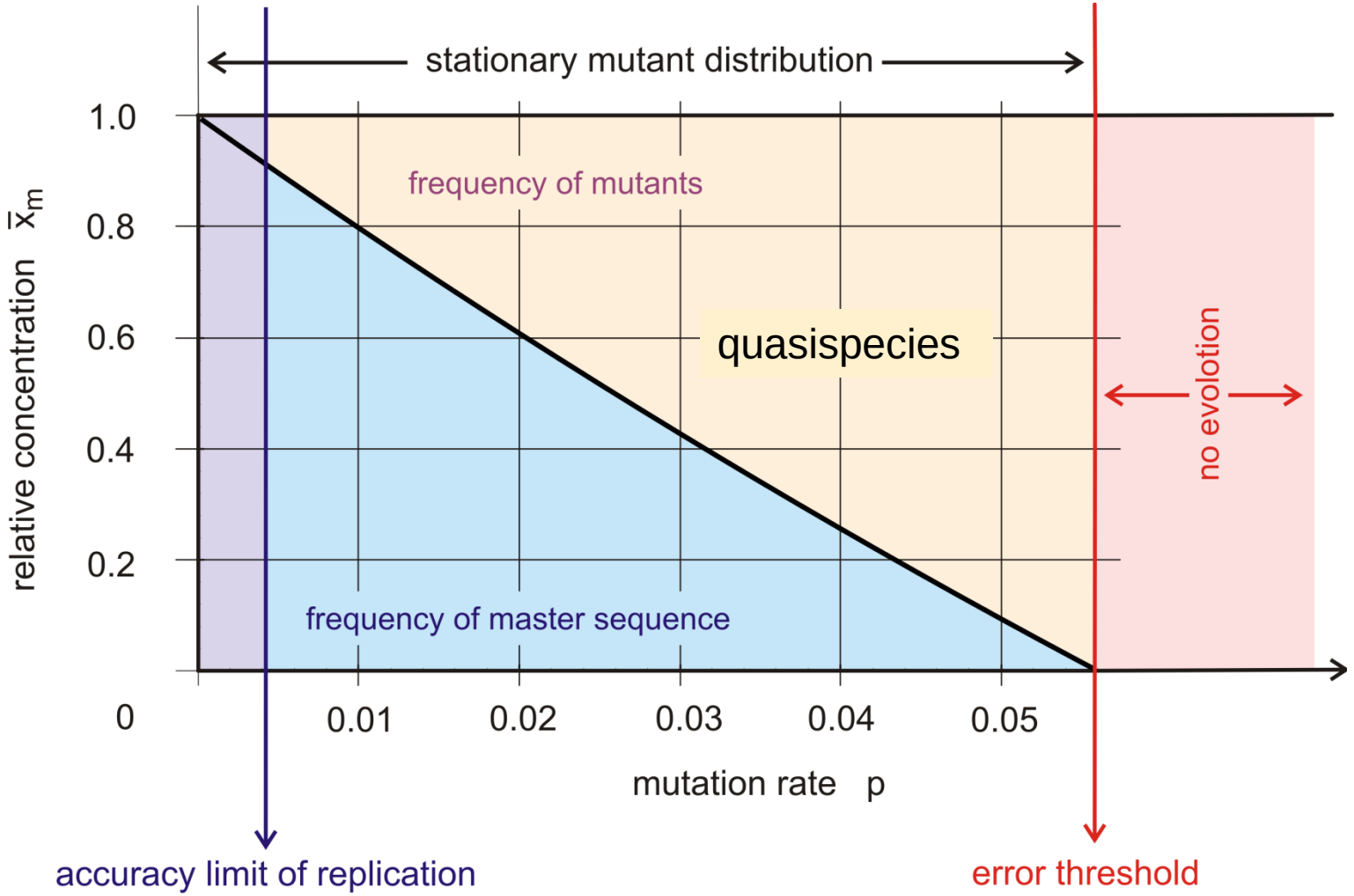
$$\frac{dx_i}{dt} = \sum_{j=1}^{n} Q_{ij} f_j x_j - x_i \Phi ; \; i = 1, 2, \ldots, n$$

$$\Phi = \sum_{j=1}^{n} f_j x_j ; \sum_{j=1}^{n} x_j = 1$$

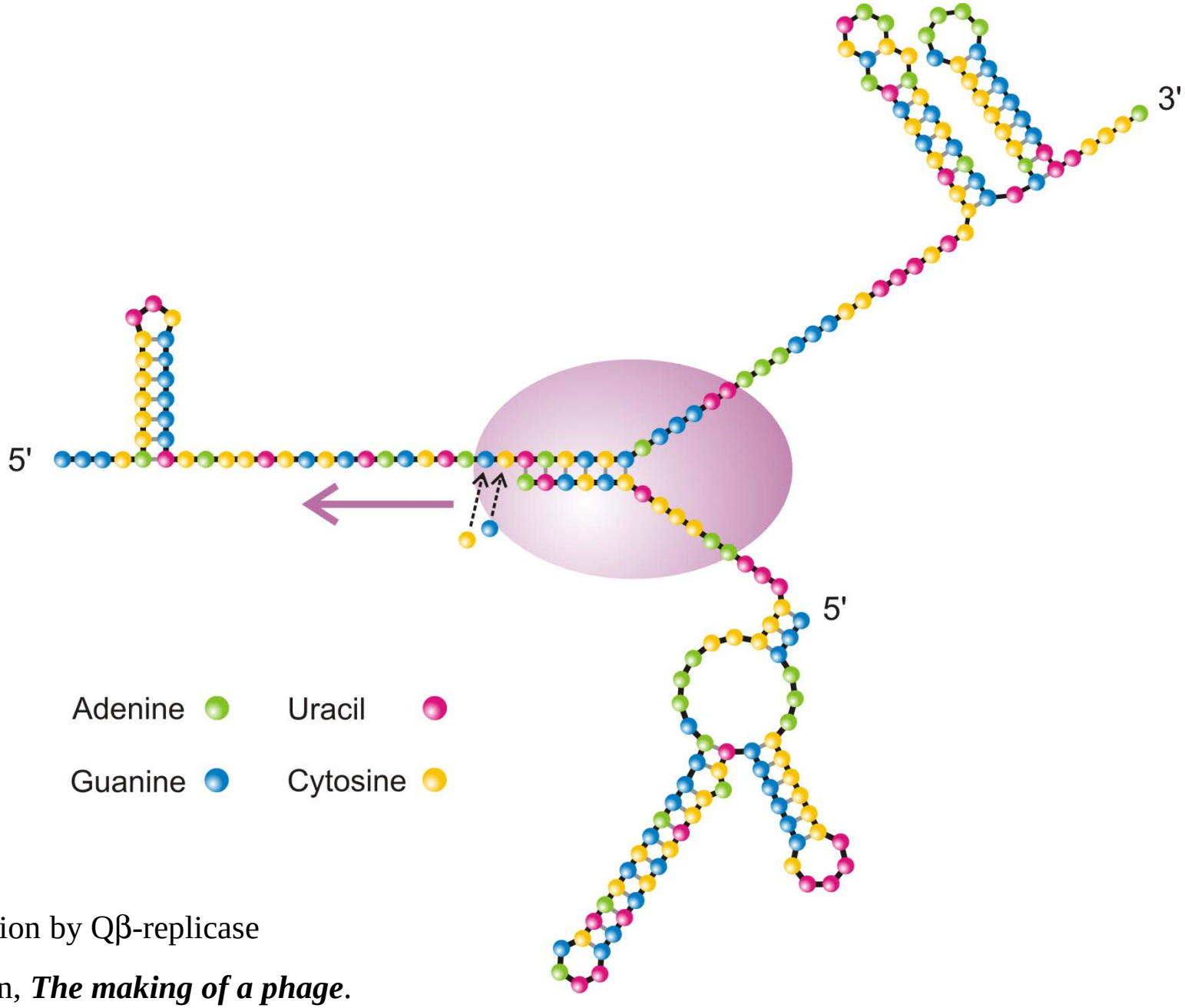


Mutation and (correct) replication as parallel chemical reactions

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

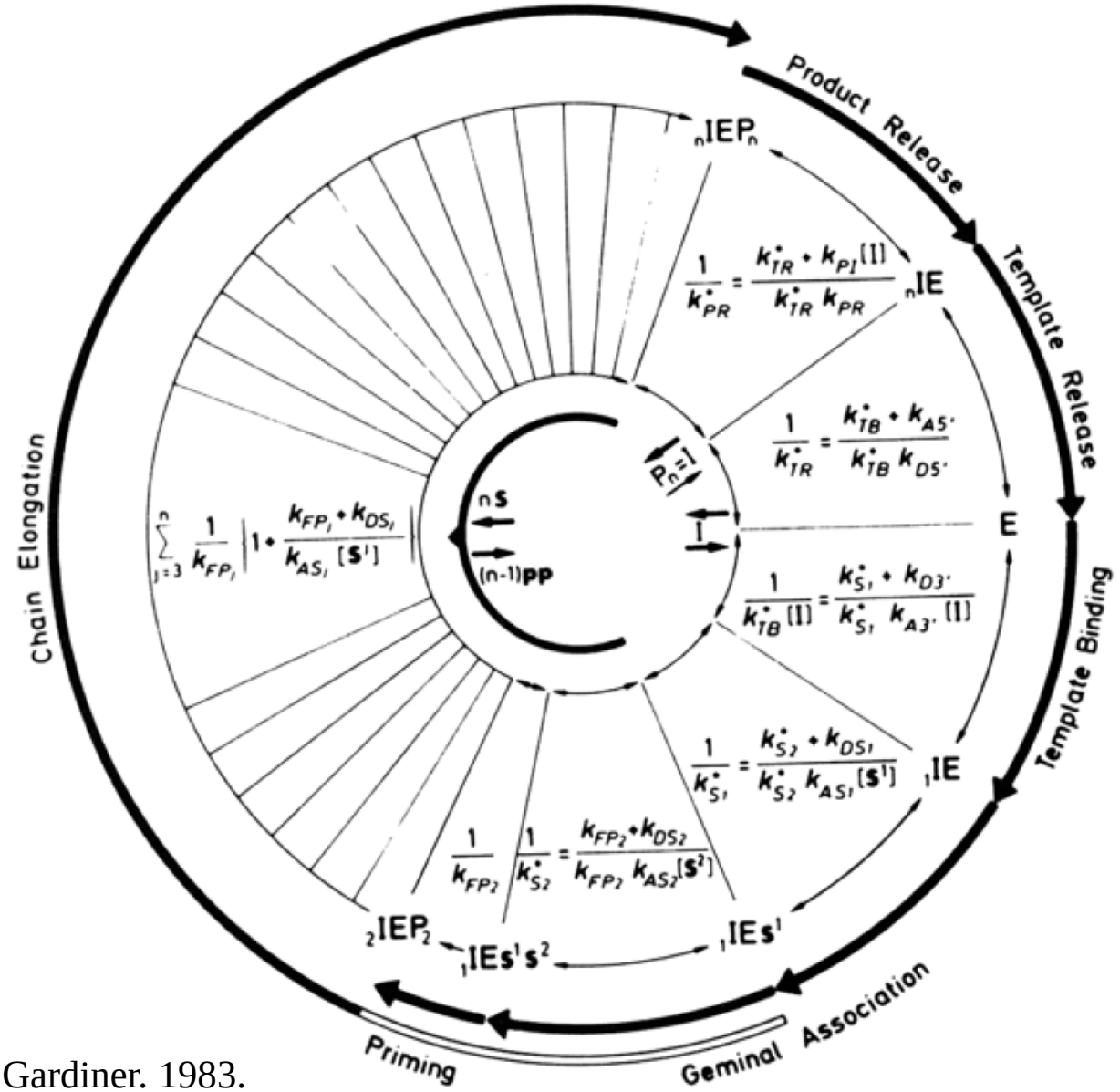


The error threshold in replication



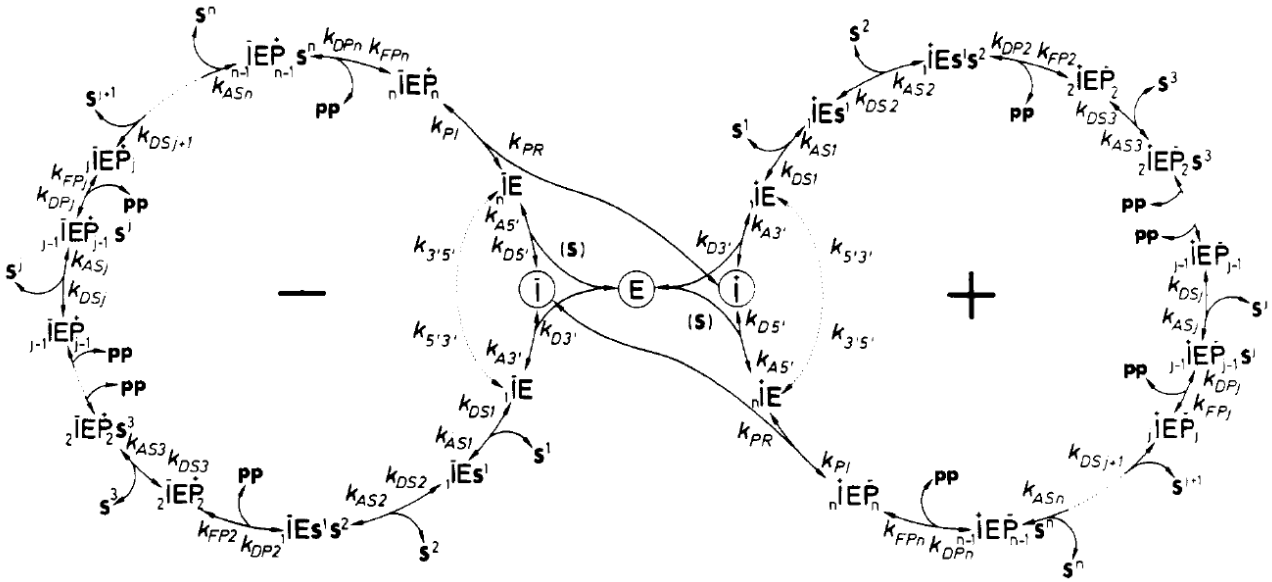
RNA replication by Qβ-replicase

C. Weissmann, ***The making of a phage***.
FEBS Letters **40** (1974), S10-S18



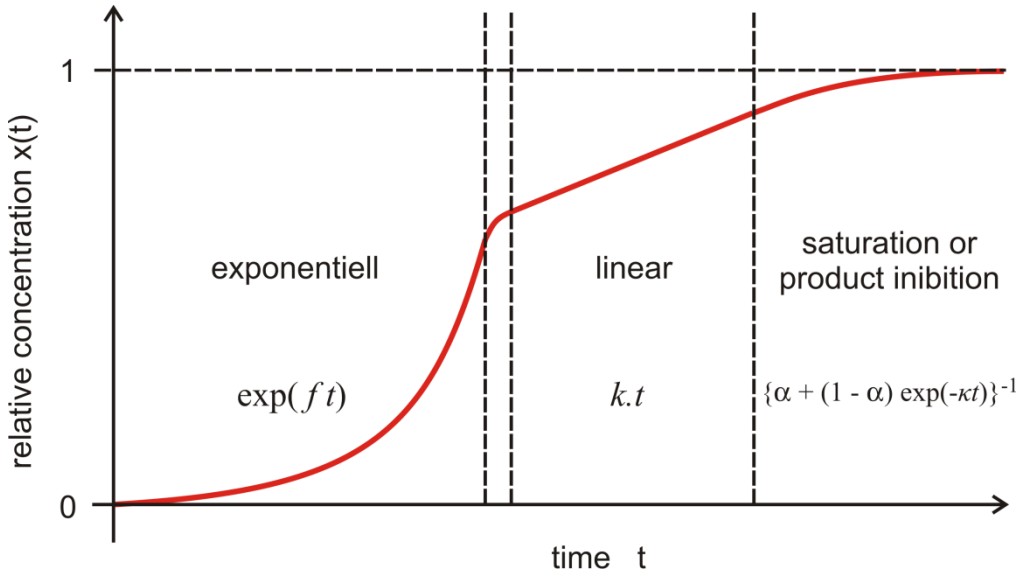
C.K. Biebricher, M.Eigen, W.C. Gardiner. 1983.
Kinetics of ribonucleic acid replication.
Biochemistry **22**:2544-2559.

Christof K. Biebricher,
1941-2009



Kinetics of RNA replication

C.K. Biebricher, M. Eigen, W.C. Gardiner, Jr.
Biochemistry **22**:2544-2559, 1983



$$
\begin{aligned}
X_+ + E &\underset{h_2^+}{\overset{h_1^+}{\rightleftharpoons}} EX_+ \,, \\
EX_+ + 2\,A &\xrightarrow{g_+} I_-EX_+ \,, \\
I_-EX_+ + (n-2)\,A &\xrightarrow{k_+} X_-EX_+ \,, \\
X_-EX_+ &\underset{d_2^+}{\overset{d_1^+}{\rightleftharpoons}} X_- + EX'_+ \,, \\
EX'_+ &\underset{b_2^+}{\overset{b_1^+}{\rightleftharpoons}} X_+ + E \,, \\
X_- + E &\underset{h_2^-}{\overset{h_1^-}{\rightleftharpoons}} EX_- \,, \\
EX_- + 2\,A &\xrightarrow{g_-} I_+EX_- \,, \\
I_+EX_- + (n-2)\,A &\xrightarrow{k_-} X_-EX_+ \,, \\
X_+EX_- &\underset{d_2^-}{\overset{d_1^-}{\rightleftharpoons}} X_+ + EX'_- \,, \\
EX'_- &\underset{b_2^-}{\overset{b_1^-}{\rightleftharpoons}} X_- + E \,.
\end{aligned}
$$

$$
\begin{aligned}
\frac{da}{dt} &= -\,2\,(g_+y_+ + g_-y_-)\,a^2 \,-\, (n-2)\,(k_+m_+ + k_-m_-)\,a^{n-2} \\
\frac{de}{dt} &= -\,(h_1^+x_+ + h_1^-x_- + b_2^+x_+ + b_2^-x_-)\,e + \\
&\quad + h_2^+y_+ + h_2^-y_- + b_1^+z_+ + b_1^-z_- \\
\frac{dx_+}{dt} &= -\,(h_1^+e + b_2^+e + d_2^-z_-)\,x_+ \,+\, h_2^+y_+ + b_1^+z_+ + d_1^-w_- \\
\frac{dy_+}{dt} &= -\,(h_2^+ + g_+a^2)\,y_+ \,+\, h_1^+x_+e \\
\frac{dm_+}{dt} &= -\,k_+a^{n-2}m_+ \,+\, g_+a^2y_+ \\
\frac{dw_+}{dt} &= -\,d_1^+w_+ \,+\, d_2^+x_-z_+ \,+\, k_+a^{n-2}m_+ \\
\frac{dz_+}{dt} &= -\,(b_1^+ + d_2^+x_-)\,z_+ \,+\, d_1^+w_+ \,+\, b_2^+x_+e \\
\frac{dx_-}{dt} &= -\,(h_1^-e + b_2^-e + d_2^+z_+)\,x_- \,+\, h_2^-y_- + b_1^-z_- + d_1^+w_+ \\
\frac{dy_-}{dt} &= -\,(h_2^- + g_-a^2)\,y_- \,+\, h_1^-x_-e \\
\frac{dm_-}{dt} &= -\,k_-a^{n-2}m_- \,+\, g_-a^2y_- \\
\frac{dw_-}{dt} &= -\,d_1^-w_- \,+\, d_2^-x_+z_- \,+\, k_-a^{n-2}m_- \\
\frac{dz_-}{dt} &= -\,(b_1^- + d_2^-x_+)\,z_- \,+\, d_1^-w_- \,+\, b_2^-x_-e \,.
\end{aligned}
$$

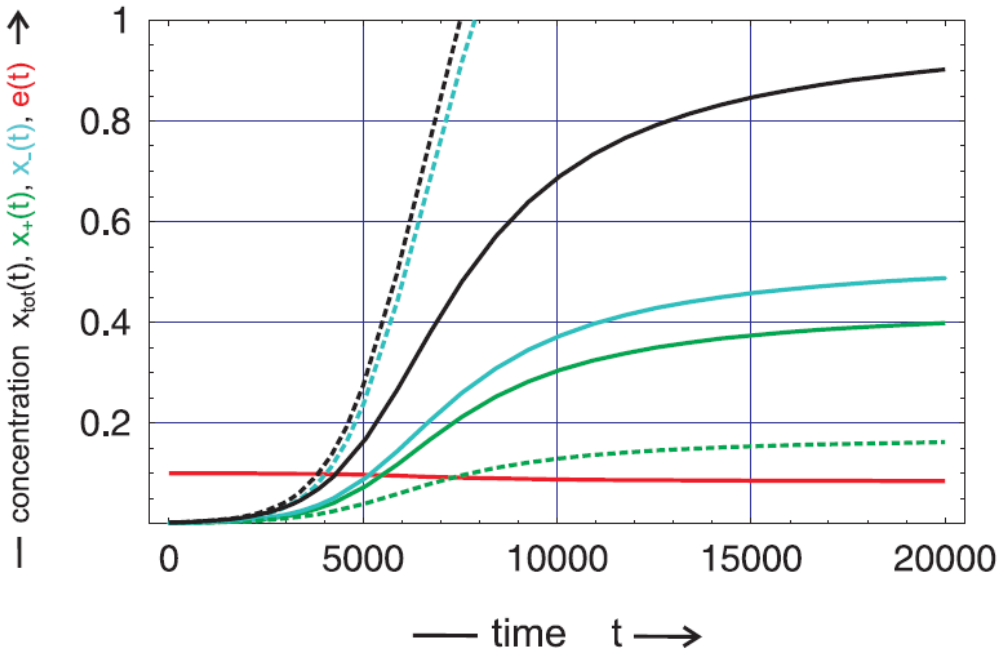
Paul E. Phillipson, Peter Schuster. 2009.
Modeling by nonlinear differential equations.
Dissipative and conservative processes.
World Scientific Publishing, Hackensack, NJ.

replicase e(t)

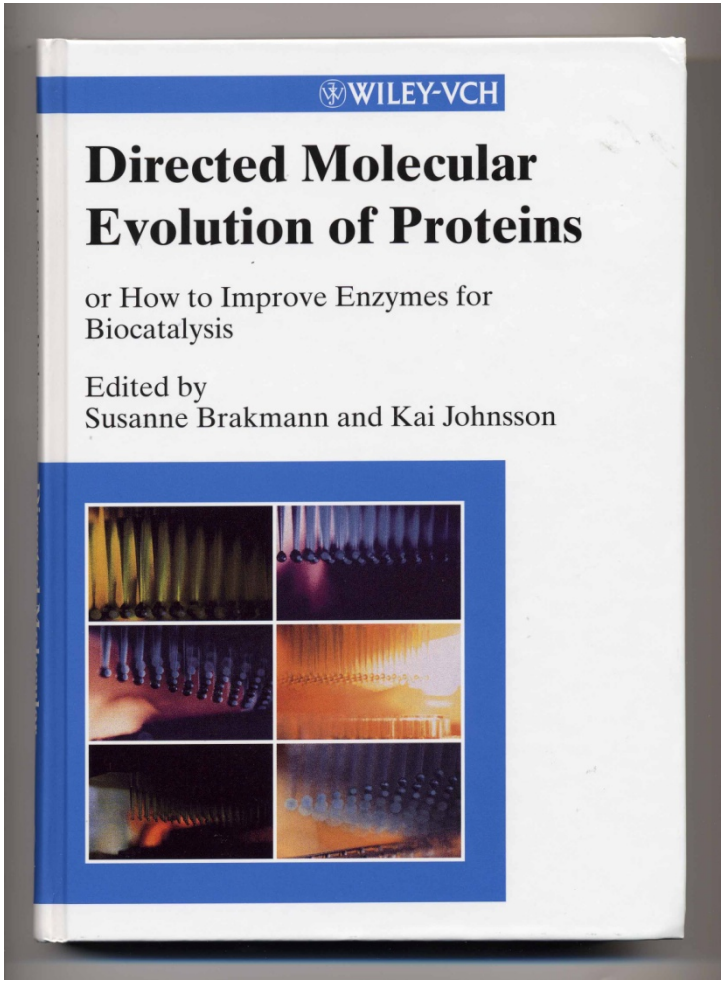
plus strand $x_+(t)$
minus strand $x_-(t)$

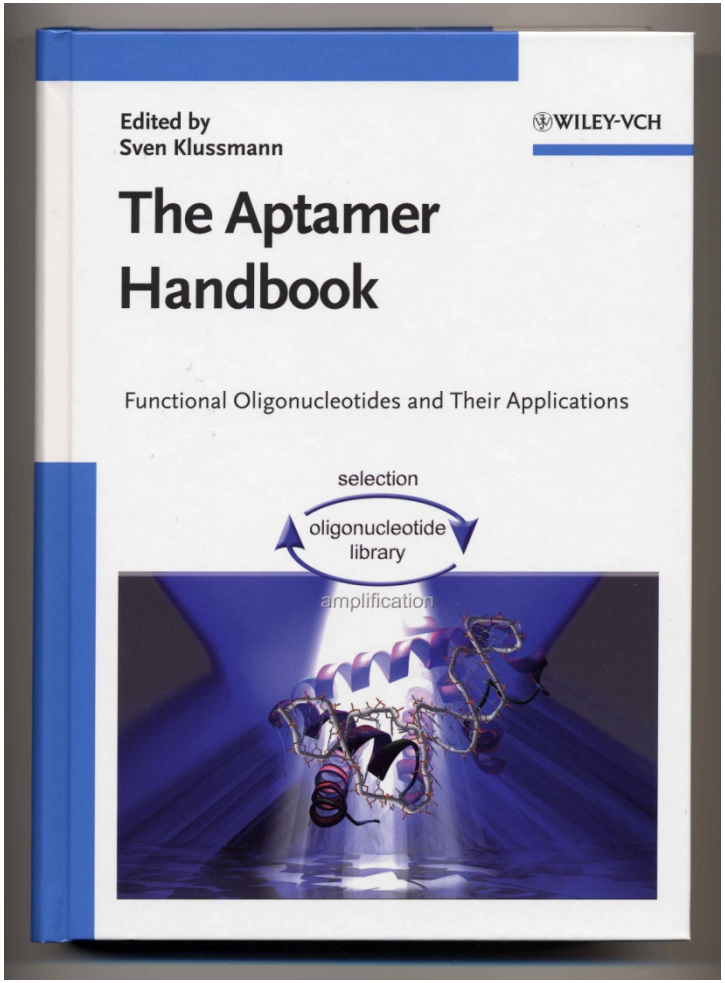
total RNA concentration
$x_{tot}(t) = x_+(t) + x_-(t)$

complemetary replication



Paul E. Phillipson, Peter Schuster. 2009.
Modeling by nonlinear differential equations.
Dissipative and conservative processes.
World Scientific Publishing, Hackensack, NJ.

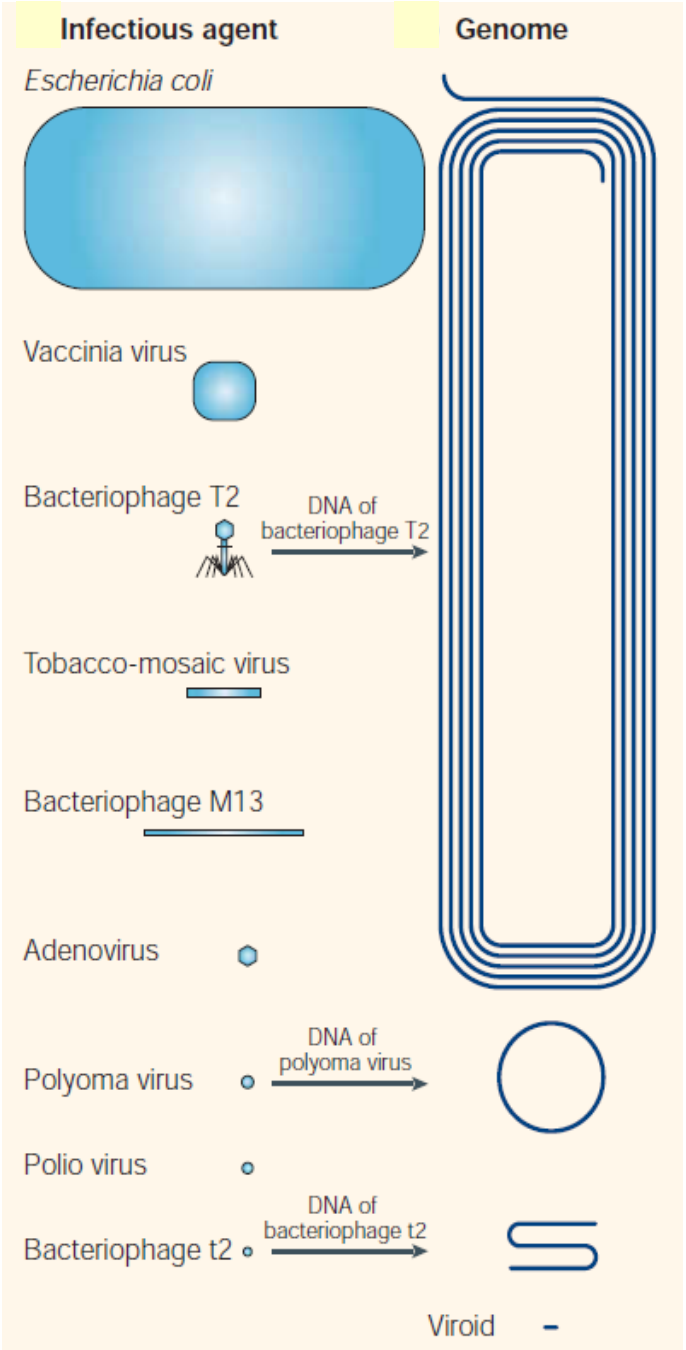




Application of molecular evolution to problems in biotechnology

Viroids

Viroids: circular RNAs 246 - 401 nt long
infect inclusively plants

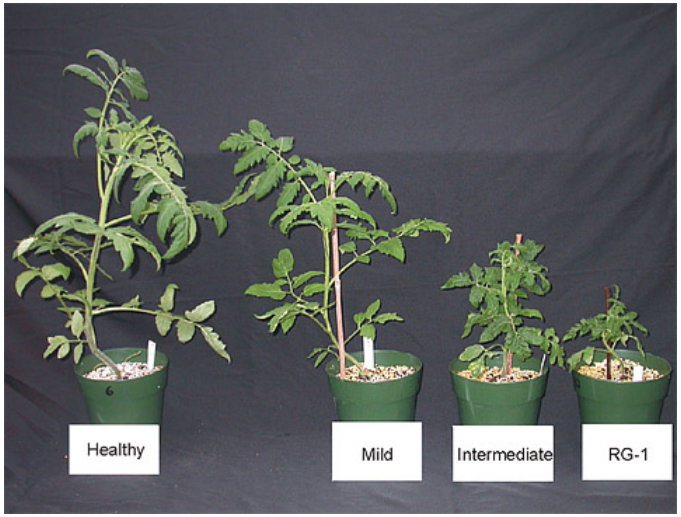


Theodor O. Diener. 2003. Discovering viroids –
A personal perspective.
Nat.Rev.Microbiology 1:75-80.

José-Antonio Daròs, Santiago F. Elena, Ricardo Flores.
2006. Viroids: An Ariadne‘s thread thorugh the
RNA labyrinth. EMBO Reports 7:593-598.

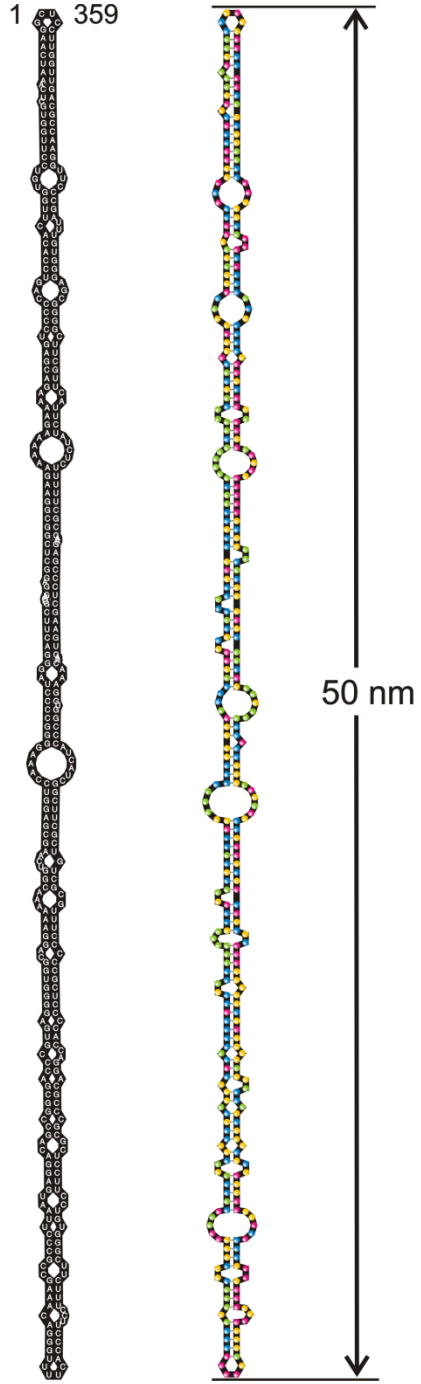
Ricardo Flores *et al*. 2009. Viroid replication: Rolling
circles, enzymes and ribozymes.
Viruses 2009:317-334.

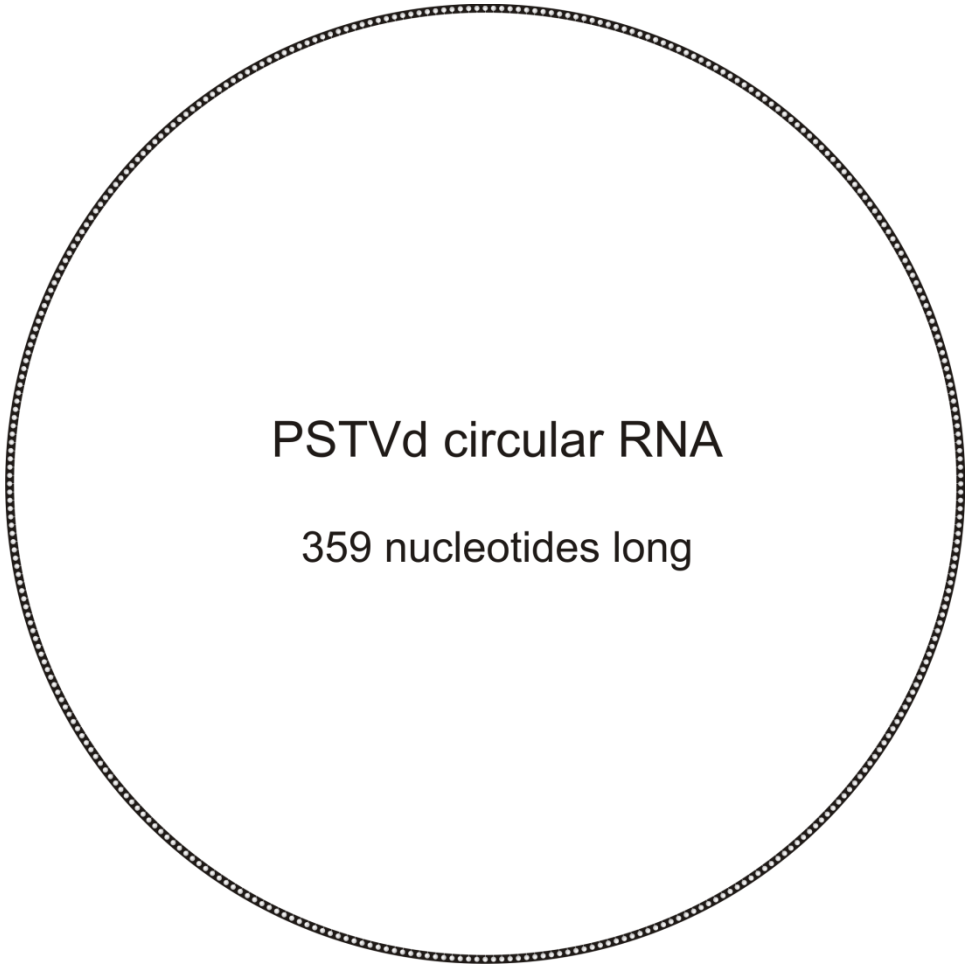
J. Demez. European and mediterranean plant protection organization archive. France



R.W. Hammond, R.A. Owens. Molecular Plant Pathology Laboratory, US Department of Agriculture

Plant damage by viroids





Adenine

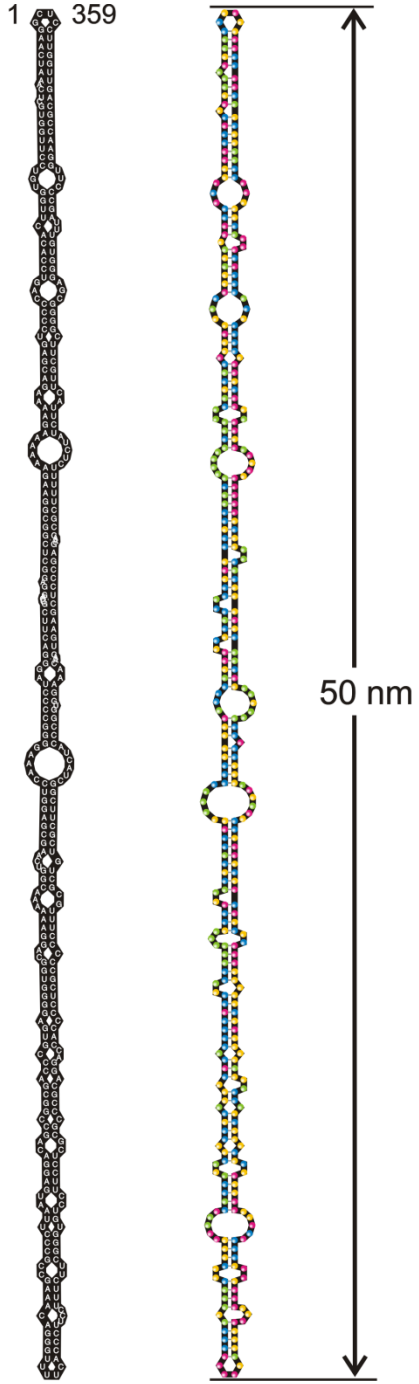
Uracil

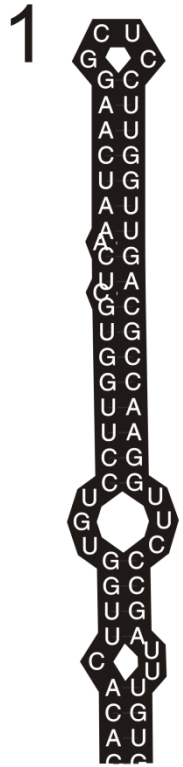
Guanine

Cytosine

Nucleotide sequence and secondary structure of the potato spindle tuber viroid RNA

H.J.Gross, H. Domdey, C. Lossow, P Jank, M. Raba, H. Alberty, and H.L. Sänger. *Nature* **273**:203-208 (1978)





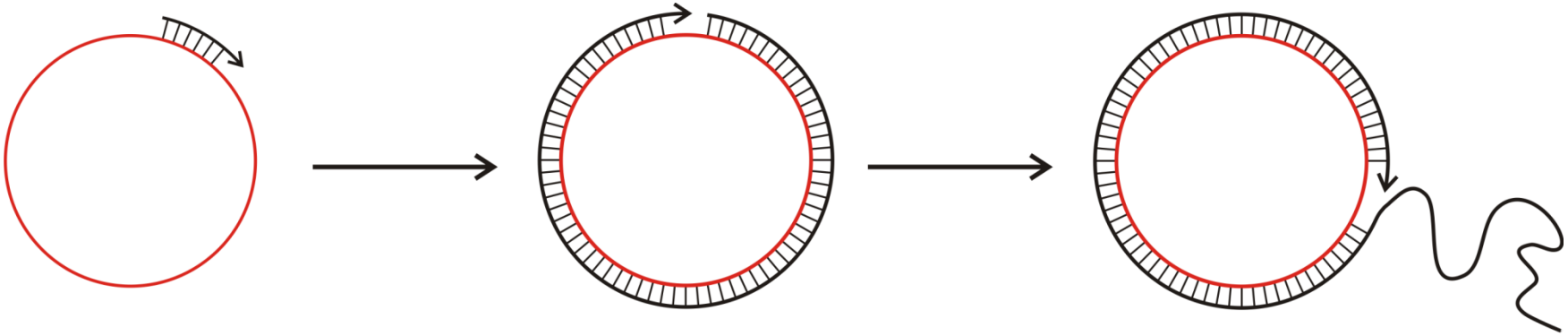
Vienna RNA Package 1.8.2

Biochemically supported structure

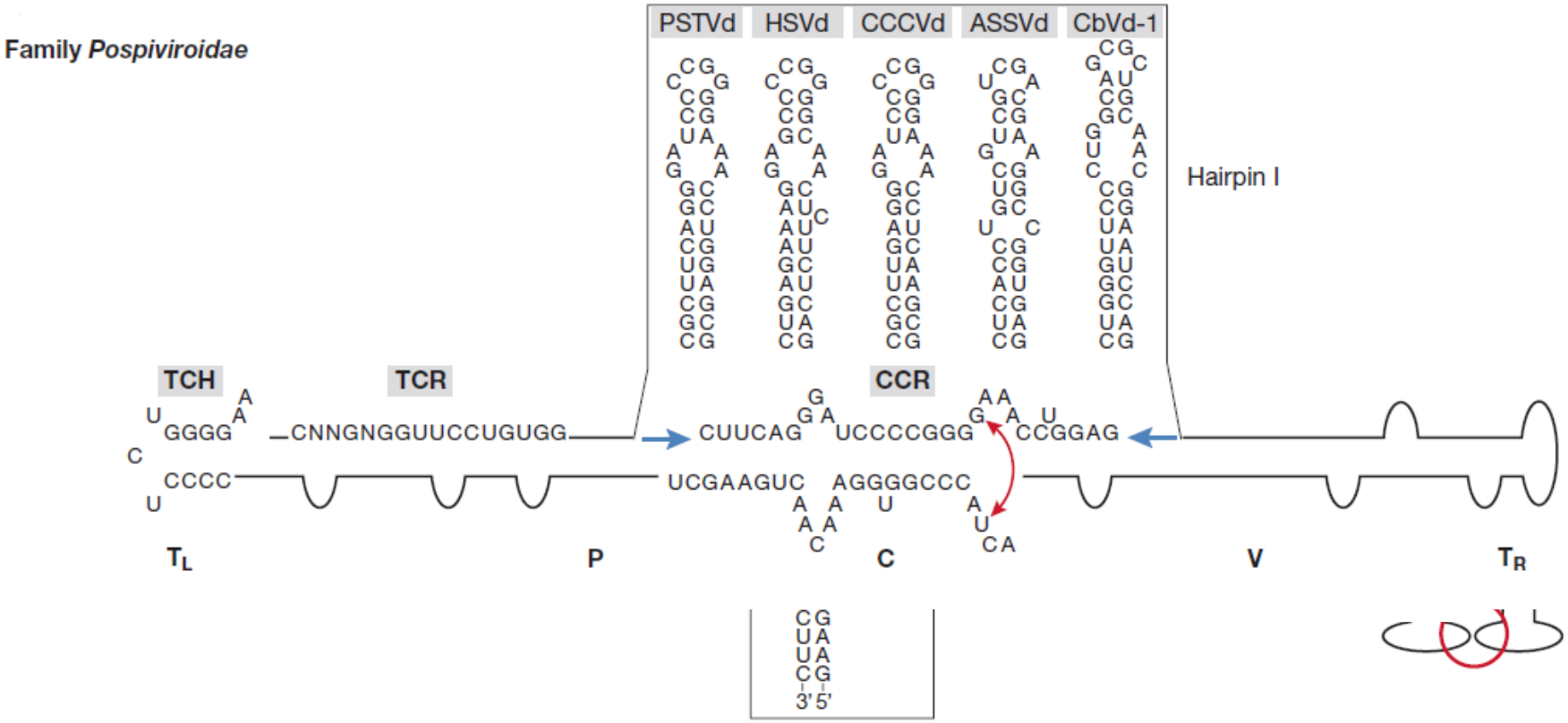
- Adenine
- Uracil
- Guanine
- Cytosine

Nucleotide sequence and secondary structure of the potato spindle tuber viroid RNA

H.J.Gross, H. Domdey, C. Lossow, P Jank, M. Raba, H. Alberty, and H.L. Sänger. *Nature* **273**:203-208 (1978)

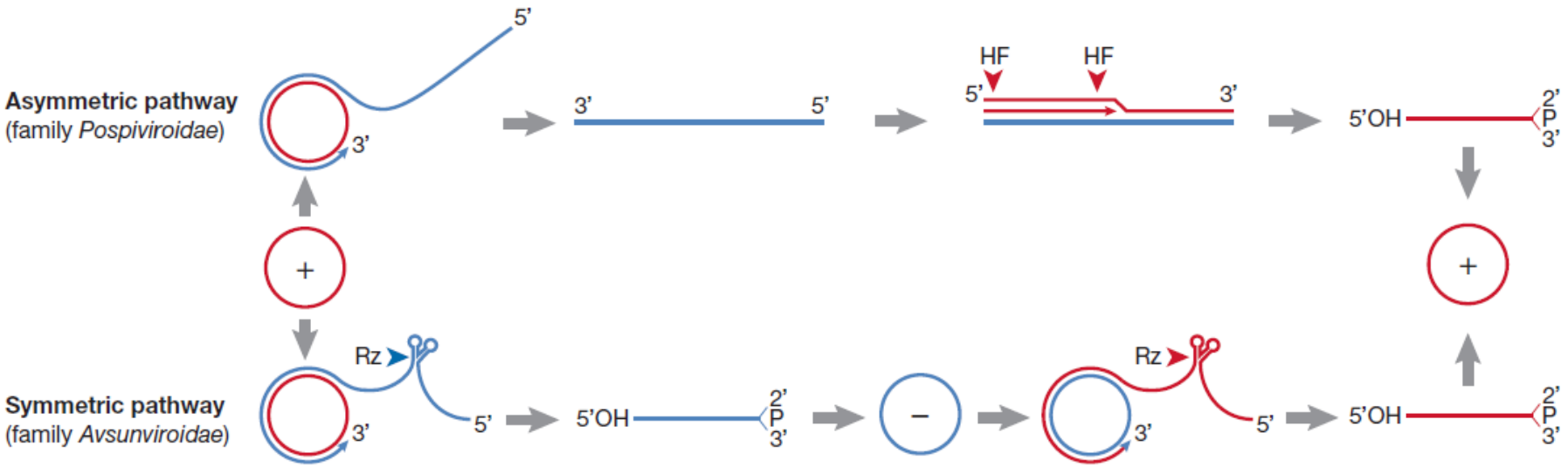


The principle of viroid replication: Rolling circle



The two major classes of viroids .

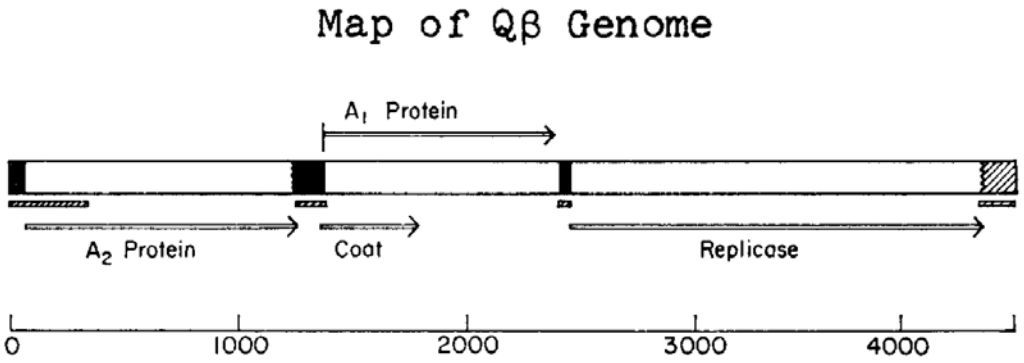
José-Antonio Daròs, Santiago F. Elena, Ricardo Flores. 2006. Viroids: An Adriadne‘s thread into the RNA labyrinth. *EMBO Reports* **7**:593-598.

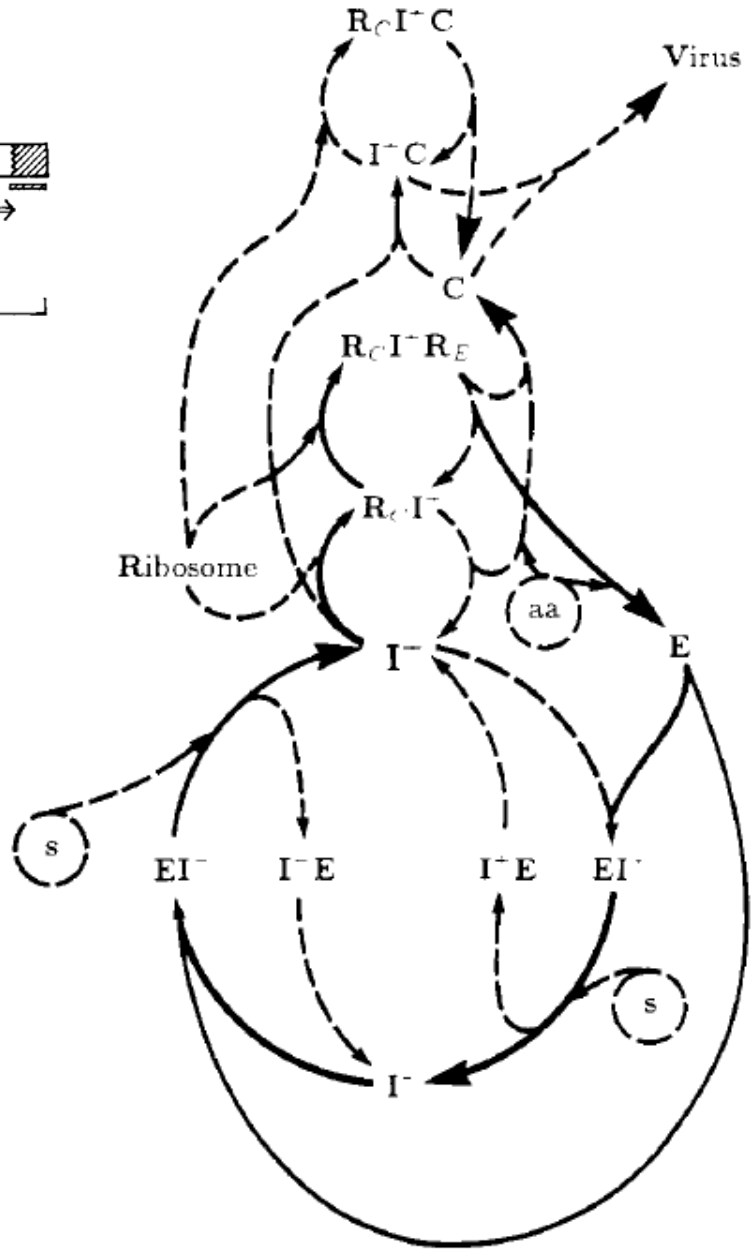


Replication in the two major classes of viroids .

José-Antonio Daròs, Santiago F. Elena, Ricardo Flores. 2006.
Viroids: An Adriadne‘s thread into the RNA labyrinth. *EMBO Reports* **7**:593-598.

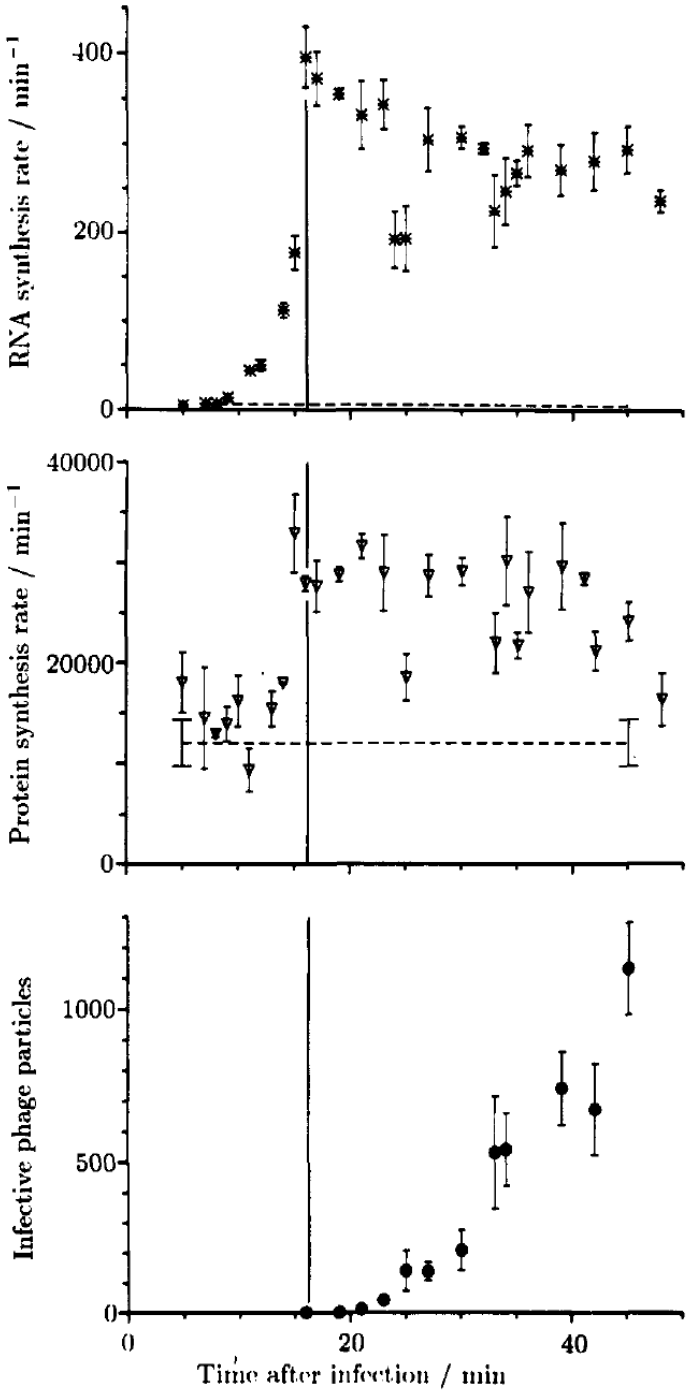
Viruses





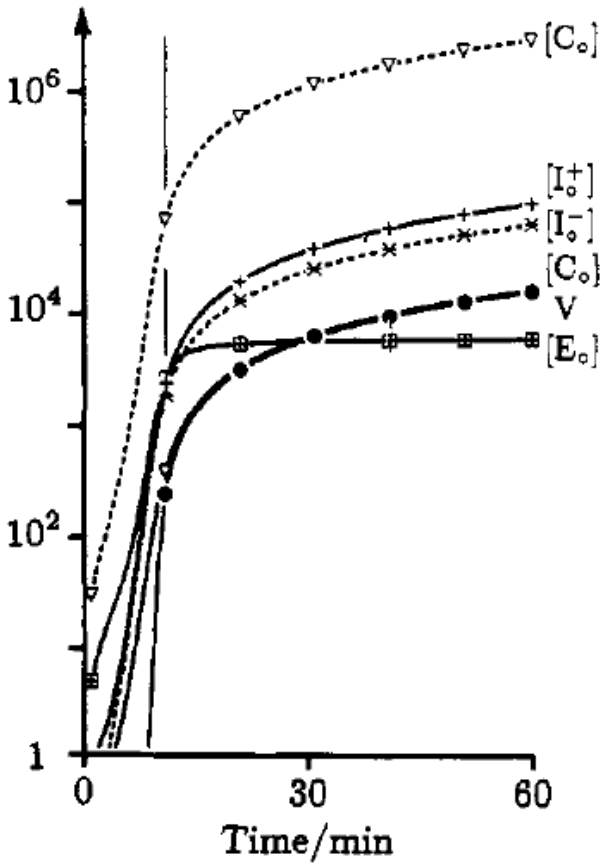
M. Eigen, C.K. Biebricher, M. Gebinoga, W.C. Gardiner. 1991. The hypercycle. Coupling of RNA and protein biosynthesis in the infection of an RNA bacteriophage. *Biochemistry* **30**:11005-11018.

Qβ phage infection of *Escherichia coli* cells.

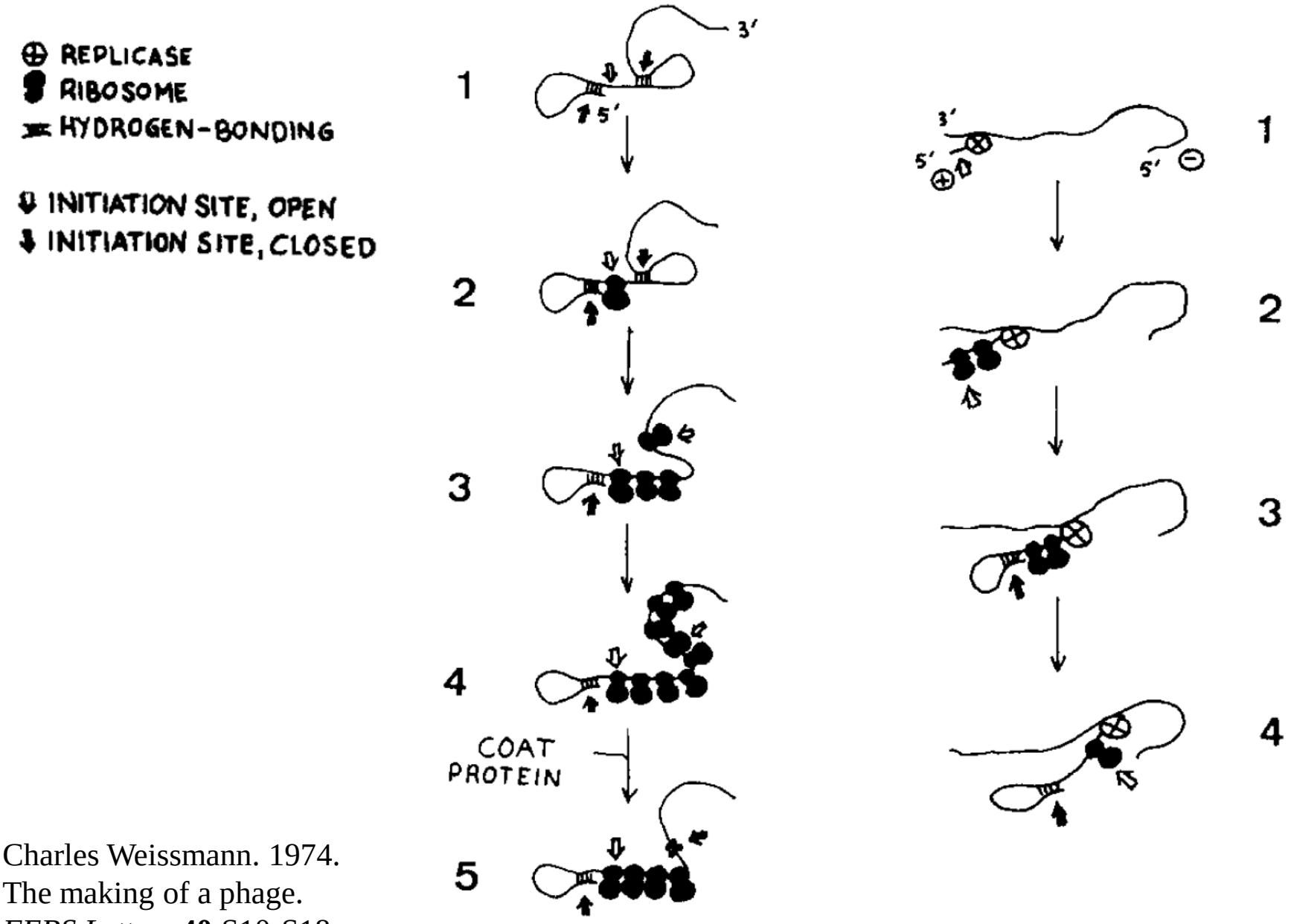


M. Eigen, C.K. Biebricher, M. Gebinoga, W.C. Gardiner. 1991. The hypercycle. Coupling of RNA and protein biosynthesis in the infection of an RNA bacteriophage. *Biochemistry* **30**:11005-11018.

Experimental rate profiles.



Computer simulation of the infection process



Charles Weissmann. 1974.
The making of a phage.
FEBS Letters **40**:S10-S18.

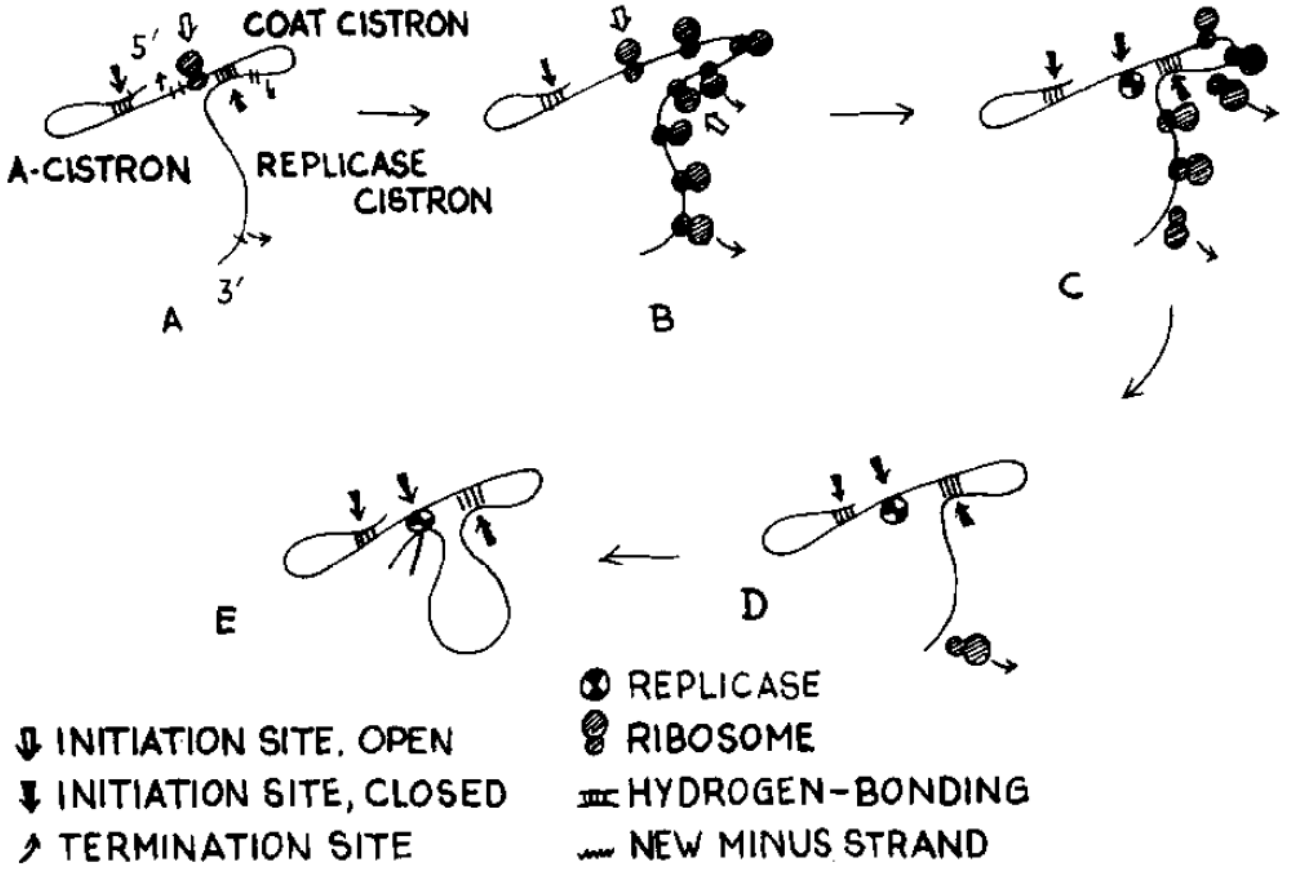
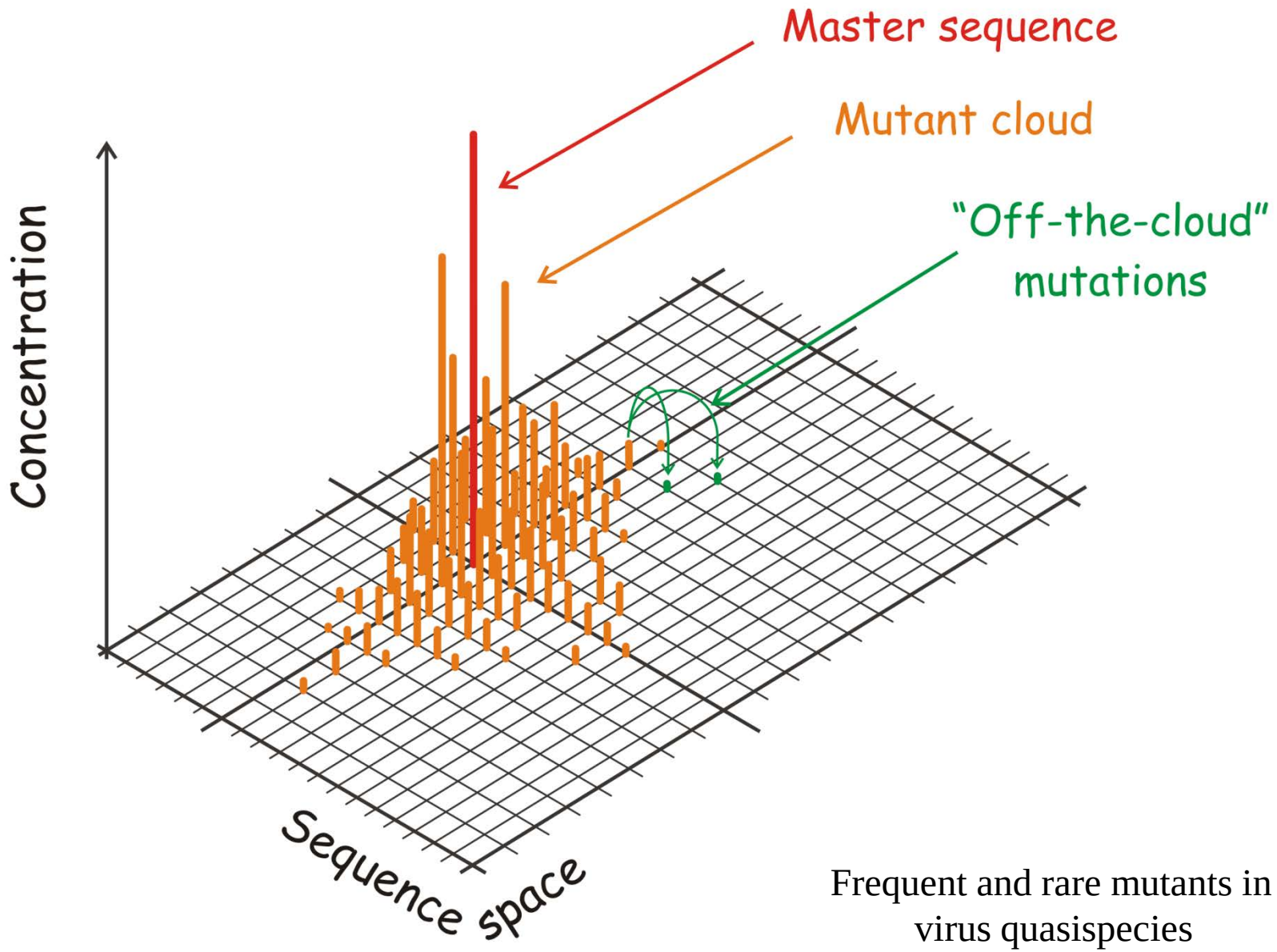
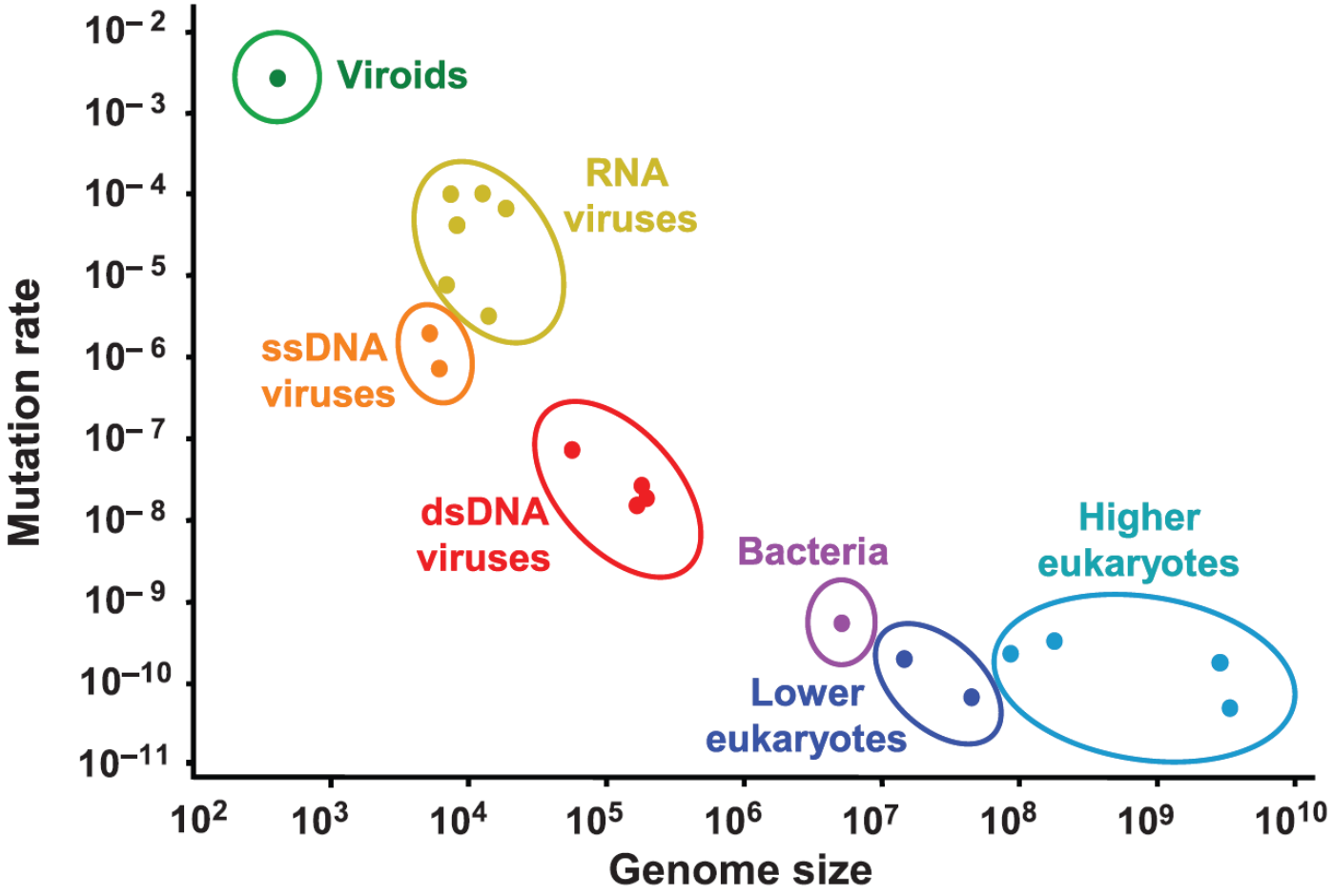


Fig. 3. Transition of phage RNA from polysome to replicating complex – repressor function of Qβ viral replicase.
(A) Ribosomes attach to the RNA at the coat initiation site. The initiation site of the replicase cistron is unavailable because of the secondary structure of the RNA. (B) Translation of the coat cistron ensues and the initiation site of the replicase cistron is exposed. The replicase cistron is translated. (C) When replicase becomes available, it attaches to the initiation site of the coat protein and blocks attachment of ribosomes in this position. The RNA refolds, preventing initiation at the replicase cistron. (D) The RNA is cleared of ribosomes. (E) Replicase can now attach to the 3′ terminus and initiate synthesis of the minus strand. The A cistron initiation site is at all times unaccessible to ribosomes because of the secondary structure of the mature RNA (cf. fig. 2) (from ref. [64]).



Frequent and rare mutants in virus quasispecies



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

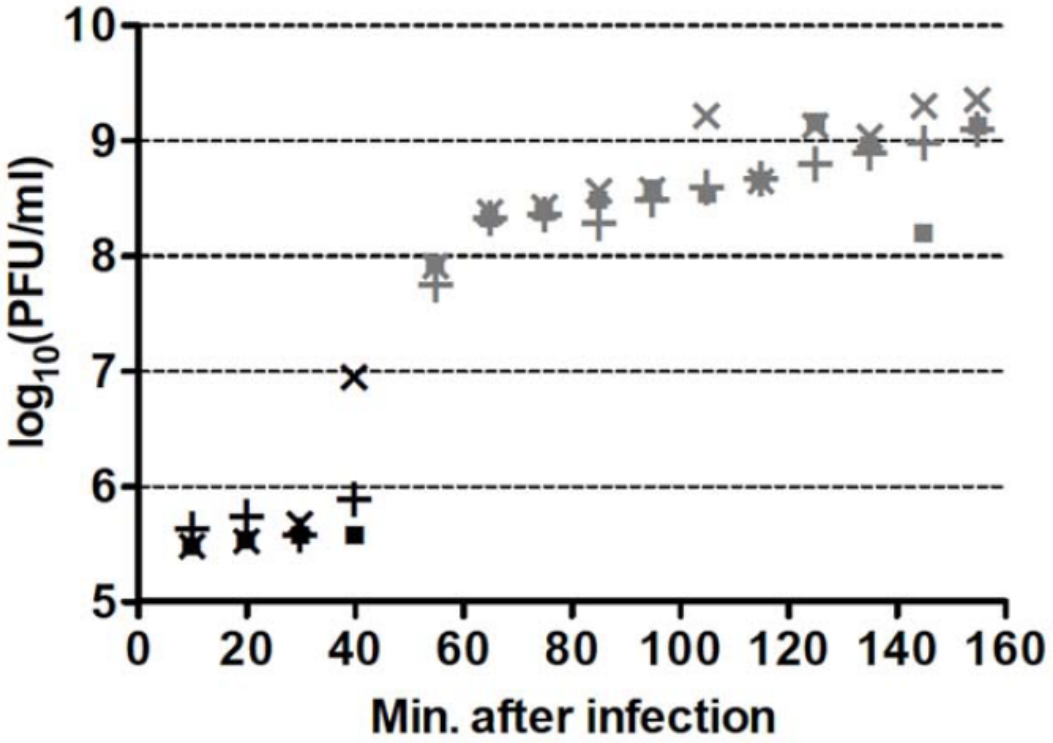
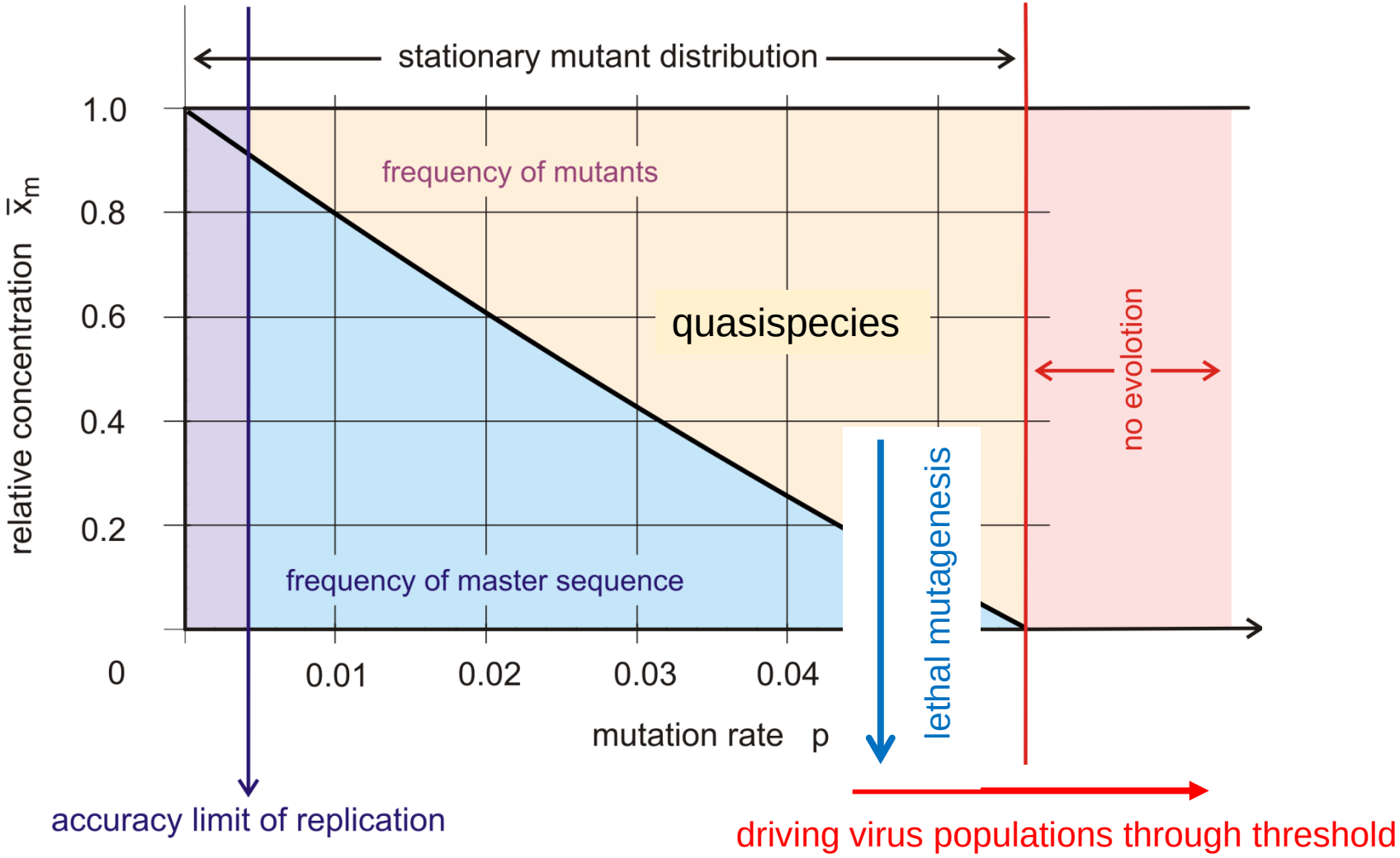


Figure 2. Representative one-step growth curves of wt Qß in RTH cells. Individual cultures are represented by different symbols while black and gray colors represent two different culture dilutions. PFU = plaque-forming units. Similar curves were obtained for each of the four wt Qß isolates used to generate one-step growth lysates.
doi:10.1371/journal.pgen.1002832.g002

L. Garcia-Villada, J.W. Drake. 2012. The three faces of riboviral spontaneous mutation: Spectrum, mode of genome replication, and mutation rate. *PLoS Genetics* **8**:e1002832.



The error threshold in replication

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Virus Research 107 (2005) 115–116
Virus Research
www.elsevier.com/locate/virusres

Preface

Antiviral strategy on the horizon

Error catastrophe had its conceptual origins in the middle of the XXth century, when the consequences of mutations on enzymes involved in protein synthesis, as a theory of aging. In those times biological processes were generally perceived differently from today. Infectious diseases were regarded as a fleeting nuisance which would be eliminated through the use of antibiotics and antiviral agents. Microbial variation, although known in some cases, was not thought to be a significant problem for disease control. Variation in differentiated organisms was seen as resulting essentially from exchanges of genetic material associated with sexual reproduction. The problem was to unveil the mechanisms of inheritance, expression of genetic information and metabolism. Few saw that genetic change is occurring at present in all organisms, and still fewer recognized Darwinian principles as essential to the biology of pathogenic viruses and cells. Population geneticists rarely used bacteria or viruses as experimental systems to define concepts in biological evolution. The extent of genetic polymorphism among individuals of the same biological species came as a surprise when the first results on comparison of electrophoretic mobility of enzymes were obtained. With the advent of in vitro DNA recombination, and rapid nucleic acid sequencing techniques, molecular analyses of genomes reinforced the conclusion of extreme inter-individual genetic variation within the same species. Now, due largely to spectacular progress in comparative genomics, we see cellular DNAs, both prokaryotic and eukaryotic, as highly dynamic. Most cellular processes, including such essential information-bearing and transferring events as genome replication, transcription and translation, are increasingly perceived as inherently inaccurate. Viruses, and in particular RNA viruses, are among the most extreme examples of exploitation of replication inaccuracy for survival.

Error catastrophe, or the loss of meaningful genetic information through excess genetic variation, was formulated in quantitative terms as a consequence of quasispecies theory, which was first developed to explain self-organization and adaptability of primitive replicons in early stages of life. Recently, a conceptual extension of error catastrophe that could be defined as "induced genetic deterioration" has emerged as a possible antiviral strategy. This is the topic of the current special issue of ***Virus Research***.

Few would nowadays doubt that one of the major obstacles for the control of viral disease is short-term adaptability of viral pathogens. Adaptability of viruses follows the same Darwinian principles that have shaped biological evolution over eons, that is, repeated rounds of reproduction with genetic variation, competition and selection, often perturbed by random events such as statistical fluctuations in population size. However, with viruses the consequences of the operation of these very same Darwinian principles are felt within very short times. Short-term evolution (within hours and days) can be also observed with some cellular pathogens, with subsets of normal cells, and cancer cells. The nature of RNA viral pathogens begs for alternative antiviral strategies, and forcing the virus to cross the critical error threshold for maintenance of genetic information is one of them.

The contributions to this volume have been chosen to reflect different lines of evidence (both theoretical and experimental) on which antiviral designs based on genetic deterioration inflicted upon viruses are being constructed. Theoretical studies have explored the copying fidelity conditions that must be fulfilled by any information-bearing replication system for the essential genetic information to be transmitted to progeny. Closely related to the theoretical developments have been numerous experimental studies on quasispecies dynamics and their multiple biological manifestations. The latter can be summarized by saying that RNA viruses, by virtue of existing as mutant spectra rather than defined genetic entities, remarkably expand their potential to overcome selective pressures intended to limit their replication. Indeed, the use of antiviral inhibitors in clinical practice and the design of vaccines for a number of major RNA virus-associated diseases, are currently presided by a sense of uncertainty. Another line of growing research is the enzymology of copying fidelity by viral replicases, aimed at understanding the molecular basis of mutagenic activities. Error catastrophe as a potential new antiviral strategy received an important impulse by the observation that ribavirin (a licensed antiviral nucleoside analogue) may be exerting, in some systems, its antiviral activity through enhanced mutagenesis. This has encouraged investigations on new mutagenic base analogues, some of them used in anticancer chemotherapy. Some chapters summarize these important biochemical studies on cell entry pathways and metabolism of mutagenic agents, that may find new applications as antiviral agents.

This volume intends to be basically a progress report, an introduction to a new avenue of research, and a realistic appraisal of the many issues that remain to be investigated. In this respect, I can envisage (not without many uncertainties) at least three lines of needed research: (i) One on further understanding of quasispecies dynamics in infected individuals to learn more on how to apply combinations of virus-specific mutagens and inhibitors in an effective way, finding synergistic combinations and avoiding antagonistic ones as well as severe clinical side effects. (ii) Another on a deeper understanding of the metabolism of mutagenic agents, in particular base and nucleoside analogues. This includes identification of the transporters that carry them into cells, an understanding of their metabolic processing, intracellular stability and alterations of nucleotide pools, among other issues. (iii) Still another line of needed research is the development of new mutagenic agents specific for viruses, showing no (or limited) toxicity for cells. Some advances may come from links with anticancer research, but others should result from the designs of new molecules, based on the structures of viral polymerases. I really hope that the reader finds this issue not only to be an interesting and useful review of the current situation in the field, but also a stimulating exposure to the major problems to be faced.

The idea to prepare this special issue came as a kind invitation of Ulrich Desselberger, former Editor of ***Virus Research***, and then taken enthusiastically by Luis Enjuanes, recently appointed as Editor of ***Virus Research***. I take this opportunity to thank Ulrich, Luis and the Editor-in-Chief of ***Virus Research***, Brian Mahy, for their continued interest and support to the research on virus evolution over the years.

My thanks go also to the 19 authors who despite their busy schedules have taken time to prepare excellent manuscripts, to Elsevier staff for their prompt responses to my requests, and, last but not least, to Ms. Lucía Horrillo from Centro de Biología Molecular "Severo Ochoa" for her patient dealing with the correspondence with authors and the final organization of the issue.

Esteban Domingo
Universidad Autónoma de Madrid
Centro de Biología Molecular "Severo Ochoa"
Consejo Superior de Investigaciones Científicas
Cantoblanco and Valdeolmos
Madrid, Spain
Tel.: +34 91 497 84858/9; fax: +34 91 497 4799
E-mail address: edomingo@cbm.uam.es

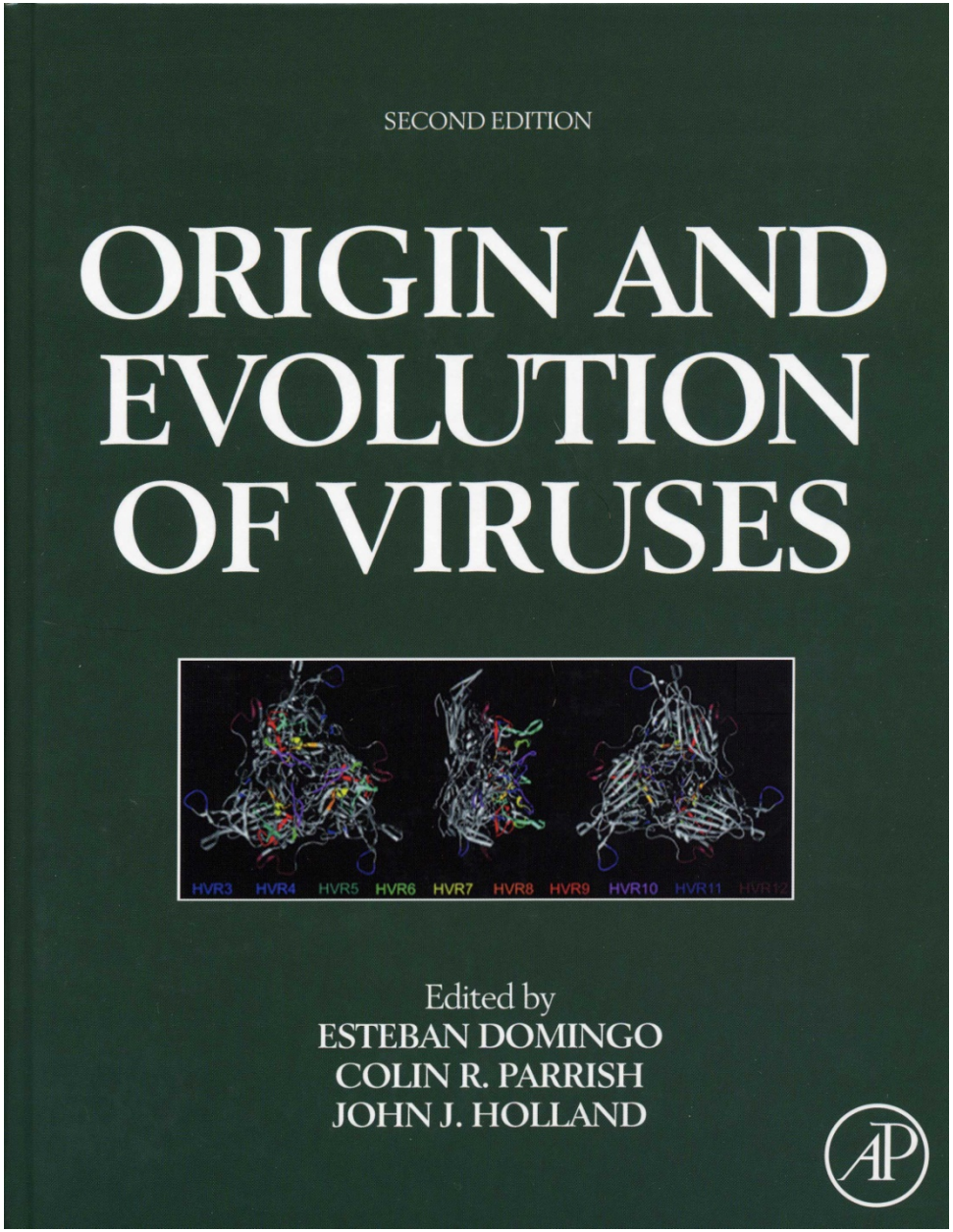
Available online 8 December 2004

doi:10.1016/j.virusres.2004.11.001

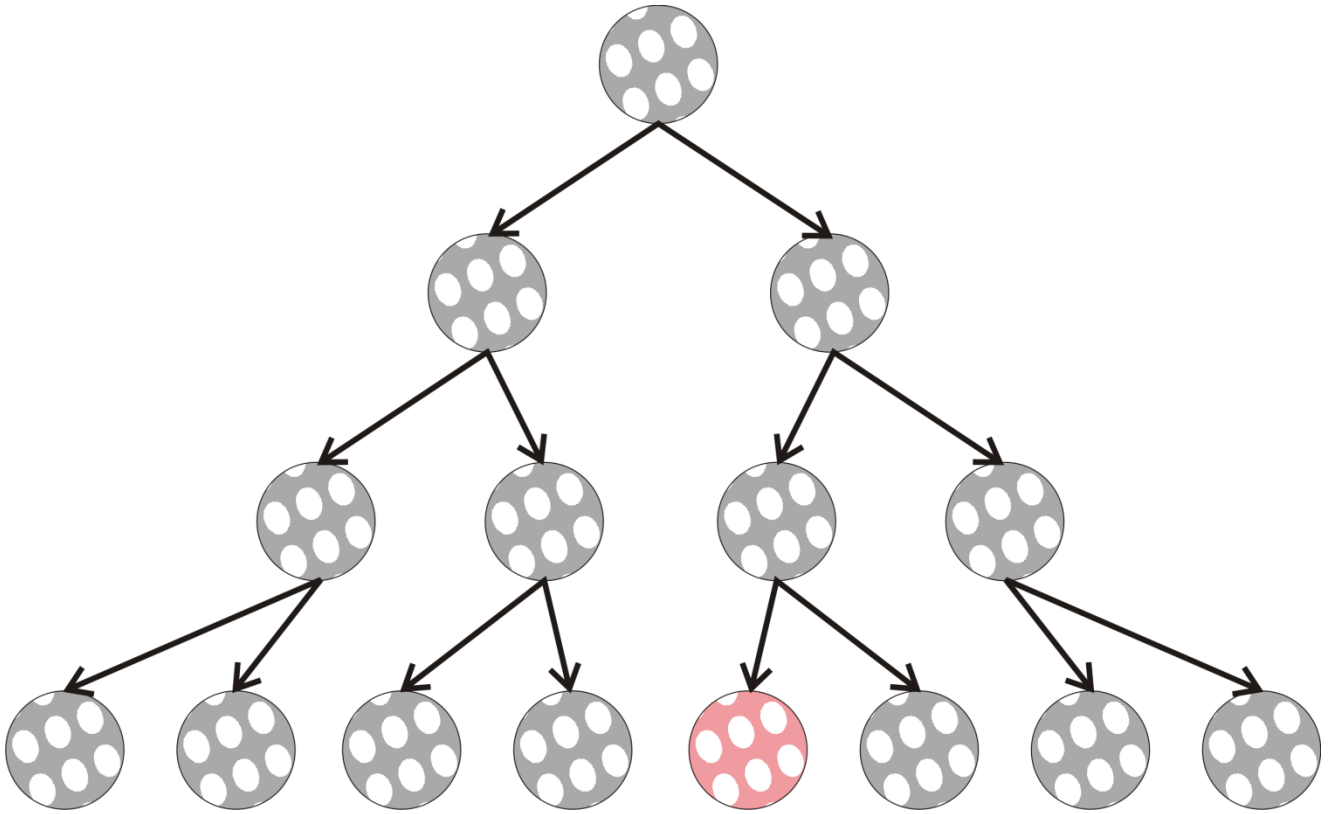
Esteban Domingo
1943 -

Application of quasispecies theory to the fight against viruses

Molecular evolution of viruses



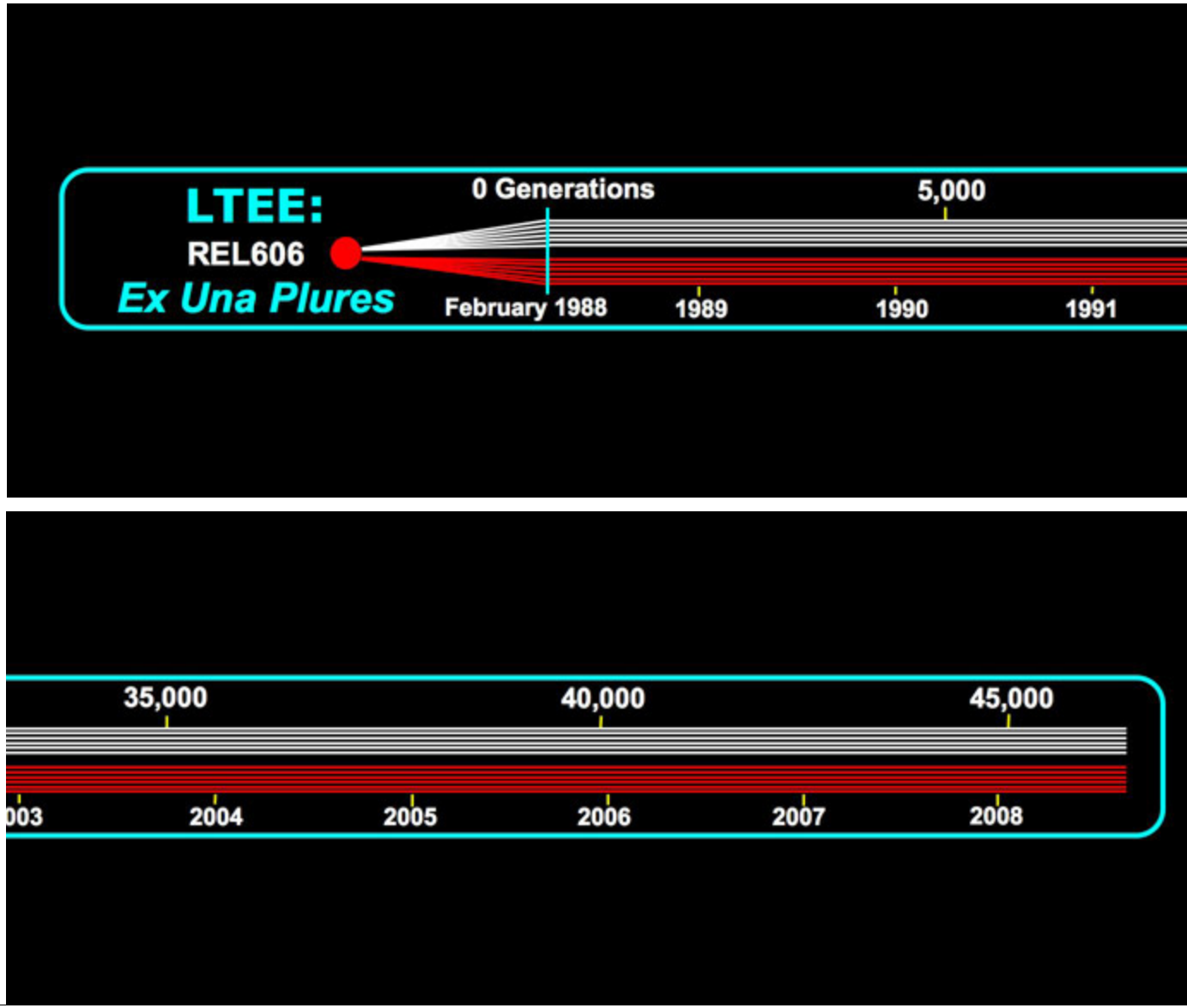
Bacteria



Complex replication dynamics, metabolism, and regulation efficiency are cast into fitness values

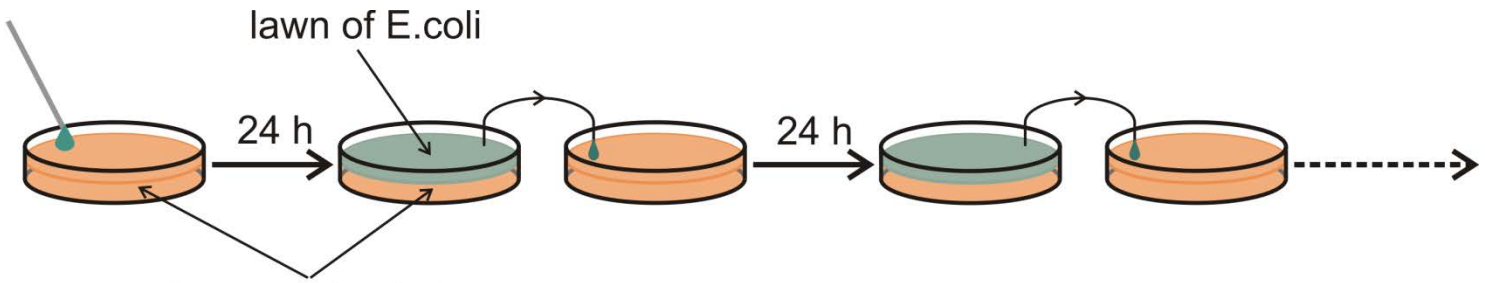
Bacterial evolution in cell-lines

Richard Lenski, 1956 -



Bacterial evolution under controlled conditions: A twenty years experiment.

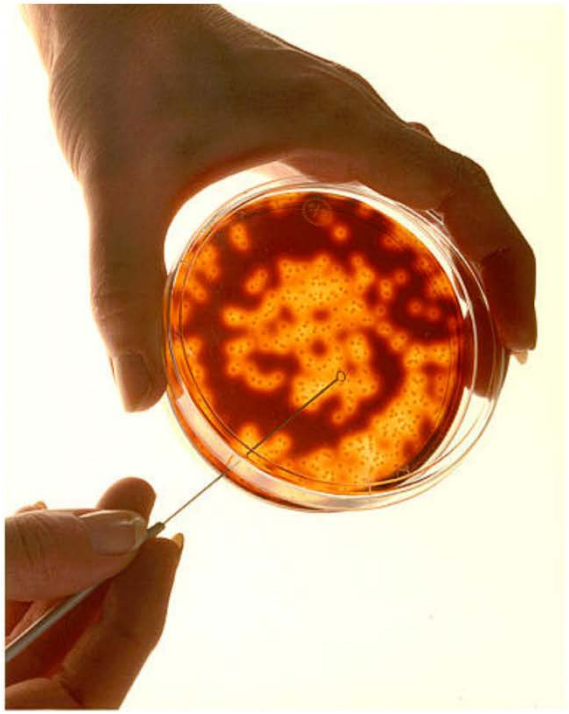
Richard Lenski, University of Michigan, East Lansing



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

1 day	≈	6.67 generations
1 month	≈	200 generations
1 year	≈	**2400 generations**

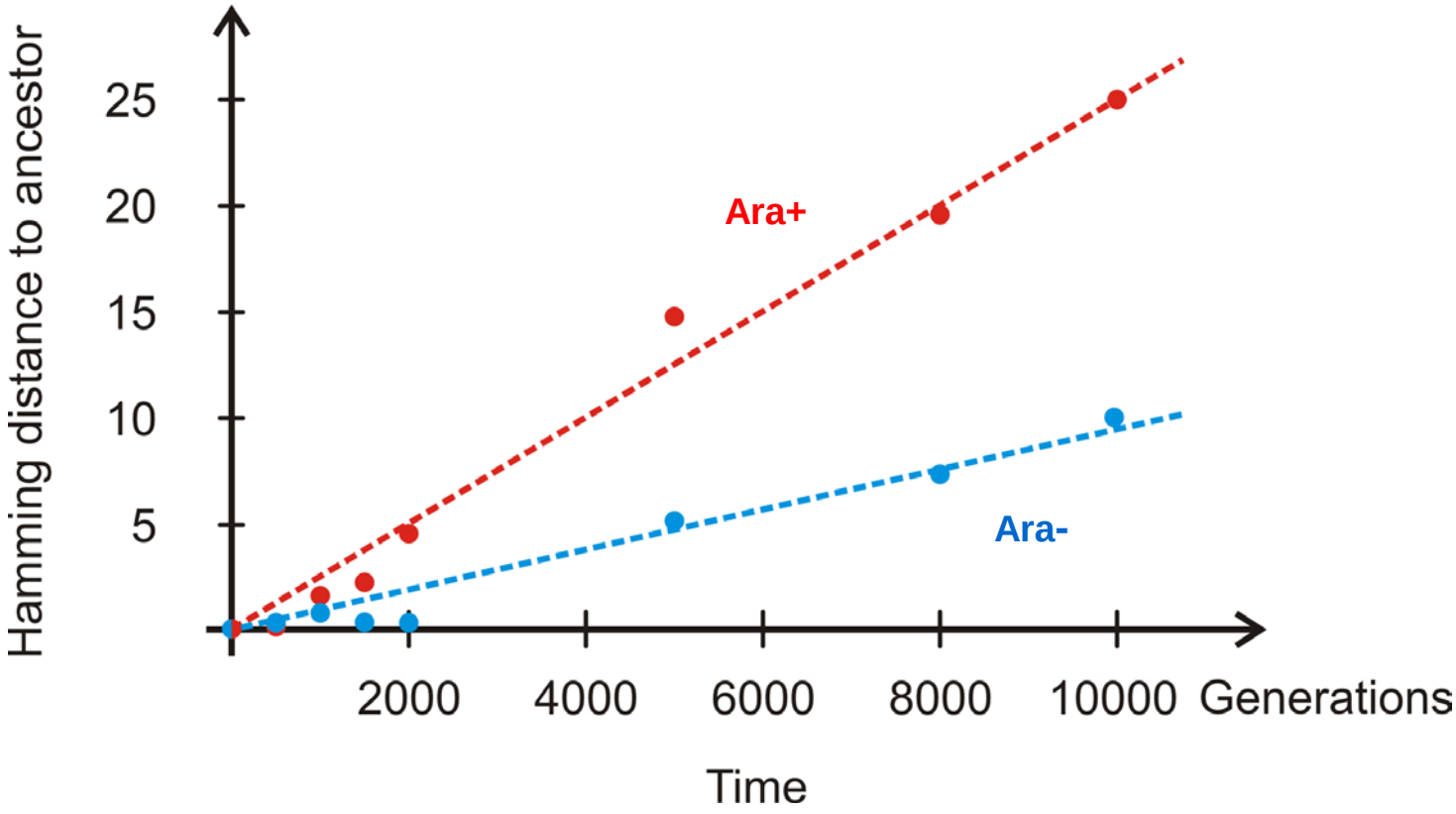
Serial transfer of bacterial cultures in Petri dishes



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



Variation of genotypes in a bacterial serial transfer experiment

D. Papadopoulos, D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, M. Blot. *Genomic evolution during a 10,000-generation experiment with bacteria*. Proc.Natl.Acad.Sci.USA **96** (1999), 3807-3812

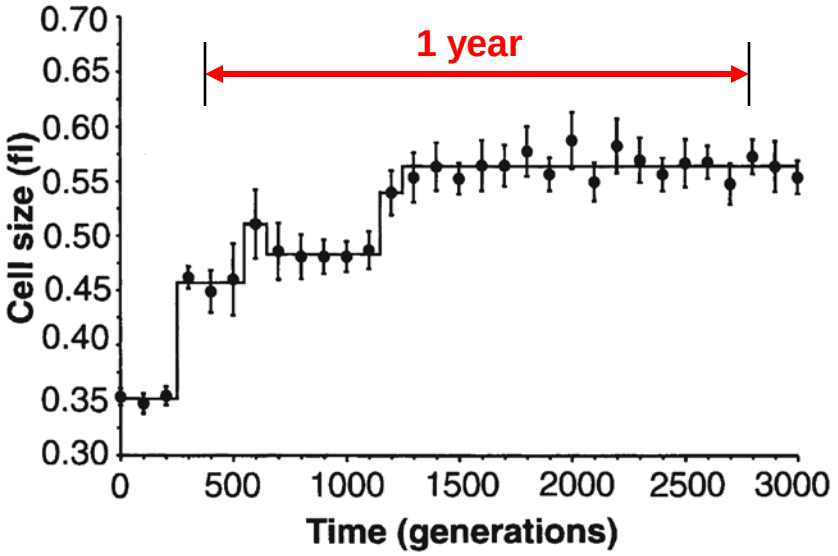
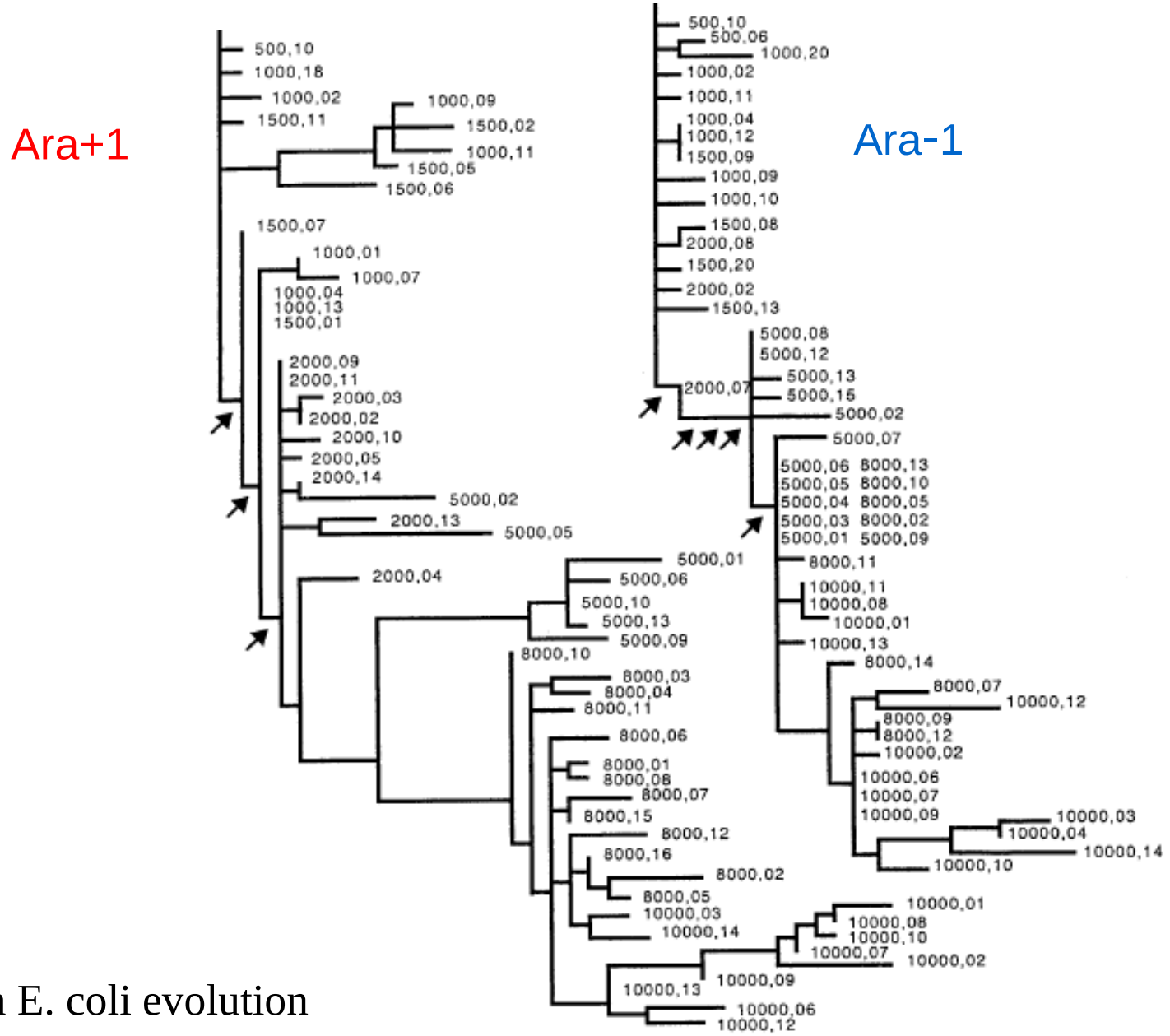


Fig. 1. Change in average cell size (1 fl = 10^{-15} L) in a population of *E. coli* during 3000 generations of experimental evolution. Each point is the mean of 10 replicate assays (*22*). Error bars indicate 95% confidence intervals. The solid line shows the best fit of a step-function model to these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



Phylogeny in E. coli evolution

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

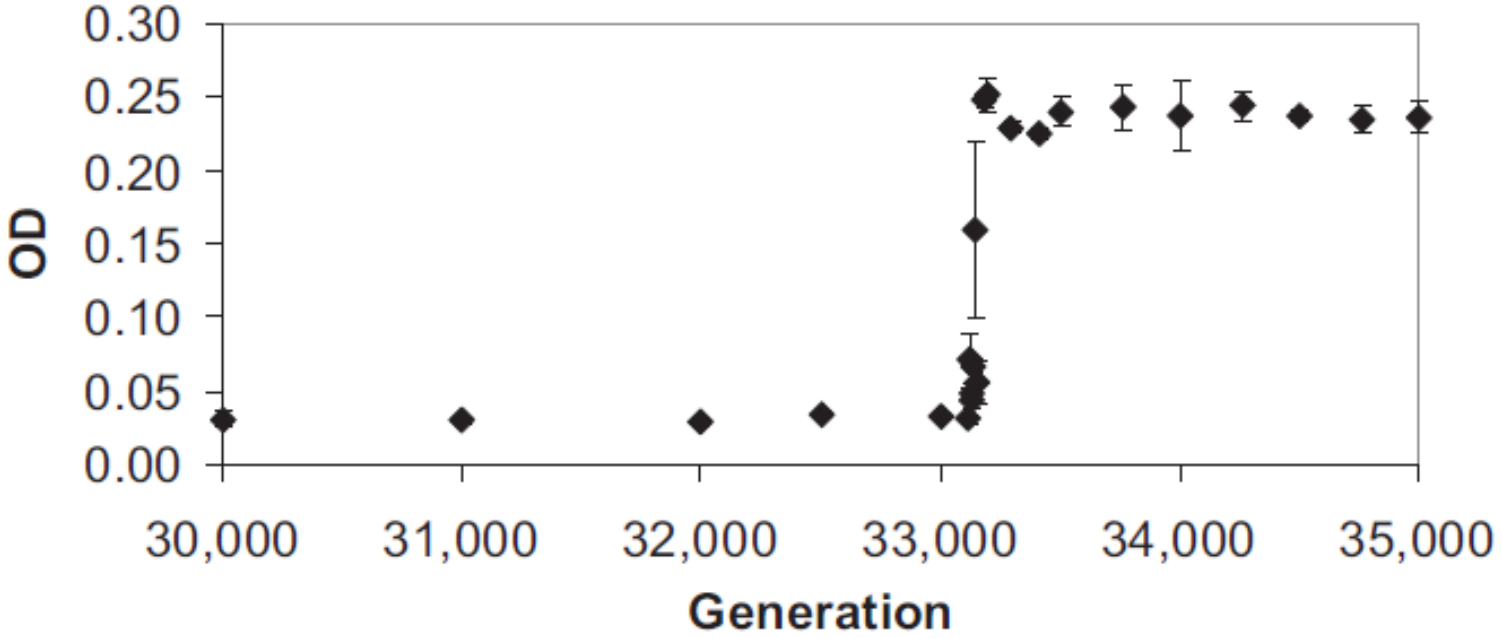


Fig. 1. Population expansion during evolution of the Cit$^+$ phenotype. Samples frozen at various times in the history of population Ara-3 were revived, and three DM25 cultures were established for each generation. Optical density (OD) at 420 nm was measured for each culture at 24 h. Error bars show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

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

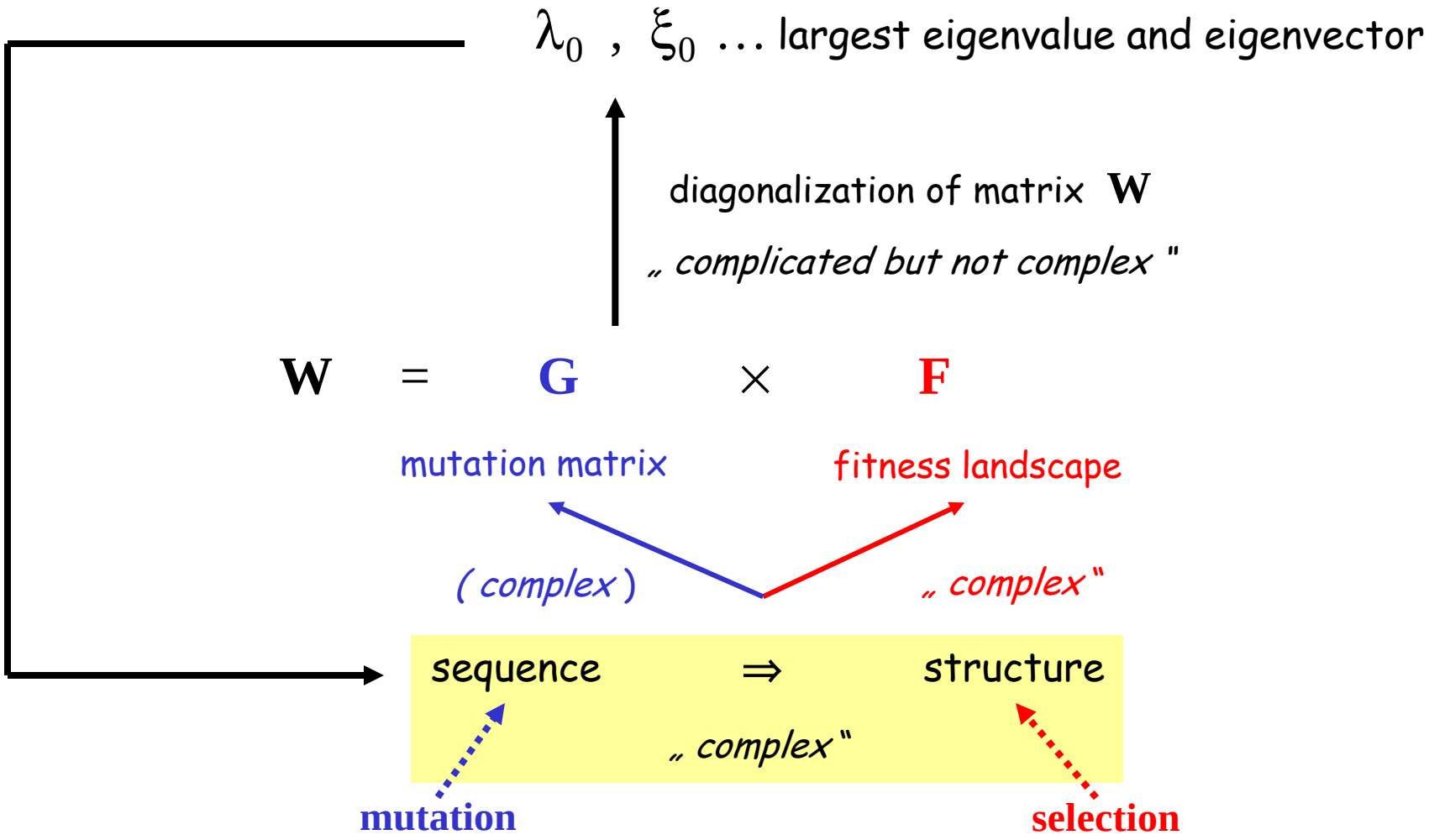
Z.D. Blount, C.Z. Borland, R.E. Lenski. 2008. Proc.Natl.Acad.Sci.USA 105:7899-7906

Table 1. Summary of replay experiments

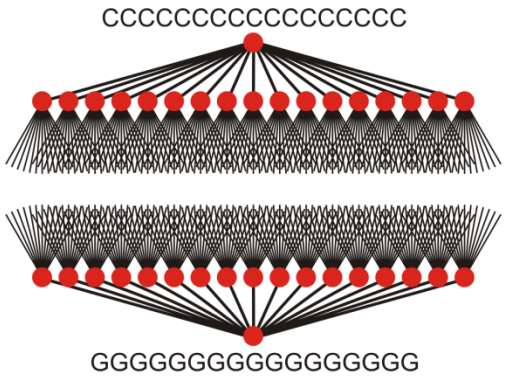
	First experiment		Second experiment		Third experiment	
Generation	Replicates	Independent Cit$^+$ mutants	Replicates	Independent Cit$^+$ mutants	Replicates	Independent Cit$^+$ mutants
Ancestor	6	0	10	0	200	0
5,000	—	—	—	—	200	0
10,000	6	0	30	0	200	0
15,000	—	—	—	—	200	0
20,000	6	0	30	0	200	2
25,000	6	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

Universality of Darwin's mechanism



Complexity in molecular evolution



sequence space

Q

genotype → phenotype → selection

Evolution as a global phenomenon in genotype space

Fitness landscapes are becoming experimentally accessible!

Protein landscapes: Yuuki Hayashi, Takuyo Aita, Hitoshi Toyota, Yuzuru Husimi, Itaru Urabe, Tetsuya Yomo. 2006. Experimental rugged fitness landscape in protein seqeunce 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

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

Facit: 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.

Quo vadis bioinformatics?

Mycoplasma pneumoniae:	genome length	820 000 bp
	# genes:	733
	# proteins (ORF):	689
	# tRNAs	37
	# rRNAs	3
	# other RNAs	4

S. Kühner, V. van Noort, M. J. Betts, A. Leo-Macias, C. Batisse, M. Rode, T. Yamada, T. Maier, S. Bader, P. Beltran-Alvarez, D. Castaño-Diez, W.-H. Chen, D. Devos, M. Güell, T. Norambuena, I. Racke, V. Rybin, A. Schmidt, E. Yus, R. Aebersold, R. Herrmann, B. Böttcher, A. S. Frangakis, R. B. Russell, L. Serrano, P. Bork, and A.-C. Gavin. 2009.
Proteome organization in a genome-reduced bacterium. Science **326**:1235–1240.

E. Yus, T. Maier, K. Michalodimitrakis, V. van Noort, T. Yamada, W.-H. Chen, J. A. Wodke, M. Güell, S. Martínez, R. Bourgeois, S. Kühner, E. Raineri, I. Letunic, O. V. Kalinina, M. Rode, R. Herrmann, R. Gutiérez-Gallego, R. B. Russell, A.-C. Gavin, P. Bork, and L. Serrano. 2009.
Impact of genome reduction on bacterial metabolism and its regulation. Science **326**:1263–1268.

M. Güell, V. van Noort, E. Yus, W.-H. Chen, J. Leigh-Bell, K. Michalodimitrakis, T. Yamada, M. Arumugam, T. Doerks, S. Kühner, M. Rode, M. Suyama, S. Schmidt, A.-C. Gavin, P. Bork, and L. Serrano. 2009.
Transcriptome complexity in a genome-reduced bacterium. Science **326**:1268–1271.

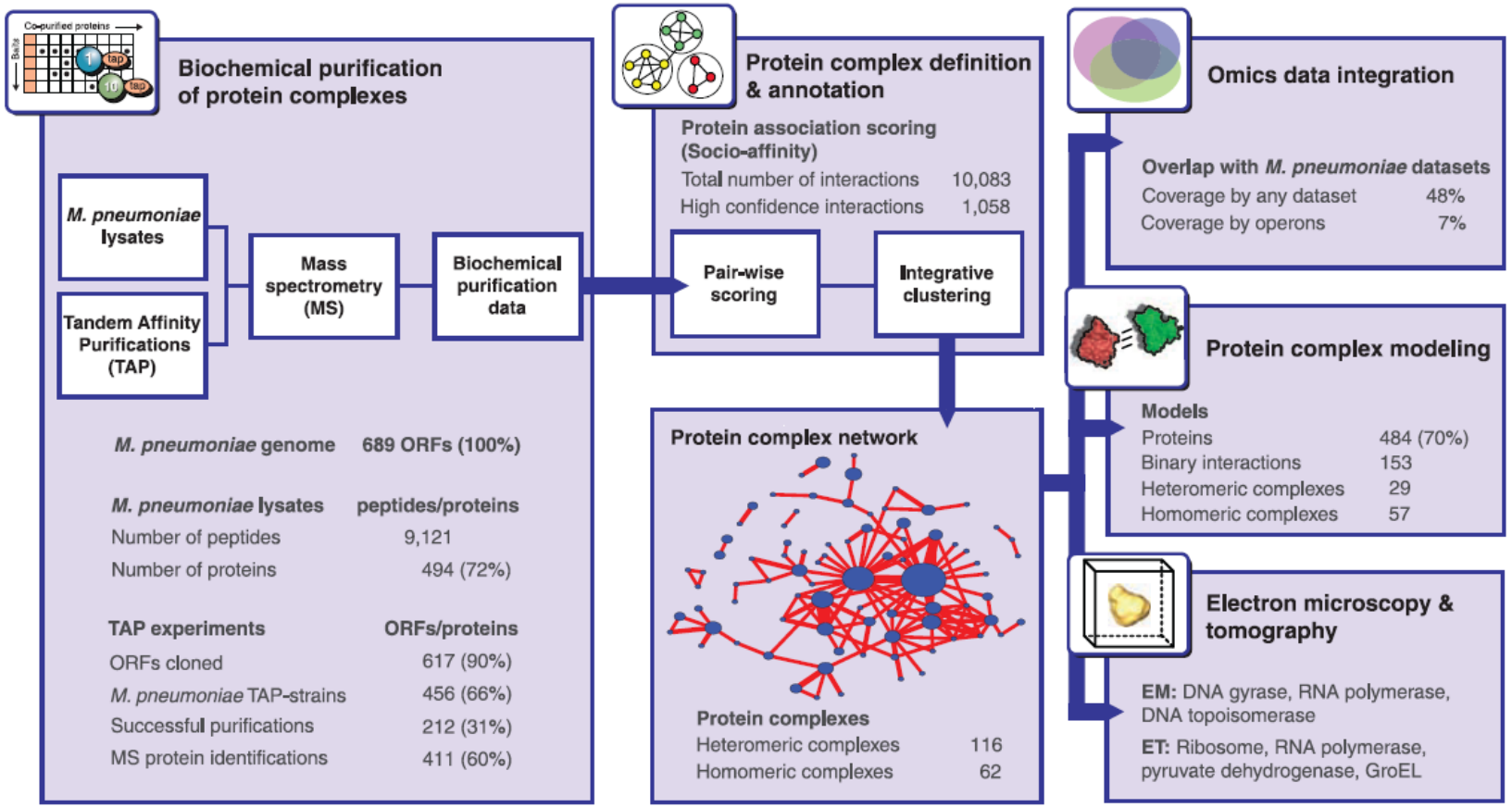


Fig. 1. Synopsis of the genome-wide screen of complexes in *M. pneumoniae*.

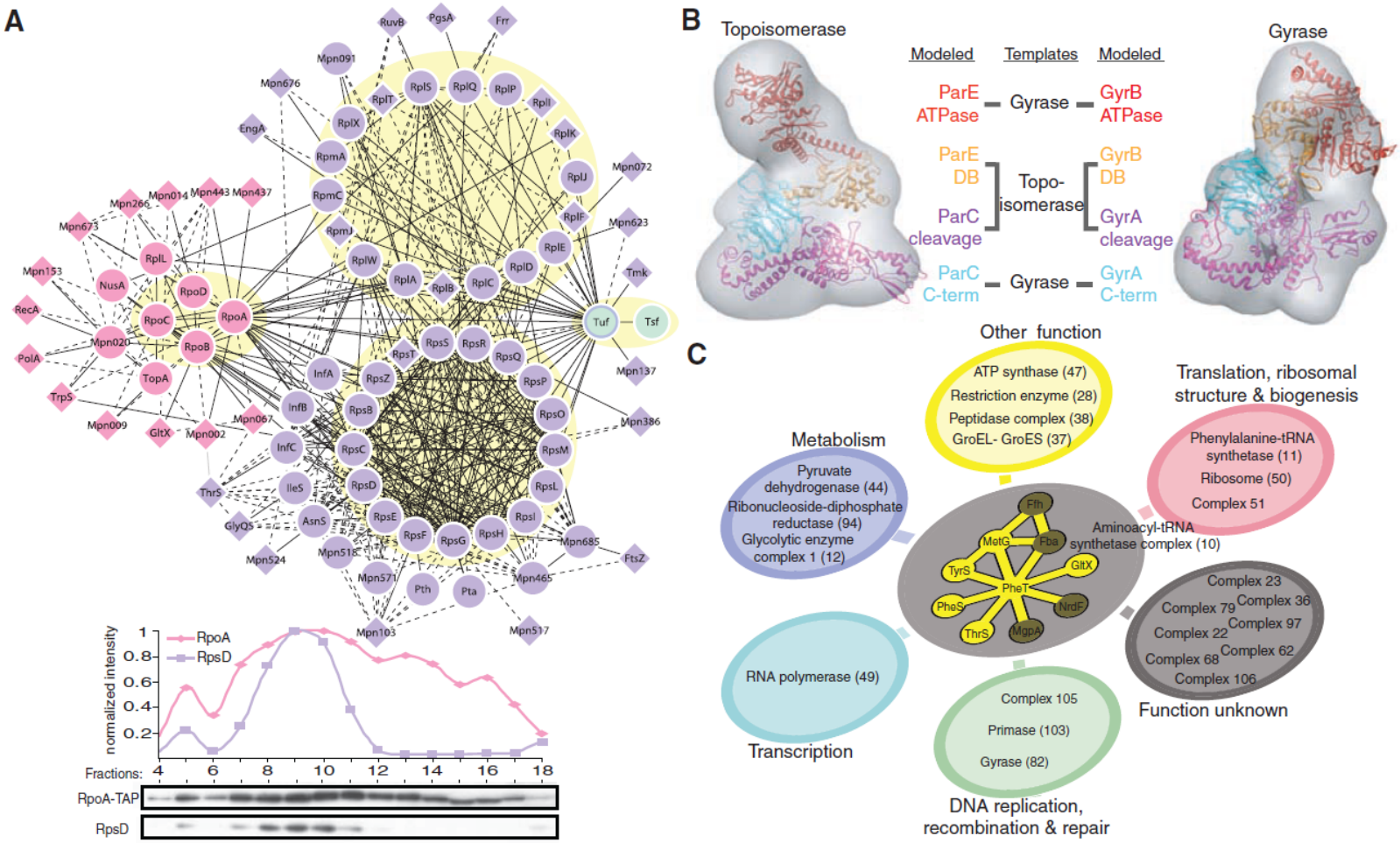


Fig. 3. Higher level of proteome organization. (**A**) The RNA polymerase–ribosome assembly. Core components are represented by circles, attachments by diamonds. The line attribute corresponds to socio-affinity indices: dashed lines, 0.5 to 0.86; plain lines, >0.86. Color code and shaded yellow circles around groups of proteins refer to individual complexes: RNA polymerase (pink), ribosome (purple), and translation elongation factor (green). The bottom graph shows that the ribosomal protein RpsD (23 kD) and the α subunit of the RNA polymerase, RpoA-TAP (57 kD), co-elute in high molecular weight fractions (MD range) during gel filtration chromatography. (**B**) DNA topoisomerase (diameter ~ 12 nm) is a heterodimer in bacteria: ParE (ATPase and DNA binding domains) and ParC (cleavage and C-terminal domains). The interaction between ParE-DNA–binding and ParC–cleavage domains was modeled by using yeast topoisomerase II as a template [Protein Data Bank (PDB) code 2rgr], and ParE-ATPase and ParC–C-terminal domains were modeled separately on structures of gyrase homologs (PDB 1kij and 1suu). All four domains were fitted into the electron microscopy density. Gyrase (~12 nm) is similarly split in bacteria into GyrA/GyrB, which are paralogs of ParE/ParC, and was modeled and fitted by using PDB 1bjt as a template for the GyrB-DNA–binding and GyrA-cleavage domains interaction. (**C**) Protein multifunctionality in *M. pneumoniae* illustrated with the AARS complexes.

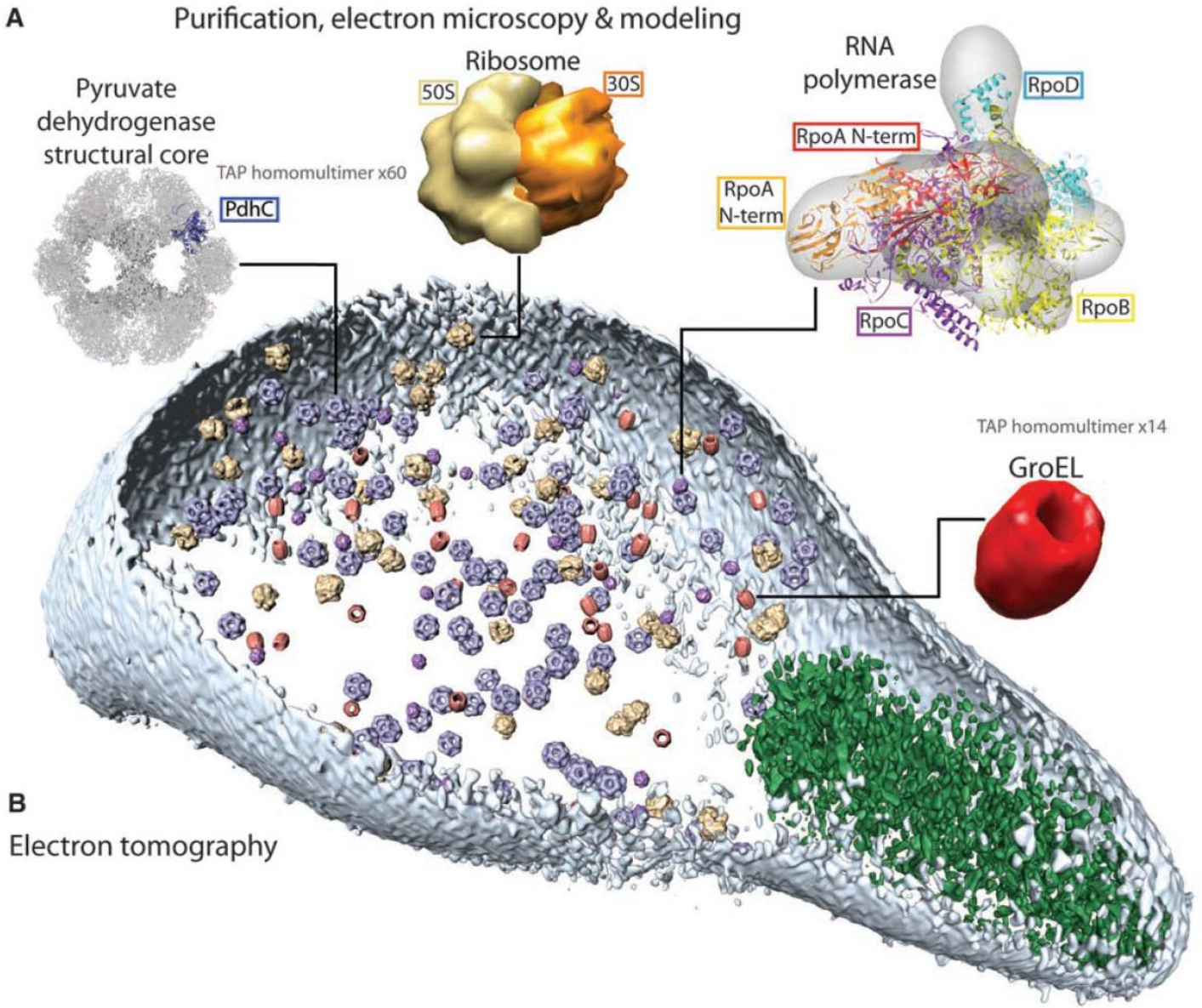
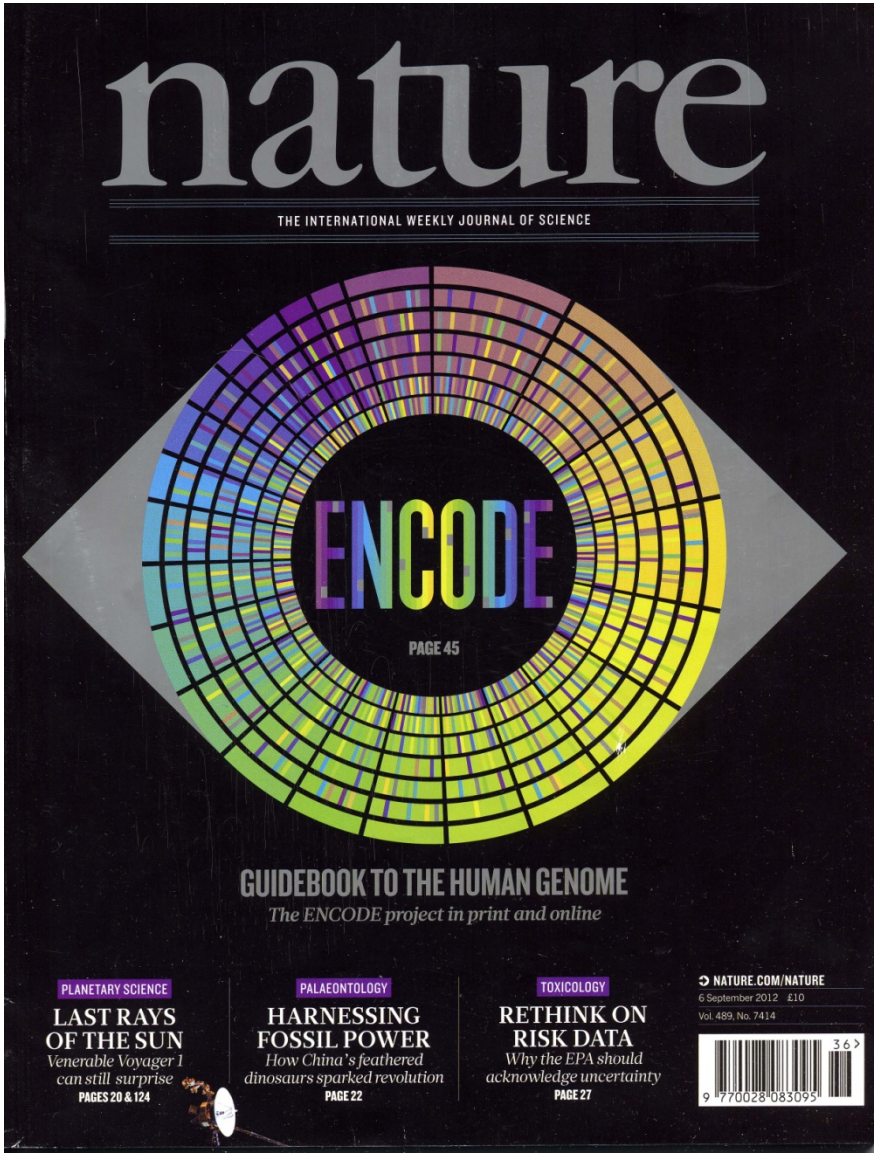


Fig. 4. From proteomics to the cell. By a combination of pattern recognition and classification algorithms, the following TAP-identified complexes from *M. pneumoniae*, matching to existing electron microscopy and x-ray and tomogram structures (**A**), were placed in a whole-cell tomogram (**B**): the structural core of pyruvate dehydrogenase in blue (~23 nm), the ribosome in yellow (~26 nm), RNA polymerase in purple (~17 nm), and GroEL homomultimer in red (~20 nm). Cell dimensions are ~300 nm by 700 nm. The cell membrane is shown in light blue. The rod, a prominent structure filling the space of the tip region, is depicted in green. Its major structural elements are HMW2 (Mpn310) in the core and HMW3 (Mpn452) in the periphery, stabilizing the rod (*42*).The individual complexes (A) are not to scale, but they are shown to scale within the bacterial cell (B).



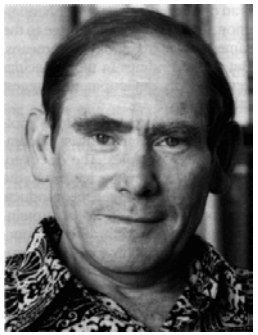
ENCyclopedia **O**f **D**NA **E**lements

Advantages of the molecular approach

1. Complex reproduction mechanisms are readily included.

2. Gene regulation – DNA or RNA based – is chemical kinetics!

3. Accounting for epigenetic effects requires just the simultaneous consideration of several generations.



Sydney Brenner, 1927 -

What else is epigenetics than a funny form of enzymology ? Each protein, after all, comes from some piece of DNA.

Universität Wien

Coworkers

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

Günter Wagner, Yale University, CT

Walter Fontana, Harvard Medical School, MA

Martin Nowak, Harvard University, MA

Christian Reidys, University of Southern Denmark, Odense, DM

Sebastian Bonhoeffer, Eidgenössische Technische Hochschule, Zürich, CH

Christian Forst, Texas Medical Center, Dallas, TX

Thomas Wiehe, Universität Köln, DE

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

Happy birthday bioinfomatics in Leipzig !

Thank you for your attention!

Web-Page for further information:

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