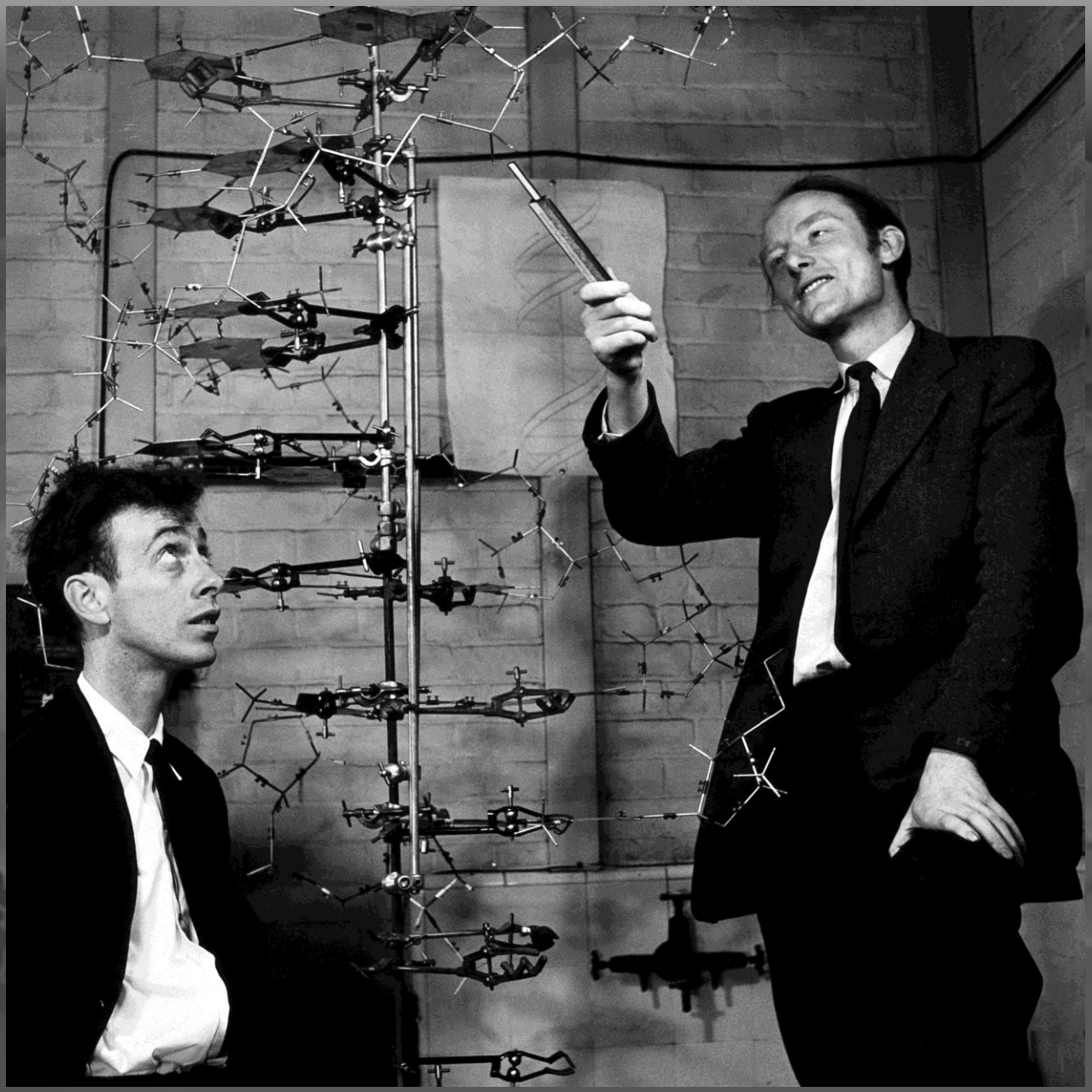
LATVIJAS UNIVERSITĀTE
UNIVERSITY OF LATVIA
ANNO 1919

DNS kā informācijas nesējs

Dr. biol. Nils Rostoks
LU Bioloģijas fakultāte

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111110011011000001



ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111110011011000001

Leading Edge
BenchMarks

Cell

1953: When Genes Became "Information"

Matthew Cobb[1,*]
[1]Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK
*Correspondence: cobb@manchester.ac.uk
http://dx.doi.org/10.1016/j.cell.2013.04.012

In 1953, Watson and Crick not only described the double-helix structure of DNA, but also embraced the idea that genes contained a code that expresses information and thereby changed our view of life. This article traces how these ideas entered biological thinking and highlights the connections between different branches of science at the time, exploring the power of metaphor in science.

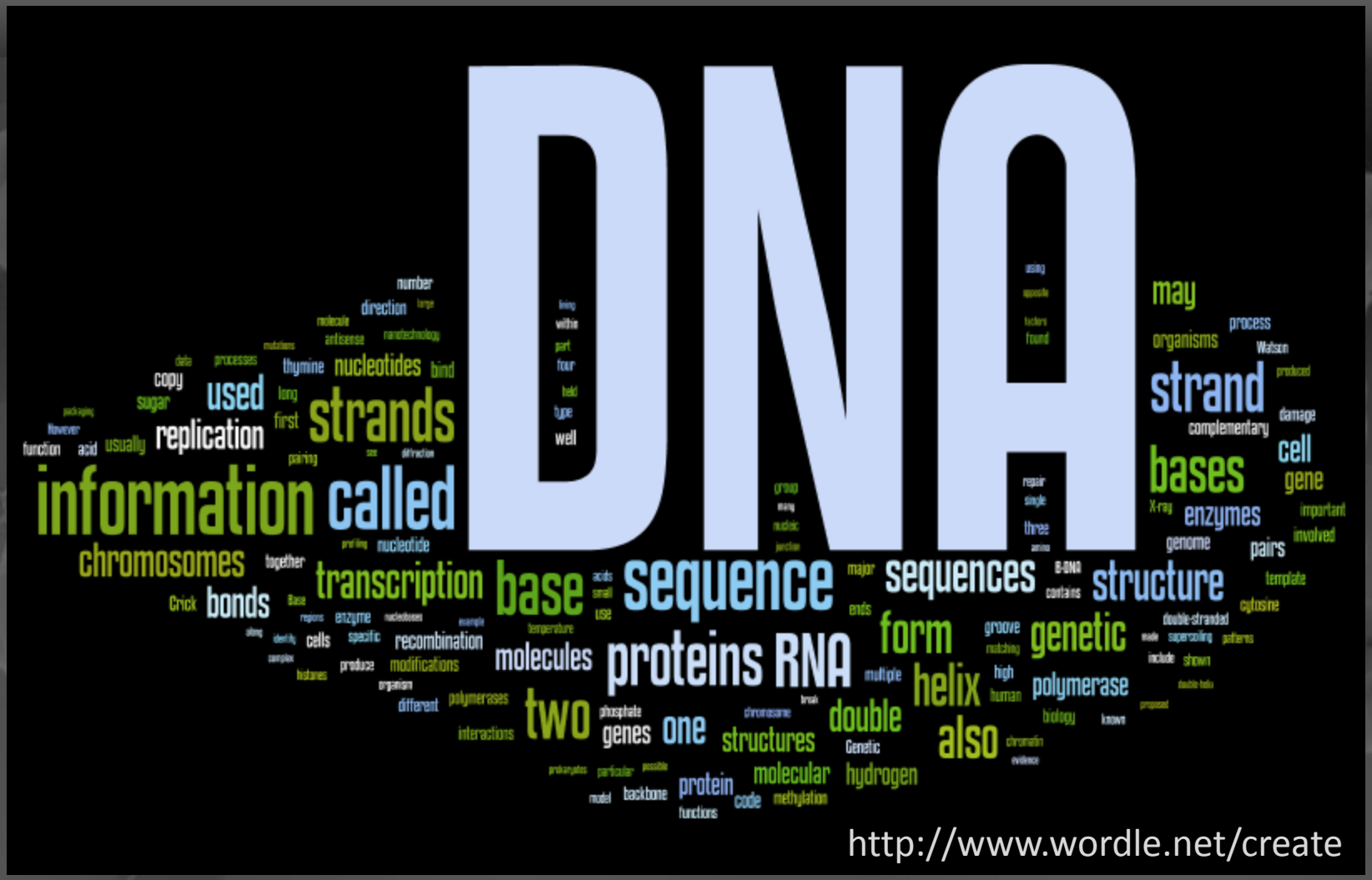
Introduction

Sixty years ago, James Watson and Francis Crick described the double-helix structure of DNA. The double helix famously led them to state that "it has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." However, replication is merely one aspect of the hereditary material is DNA—neither of these ideas existed as biological concepts. The way these fundamental ideas entered biology in the 1940s reveals surprising interconnections between different aspects of science at the time.

Schrödinger's Code

In February 1943, the Nobel-Prize-winning physicist Erwin Schrödinger has been the subject of argument among participants and historians (e.g., Kay, 2000; Pauling, 1987; Yoxen, 1979). Whatever the case, the book did inspire many of those who came to focus their lives on the structure of DNA and of genes—James Watson, Francis Crick, Maurice Wilkins, and Seymour Benzer, among others. Above all, Schrödinger was the first person in the 20th century to explicitly

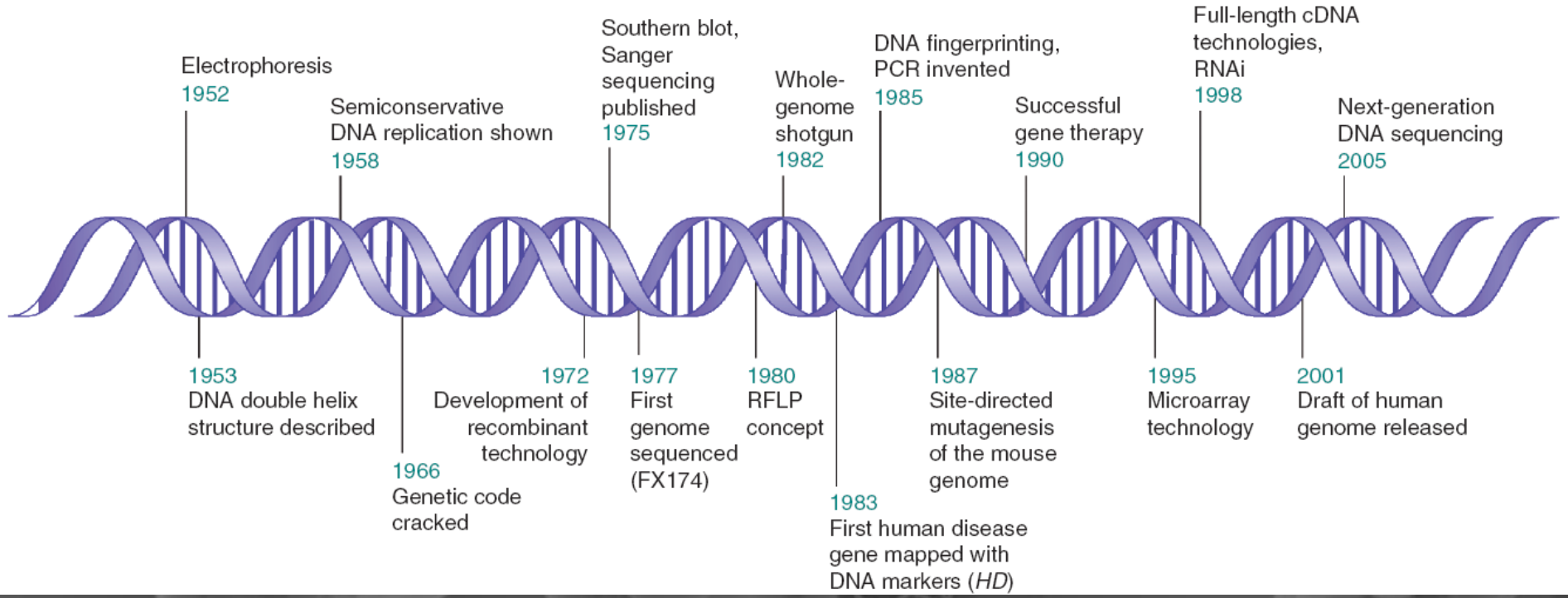
ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111110011011000001



http://www.wordle.net/create

https://en.wikipedia.org/wiki/DNA

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111110011011000001



Betz et al. (2012) JID 132: 906

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111110011011000001

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

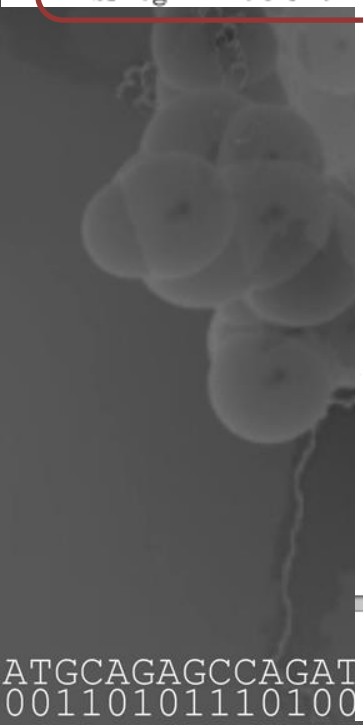
A Structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position



BASIS OF TECHNICAL EDUCATION

GENERAL education to-day should be planned so as to enable the ordinary citizen to adapt himself to the needs of technological society and to understand what is happening and what is required of him. This was the theme of an international conference convened by the United Nations Educational, Cultural and Scientific Organization at Unesco House in June 1950*.

Broadly, the Conference found that organized social foresight is essential to enable the educational system of a country to prepare children for the type of life and work they are likely to encounter, and that a substantial development of technical education

* Education in a Technological Society: a Preliminary International Survey of the Nature and Efficacy of Technical Education. (Tensions and Technology Series.) Pp. 76. (Paris: Unesco; London: H.M.S.O., 1952.) 200 francs; 4*s.*; 75 cents.

GENETICAL IMPLICATIONS OF THE STRUCTURE OF DEOXYRIBONUCLEIC ACID

By J. D. WATSON and F. H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge

THE importance of deoxyribonucleic acid (DNA) within living cells is undisputed. It is found in all dividing cells, largely if not entirely in the nucleus, where it is an essential constituent of the chromosomes. Many lines of evidence indicate that it is the carrier of a part of (if not all) the genetic specificity of the chromosomes and thus of the gene itself.

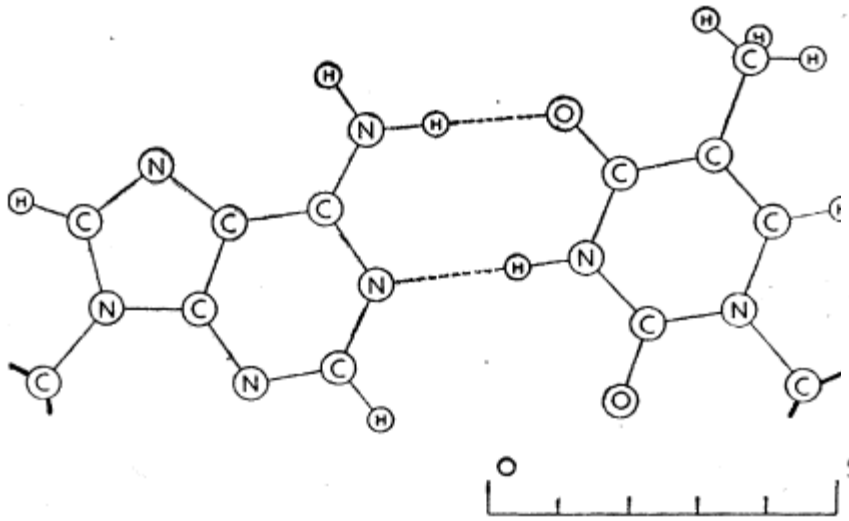
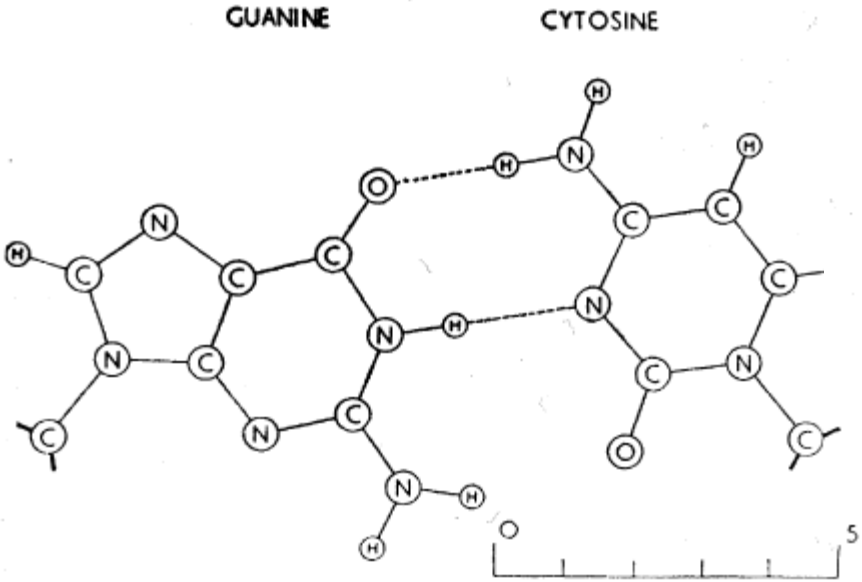


Fig. 4. Pairing of adenine and thymine. Hydrogen bonds are shown dotted. One carbon atom of each sugar is shown



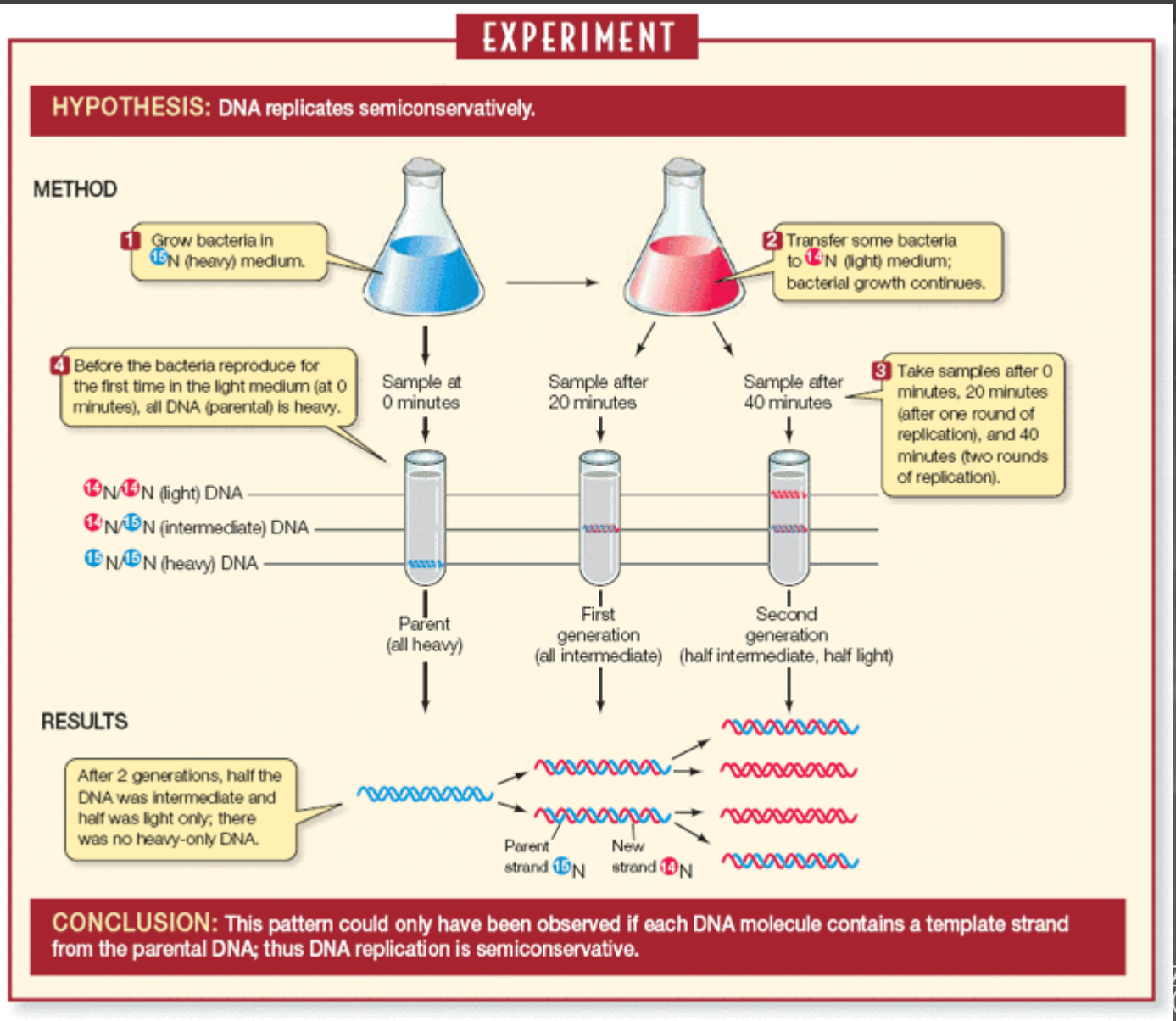
The phosphate-sugar backbone of our model is completely regular, but any sequence of the pairs of bases can fit into the structure. It follows that in a long molecule many different permutations are possible, and it therefore seems likely that the precise sequence of the bases is the code which carries the genetical information. If the actual order of the

molecules.

Now our model for deoxyribonucleic acid is, in effect, a *pair* of templates, each of which is complementary to the other. We imagine that prior to duplication the hydrogen bonds are broken, and the two chains unwind and separate. Each chain then acts as a template for the formation on to itself of a new companion chain, so that eventually we shall have *two* pairs of chains, where we only had one before. Moreover, the sequence of the pairs of bases will have been duplicated exactly.

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011100111101011111111001101100001

Meselson un *Stahl* eksperiments



«*RNA Tie*» klubs

- Kā DNS var kodēt proteīnus un kā tur iesaistīta RNS?
- Zinātnieku džentlmeņu klubs (20 biedri = 20 aminoskābes) – izveidots pēc *James Watson* (Pro) un *Georg Gamov* (Ala) idejas

Ģenētiskais kods

- Kā 4 nukleotīdi var kodēt 20 aminoskābes?
- Minimālā pietiekamā kombinācija ir triplets (4^3=64)
- Nobela prēmija 1968

Marshall Nirenberg izmanto RNS homopolimēru poly U, lai *in vitro* sintezētu peptīdu, kas sastāv no Phe

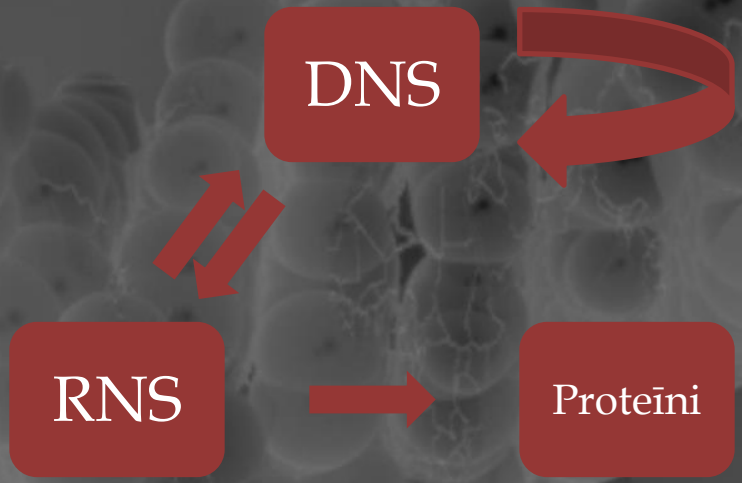
Har Gobind Khorana sintezē dažādas secības RNS un atšifrē pārējos tripletus

Robert Holley atklāj tRNS, kas nodrošina katram tripletam atbilstošu aminoskābi

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111100110110000 1

Replikācija, transkripcija un translācija

Molekulārās bioloģijas centrālā dogma

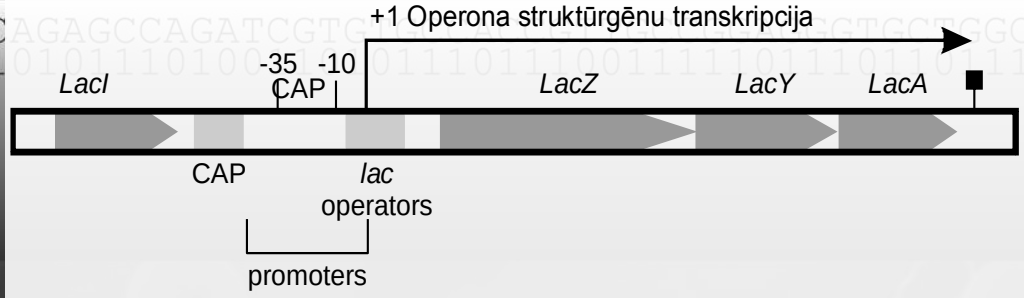


ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111100110110000 1

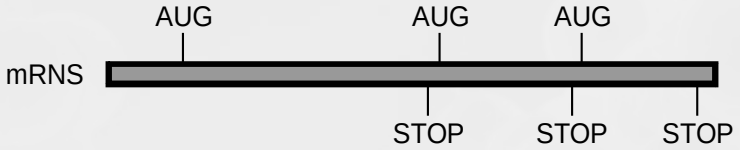
Gēnu ekspresijas regulācija

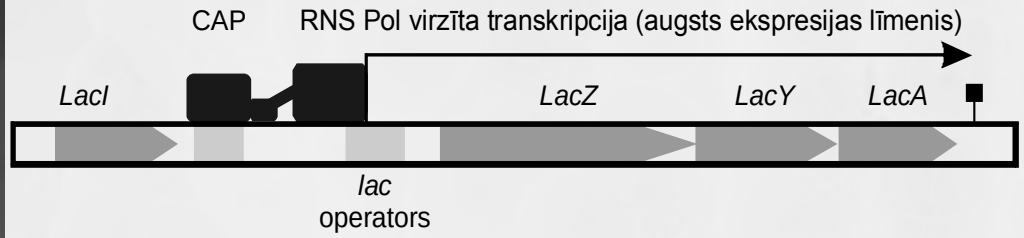
- Vai gēnu transkripcija un translācija tiek regulēta?
- Gēnu ekspresija tiek regulēta organisma ontoģenēzes laikā, kā arī atbildot uz apkārtējās vides signāliem
- Kā notiek gēnu ekspresijas regulācija?
- Nobela prēmija 1965 par *Escherichia coli lac* katabolisma operona regulācijas atklāšanu (*Jacques Monod*, *Francois Jacob*, *Andre Lwoff*)

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111100110110000 1

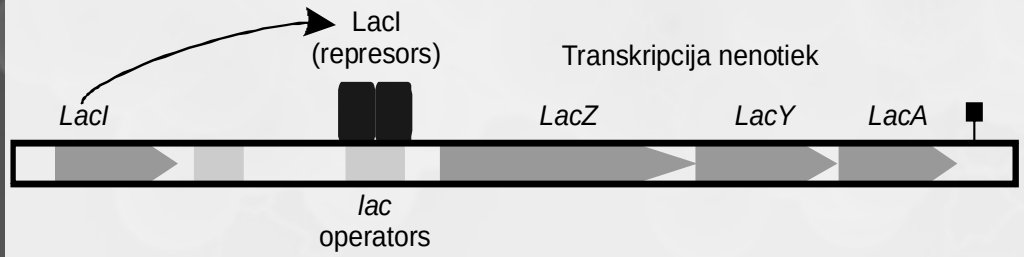


Laktozes operona shēma. Operonu veido 4 gēni. *LacI* gēns tiek transkribēts no sava promotera. *LacZ*, *LacY* un *LacA* gēni tiek transkribēti no viena promotera un proteīni tiek sintezēti no vienas mRNS. Operons kodē laktozes transportam un katabolismam nepieciešamos proteīnus. Optimālos augšanas apstākļos, ja vidē ir glikoze, laktozes operons ir izslēgts. Ja glikozes daudzums vidē samazinās, bet pieejama laktoze, operons tiek ieslēgts un šūnas sāk uzņemt un pārstrādāt laktozi.

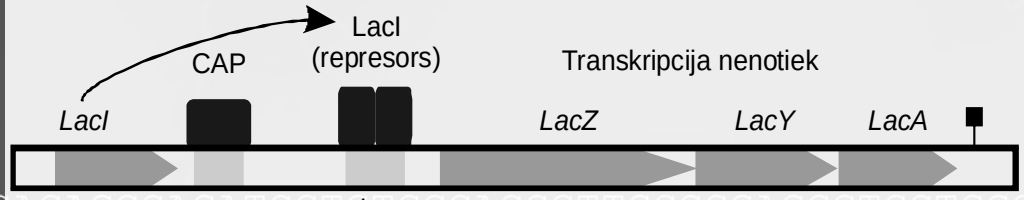




Zema glikozes koncentrācija vidē. Pieejama laktoze.



Augsta glikozes koncentrācija. Laktoze nav pieejama.



Zema glikozes koncentrācija. Laktoze nav pieejama.

Kā noteikt DNS kodēto informāciju?

- Gēni ir noteiktas nukleotīdu sekvences, kuras kodē noteiktas aminoskābju virknes, vai arī veido gēnu ekspresiju regulējošus elementus
- DNS sekvenēšanas un rekombinantās DNS metodes izstrādā 70-tajos gados (Nobela prēmija *Walter Gilbert*, *Fred Sanger*, *Paul Berg*)
- Populārākā ir didezoksiterminatoru sekvenēšanas metode, kur izmanto vēl šodien
- Nākamās paaudzes sekvenēšanas tehnoloģijas

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111100110110000 1

DNS sekvences un datori

Volume 4 Number 11 November 1977 Nucleic Acids Research

Sequence data handling by computer

R.Staden

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

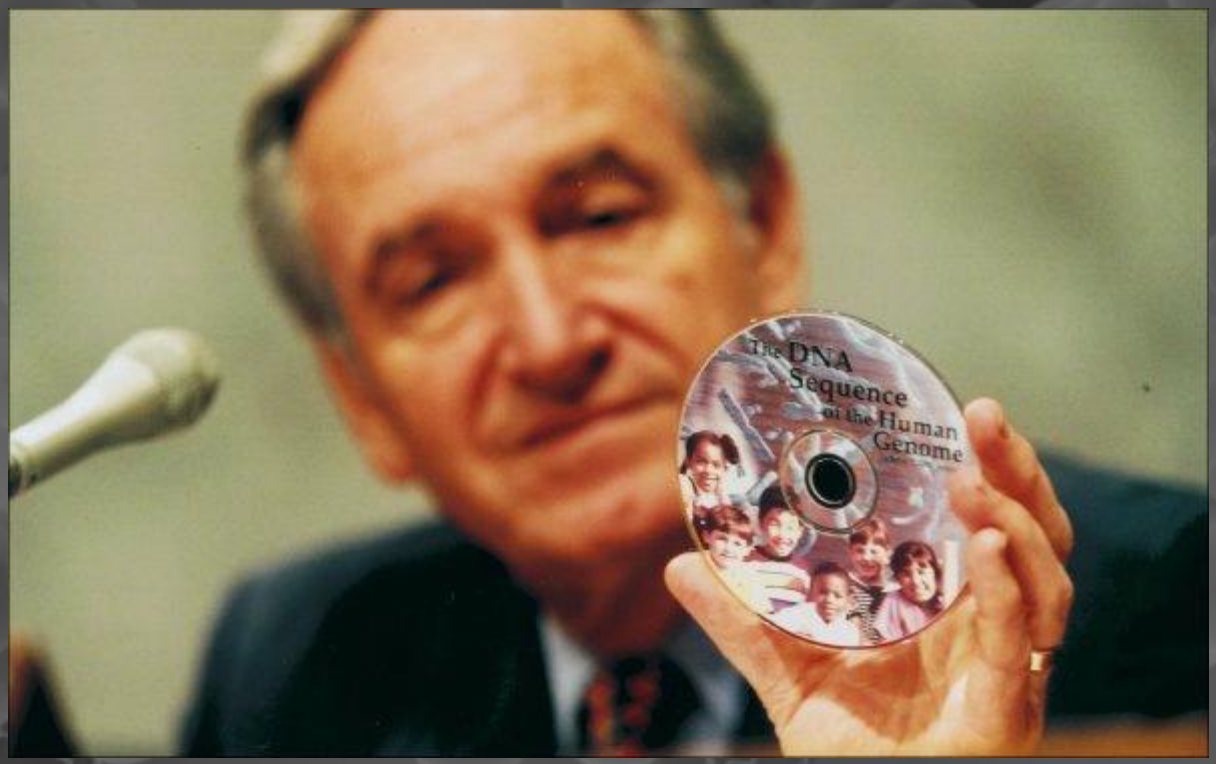
Received 10 October 1977

ABSTRACT

The speed of the new DNA sequencing techniques has created a need for computer programs to handle the data produced. This paper describes simple programs designed specifically for use by people with little or no computer experience. The programs are for use on small computers and provide facilities for storage, editing and analysis of both DNA and amino acid sequences. A magnetic tape containing these programs is available on request.

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111100110110000 1

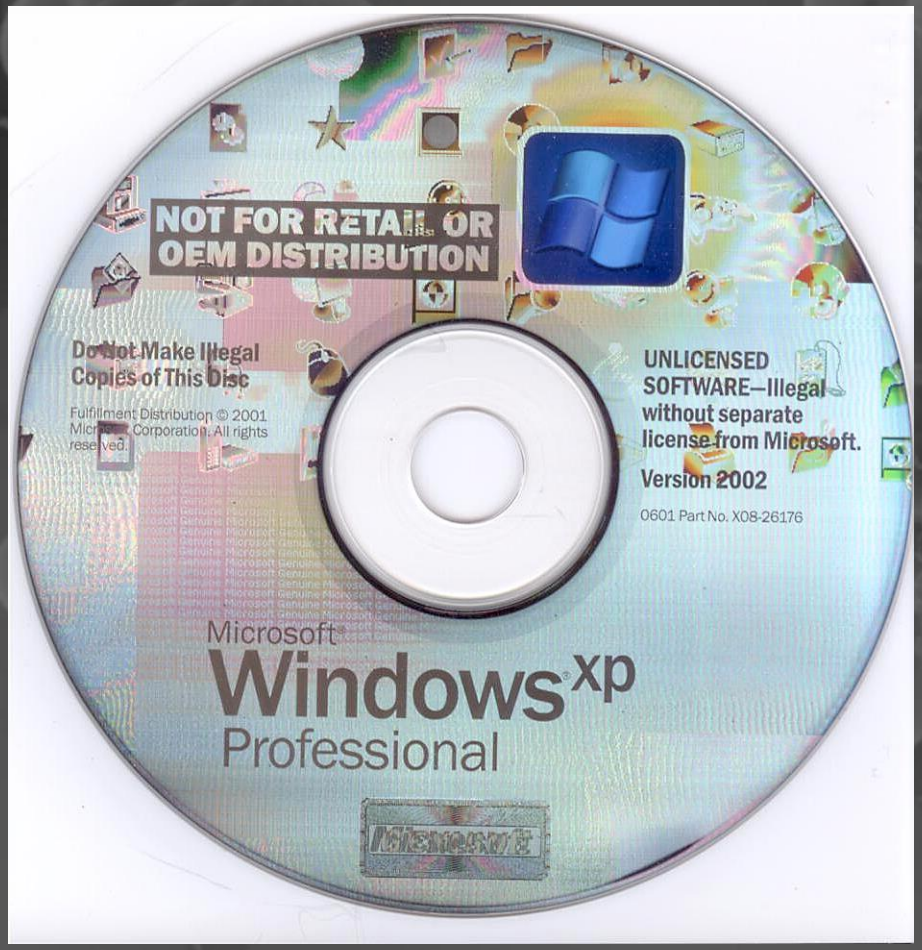
Cilvēka genoma sekvence

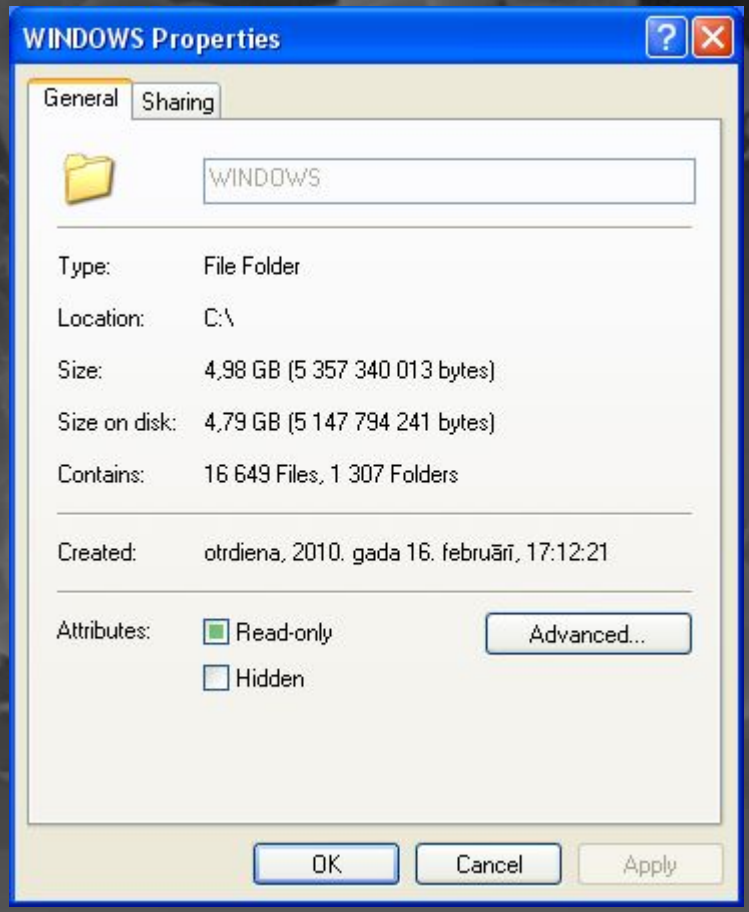


http://www.genome.gov/dmd/img.cfm?node=Photos/People/People%20Outside%20NHGRI&id=79269

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111100110110000 1

Cilvēka genoms un MS Windows operētājsistēma?





ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111100110110000

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111100110110000 1

Cilvēka genoms un MS Windows operētājsistēma?

- Diploīds genoms – apmēram 6 x 10^9 nukleotīdu
- Viena nukleotīda kodēšanai binārā formā nepieciešami vismaz 2 biti:

  A = 00, C = 01, G = 10, T = 11
- 8 biti = 1 baits
- 1 baits var kodēt 4 nukleotīdus
- Diploīds cilvēka genoms binārā formā ir 1.5 x 10^9 baiti (datora darbībai vajag vairāk koda nekā dzīvības nodrošināšanai)

http://www.tmsoft.com/article-genome.html

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111100110110000 1

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111100110110000 1

Genoms un IT

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111100110110000 1

Baktēriju šūna ar sintezētu genomu

Sciencexpress Research Article

Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome

Daniel G. Gibson,[1] John I. Glass,[1] Carole Lartigue,[1] Vladimir N. Noskov,[1] Ray-Yuan Chuang,[1] Mikkel A. Algire,[1] Gwynedd A. Benders,[2] Michael G. Montague,[1] Li Ma,[1] Monzia M. Moodie,[1] Chuck Merryman,[1] Sanjay Vashee,[1] Radha Krishnakumar,[1] Nacyra Assad-Garcia,[1] Cynthia Andrews-Pfannkoch,[1] Evgeniya A. Denisova,[1] Lei Young,[1] Zhi-Qing Qi,[1] Thomas H. Segall-Shapiro,[1] Christopher H. Calvey,[1] Prashanth P. Parmar,[1] Clyde A. Hutchison III,[2] Hamilton O. Smith,[2] J. Craig Venter[1,2*]

[1]The J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA. [2]The J. Craig Venter Institute, 10355 Science Center Drive, San Diego, CA 92121, USA.

[*]To whom correspondence should be addressed. E-mail: jcventer@jcvi.org

We report the design, synthesis and assembly of the 1.08-Mbp *Mycoplasma mycoides* JCVI-syn1.0 genome starting from digitized genome sequence information and its transplantation into a *Mycoplasma capricolum* recipient cell to create new *Mycoplasma mycoides* cells that are controlled only by the synthetic chromosome. The only DNA in the cells is the designed synthetic DNA sequence, including "watermark" sequences and other designed gene deletions and polymorphisms, and mutations

capable of independent growth in the laboratory. More than 100 of the 485 protein-coding genes of *M. genitalium* are dispensable when disrupted one-at-a-time (*4–6*).

We developed a strategy for assembling viral sized pieces to produce large DNA molecules that enabled us to assemble a synthetic *M. genitalium* genome in four stages from chemically synthesized DNA cassettes averaging about 6 kb in size. This was accomplished through a combination of in vitro enzymatic methods and in vivo recombination in

ATGCAGAGCCAGATCGTGTGCCACCGTTGCCGGAGGGTGCTGGCCTACCCGTCAGGGGCGCCCATGCTGGAATTC
0011010111010011010111011100111110111011011110011110101111111100110110000 1

Informācijas uzglabāšana DNS sekvencē

Sciencexpress

Brevia

Next-Generation Digital Information Storage in DNA

George M. Church,[1,2] Yuan Gao,[3] Sriram Kosuri[1,2]*

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [2]Wyss Institute for Biolo Inspired Engineering, Boston, MA 02115, USA. [3]Department of Biomedical Engineering, Johns Hopki University, Baltimore, MD 21205, USA.

*To whom correspondence should be addressed. E-mail: sri.kosuri@wyss.harvard.edu

As digital information continues to accumulate, higher density and long-er-term storage solutions are necessary (*1*). DNA has many potential advantages as a medium for immutable, high latency information storage needs (*2*). For example, DNA storage is very dense. At theoretical max-imum, DNA can encode two bits per nucleotide (nt) or 455 exabytes per gram of ssDNA (*3*). Unlike most digital storage media, DNA storage is not restricted to a planar layer, and is often readable despite degradation in non-ideal conditions over millennia (*4*, *5*). Finally, DNA's essential biological role provides access to natural reading and writing enzymes and ensures that DNA will remain a readable standard for the foreseea-ble future.

Storing messages in DNA was first demonstrated in 1988 (*6*) and the

for all but century-sca
sis and sequencing ha
fold per year, respect
fold per year (*11*). I
becoming available,
information (*12*). Our
combined with librar
compatible with futu
Reciprocally, large-sc
could accelerate deve
(*13*). Future work co

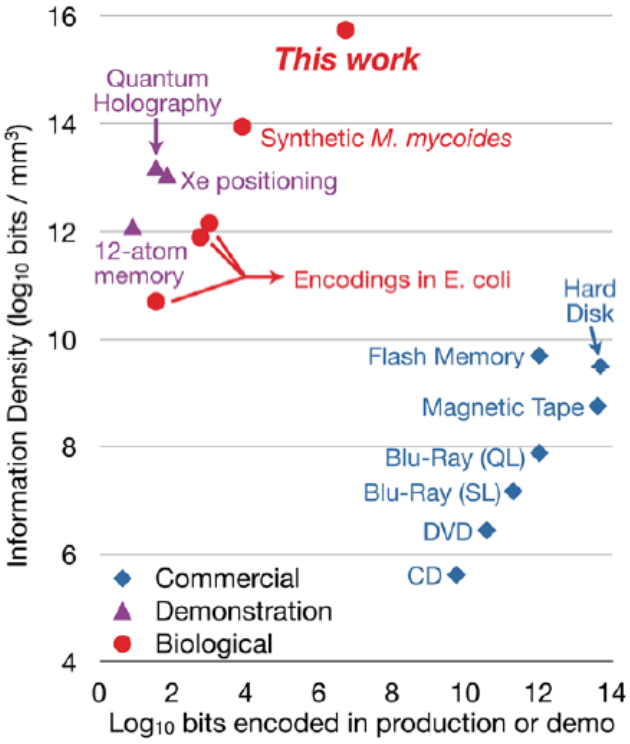


Fig. 1. Comparison to other measured by the $\log_{10}$ of bits encoded in the report or commercial technologies. We plotted information density ($\log_{10}$ of bits/mm^3) versus current scalability as unit (*3*).

Vai DNS ir vienīgā informācija?