

# Uvod u molekularnu biologiju i genetiku

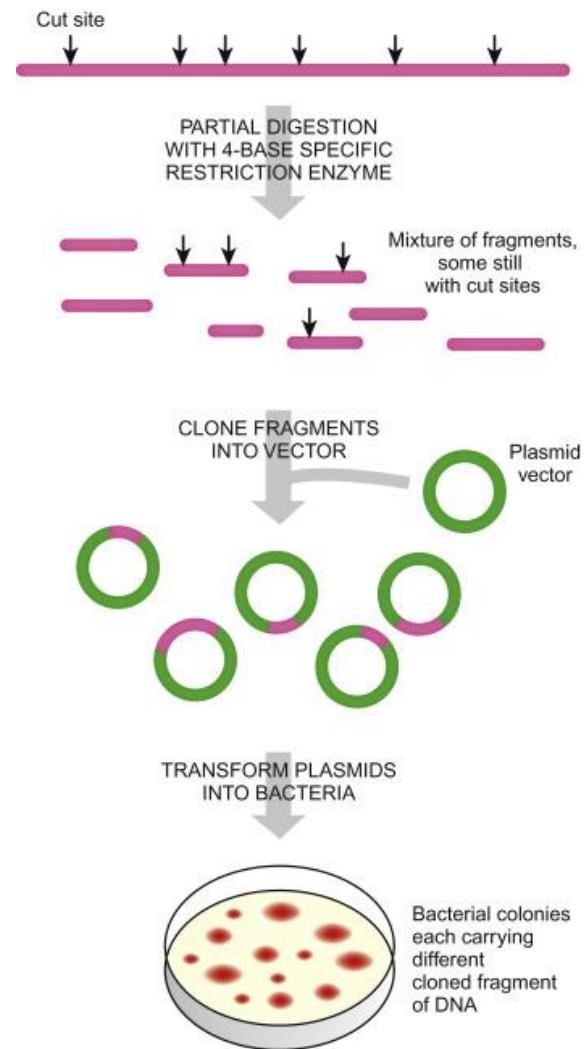


## 8. DNA biblioteke i sekvenciranje genoma

# Genomske DNA biblioteke: *in vivo* sustav umnažanja genoma

- za istraživanje genoma nesekvenciranih vrsta
- prvi korak u *de novo* sekvenciranju (*NGS-next generation sequencing*) nepoznatih genoma

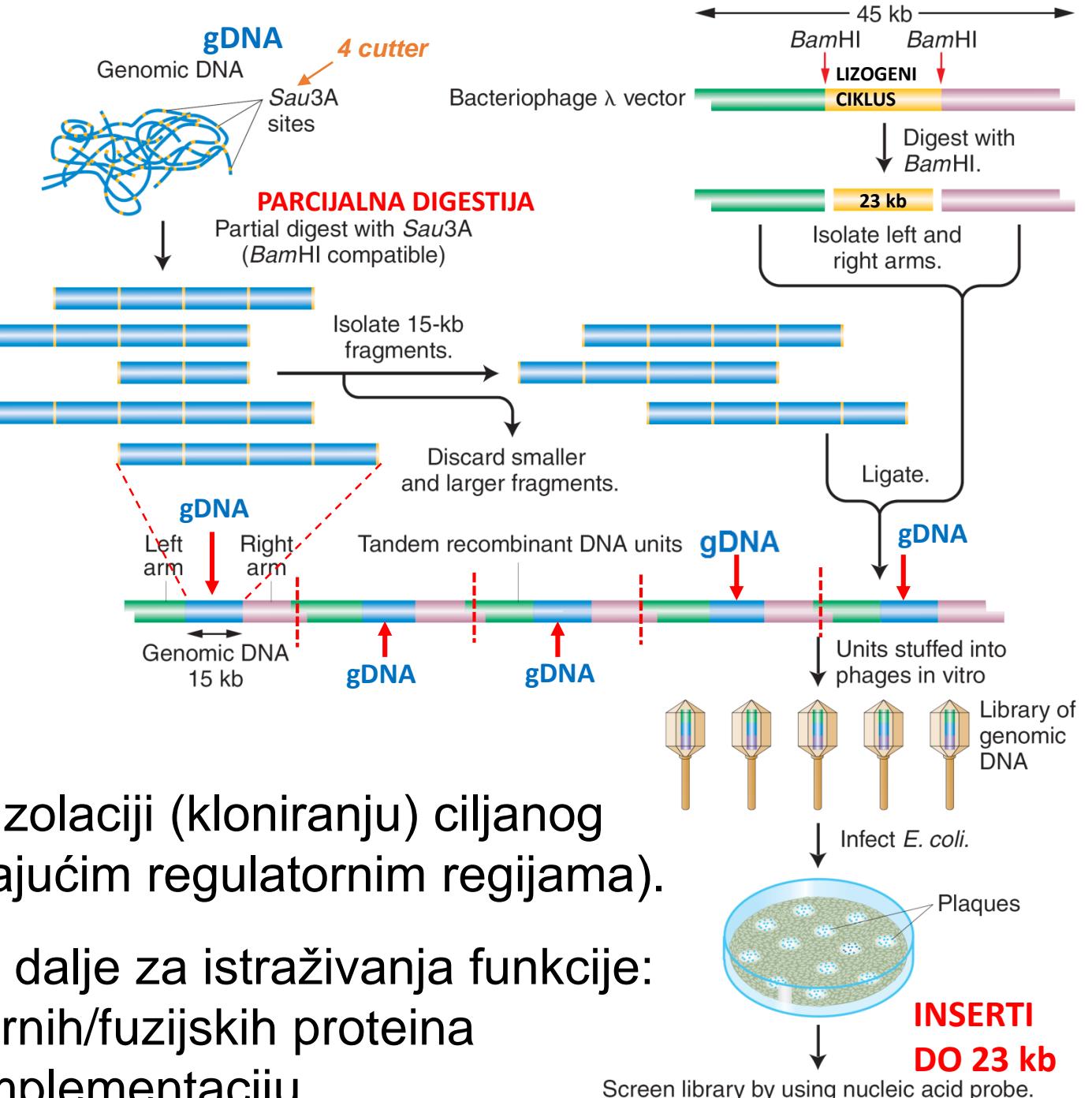
Table 17.1 Cloning Vectors				
Vector	Form of DNA	Host	Capacity	Uses
Plasmid	Circular	<i>E. coli</i>	<15 kb	Subcloning and cDNA libraries
Lambda	Linear phage chromosome	<i>E. coli</i>	<23 kb	cDNA and genomic libraries
Cosmid	Circular	<i>E. coli</i>	30–45 kb	Genomic libraries
BAC	Bacterial chromosome	<i>E. coli</i>	100–200 kb	Genomic libraries
YAC	Yeast chromosome	<i>S. cerevisiae</i>	200–2000 kb	Genomic libraries



Genom se fragmentira enzimatski (RE) ili npr. ultrazvukom (eng. *sonication*), a fragmeni se ligiraju u neki od vektora.

Vektori se razlikuju po kapacitetu tj. veličini DNA fragmenata koje mogu primiti.

# Genomske DNA biblioteke: *in vivo*



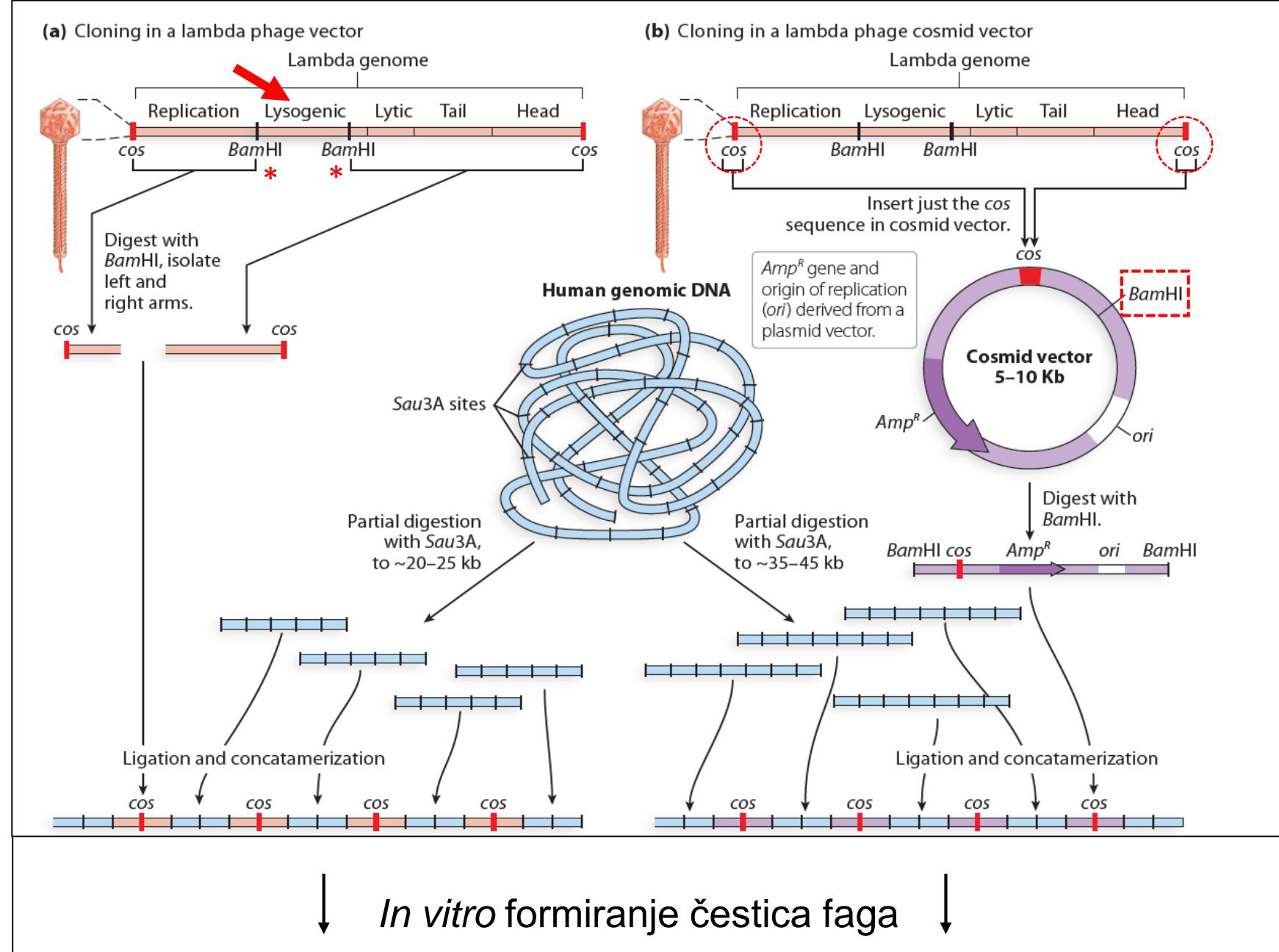
**BIBLIOTEKA  
U  $\lambda$ -FAGU  
(BEZ GENA  
ZA LIZOGENI  
CIKLUS)**

DNA biblioteke služe izolaciji (kloniranju) ciljanog gena (CDS sa pripadajućim regulatornim regijama).

Izolirani gen koristi se dalje za istraživanja funkcije: npr. konstrukcija kimernih/fuzijskih proteina (lokalizacija) ili za komplementaciju.

# BIBLIOTEKA λ FAGA: uklone se geni za lizogeni ciklus

# KOZMIDNA BIBLIOTEKA: ukloni se cijeli genom faga osim cos mesta



BIBLIOTEKA λ FAGA:

KOZMIDNA BIBLIOTEKA:

*In vitro* pakiranje u čestice faga:

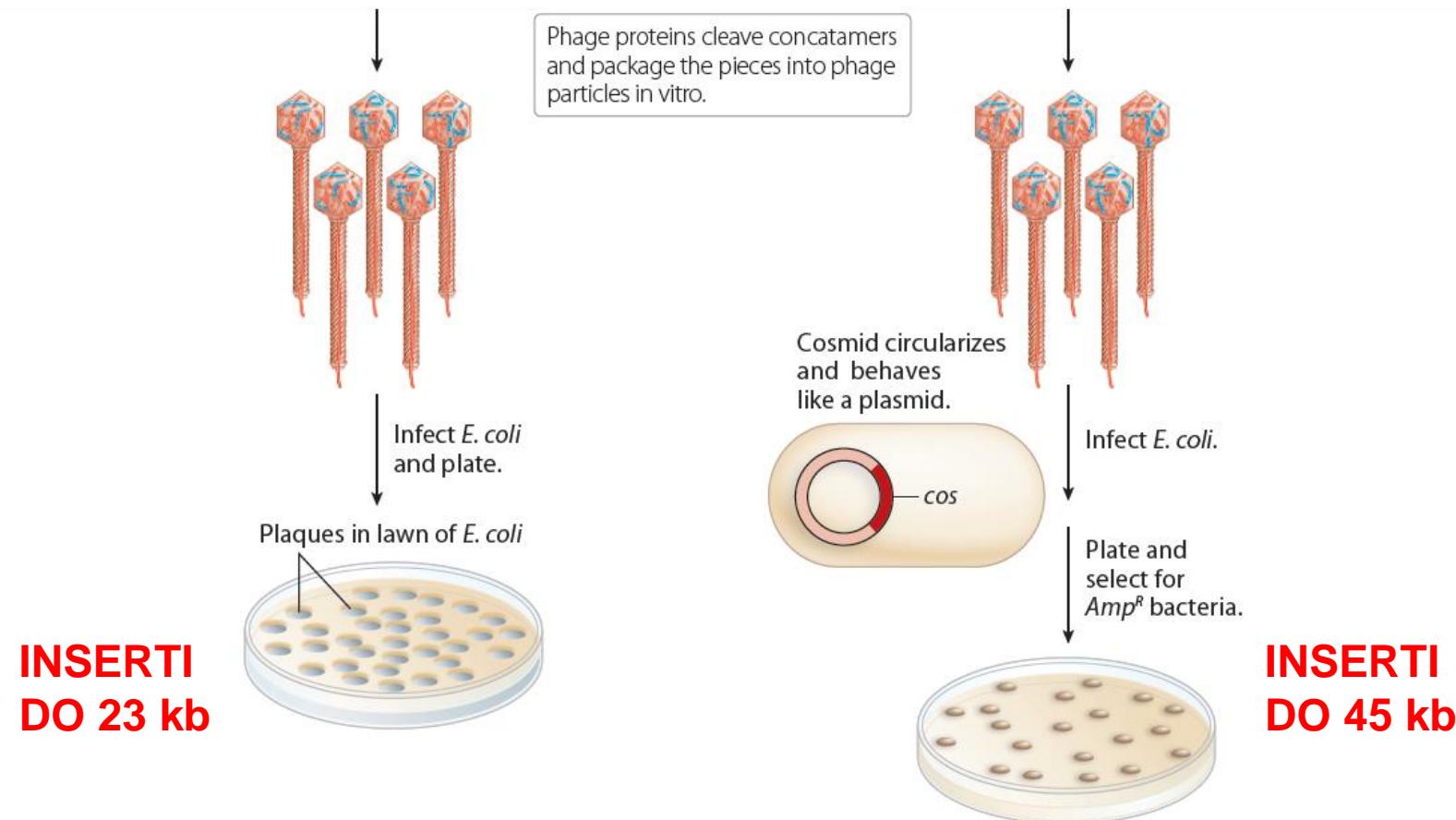
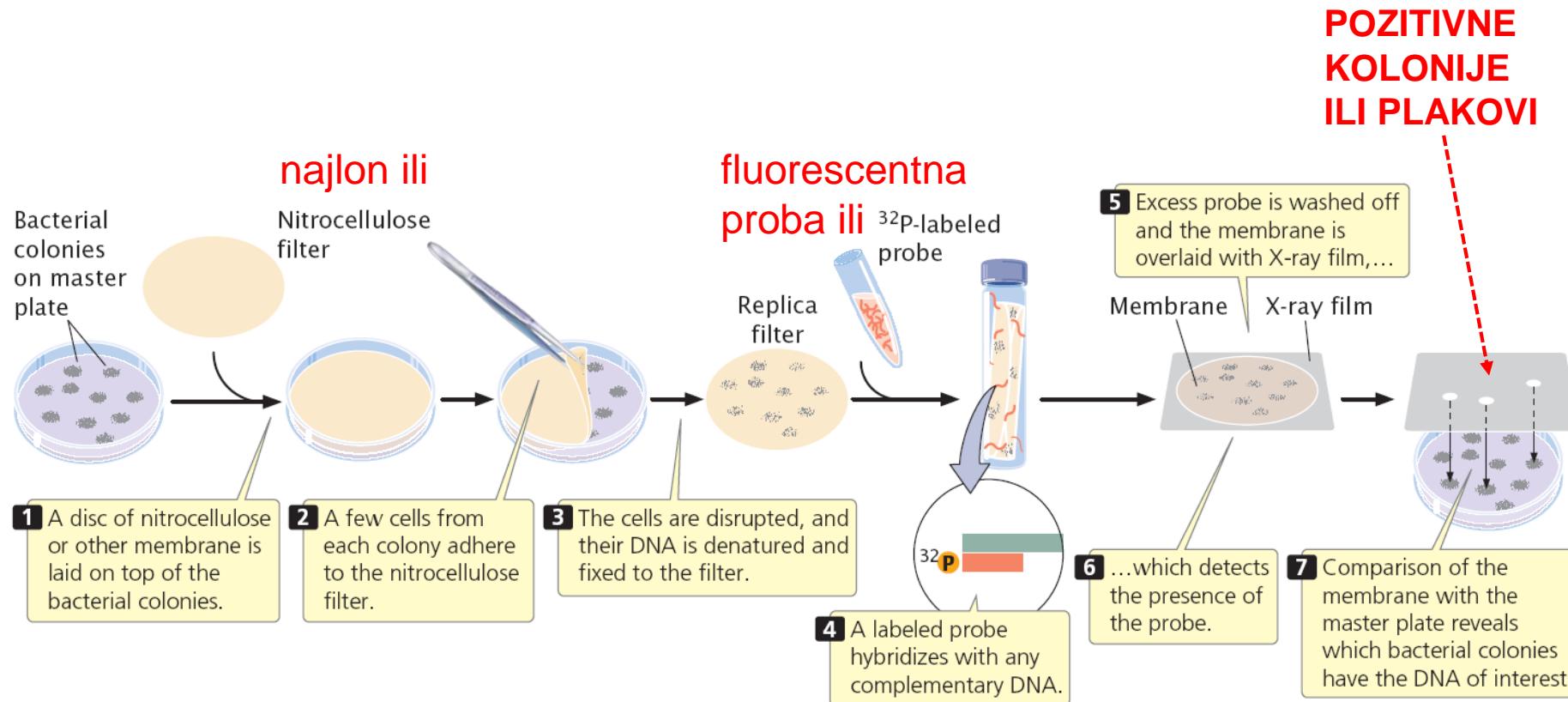
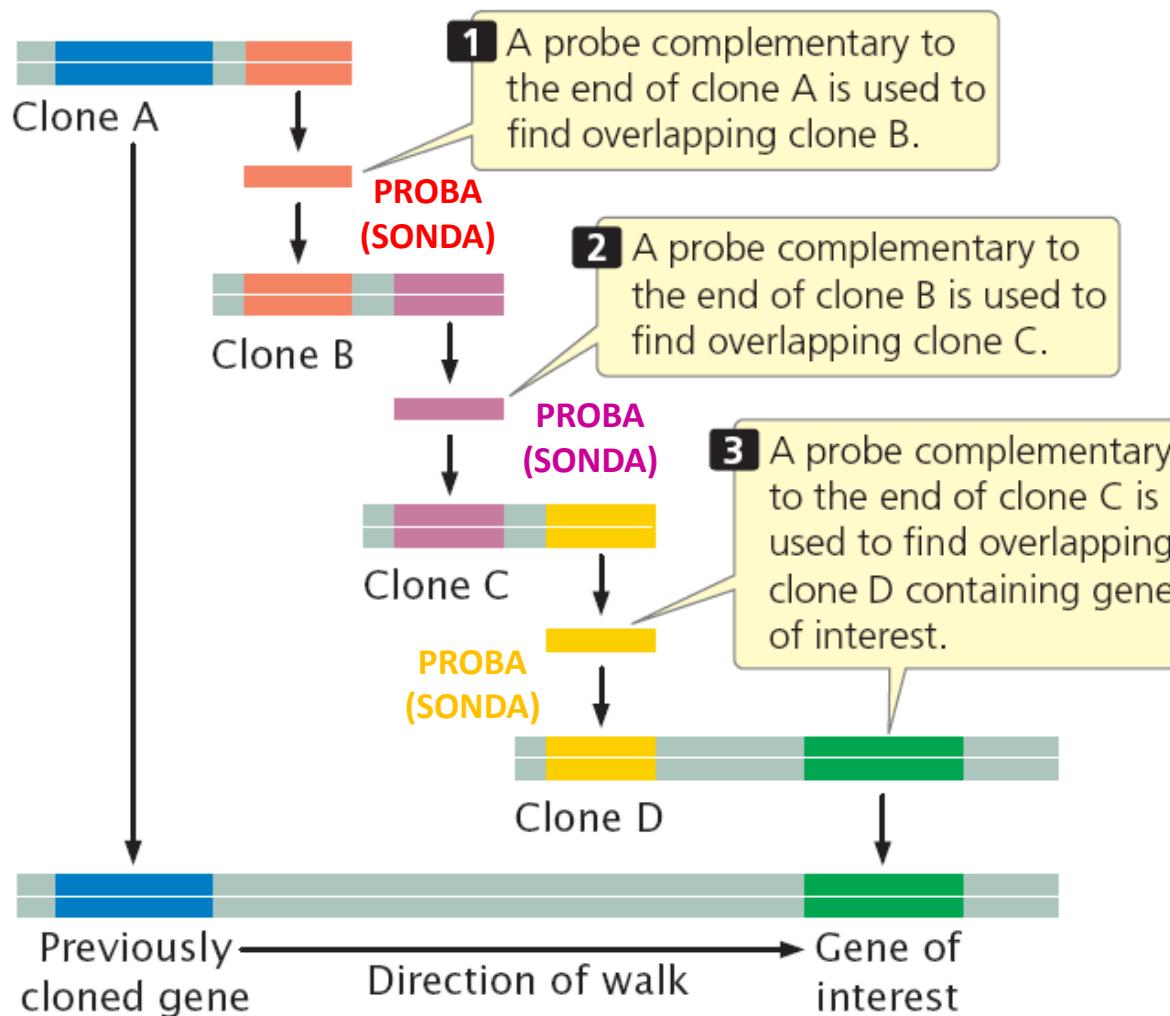


Figure 17.8 Cloning in bacteriophage vectors.

# Pretraživanje genomskih ili cDNA biblioteka za gen od interesa: tzv. “colony lift”



# Chromosome walking: traženje susjednih klonova u DNA biblioteci



**Conclusion:** By making probes complementary to areas of overlap between cloned fragments in a genomic library, we can connect a gene of interest to a previously mapped, linked gene.

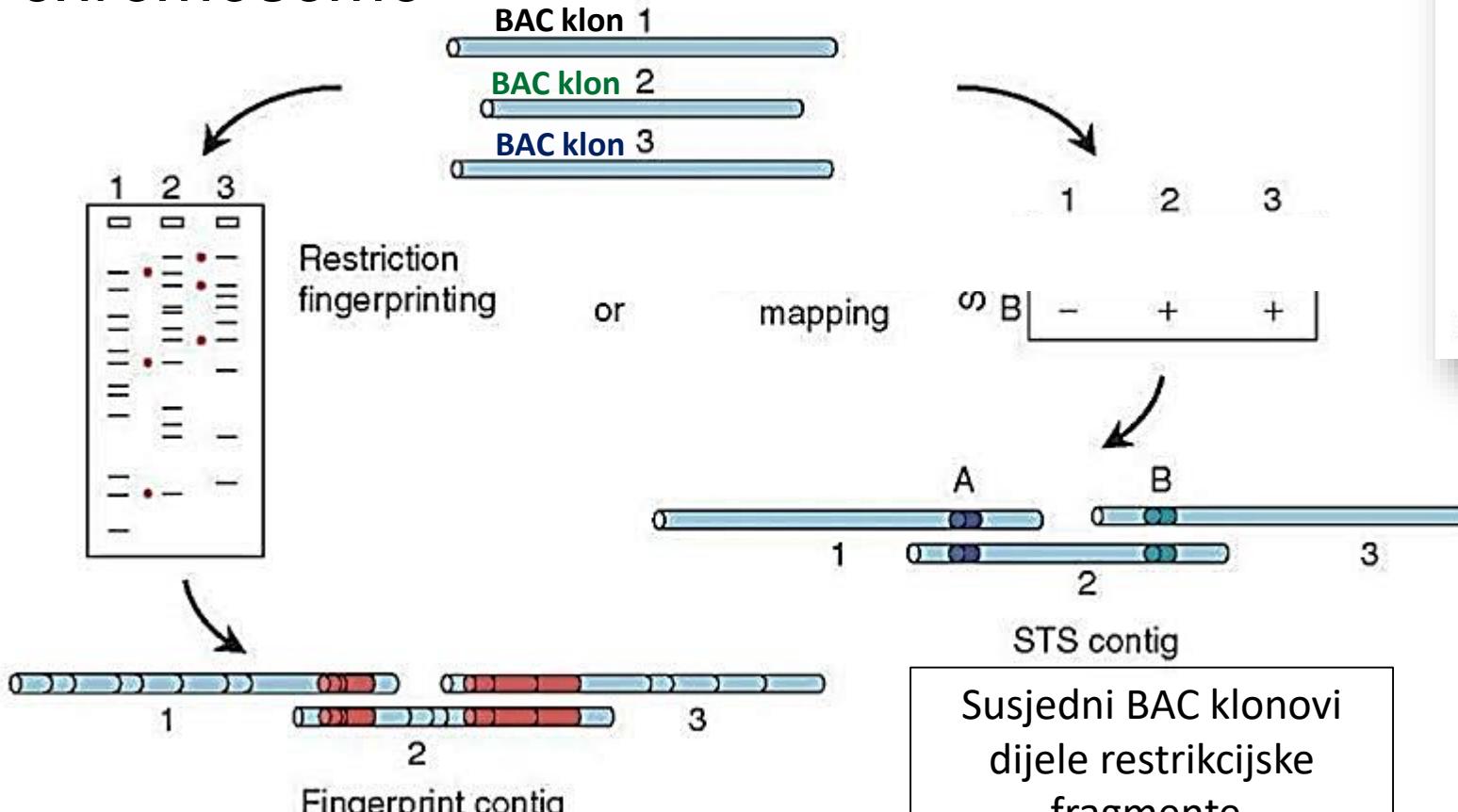
# Različite varijante DNA biblioteka: razlikuju se po kapacitetu (koliko klon biblioteke može ugraditi DNA)

**Table 19.2** Comparison of plasmids, phage lambda vectors, cosmids, and bacterial artificial chromosomes

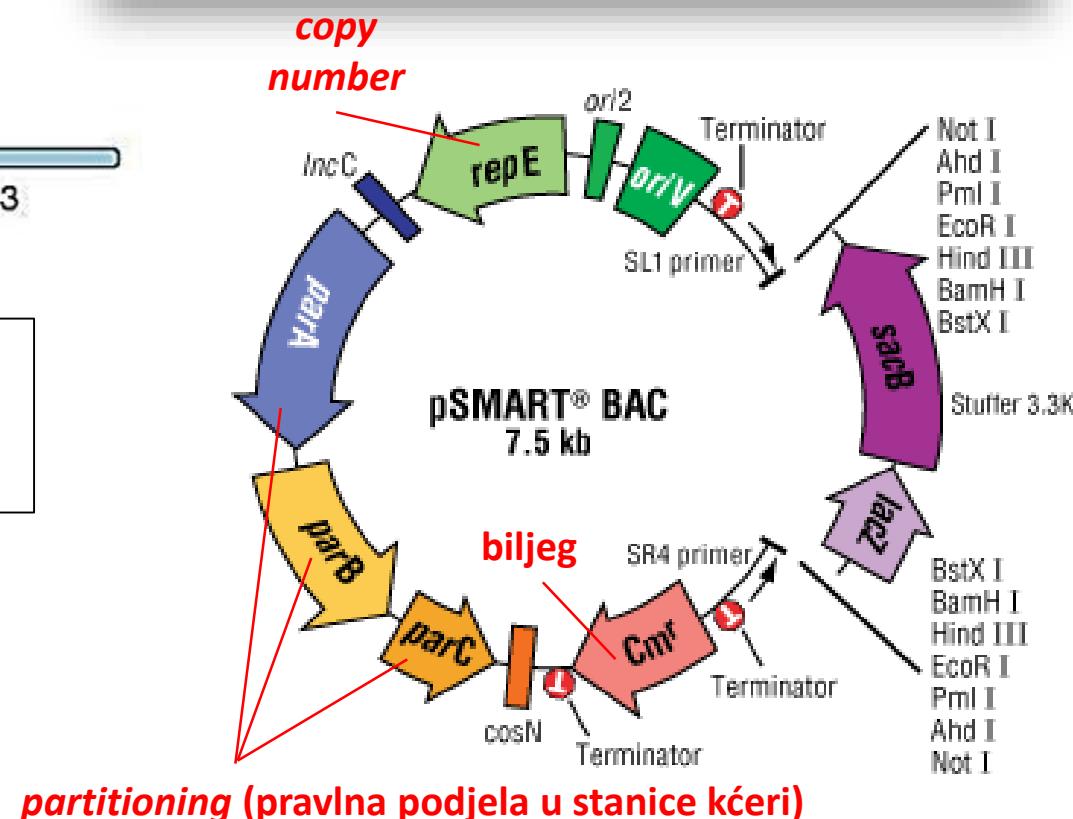
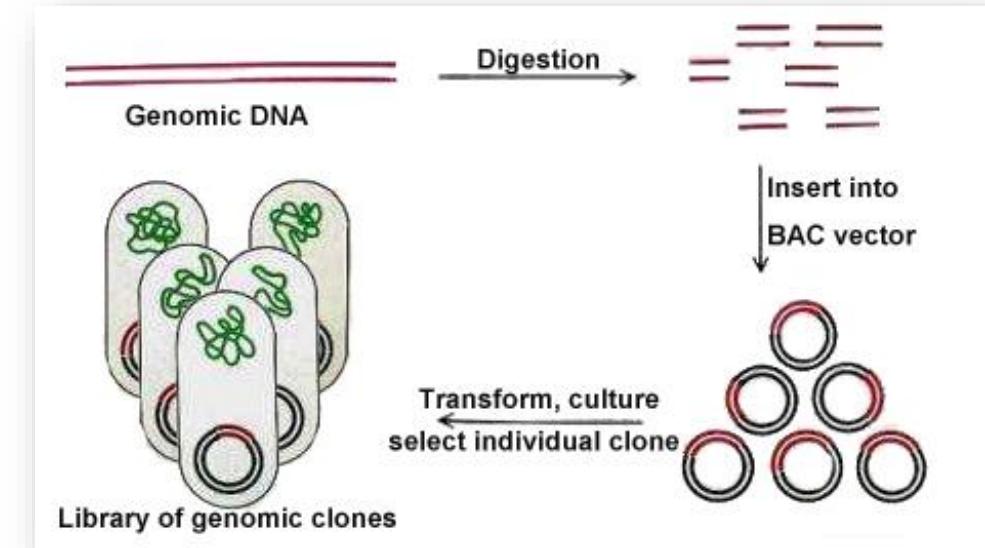
Cloning Vector	Size of DNA That Can Be Cloned	Method of Propagation	Introduction to Bacteria
1. Plasmid ( <b>PLAZMID</b> )	As large as 15 kb	Plasmid replication	Transformation
2. Phage lambda ( <b>λ-FAG</b> )	As large as 23 kb	Phage reproduction	Phage infection
3. Cosmid ( <b>KOZMID*</b> )	As large as 44 kb	Plasmid reproduction	Phage infection
4. Bacterial artificial chromosome ( <b>BAC</b> )	As large as 300 kb	Plasmid reproduction	Electroporation

\* Plazmid sa cos slijedom (*cohesive ends*) iz lambda faga.

# BAC DNA biblioteka: *bacterial artificial chromosome*



**CONTIG – kontinuirani slijed DNA dobiven preklapanjem pojedinih BAC klonova**



# Različite varijante DNA biblioteka: razlikuju se po kapacitetu (koliko klon biblioteke može ugraditi DNA)

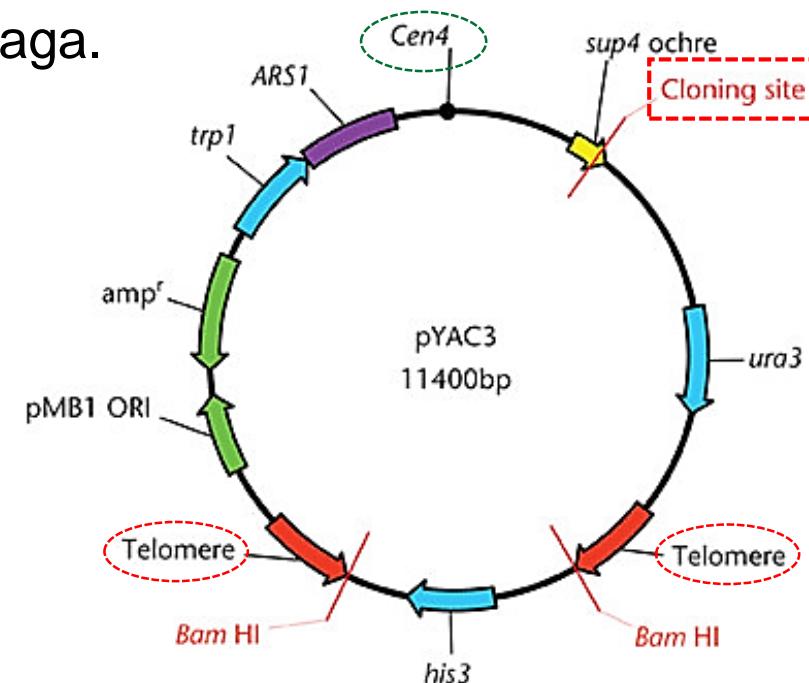
**Table 19.2**

Comparison of plasmids, phage lambda vectors, cosmids, and bacterial artificial chromosomes

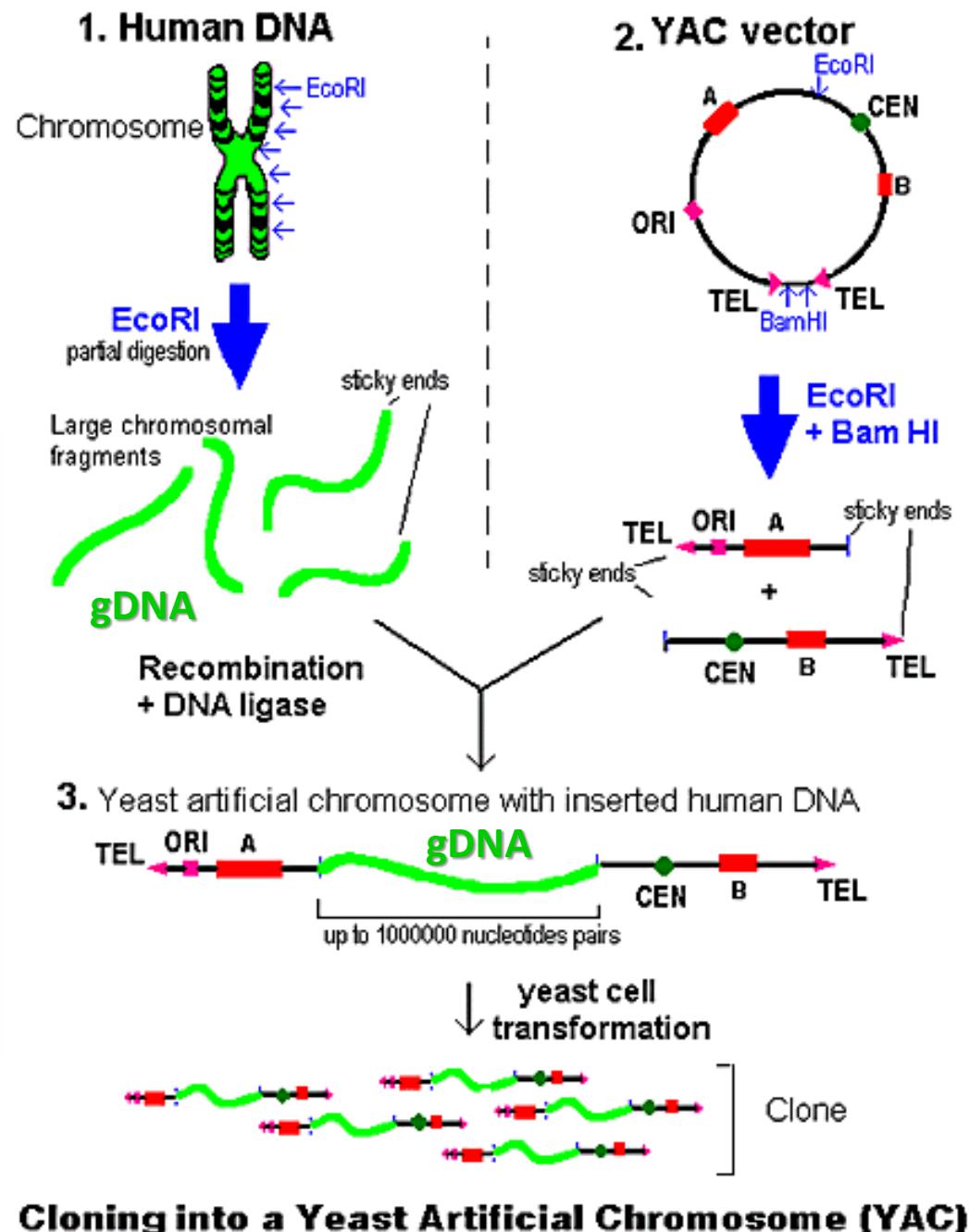
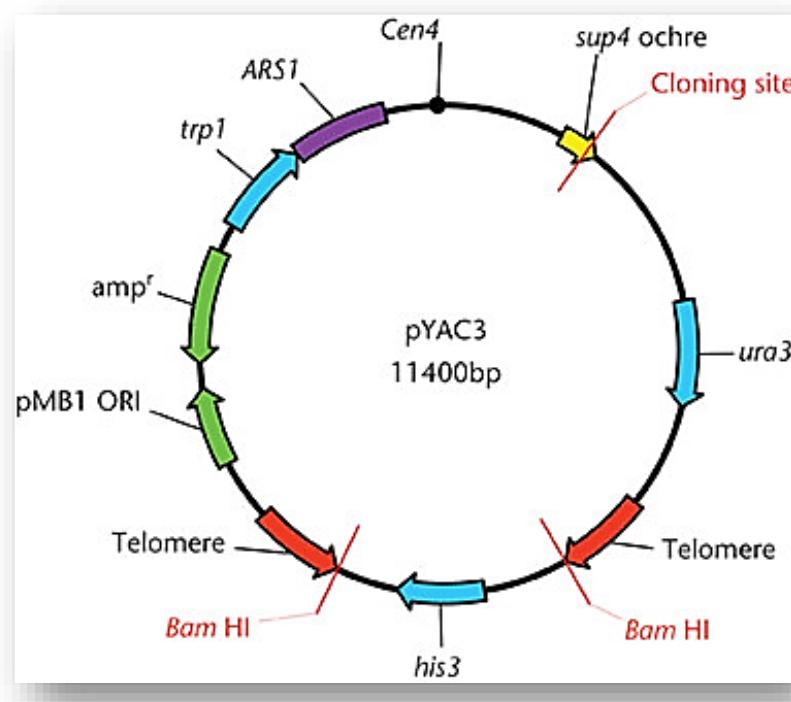
Cloning Vector	Size of DNA That Can Be Cloned	Method of Propagation	Introduction to Bacteria
1. Plasmid ( <b>PLAZMID</b> )	As large as 15 kb	Plasmid replication	Transformation
2. Phage lambda ( <b>λ-FAG</b> )	As large as 23 kb	Phage reproduction	Phage infection
3. Cosmid ( <b>KOZMID*</b> )	As large as 44 kb	Plasmid reproduction	Phage infection
4. Bacterial artificial chromosome ( <b>BAC</b> )	As large as 300 kb	Plasmid reproduction	Electroporation

\* Plazmid sa cos slijedom (*cohesive ends*) iz lambda faga.

5. YAC (Yeast Artificial Chromosome): ima kapacitet 100 kb – 3 Mb



# Kreiranje YAC biblioteke čovjeka:



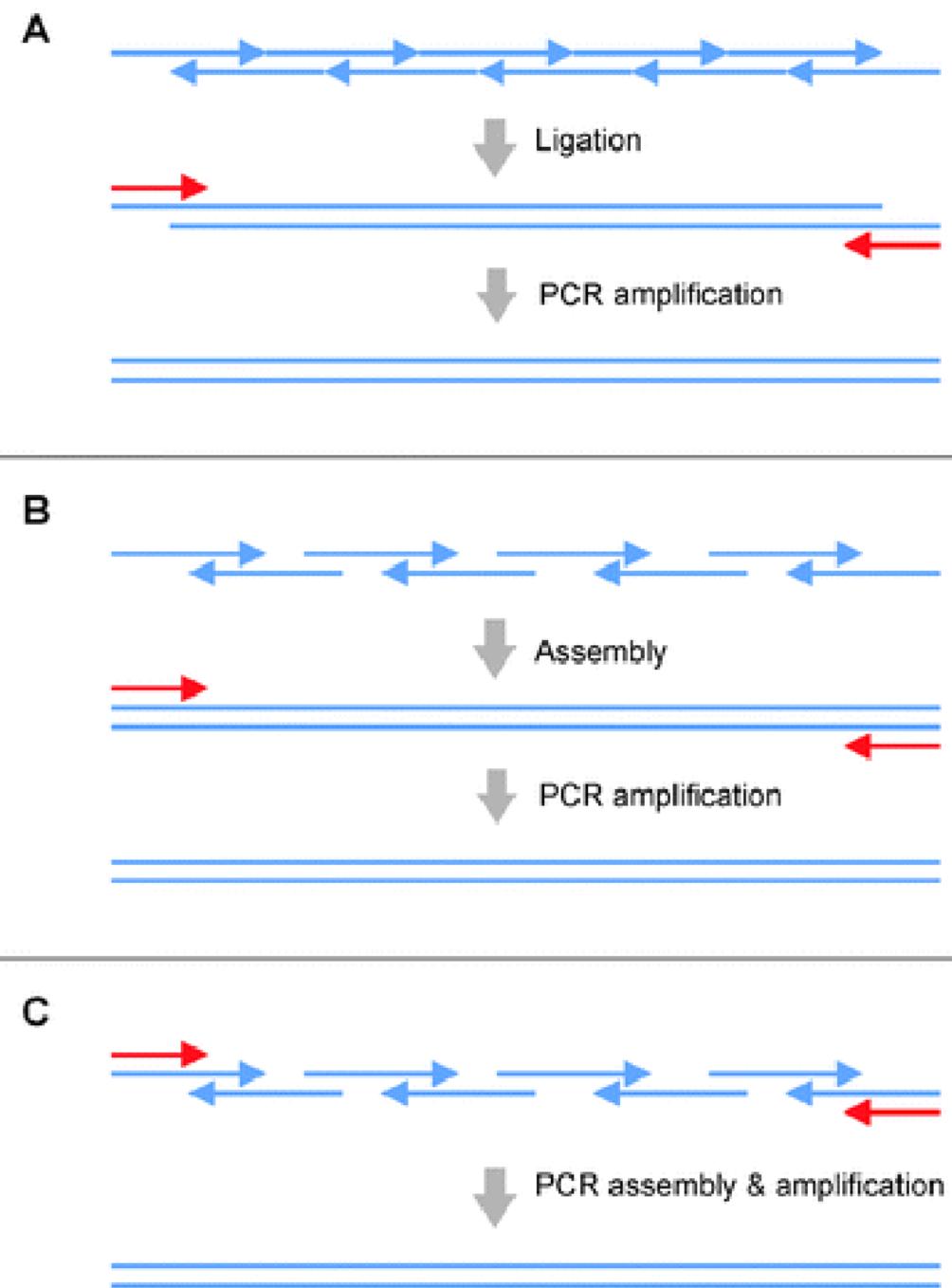
Kod sekvenciranih vrsta danas u principu nisu više potrebne DNA biblioteke jer se geni mogu sintetizirati *de novo* u DNA sintetizatoru (*in vitro*):

The screenshot shows the GENEWIZ website's services page. At the top, there are links for Live Chat, Contact, Language, Login, and Register. Below that is a navigation bar with tabs for SERVICES, RESEARCH AREAS, RESOURCES, COMPANY, and a search icon. Under SERVICES, there are links for Next Generation Sequencing, Sanger Sequencing, Gene Synthesis, Molecular Genetics, Cloning & Mutagenesis, Plasmid DNA Preparation, Oligo Synthesis, and GLP-Compliant Services. A banner below the navigation bar says "Large sequences (genomic synthetic DNA or large genes)". The main content area displays a table showing estimated completion times based on synthetic gene length. Red annotations highlight the "DUŽINA GENA" (Length of gene) column and the "VRIJEME ISPORUKE" (Delivery time) column.

Synthetic Gene Length	Estimated Completion Time*
< 1.5 kb	8-10 Business Days
1.5 kb - 3 kb	10-12 Business Days <i>Now Faster!</i>
3 kb - 5 kb	15-20 Business Days
5 kb - 6 kb	20-25 Business Days
6 kb - 7 kb	25-30 Business Days
7 kb - 8 kb	30-35 Business Days
8 kb - 10 kb	35-40 Business Days
> 10 kb	Custom Quote

<https://www.genewiz.com/en-GB/Public/Services/Gene-Synthesis/Standard>

Kod sekvenciranih vrsta danas u principu nisu više potrebne DNA biblioteke jer se geni mogu sintetizirati *de novo* u DNA sintetizatoru (*in vitro*):



# Sekvenci(oni)ranje DNA

1. Maxam-Gilbert sekvenciraje (prvo)
2. Sanger (dideoksi) sekvenciranje
3. Sekvenciranje nove generacije  
*(next-generation sequencing tzv. NGS)*

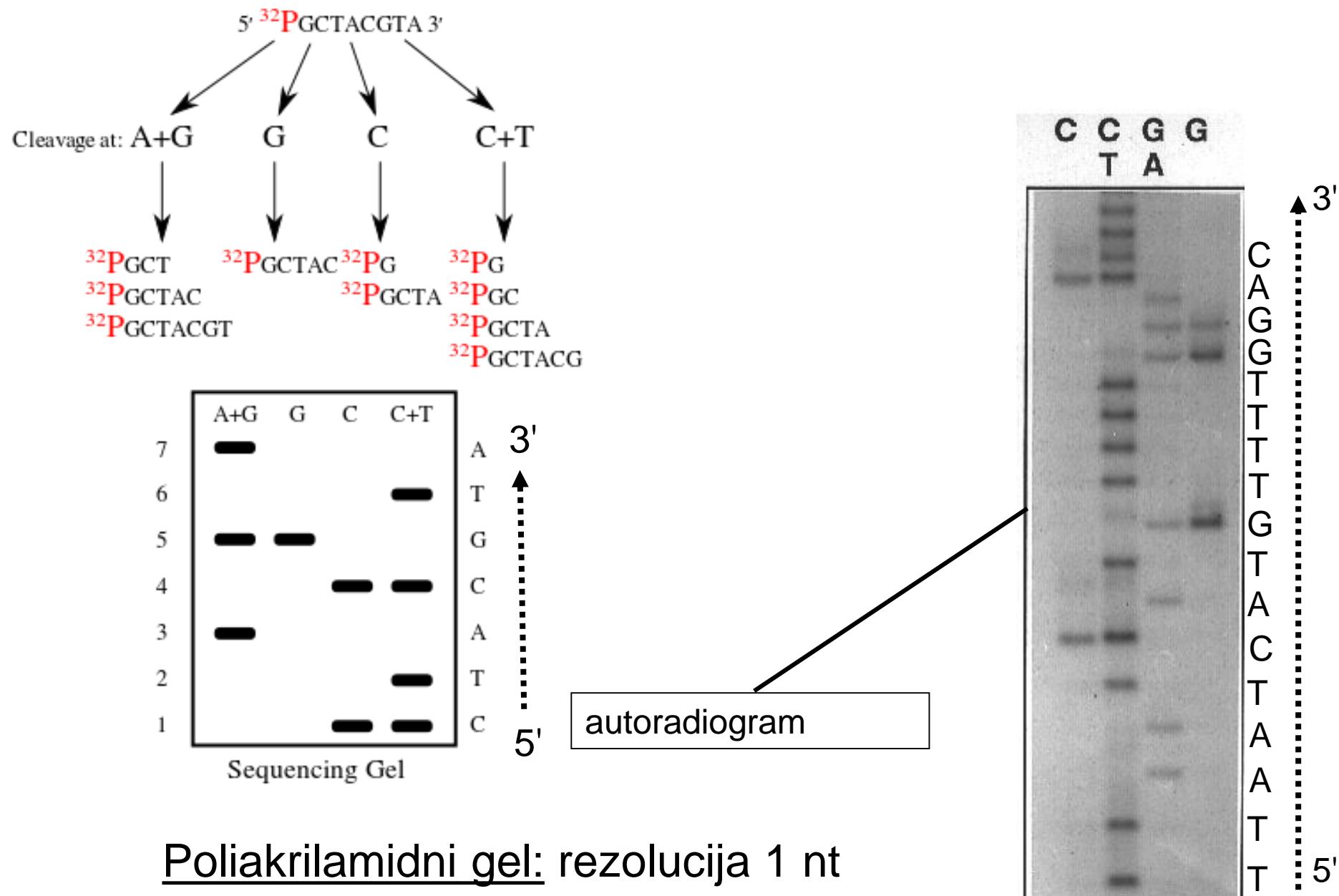


Visokoprotično (masivno paralelno)  
sekvenciranje  
*High Throughput Sequencing (HTS)*

## Metoda kemijskog cijepanja (Maxam-Gilbert): 1976.

- jednolančana DNA (ssDNA) obilježi se s  $^{32}\text{P}$  na 5' kraju
- induciranje kemijskog cijepanja lanca ispred 1 ili 2 vrste baze → u prosjeku 1 cijepanje po molekuli ssDNA
- 4 odvojene reakcije              C, C i T, G i A, G  
                                        1.    2.    3.    4.
- nema koraka PCR umnožavanja

# Metoda kemijskog cijepanja (Maxam-Gilbert): 1976.



# Metoda kemijskog cijepanja (Maxam-Gilbert): 1976.



# Dideoksi metoda (Sangerova metoda):

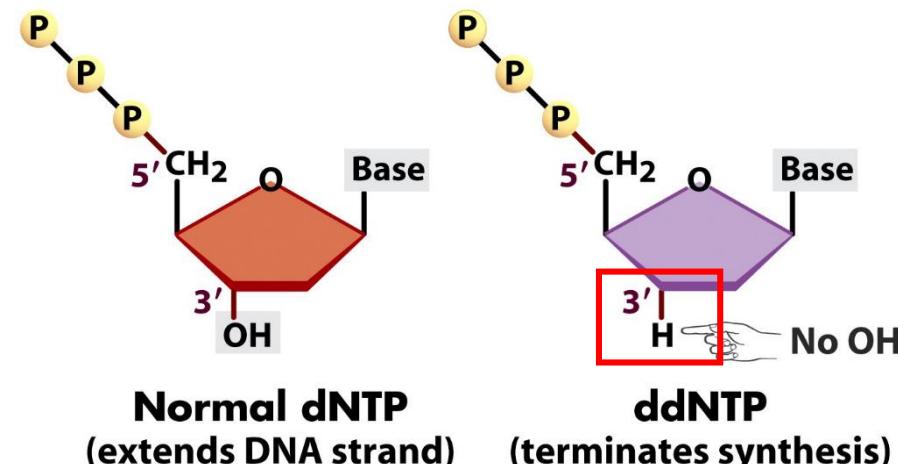
- donedavno *state of the art* (najraširenija) metoda → genom kvasca, ljudski genom, genom uročnjaka (*Arabidopsis*)...



Fred Sanger

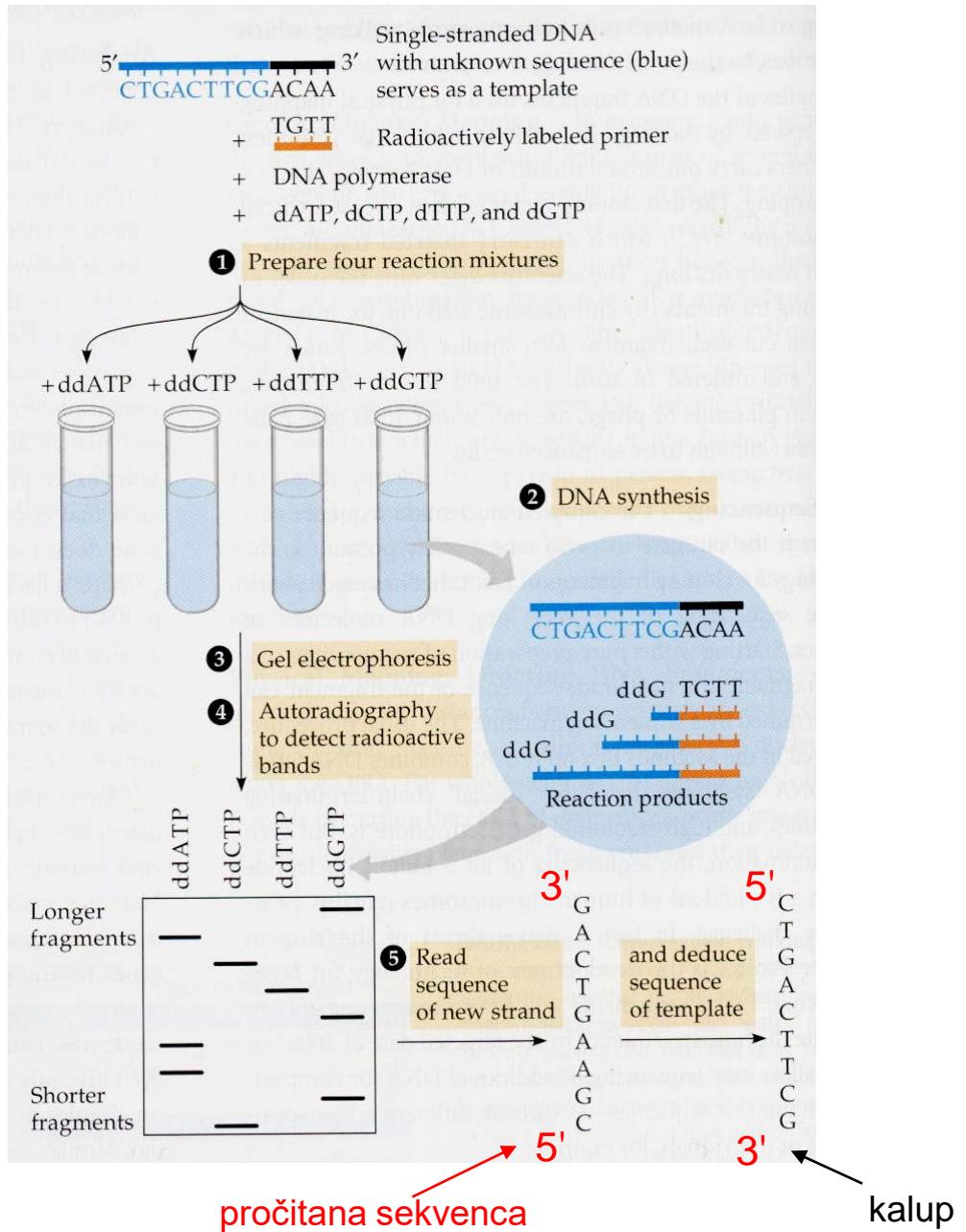
## Starija varijanta dideoksi metode:

- temelji se na *in vitro* reakciji sinteze DNA (PCR):
  - DNA klup (jednolančani)
  - jedan primer/početnica u reakciji (obilježen\*)
  - 4 dNTP-a + 1 ddNTP po reakciji (4 odvojene reakcije)
  - DNA polimeraza I



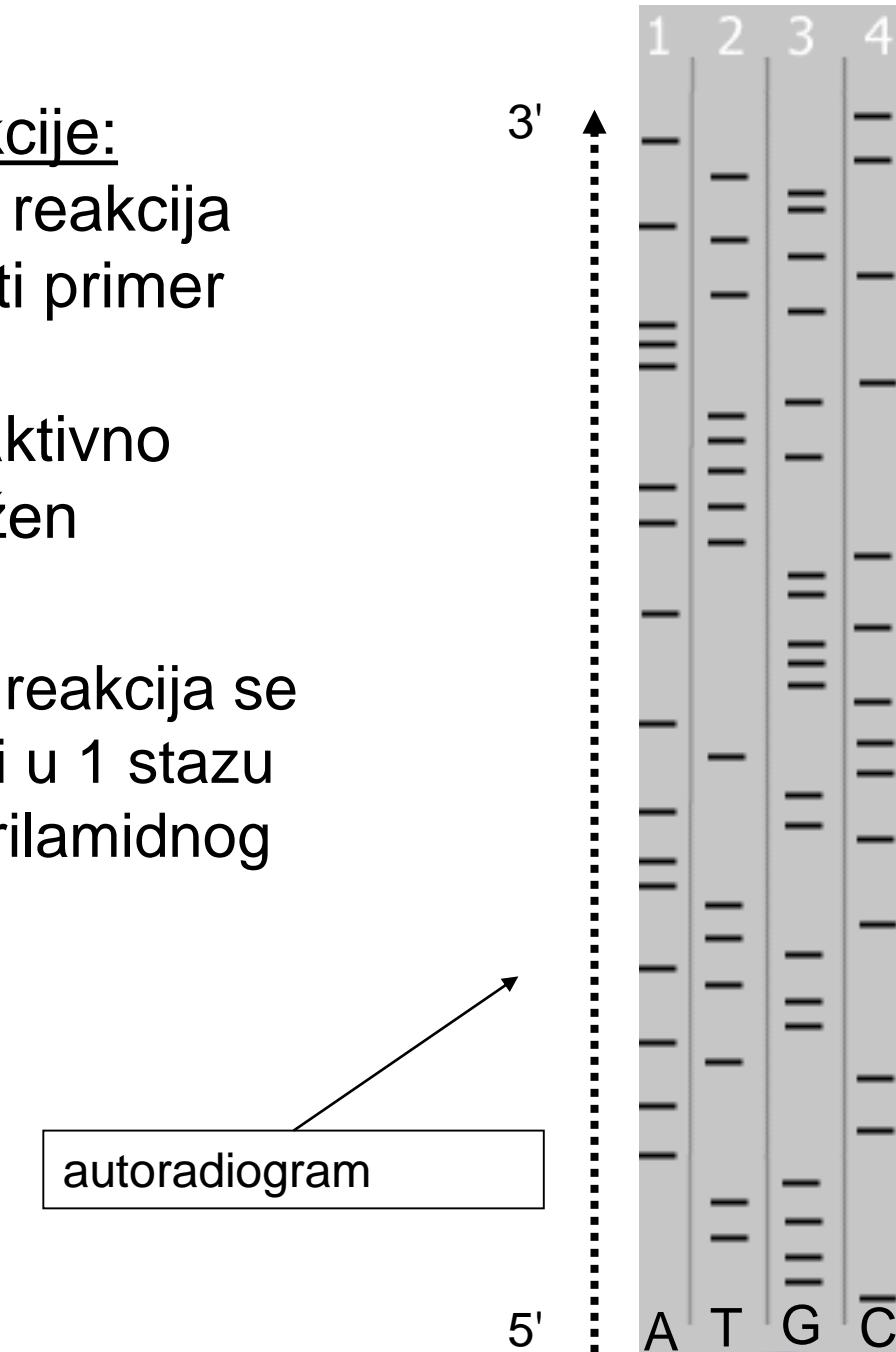
\* radioaktivno ili fluorescentno

# Starija varijanta dideoksi metode:

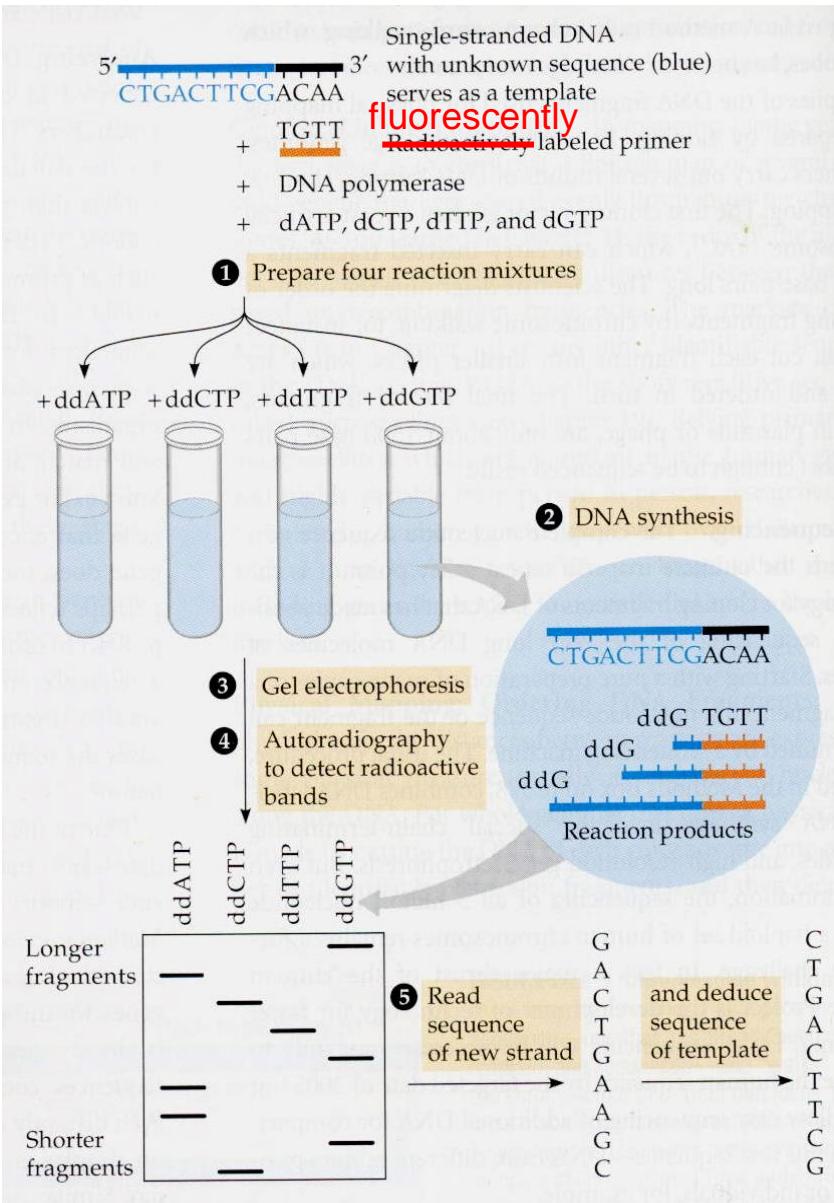


- **4 reakcije:** svaka reakcija ima isti primer koji je radioaktivno obilježen

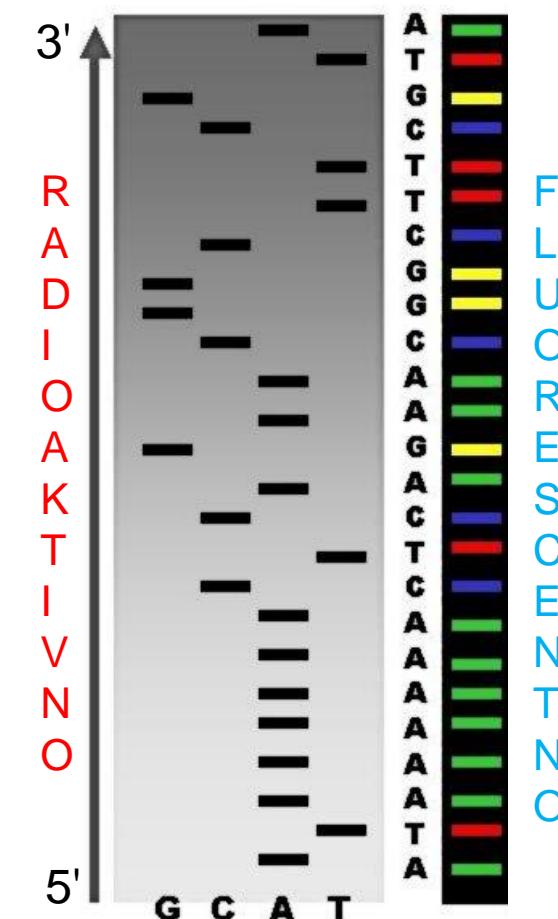
- svaka reakcija se nanosi u 1 stazu poliakrilamidnog gela



# Starija varijanta dideoksi metode:



- **fluorescentno obilježen primer**
- **4 reakcije:** svaka reakcija ima isti primer, ali obilježen različitom bojom (4 boje)
- 4 reakcije se spoje i “voze” u istoj stazi poliakrilamidnog gela
- **automatizacija:** detektor “čita” slijed DNA prema boji fragmenta
- ovako se može pročitati 800-1000 nt



# “ddGTP” reakcija: linearno umnažanje

5' -GAATGTCCTTCTCTAAGTCCTAAAG  
3' -GGAGACTTACAGGAAAGAGAGATTCAAGGATTCAAGGAGGCCTACCATGAAAGATCAAG-5'

5' -GAATGTCCTTCTCTAAGTCCTAAAGTCCTCCG  
3' -GGAGACTTACAGGAAAGAGAGATTCAAGGATTCAAGGAGGCCTACCATGAAAGATCAAG-5'

5' -GAATGTCCTTCTCTAAGTCCTAAAGTCCTCCCG  
3' -GGAGACTTACAGGAAAGAGAGATTCAAGGATTCAAGGAGGCCTACCATGAAAGATCAAG-5'

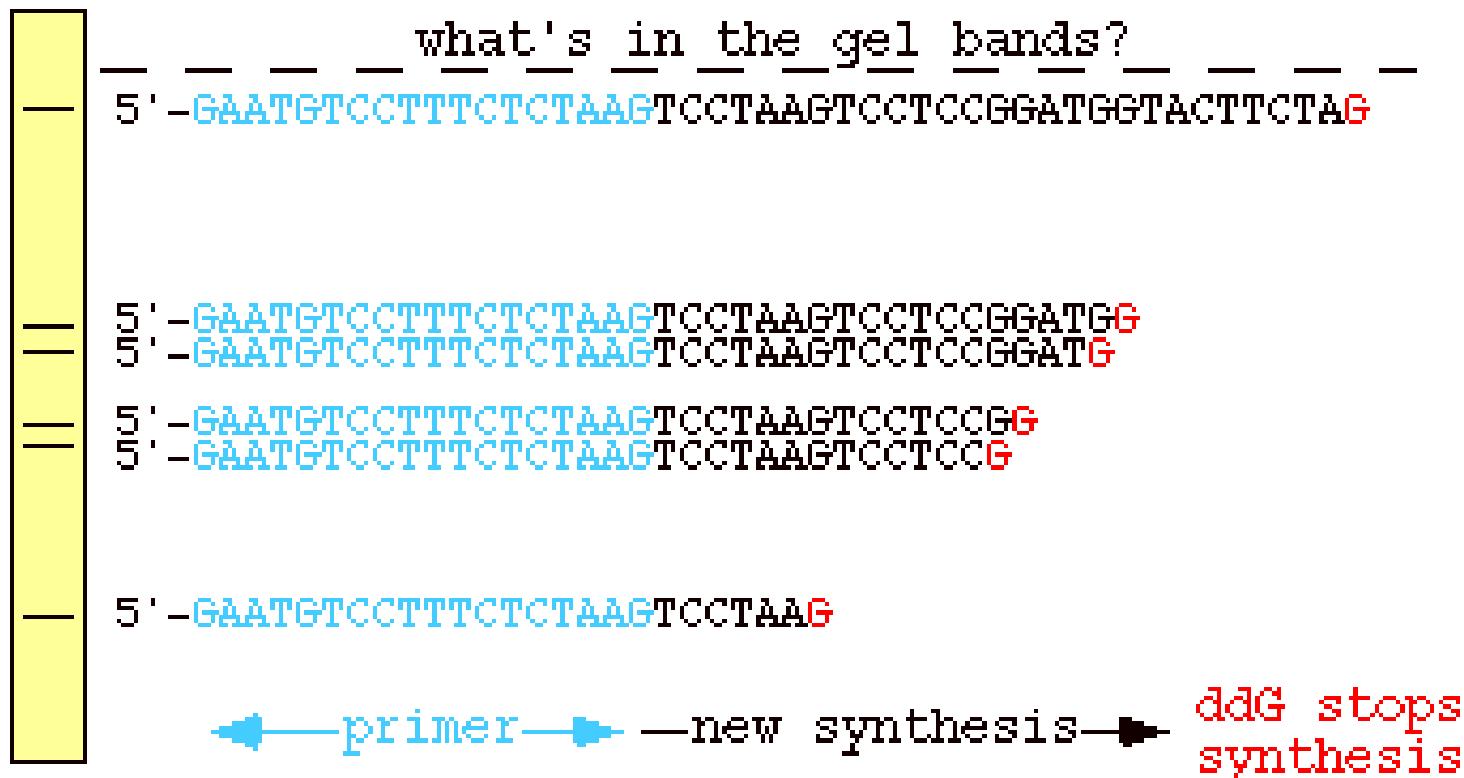
5' -GAATGTCCTTCTCTAAGTCCTAAAGTCCTCCGGATG  
3' -GGAGACTTACAGGAAAGAGAGATTCAAGGATTCAAGGAGGCCTACCATGAAAGATCAAG-5'

5' -GAATGTCCTTCTCTAAGTCCTAAAGTCCTCCGGATGG  
3' -GGAGACTTACAGGAAAGAGAGATTCAAGGATTCAAGGAGGCCTACCATGAAAGATCAAG-5'

5' -GAATGTCCTTCTCTAAGTCCTAAAGTCCTCCGGATGGTACTTCTAG  
3' -GGAGACTTACAGGAAAGAGAGATTCAAGGATTCAAGGAGGCCTACCATGAAAGATCAAG-5'

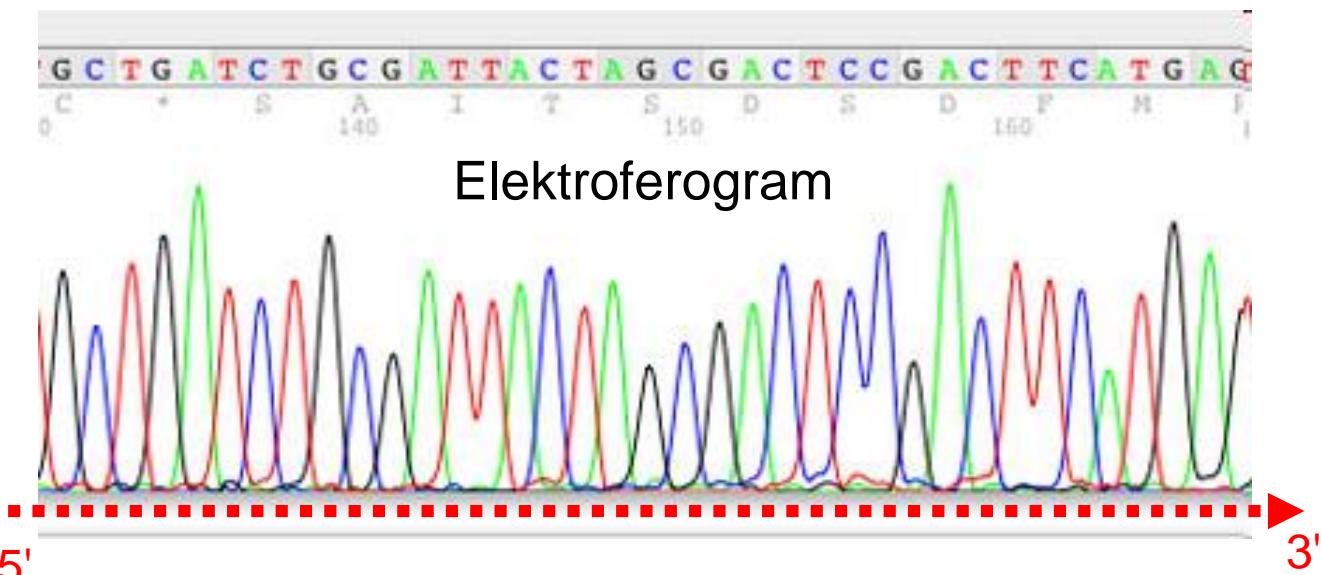
# "ddGTP" reakcija: linearno umnažanje

Polyacrylamide gel electrophoresis of the "G" reaction



# Suvremena varijanta didoksi metode: fluorescentno obilježeni ddNTP-ovi (ne primjeri kao u prošloj varijanti!)

- temelji se na jednoj PCR reakciji (ne 4 kao prije):
  - DNA klup
  - jedan neobilježeni primer (linearno umnažanje)
  - 4 dNTP-a + 4 različito fluorescentno obilježena ddNTP-a (bitan je omjer)
  - Taq DNA polimeraza
- kapilarna elektroforeza + detektor fluorescencije

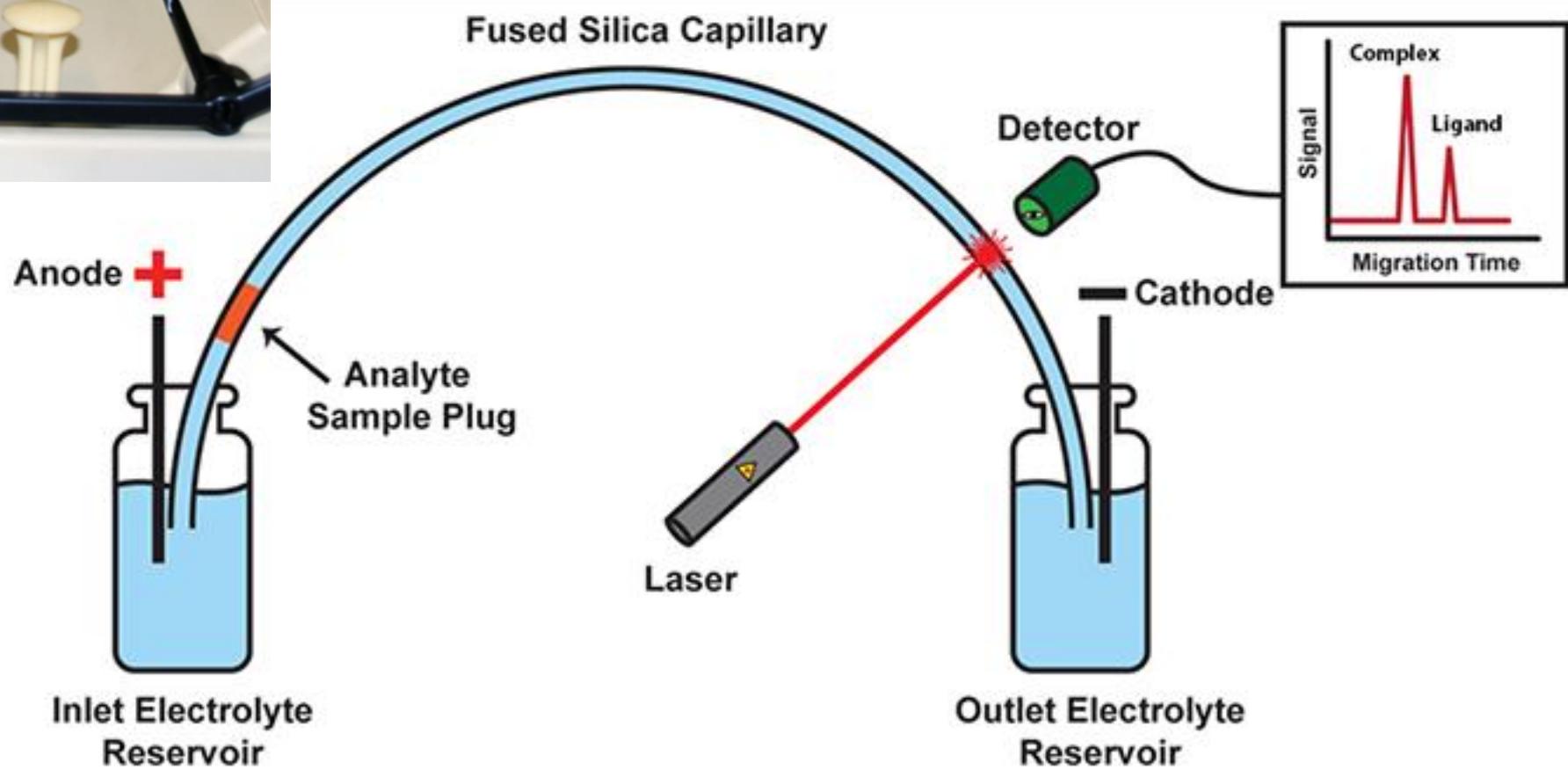
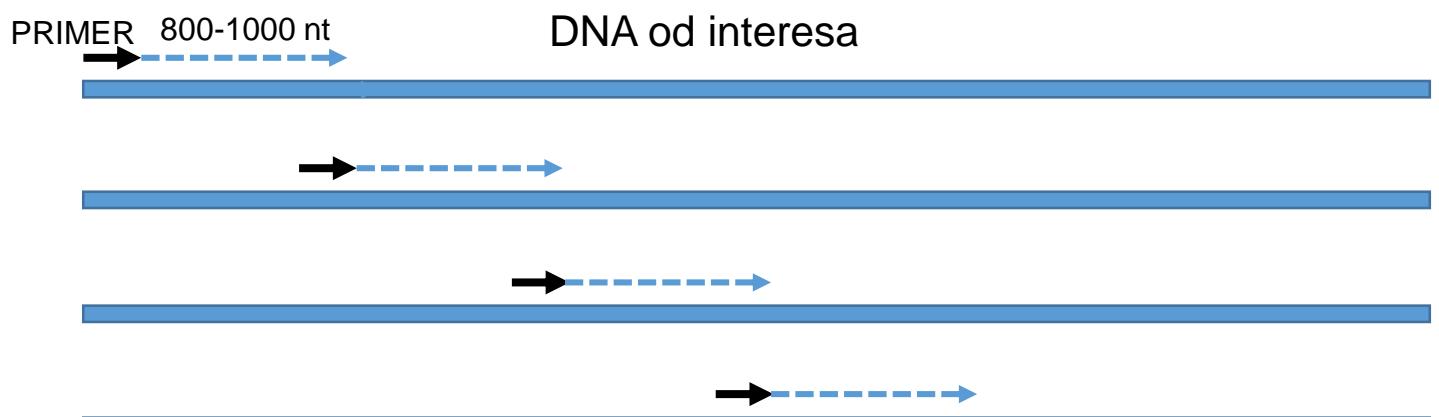
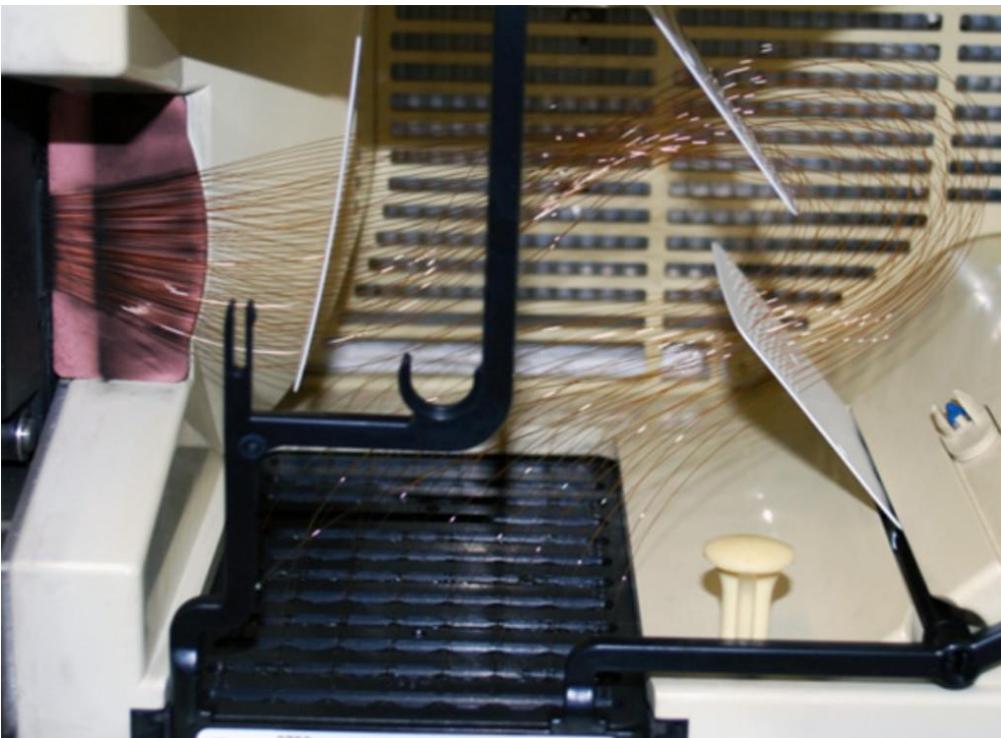


Elektroferogram

- 800-1000 nt se može pročitati u jednoj reakciji



sekvencer (sekvator)



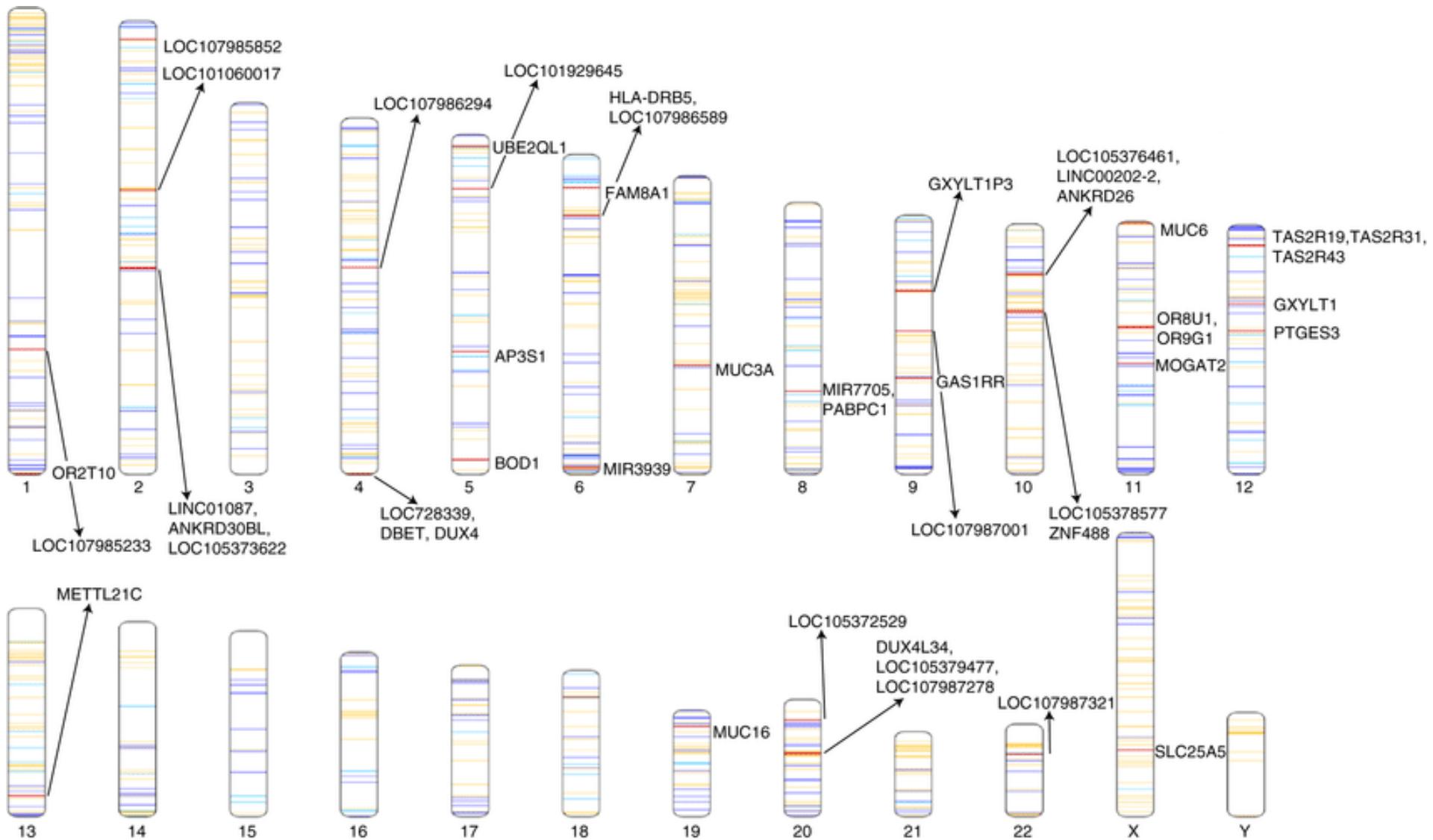
# Sanger-ova dideoksi metoda sekvenciranja:



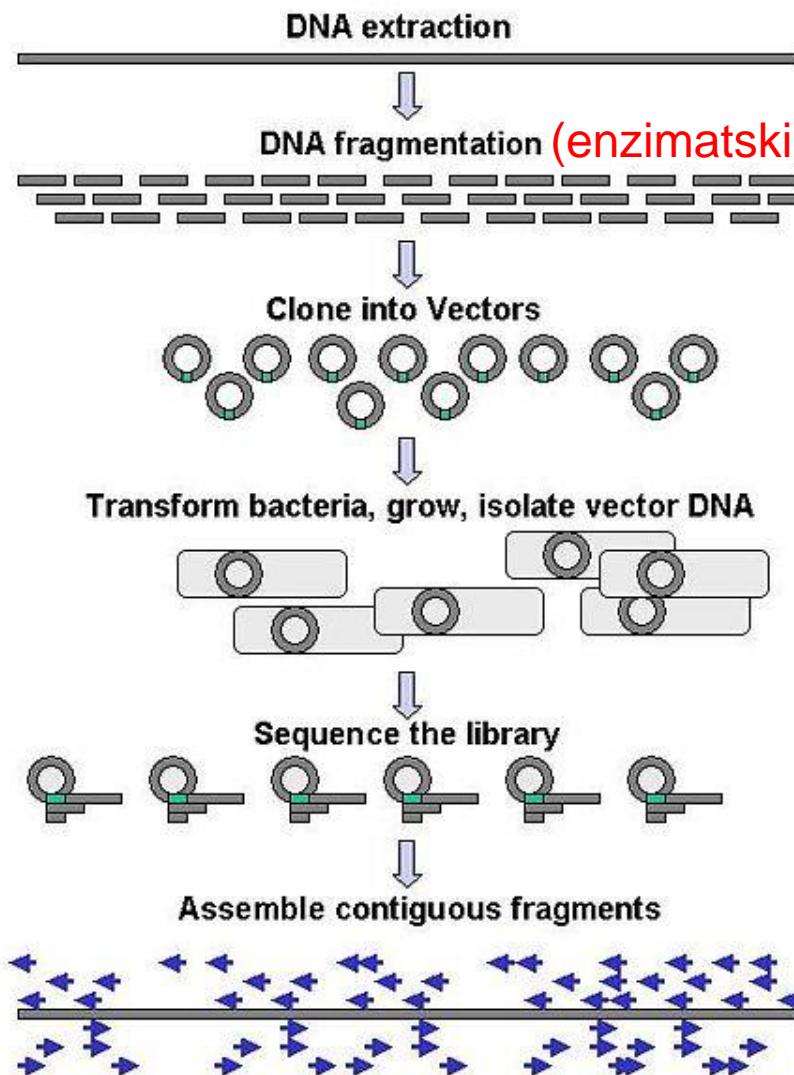
HHMI

<http://www.hhmi.org/bioInteractive/sanger-method-dna-sequencing>

# “Klasično” sekvenciranje genoma: 1987.-2001.



# “Klasično” sekvenciranje genoma: 1987.-2001.



- enzimska fragmentacija genoma i izrada DNA biblioteka
- mapiranje fragmenata na postojeću genetičku kartu

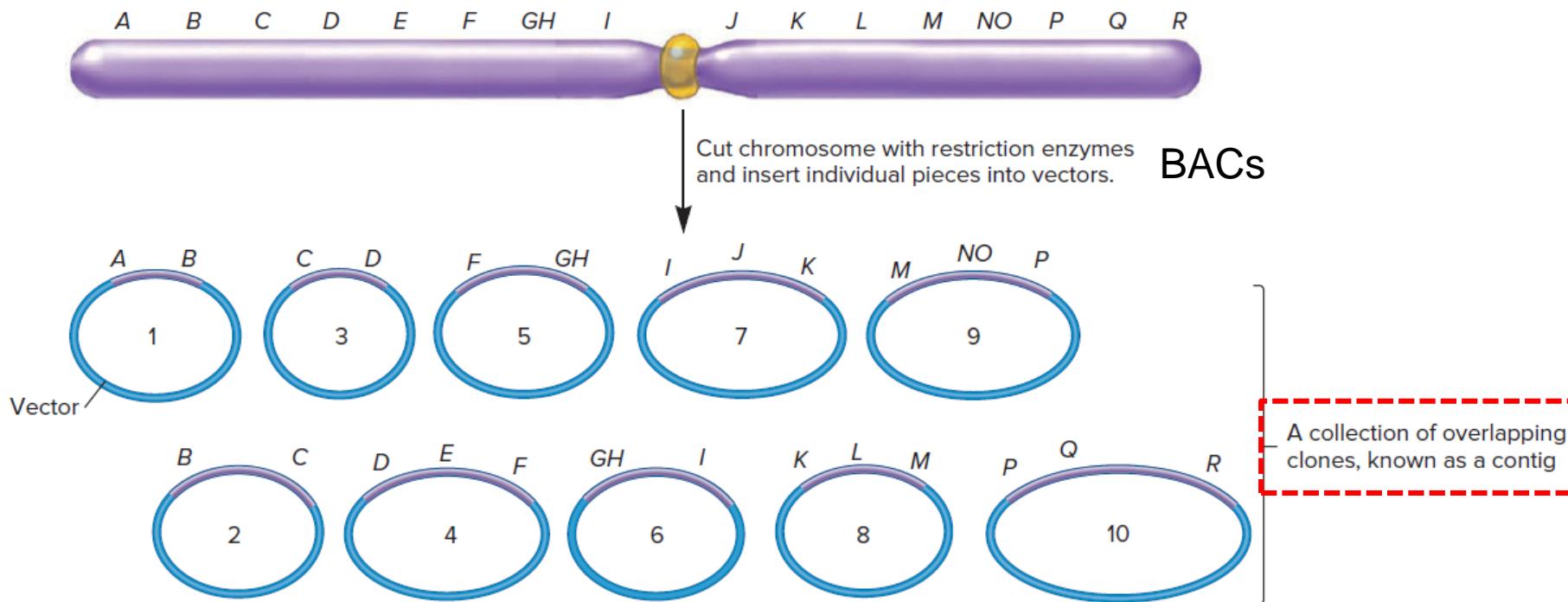


Francis Collins, NIH

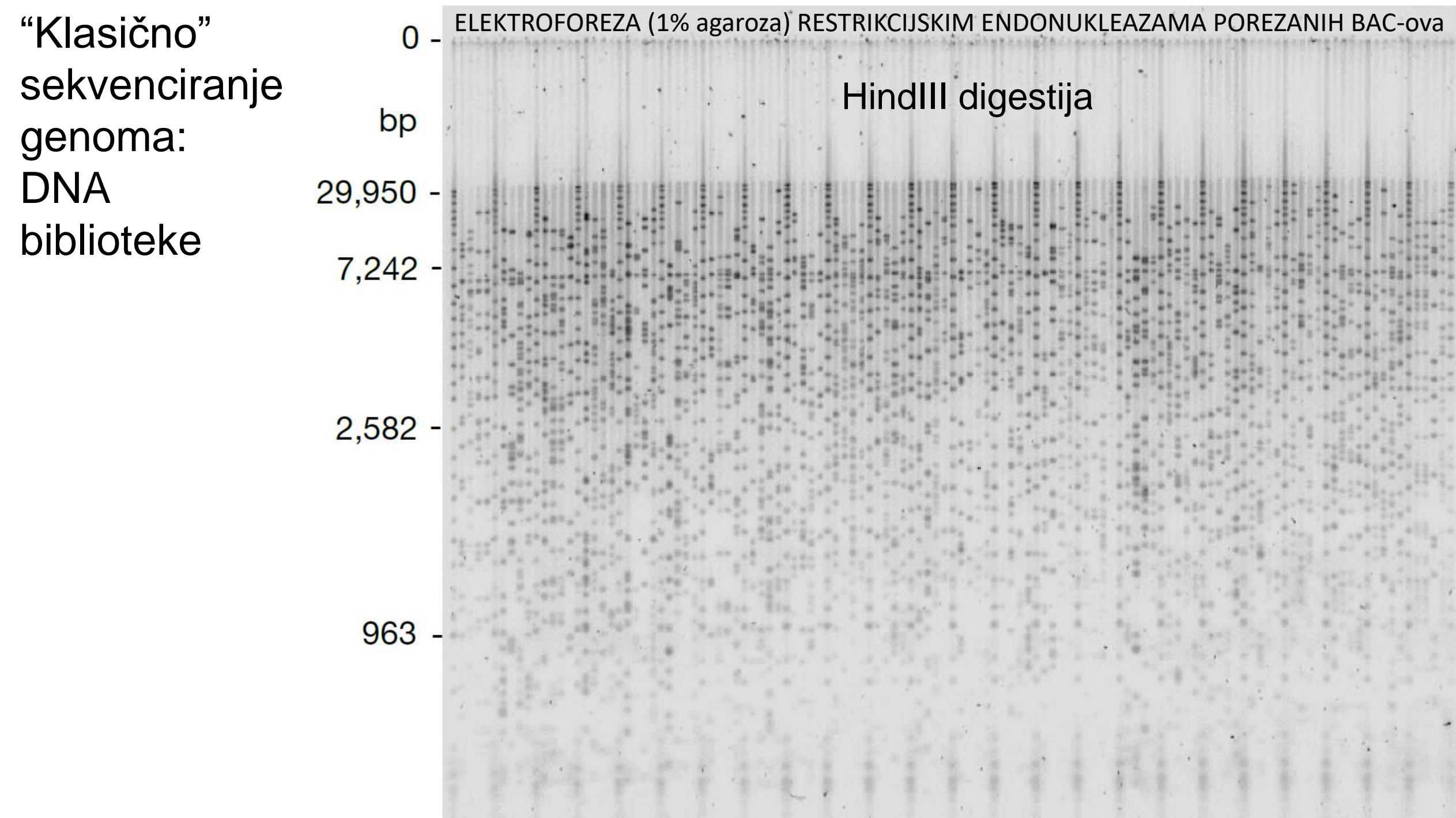
Deset godina je trebalo da se ovako sastavi sekvenca ljudskog genoma.

# “Klasično” sekvenciranje genoma: DNA biblioteke

- *contig*: kontinuirani slijed klonova neke DNA biblioteke



Svaki klon se mogao pozicionirati (mapirati) na pojedini kromosom pomoću od prije poznatih (genetičkih ili DNA) markera.



# Klasično sekvenciranje genoma: DNA biblioteke

**Table 2 Status of FPC database after automated assembly and manual editing**

a { Ctg18846 of jmcpthers\_Humanmap, Clones 287 of 836, Markers 193 of 193, Sequenced 77, Length 1552

b { 8-Cda0xh11 RH122469 8-D8S1962 D8S1436E 8-SHGC-14258 8-stSG10128 8-AFM185xe9 RH112919 8-A005A44 8-SGC35843  
 8-sts-H33187 RH112923 8-sts-N48595 RH106071 RH69712 8-stSG21444 8-D8S275 8-AFM126ZG5 8-stSG15627 8-stSG31  
 8-stSG9774 RH122498 8-sts-G20703 RH35282 8-stSG2293 8-stSG35987 8-D8S1917 8-SHG-13042 8-WIAF-96 8-stSG43c  
 D8S1408E 8-D8S1158 RH76944 A006C31 8-D8S1456 RH105891 RH106252 8-SHG-37027 RH120996 8-AFM8315YQ5 8-stSG51  
 D1751978 8-sts-N55108 8-stSG50241 ebF31g02.b1 8-RH100159 8-A006H11 8\_3640F GDB:197842 8-stSG47162 RH122077  
 8-H69841 8-stSG21232 8-WI-16808 RH83842 8-stSG47484 ebF40a02.b1 8-sts-X62167 8-D8S525 8-D8S1976 D8S333E

c { N0381H08\* N0157I04+ N0021K19 N0278015\* N0099H20 N0653B10+ N0443F22\* N0517F08-  
 N0486C06+ N0792N11\* N0303B18+ N0659D20\* N0181D19+ N0370E22\* N0586I07+ N0510I16+  
 N0352B03\* N0638004 N0080H18+ N0661E18\* N0661J03\* N0699E07\* N0509L05 N0656A16+  
 N0038L13 N0423H14+ N0172E10+ N0034M16 N0647B24+ N0785J03\* N0643H21+ N0731H10 N0051M18-  
 N0119K07 N0096I13+ N0286F13+ N0392J24+ N0562L10 N0277G05\* N0354C09+ N0296016  
 N0417F24 N0286G10 N0571J10 N0182A18 N0426C08+ N0685K06 N0644H20\* N0494015  
 N0036K24\* N0062F09+ N0028G16 N0669A11 N0008H12+ N0008H12+ N0388A07\* N0494015  
 N047C24 N0082J02+ N0363E06+ N0690A11 N0339A23\* N0653D22\* N0627J18+ N0674F15+ N0736J17\* NC  
 N0682K24\* N0035L24 N0742F19 N0741F12+ N0163M17+ N0171L17+ N0674A24+ N0809H11- NC  
 N02G05\* N0075G10+ N0035L24 N0128I16 N0117H07+ M2003P13 N0643H07 N0170P03+ N076:  
 N0624P08 N046BT11+ N0720F01+ N0436T21+ N0390M09 N0678C22 N0320G23 N054:  
 N0813B11+ N0079N04+ N0336E22+ N0733M18\* N0309F04+ N0353B05 N0083M2  
 N0663N03 N0764J10 N0698C24+ N0520I03 N0577I23\* N0813B08 N0304E0-  
 N0125J17 N0019D14+ N0067F19+ N0256A09 N0578H14+ N0097P18

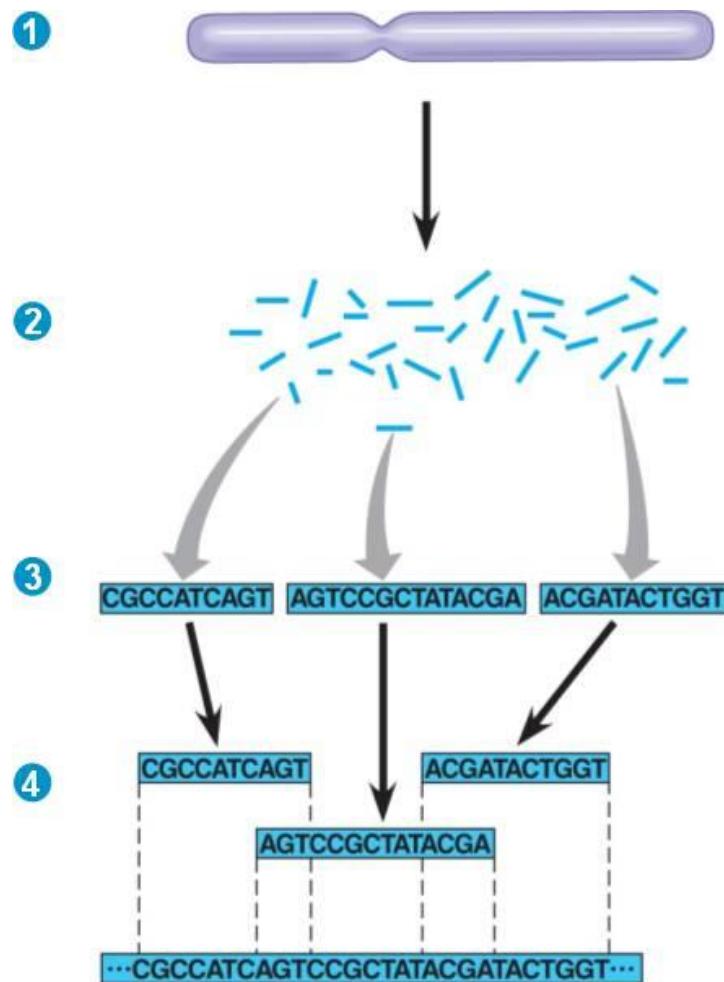
d { JMF-8q21.1 AC012533 JMF-8q21 AC026617:N0769H01 AC024371 AC027701 JMF-8q21 JMF-8q21.1 JMF-9q22 JM  
 AC022598 JMF-8q21.1 COX\_8 AC022647 JMF-8q21 AC010000 COX\_8 JMF-8q21.1 AC069359 AC069139 RD022821 AC  
 COX\_8 COX\_8 COX\_8 JMF-8q21.1 AC018616 AC013542 JMF-8q21.1 JMF-8q21 AC073282  
 JMF-8q21 AC022778 AC009902 JMF-8q21.1 JMF-8q21 AC018882 AC019252 AC060765 AC018443  
 AC024367 COX\_8 COX\_14 AF181449:G0402C07 JMF-8q21 JM02260 AC016875

e { 8-stSG10097 8-Cda01e03 8-stSG2255 8-AFM185xe9  
 8-stSG27065 8-stSG4799 8-stSG42447 8-sts-X62167  
 8-stSG54782 8-stSG40118 8-stSG53462 8-A005A44  
 8-WI-6959 8-AFM255yb1 8-WI-16645 8-stSG15627  
 8-sts-J04156 8-WI-15063 8-Cda0xh11 8-stSG47162

	Automated assembly	Manually edited database
Date	December 1999	September 2000
BAC clones in FPC	283,287	372,264
Number of contigs	7,133	1,447
Clones in contigs	264,555	295,828
Number of singletons*	18,732	76,436
Contigs containing:		
>25 clones	3,012	912
9–25 clones	1,844	260
3–9 clones	1,957	204
2 clones	887	71

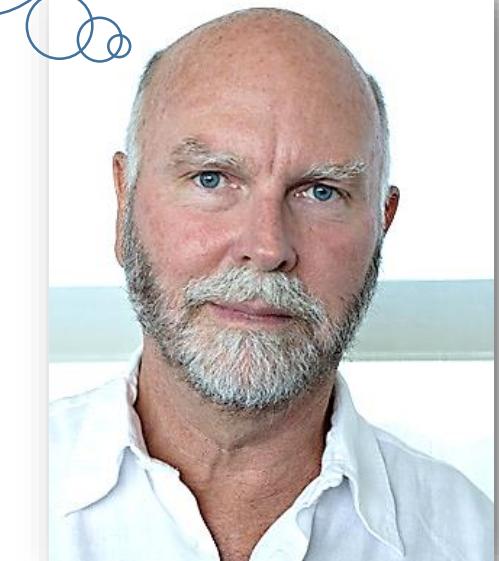
\* Clones not incorporated into any contig; see text.

# „Shotgun“ sekvenciranje genoma:



- fragmentacija genoma fizičkim putem: „*brute force approach*“ (ultrazvuk)
- slaganje fragmenata u cjeloviti genom (bez DNA biblioteke): bioinformatički alati

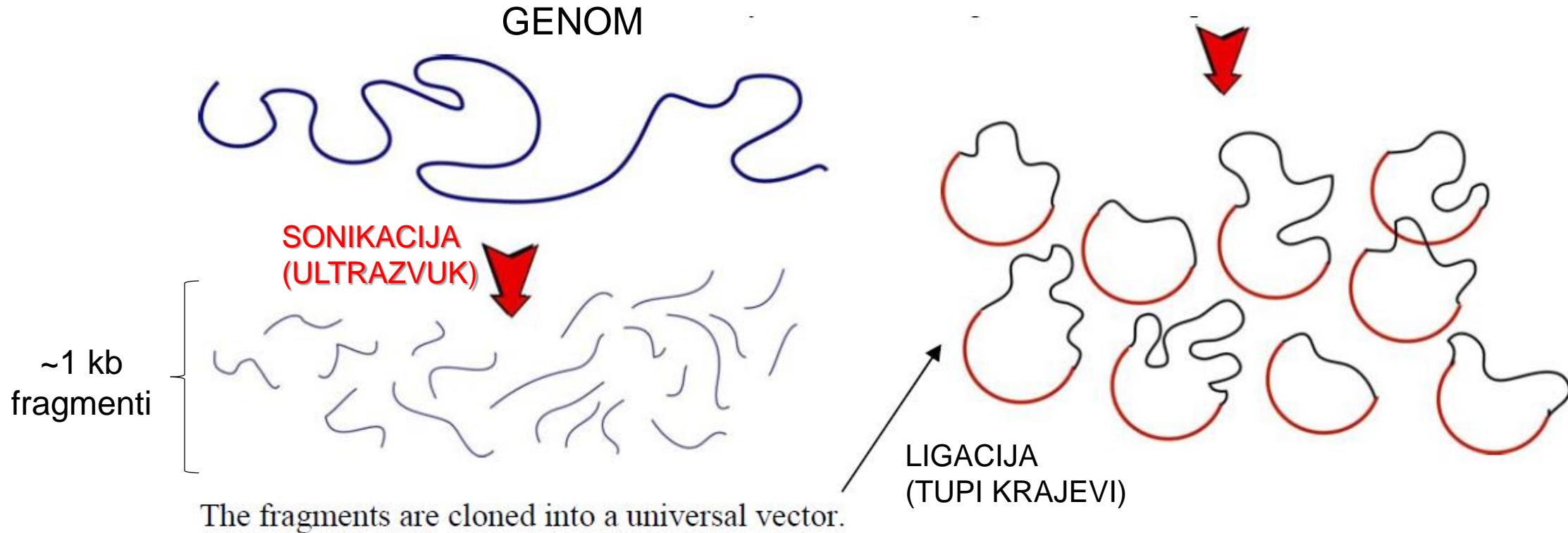
Dvije godine je trebalo da se ovako sastavi sekvenca ljudskog genoma.



Craig Venter



# „Shotgun“ sekvenciranje genoma:

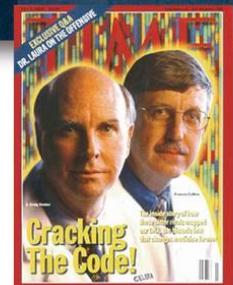
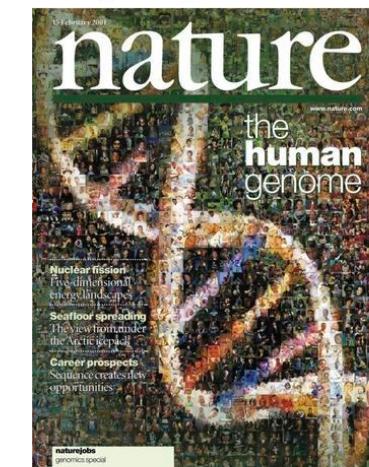


- nakon fragmentacije, a prije ligacije, fragmenti se tretiraju s Klenow fragmentom (podjedinica DNA polimeraze I) ili T4 DNA polimerazom da se dobiju tupi krajevi

# „Shotgun“ sekvenciranje genoma:



- svaki klon se sekvencira dvosmjerno
- cilj je posložiti sve sekvene u kontinuirani slijed klonova (bez praznina)

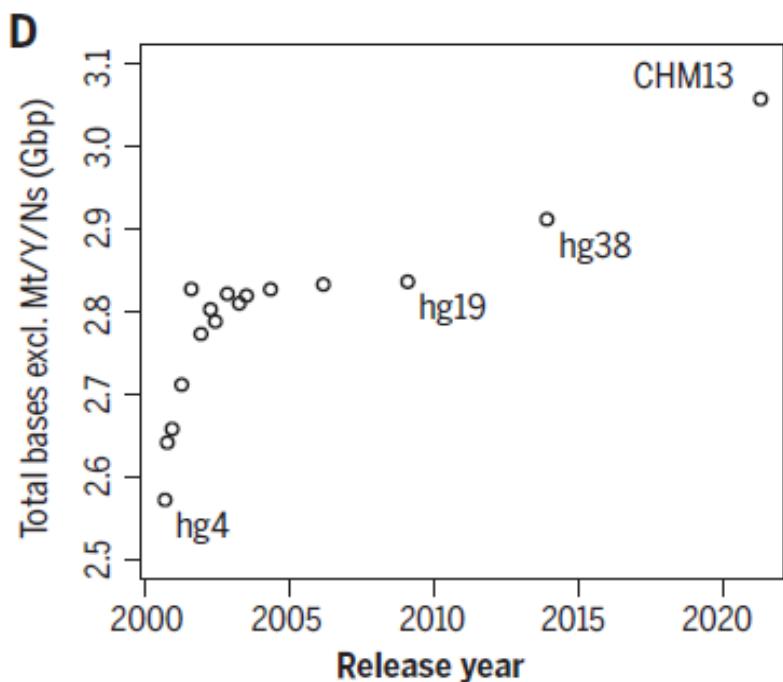


2001.

# Potpuna sekvenca ljudskog genoma: $3.055 \times 10^9$ pb

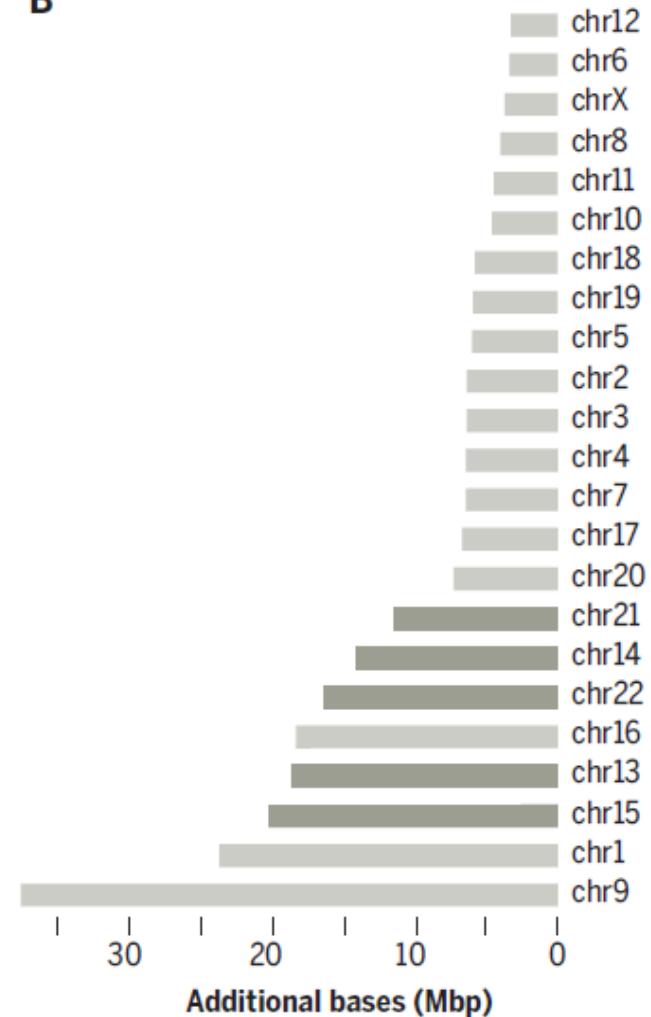


Nurk et al. (2022) Science



Telomere-to-Telomere (T2T) Consortium  
Human Pangenome Reference Consortium

B



# The Human Pangenome Project: a global resource to map genomic diversity

<https://doi.org/10.1038/s41586-022-04601-8>

Received: 30 August 2021

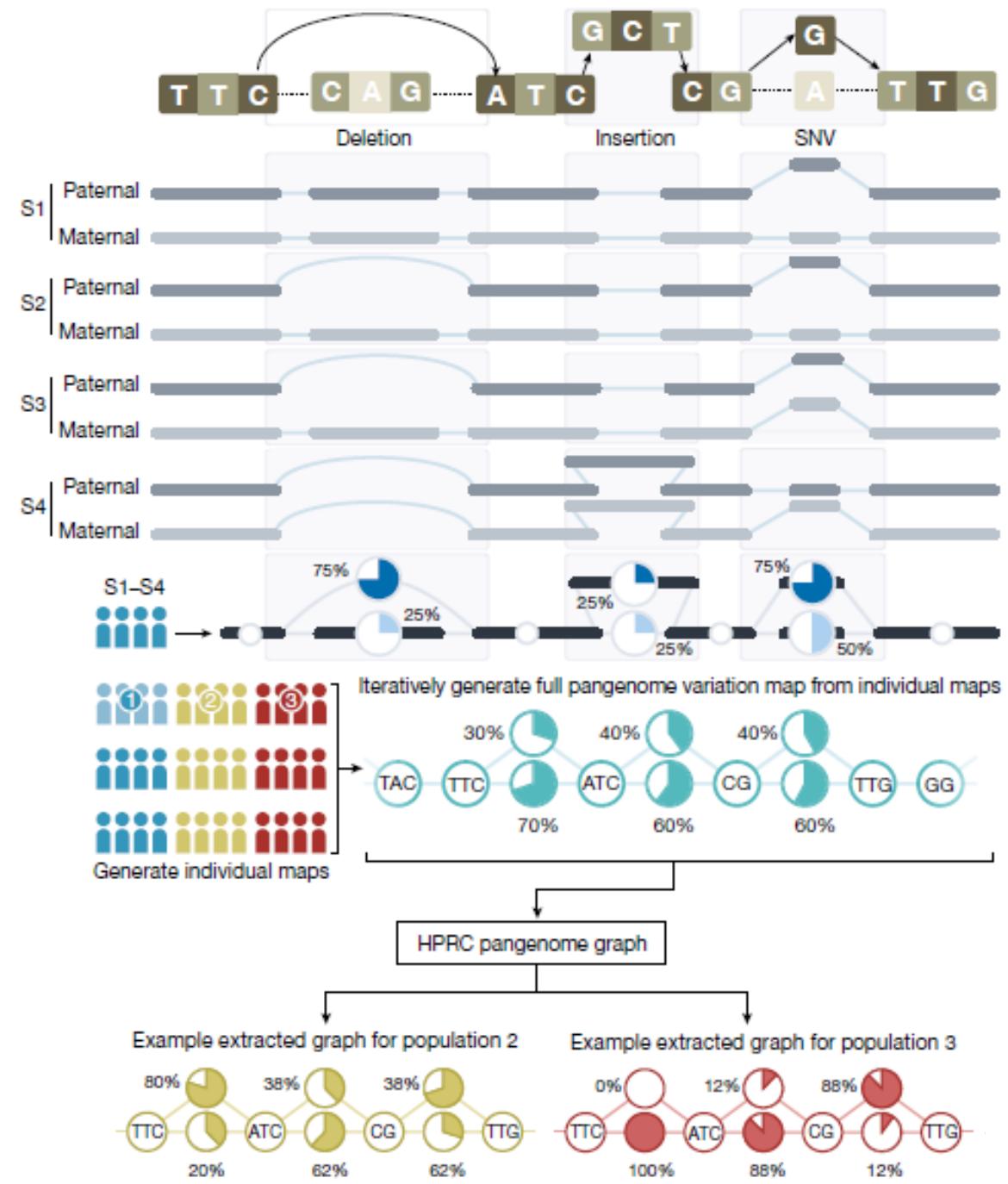
Accepted: 1 March 2022

Published online: 20 April 2022

Check for updates

Ting Wang<sup>1,2,3</sup>, Lucinda Antonacci-Fulton<sup>3</sup>, Kerstin Howe<sup>4</sup>, Heather A. Lawson<sup>1</sup>, Julian K. Lucas<sup>5</sup>, Adam M. Phillippy<sup>6</sup>, Alice B. Popejoy<sup>7</sup>, Mobin Asrif<sup>8</sup>, Caryn Carson<sup>1,2,3</sup>, Mark J. P. Chaisson<sup>8</sup>, Xian Chang<sup>5</sup>, Robert Cook-Deegan<sup>9</sup>, Adam L. Felsenfeld<sup>10</sup>, Robert S. Fulton<sup>3</sup>, Erik P. Garrison<sup>11</sup>, Nanibaa' A. Garrison<sup>12,13,14</sup>, Tina A. Graves-Lindsay<sup>3</sup>, Hanlee Ji<sup>15</sup>, Eimear E. Kenny<sup>16,17,18</sup>, Barbara A. Koenig<sup>19</sup>, Daofeng Li<sup>1,2,3</sup>, Tobias Marschall<sup>20</sup>, Joshua F. McMichael<sup>3</sup>, Adam M. Novak<sup>5</sup>, Deepak Purushotham<sup>1,2,3</sup>, Valerie A. Schneider<sup>21</sup>, Baergen I. Schultz<sup>10</sup>, Michael W. Smith<sup>10</sup>, Heidi J. Sofia<sup>10</sup>, Tsachy Weissman<sup>22</sup>, Paul Flicek<sup>23</sup>, Heng Li<sup>24,25</sup>, Karen H. Miga<sup>5</sup>, Benedict Paten<sup>6</sup>, Erich D. Jarvis<sup>26,27</sup>, Ira M. Hall<sup>28</sup>, Evan E. Eichler<sup>29,30</sup>, David Haussler<sup>5,31</sup> & the Human Pangenome Reference Consortium\*

- reprezentiranost alelnih varijanti u (svim) populacijama
- jasno razlikovanje haploitipova (haploidnih genotipova): jednog naslijedjenog od majke, drugog od oca



# Sekvenci(oni)ranje genoma čovjeka

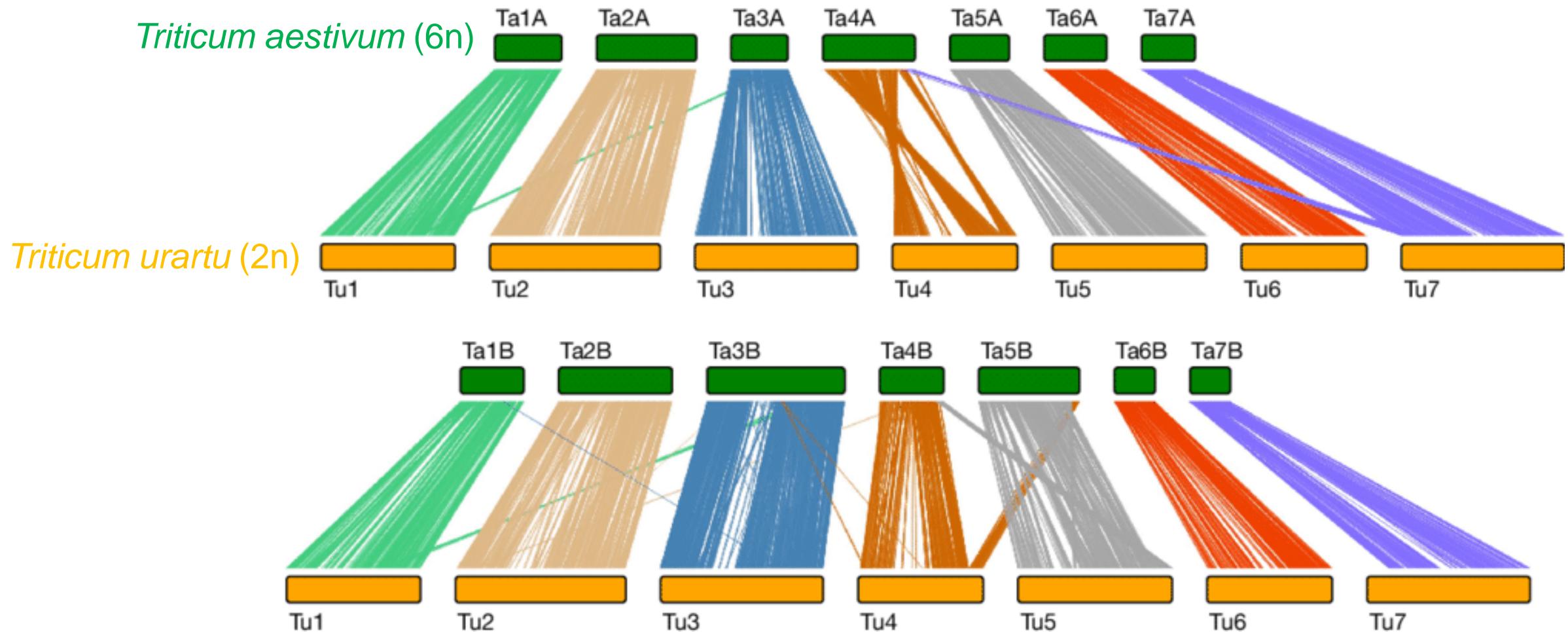


HHMI

<http://www.hhmi.org/bioInteractive/human-genome-sequencing>

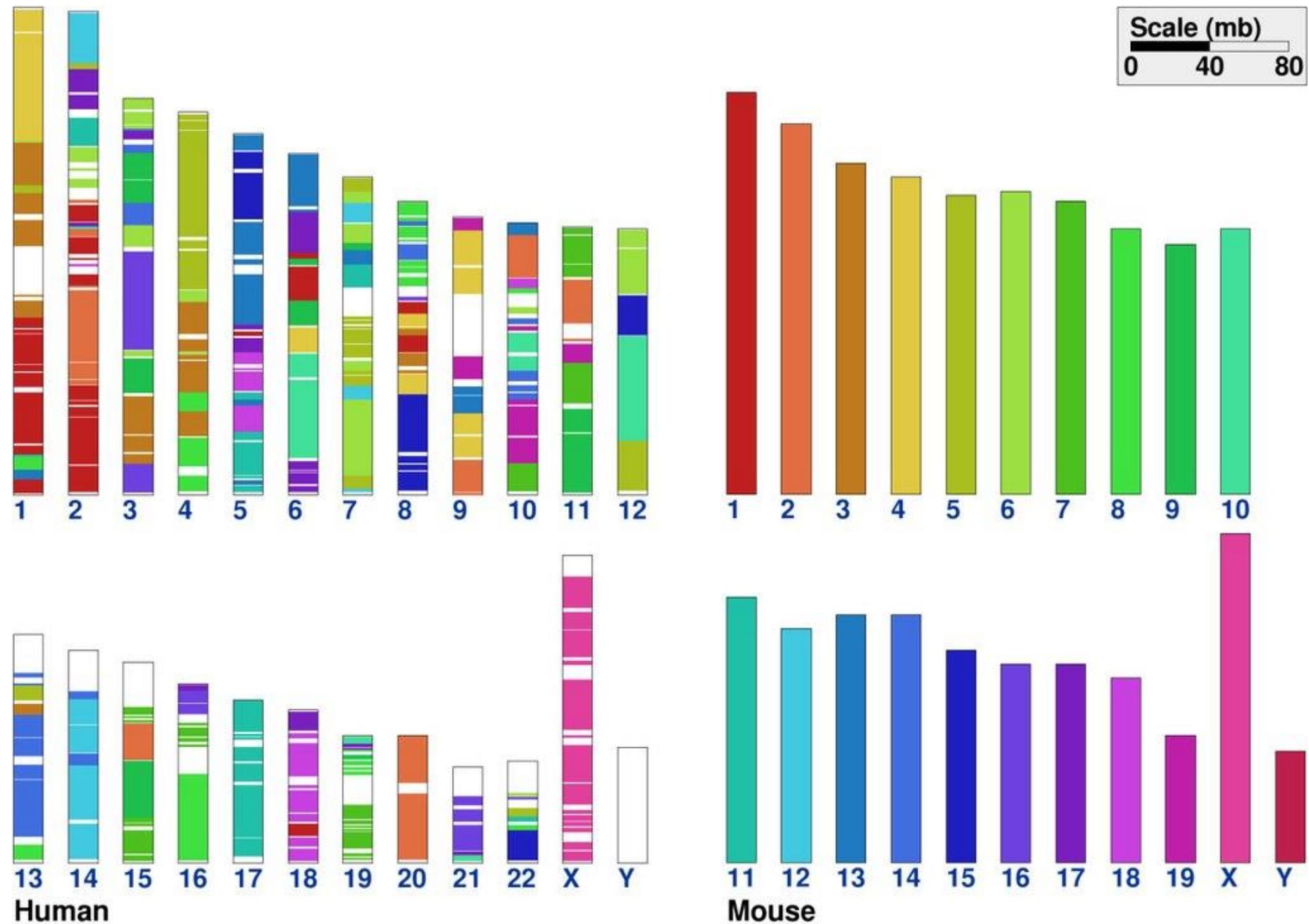
# Sintenija (eng. *synteny*): kolinearnost gena u genomima

- redoslijed gena na kromosomima srodnih vrsta je očuvan (u pravilu)

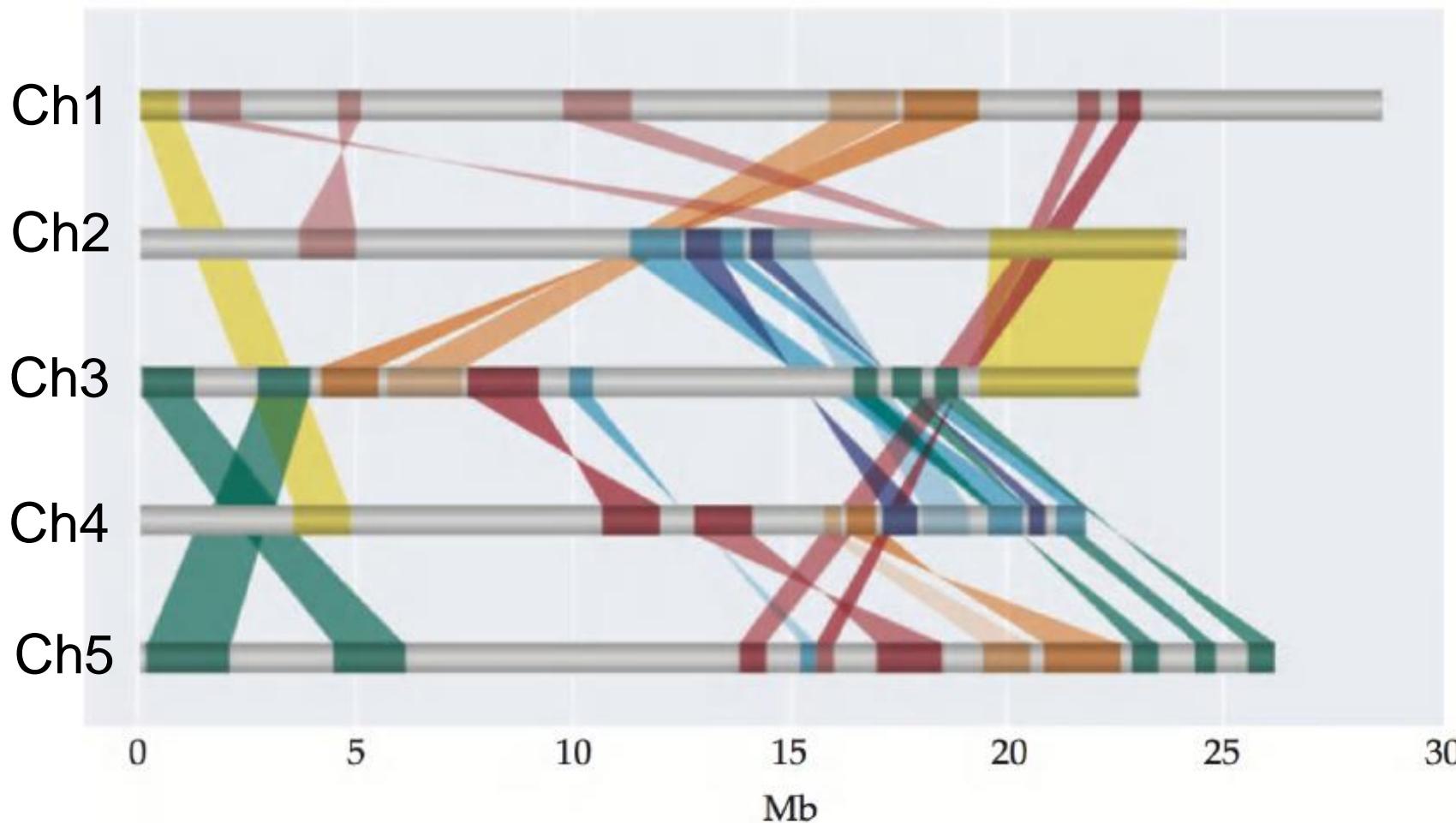


# Sintenija (eng. synteny):

- miš vs. čovjek



# Duplikacije u genomu otkrivene sekvenciranjem: primjer uročnjaka



**Figure 1.16 Chromosome duplications in the *Arabidopsis thaliana* genome.**

# Genomi svih modelnih organizama su sekvencirani:

**TABLE 1.3 Comparison of Gene Content in Some Representative Genomes**

	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>Drosophila<sup>a</sup></i>	<i>C. elegans</i>	<i>A. thaliana<sup>a</sup></i>	<i>H. sapiens<sup>a</sup></i>
Genome size <sup>a</sup> (Mb)	4.6	12.0	120+	97	115+	3,000+
Number of genes <sup>b</sup>	4,300	6,250	13,600	18,425	25,500	22,000
Average gene density (kb)	1.1	1.9	8.8	5.3	4.5	135
Number of gene families <sup>c</sup>	2,500	4,500	8,000	9,500	11,000	10,000



- sekvenciraju se genomi kultiviranih biljaka, domaćih životinja, izumrlih vrsta, koproliti (fossilizirani izmet), okolišna DNA...

 Search NCBI ...

Log in

Datasets

Taxonomy

Genome

Gene

Command-line tools

Documentation

BETA

# Genome

Download a genome data package including genome, transcript and protein sequence, annotation and a data report

Selected taxa

 Enter one or more taxonomic names

Please enter a taxonomic name or try one of these examples

**EUKARYOTES**

*Homo sapiens*  
*Mus musculus*  
*Arabidopsis thaliana*  
*Sus scrofa*

**BACTERIA**

*Escherichia coli*  
*Staphylococcus aureus*  
*Pseudomonas aeruginosa*  
*Mycobacterium tuberculosis*

**ARCHAEA**

*Pyrococcus furiosus*  
*Halobacterium salinarum*  
*Haloferax volcanii*  
*Thermoplasma acidophilum*

**VIRUSES**

*Human immunodeficiency virus 1*  
*Influenza A virus*  
*Hepatitis B virus*  
*Rotavirus A*

<https://www.ncbi.nlm.nih.gov/data-hub/genome/>

## *Next generation sequencing (NGS):*

- Pyrosequencing/454 (*Roche*)
  - Reversible dye-terminators technology (**Illumina**), sekvenciranje sintezom DNA
    - Sequencing by ligation/SOLiD technology (*Appl. Biosystems*)
    - Heliscope sequencing (*Helicos*)
    - Ion semiconductor sequencing (*Ion Torrent Systems*)
      - *Pacific Biosciences-PacBio (long read sequencing)*
      - *Oxford Nanopore (long read sequencing)*

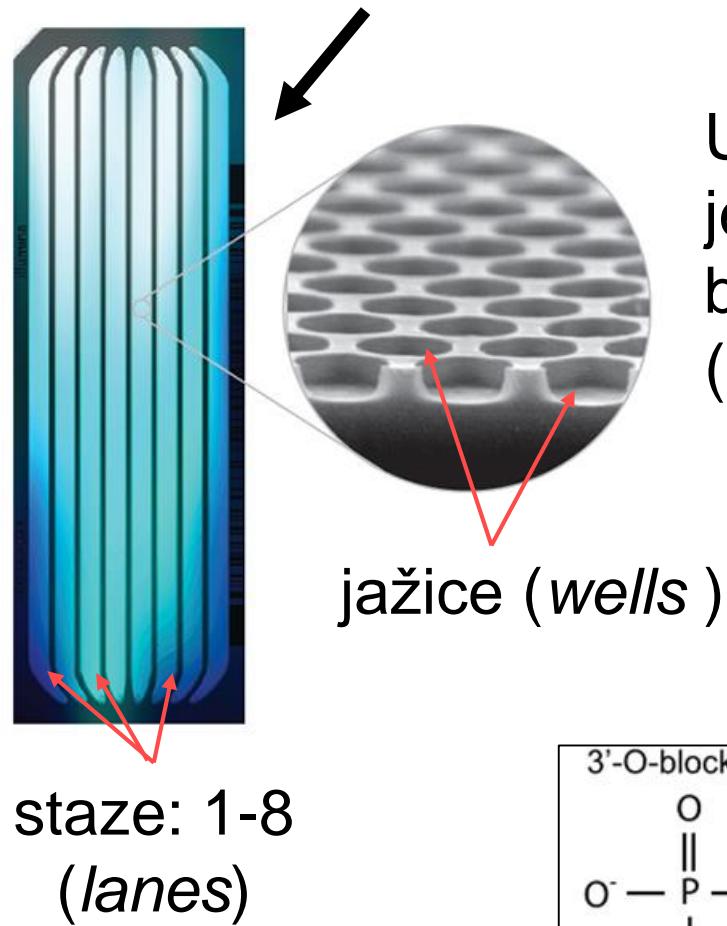
### Prednosti:

- manja cijena, visoka protočnost (*high-throughput*) → velika količina dobivenih podataka (tj. broj pročitanih nukleotida)

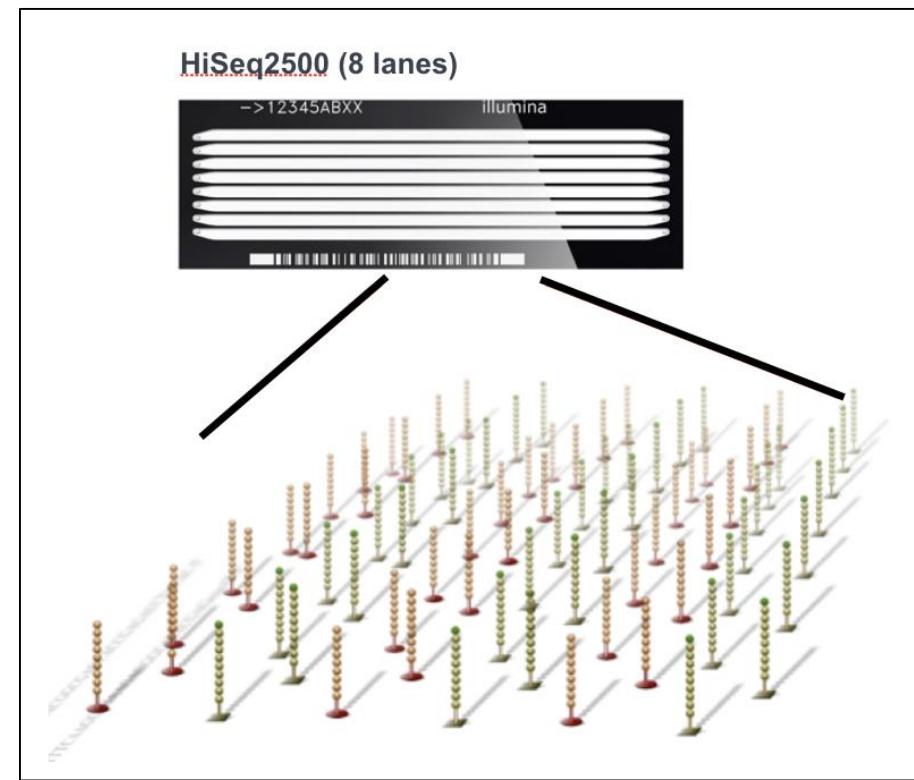
### Mane:

- čitaju se kraći sljedovi (50-150 nt); više grešaka u čitanju

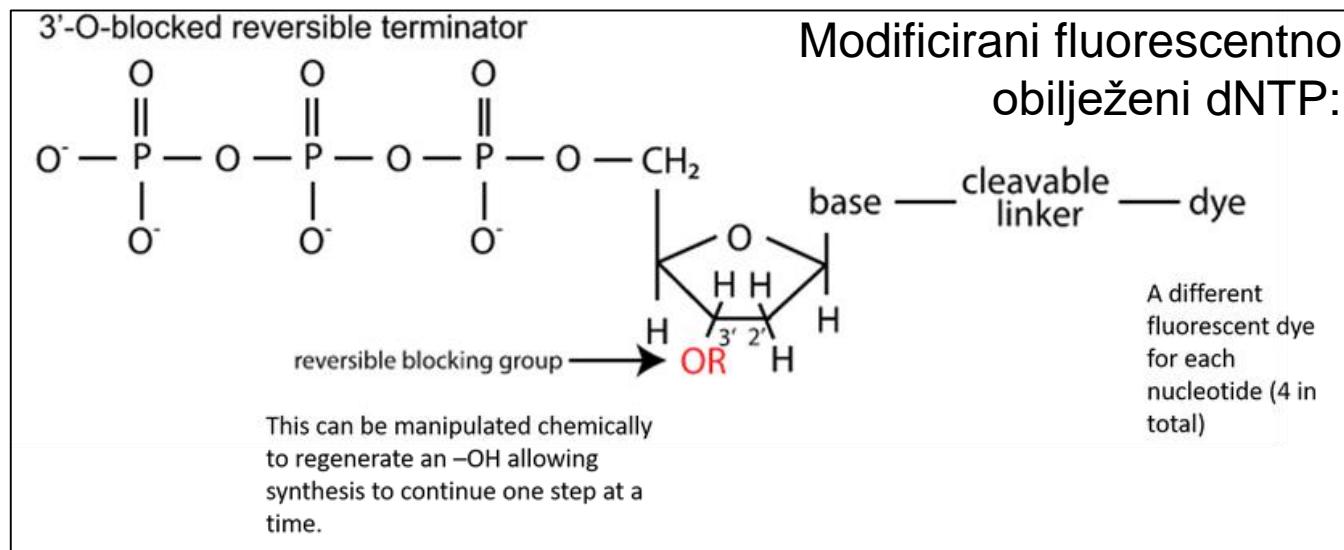
# Illumina flow-cell („pločica”):



U svakoj jažici je jedan fragment biblioteke (u prosjeku).



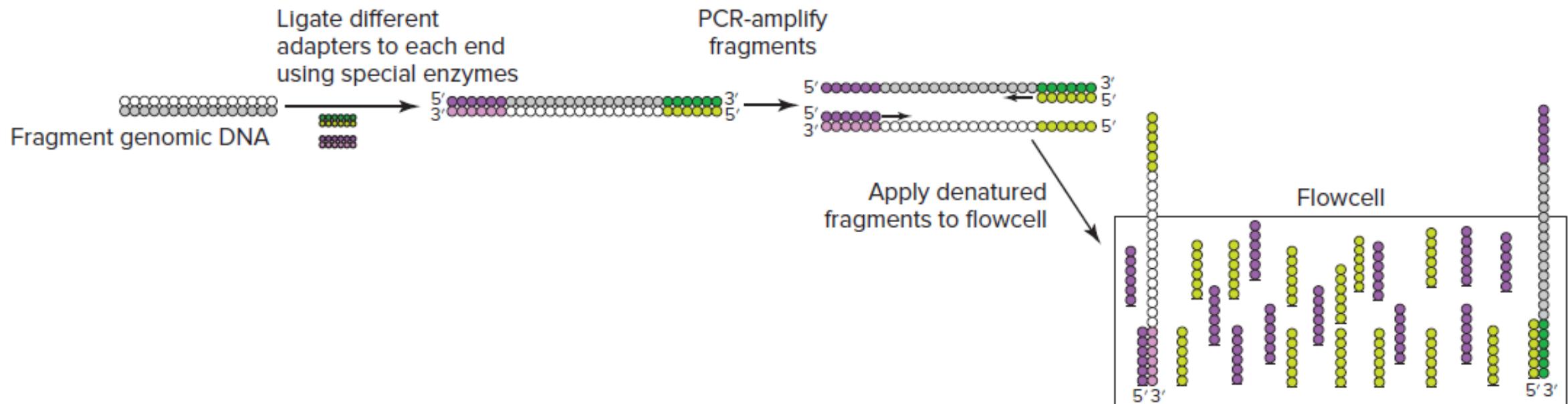
Specijalne DNA polimeraze!

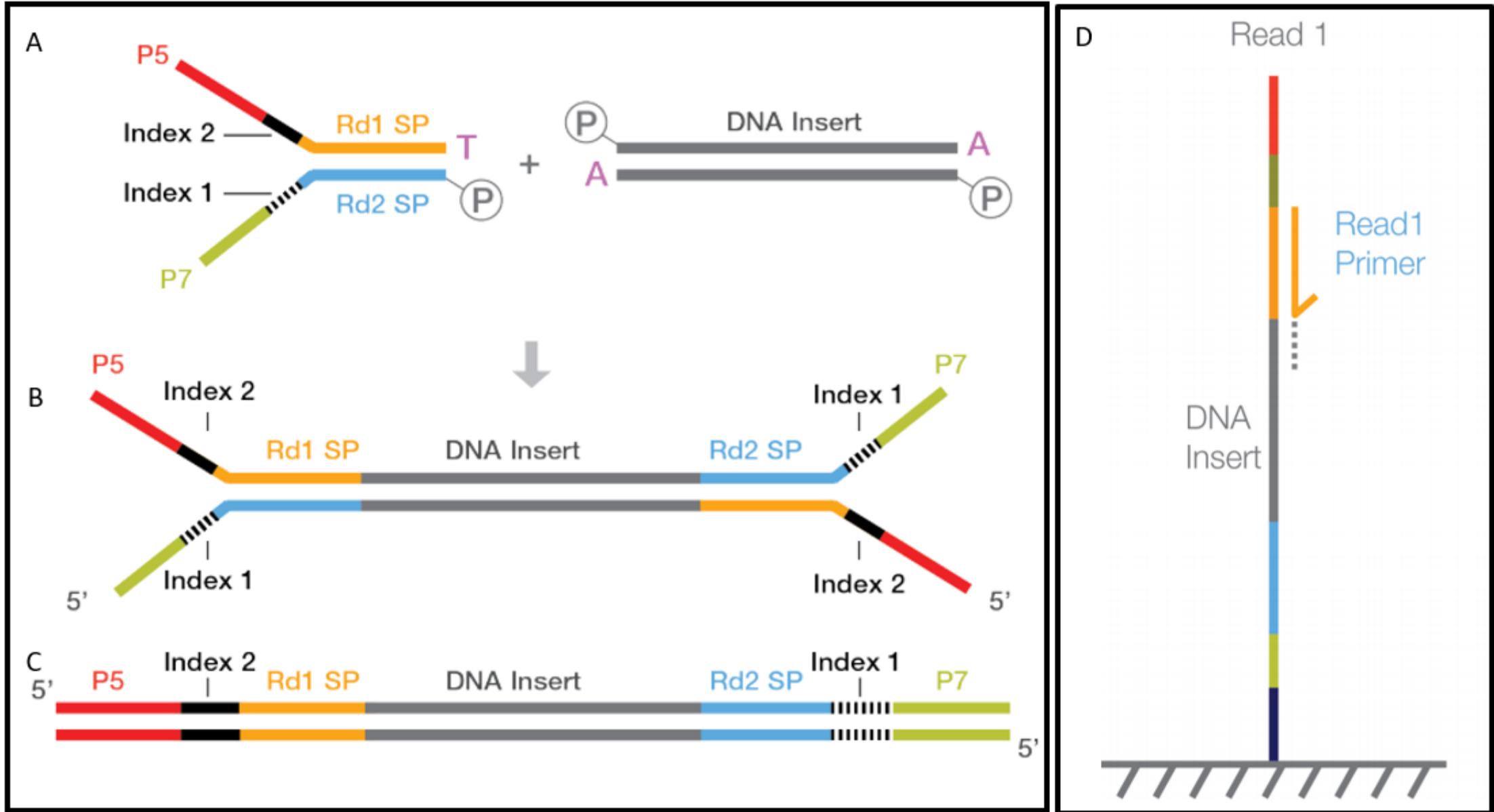


# Illumina: sekvenciranje DNA sintezom (SBS - sequencing by synthesis)

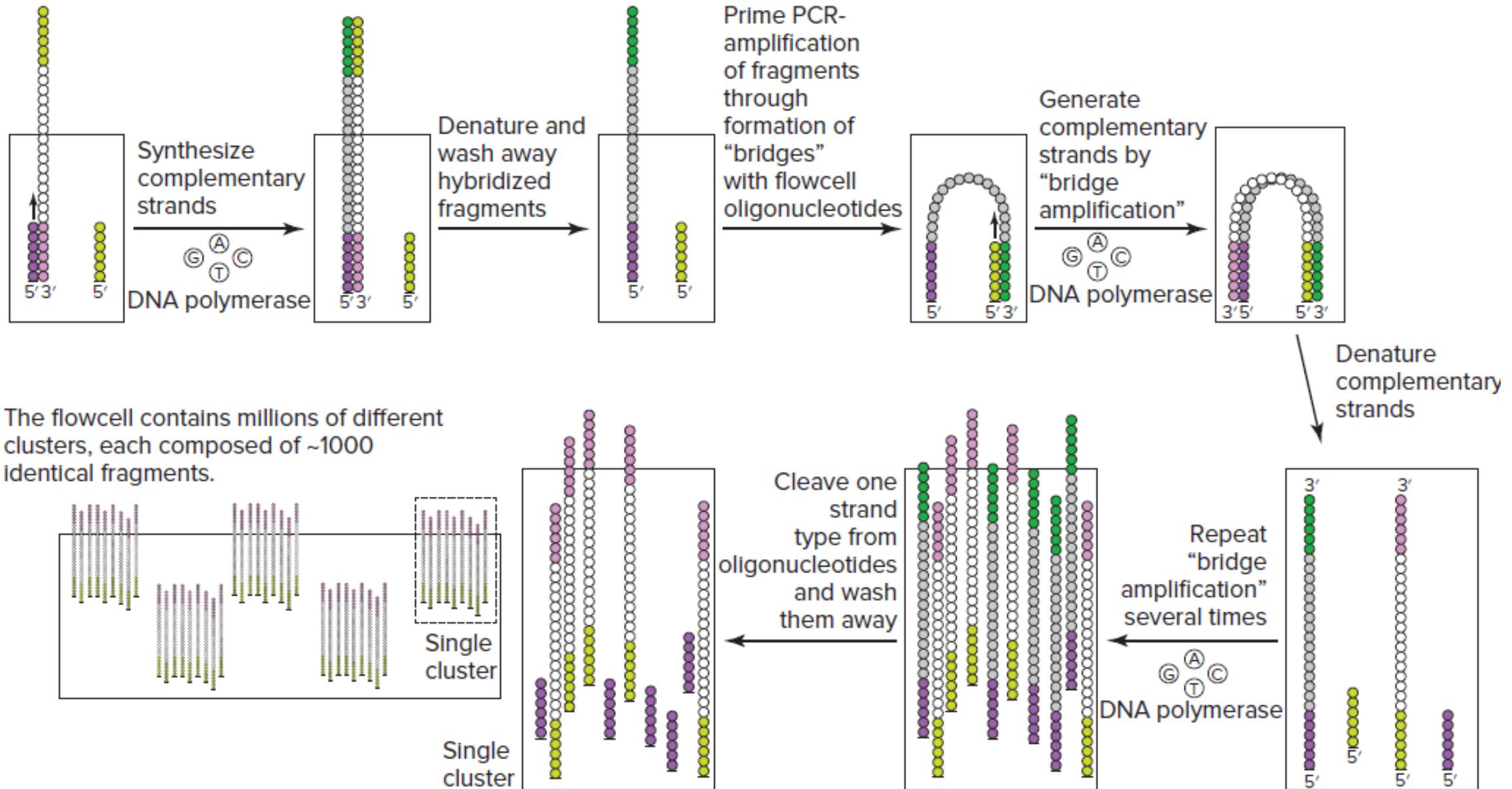
- gDNA se nakon izolacije fragmentira na manje fragmente (RE ili ulztrazvukom): 200-300 nt
- ova tehnologija zasad može pouzdano pročitati niz od 50-150nt

## (a) Sample preparation

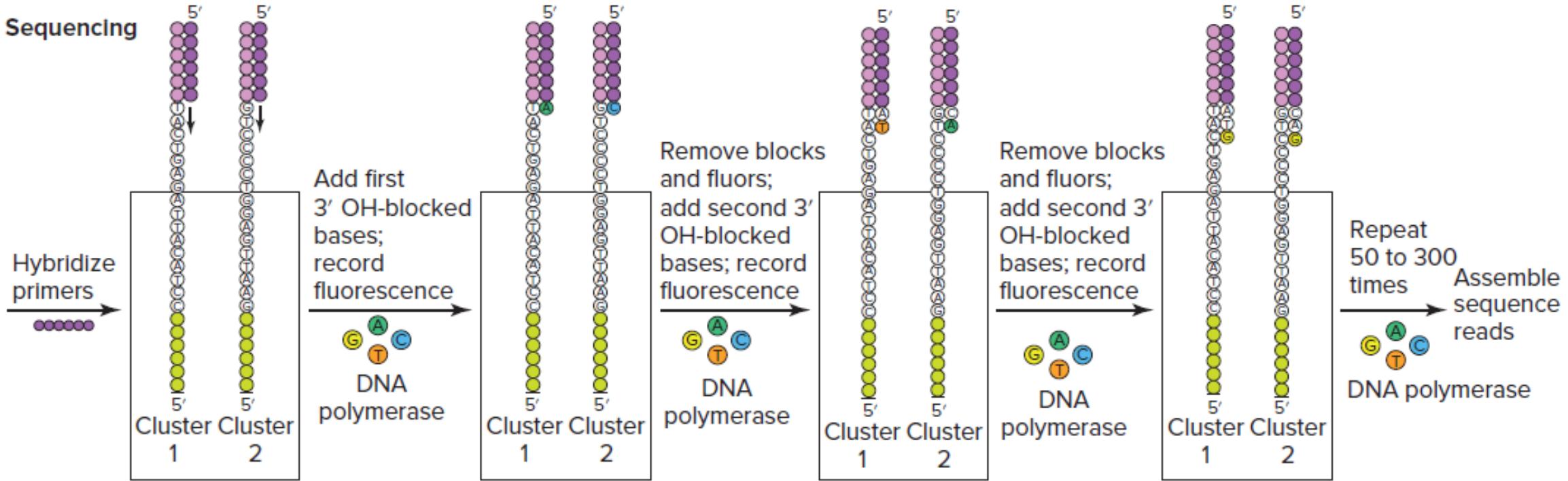




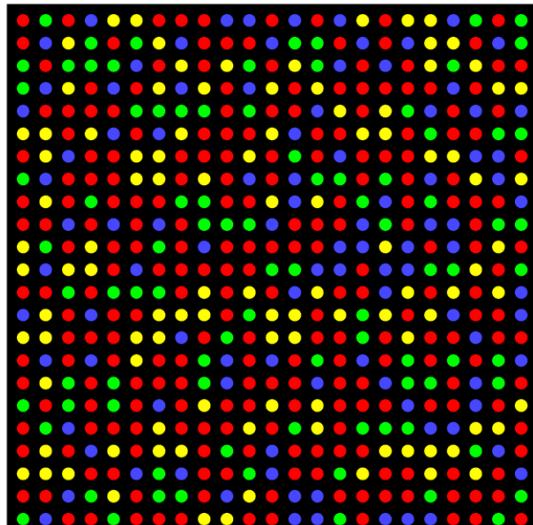
(b) Cluster generation



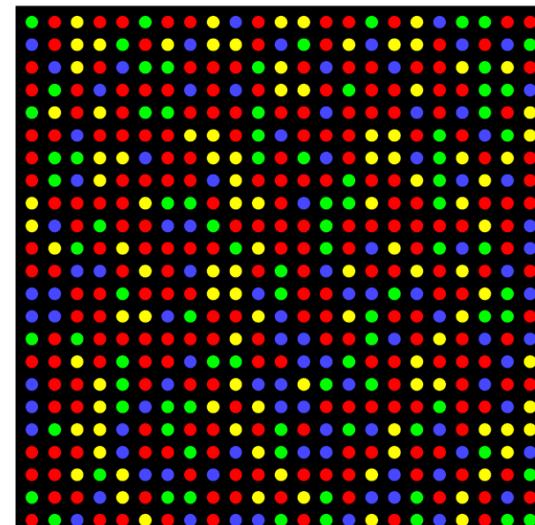
(c) Sequencing



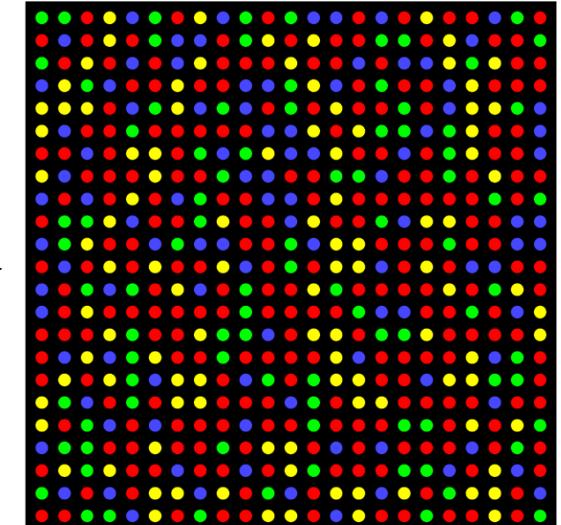
NAKON UGRADNJE 1. NUKLEOTIDA



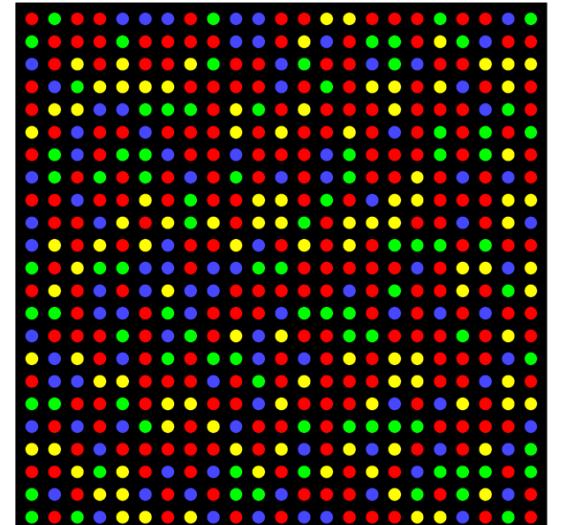
NAKON UGRADNJE 2. NUKLEOTIDA



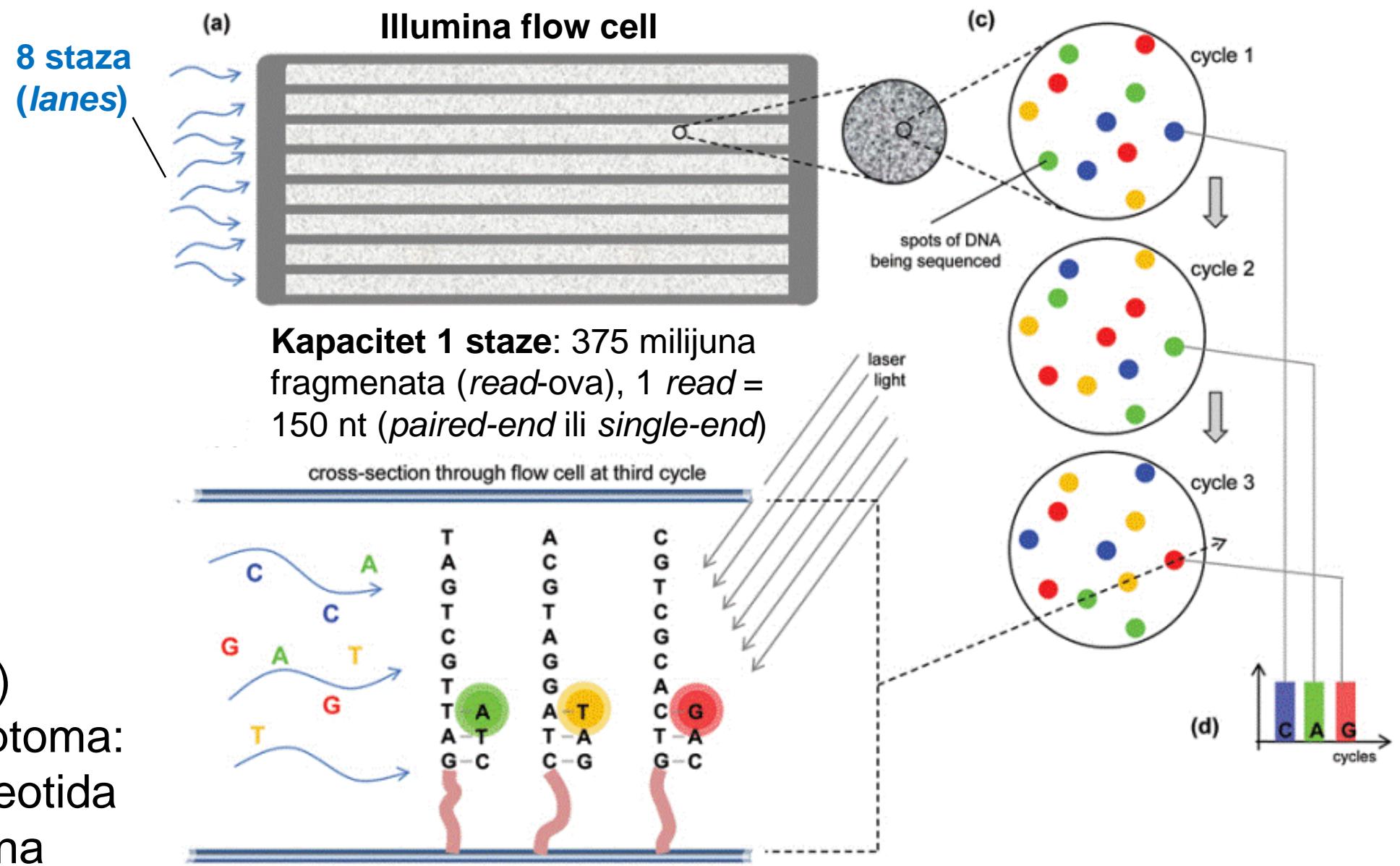
NAKON UGRADNJE 3. NUKLEOTIDA



NAKON UGRADNJE 4. NUKLEOTIDA



# DNA sekvenciranje:



Obuhvat (coverage)  
pročitanog transkriptoma:  
broj pročitanih nukleotida  
/ procijenjena veličina  
transkriptoma (10-100 x)

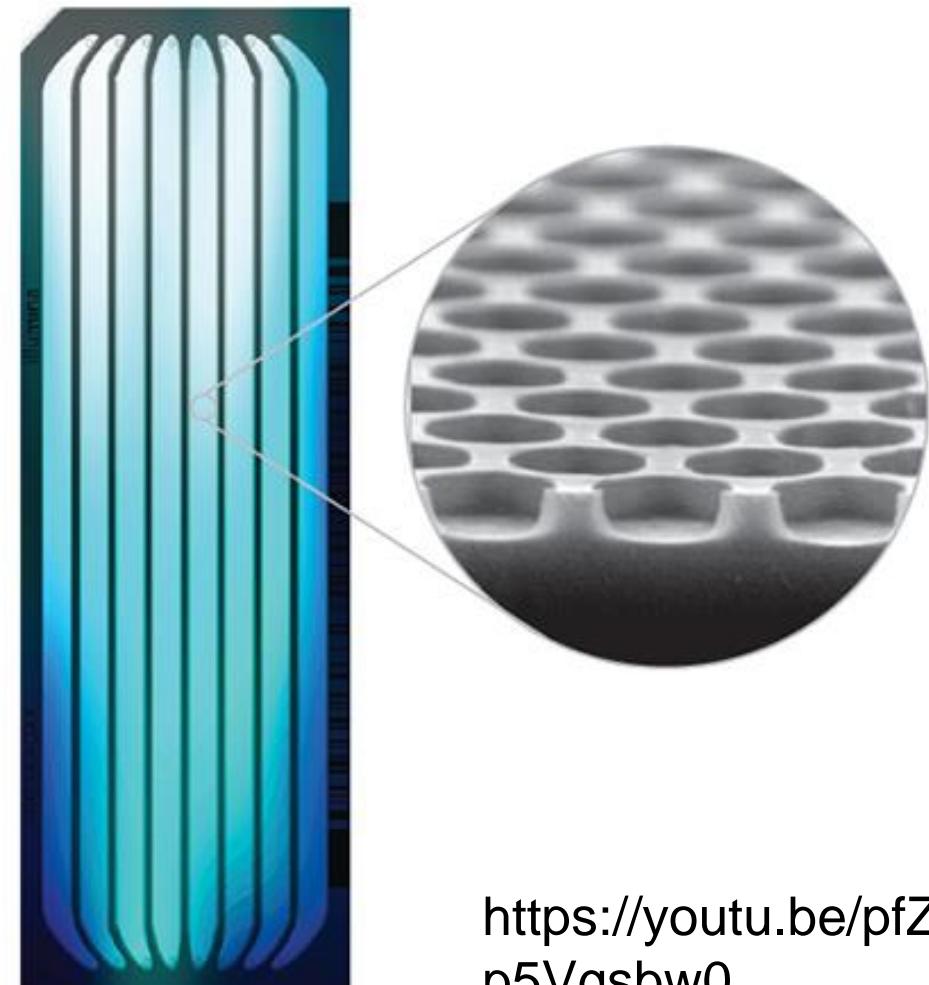
# Illumina flow-cell: količina dobivenih podataka

- $\sim 400 \times 10^6$  jažica po stazi *flow-cell-a*
- $x 8$  staza =  $3.2 \times 10^9$  fragmenata biblioteke (*reads*)
- po 1 fragmentu, može se pročitati 150 nukleotida u dva smjera
- ukupno:  $960 \times 10^9$  nukleotida!

eng. *Coverage* (pokrivenost/obuhvat):

$960 \times 10^9 / 3 \times 10^9$  (ljudski genom)

= 320x („pokriven” genom)



<https://youtu.be/pfZp5Vgsbw0>

# Illumina Sequencing Technology

						
Key Methods	Amplicon, targeted RNA, small RNA, and targeted gene panel sequencing.	Small genome, amplicon, and targeted gene panel sequencing.	Everyday exome, transcriptome, and targeted resequencing.	Production-scale genome, exome, transcriptome sequencing, and more.	Population- and production-scale whole-genome sequencing.	Same as HiSeq
Maximum Output	7.5 Gb	15 Gb	120 Gb	1500 Gb	1800 Gb	1 - 6 Tb
Maximum Reads per Run	25 million	25 million <sup>†</sup>	400 million	5 billion	6 billion	6.6 billion
Maximum Read Length	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp
Run Time	4–24 hours	4–55 hours	12–30 hours	<1–3.5 days (HiSeq 3000/HiSeq 4000) 7 hours–6 days (HiSeq 2500)	<3 days	19 - 40 hrs
Benchtop Sequencer	Yes	Yes	Yes	No	No	no

Nedostaci: mogu se pročitati samo vrlo kratki fragmenti u kontinuitetu, učestalost grešaka je veća nego kod Sanger (dideoksi) sekvenciranja

# *Next generation sequencing (NGS):*

*Illumina: sequencing by synthesis*

<https://www.youtube.com/watch?v=jFCD8Q6qSTM>

<https://www.youtube.com/watch?v=womKfikWIxM>

[https://www.youtube.com/watch?annotation\\_id=annotation\\_228575861&feature=iv&src\\_vid=womKfikWIxM&v=fCd6B5HRaZ8](https://www.youtube.com/watch?annotation_id=annotation_228575861&feature=iv&src_vid=womKfikWIxM&v=fCd6B5HRaZ8)

# Next generation sequencing:



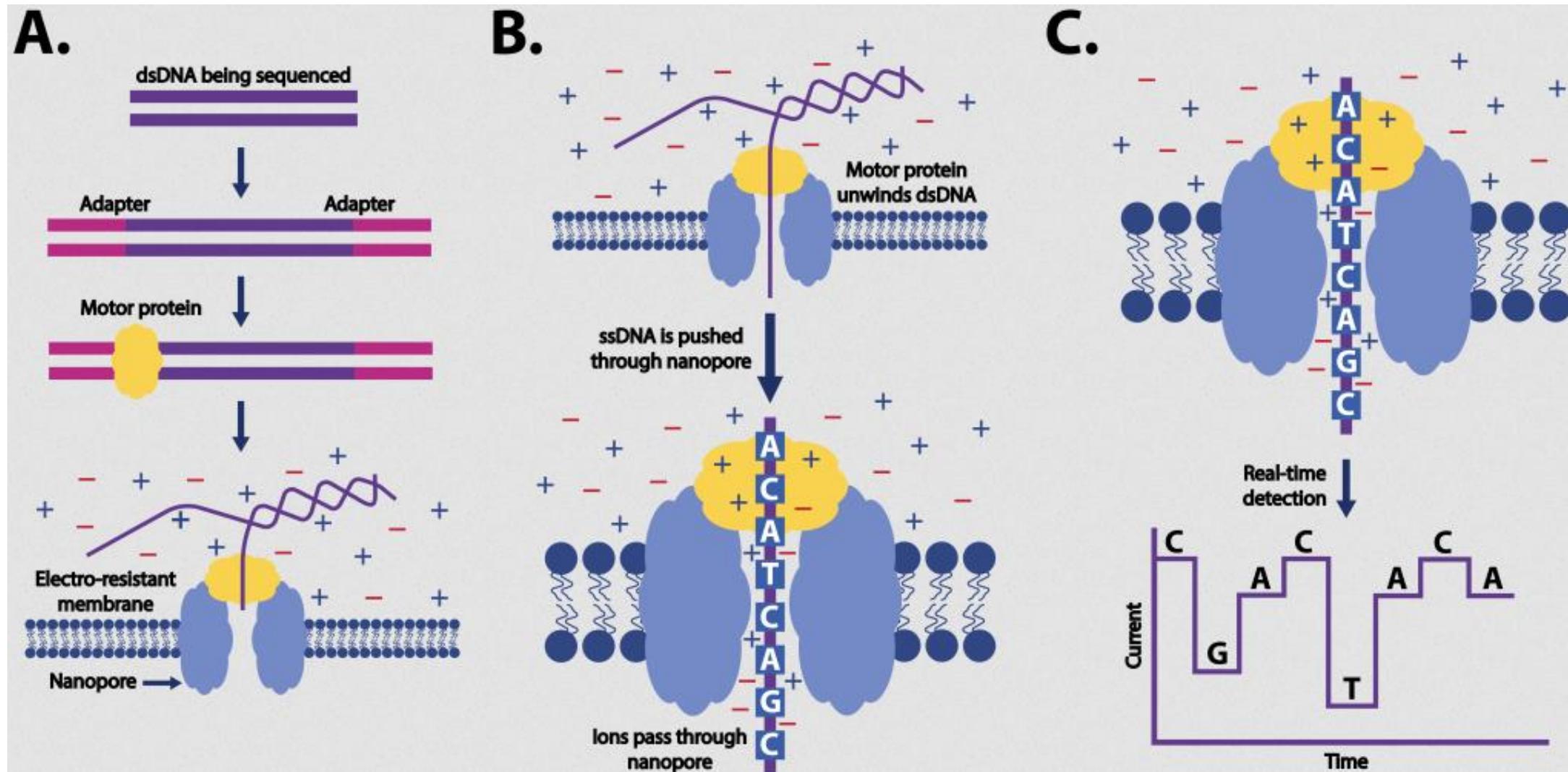
Nove strategije sekvenciranja: *Oxford nanopore*

- cilj: dobiti što je moguće dulje pročitane sljedove
- sekvenciranje kroz proteinsku poru
- promjena u električnom signalu pri prolasku DNA kroz poru prevodi se u slijed nukleotida (nema PCR-a/sinteze DNA, obilježavanja primer-a ili ddNTP-ova)



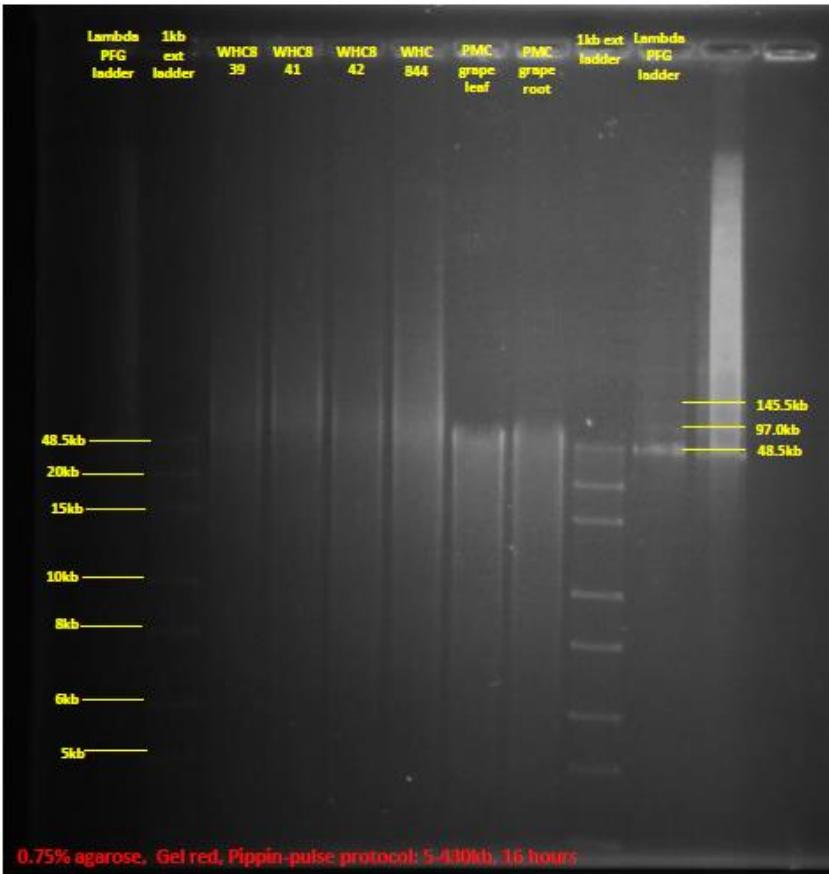
Direktno čitanje RNA i proteina.





- ovom tehnologijom moguće je rutinski pročitati 10-100 kb u nizu
- uspjelo se pročitati i kontinuirani fragment od 4 Mb

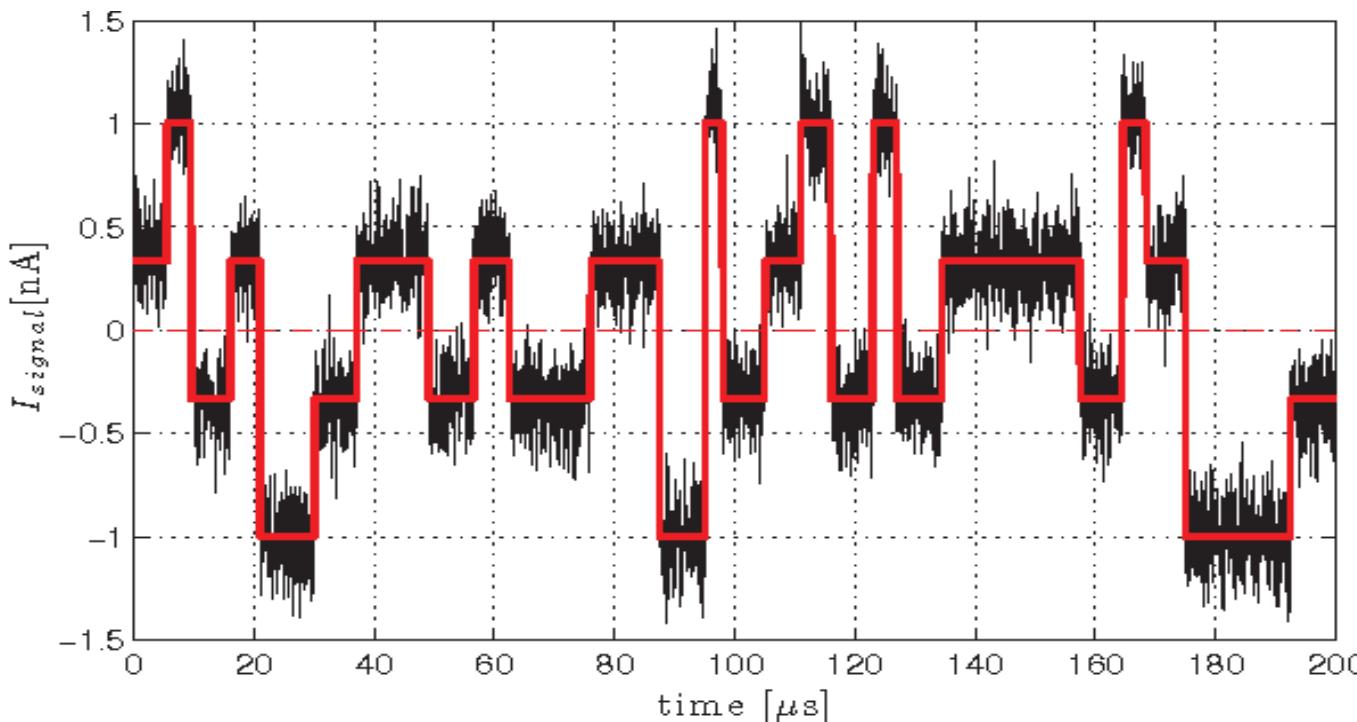
# Primjer: uzorak vinove loze poslan na ON sekvenciranje



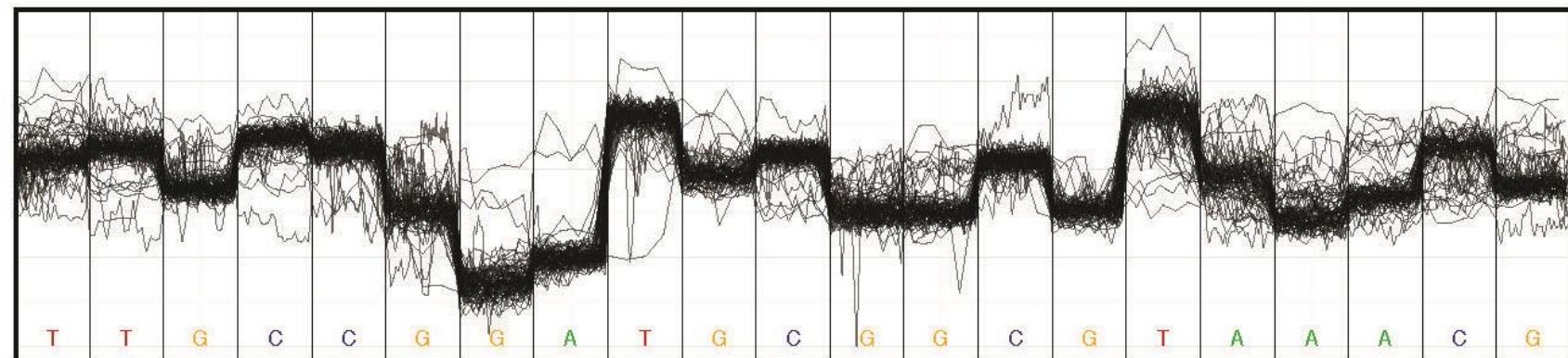
Criteria	ON002-DNA-R00392	
	All reads:	Reads with Q $\geq$ 7:
total gigabases:	73.197	67.851
total reads:	5,444,954	4,877,259
N50 length:	19,734	19,902
mean length:	13,443	13,912
median length:	12,240	12,782
max length:	903,223	516,703
mean q:	12.1	13
median q:	12.8	13.2
reads:		
'>10kb':	3,148,572	2,925,452
'>20kb':	1,301,518	1,222,417
'>50kb':	15,656	14,645
'>100kb':	188	61
'>200kb':	49	10
'>500kb':	10	1

# Nanopore sekvenciranje: sekvenciranje bez sinteze DNA

- jakost struje → slijed nukleotida (indirektna identifikacija nukleotida)



- učestalost grešaka je veća nego kod Illumine
- omogućuje slaganje haplotipova
- lakše sekvenciranje repetitivnih regija genoma

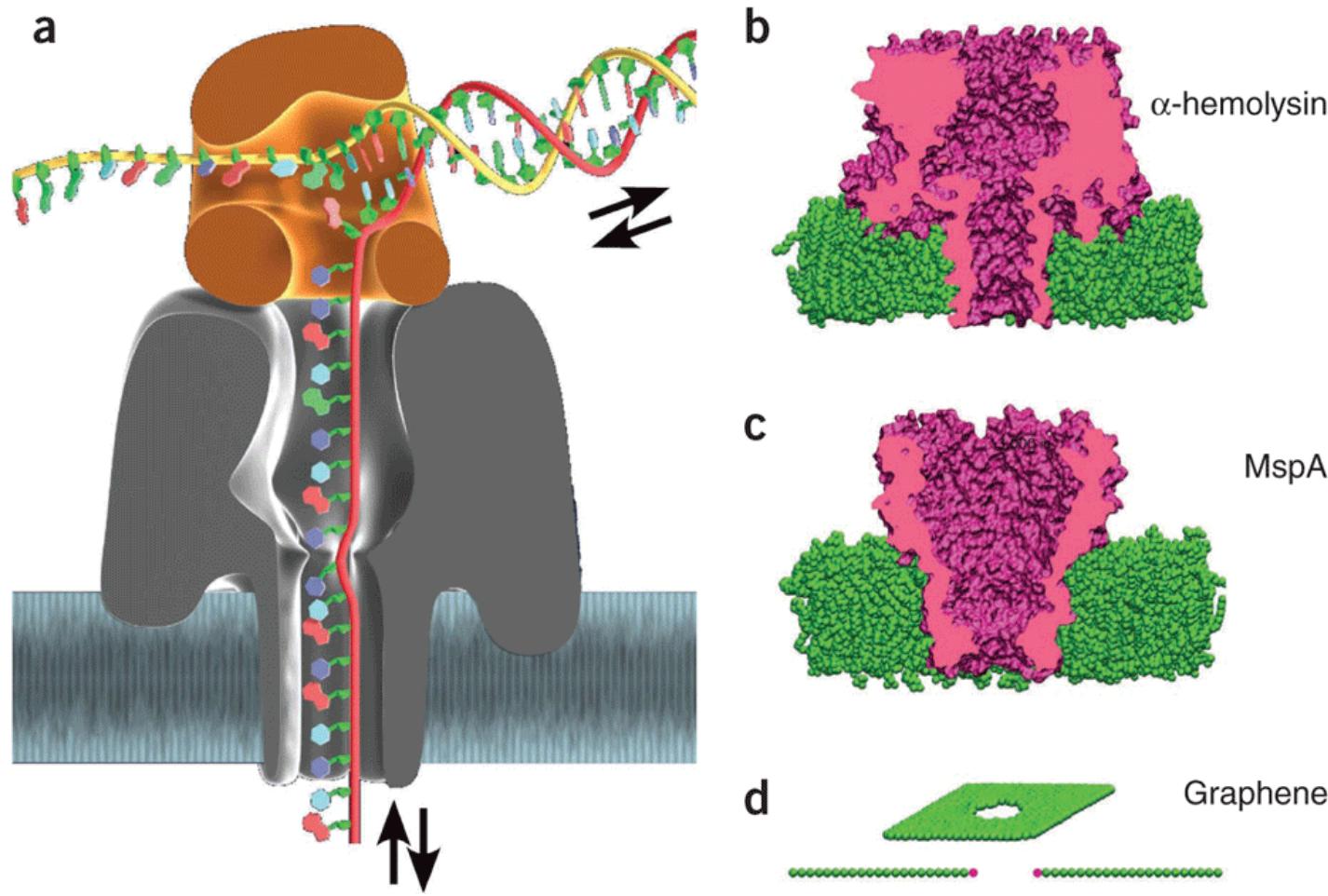


# Sekvenciranje na terenu:



Ricardo Funari/Zibra project

Next generation sequencing (NGS):



Hvala na pažnji!