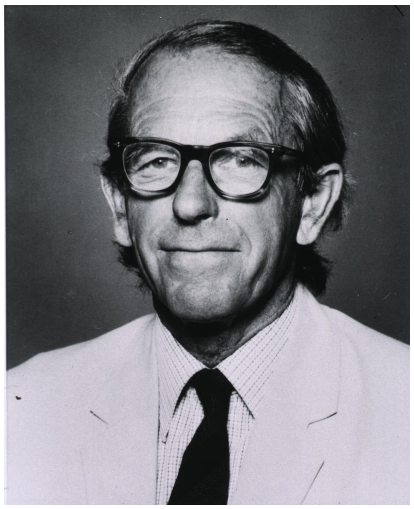
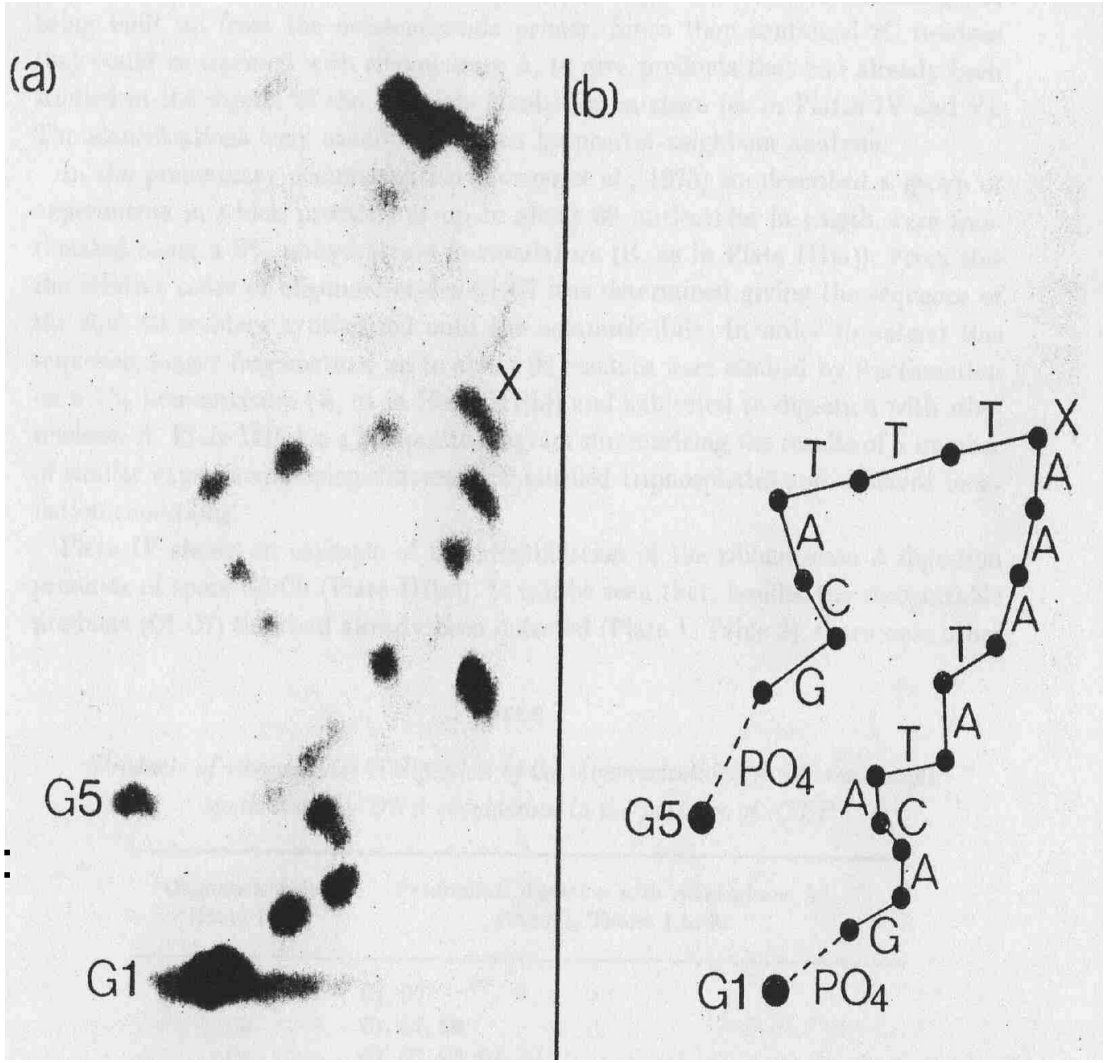
DNS szekvenálás





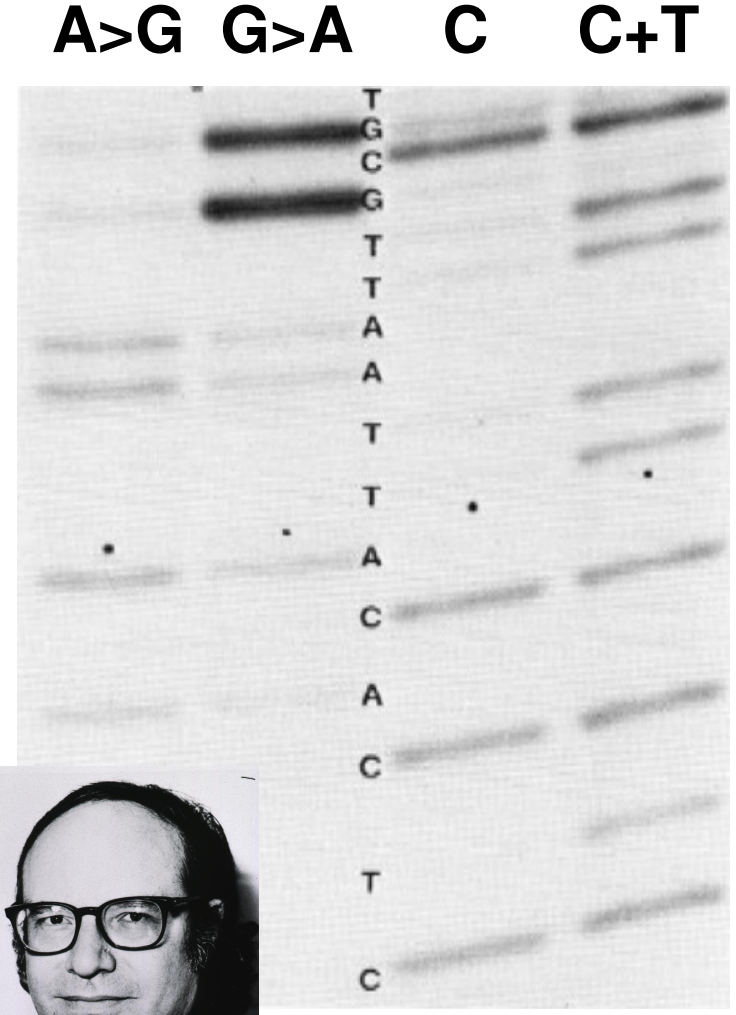
1. Sanger első módszere (+/- módszer?) (rNTP beépítés és kémiai/enzimes degradáció) 1974:
• Enzimes degradáció, papírkromatográfia majd elektroforézis merőlegesen jellegzetes mintázat, amiből a szekvencia körülményesen, de kikövetkeztethető (ujjlenyomat).

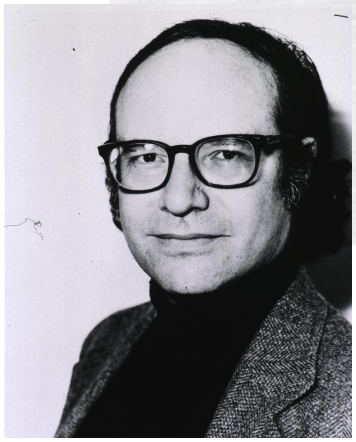
DNS szekvenálás

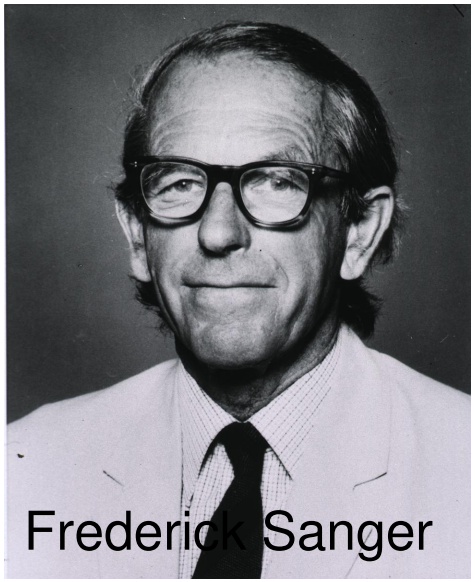
2. Maxam-Gilbert módszer (részleges kémiai degradáció) 1977.

• Különbségtétel a két-két bázis között:
-Metilezés dimetil szulfáttal - G 5x gyorsabban metileződik, mint A); Instabil glikozidkötés miatt a nukleobázis melegítésre lehasad semleges pH-n (0.5M HCl, 0°C szinte csak A depurinálódik.)
-Reakció hidrazinnal (C, T is reagál, de 2M NaCl-ban csak a C!); Nukleobázis elimináció + hidrazon képződés a cukorból.

-lúgos hidrolízis (0.1M NaOH v. 0.5M piperidin elhasítja a láncot a hiányzó nukleobázisnál.



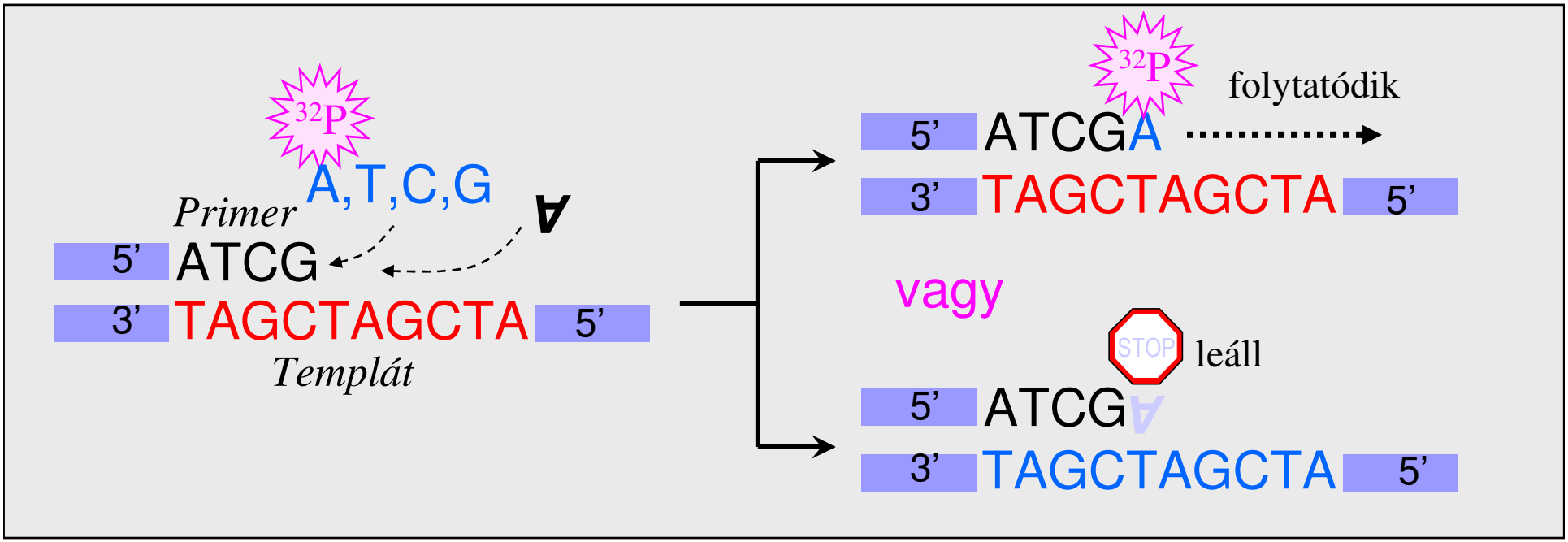




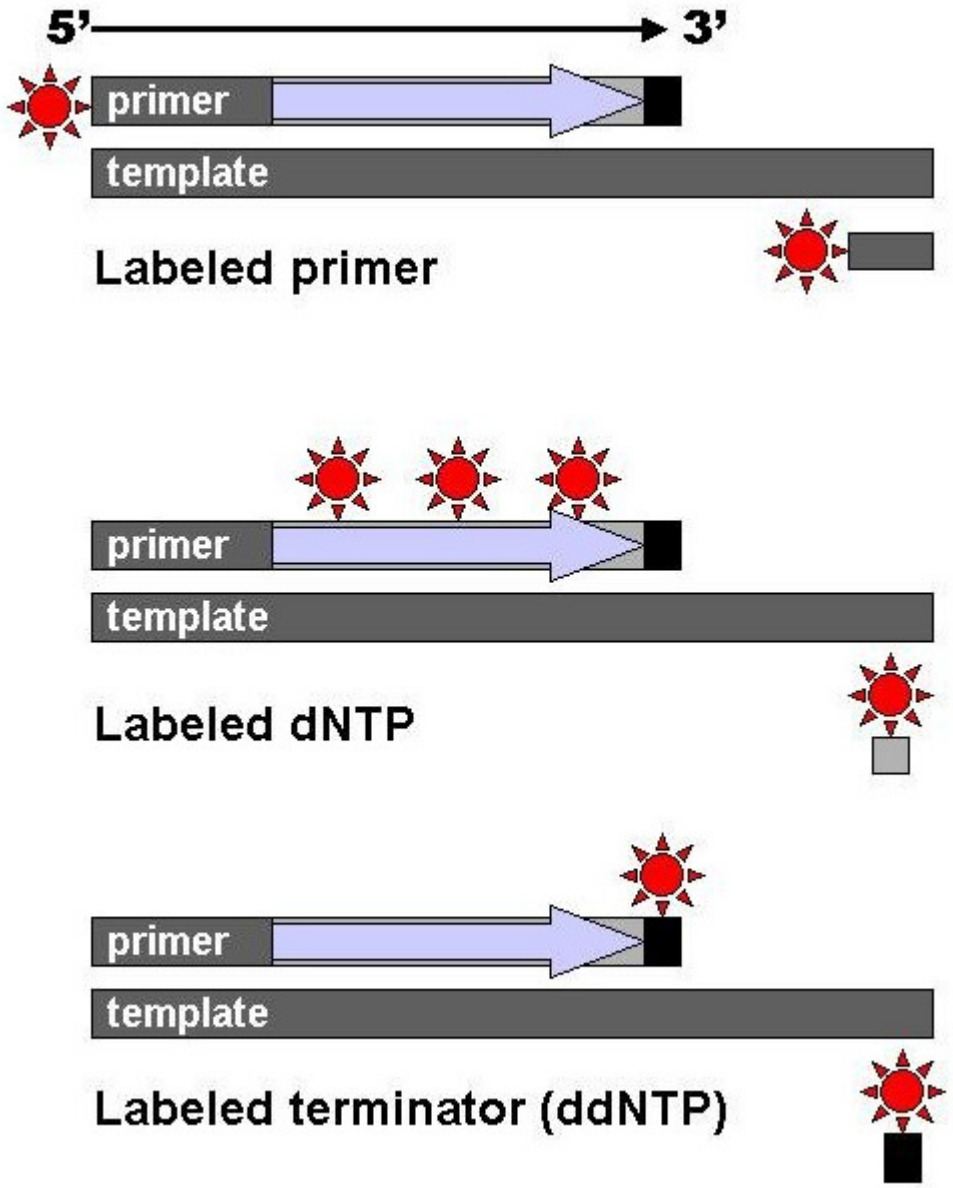
3. Sanger módszer (terminációs módszer) 1977.
Eredeti változat, radioaktív jelölés

^{32}P A Radioaktívan jelölt dATP

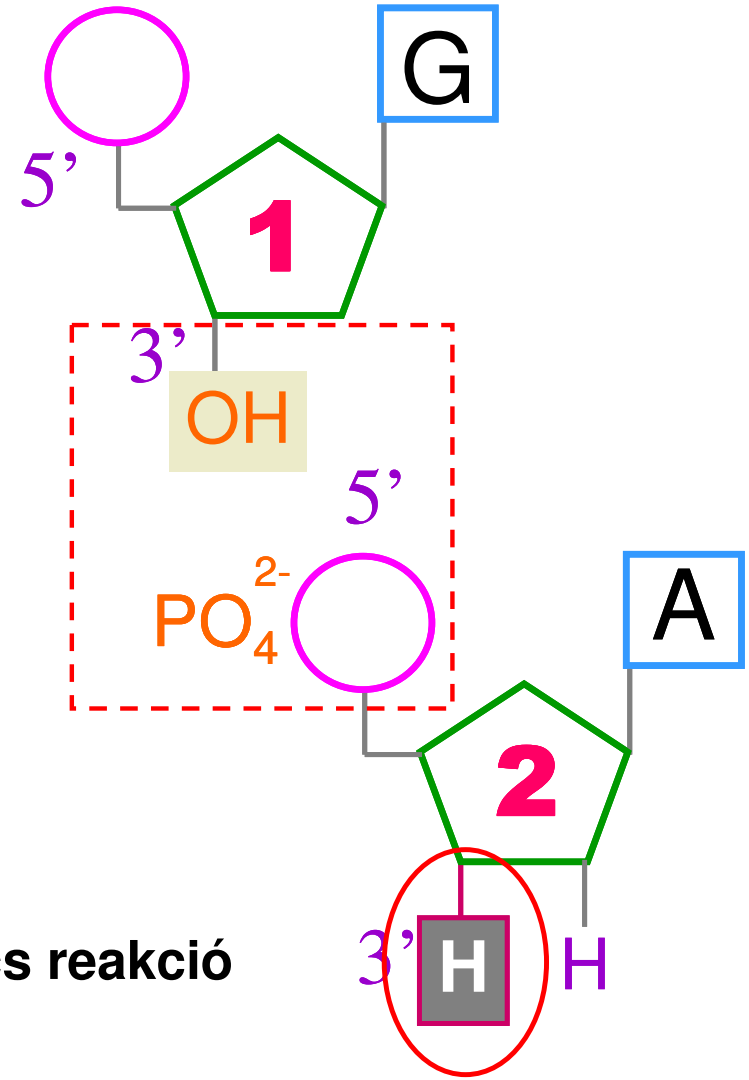
A = ddATP



Különböző méretű fragmenseket produkál


5'
3'
primer
template
Labeled primer
primer
template
Labeled dNTP
primer
template
Labeled terminator (ddNTP)

DNS szekvenálás

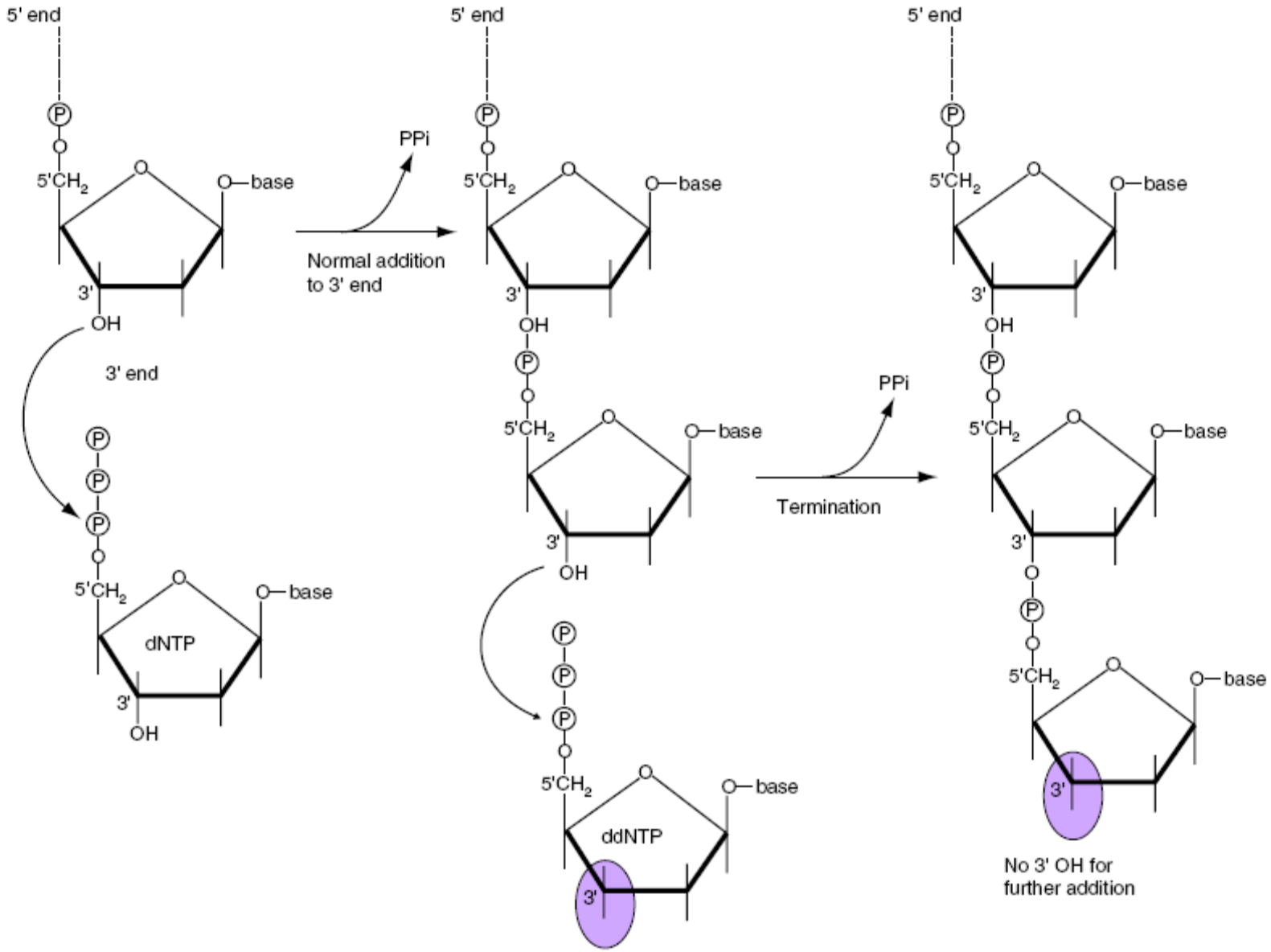


Foszfodiészter kötés

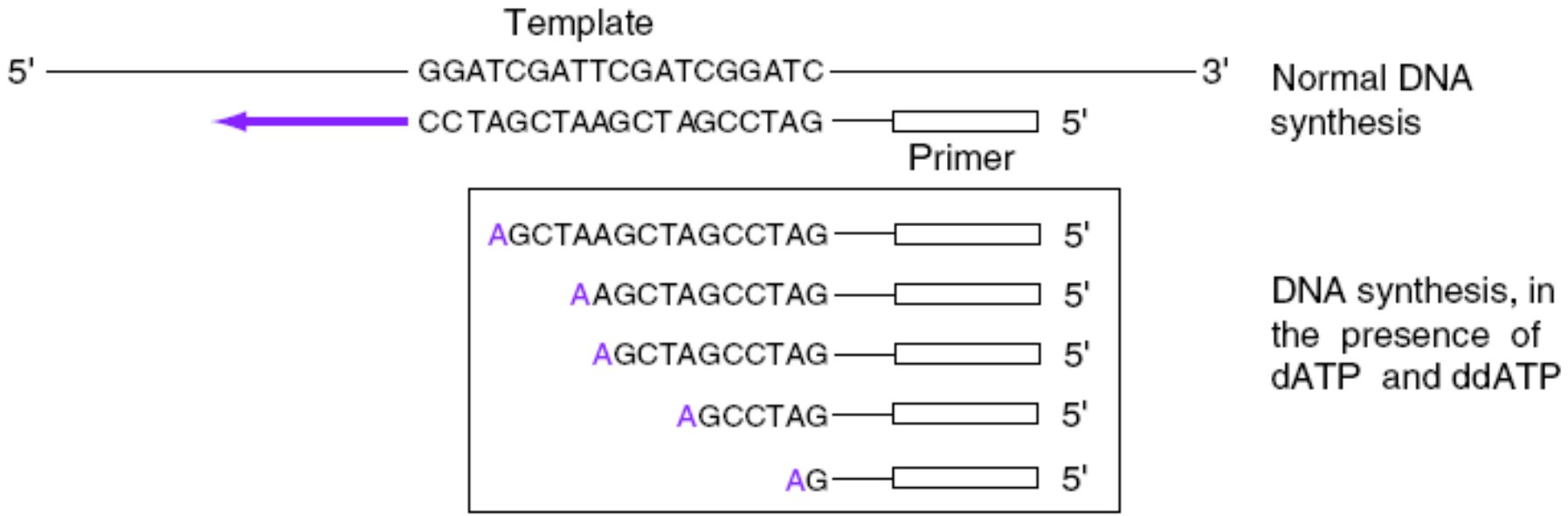
***Di*deoxynucleotide (ddATP)**

Nincs reakció

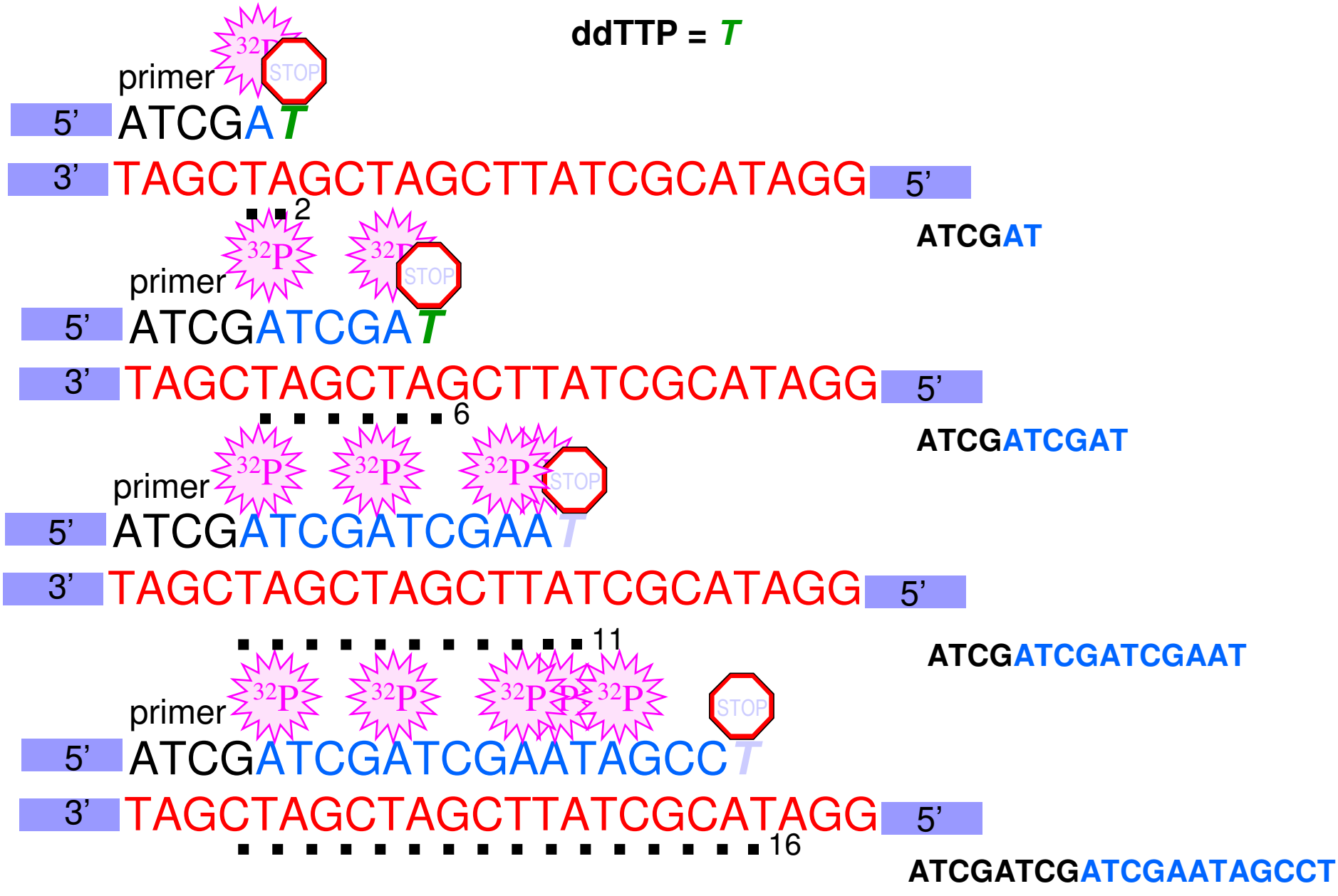
DNS szekvenálás



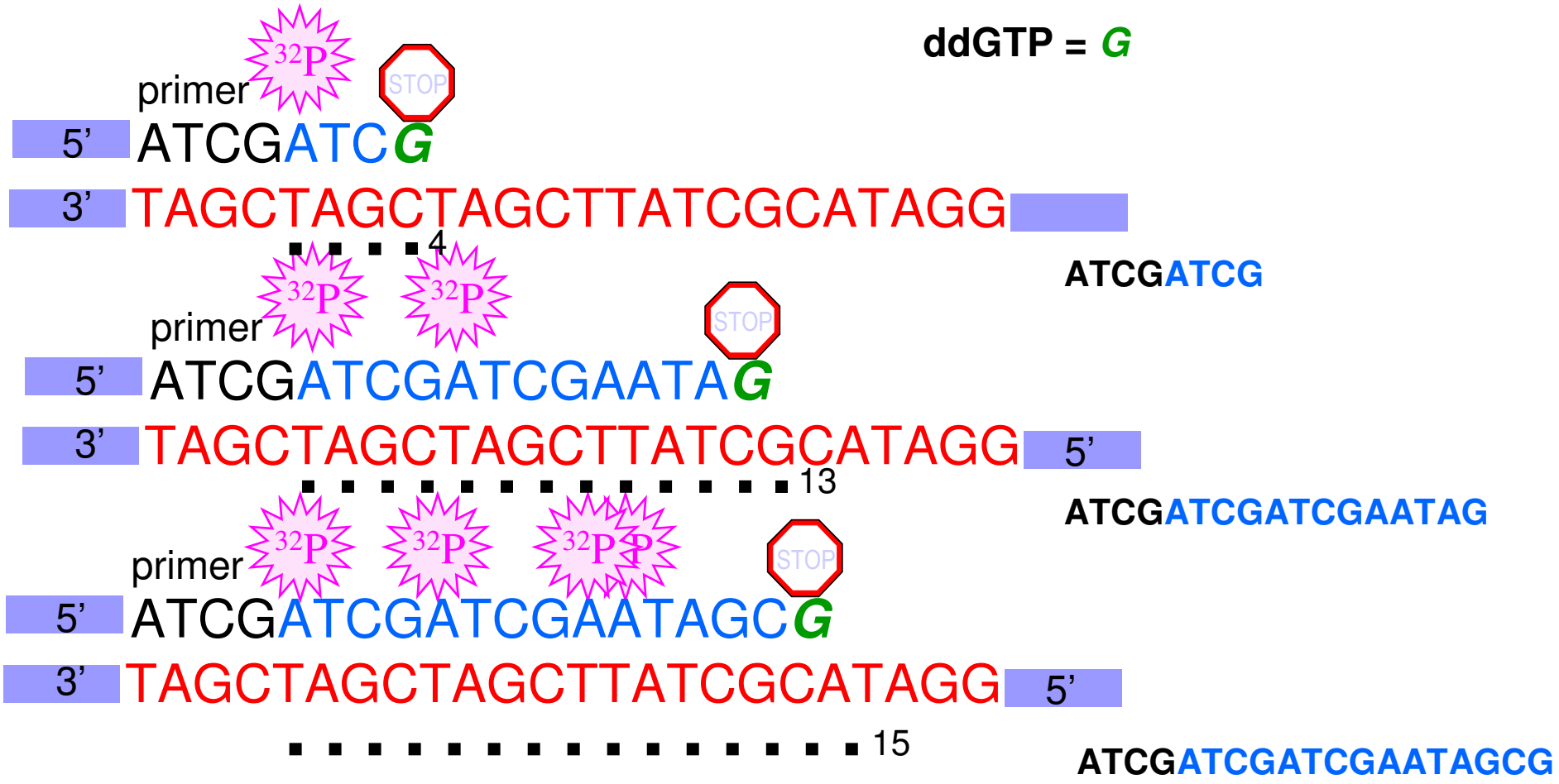
DNS szekvenálás



DNS szekvenálás



DNS szekvenálás

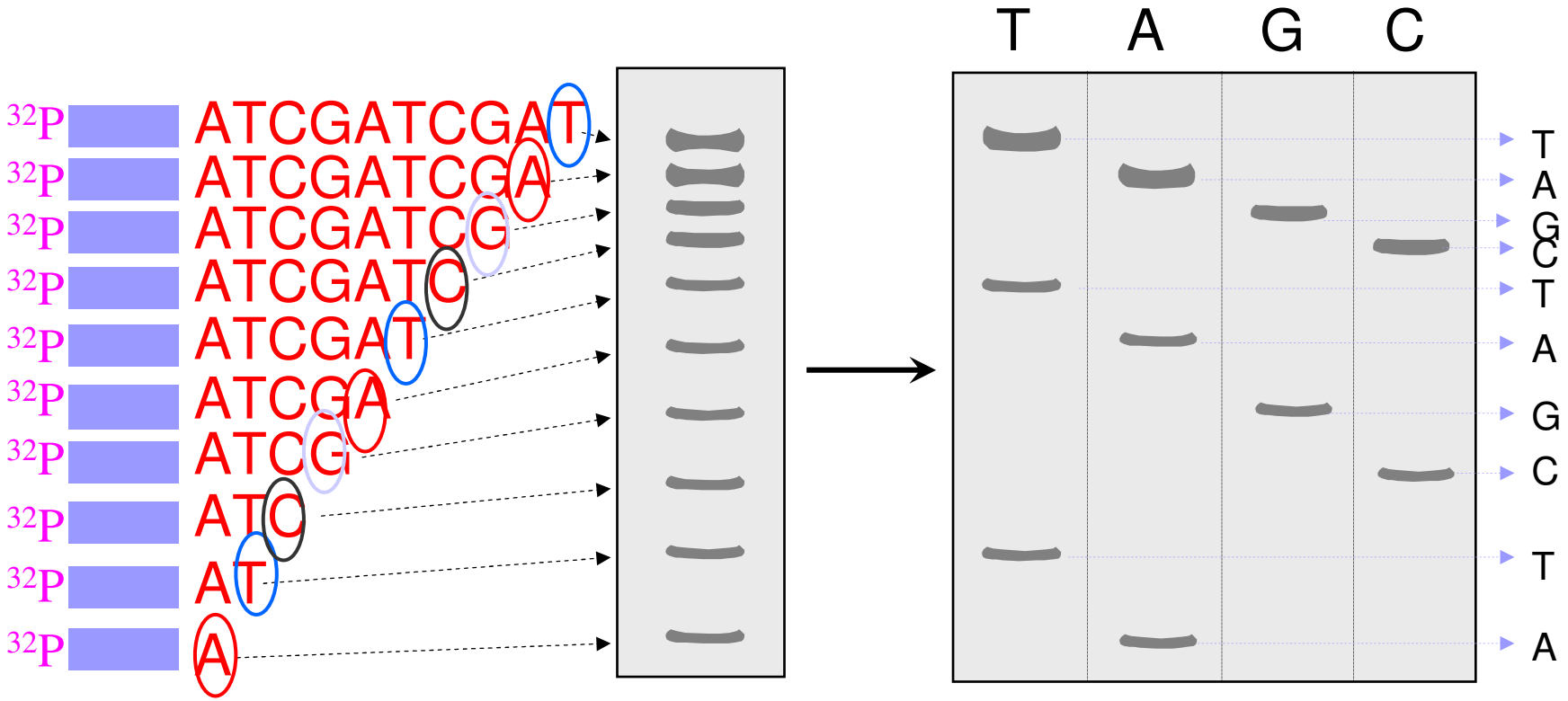


Hogyan határozzuk meg a DNS szekvenciát ?

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
A	1				5				9	10		12					17	
T		2				6					11					16		
G				4				8					13		15			
C			3				7							14				18

Hogyan határozzuk meg a DNS szekvenciát ?

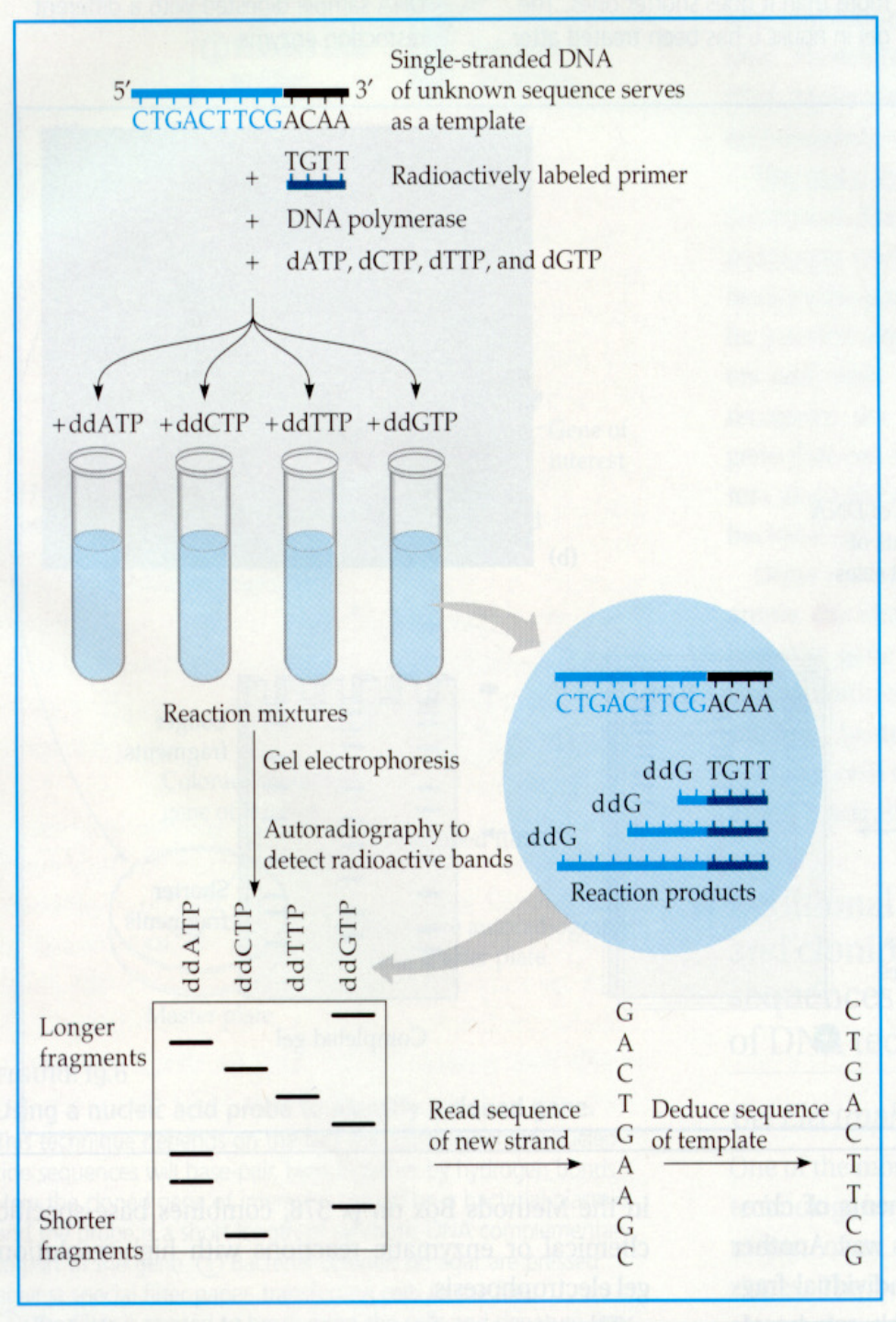
A DNS fragmensek egy nukleotidban különböznek, s elválaszthatóak gél elektroforézissel:



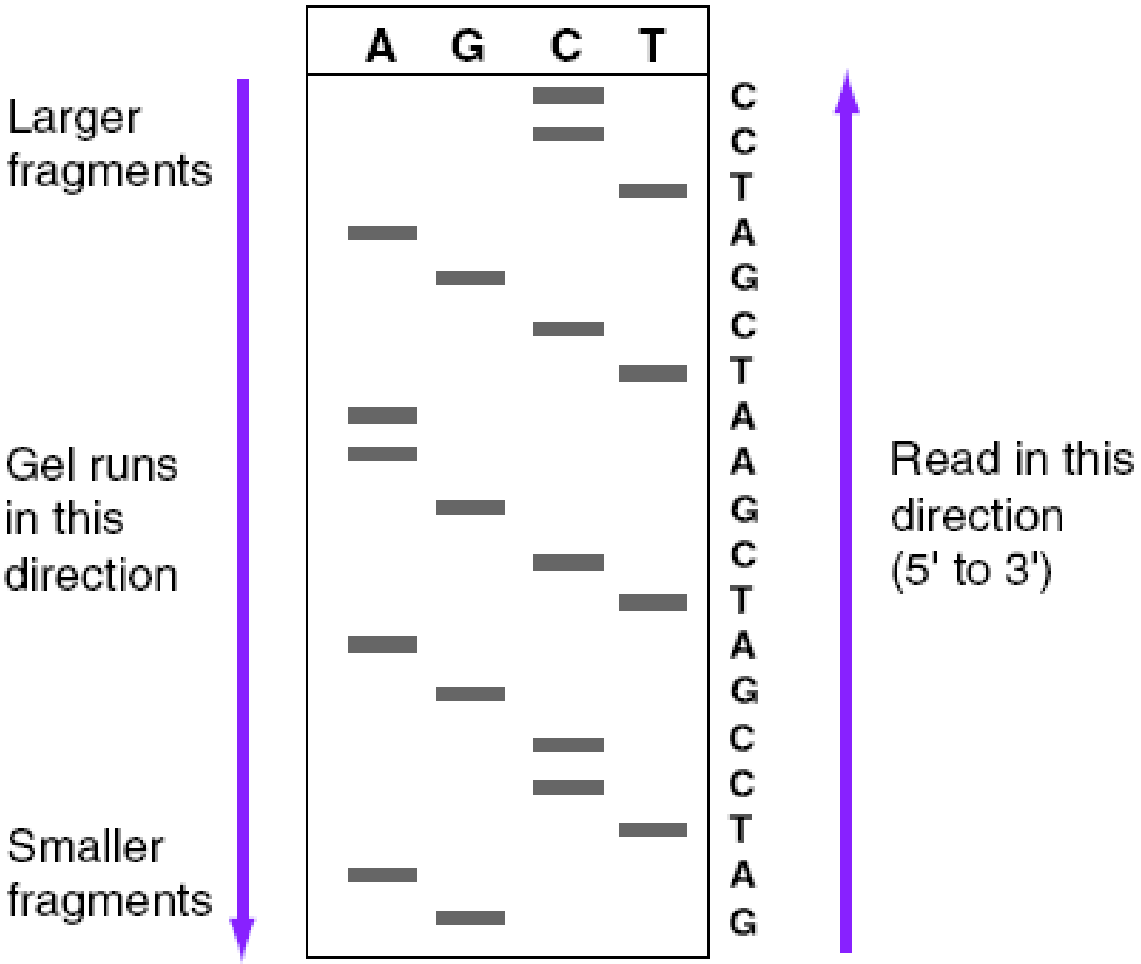
De, e csíkok alapján nem azonosíthatóak a terminális nukleotidok.

DNS szekvenálás

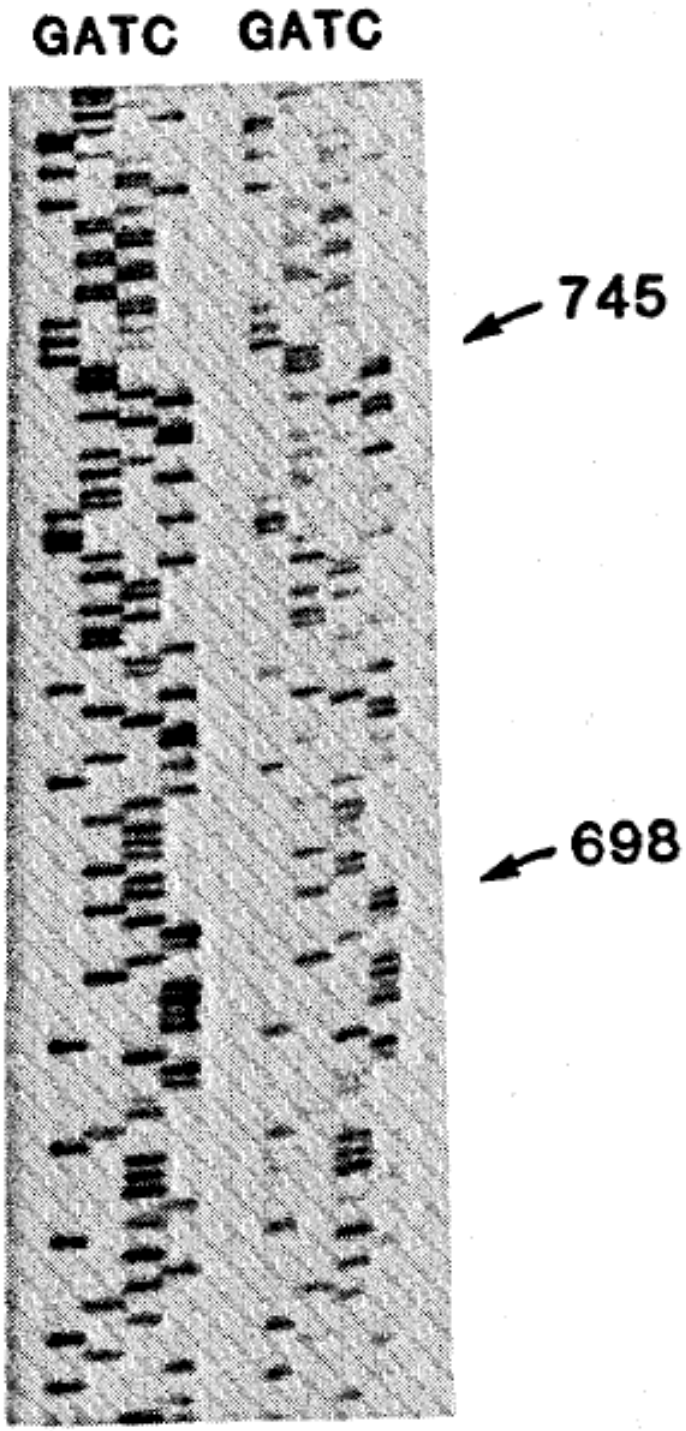
PRINCIPLES OF
DNA SEQUENCING
BY THE DIDEOXY
CHAIN TERMINATION
METHOD OF
SANGER *ET AL.*



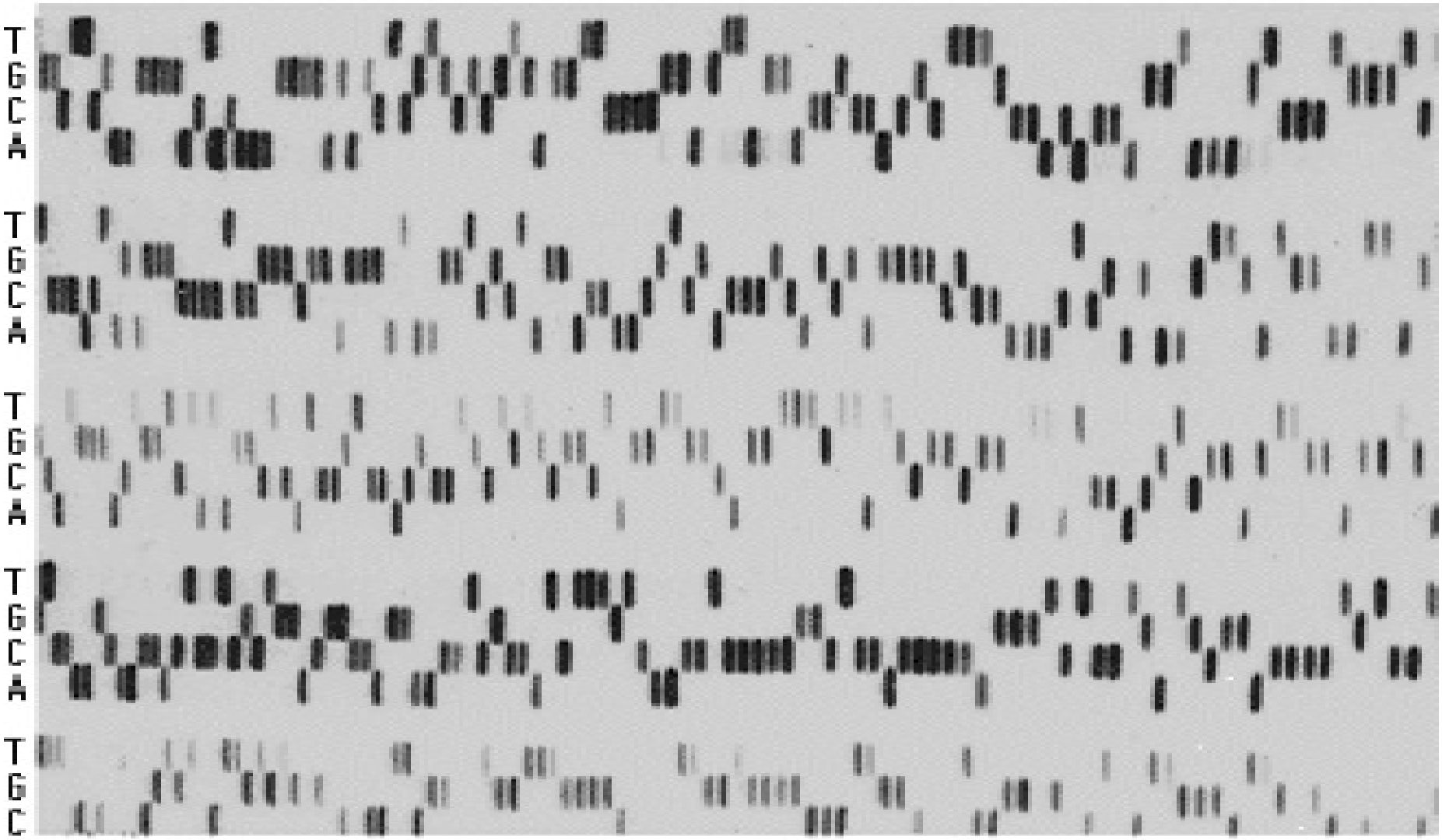
Szekvenálás radioaktívan-jelölt ddNTP-kkel



DNS szekvenálás

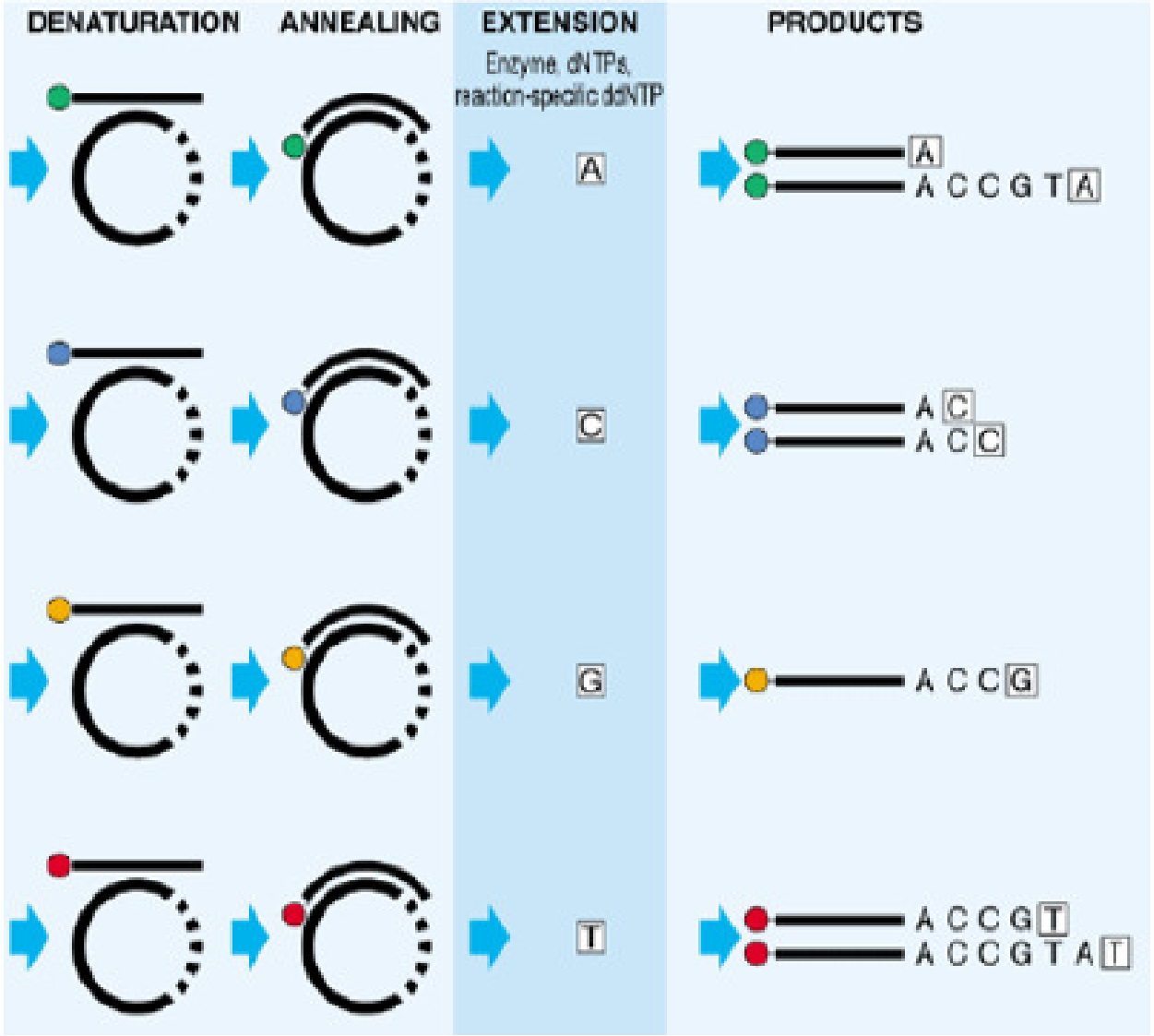


DNS szekvenálás



Ugyanaz a módszer, de a primer végén fluoreszcens festék a P^{32} helyett.
Elvileg egy festékkel nem sok különbség. De négy különböző festékkel már igen!

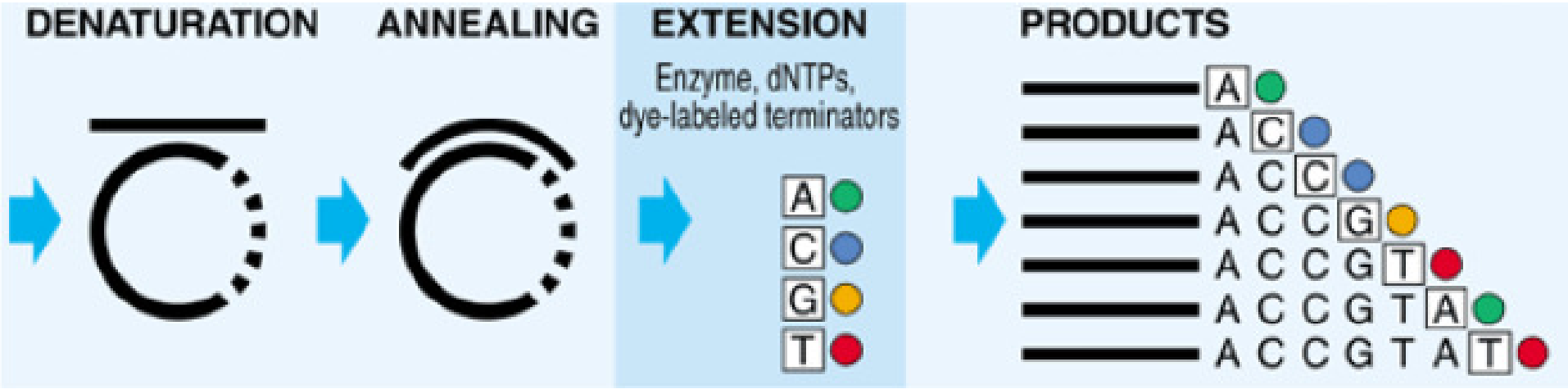
DNS szekvenálás



Ugyanaz a módszer, de a primer végén fluoreszcens festék a P^{32} helyett.
Elvileg egy festékkel nem sok különbség. De négy különböző festékkel már igen!

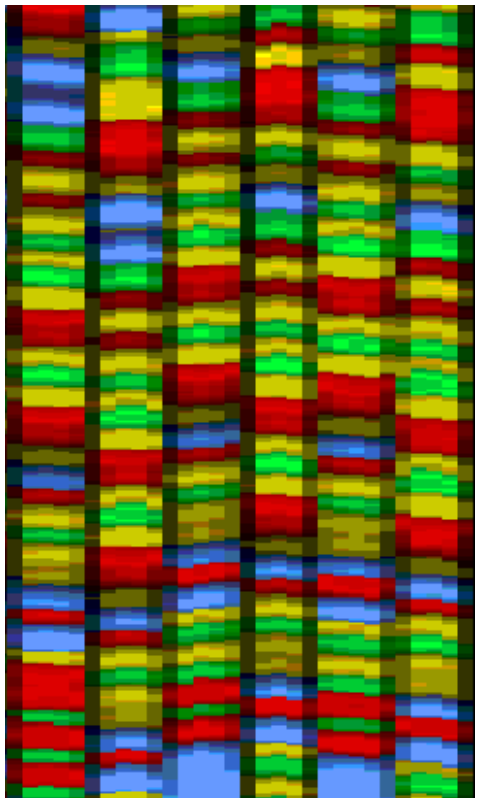
Modern DNS szekvenálás

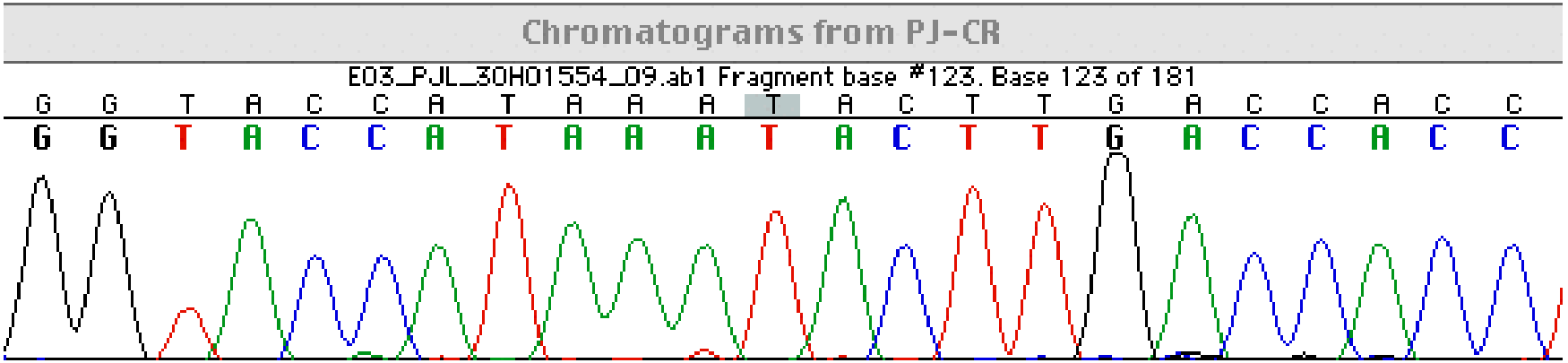
Sanger módszer (terminációs módszer) 1987. Prober és mtsai harmadik változat, fluoreszcens didezoxi festékek



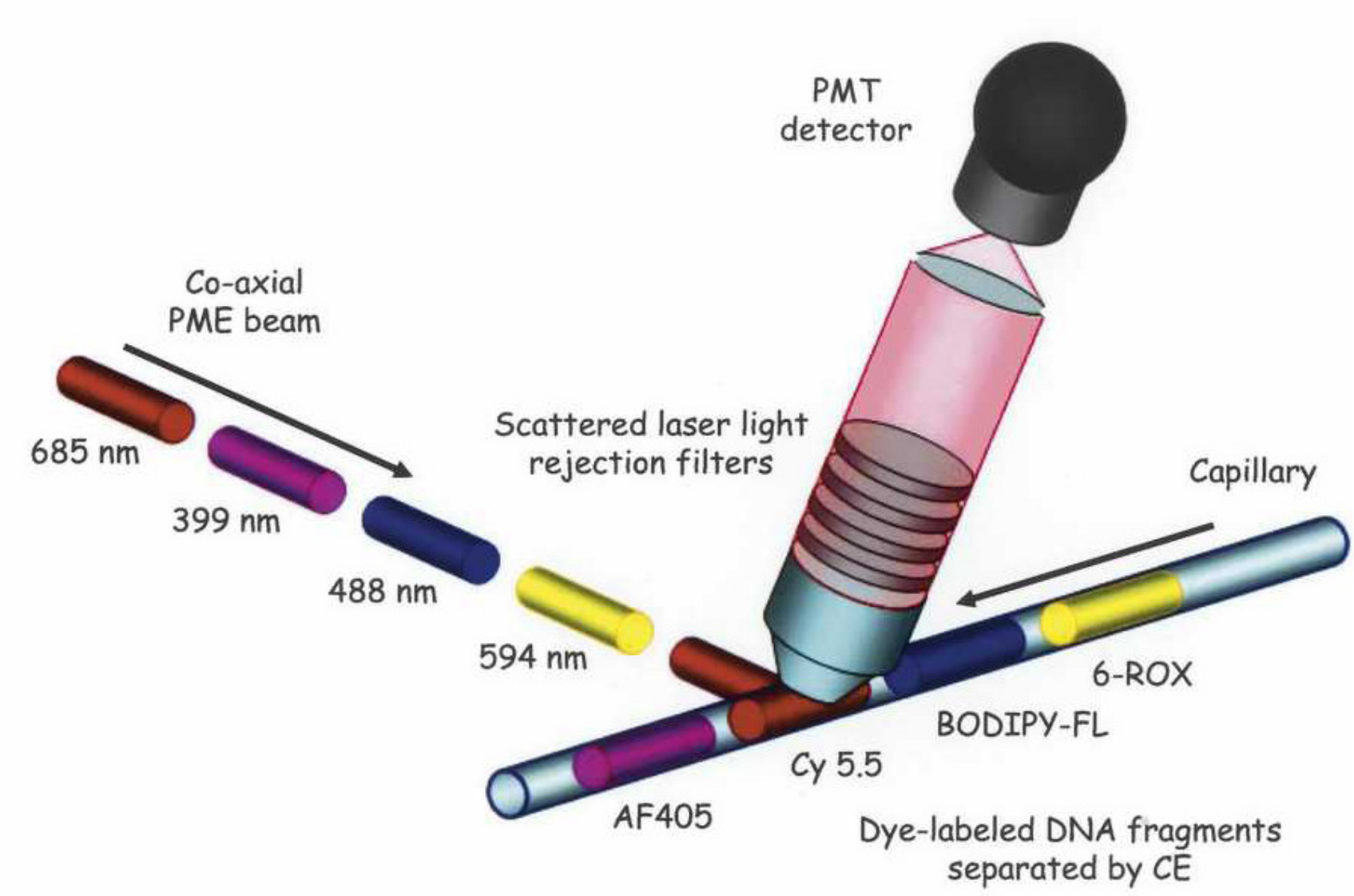
Modern DNS szekvenálás

Sanger módszer (terminációs módszer) 1987. Prober és mtsai harmadik változat, fluoreszcens didezoxi festékek





AUTOMATED DNA SEQUENCING



AUTOMATED DNA SEQUENCING

Using fluorescent terminators or primers

(ABI, Pharmacia ALF express, Li-Cor

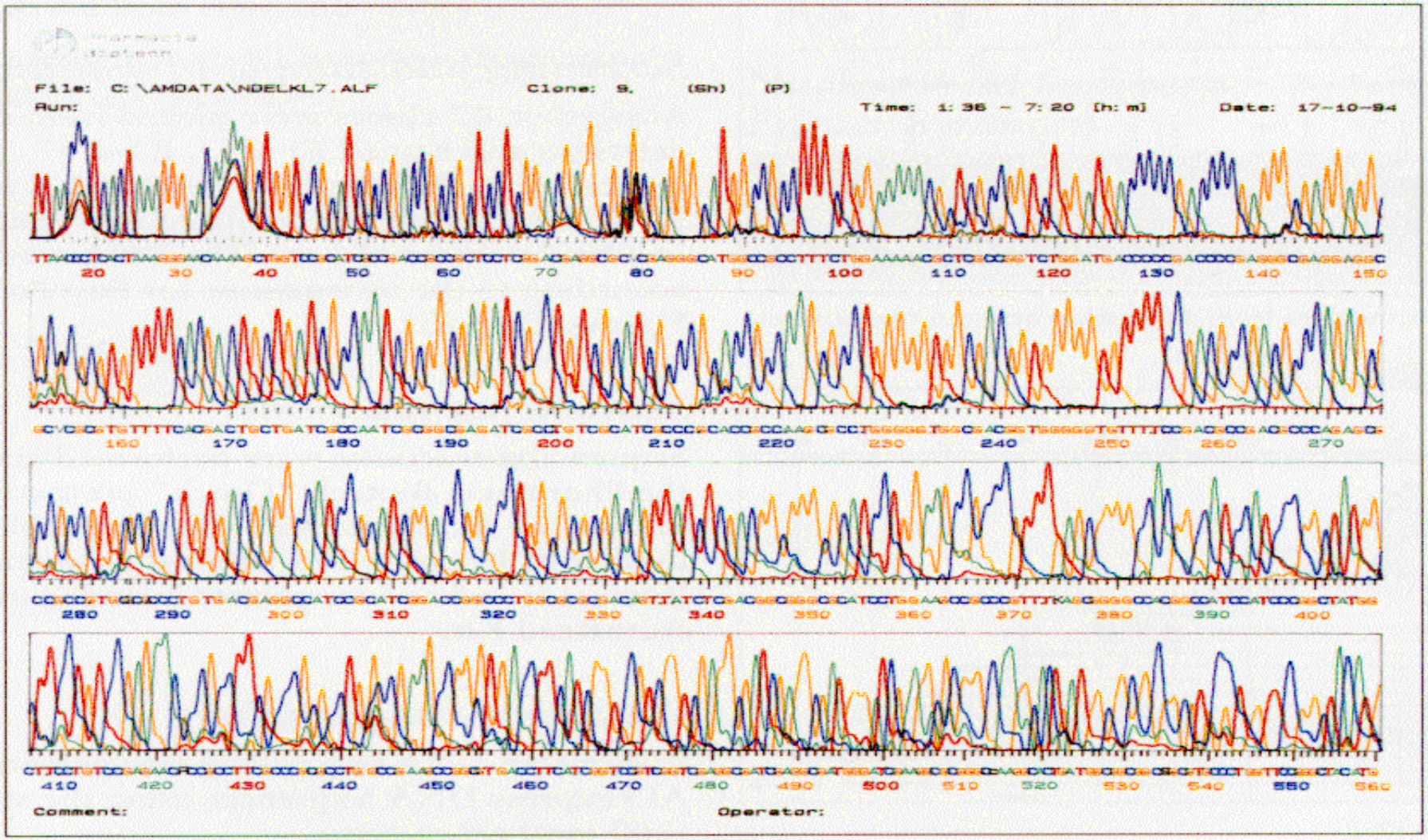
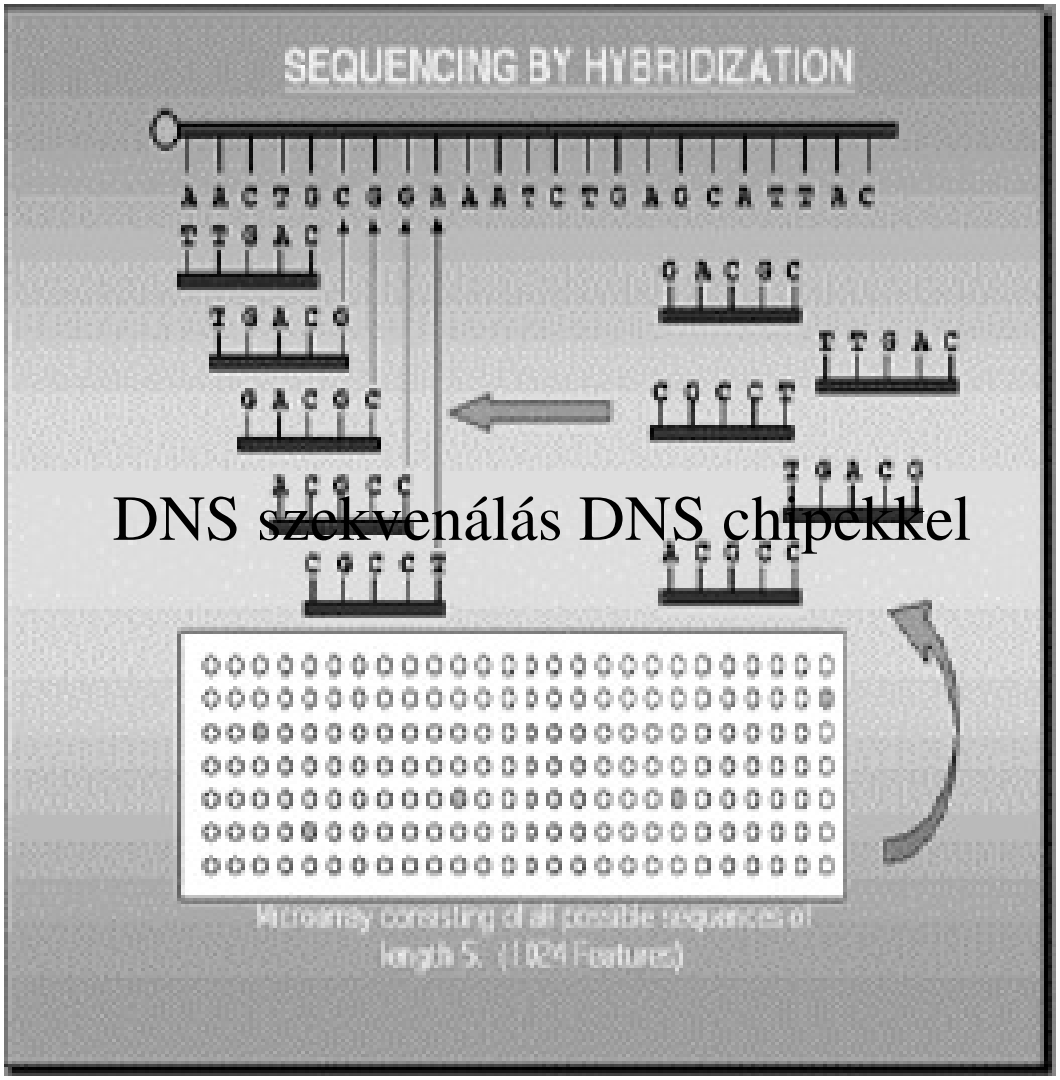


Fig. 2. Printout.

DNS szekvenálás DNS chipekkel



DNS szekvenálás DNS chipekkel

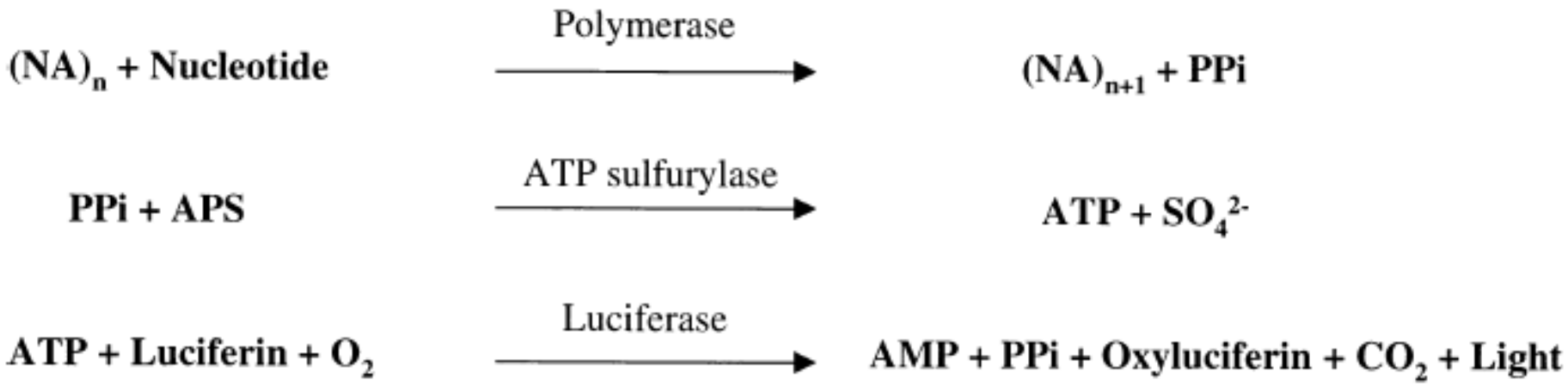
Hátrány: hibridizációs különbségek és ismétlődések gondjai!

Piroszekvenálás (pirofoszfátképződés mérése)

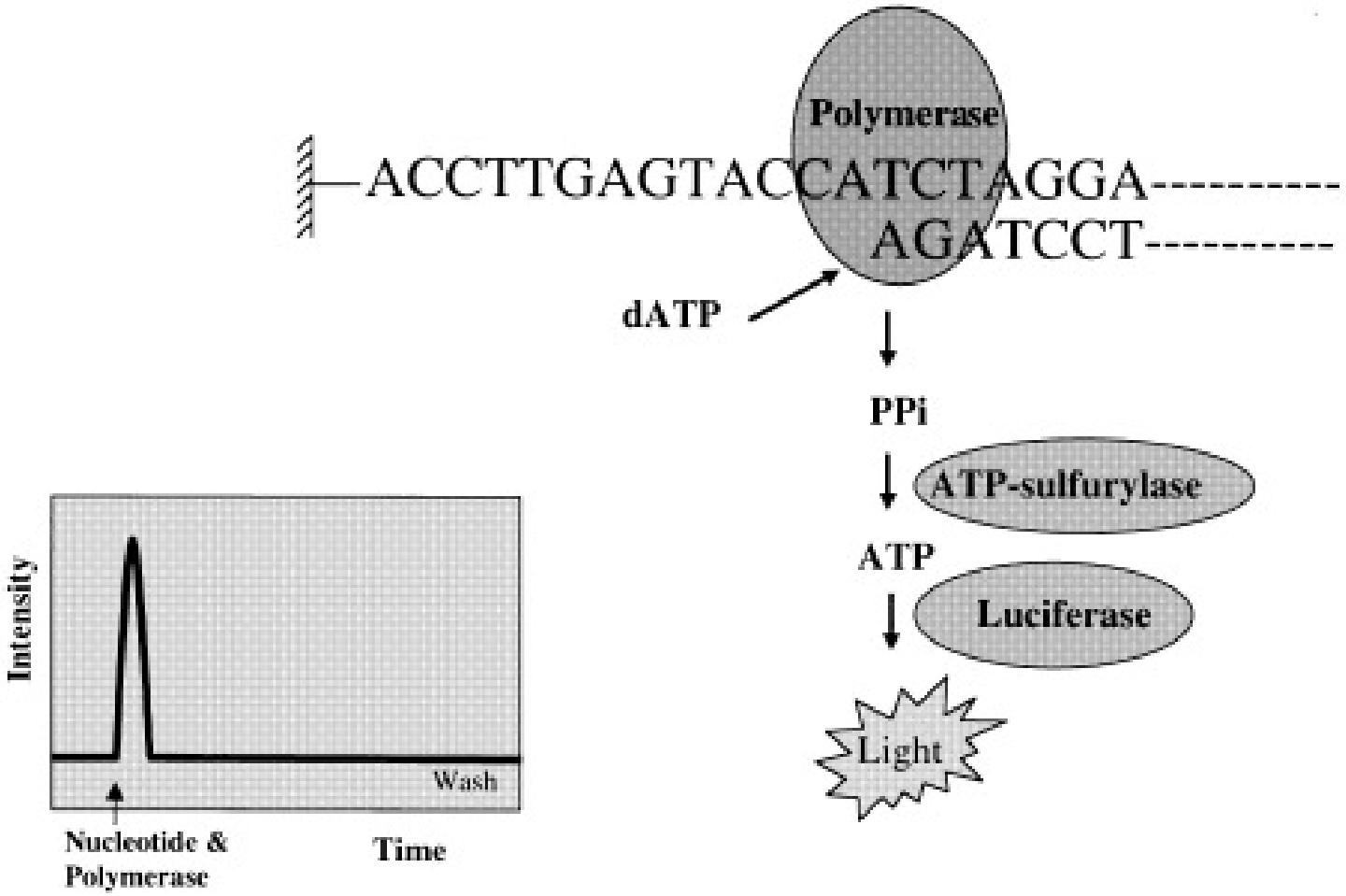
Pyrosequencing is a nonfluorescence technique that measures the release of inorganic pyrophosphate, which is proportionally converted into visible light by a series of enzymatic reactions (Ronaghi et al. 1996, 1998). Unlike other sequencing approaches that use 3'-modified dNTPs to terminate DNA synthesis, the pyrosequencing assay manipulates DNA polymerase by single addition of dNTPs in limiting amounts. Upon addition of the complementary dNTP, DNA polymerase extends the primer and pauses when it encounters a noncomplementary base. DNA synthesis is reinitiated following the addition of the next complementary dNTP in the dispensing cycle. The light generated by the enzymatic cascade is recorded as a series of peaks called a pyrogram, which corresponds to the order of complementary dNTPs incorporated and reveals the underlying DNA sequence.

The 454 Corporation has recently introduced a whole genome sequencing strategy by integrating pyrosequencing with their PicoTiterPlate (PTP) platform, which has been shown to amplify and image approximately 300,000 PCR templates captured on Sepharose beads (Leamon et al. 2003).

Piroszekvenálás (pirofoszfátképződés mérése)

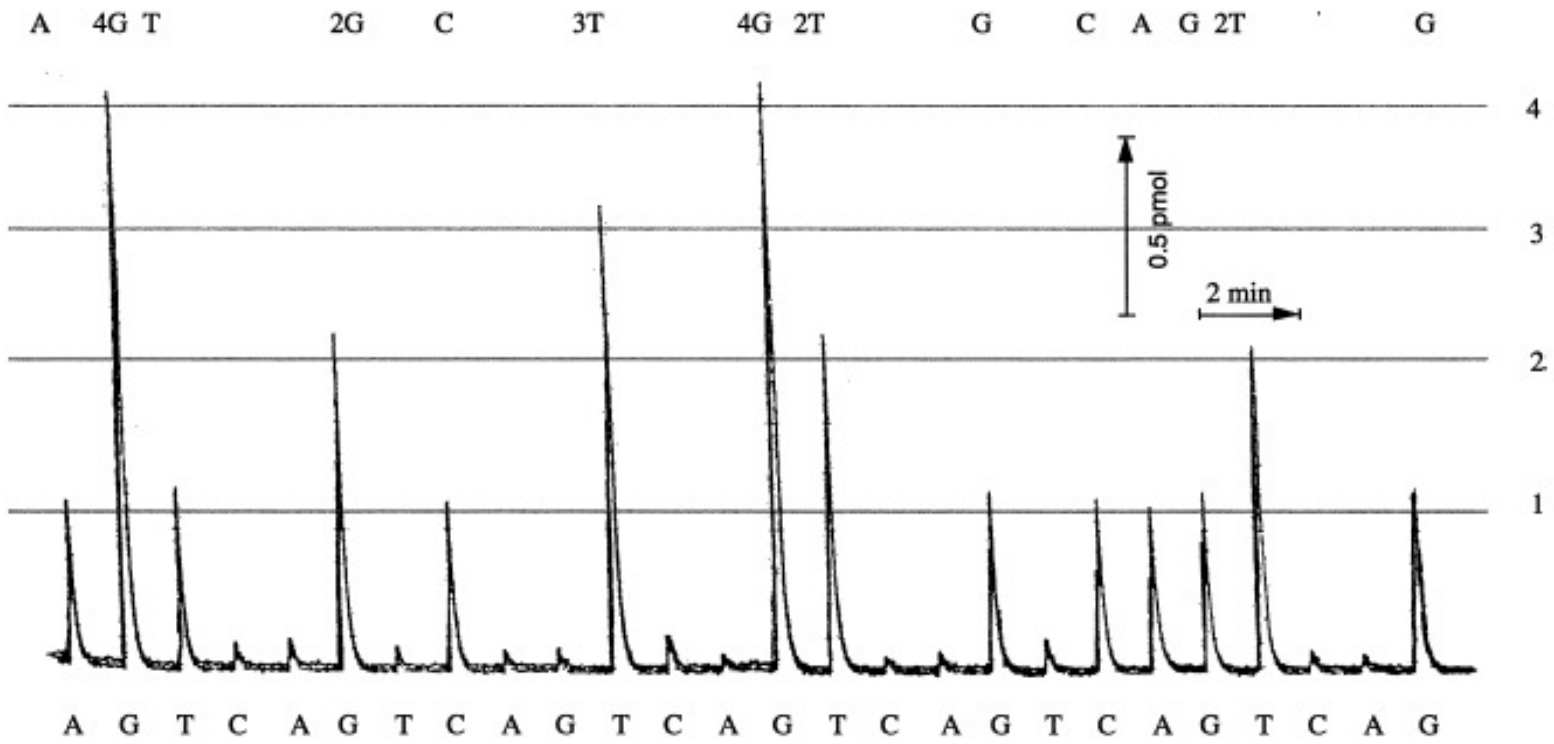


Piroszekvenálás (pirofoszfátképződés mérése)



Schematic representation of the progress of the enzyme reaction in solid-phase pyrosequencing. The four different nucleotides are added stepwise to the immobilized primed DNA template and the incorporation event is followed using the enzyme ATP sulfurylase and luciferase. After each nucleotide addition, a washing step is performed to allow iterative addition.

Piroszekvenálás (pirofoszfátképződés mérése)



Pyrogram of the raw data obtained from liquid-phase pyrosequencing. Proportional signals are obtained for one, two, three, and four base incorporations. Nucleotide addition, according to the order of nucleotides, is indicated below the pyrogram and the obtained sequence is indicated above the pyrogram.


Z400
GS Junior
Roche

Workflow

The GS Junior System offers an end-to-end sequencing solution from sample preparation and sequence generation through data analysis. Robust protocols with minimal handling steps make the workflow ideally suited for individual labs. Produce libraries in less then half a day with easy-to-follow sample preparation protocols. Use only general laboratory equipment without the need to purchase tons of additional supplies. Perform overnight sequencing and data processing with a quick 10-hour instrument run time. Go from sequence data to publishable result with straightforward tools for de novo assembly, mapping and amplicon variant analysis.

How It Works?

1. Sample Input and Fragmentation

The GS Junior System supports the sequencing of samples from a wide variety of starting materials including genomic DNA, PCR products, BACs, and cDNA. Samples such as genomic DNA and BACs are randomly fragmented into small, 300- to 800-basepair pieces. For smaller samples, such as small non-coding RNA or PCR amplicons, fragmentation is not required. Instead, PCR products amplified using Genome Sequencer fusion primers may be immobilized onto DNA capture beads and clonally amplified as shown below under "One Fragment = One Bead".

2. Library Preparation
Using a series of standard molecular biology techniques, short DNA adaptors are added to each library fragment. These adaptors are then used in subsequent quantification, amplification, and sequencing steps.

3. One Fragment = One Bead
The single-stranded DNA library is immobilized onto specifically designed DNA Capture Beads. Each bead carries a unique single-stranded DNA library fragment. The bead-bound library is emulsified with amplification reagents in a water-in-oil mixture resulting in microreactors containing just one bead with one unique sample-library fragment.

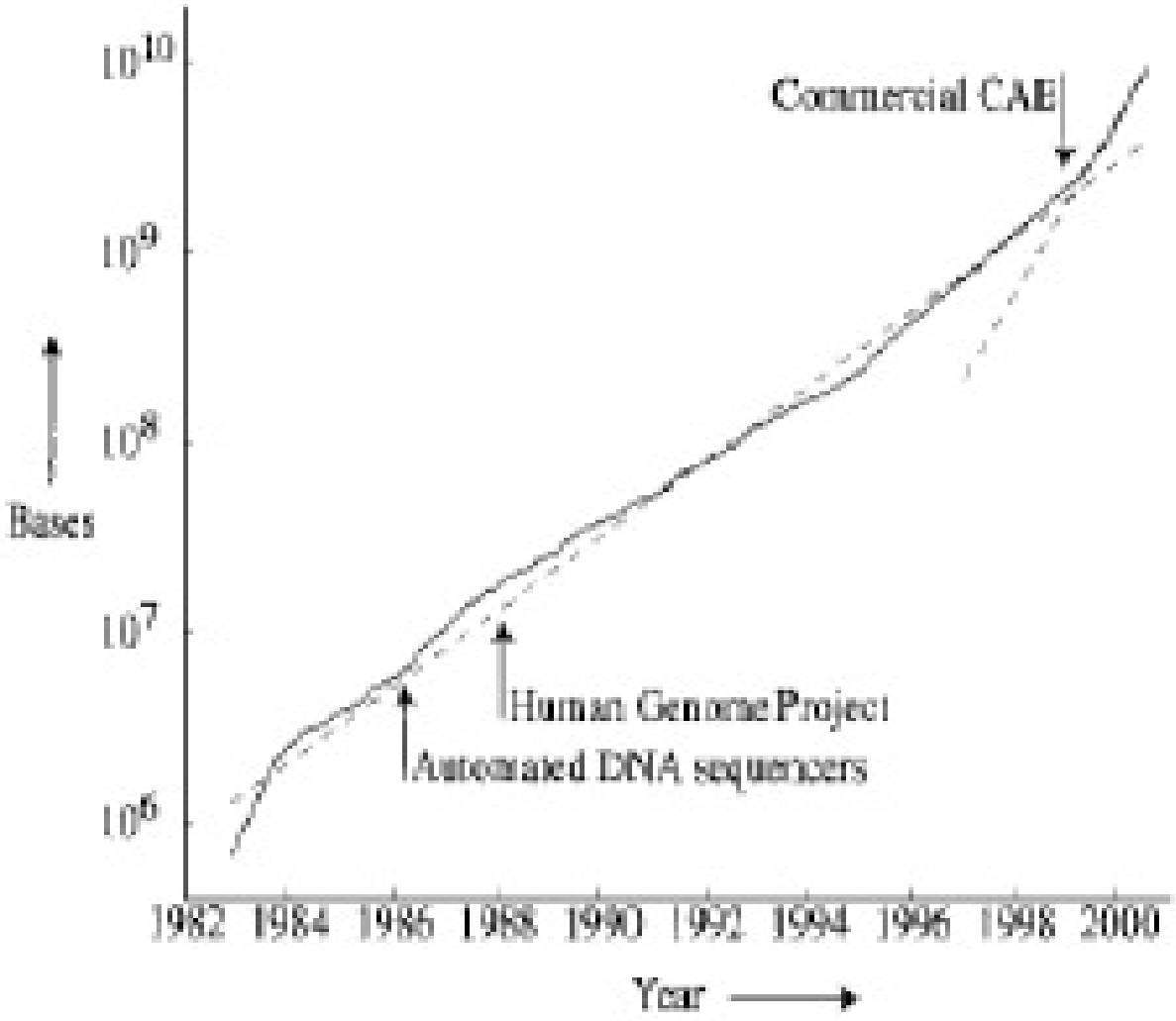
4. emPCR (Emulsion PCR) Amplification
Each unique sample library fragment is clonally amplified within its own microreactor, excluding competing or contaminating sequences. Amplification of the entire fragment collection is carried out in parallel; for each fragment, this produces several million copies of the original fragment per bead. Subsequently, the emulsions are broken to facilitate collection of the amplified fragments bound to their specific beads.

5. One Bead = One Read

The clonally amplified fragments are enriched and loaded onto a PicoTiterPlate device for sequencing. The diameter of the PicoTiterPlate wells allows for only one bead per well. After addition of sequencing enzymes and reagents, the fluidics subsystem of the Genome Sequencer System serially flows nucleotides in a fixed order (i.e. first T, then A, and so on) across the hundreds of thousands of wells containing one bead each. Addition of one (or more) nucleotide(s) complementary to the template strand results in a chemiluminescent signal recorded by the CCD camera of the Genome Sequencer System. The intensity of the resulting signal is proportional to the number of bases incorporated.

6. Data Analysis

The combination of signal intensity and positional information generated across the PicoTiterPlate device allows the software to determine the sequence of 100,000 individual reads per 10-hour instrument run simultaneously. For sequencing data analysis, three different bioinformatics tools are supplied that readily support the following applications: de novo genome assembly up to 3 Gb; resequencing/mapping genomes of any size; and amplicon variant detection by comparison with a known reference sequence.


10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
Bases
Commercial CAE
Human Genome Project
Automated DNA sequencers
1982 1984 1986 1988 1990 1992 1994 1996 1998 2000
Year

TOTAL GENOME SEQUENCING PROJECTS

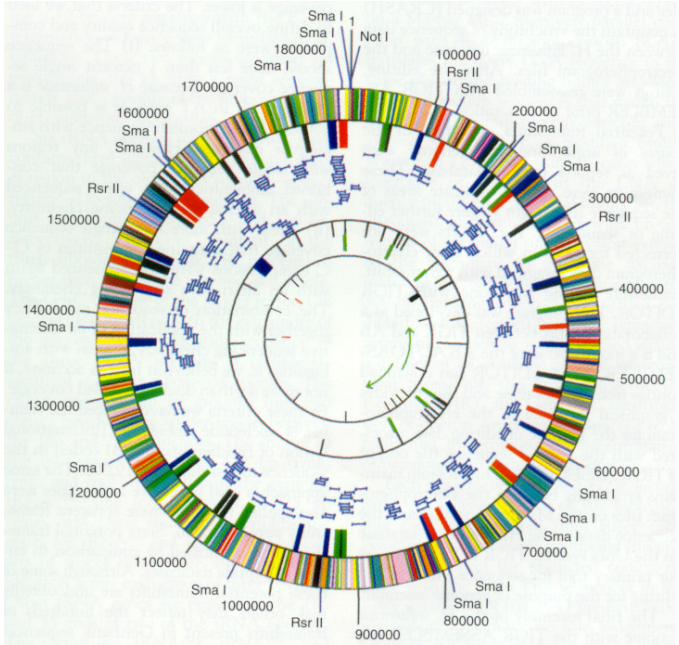


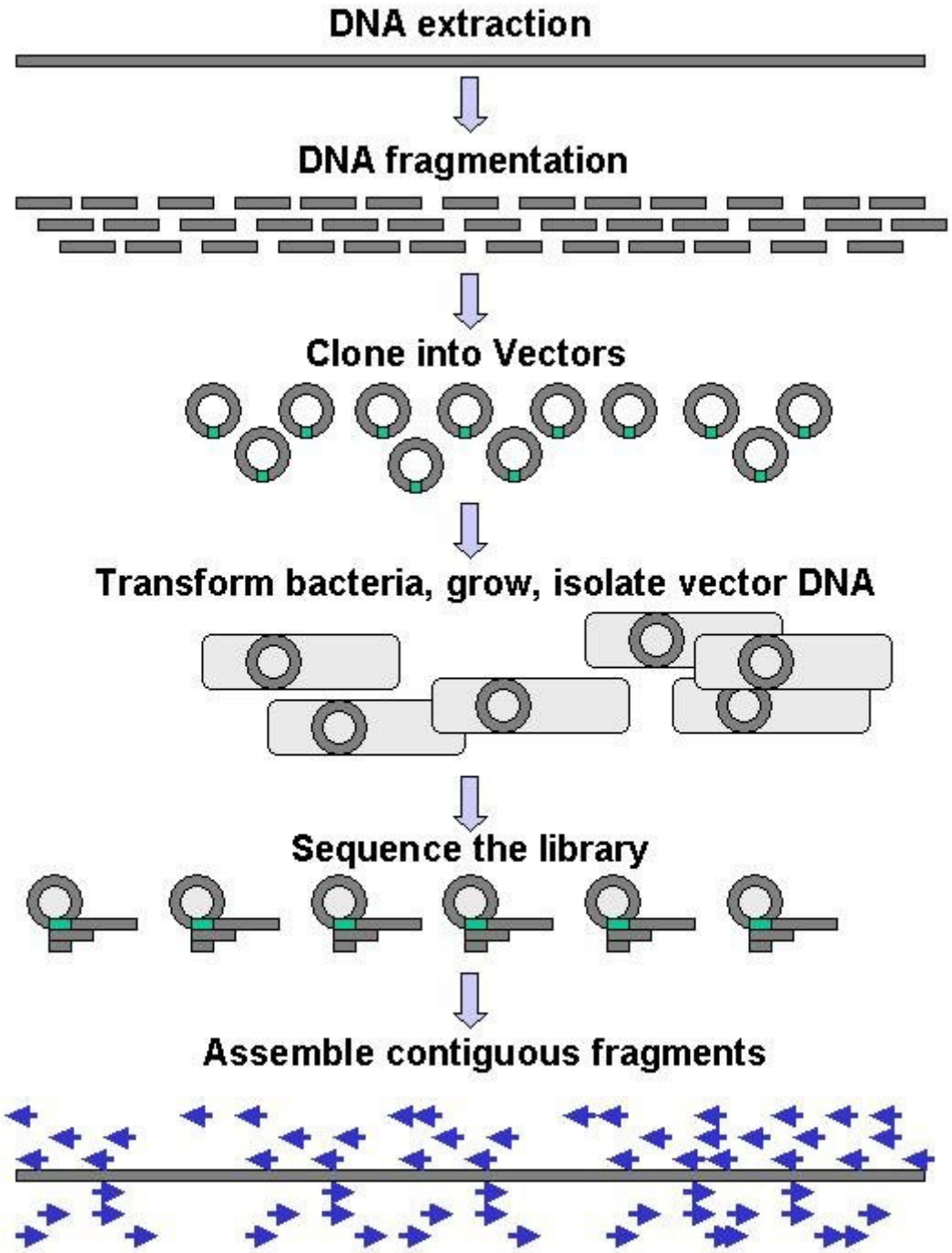
Fig. 1. A circular representation of the *H. influenzae* Rd chromosome illustrating the location of eac

Haemophilus influenzae
TIGR (1995)

28,984 sequencing reactions
84% success
Ave. read length 485 bp

Genome size	1.8301 Mb
Predicted ORFs	1743
Unassigned	42%

No. of authors 39


DNA extraction
DNA fragmentation
Clone into Vectors
Transform bacteria, grow, isolate vector DNA
Sequence the library
Assemble contiguous fragments

Box 7.1 Estimates of the required size of genomic libraries

Organism	Genome size	Vector type	Insert size	P	Library size
Bacterium	4×10^6 bases	plasmid	4 kb	0.99	4.6×10^3
		lambda replacement	18 kb	0.99	1.0×10^3
		cosmid	40 kb	0.99	458
		BAC	300 kb	0.99	59
Mammal	3×10^9 bases	plasmid	4 kb	0.99	3.5×10^6
		lambda replacement	18 kb	0.99	7.7×10^5
		cosmid	40 kb	0.99	3.5×10^5
		BAC	300 kb	0.99	4.6×10^4

The values shown for the genome sizes of bacteria and mammals are examples for the purpose of this calculation. The actual genome sizes vary quite widely from one organism to another. The insert sizes for specific vectors will also vary.

Table 1–1 Some Genomes That Have Been Completely Sequenced

SPECIES	SPECIAL FEATURES	HABITAT	GENOME SIZE (1000s OF NUCLEOTIDE PAIRS PER HAPLOID GENOME)	ESTIMATED NUMBER OF GENES CODING FOR PROTEINS
BACTERIA				
Mycoplasma genitalium	has one of the smallest of all known cell genomes	human genital tract	580	468
Synechocystis sp.	photosynthetic, oxygen-generating (cyanobacterium)	lakes and streams	3573	3168
Escherichia coli	laboratory favorite	human gut	4639	4289
Helicobacter pylori	causes stomach ulcers and predisposes to stomach cancer	human stomach	1667	1590
Bacillus anthracis	causes anthrax	soil	5227	5634
Aquifex aeolicus	lithotrophic; lives at high temperatures	hydrothermal vents	1551	1544
Streptomyces coelicolor	source of antibiotics; giant genome	soil	8667	7825
Treponema pallidum	spirochete; causes syphilis	human tissues	1138	1041
Rickettsia prowazekii	bacterium most closely related to mitochondria; causes typhus	lice and humans (intracellular parasite)	1111	834
Thermotoga maritima	organotrophic; lives at very high temperatures	hydrothermal vents	1860	1877

Genome size and gene number vary between strains of a single species, especially for bacteria and archaea. The table shows data for particular strains that have been sequenced. For eucaryotes, many genes can give rise to several alternative variant proteins, so that the total number of proteins specified by the genome is substantially greater than the number of genes.

Table 1–1 Some Genomes That Have Been Completely Sequenced

SPECIES	SPECIAL FEATURES	HABITAT	GENOME SIZE (1000s OF NUCLEOTIDE PAIRS PER HAPLOID GENOME)	ESTIMATED NUMBER OF GENES CODING FOR PROTEINS
ARCHAEA				
Methanococcus jannaschii	lithotrophic, anaerobic, methane-producing	hydrothermal vents	1664	1750
Archaeoglobus fulgidus	lithotrophic or organotrophic, anaerobic, sulfate-reducing	hydrothermal vents	2178	2493
Nanoarchaeum equitans	smallest known archaean; anaerobic; parasitic on another, larger archaean	hydrothermal and volcanic hot vents	491	552
EUCARYOTES				
Saccharomyces cerevisiae (budding yeast)	minimal model eucaryote	grape skins, beer	12,069	~6300
Arabidopsis thaliana (Thale cress)	model organism for flowering plants	soil and air	~142,000	~26,000
Caenorhabditis elegans (nematode worm)	simple animal with perfectly predictable development	soil	~97,000	~20,000
Drosophila melanogaster (fruit fly)	key to the genetics of animal development	rotting fruit	~137,000	~14,000
Homo sapiens (human)	most intensively studied mammal	houses	~3,200,000	~24,000

Genome size and gene number vary between strains of a single species, especially for bacteria and archaea. The table shows data for particular strains that have been sequenced. For eucaryotes, many genes can give rise to several alternative variant proteins, so that the total number of proteins specified by the genome is substantially greater than the number of genes.

Phage λ
50 kb
2 pages

Escherichia coli
(bacteria)
4.7 Mb
200 pages

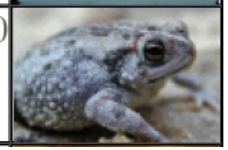
Saccharomyces cerevisiae
(yeast)
12.5 Mb
500 pages

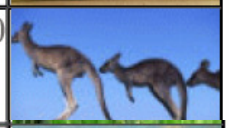
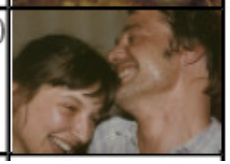
Caenorhabditis elegans
(nematode)
Arabidopsis thaliana
(plant)
100 Mb
3 volumes

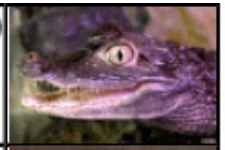
Drosophila melanogaster
(fruit fly)
165 Mb
5 volumes

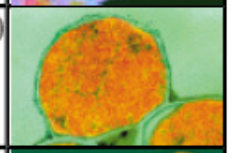
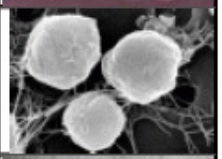
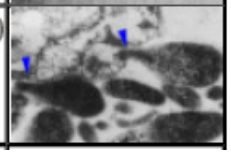
Human being
3000 Mb
80 volumes

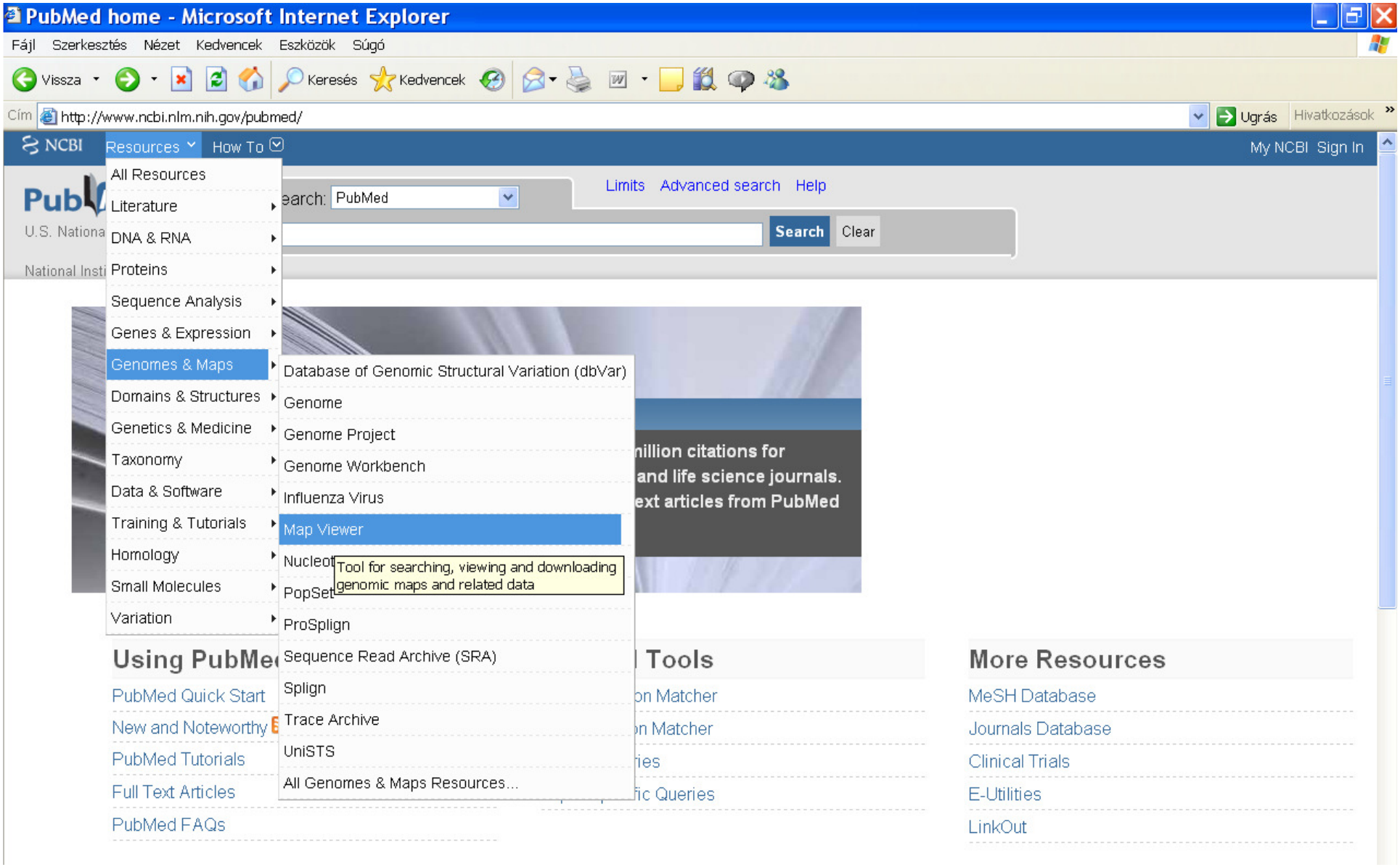
25 kb per page
1500 pages
per volume
(2 inches thick)

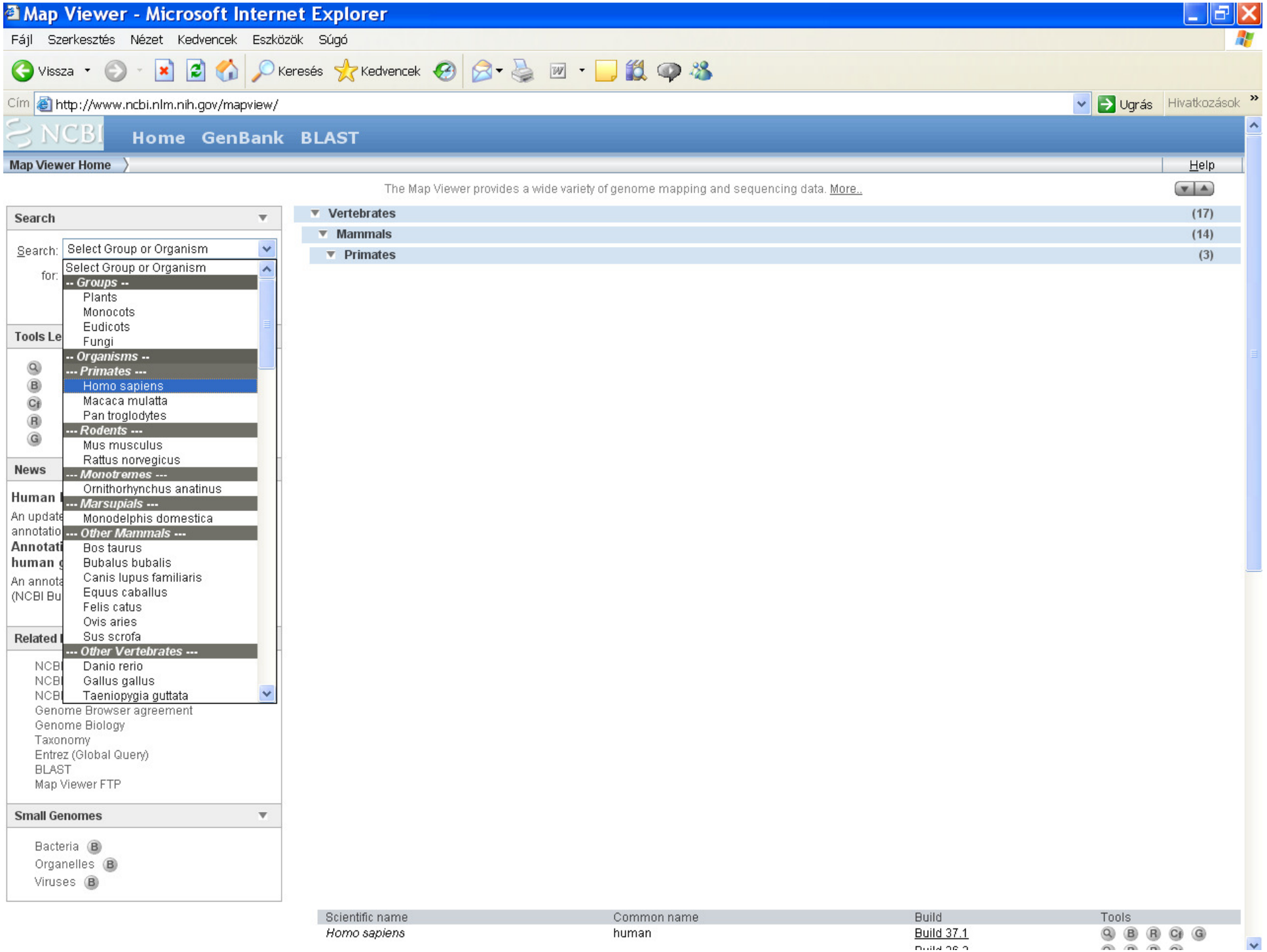
Amoeba dubia	670 000 000 000	
Amoeba proteus	290 000 000 000	
Lilium longiflorum	90 000 000 000	
Pinus resinosa	68 000 000 000	
Triturus cristatus	20 600 000 000	
Allium cepa	18 000 000 000	
Podisma pedestris	18 000 000 000	
Paramecium caudatum	8 600 000 000	
Bufo bufo	6 900 000 000	

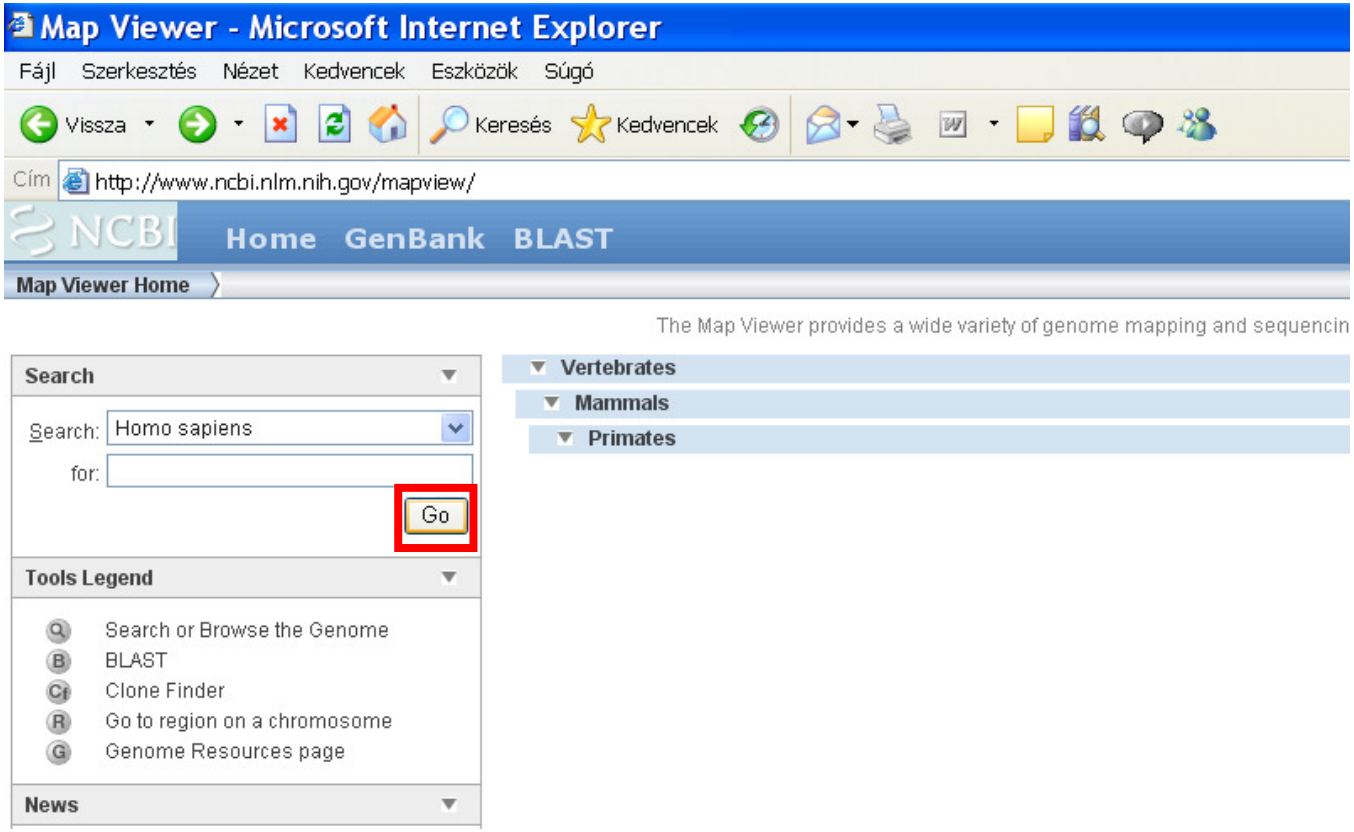
Zea mays	5 000 000 000	
Macropus giganteus	4 154 800 000	
Bos taurus	3 651 500 000	
Pongo pygmaeus	3 607 100 000	
Pan troglodytes	3 577 500 000	
Macaca mulatta	3 543 000 000	
Gorilla gorilla	3 523 200 000	
Papio hamadryas	3 478 800 000	
Mus musculus	3 454 200 000	
Homo sapiens	3 400 000 000	

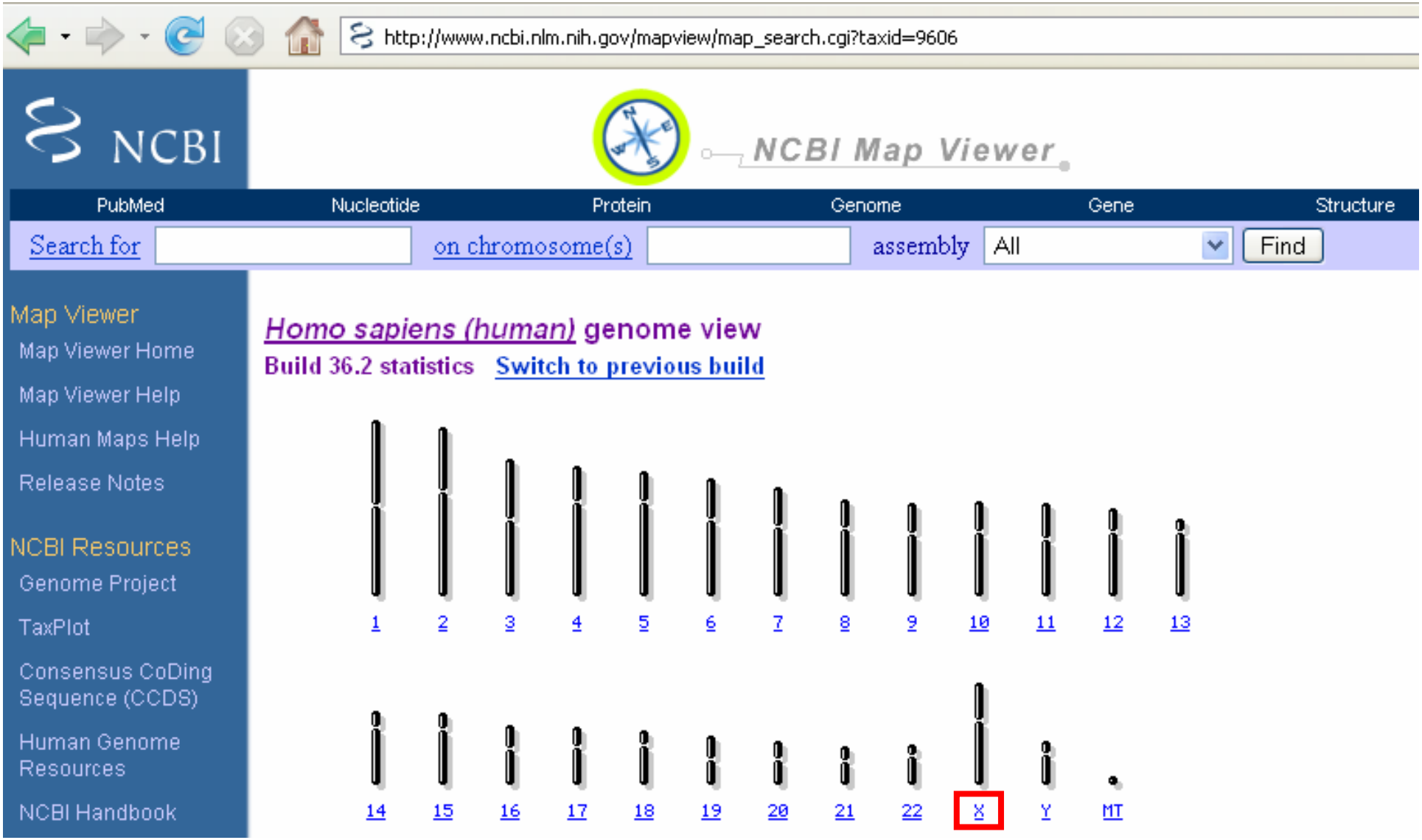
Caiman crocodylus	2 600 000 000	
Parascaris equorum	2 500 000 000	
Microtus pennsylvanicus	2 477 100 000	
Boa constrictor	2 100 000 000	
Danio rerio	1 900 000 000	
Aplysia californica	1 800 000 000	
Python reticulatus	1 700 000 000	
Cyprinus carpio	1 700 000 000	

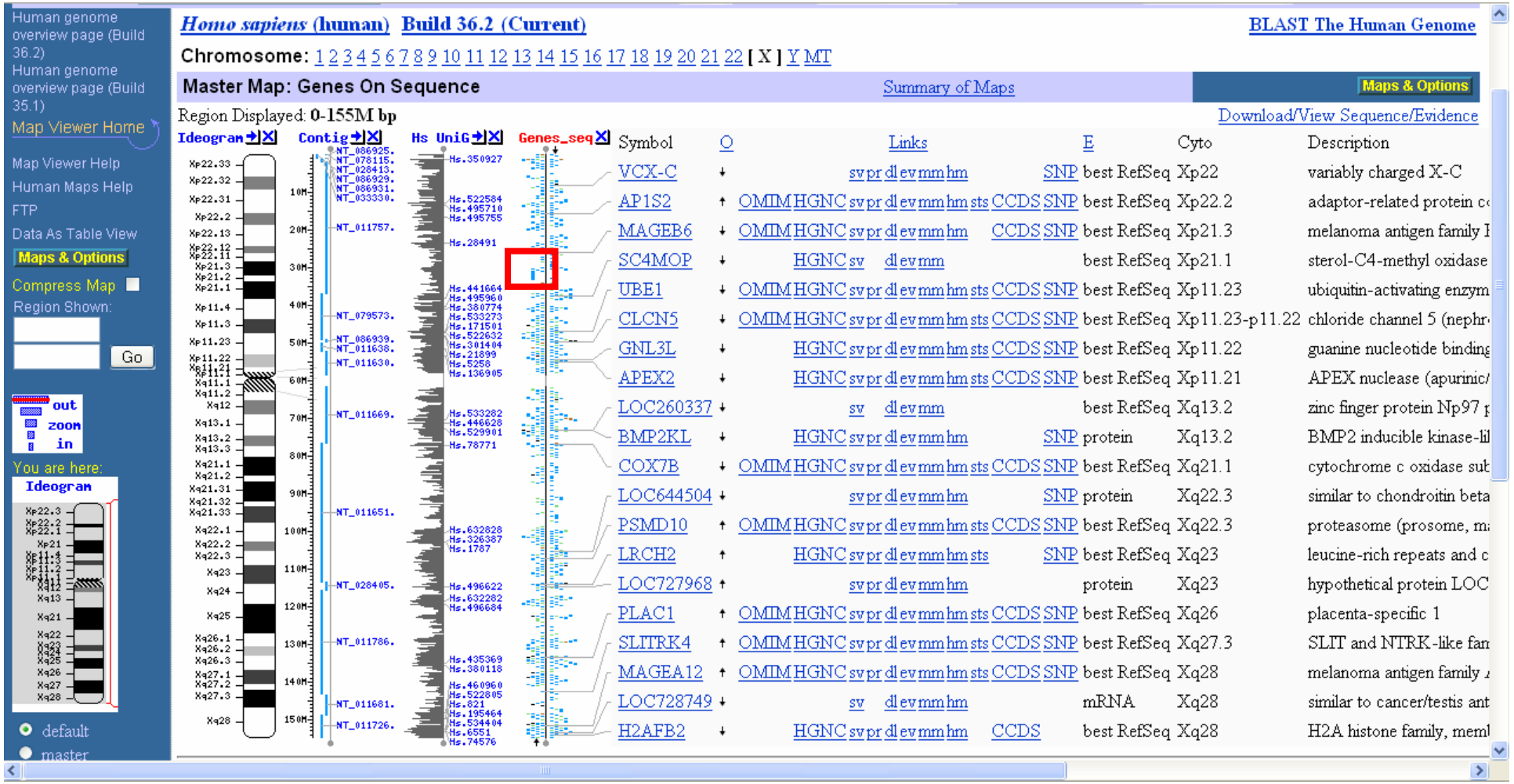
Escherichia coli	4 639 221	
Mycobacterium tuberculosis	4 397 000	
Mycobacterium leprae	2 800 000	
Haemophilus influenzae	1 830 137	
Helicobacter pylori	1 667 867	
Methanococcus jannaschii	1 664 974	
Borrelia garinii	953 000	
Mycoplasma genitalium	580 000	

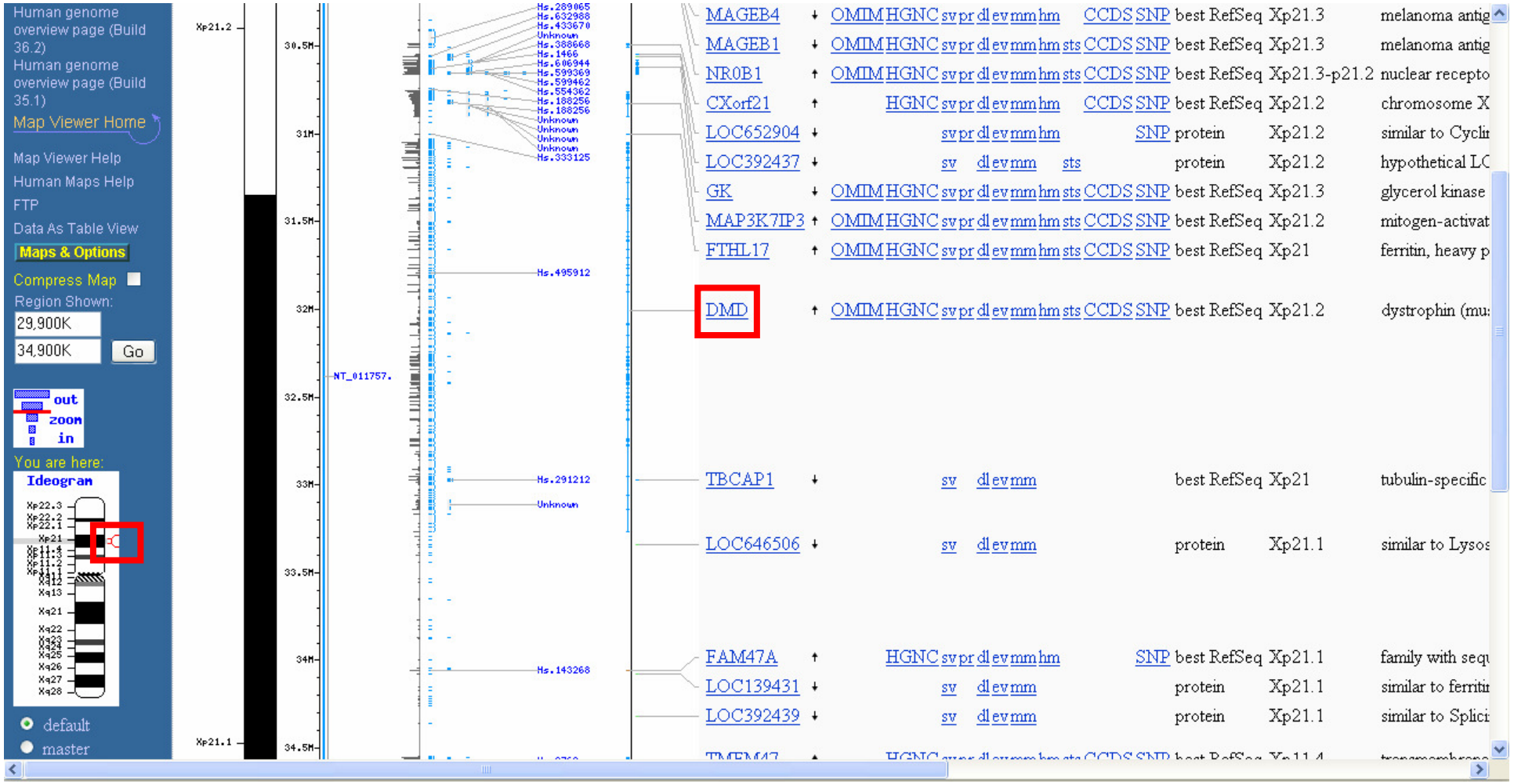

PubMed home - Microsoft Internet Explorer
Fájl Szerkesztés Nézet Kedvencek Eszközök Súgó
Vissza Keresés Kedvencek
Cím http://www.ncbi.nlm.nih.gov/pubmed/
Ugrás Hivatkozások
NCBI Resources How To
My NCBI Sign In
All Resources
Literature
DNA & RNA
Proteins
Sequence Analysis
Genes & Expression
Genomes & Maps
Domains & Structures
Genetics & Medicine
Taxonomy
Data & Software
Training & Tutorials
Homology
Small Molecules
Variation
Database of Genomic Structural Variation (dbVar)
Genome
Genome Project
Genome Workbench
Influenza Virus
Map Viewer
Nucleot
PopSet
ProSplign
Sequence Read Archive (SRA)
Splign
Trace Archive
UniSTS
All Genomes & Maps Resources...
Tool for searching, viewing and downloading genomic maps and related data
Limits Advanced search Help
PubMed
Search
Clear
U.S. Nationa
National Insti
illion citations for
and life science journals.
ext articles from PubMed
Using PubMe
PubMed Quick Start
New and Noteworthy
PubMed Tutorials
Full Text Articles
PubMed FAQs
Tools
on Matcher
n Matcher
ies
ic Queries
More Resources
MeSH Database
Journals Database
Clinical Trials
E-Utilities
LinkOut

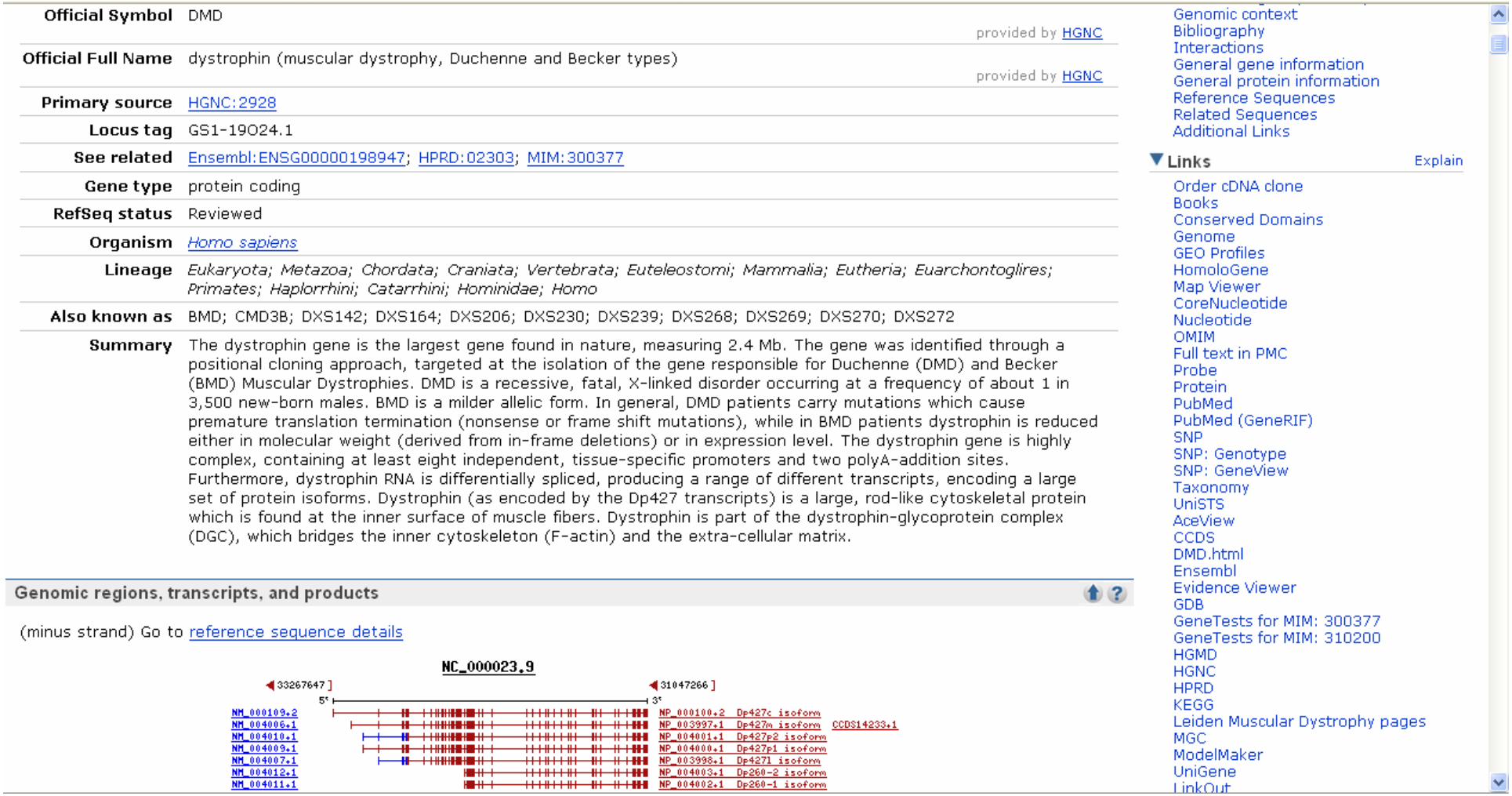

Map Viewer - Microsoft Internet Explorer
Fájl Szerkesztés Nézet Kedvencek Eszközök Súgó
Vissza Keresés Kedvencek
Cím http://www.ncbi.nlm.nih.gov/mapview/
Ugrás Hivatkozások
NCBI Home GenBank BLAST
Map Viewer Home
Help
The Map Viewer provides a wide variety of genome mapping and sequencing data. More..
Search
Search: Select Group or Organism
for:
Select Group or Organism
-- Groups --
Plants
Monocots
Eudicots
Fungi
-- Organisms --
--- Primates ---
Homo sapiens
Macaca mulatta
Pan troglodytes
--- Rodents ---
Mus musculus
Rattus norvegicus
--- Monotremes ---
Ornithorhynchus anatinus
--- Marsupials ---
Monodelphis domestica
--- Other Mammals ---
Bos taurus
Bubalus bubalis
Canis lupus familiaris
Equus caballus
Felis catus
Ovis aries
Sus scrofa
--- Other Vertebrates ---
Danio rerio
Gallus gallus
Taeniopygia guttata
Tools
News
Related
Genome Browser agreement
Genome Biology
Taxonomy
Entrez (Global Query)
BLAST
Map Viewer FTP
Small Genomes
Bacteria B
Organelles B
Viruses B
Vertebrates (17)
Mammals (14)
Primates (3)
Scientific name
Common name
Build
Tools
Homo sapiens
human
Build 37.1


Map Viewer - Microsoft Internet Explorer
Fájl Szerkesztés Nézet Kedvencek Eszközök Súgó
Vissza
Keresés
Kedvencek
Cím
http://www.ncbi.nlm.nih.gov/mapview/
NCBI
Home GenBank BLAST
Map Viewer Home
The Map Viewer provides a wide variety of genome mapping and sequencin
Search
Search: Homo sapiens
for:
Go
Tools Legend
Search or Browse the Genome
BLAST
Clone Finder
Go to region on a chromosome
Genome Resources page
News
Vertebrates
Mammals
Primates

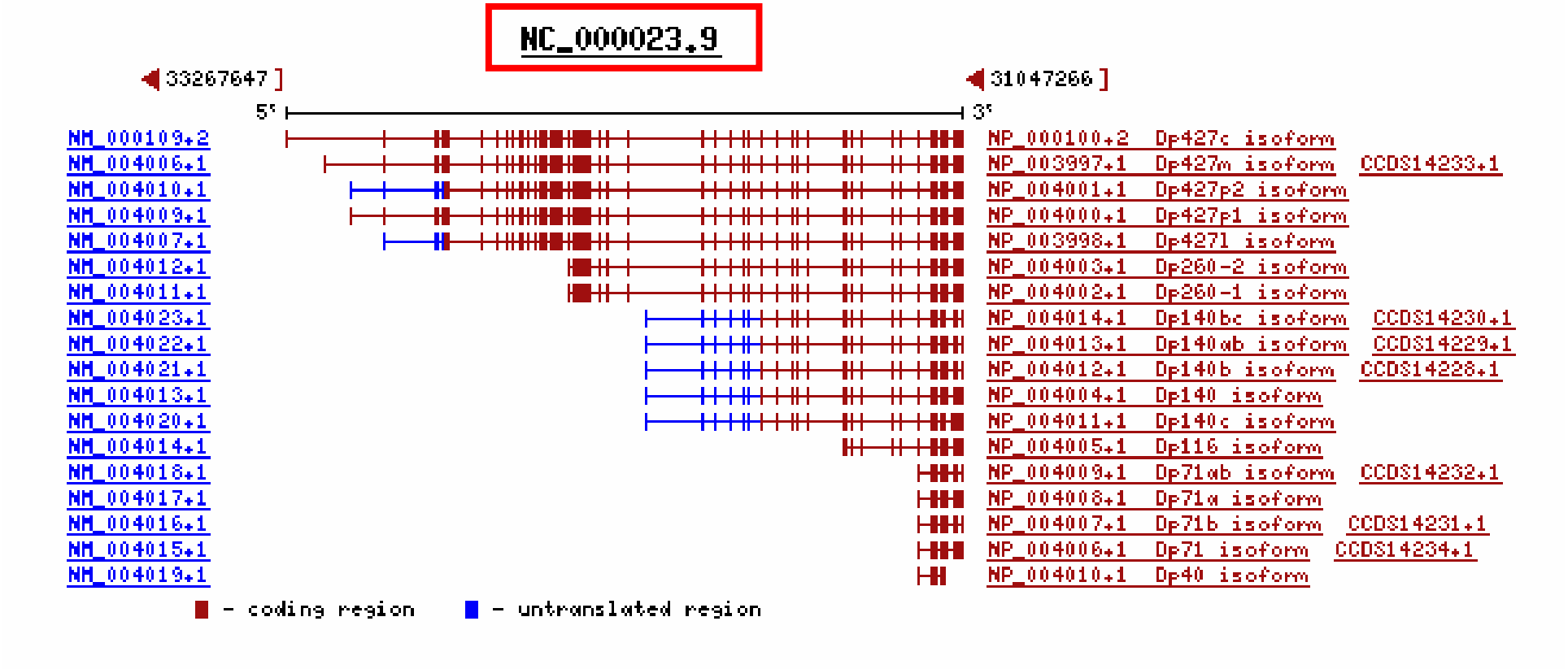

http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9606
NCBI
NCBI Map Viewer
PubMed
Nucleotide
Protein
Genome
Gene
Structure
Search for
on chromosome(s)
assembly
All
Find
Map Viewer
Map Viewer Home
Map Viewer Help
Human Maps Help
Release Notes
NCBI Resources
Genome Project
TaxPlot
Consensus CoDing Sequence (CCDS)
Human Genome Resources
NCBI Handbook
Homo sapiens (human) genome view
Build 36.2 statistics
Switch to previous build
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
MT


Human genome overview page (Build 36.2)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
Go
out
zoom
in
You are here:
Ideogram
default
master
Homo sapiens (human) Build 36.2 (Current)
BLAST The Human Genome
Chromosome: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 [X] Y MT
Master Map: Genes On Sequence
Summary of Maps
Maps & Options
Region Displayed: 0-155M bp
Download/View Sequence/Evidence
Ideogram
Contig
Hs UniG
Genes_seq
Symbol O Links E Cyto Description
VCX-C sv pr dl ev mm hm SNP best RefSeq Xp22 variably charged X-C
AP1S2 OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xp22.2 adaptor-related protein c
MAGEB6 OMIM HGNC sv pr dl ev mm hm CCDS SNP best RefSeq Xp21.3 melanoma antigen family I
SC4MOP HGNC sv dl ev mm best RefSeq Xp21.1 sterol-C4-methyl oxidase
UBE1 OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xp11.23 ubiquitin-activating enzym
CLCN5 OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xp11.23-p11.22 chloride channel 5 (nephr
GNL3L HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xp11.22 guanine nucleotide bindin
APEX2 HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xp11.21 APEX nuclease (apurinic/
LOC260337 sv dl ev mm best RefSeq Xq13.2 zinc finger protein Np97 p
BMP2KL HGNC sv pr dl ev mm hm SNP protein Xq13.2 BMP2 inducible kinase-li
COX7B OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xq21.1 cytochrome c oxidase sub
LOC644504 sv pr dl ev mm hm SNP protein Xq22.3 similar to chondroitin beta
PSMD10 OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xq22.3 proteasome (prosome, m
LRCH2 HGNC sv pr dl ev mm hm sts SNP best RefSeq Xq23 leucine-rich repeats and c
LOC727968 sv pr dl ev mm hm protein Xq23 hypothetical protein LOC
PLAC1 OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xq26 placenta-specific 1
SLITRK4 OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xq27.3 SLIT and NTRK-like fam
MAGEA12 OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xq28 melanoma antigen family
LOC728749 sv dl ev mm hm mRNA Xq28 similar to cancer/testis ant
H2AFB2 HGNC sv pr dl ev mm hm CCDS best RefSeq Xq28 H2A histone family, meml


Human genome overview page (Build 36.2)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
29,900K
34,900K
Go
out
zoom
in
You are here:
Ideogram
default
master
Xp21.2
Xp21.1
30.5M
31M
31.5M
32M
32.5M
33M
33.5M
34M
34.5M
NT_011757.
Hs.289065
Hs.632988
Hs.433670
Unknown
Hs.388668
Hs.1466
Hs.606944
Hs.599369
Hs.599462
Hs.554362
Hs.188256
Hs.188256
Unknown
Unknown
Unknown
Unknown
Hs.333125
Hs.495912
Hs.291212
Unknown
Hs.143268
MAGEB4 OMIM HGNC sv pr dl ev mm hm CCDS SNP best RefSeq Xp21.3 melanoma antig
MAGEB1 OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xp21.3 melanoma antig
NR0B1 OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xp21.3-p21.2 nuclear recepto
CXorf21 HGNC sv pr dl ev mm hm CCDS SNP best RefSeq Xp21.2 chromosome X
LOC652904 sv pr dl ev mm hm SNP protein Xp21.2 similar to Cyclir
LOC392437 sv dl ev mm sts protein Xp21.2 hypothetical LC
GK OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xp21.3 glycerol kinase
MAP3K7IP3 OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xp21.2 mitogen-activat
FTHL17 OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xp21 ferritin, heavy p
DMD OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq Xp21.2 dystrophin (mu
TBCAP1 sv dl ev mm best RefSeq Xp21 tubulin-specific
LOC646506 sv dl ev mm protein Xp21.1 similar to Lysos
FAM47A HGNC sv pr dl ev mm hm SNP best RefSeq Xp21.1 family with sequ
LOC139431 sv dl ev mm protein Xp21.1 similar to ferritir
LOC392439 sv dl ev mm protein Xp21.1 similar to Splici



There is an enormous amount of information here.


NC_000023.9
33267647
31047266
5'
3'
NM_000109.2 NP_000100.2 Dp427c isoform
NM_004006.1 NP_003997.1 Dp427m isoform CCDS14233.1
NM_004010.1 NP_004001.1 Dp427p2 isoform
NM_004009.1 NP_004000.1 Dp427p1 isoform
NM_004007.1 NP_003998.1 Dp427l isoform
NM_004012.1 NP_004003.1 Dp260-2 isoform
NM_004011.1 NP_004002.1 Dp260-1 isoform
NM_004023.1 NP_004014.1 Dp140bc isoform CCDS14230.1
NM_004022.1 NP_004013.1 Dp140ab isoform CCDS14229.1
NM_004021.1 NP_004012.1 Dp140b isoform CCDS14228.1
NM_004013.1 NP_004004.1 Dp140 isoform
NM_004020.1 NP_004011.1 Dp140c isoform
NM_004014.1 NP_004005.1 Dp116 isoform
NM_004018.1 NP_004009.1 Dp71ab isoform CCDS14232.1
NM_004017.1 NP_004008.1 Dp71a isoform
NM_004016.1 NP_004007.1 Dp71b isoform CCDS14231.1
NM_004015.1 NP_004006.1 Dp71 isoform CCDS14234.1
NM_004019.1 NP_004010.1 Dp40 isoform
- coding region
- untranslated region

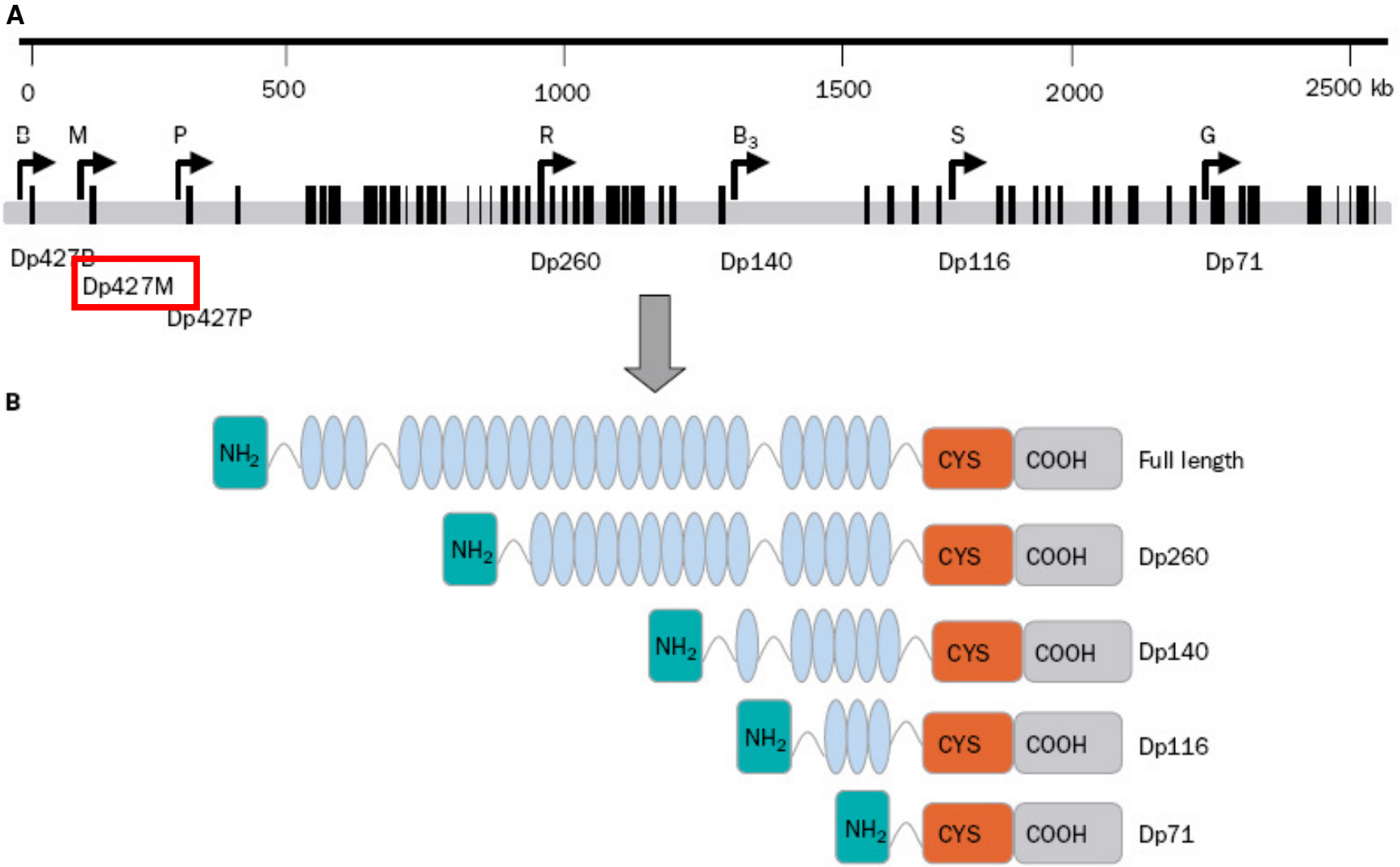


Figure 1. A: Genomic organisation of the dystrophin gene, located in Xp21. The black vertical lines represent the 79 exons of the dystrophin gene distributed over about 2·5 million bases. The arrows indicate the various promoters: in particular are brain (B), muscle (M), and Purkinje (P) promoters; R, B_3, S, and G represent the Dp260 (retinal), Dp140 (brain), Dp116 (Schwann cells), and Dp71 (general) promoters. B: The domain composition of the various dystrophin proteins is indicated. The amino-terminal domain is followed by the spectrin like domain, the cysteine rich, and the carboxy-terminal domain.

Francesco Muntoni, Silvia Torelli, and Alessandra Ferlini, *Lancet Neurol.*, 2003, 2, 731-740.

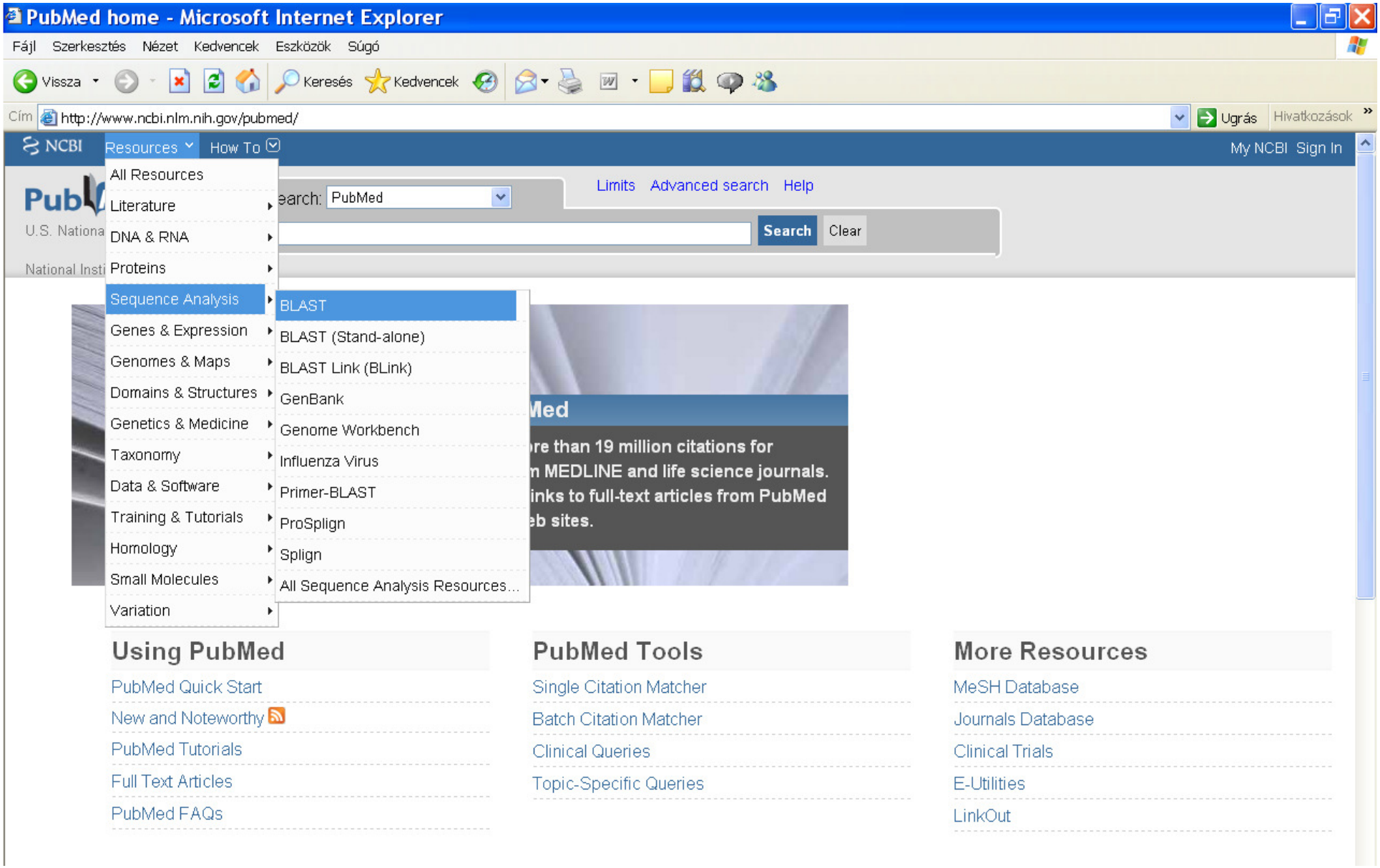
6. Feladat:

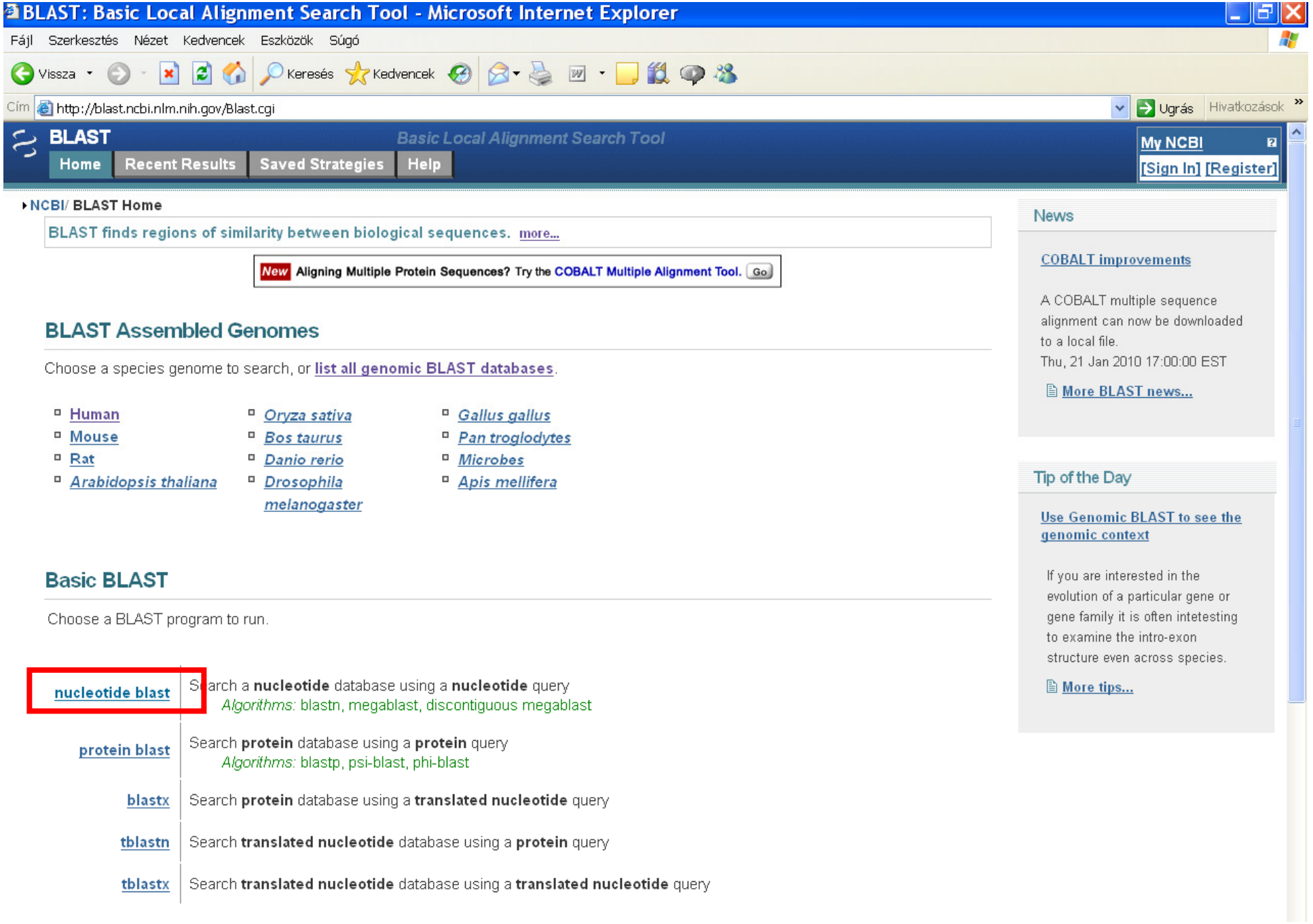
A Duchenne-féle izomsorvadásos betegségért felelős DMD gén az X-kromoszóma 21-es lókuszán található. A betegség általában egy-egy kódoló régió hiányának köszönhető, aminek következtében a gén által kódolt disztrofin fehérje nem termelődik, ezért az izomsejtek fokozatosan elhalnak. A humán genom projektnek köszönhetően a gén teljes szekvenciája ismert. Szeretnénk kideríteni, hogy a gén 1. exonja (szekvencia a következő oldalon) megvan-e a páciens DNS-ében? Ehhez szükség lenne az 1. exon szekvenciájából egy olyan 30 bázispárból álló részletre, mely egyedi a humán genomban. Ez annyit jelent, hogy a humán genomban a leghosszabb nemspecifikus kötődés(ek)re jellemző olvadáspont(ok) legalább 15°C-kal legyen(ek) kisebb(ek) mint a 30 bázispáros marker olvadáspontja.

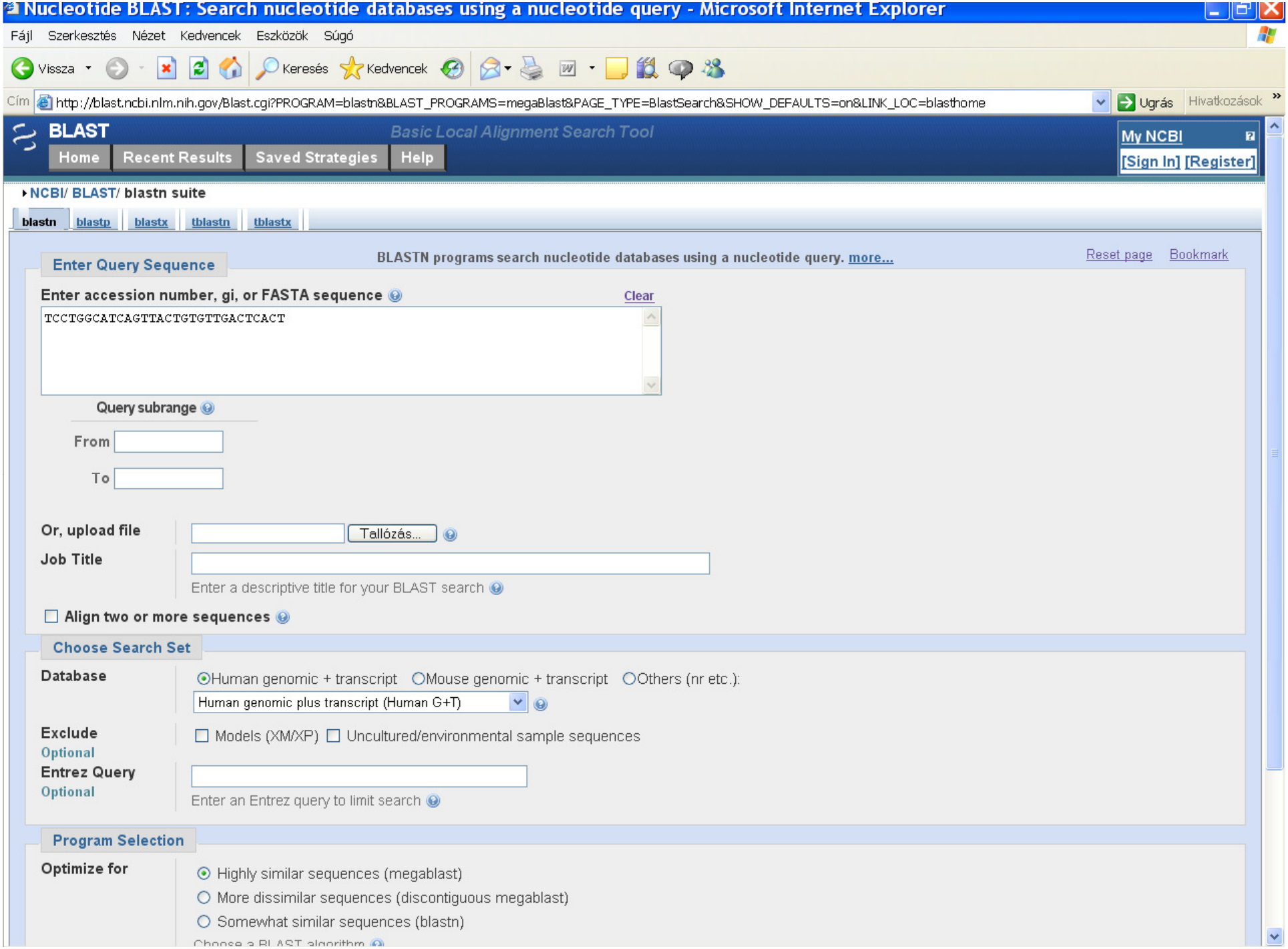
Keressen egy ilyen markert és szekvenciáját, valamint jellemzését küldje el e-mailben a gyurcsik@chem.u-szeged.hu címre! A file neve a monogramja és egy utána írt 6.doc legyen.

A DMD gén 1. exonjának DNS szekvenciája.

TCCTGGCATCAGTTACTGTGTTGACTCACTCAGTGTTGGGATCACTCAC
TTTCCCCCTACAGGACTCAGATCTGGGAGGCAATTACCTTCGGAGAAAA
ACGAATAGGAAAAACTGAAGTGTTACTTTTTTTTAAAGCTGCTGAAGTTT
GTTGGTTTCTCATTGTTTTTTAAGCCTACTGGAGCAATAAAGTTTGAAGA
ACTTTTACCAGGTTTTTTTTTATCGCTGCCTTGATATACACTTTTCAAAA
TGCTTTGGTGGGAAGAAGTAGAGGACTGTT

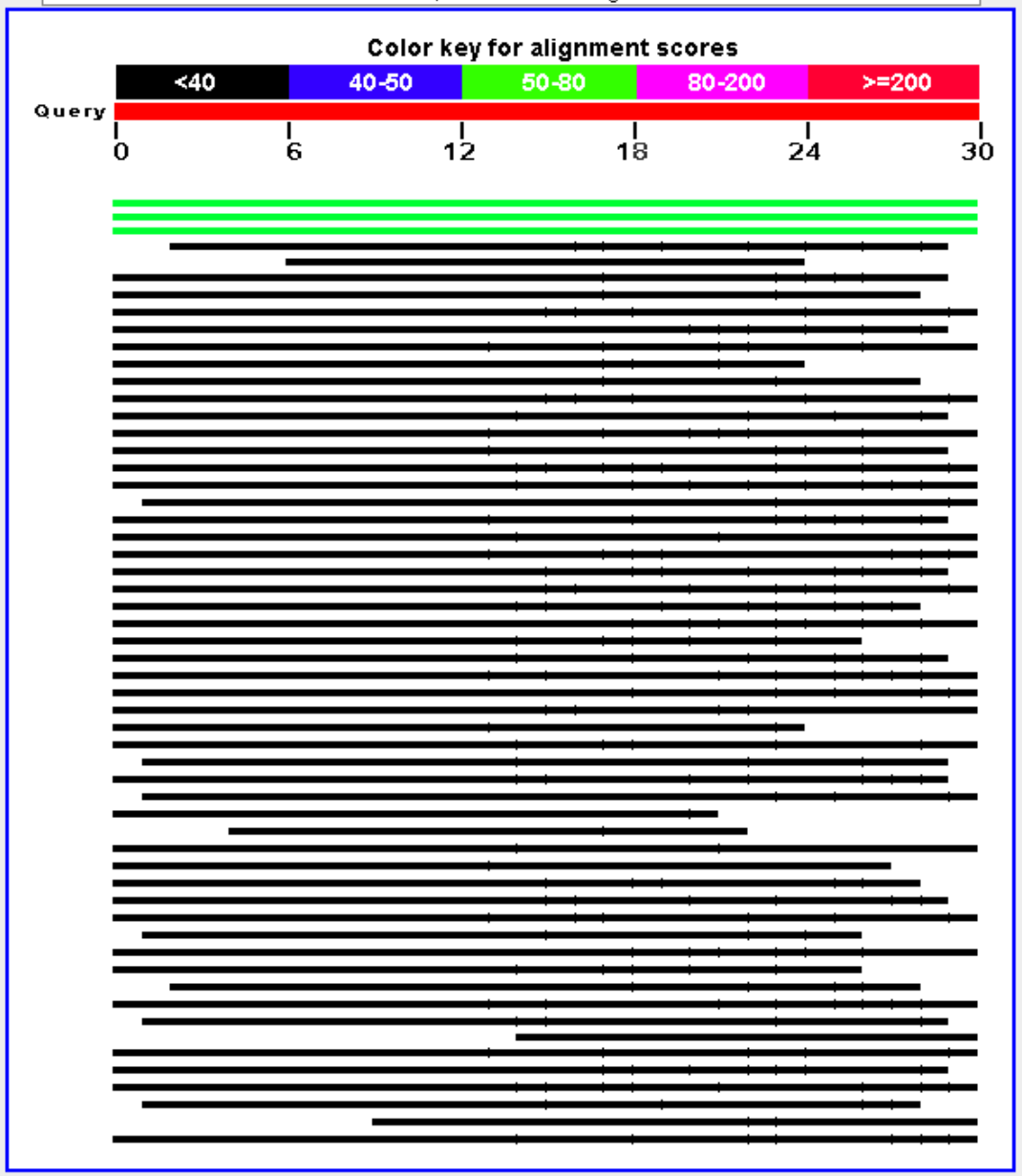

PubMed home - Microsoft Internet Explorer
Fájl Szerkesztés Nézet Kedvencek Eszközök Súgó
Vissza Keresés Kedvencek
Cím http://www.ncbi.nlm.nih.gov/pubmed/
Ugrás Hivatkozások
NCBI Resources How To
My NCBI Sign In
All Resources
Literature
DNA & RNA
Proteins
Sequence Analysis
Genes & Expression
Genomes & Maps
Domains & Structures
Genetics & Medicine
Taxonomy
Data & Software
Training & Tutorials
Homology
Small Molecules
Variation
BLAST
BLAST (Stand-alone)
BLAST Link (BLink)
GenBank
Genome Workbench
Influenza Virus
Primer-BLAST
ProSplign
Splign
All Sequence Analysis Resources...
Limits Advanced search Help
PubMed
Search
Clear
Using PubMed
PubMed Quick Start
New and Noteworthy
PubMed Tutorials
Full Text Articles
PubMed FAQs
PubMed Tools
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
Topic-Specific Queries
More Resources
MeSH Database
Journals Database
Clinical Trials
E-Utilities
LinkOut


BLAST: Basic Local Alignment Search Tool - Microsoft Internet Explorer
Fájl Szerkesztés Nézet Kedvencek Eszközök Súgó
Vissza Keresés Kedvencek
Cím http://blast.ncbi.nlm.nih.gov/Blast.cgi
Ugrás Hivatkozások
BLAST
Basic Local Alignment Search Tool
Home
Recent Results
Saved Strategies
Help
My NCBI
[Sign In] [Register]
NCBI/ BLAST Home
BLAST finds regions of similarity between biological sequences. more...
New Aligning Multiple Protein Sequences? Try the COBALT Multiple Alignment Tool. Go
BLAST Assembled Genomes
Choose a species genome to search, or list all genomic BLAST databases.
Human
Mouse
Rat
Arabidopsis thaliana
Oryza sativa
Bos taurus
Danio rerio
Drosophila melanogaster
Gallus gallus
Pan troglodytes
Microbes
Apis mellifera
Basic BLAST
Choose a BLAST program to run.
nucleotide blast
Search a nucleotide database using a nucleotide query
Algorithms: blastn, megablast, discontiguous megablast
protein blast
Search protein database using a protein query
Algorithms: blastp, psi-blast, phi-blast
blastx
Search protein database using a translated nucleotide query
tblastn
Search translated nucleotide database using a protein query
tblastx
Search translated nucleotide database using a translated nucleotide query
News
COBALT improvements
A COBALT multiple sequence alignment can now be downloaded to a local file.
Thu, 21 Jan 2010 17:00:00 EST
More BLAST news...
Tip of the Day
Use Genomic BLAST to see the genomic context
If you are interested in the evolution of a particular gene or gene family it is often intetesting to examine the intro-exon structure even across species.
More tips...


Nucleotide BLAST: Search nucleotide databases using a nucleotide query - Microsoft Internet Explorer
Fájl Szerkesztés Nézet Kedvencek Eszközök Súgó
Vissza Keresés Kedvencek
Cím http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome
Ugrás Hivatkozások
BLAST
Basic Local Alignment Search Tool
Home Recent Results Saved Strategies Help
My NCBI
[Sign In] [Register]
NCBI/ BLAST/ blastn suite
blastn blastp blastx tblastn tblastx
BLASTN programs search nucleotide databases using a nucleotide query. more...
Reset page Bookmark
Enter Query Sequence
Enter accession number, gi, or FASTA sequence
Clear
TCCTGGCATCAGTTACTGTGTTGACTCACT
Query subrange
From
To
Or, upload file
Tallózás...
Job Title
Enter a descriptive title for your BLAST search
Align two or more sequences
Choose Search Set
Database
Human genomic + transcript
Mouse genomic + transcript
Others (nr etc.):
Human genomic plus transcript (Human G+T)
Exclude
Optional
Models (XM/XP)
Uncultured/environmental sample sequences
Entrez Query
Optional
Enter an Entrez query to limit search
Program Selection
Optimize for
Highly similar sequences (megablast)
More dissimilar sequences (discontiguous megablast)
Somewhat similar sequences (blastn)

Distribution of 1799 Blast Hits on the Query Sequence

Mouse-over to show defline and scores, click to show alignments



Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
Transcripts							
NM_004006.1	Homo sapiens dystrophin (DMD), transcript variant Dp427m, mRNA	60.0	60.0	100%	6e-08	100%	G M
Genomic sequences [show first]							
NT_167197.1	Homo sapiens chromosome X genomic contig, GRCh37 reference pri	60.0	518	100%	6e-08	100%	
NW_001842360.1	Homo sapiens chromosome X genomic contig, alternate assembly (b	60.0	409	100%	6e-08	100%	
NT_023133.13	Homo sapiens chromosome 5 genomic contig, GRCh37 reference pri	36.2	498	90%	0.91	100%	
NW_001838960.2	Homo sapiens chromosome 5 genomic contig, alternate assembly (b	36.2	36.2	60%	0.91	100%	
NT_011651.17	Homo sapiens chromosome X genomic contig, GRCh37 reference pri	34.2	543	96%	3.6	100%	
NT_079573.4	Homo sapiens chromosome X genomic contig, GRCh37 reference pri	34.2	171	93%	3.6	100%	
NT_011387.8	Homo sapiens chromosome 20 genomic contig, GRCh37 reference p	34.2	543	100%	3.6	100%	
NT_030059.13	Homo sapiens chromosome 10 genomic contig, GRCh37 reference p	34.2	1340	96%	3.6	100%	
NT_005403.17	Homo sapiens chromosome 2 genomic contig, GRCh37 reference pri	34.2	1443	100%	3.6	100%	
NW_001842386.2	Homo sapiens chromosome X genomic contig, alternate assembly (b	34.2	220	80%	3.6	100%	
NW_001842361.2	Homo sapiens chromosome X genomic contig, alternate assembly (b	34.2	145	93%	3.6	100%	
NW_001838652.1	Homo sapiens chromosome 20 genomic contig, alternate assembly (	34.2	409	100%	3.6	100%	
NW_001837986.1	Homo sapiens chromosome 10 genomic contig, alternate assembly (	34.2	276	96%	3.6	100%	
NW_001838860.1	Homo sapiens chromosome 2 genomic contig, alternate assembly (b	34.2	434	100%	3.6	100%	
NT_025028.14	Homo sapiens chromosome 18 genomic contig, GRCh37 reference p	32.2	300	96%	14	100%	
NT_026437.12	Homo sapiens chromosome 14 genomic contig, GRCh37 reference p	32.2	1193	100%	14	100%	
NT_029419.12	Homo sapiens chromosome 12 genomic contig, GRCh37 reference p	32.2	1352	100%	14	100%	
NT_009237.18	Homo sapiens chromosome 11 genomic contig, GRCh37 reference p	32.2	589	96%	14	100%	
NT_008470.19	Homo sapiens chromosome 9 genomic contig, GRCh37 reference pri	32.2	997	96%	14	100%	

>ref|NT_023133.13| Homo sapiens chromosome 5 genomic contig, GRCh37 reference primary assembly
Length=25716533

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Features in this part of subject sequence:
hypothetical protein LOC375484

Score = 36.2 bits (18), Expect = 0.91
Identities = 18/18 (100%), Gaps = 0/18 (0%)
Strand=Plus/Plus

```
Query  7         CATCAGTTACTGTGTTGA  24
                 ||||||||||||||||||
Sbjct  20573499  CATCAGTTACTGTGTTGA  20573516
```

Features flanking this part of subject sequence:
14909 bp at 5' side: hypothetical protein LOC202134
47766 bp at 3' side: hypothetical protein isoform 1

Score = 30.2 bits (15), Expect = 56
Identities = 15/15 (100%), Gaps = 0/15 (0%)
Strand=Plus/Plus

```
Query  10        CAGTTACTGTGTTGA  24
                 |||||||||||||||
Sbjct  20367234  CAGTTACTGTGTTGA  20367248
```

Features in this part of subject sequence:
delta-sarcoglycan isoform 1
delta-sarcoglycan isoform 2

Score = 28.2 bits (14), Expect = 221
Identities = 14/14 (100%), Gaps = 0/14 (0%)
Strand=Plus/Minus

```
Query  7       CATCAGTTACTGTG  20
               ||||||||||||||
Sbjct  958772  CATCAGTTACTGTG  958759
```

Features flanking this part of subject sequence:
62771 bp at 5' side: hypothetical protein LOC408263