Darwin, self-organization and molecules

Peter Schuster

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

Darwin and the Origin of Life

San Sebastian - Donostia, 22.05.2009

Web-Page for further information:

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

Color patterns on animal skins and wings

1. Pattern formation in physics and chemistry
2. Pattern formation in biology
3. Darwins natural selection and the origin of life
4. Molecular biology and evolution
5. Evolution in the test tube
6. Complexity in biology

1. **Pattern formation in physics and chemistry**
2. Pattern formation in biology
3. Darwins natural selection and the origin of life
4. Molecular biology and evolution
5. Evolution in the test tube
6. Complexity in biology

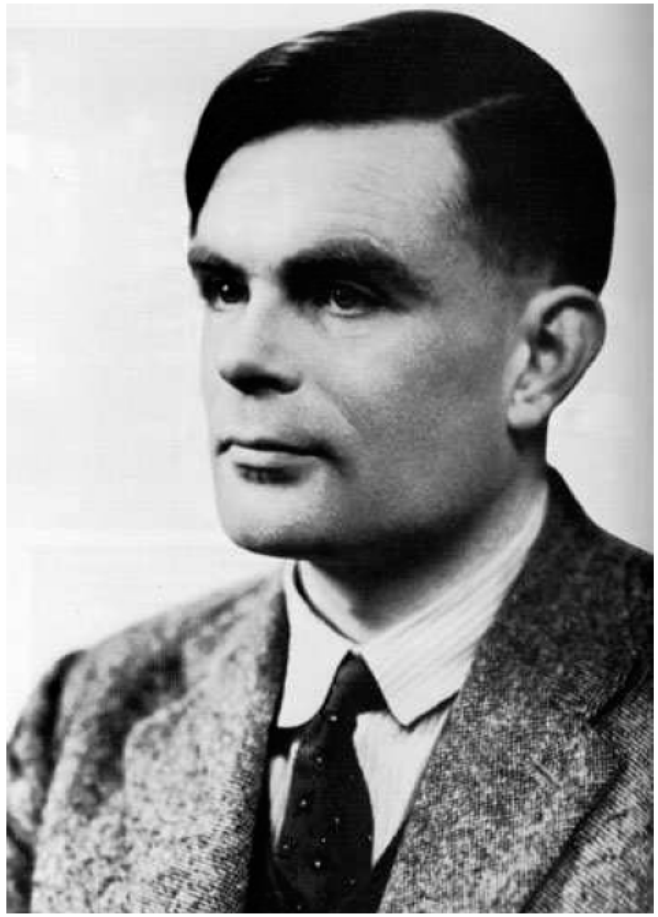
$$\frac{\partial u}{\partial t} = D_u \nabla^2 u + f(u,v)$$

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

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

Change in local concentration =

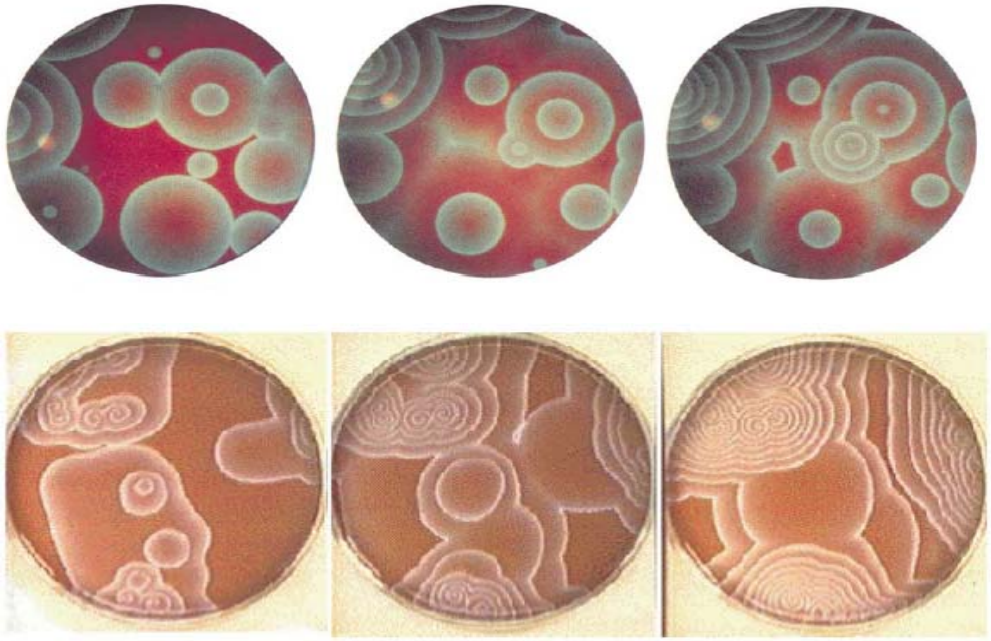
= diffusion + chemical reaction



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



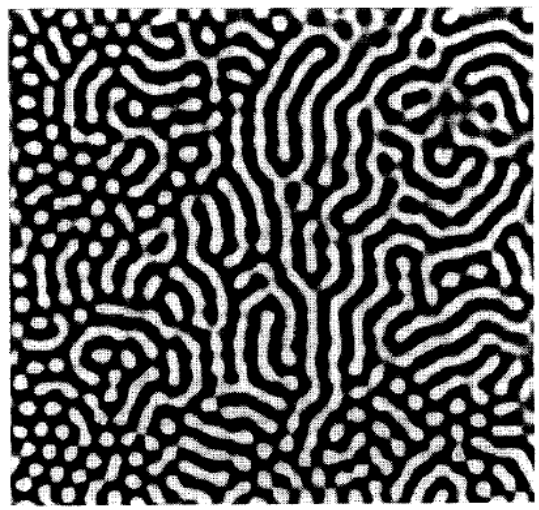
Belousov-Zhabotinskii reaction 1959

Pattern formation through chemical self-organization:

Liesegang rings through precipitation from oversaturated solutions,

space-time patterns of the Belousov-Zhabotinskii reaction,

and stationary Turing pattern.



Turing pattern:
Boissonade, De Kepper 1990

1. Pattern formation in physics and chemistry
2. **Pattern formation in biology**
3. Darwins natural selection and the origin of life
4. Molecular biology and evolution
5. Evolution in the test tube
6. Complexity in biology

mother

presumptive father

daughter

Skin patterns in an inbred strain of cats

Parents and daughter

Bates' mimicry

Müller's mimicry

Different forms of mimicry observed in nature

milk snake

Bates' mimicry

false coral snake

coral snake

Emsley's or Mertens' mimicry

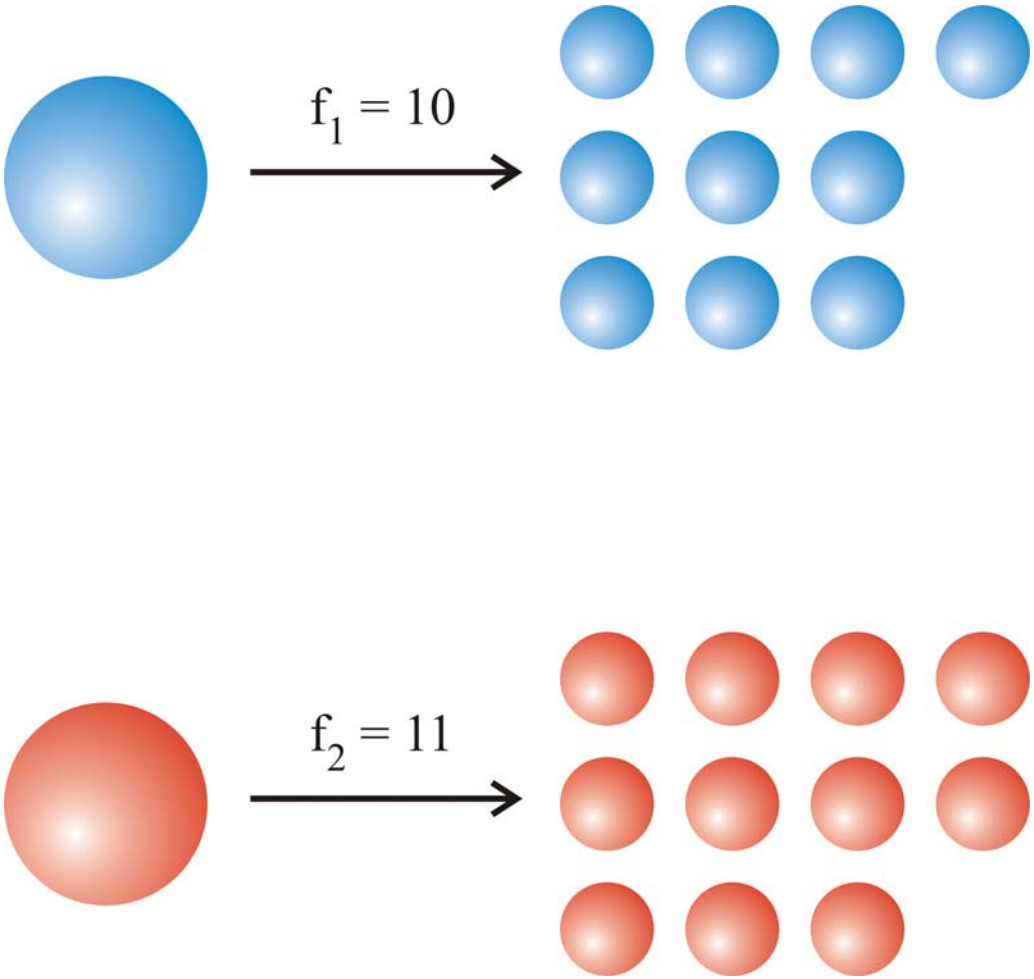
Different forms of mimicry observed in nature

1. Pattern formation in physics and chemistry
2. Pattern formation in biology
3. **Darwins natural selection and the origin of life**
4. Molecular biology and evolution
5. Evolution in the test tube
6. Complexity in biology

Three necessary conditions for Darwinian evolution:

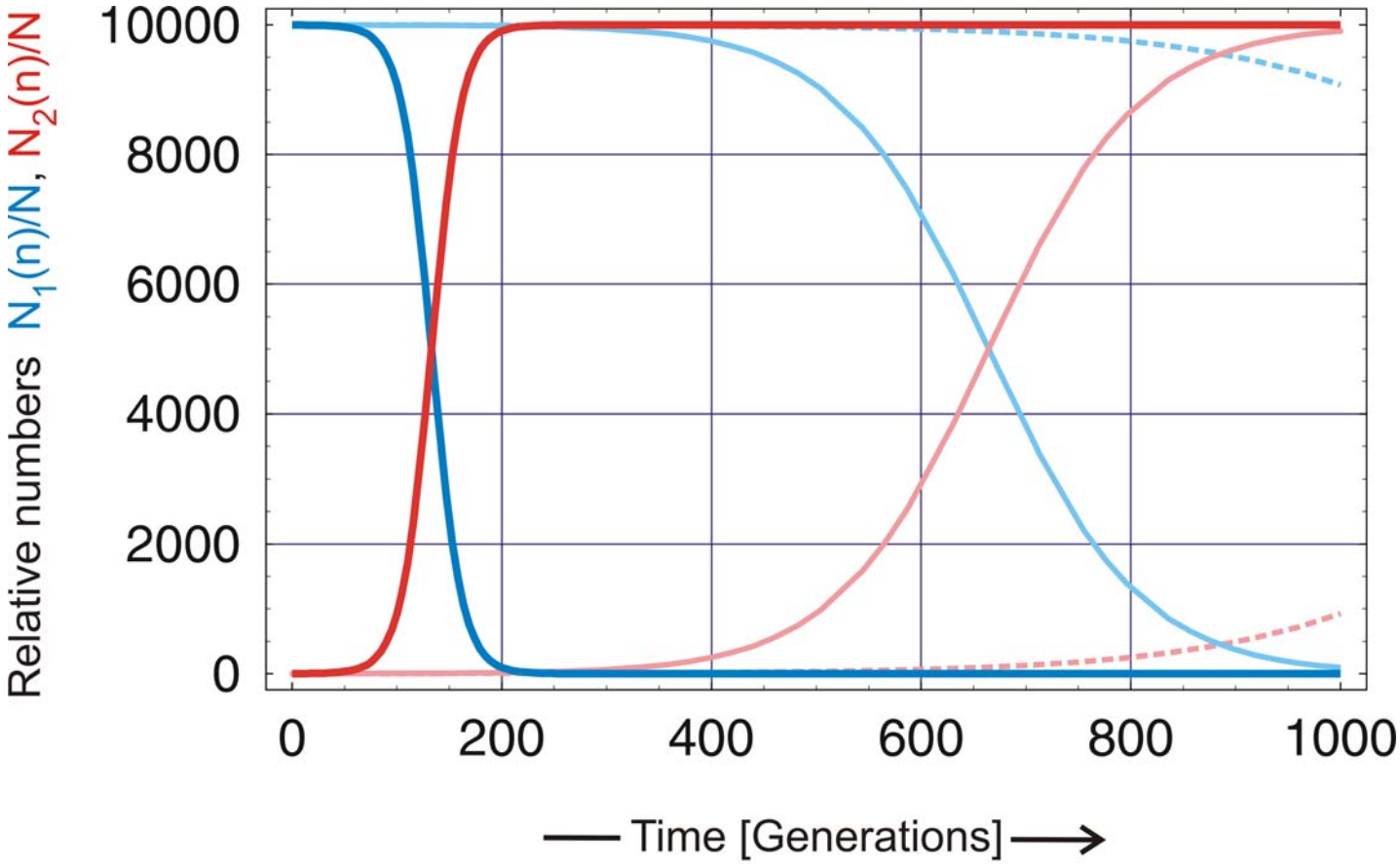
1. **Multiplication**
2. **Variation**
3. **Selection**

Empirically recognized principle of natural selection



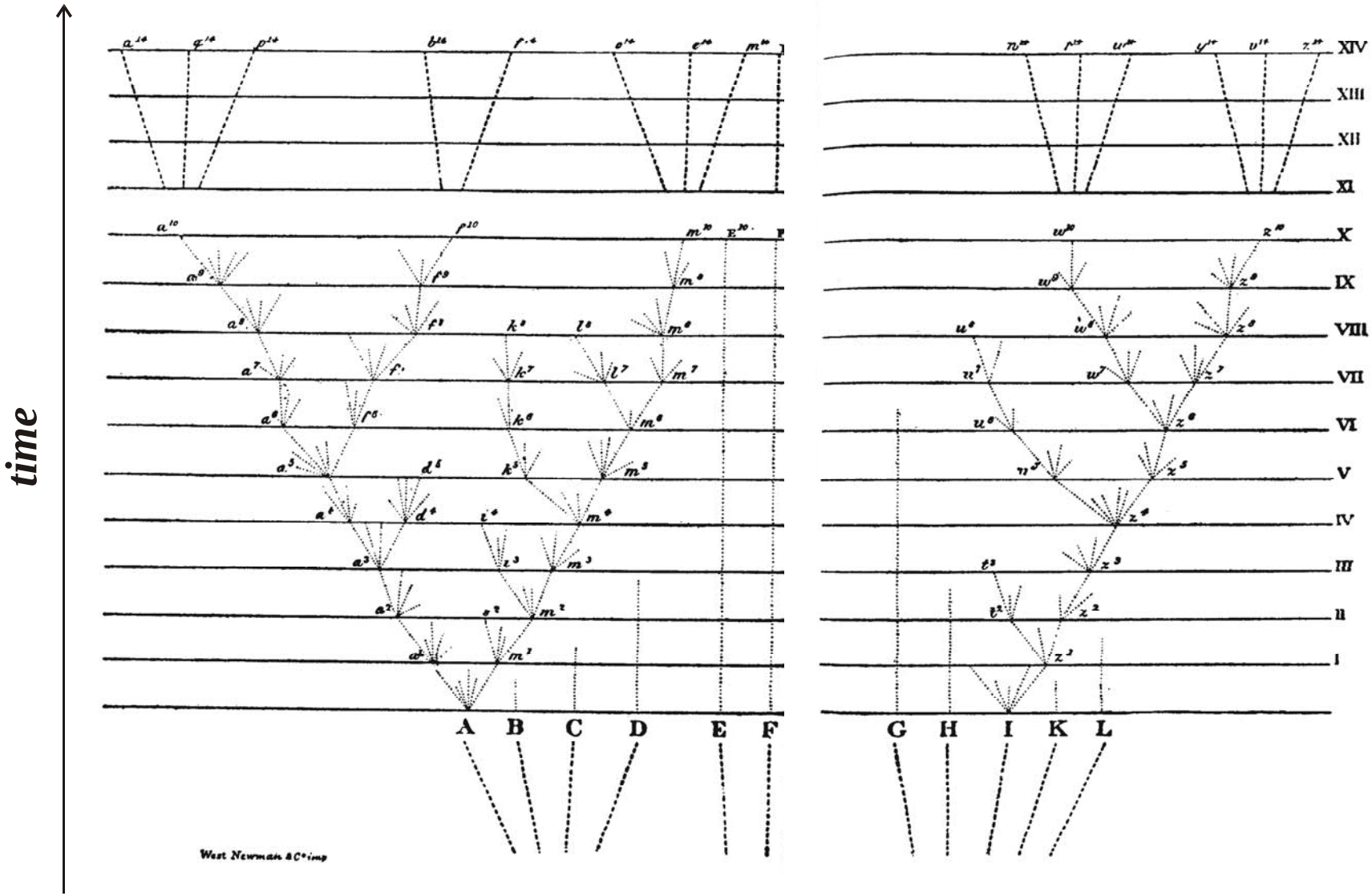
$$s = \frac{f_2 - f_1}{f_1} = 0.1$$

Two variants with a mean progeny of ten or eleven descendants

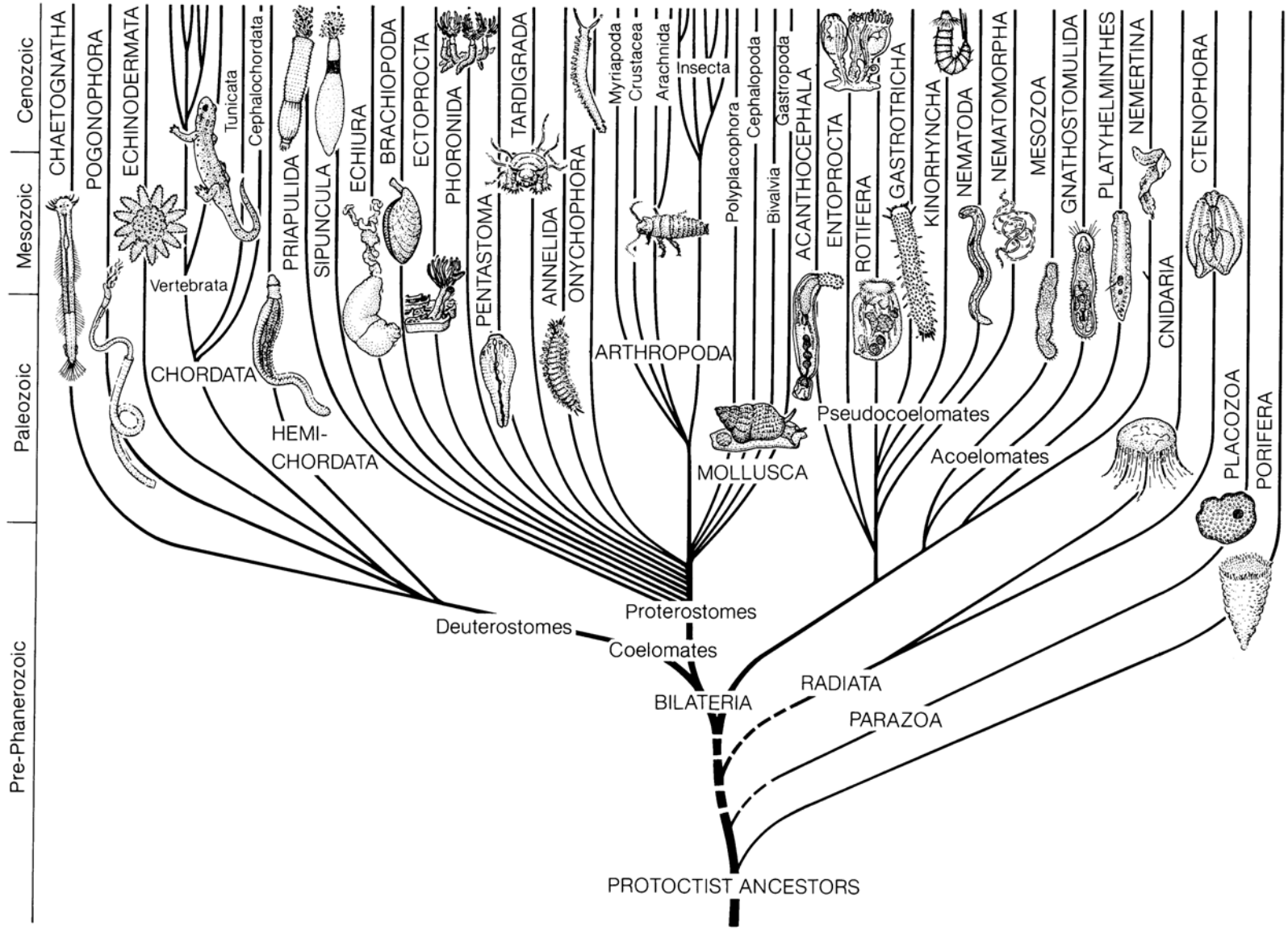


$$N_1(0) = 9999,\ N_2(0) = 1; \quad s = 0.1, 0.02, 0.01$$

Selection of advantageous mutants in populations of N = 10 000 individuals



Charles Darwin, *The Origin of Species*, 6th edition.
Everyman‘s Library, Vol.811, Dent London, pp.121-122.



Modern phylogenetic tree: Lynn Margulis, Karlene V. Schwartz. *Five Kingdoms. An Illustrated Guide to the Phyla of Life on Earth*. W.H. Freeman, San Francisco, 1982.

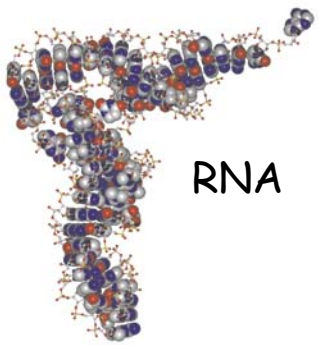
1. Pattern formation in physics and chemistry
2. Pattern formation in biology
3. Darwins natural selection and the origin of life
4. **Molecular biology and evolution**
5. Evolution in the test tube
6. Complexity in biology

Genotype, Genome

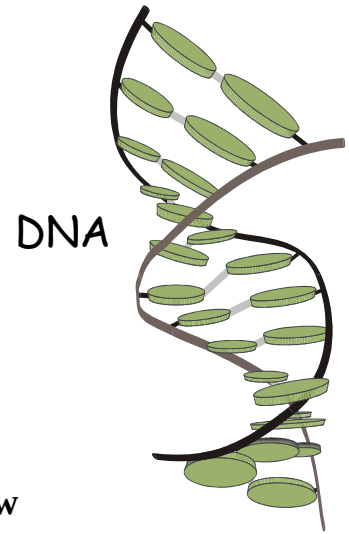
GCGGATTTAGCTCAGTTGGGAGAGCGCCAGACTGAAGATCTGGAGGTCCTGTGTTCGATCCACAGAATTCGCACCA

systems biology

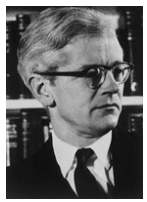
'the new biology is the chemistry of living matter'



RNA



DNA



John Kendrew

Phenotype

Thomas Cech

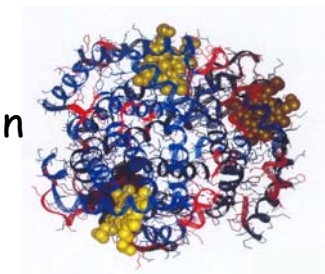
RNA catalysis

Manfred Eigen

Linus Pauling and Emile Zuckerkandl

molecular evolution

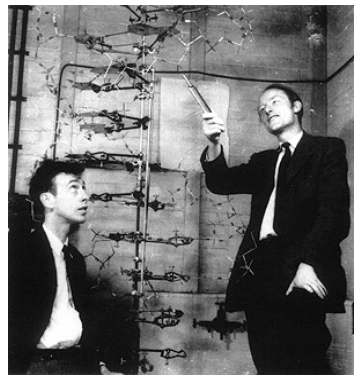
protein



Gerhard Braunitzer

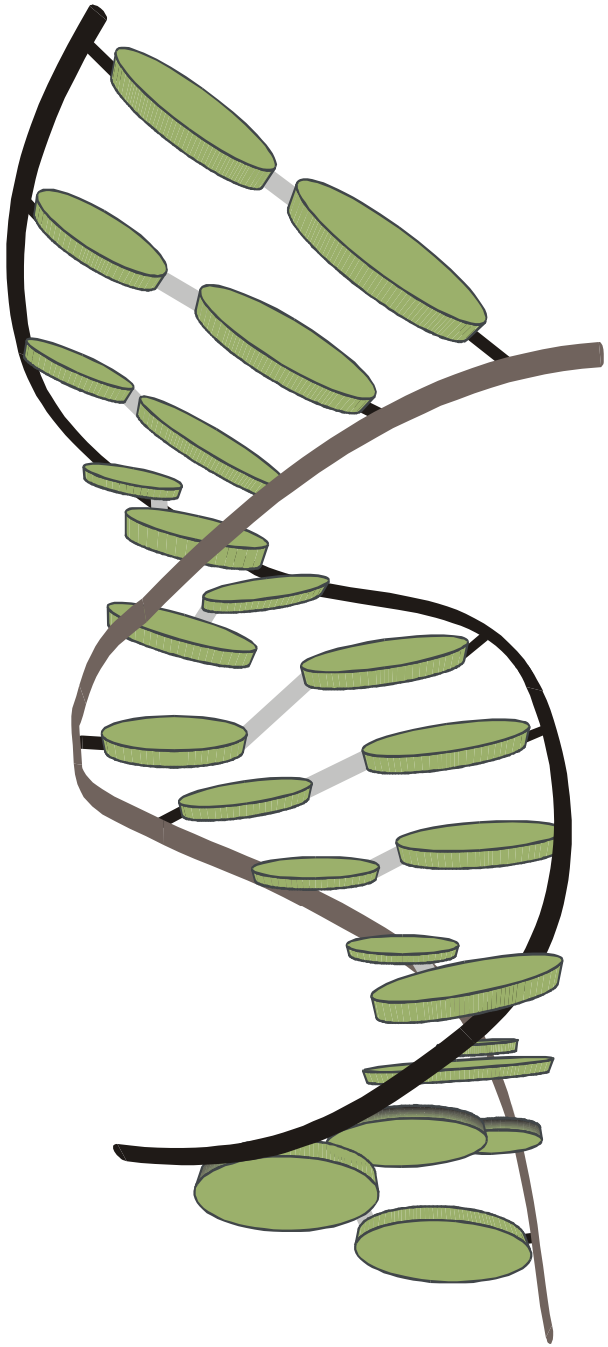
hemoglobin sequence

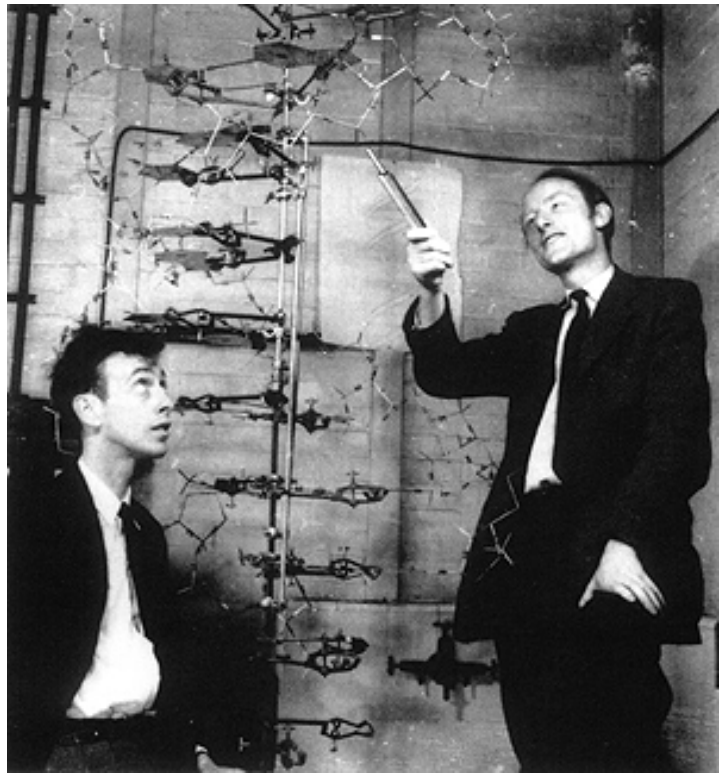
Max Perutz



James D. Watson und Francis H.C. Crick

DNA structure



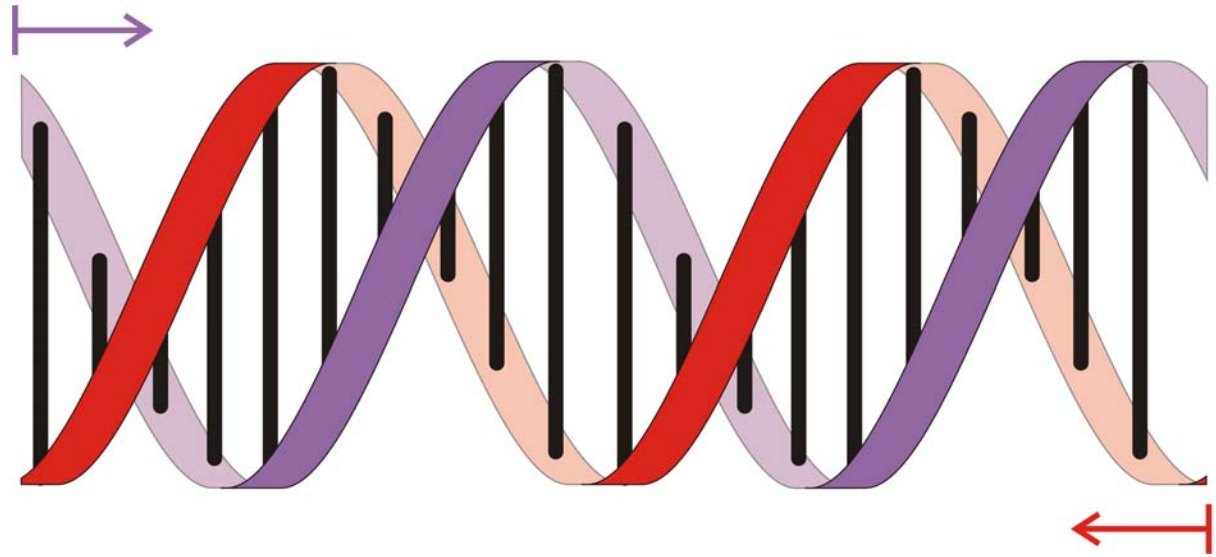


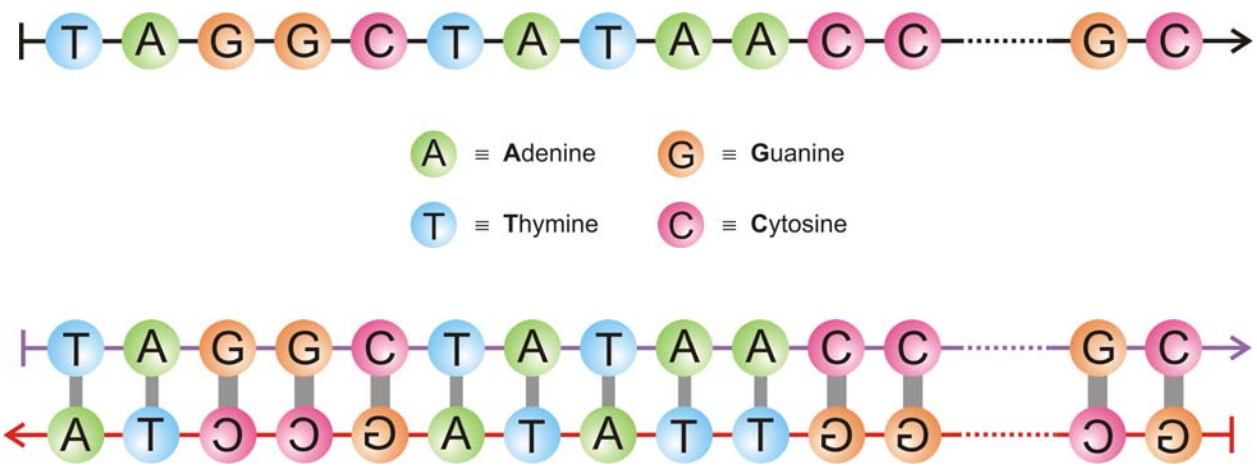
James D. Watson, 1928-, and Francis H.C. Crick, 1916-2004

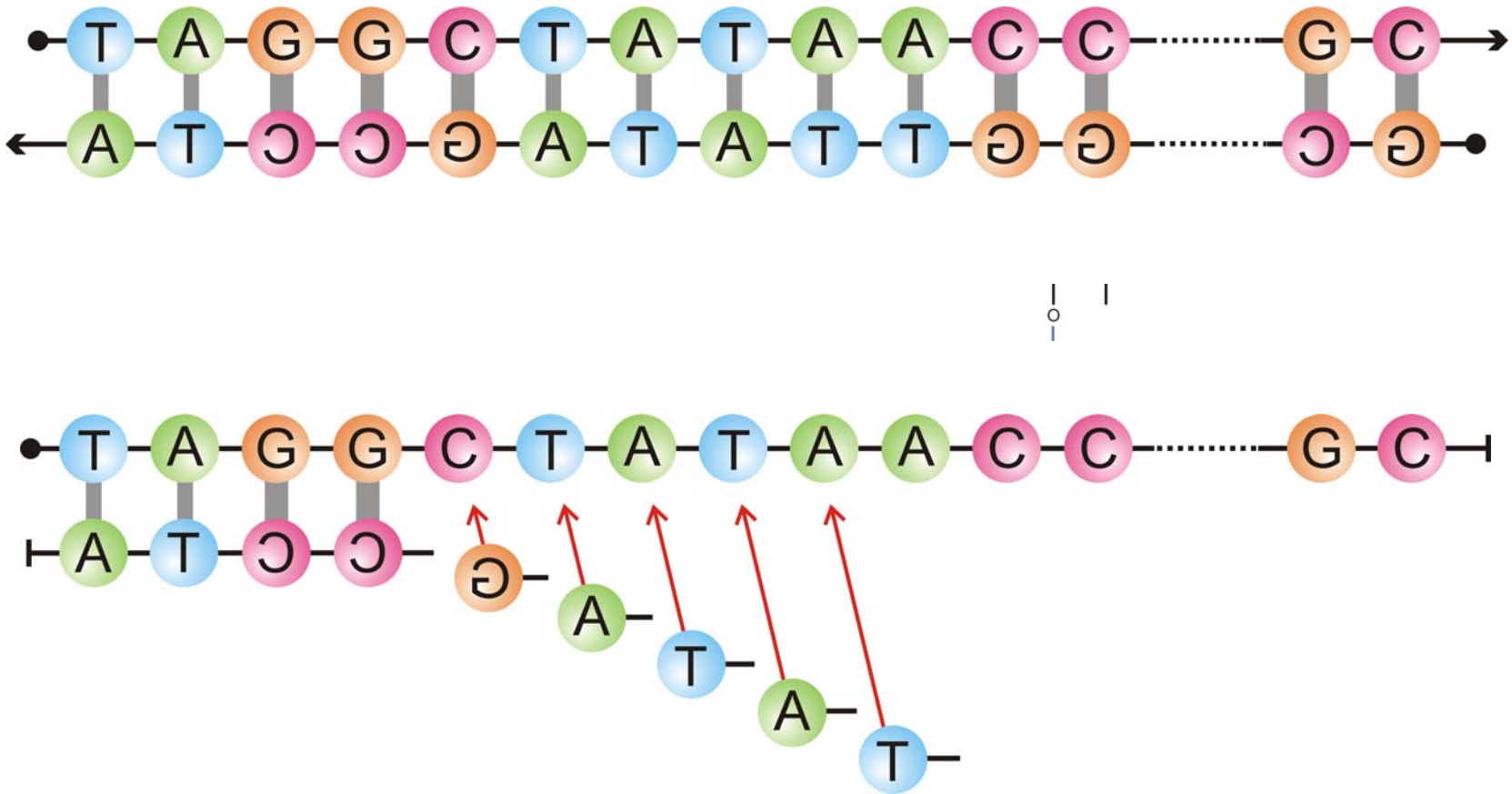
Nobel prize 1962

1953 – 2003 fifty years double helix

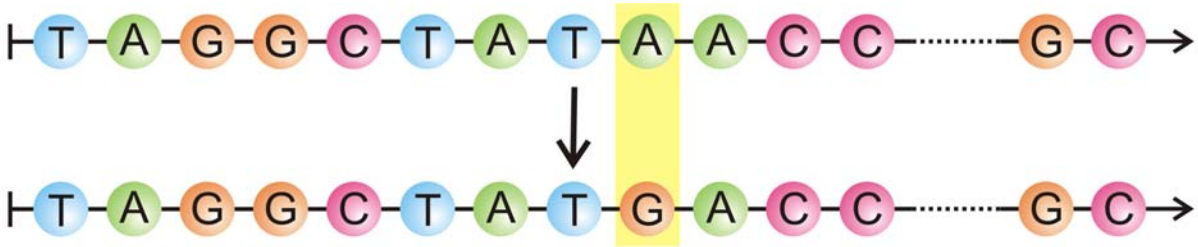
The three-dimensional structure of a short double helical stack of B-DNA



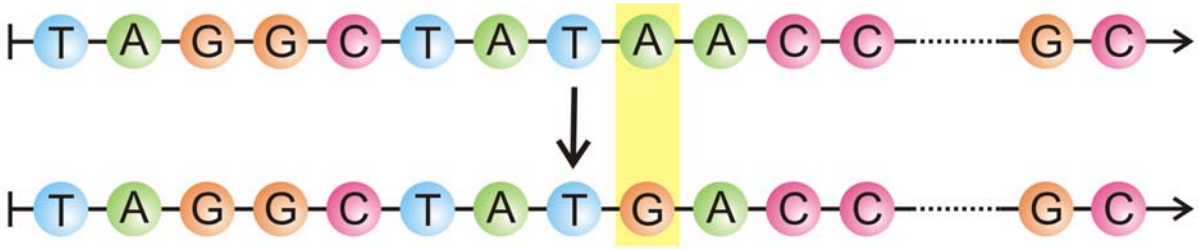

T A G G C T A T A A C C G C
A ≡ Adenine
G ≡ Guanine
T ≡ Thymine
C ≡ Cytosine
T A G G C T A T A A C C G C
A T C C G A T A T T G G C G



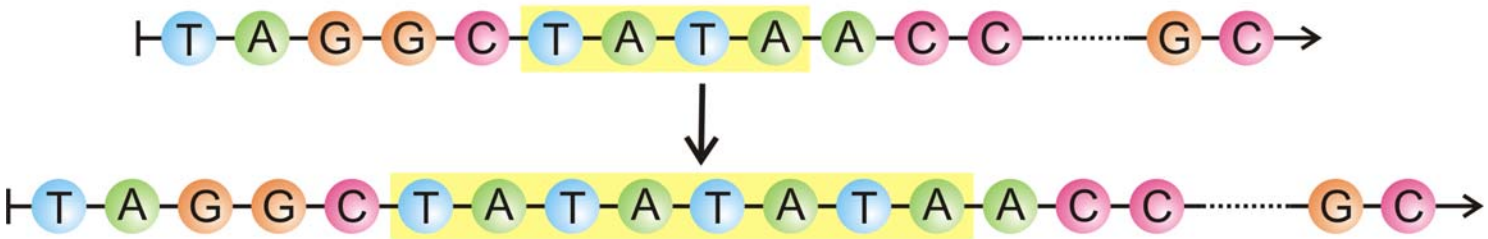
DNA structure and DNA replication



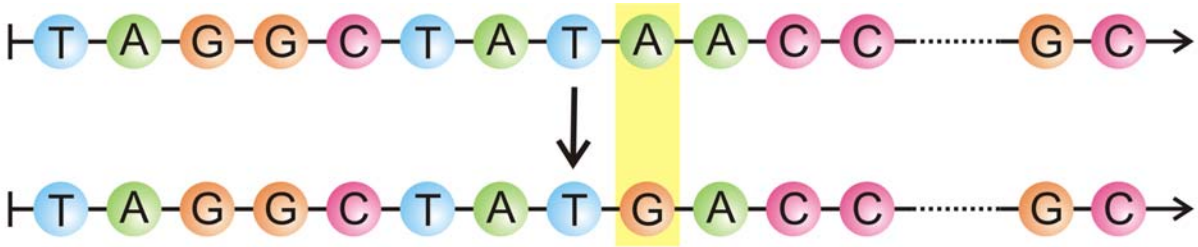
Point mutation



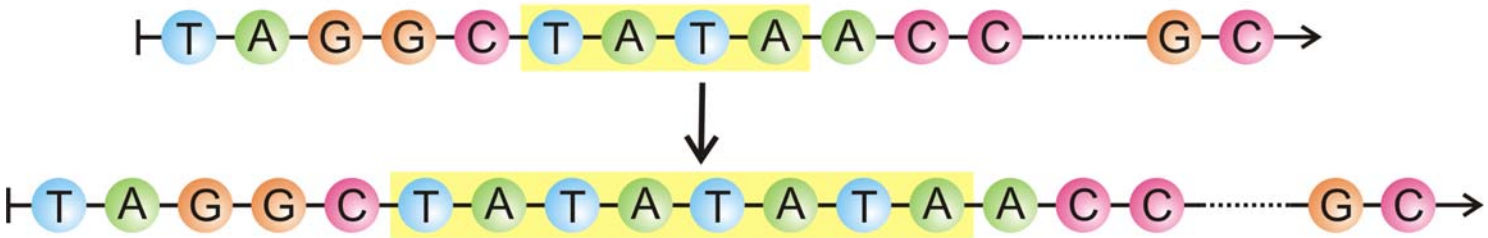
Point mutation



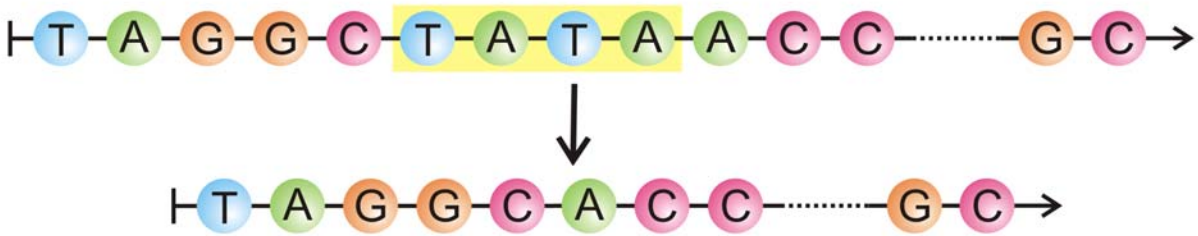
Insertion



Point mutation



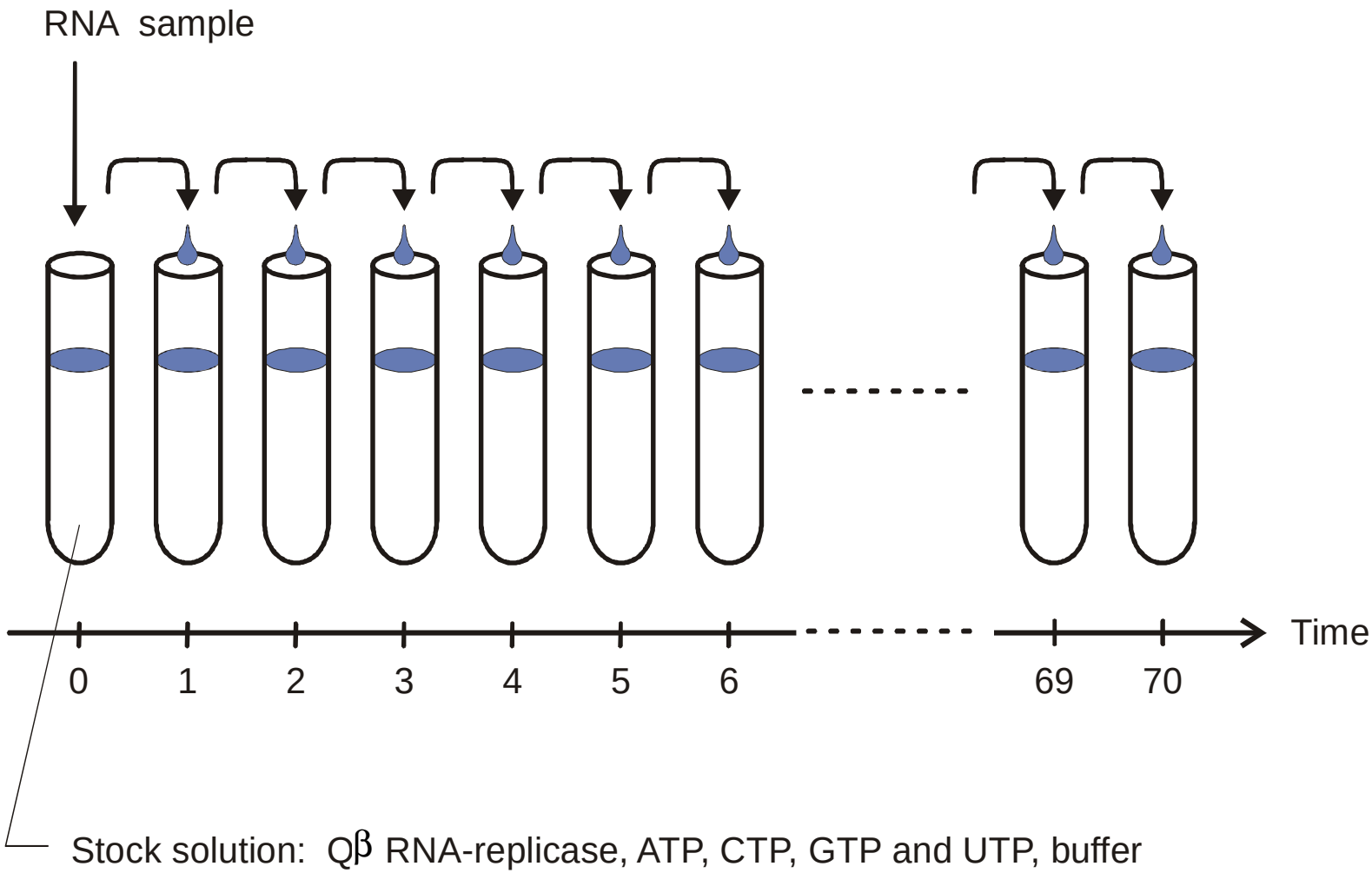
Insertion



Deletion

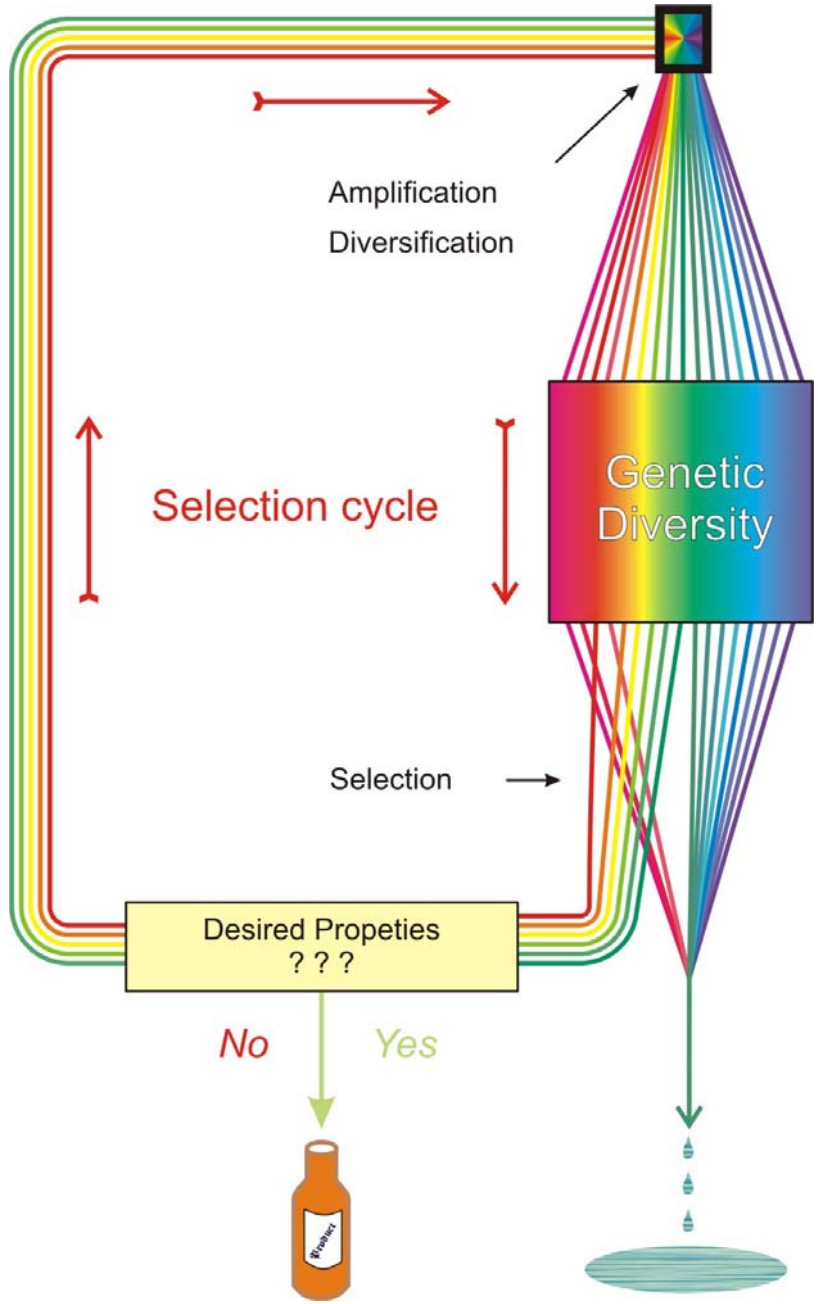
Reconstruction of phylogenies through comparison of molecular sequence data

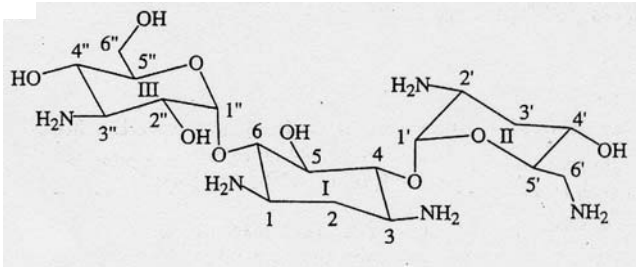
1. Pattern formation in physics and chemistry
2. Pattern formation in biology
3. Darwins natural selection and the origin of life
4. Molecular biology and evolution
5. **Evolution in the test tube**
6. Complexity in biology



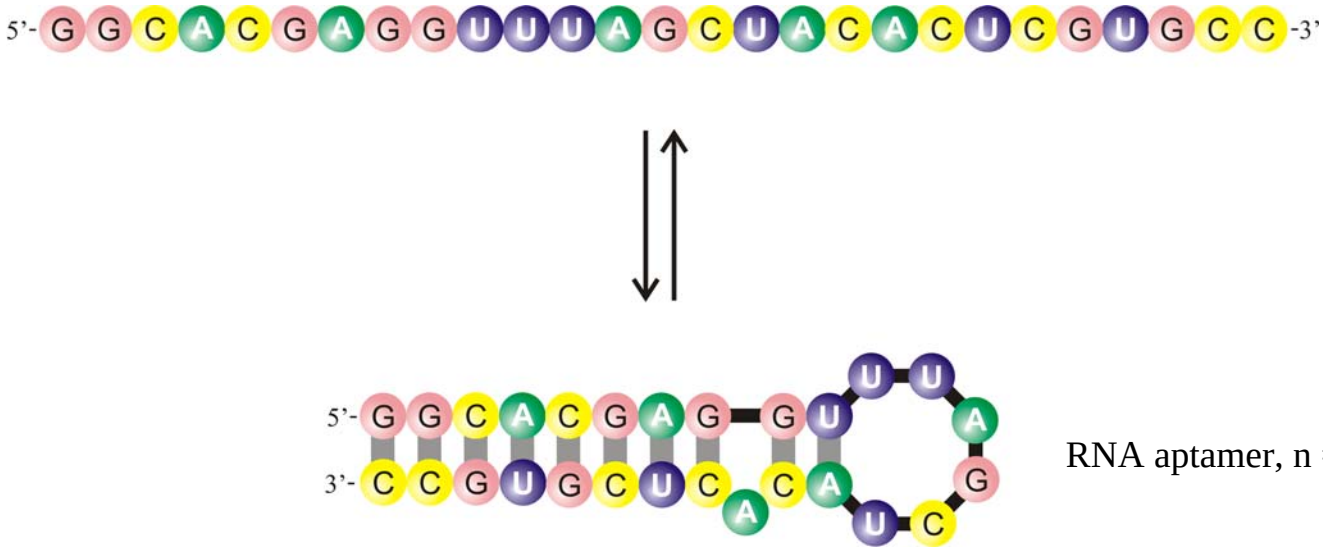
Application of the serial transfer technique to RNA-evolution in the test tube

An example of ‘artificial selection’ with RNA molecules or ‘breeding’ of biomolecules





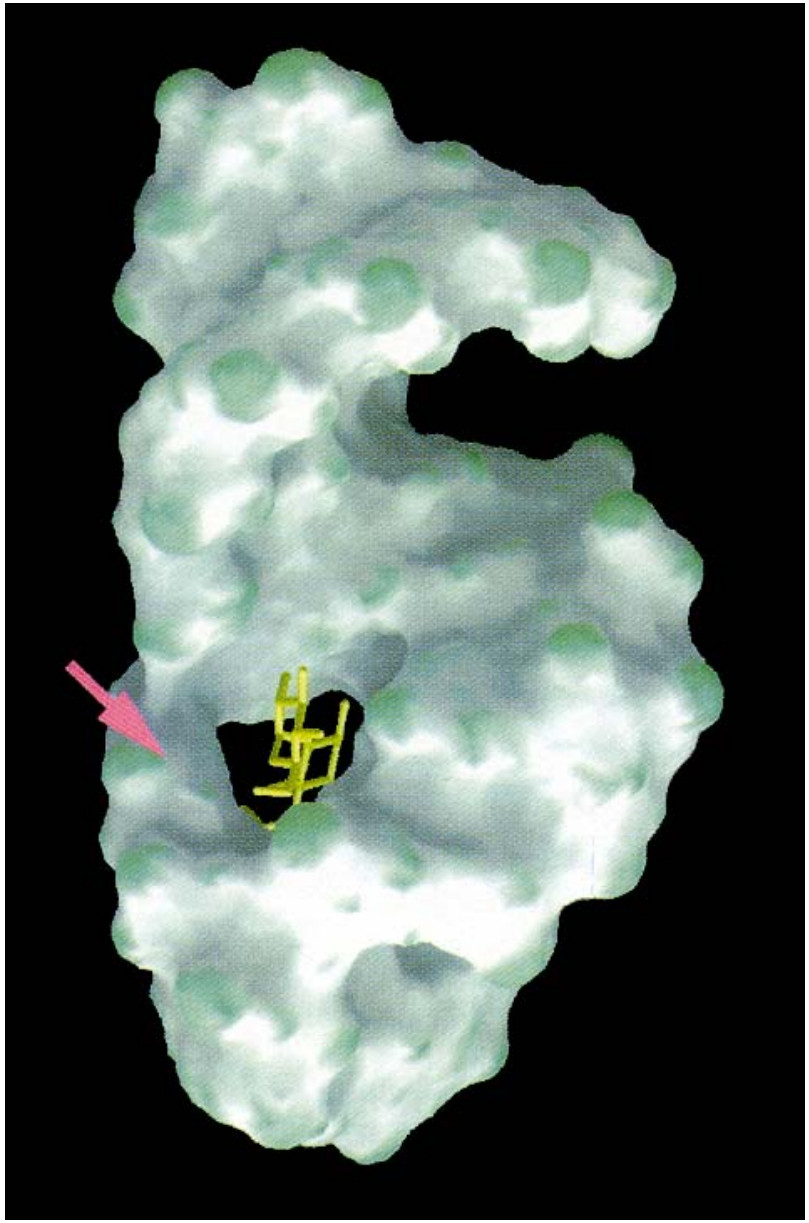
tobramycin



RNA aptamer, n = 27

Formation of secondary structure of the tobramycin binding RNA aptamer with **K_D = 9 nM**

L. Jiang, A. K. Suri, R. Fiala, D. J. Patel, ***Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex.*** Chemistry & Biology **4**:35-50 (1997)



The three-dimensional structure of the tobramycin aptamer complex

L. Jiang, A. K. Suri, R. Fiala, D. J. Patel, Chemistry & Biology **4**:35-50 (1997)

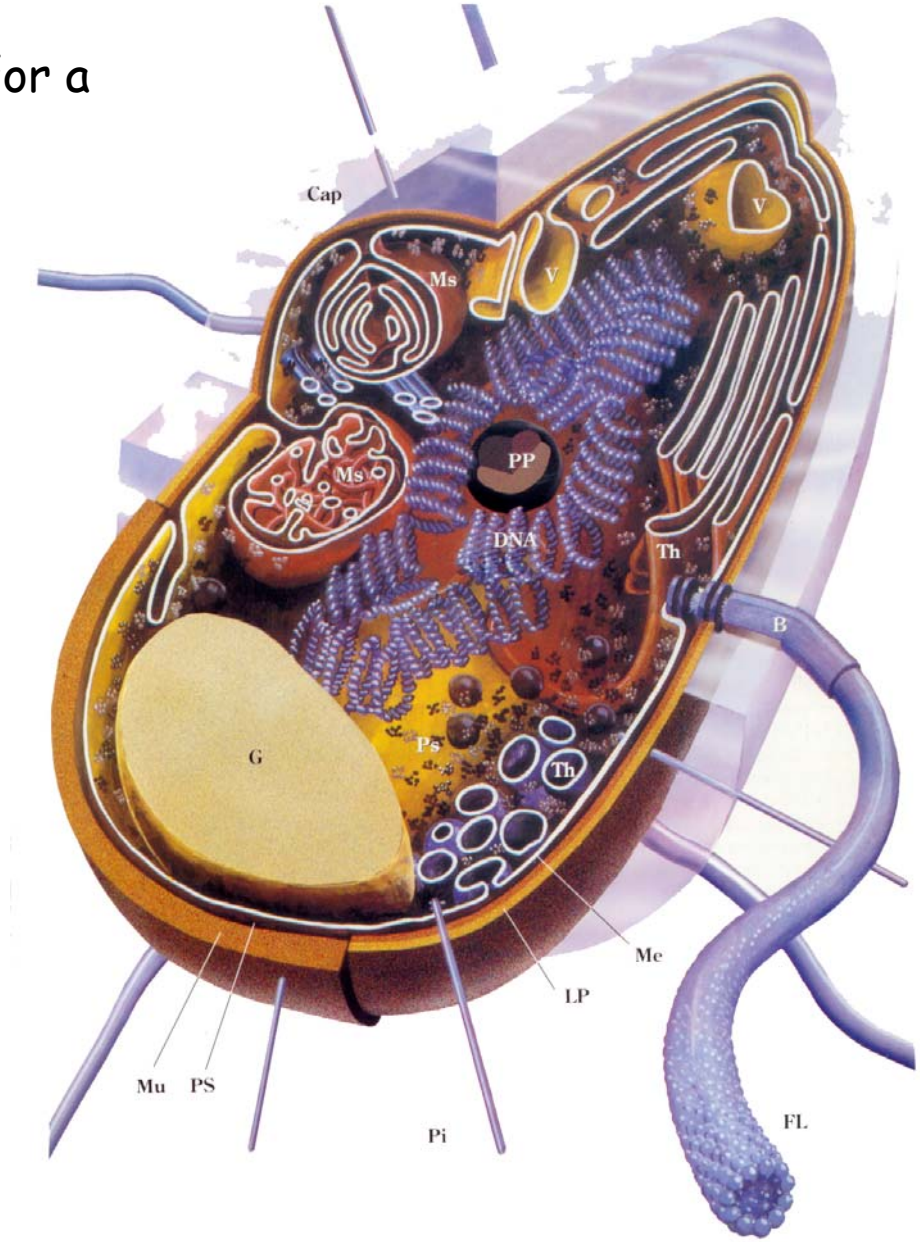
1. Pattern formation in physics and chemistry
2. Pattern formation in biology
3. Darwins natural selection and the origin of life
4. Molecular biology and evolution
5. Evolution in the test tube
6. **Complexity in biology**

The bacterial cell as an example for a simple form of autonomous life

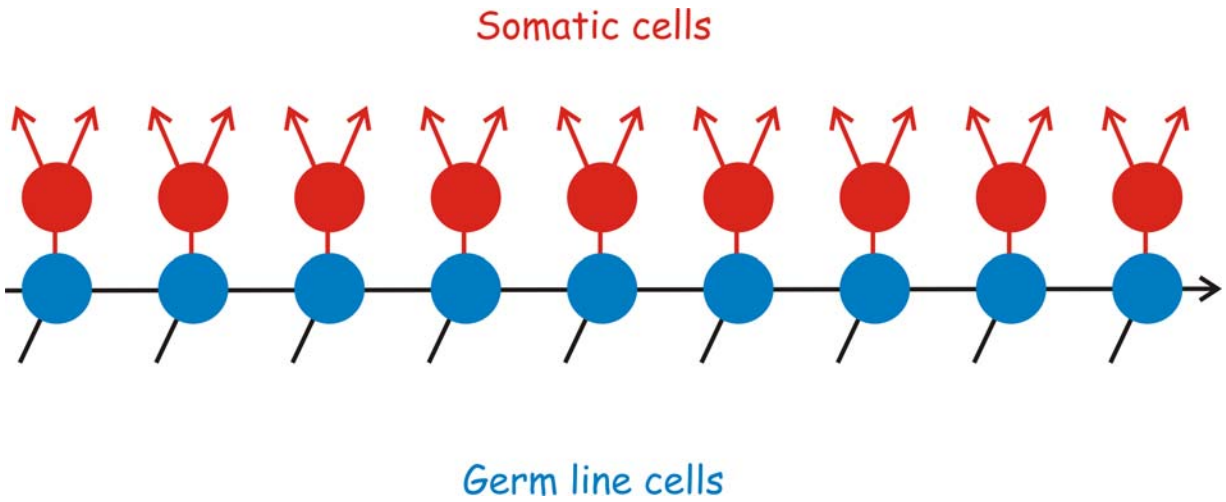
Escherichia coli genome:

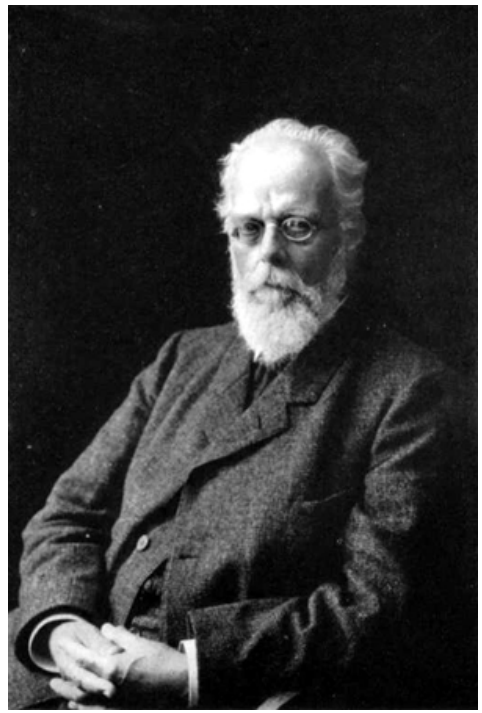
4 million nucleotides

4460 genes



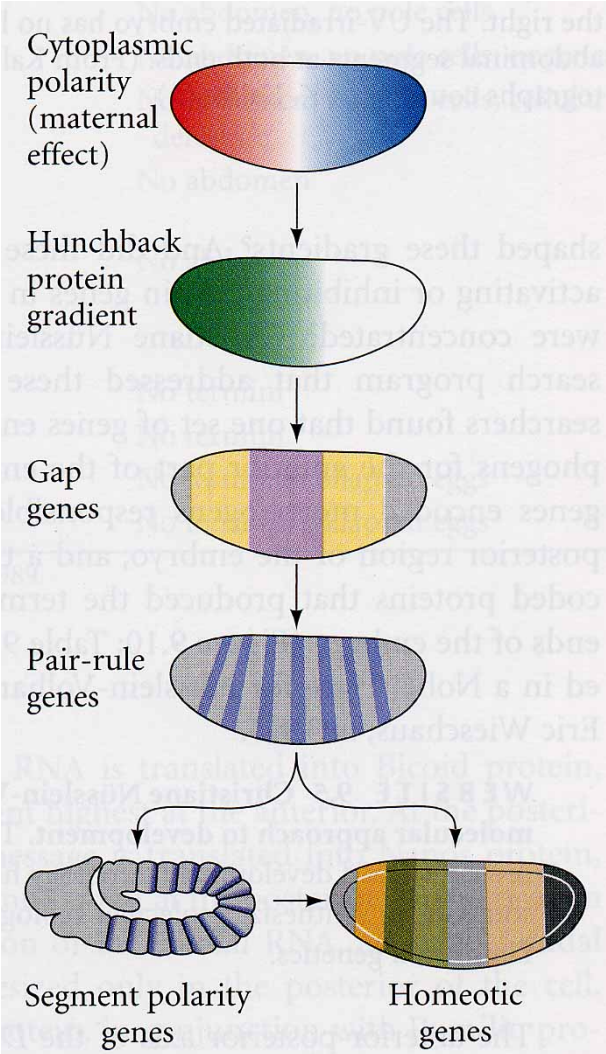
The structure of the bacterium *Escherichia coli*

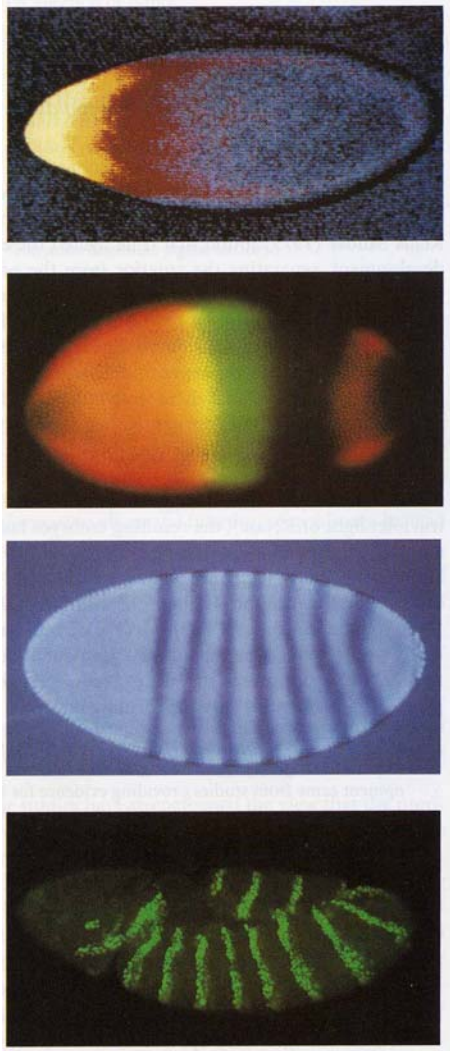


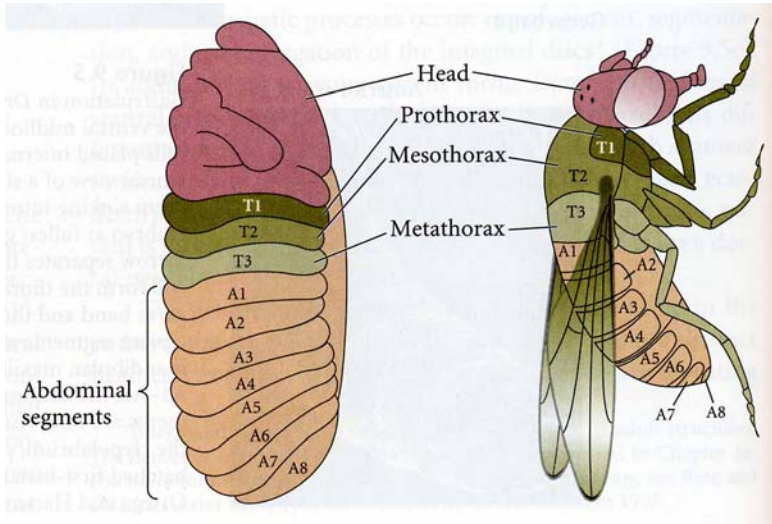


August Weismann, 1834-1914

Separation of germ line and soma







Cascades, A ⇒ B ⇒ C ⇒ ... , and networks of genetic control

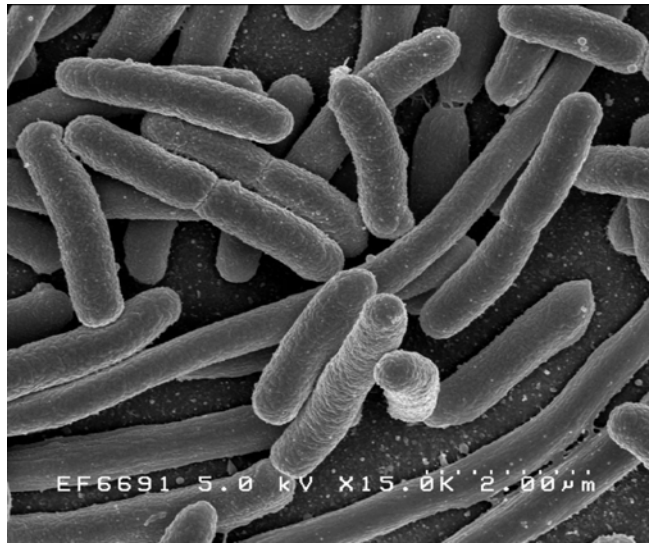
Turing pattern resulting from reaction-diffusion equation ?

Intercelluar communication creating positional information

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

E. coli:	Genome length	4×10^6 nucleotides
	Number of cell types	1
	Number of genes	4 460

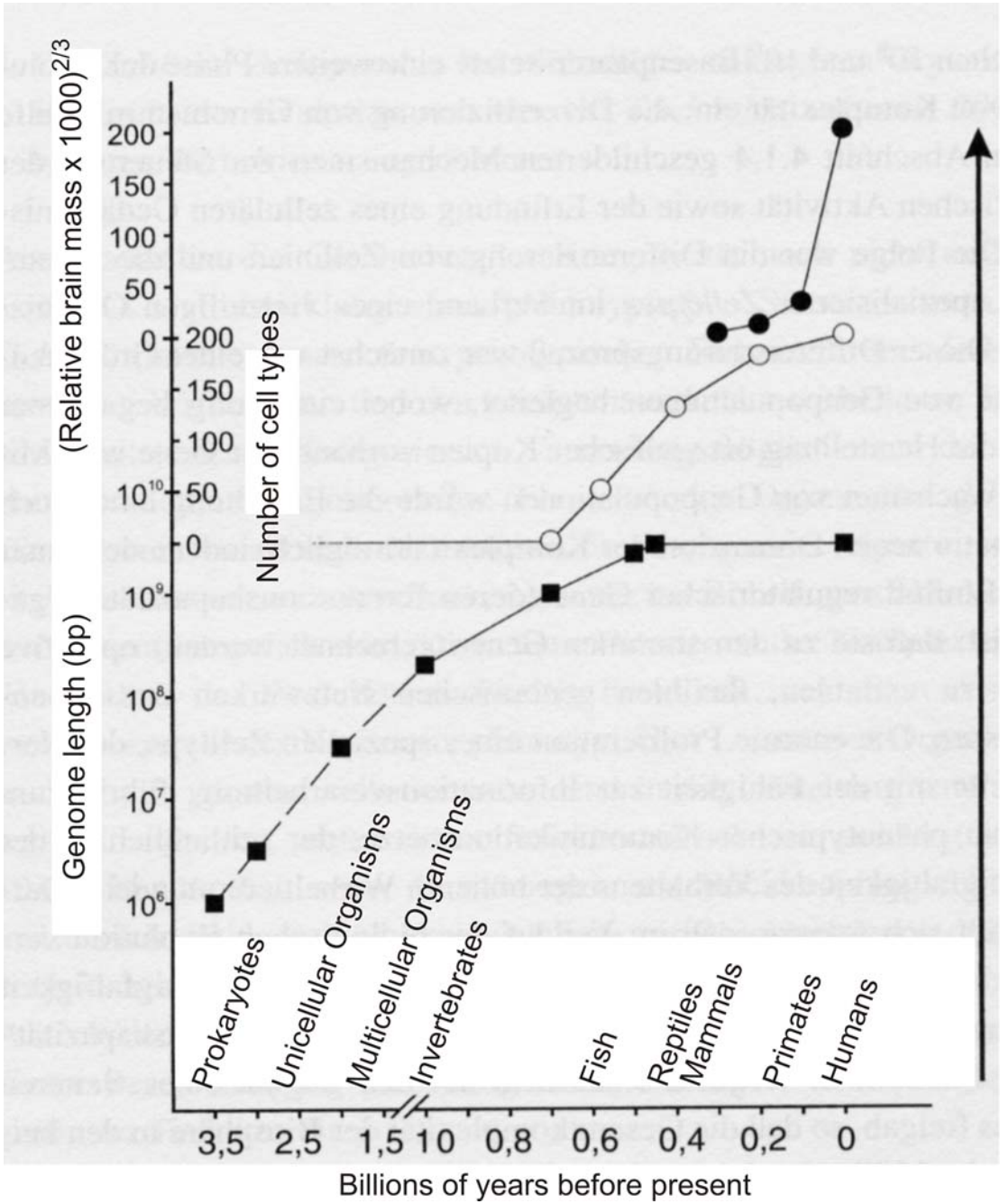
Four books, 300 pages each



Man:	Genome length	3×10^9 nucleotides
	Number of cell types	200
	Number of genes	≈ 30 000

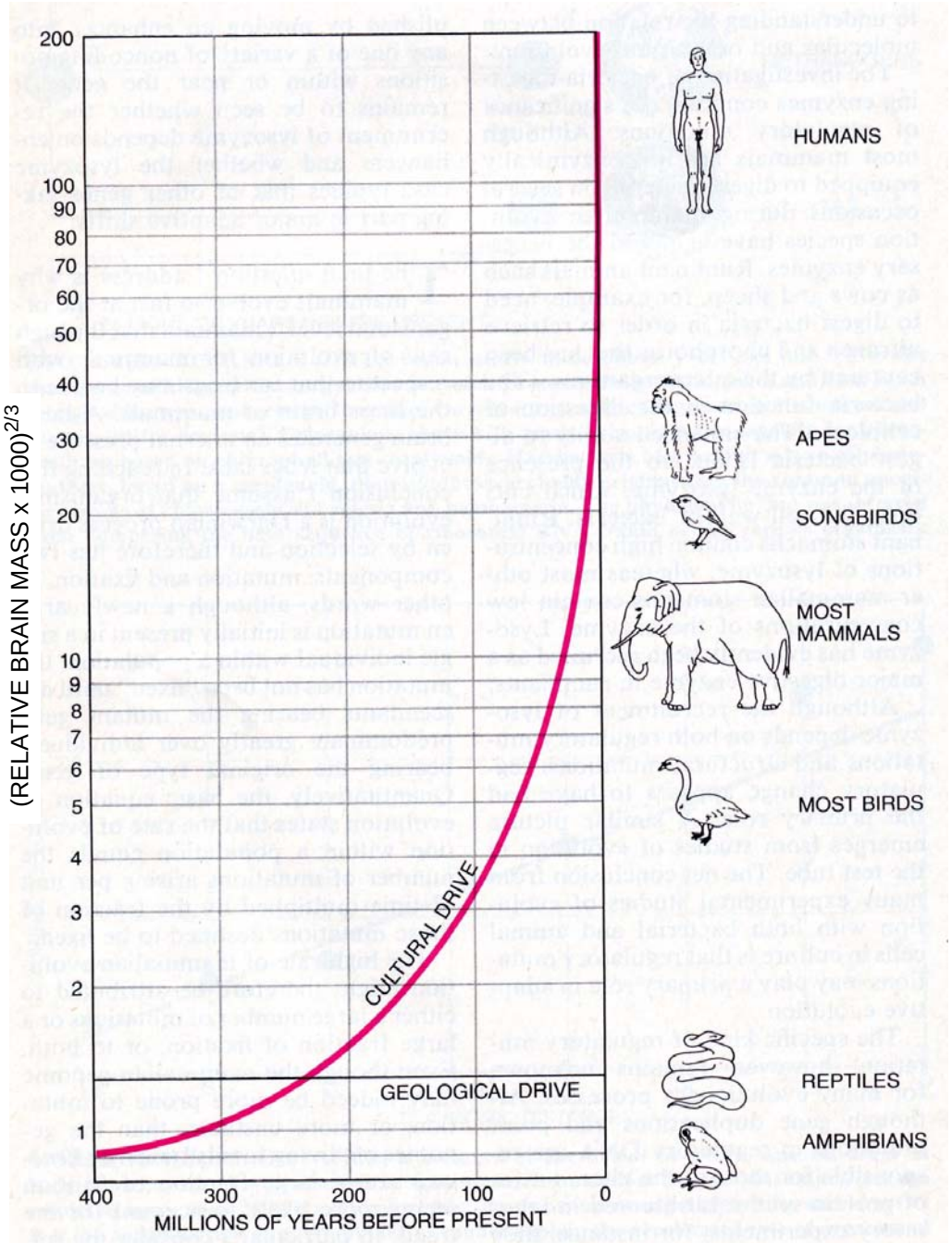
A library of 3000 volumes,
300 pages each

Complexity in biology

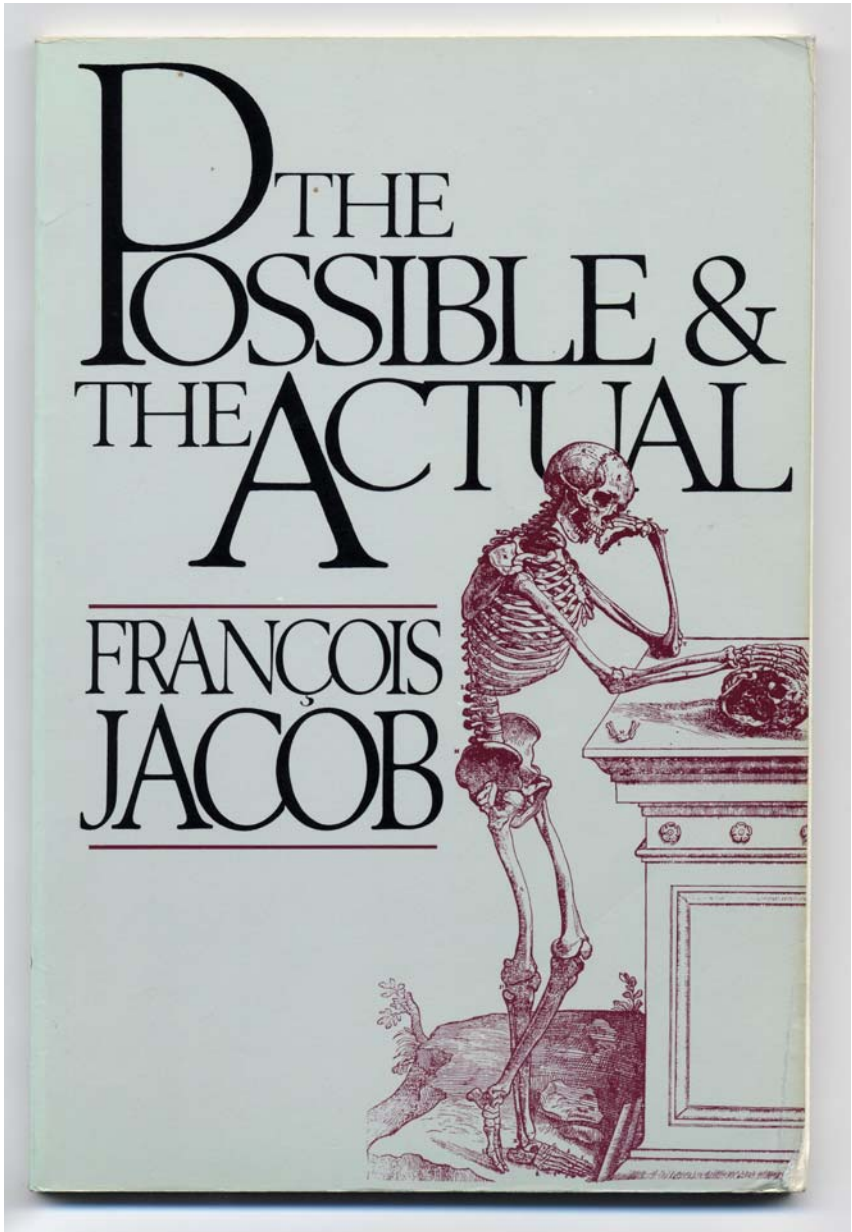


Wolfgang Wieser. 1998. ‚*Die Erfindung der Individualität*‘ oder ‚*Die zwei Gesichter der Evolution*‘. Spektrum Akademischer Verlag, Heidelberg 1998

BRITISH TIT



Alan C. Wilson.1985. The molecular basis of evolution.
Scientific American **253**(4):148-157.



Evolution does not design with the eyes of an engineer, evolution works like a tinkerer.

François Jacob. *The Possible and the Actual*. Pantheon Books, New York, 1982, and

Evolutionary tinkering. *Science* **196** (1977), 1161-1166.

The difficulty to define the notion of „gene".

Helen Pearson,
Nature **441**: 399-401, 2006

WHAT IS A GENE?

The idea of genes as beads on a DNA string is fast fading. Protein-coding sequences have no clear beginning or end and RNA is a key part of the information package, reports **Helen Pearson**.

'Gene' is not a typical four-letter word. It is not offensive. It is never bleeped out of TV shows. And where the meaning of most four-letter words is all too clear, that of gene is not. The more expert scientists become in molecular genetics, the less easy it is to be sure about what, if anything, a gene actually is.

Rick Young, a geneticist at the Whitehead Institute in Cambridge, Massachusetts, says that when he first started teaching as a young professor two decades ago, it took him about two hours to teach fresh-faced undergraduates what a gene was and the nuts and bolts of how it worked. Today, he and his colleagues need three months of lectures to convey the concept of the gene, and that's not because the students are any less bright. "It takes a whole semester to teach this stuff to talented graduates," Young says. "It used to be we could give a one-off definition and now it's much more complicated."

In classical genetics, a gene was an abstract concept — a unit of inheritance that ferried a characteristic from parent to child. As biochemistry came into its own, those characteristics were associated with enzymes or proteins, one for each gene. And with the advent of molecular biology, genes became real, physical things — sequences of DNA which when converted into strands of so-called messenger RNA could be used as the basis for building their associated protein piece by piece. The great coiled DNA molecules of the chromosomes were seen as long strings on which gene sequences sat like discrete beads.

This picture is still the working model for many scientists. But those at the forefront of genetic research see it as increasingly old-fashioned — a crude approximation that, at best, hides fascinating new complexities and, at worst, blinds its users to useful new paths of enquiry.

Information, it seems, is parceled out along chromosomes in a much more complex way than was originally supposed. RNA molecules are not just passive conduits through which the gene's message flows into the world but active regulators of cellular processes. In some cases, RNA may even pass information across generations — normally the sole preserve of DNA.

An eye-opening study last year raised the possibility that plants sometimes rewrite their DNA on the basis of RNA messages inherited from generations past[1]. A study on page 469 of this issue suggests that a comparable phenomenon might occur in mice, and by implication in other mammals[2]. If this type of phenomenon is indeed widespread, it "would have huge implications," says evolutionary geneticist Laurence Hurst at the University of Bath, UK.

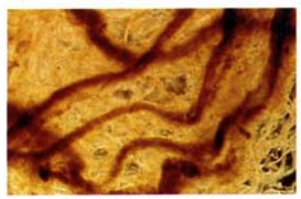
"All of that information seriously challenges our conventional definition of a gene," says molecular biologist Bing Ren at the University of California, San Diego. And the information challenge is about to get even tougher. Later this year, a glut of data will be released from the international Encyclopedia of DNA Elements (ENCODE) project. The pilot phase of ENCODE involves scrutinizing roughly 1% of the human genome in unprecedented detail; the aim is to find all the sequences that serve a useful purpose and explain what that purpose is. "When we started the ENCODE project I had a different view of what a gene was," says contributing researcher Roderic Guigo at the Center for Genomic Regulation in Barcelona. "The degree of complexity we've seen was not anticipated."

"We've come to the realization that the genome is full of overlapping transcripts."
— Phillip Kapranov

Under fire

The first of the complexities to challenge molecular biology's paradigm of a single DNA sequence encoding a single protein was alternative splicing, discovered in viruses in 1977 (see 'Hard to track', overleaf). Most of the DNA sequences describing proteins in humans have a modular arrangement in which exons, which carry the instructions for making proteins, are interspersed with non-coding introns. In alternative splicing, the cell snips out introns and sews together the exons in various different orders, creating messages that can code for different proteins. Over the years geneticists have also documented overlapping genes, genes within genes and countless other weird arrangements (see 'Muddling over genes', overleaf).

Alternative splicing, however, did not in itself require a drastic reappraisal of the notion of a gene; it just showed that some DNA sequences could describe more than one protein. Today's assault on the gene concept is more far reaching, fuelled largely by studies that show the previously unimagined scope of RNA.



Spools of DNA (above) still harbour surprises, with one protein-coding gene often overlapping the next.

P. PLAILLY/SPL

The one gene, one protein idea is coming under particular assault from researchers who are comprehensively extracting and analysing the RNA messages, or transcripts, manufactured by genomes, including the human and mouse genome. Researchers led by Thomas Gingeras at the company Affymetrix in Santa Clara, California, for example, recently studied all the transcripts from ten chromosomes across eight human cell lines and worked out precisely where on the chromosomes each of the transcripts came from[3].

The picture these studies paint is one of mind-boggling complexity. Instead of discrete genes dutifully mass-producing identical RNA transcripts, a teeming mass of transcription converts many segments of the genome into multiple RNA ribbons of differing lengths. These ribbons can be generated from both strands of DNA, rather than from just one as was conventionally thought. Some of these transcripts come from regions of DNA previously identified as holding protein-coding genes. But many do not. "It's somewhat revolutionary," says Gingeras's colleague Phillip Kapranov. "We've come to the realization that the genome is full of overlapping transcripts."

Other studies, one by Guigo's team[4], and one by geneticist Rotem Sorek[5], now at Tel Aviv University, Israel, and his colleagues, have hinted at the reasons behind the mass of transcription. The two teams investigated occasional reports that transcription can start at a DNA sequence associated with one protein and run straight through into the gene for a completely different protein, producing a fused transcript. By delving into databases of human RNA transcripts, Guigo's team estimate that 4–5% of the DNA in regions conventionally recognized as genes is transcribed in this way. Producing fused transcripts could be one way for a cell to generate a greater variety of proteins from a limited number of exons, the researchers say.

Many scientists are now starting to think that the descriptions of proteins encoded in DNA know no borders — that each sequence reaches into the next and beyond. This idea will be one of the central points to emerge from the ENCODE project when its results are published later this year.

Kapranov and others say that they have documented many examples of transcripts in which protein-coding exons from one part of the genome combine with exons from another

ENCODE stands for
ENCyclopedia Of DNA Elements.

ENCODE Project Consortium.
Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.
Nature **447**:799-816, 2007

Web-Page for further information:

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