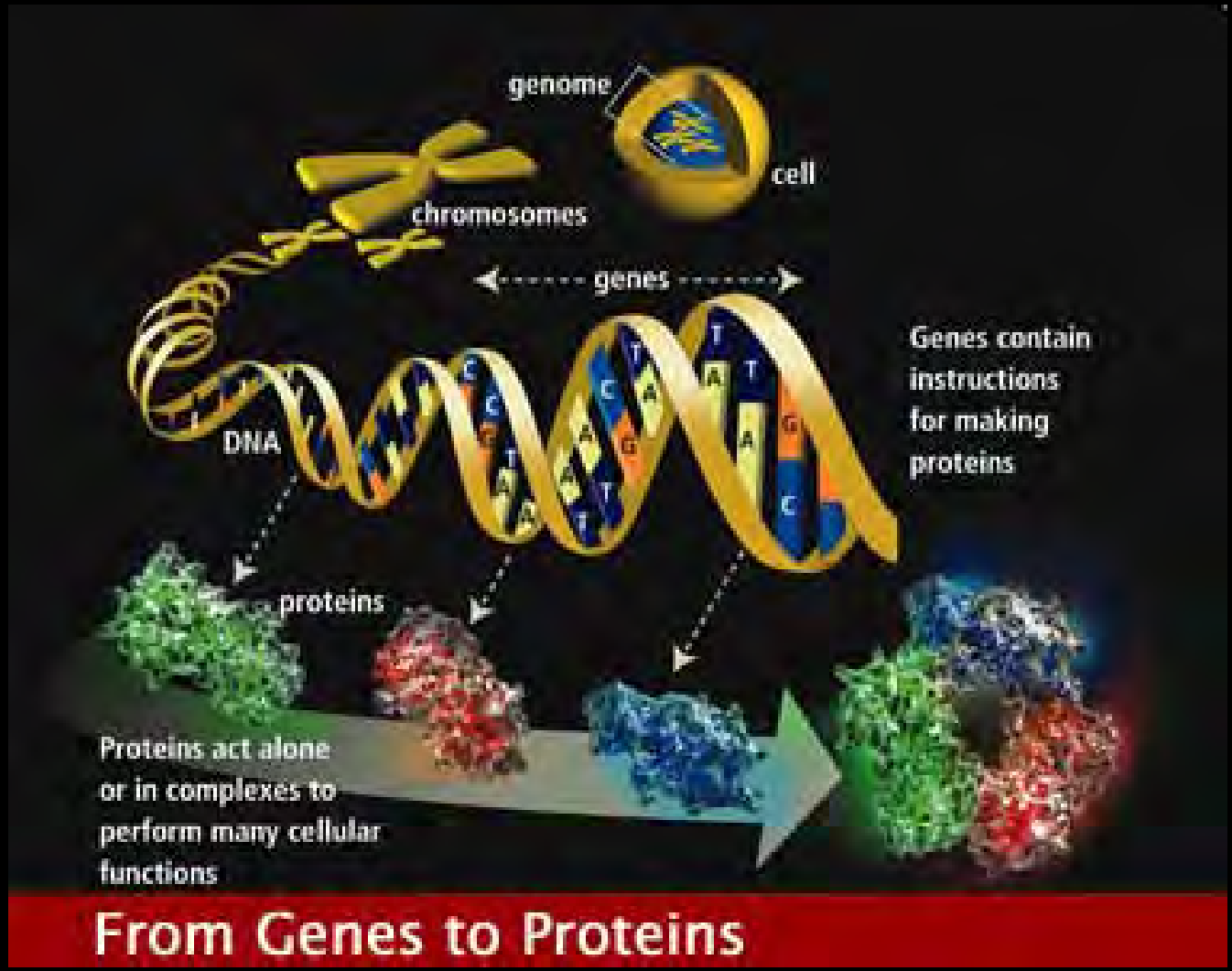
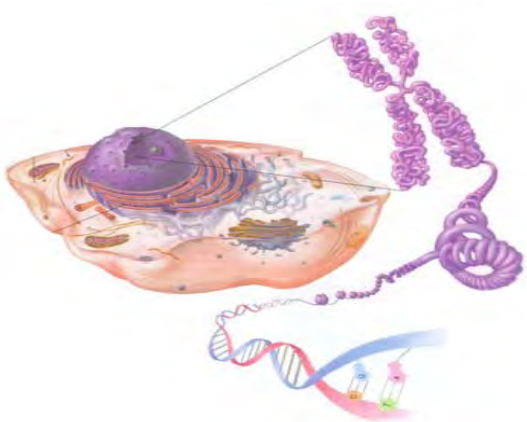
Od gena do genoma...

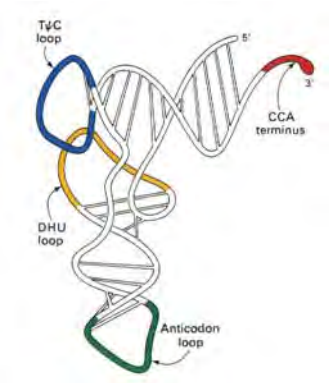


....od genoma do proteoma

Genom, Transkriptom, Proteom



Genom – ukupna genetska informacija organizma, predstavlja celokupnu sekvencu baza u DNK (ili, kod mnogih virusa, RNK)
Gen - određeni deo DNK koji nosi informaciju za RNK ili protein.



Transkriptom - ukupne iRNK eksprimirane u ćeliji.

Proteom – svi proteini eksprimirani u ćeliji.

"Kad bih vam iščitavao genom brzinom od jedne reči u sekundi tokom osam sati dnevno, bio bi mi potreban ceo vek. Kad bih ispisao ljudski genom slovima veličine jednog milimetra, moj bi tekst bio dužine reke Dunav. Divovski je to dokument, golema knjiga, tekst neobične dužine koji stane u jedro mikroskopske veličine u ćeliji tako sićušnoj da bi se mogla smestiti u glavicu čiode."

Met Ridley

(Iz knjige Genom, Autobiografija vrste u 23 poglavlja, kod nas postoji u prevodu prof. Gordana Matić, Plato,Beograd, 2001)

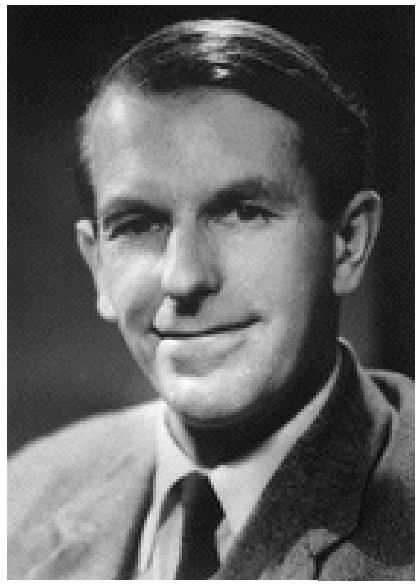
Sekvenciranje DNK

AGTCCGCGAATACAGGCTCGGT

Frederic Sanger

Nobelova nagrada za hemiju 1980:

"for their contributions concerning the determination of base sequences in nucleic acids"



Proc. Natl. Acad. Sci. USA
Vol. 74, No. 12, pp. 5463–5467, December 1977
Biochemistry

DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage ϕX174)

F. SANGER, S. NICKLEN, AND A. R. COULSON

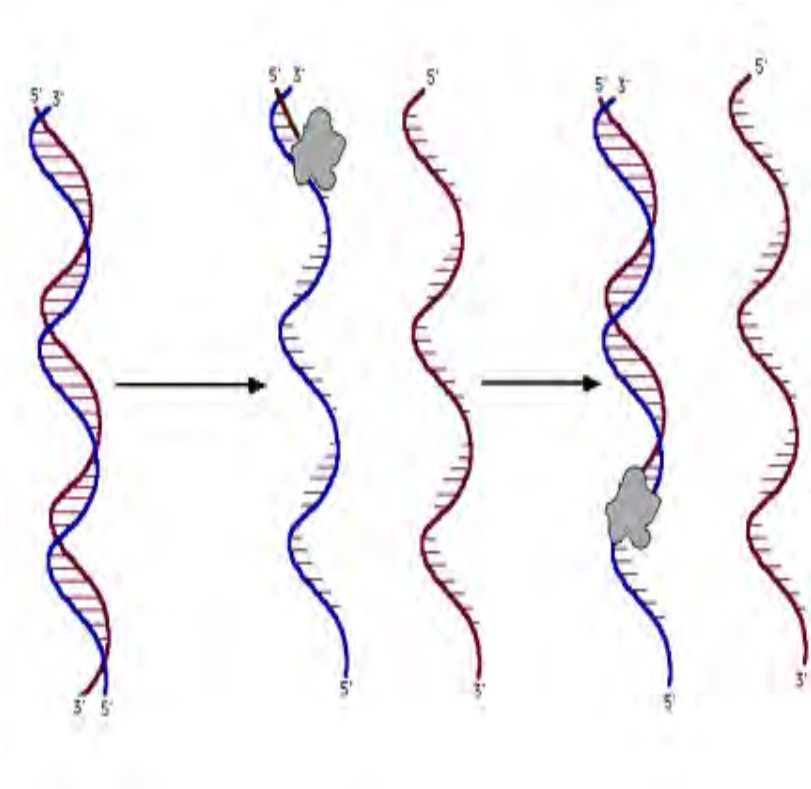
Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Sanger-ova metoda za sekvenciranje DNK

- **Reakcije sekvenciranja DNK slične PCR reakcijama!**

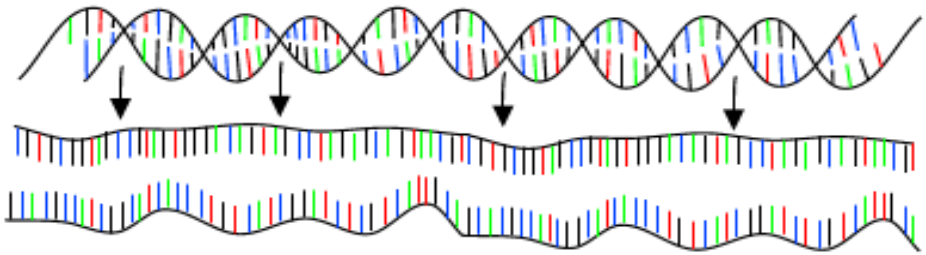
- **U reakcionoj smeši se nalaze: templatna DNK, Taq polimeraza, dNTP, 5% ddNTP-a, i "primer" (malo parče jednolančane DNK, 20-30 nukleotida, koja se hibridizuje sa templatnom DNK).**

- **Reakcija započinje zagrevanjem ....**



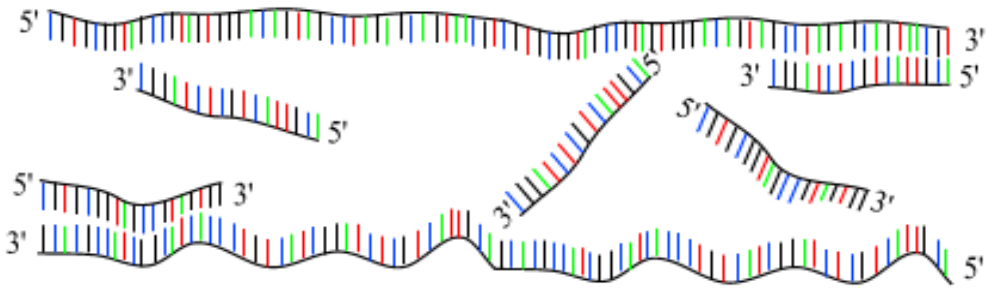
PCR : Polymerase Chain Reaction

30 - 40 cycles of 3 steps :



Step 1 : denaturation

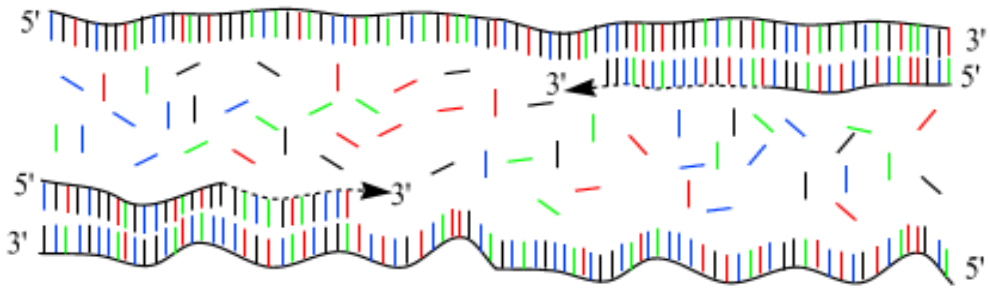
1 minut 94 °C



Step 2 : annealing

45 seconds 54 °C

forward and reverse primers !!!



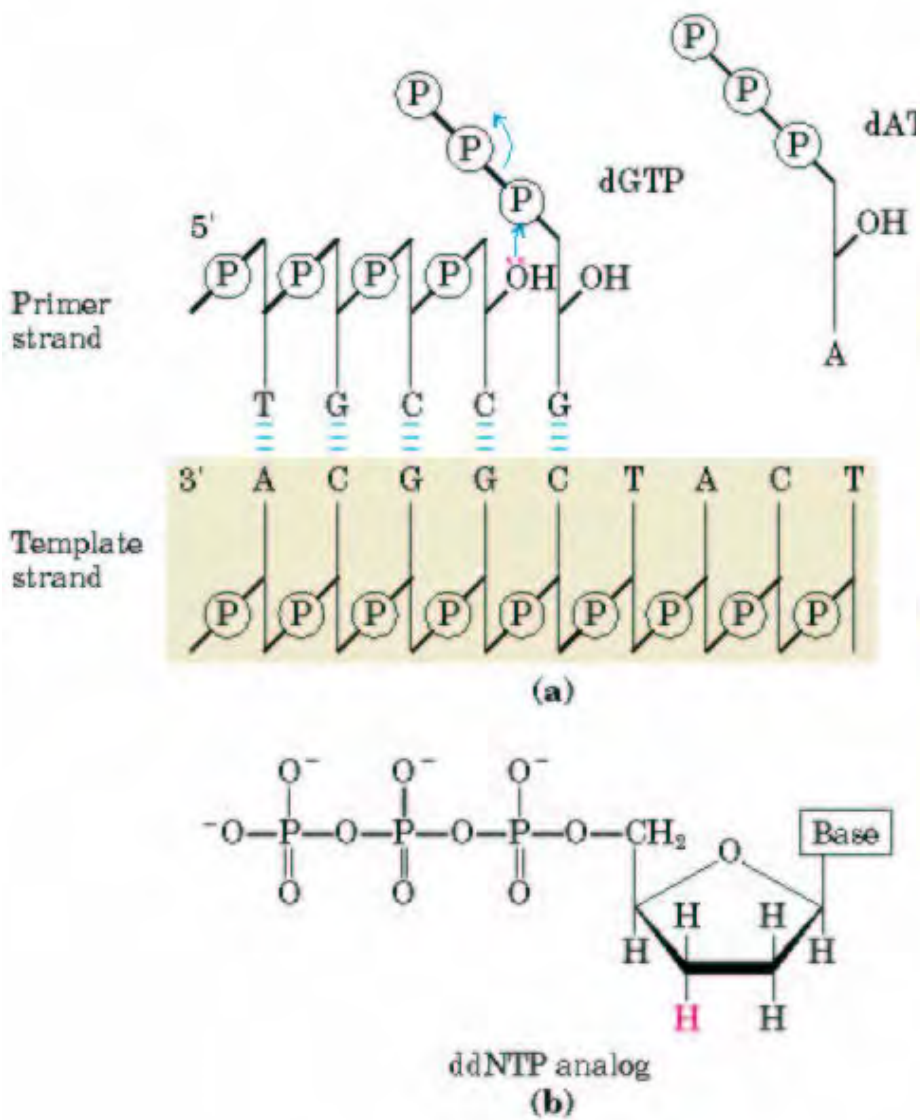
Step 3 : extension

2 minutes 72 °C

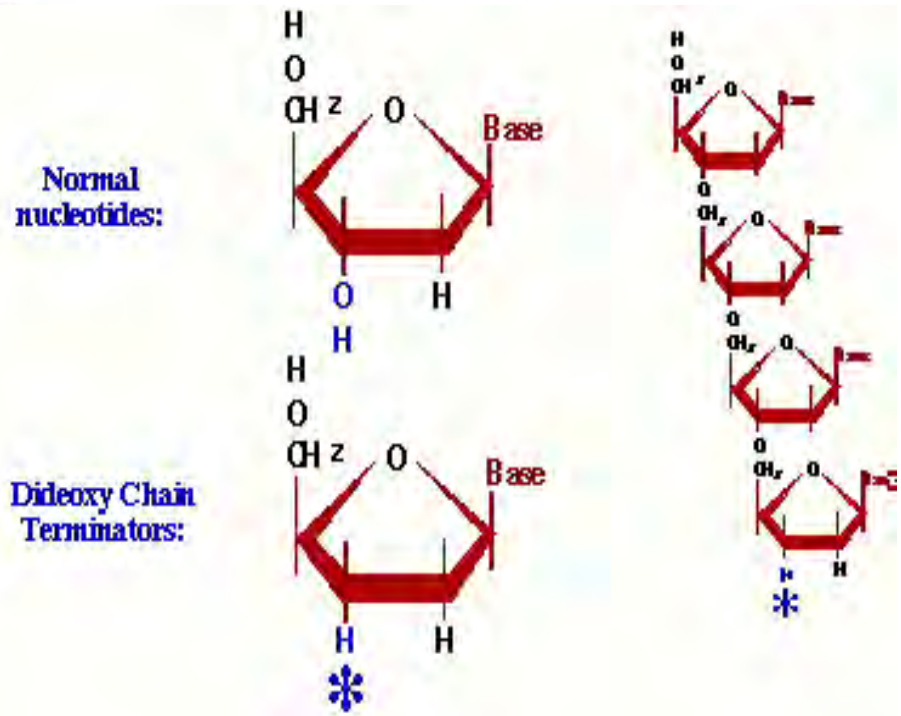
only dNTP's

(Andy Vierstraete 1999)

Sekvenciranje DNK Sanger-ovom metodom



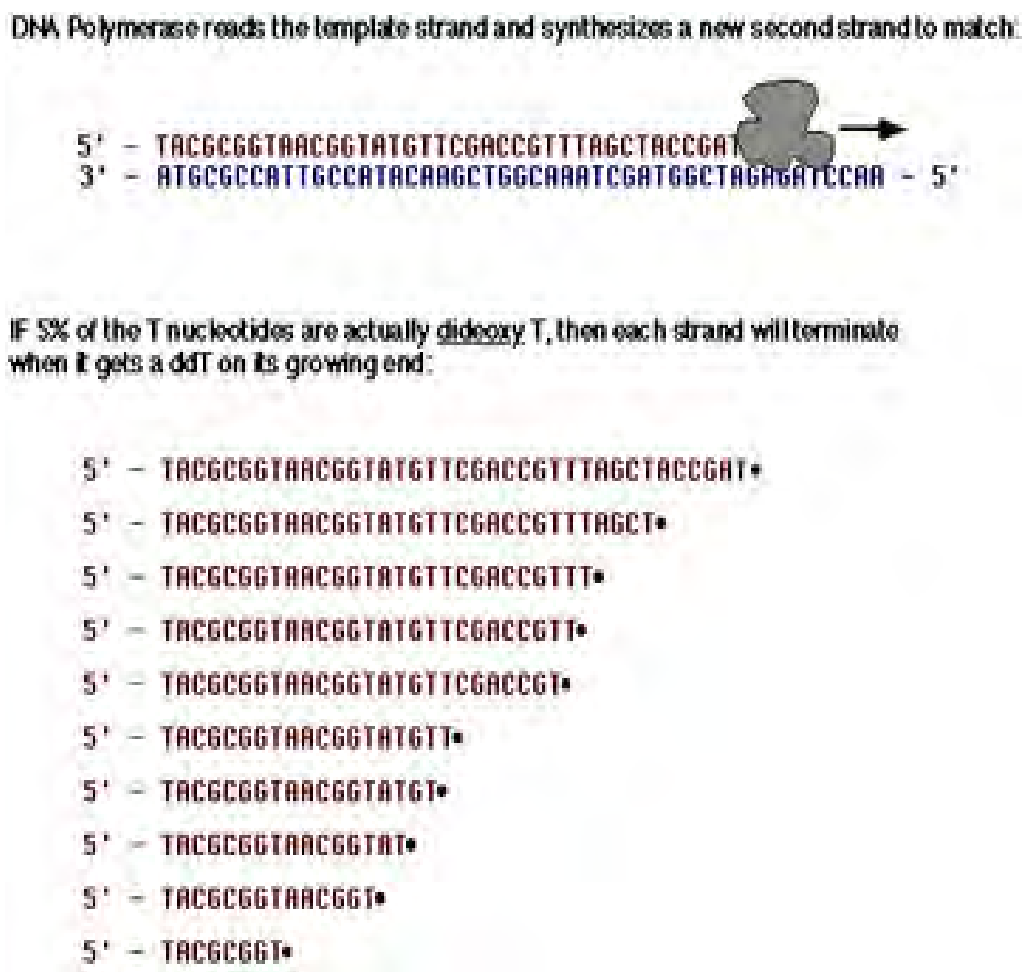
dNTP



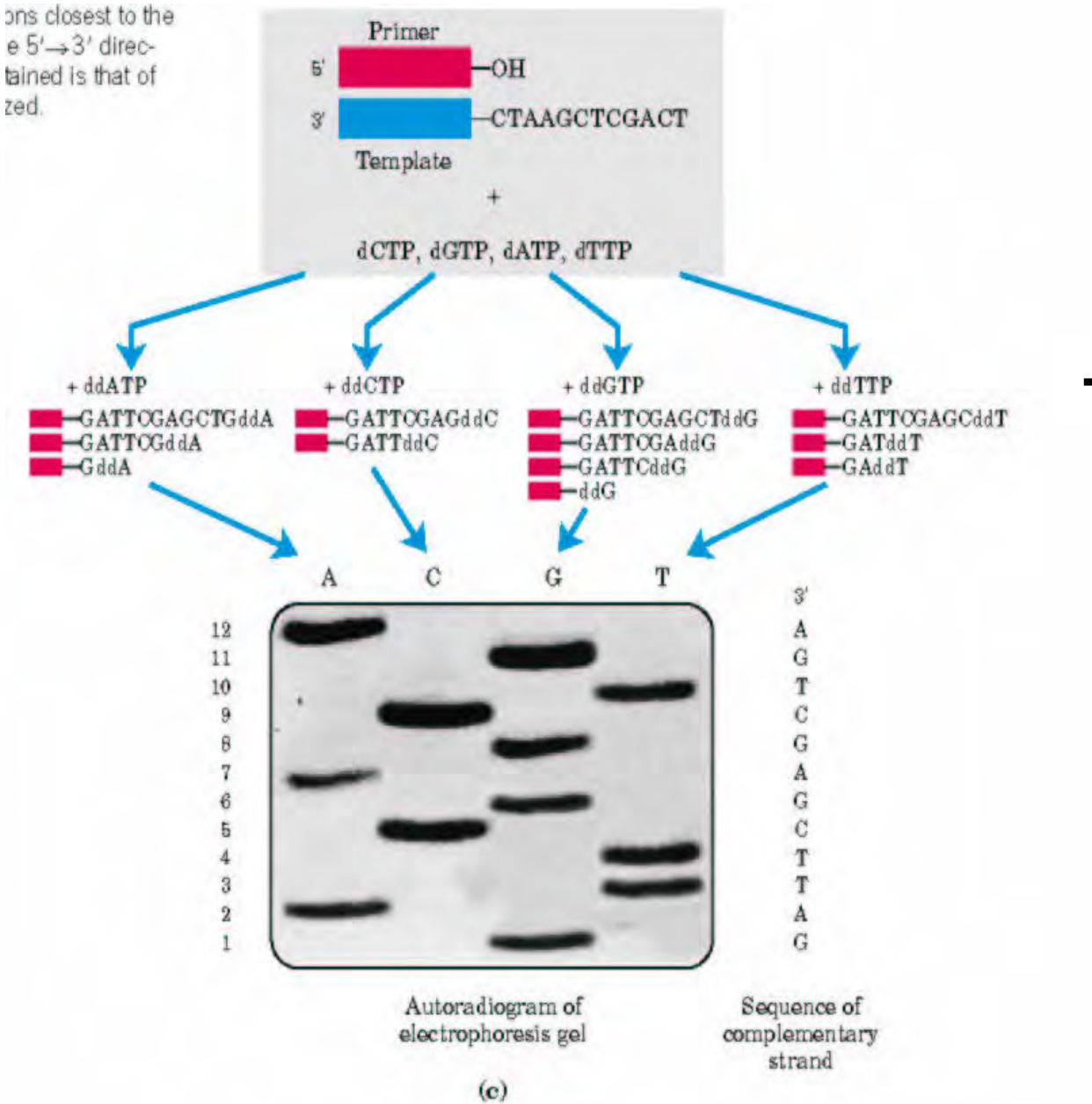
ddNTP

DNK replikacija u prisustvu 5% ddNTP-a

- U reakcionoj smeši se nalazi pored sva 4 dNTP-a, 5% ddNTP-a (primer ddTTP na slici desno!)
- Polimerizacija će se završiti kada se dodati ddNTP nađe na rastućem kraju (replikacija ide u smeru 5’ ka 3’).

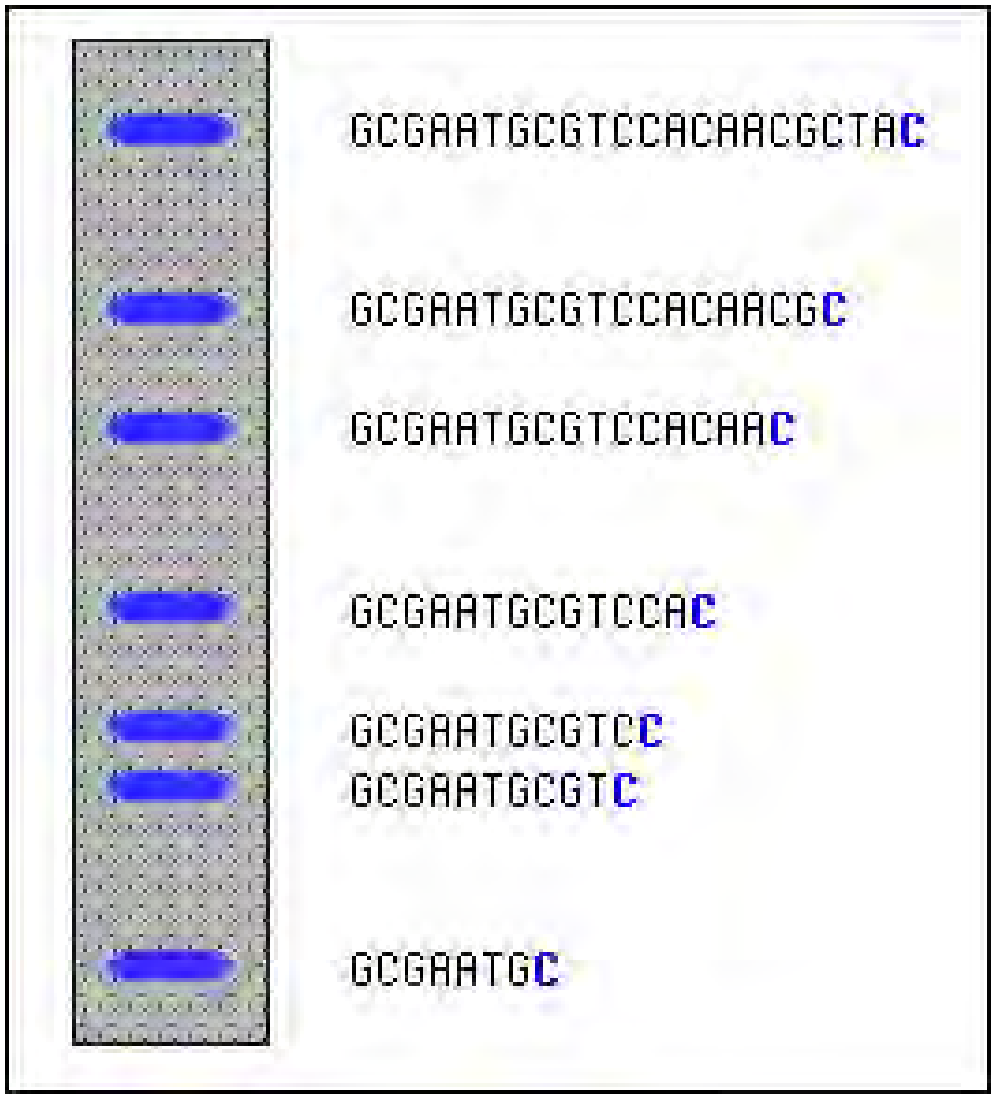


Sekvenciranje DNK Sanger-ovom metodom

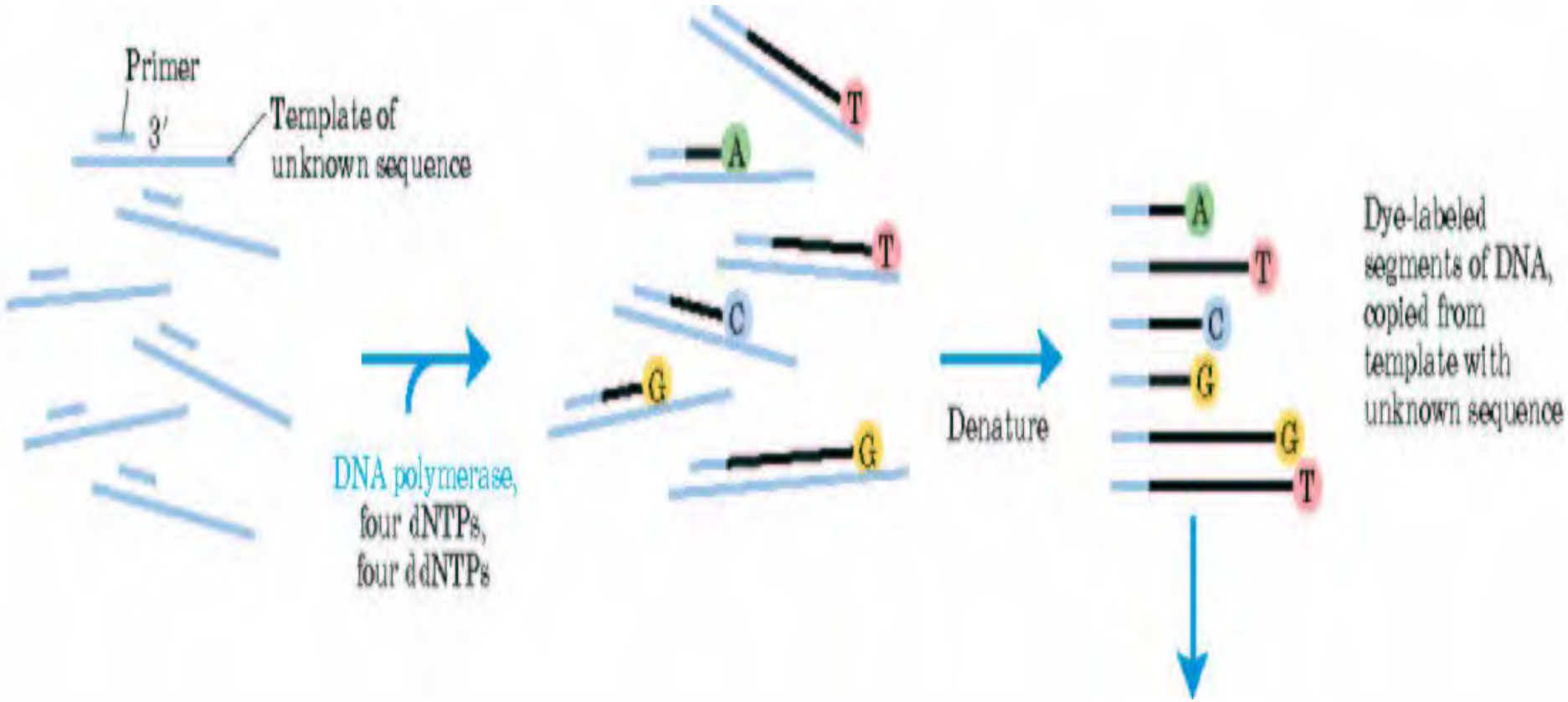


+ 5% ddNTP

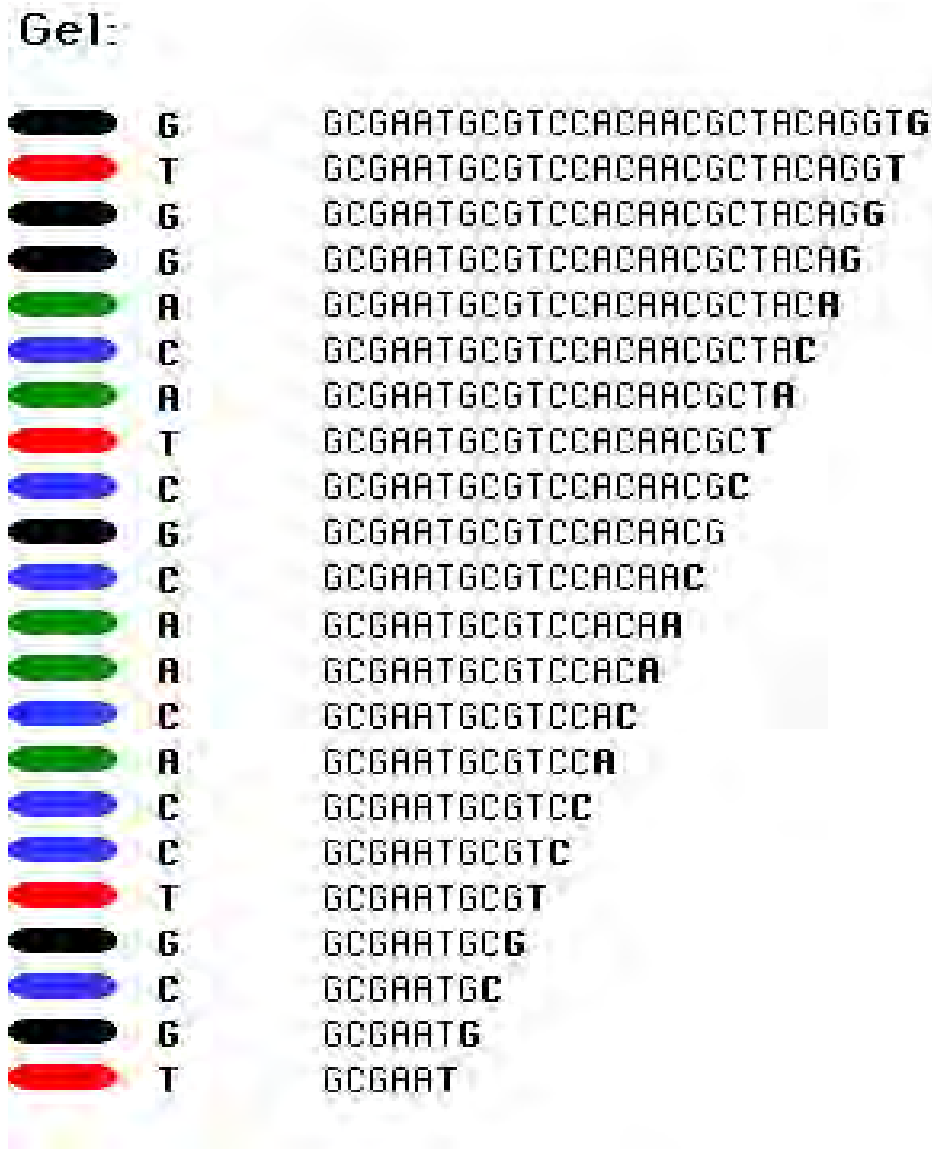
....gel elektroforegram fragmenata DNK razdvojenih po veličini



Automatsko sekvenciranje DNK: svaki ddNTP je obeležen fluorescentnom bojom

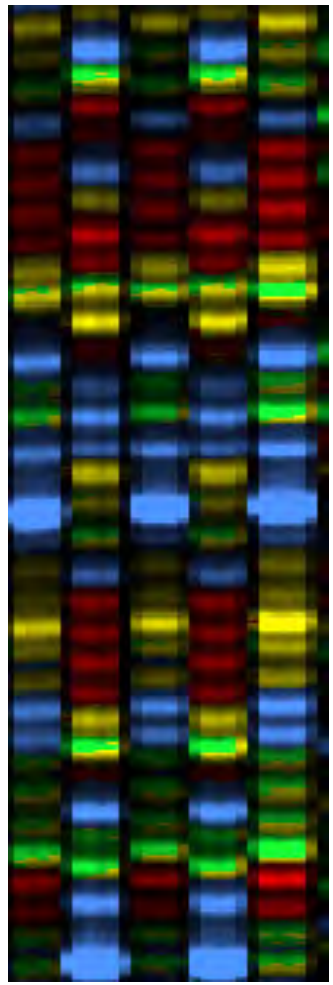


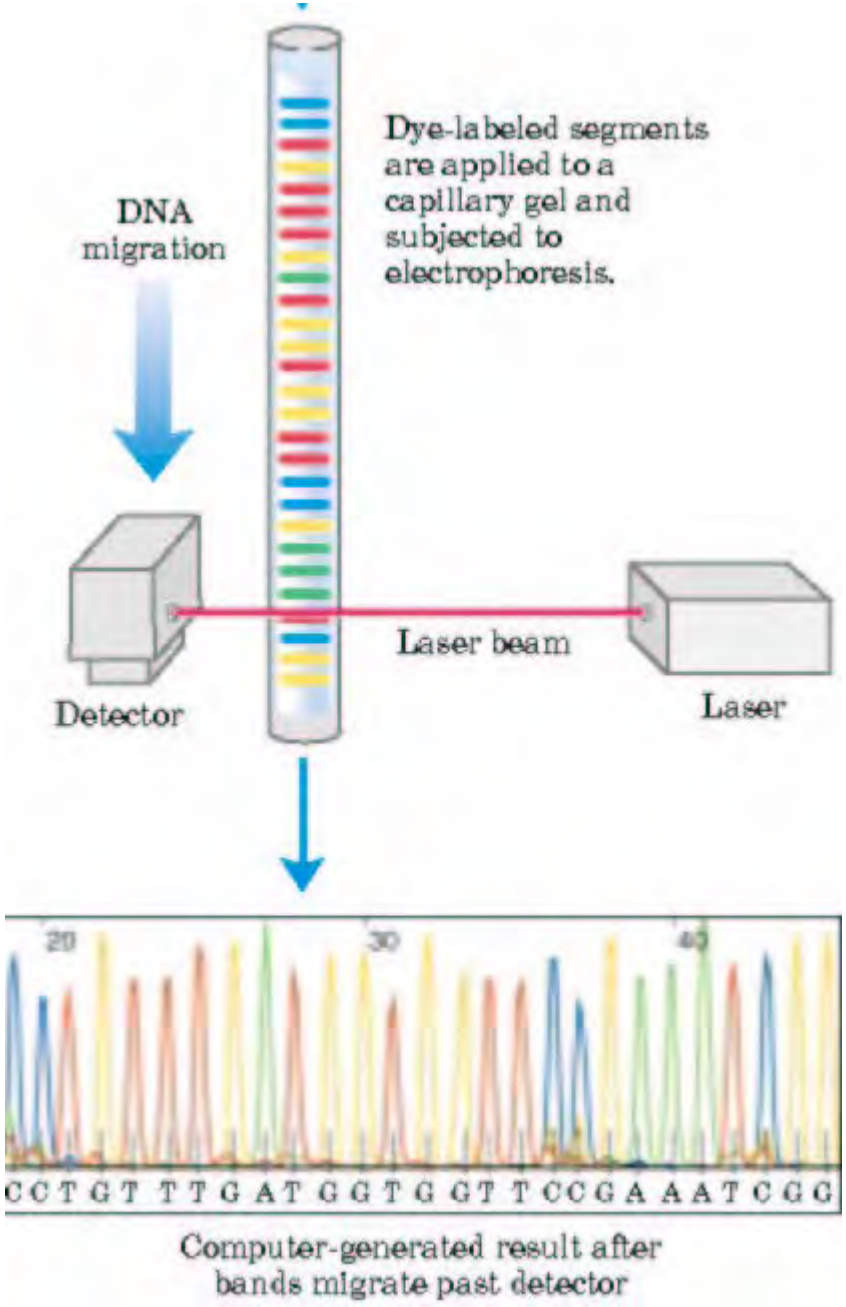
Gel elektroforegram



Sirovi podaci automatskog sekvenciranja DNK

Zeleno: **ddATP**
Crveno: **ddTTP**
Žuto: **ddGTP**
Plavo: **ddCTP**

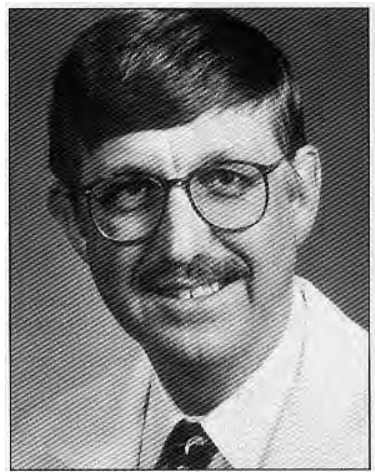




Computer-generated result after bands migrate past detector

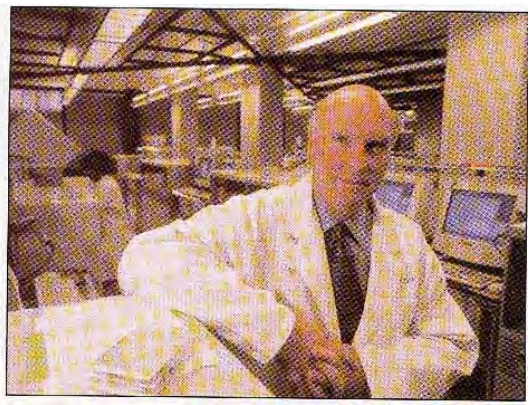
Kratka istorija sekvenciranja humanog genoma...primer kompeticije i kooperacije u nauci!!!!

- **1990 g.** (inicijativa i pripreme započele još 1984.g!) oformljen "Human Genome Project (HGP)": internacionalni projekat/konzorcijum (20 istraživačkih centara za sekvenciranje iz 6 zemalja: SAD, GB, Nemačka, Francuska, Japan, Kina) koji je planiran da traje 15 godina, ali je završen ranije!!!! Koordinatori projekta: J. Watson, potom F.S. Collins (NIH).



Francis S. Collins

- **1997.g.** privatna firma Celera objavila da započinje sa sekvenciranjem humanog genoma pod rukovodstvom J.C. Ventera!!!!



J. Craig Venter

Prvi (grubi) rezultati publikovani su u časopisima Science i Nature februara, 2001.g.!, a definitivni? 2003.g.

Nature, 409, February 15, 860-921 (2001)

nature

nature 15 February 2001

the human genome article

Nature 409, 860 - 921 (2001) © Macmillan Publishers Ltd.

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium

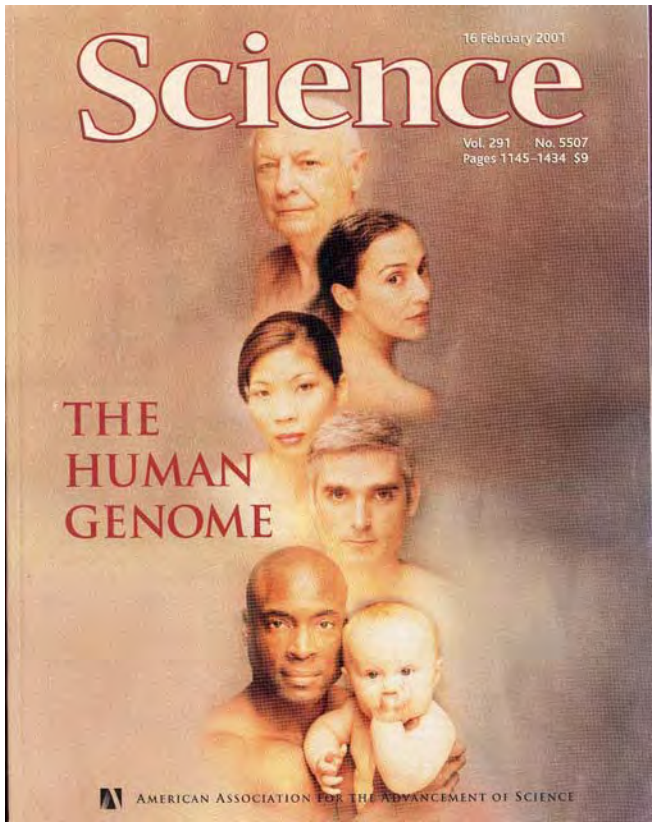
The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

Science, 291, February 16, 1304-1351 (2001)

The Sequence of the Human Genome

J. Craig Venter,[1*] Mark D. Adams,[1] Eugene W. Myers,[1] Peter W. Li,[1] Richard J. Mural,[1] Granger G. Sutton,[1] Hamilton O. Smith,[1] Mark Yandell,[1] Cheryl A. Evans,[1] Robert A. Holt,[1] Jeannine D. Gocayne,[1] Peter Amanatides,[1] Richard M. Ballew,[1] Daniel H. Huson,[1] Jennifer Russo Wortman,[1] Qing Zhang,[1] Chinnappa D. Kodira,[1] Xiangqun H. Zheng,[1] Lin Chen,[1] Marian Skupski,[1] Gangadharan Subramanian,[1] Paul D. Thomas,[1] Jinghui Zhang,[1] George L. Gabor Miklos,[2] Catherine Nelson,[3] Samuel Broder,[1] Andrew G. Clark,[4] Joe Nadeau,[5] Victor A. McKusick,[6] Norton Zinder,[7] Arnold J. Levine,[7] Richard J. Roberts,[8] Mel Simon,[9] Carolyn Slayman,[10] Michael Hunkapiller,[11] Randall Bolanos,[1] Arthur Delcher,[1] Ian Dew,[1] Daniel Fasulo,[1] Michael Flanigan,[1] Liliana Florea,[1] Aaron Halpern,[1] Sridhar Hannenhalli,[1] Saul Kravitz,[1] Samuel Levy,[1] Clark Mobarry,[1] Knut Reinert,[1] Karin Remington,[1] Jane Abu-Threideh,[1] Ellen Beasley,[1] Kendra Biddick,[1] Vivien Bonazzi,[1] Rhonda Brandon,[1] Michele Cargill,[1] Ishwar Chandramouliswaran,[1] Rosane Charlab,[1] Kabir Chaturvedi,[1] Zuoming Deng,[1] Valentina Di Francesco,[1] Patrick Dunn,[1] Karen Eilbeck,[1] Carlos Evangelista,[1] Andrei E. Gabrielian,[1] Weiniu Gan,[1] Wangmao Ge,[1] Fangcheng Gong,[1] Zhiping Gu,[1] Ping Guan,[1] Thomas J. Heiman,[1] Maureen E. Higgins,[1] Rui-Ru Ji,[1] Zhaoxi Ke,[1] Karen A. Ketchum,[1] Zhongwu Lai,[1] Yiding Lei,[1] Zhenya Li,[1] Jiayin Li,[1] Yong Liang,[1] Xiaoying Lin,[1] Fu Lu,[1] Gennady V. Merkulov,[1] Natalia Milshina,[1] Helen M. Moore,[1] Ashwinikumar K Naik,[1] Vaibhav A. Narayan,[1] Beena Neelam,[1] Deborah Nusskern,[1] Douglas B. Rusch,[1] Steven Salzberg,[12] Wei Shao,[1] Bixiong Shue,[1] Jingtao Sun,[1] Zhen Yuan Wang,[1] Aihui Wang,[1] Xin Wang,[1] Jian Wang,[1] Ming-Hui Wei,[1] Ron Wides,[13] Chunlin Xiao,[1] Chunhua Yan,[1] Alison Yao,[1] Jane Ye,[1] Ming Zhan,[1] Weiqing Zhang,[1] Hongyu Zhang,[1] Qi Zhao,[1] Liansheng Zheng,[1] Fei Zhong,[1] Wenyan Zhong,[1] Shiaoping C. Zhu,[1] Shaying Zhao,[12] Dennis Gilbert,[1] Suzanna Baumhueter,[1] Gene Spier,[1] Christine Carter,[1] Anibal Cravchik,[1] Trevor Woodage,[1] Feroze Ali,[1] Huijin An,[1] Aderonke Awe,[1] Danita Baldwin,[1] Holly Baden,[1] Mary Barnstead,[1] Ian Barrow,[1] Karen Beeson,[1] Dana Busam,[1] Amy Carver,[1] Angela Center,[1] Ming Lai Cheng,[1] Liz Curry,[1] Steve Danaher,[1] Lionel Davenport,[1] Raymond Desilets,[1] Susanne Dietz,[1] Kristina Dodson,[1] Lisa Doup,[1] Steven Ferriera,[1] Neha Garg,[1] Andres Gluecksmann,[1] Brit Hart,[1] Jason Haynes,[1] Charles Haynes,[1] Cheryl Heiner,[1] Suzanne Hladun,[1] Damon Hostin,[1] Jarrett Houck,[1] Timothy Howland,[1] Chinyere Ibegwam,[1] Jeffery Johnson,[1] Francis Kalush,[1] Lesley Kline,[1] Shashi Koduru,[1] Amy Love,[1] Felecia Mann,[1] David May,[1] Steven McCawley,[1] Tina McIntosh,[1] Ivy McMullen,[1] Mee Moy,[1] Linda Moy,[1] Brian Murphy,[1] Keith Nelson,[1] Cynthia Pfannkoch,[1] Eric Pratts,[1] Vinita Puri,[1] Hina Qureshi,[1] Matthew Reardon,[1] Robert Rodriguez,[1] Yu-Hui Rogers,[1] Deanna Romblad,[1] Bob Ruhfel,[1] Richard Scott,[1] Cynthia Sitter,[1] Michelle Smallwood,[1] Erin Stewart,[1] Renee Strong,[1] Ellen Suh,[1] Reginald Thomas,[1] Ni Ni Tint,[1] Sukyee Tse,[1] Claire Vech,[1] Gary Wang,[1] Jeremy Wetter,[1] Sherita Williams,[1] Monica Williams,[1] Sandra Windsor,[1] Emily Winn-Deen,[1] Keriellen Wolfe,[1] Jayshree Zaveri,[1] Karena Zaveri,[1] Josep F. Abril,[14] Roderic Guigó,[14] Michael J. Campbell,[1] Kimmen V. Sjolander,[1] Brian Karlak,[1] Anish Kejariwal,[1] Huaiyu Mi,[1] Betty Lazareva,[1] Thomas Hatton,[1] Apurva Narechania,[1] Karen Diemer,[1] Anushya Muruganujan,[1] Nan Guo,[1] Shinji Sato,[1]

Projekat humanog genoma (HGP)



- Šta su uradili?
- Zašto su to uradili?
- Šta to znači za čovečanstvo?

Ciljevi HGP

- Napraviti fizičku mapu za svih 24 humana hromozoma.
- Identifikovati sve gene u humanom genomu & mapirati ih na njihovim hromozomima.
- Odrediti sekvencu procenjenih 3 milijardi parova baza.
- Analizirati genetske varijacije među ljudima.
- Mapirati i sekvencirati genome model-organizama.
- Razvijati laboratorijske i računarske tehnike.
- Širenje (diseminacija) genetske informacije.
- Razmatranje etičkih, pravnih i socijalnih aspekata rezultata projekta...

Model organizmi

- Bakterije (*E. coli*, influenza, i neke druge)
- Kvasci (*Saccharomyces cerevisiae*)
- Biljke (*Arabidopsis thaliana*)
- Crvi (*Caenorhabditis elegans*)
- Vinska mušica (*Drosophila melanogaster*)
- Miš (*Mus musculus*)

Centri za sekvenciranje DNK

Izolovanje DNK

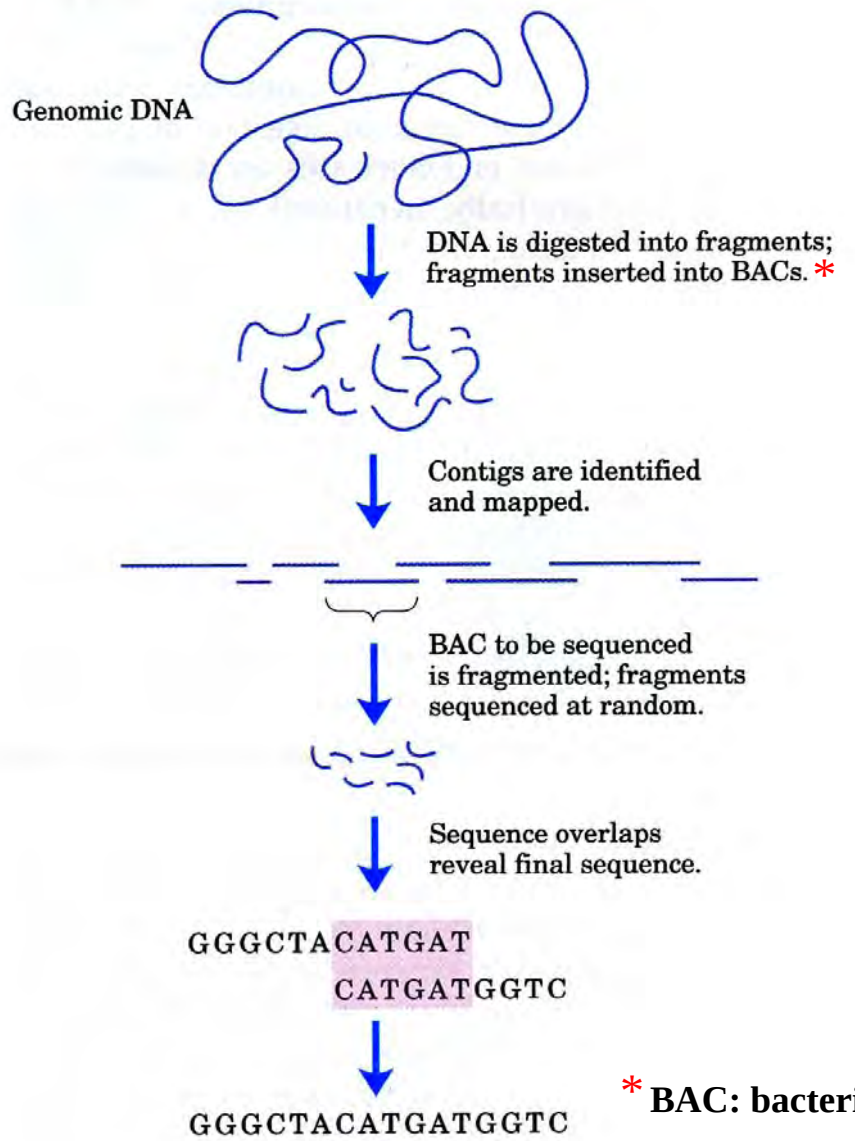
Razdvajanje uzoraka

Laboratorija

Roboti za sekvenciranje

Soba sa mašinama za sekvenciranje Celera Genomics

Strategije za sekvenciranje DNK:



HGP: polazi od fizičke mape hromozoma:

- genom je mapiran i raspodeljen među istraživačkim centrima;
-svaki klon je sekvenciran "shotgun" postupkom (tehnike za sekvenciranje omogućuju sekvenciranje 600-750 bp, a mnogi klonovi su imali više od 100 000 bp!)

Celera: "shotgun" celokupne DNK: genom se nasumično ("random") fragmentiše, fragmenti se kloniraju (umnože) i sekvenciraju. Postupak zahteva sekvenciranje 7-9x puta većeg broja baza od postojećeg: za humani genom of 3 milijarde bp, treba sekvencirati 21-27 milijardi baza da bi se dobilo adekvatno preklapanje fragmenata.

* **BAC: bacterial artificial chromosome**

Šta smo naučili iz sekvence humanog genoma?

- **U brojevima:**
  - Humani genom sadrži približno 3 x 10^9 bp.
  - Prosečni gen sadrži oko 3000 baza.
  - Ukupan broj gena je procenjen na manje od 30 000 (ranije procenjivano: 80 000 – 140 000 gena!)
  - 99.9% sekvence je isto kod svih ljudi.
  - Ne poznajemo funkcije 50% otkrivenih gena**.**

- **“Odvajanje žita od kukolja”:**
  - Manje od 2% genoma kodira proteine.
  - Ponavljajuće sekvence ("junk DNK") čine najmanje 50% humanog genoma.
  - Smatra se da ponavljajuće sekvence nemaju direktnu funkciju, ali su povezane sa strukturom i dinamikom hromozoma. Tokom vremena ove sekvence menjaju genom, kreirajući nove gene i modifikujući postojeće!

- **Kako su sekvence raspoređene?**
  - U genima bogatim delovima (“gene-dense”) "urbanim centrima" dominiraju blokovi G i C.
  - U genima siromašnim delovima, “pustinjama” dominiraju blokovi A i T.
  - Geni su raspoređeni nasumično duže genoma, sa velikim delovima ne-kodirajuće DNK između njih.
  - Delovi od oko 30 000 ponavljajućih C i G baza često se nalaze do oblasti bogate genima, stvarajući barijeru izmedju gena i "junk DNK." Ova CpG “ostrva” imaju funkciju u regulaciji aktivnosti gena.
  - Hromozom 1 ima najviše (2968), a hromozom Y najmanje (231) gena.

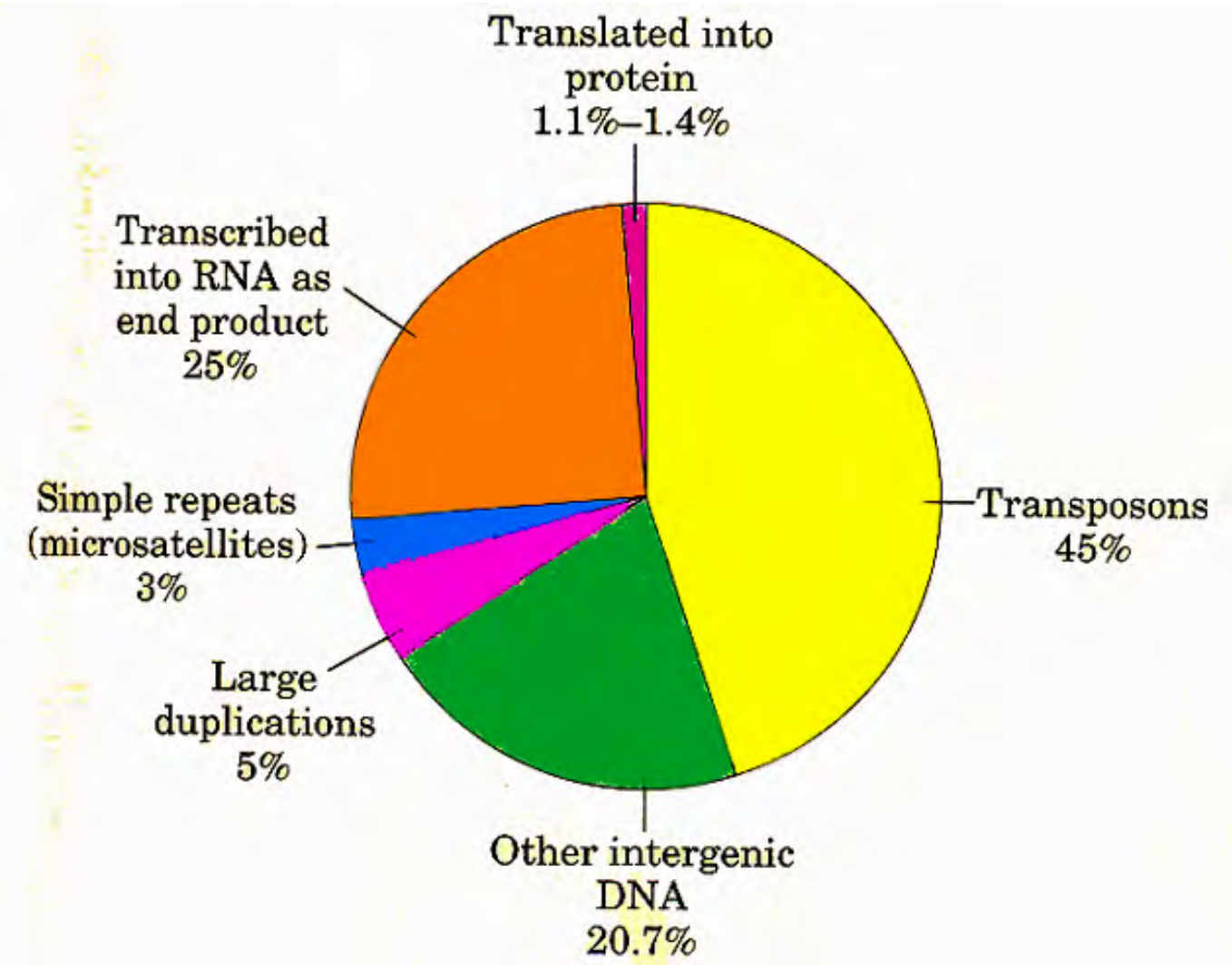
- **Humani genom u poređenju sa genomima drugih organizama:**
  - Čovek ima tri puta više vrsta proteina od npr muve ili crva zbog “alternativnog splicinga” iRnk i post-translacionih modifikacija proteina. Na ovaj način mogu da nastanu različiti proteini iz istog gena.
  - Većina proteinskih familija je ista kod čoveka i crva, muva, biljaka, ali je broj članova familija veći kod čoveka i to posebno proteina koji su uključeni u razvoj i imunitet.

- **Individualne variajacije i mutacije:**
  - Na oko 1.4 miliona lokacija razlika je u po jednoj bazi: 1 na 1000 bp (**single nucleotide polimorphism - SNPs**)

Variation between individuals

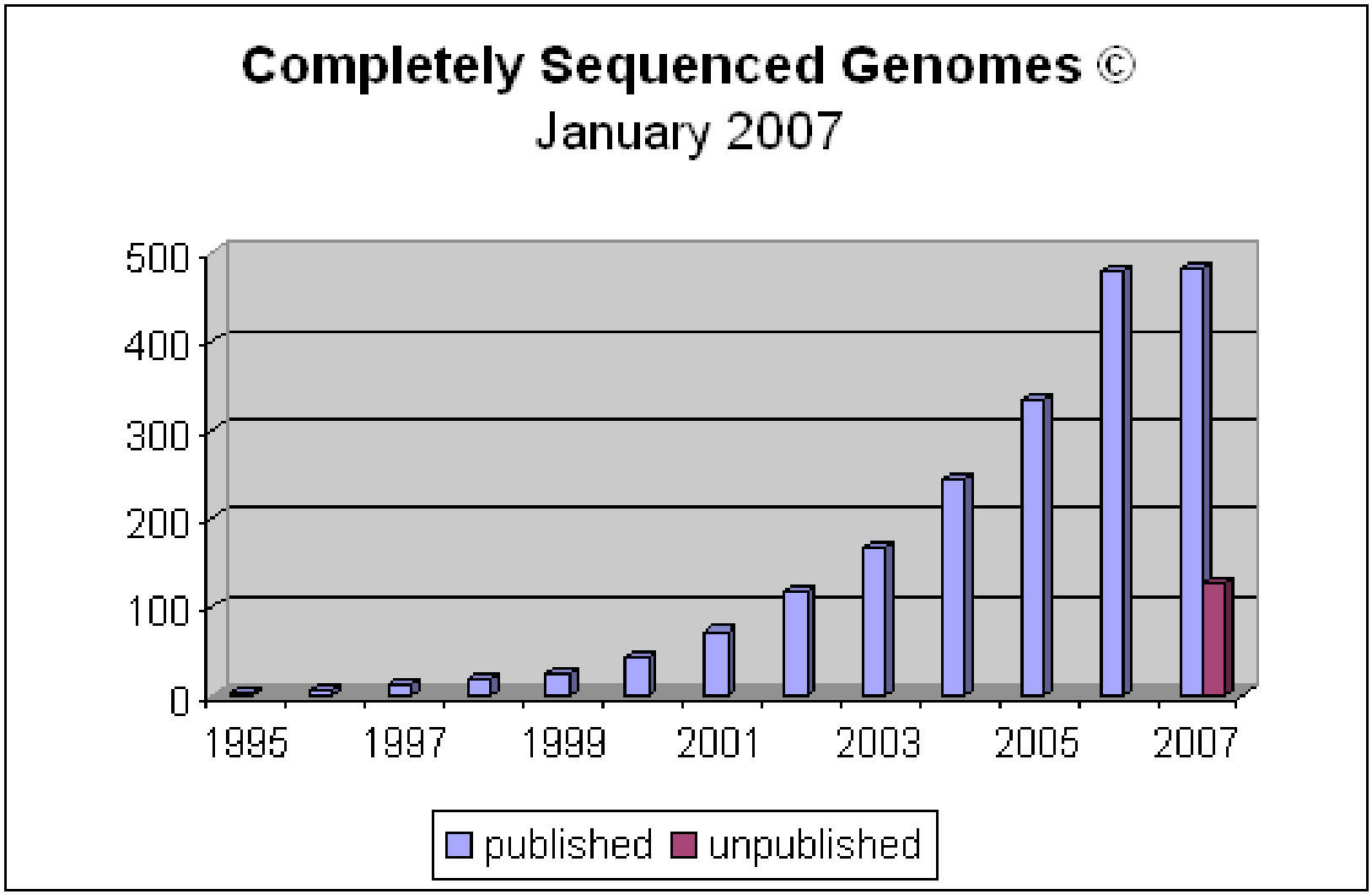
- your DNA is 99.9% the same as any other human on Earth
- humans & chimpanzees share about 98.5% of their genes
- banana DNA & human DNA are about 50% the same

Humani genom: rezime

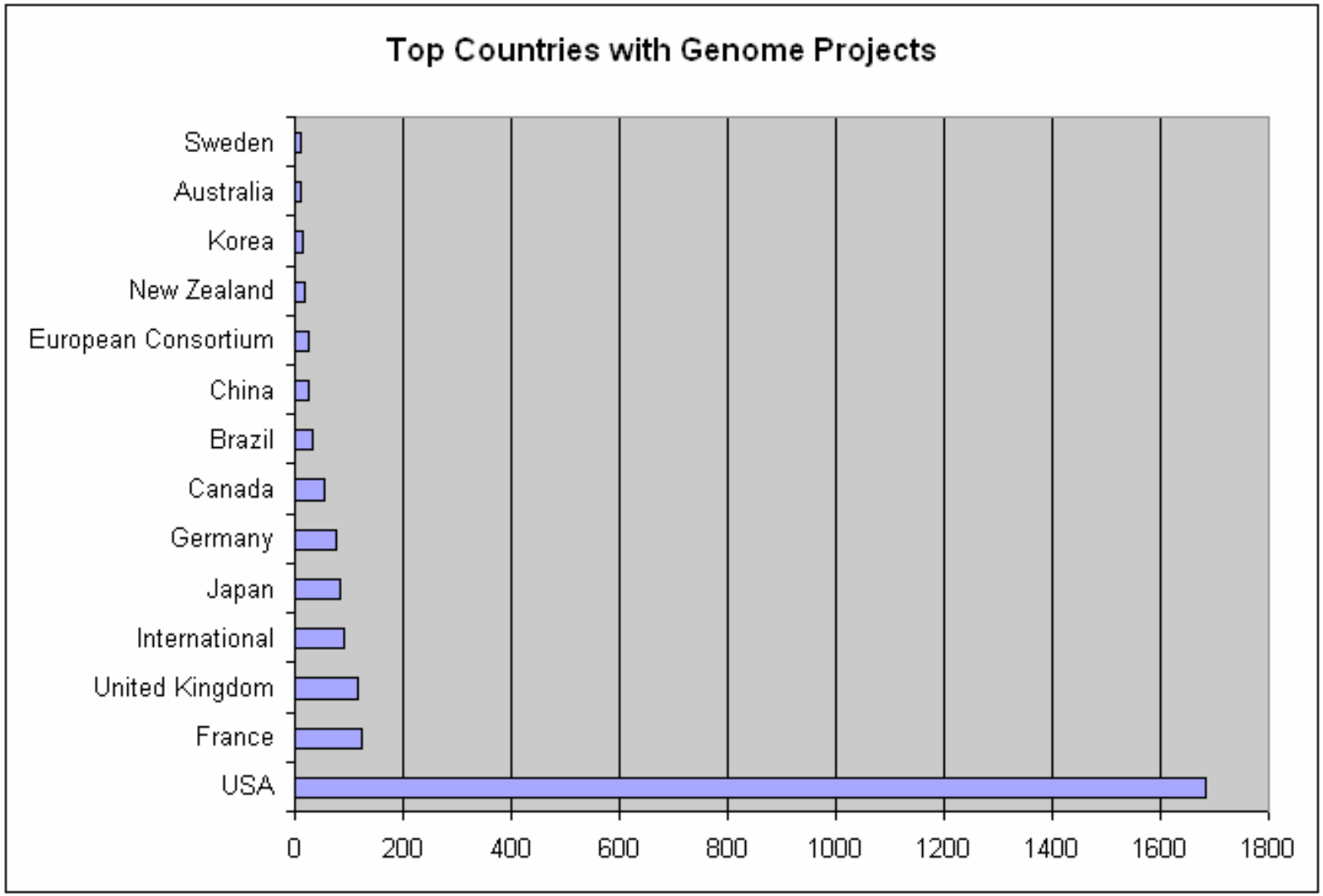


Genomes

organism	size in bp	# genes
E.coli	$4x10^6$	4,300
Anthrax bacillus	$5x10^6$	4,700
Yeast	$2x10^7$	6,000
Nematode	$8x10^7$	19,000
Fruit Fly	$2x10^8$	14,000
Mouse	$3x10^9$	30,000
Human	$3x10^9$	30,000
Rice	$4x10^{11}$	46,000



http://www.genomesonline.org

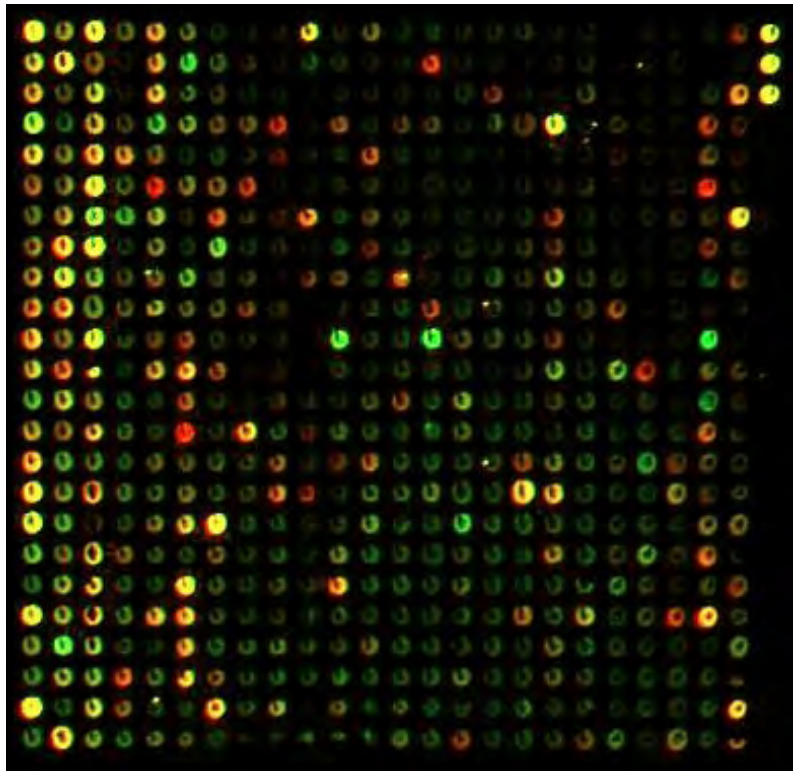


http://www.genomesonline.org

Pitanja koja se postavljaju nakon određivanja sekvence humanog genoma:

- Koje su funkcije ~30 000 gena?
- U kojim ćelijama pod kojim uslovima i u kom stepenu se ovi geni eksprimiraju?
- Kako proizvodi gena interaguju da bi nastao funkcionalni organizam?
- Koje su medicinske posledice varijanti u genima?
- Tradicionalni pristup (jedan protein jedan gen) ne daje odgovore na ova pitanja!!!

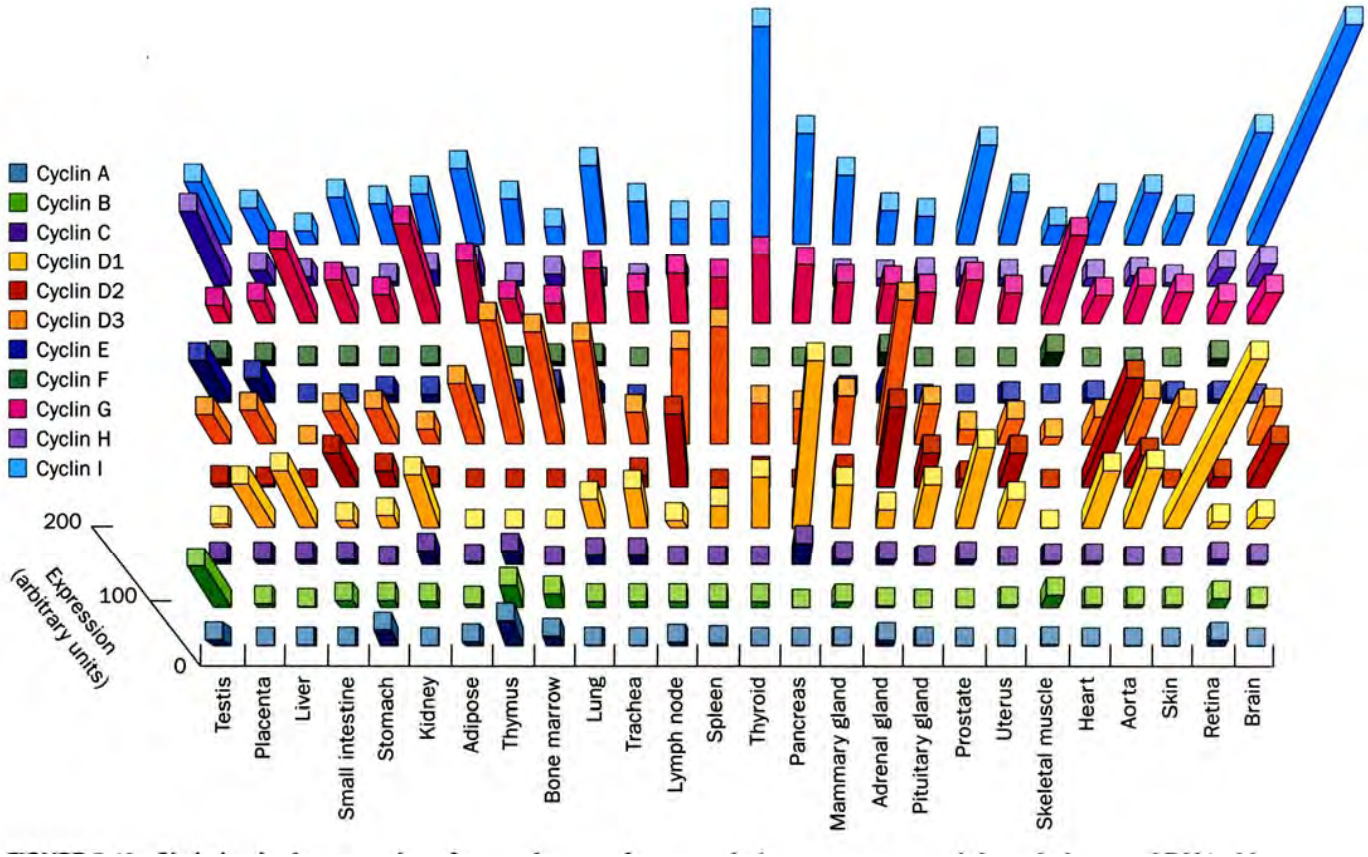
“DNA microarray” ili DNK “chip”



Geni iz kvasca eksprimirani kada je kvasac gajen sa (crveno) i bez (zeleno) glukoze.

Svaka tačka sadrži pikomole specifične sekvence DNK ili sintetičke oligonukleotide (“proba”), npr deo gena ili neki drugi deo DNK koji se koristi za hibridizaciju sa cDNK ili cRNK iz uzorka (“target – meta”).

“DNA microarray” ili DNK “chip”



Varijacije u ekspresiji gena za protein ciklin u raznim tkivima

Proteom: termin prvi put upotrebljen 1995.g.!

- Analogija sa genomom:

- Svi eksprimirani proteini u ćeliji:
  - identifikacija
  - post-translacione modifikacije
  - količina
  - lokalizacija
  - međusobne interakcije

Ispitivanje proteoma – “proteomics”: 2D elektroforeza

Combining isoelectric focusing and SDS electrophoresis sequentially in a process called **two-dimensional electrophoresis** permits the resolution of complex mixtures of proteins (Fig. 5–22). This is a more sensitive analytical method than either electrophoretic method alone. Two-dimensional electrophoresis separates proteins of identical molecular weight that differ in pI, or proteins with similar pI values but different molecular weights.

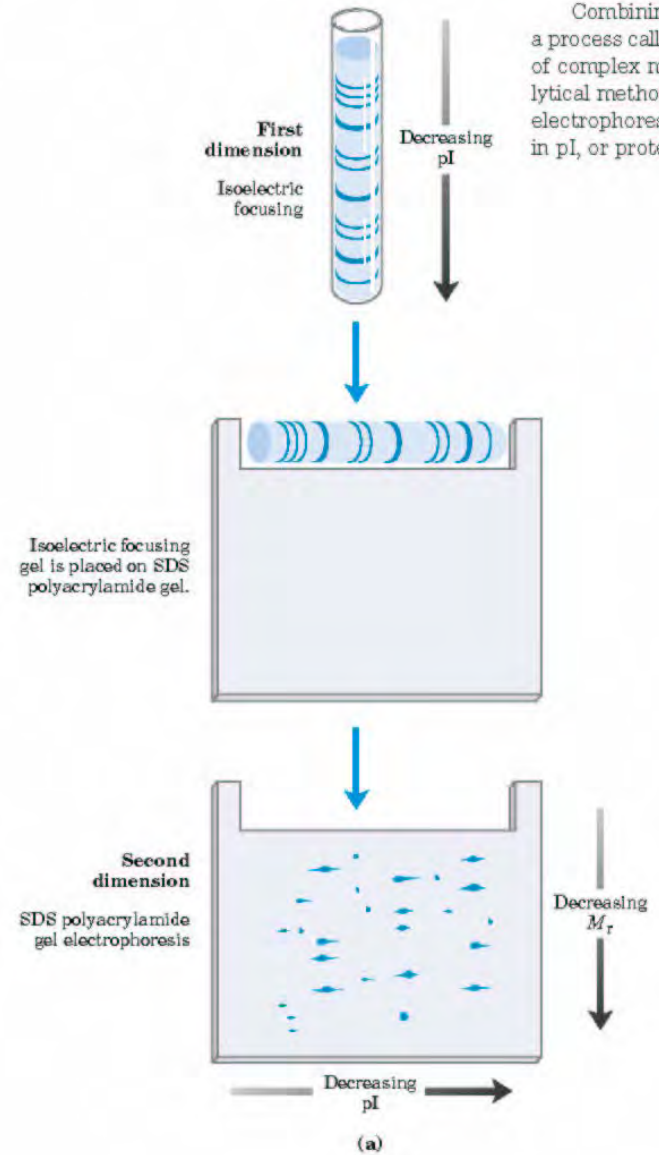


figure 5–22
Two-dimensional electrophoresis. (a) Proteins are first separated by isoelectric focusing in a cylindrical gel. The gel is then laid horizontally on a second, slab-shaped gel, and the proteins are separated by SDS polyacrylamide gel electrophoresis. Horizontal separation reflects differences in pI; vertical separation reflects differences in molecular weight. **(b)** More than 1,000 different proteins from *E. coli* can be resolved using this technique.

- 2D elektroforezom se može videti do 5000 proteina prisutnih u uzorku!!!

- Brojni, referentni 2D elektroforegrami su javno dostupni: http://www.expasy.ch/ch2d/2d-index.html.

Primena u medicini i bilogiji

- Genetske bolesti:
  - javna dostupnost sekvence humanog genoma omogućuje brzu identifikaciju gena u raznim bolestima *in silico;*
- Mete (“targets”) za lekove:
  - farmaceutska industrija je imala ograničeni broj meta za razvoj novih terapija....
  - sada se mogu naći nove mete *in silico*
- Fundamentalna biologija:
  - fundammentalna fiziologija, ćelijska biologija...


POSSIBLE OUTCOME OF THE HUMAN GENOME PROJECT? NO MORE GENETICALLY IMPERFECT PEOPLE!
LIKE STEPHEN HAWKING?
WELL... NO MORE DEPRESSED PEOPLE!
LIKE VINCENT VAN GOGH?
DNA MAP
WELL... NO MORE PHYSICALLY CHALLENGED PEOPLE!
LIKE FDR?
WELL... NO MORE BLIND PEOPLE!
LIKE STEVIE WONDER?
WELL... NO MORE PEOPLE WITH MULTIPLE HANDICAPS!
LIKE HELEN KELLER?
WELL... MAYBE THERE'S A LESSON HERE!
LIKE "THERE'S TWO SIDES TO EVERY COIN"?


"That's all Folks!"