

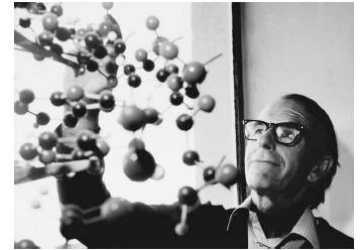
Uvod u sekvenciranje sljedeće generacije

(engl. *Next Generation Sequencing*, NGS)

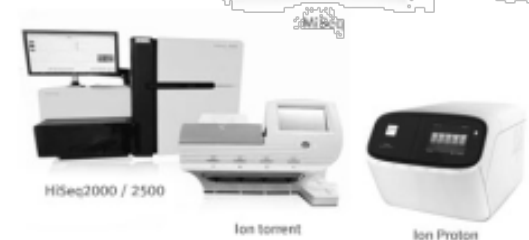
Genetska osnova količine bioaktivnih hranjivih tvari hrvatskih tradicijskih kultivara graha (BeanQual)



dr. sc. Ana Barešić,
poslijedoktorandica



- kratki pregled sekvenciranja
- NGS tehnologija
- pregled dostupnih uređaja
- primjena NGS-a
- primjer: nabava uređaja



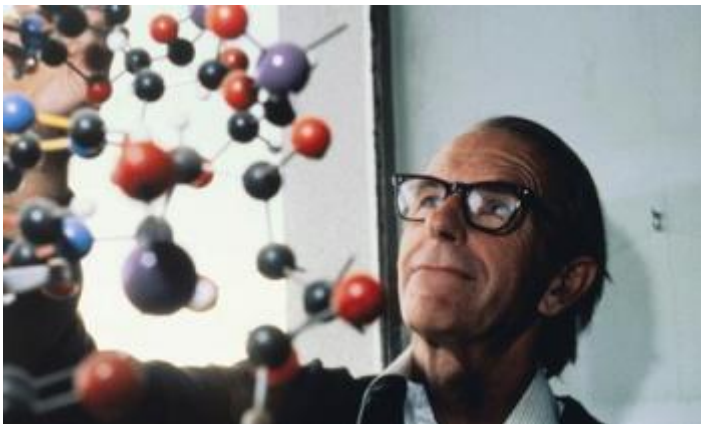
- što nas čini jedinstvenima?



F. Miescher, 1869. g.
“nuklein”



1953. g.



DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage ϕ X174)

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

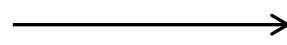
Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Contributed by F. Sanger, October 3, 1977

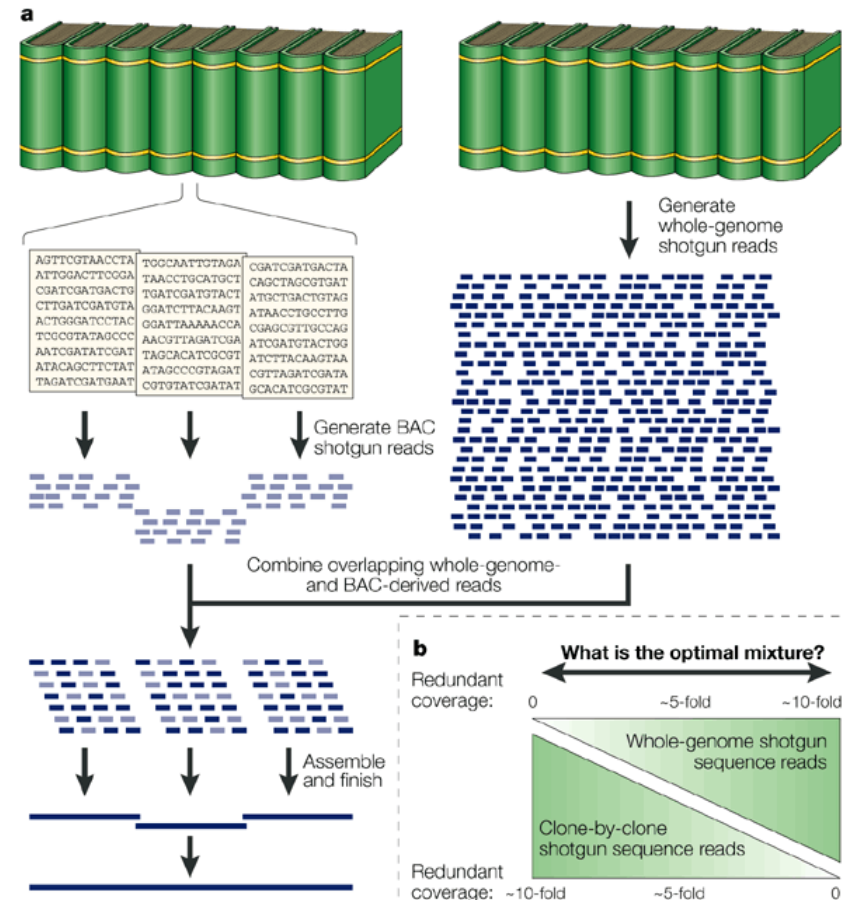
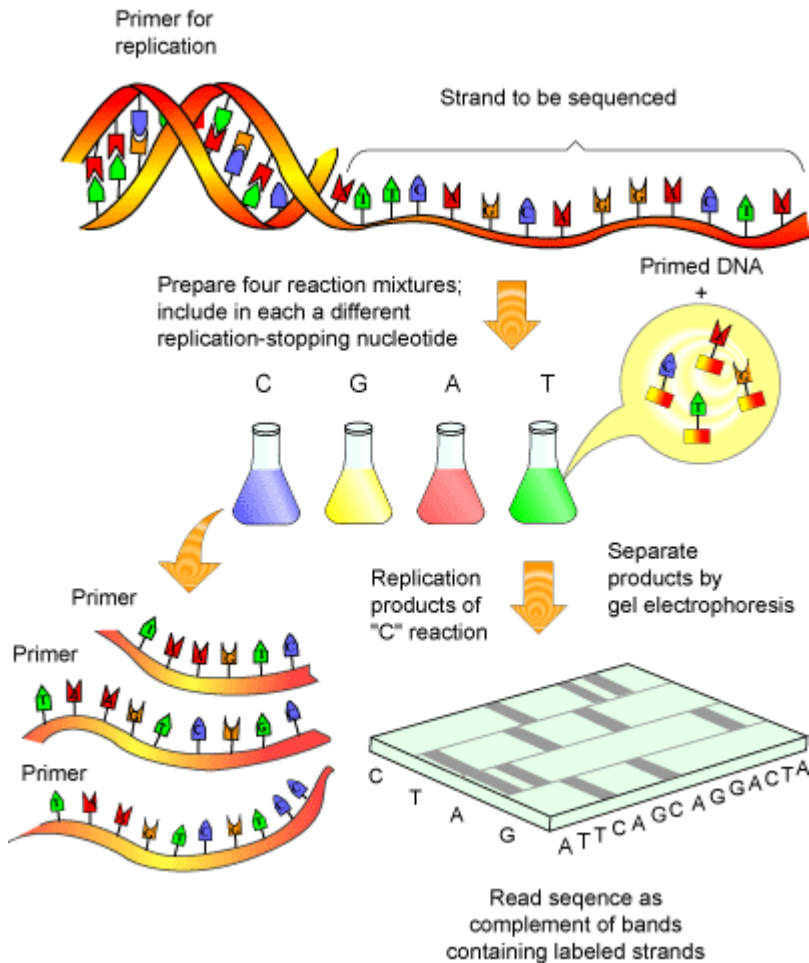
ABSTRACT A new method for determining nucleotide sequences in DNA is described. It is similar to the “plus and minus” method [Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* 94, 441–448] but makes use of the 2',3'-dideoxy and arabinonucleoside analogues of the normal deoxynucleoside triphosphates, which act as specific chain-terminating inhibitors of DNA polymerase. The technique has been applied to the DNA of bacteriophage ϕ X174 and is more rapid and more accurate than either the plus or the minus method.



Sangerova metoda (1977.)



Hierarchical nad shotgun sequencing (1996.)



Nature Reviews | Genetics

100 milijuna \$ 2001. godine na 10,000 \$ u 2011.



- 46 kromosoma – 3 milijarde parova baza
- 2 ljudi razlikuje se međusobno na 4 milijuna mjesta u genomu
- promjena na samo jednom mjestu! može dovesti do bolesti

Nucleotide sequence of bacteriophage Φ X174 DNA

F. Sanger, G. M. Air*, B. G. Barrell, N. L. Brown[†], A. R. Coulson, J. C. Fiddes, C. A. Hutchison III[‡], P. M. Slocombe[§] & M. Smith^{*}

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

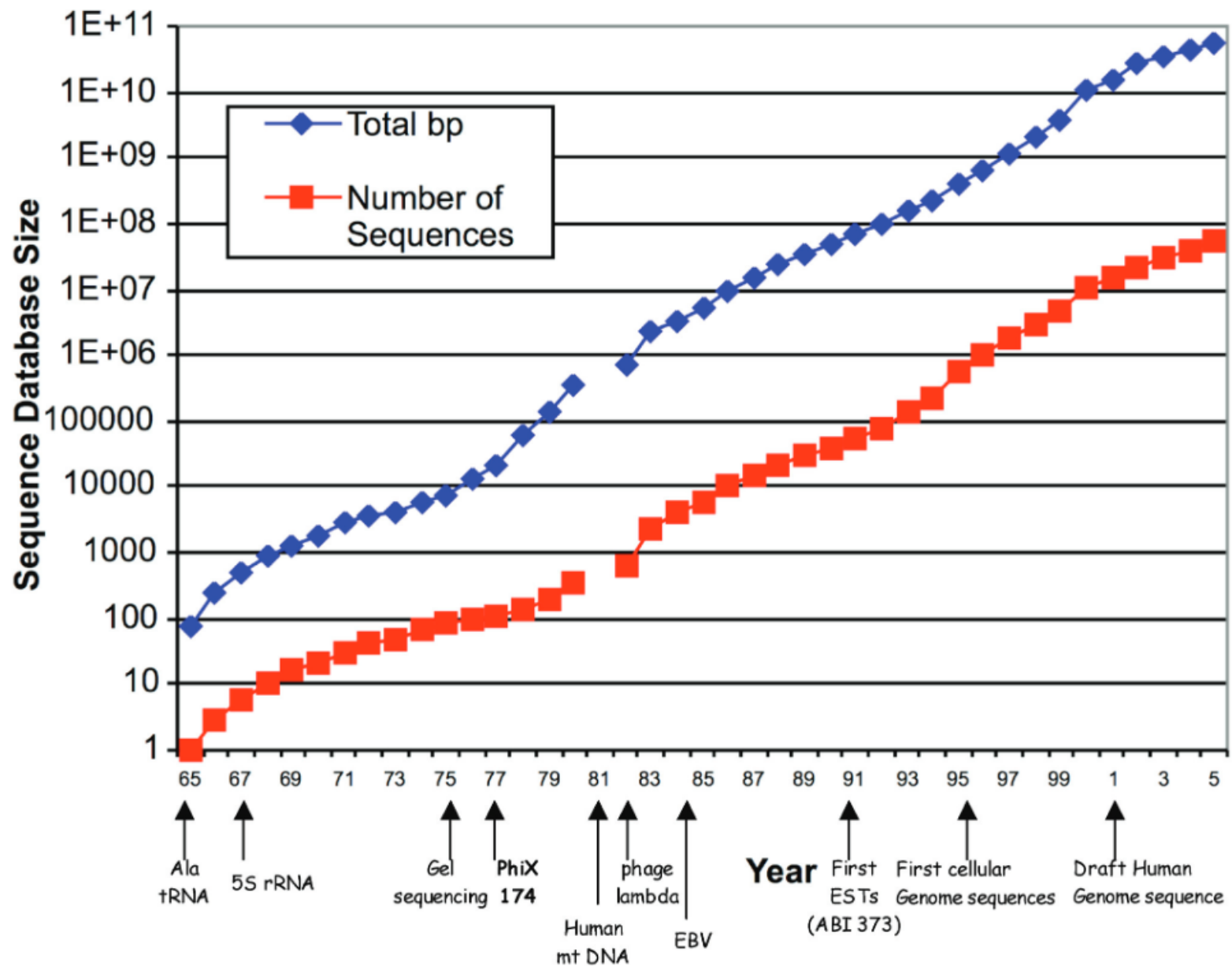
A DNA sequence for the genome of bacteriophage Φ X174 of approximately 5,375 nucleotides has been determined using the rapid and simple 'plus and minus' method. The sequence identifies many of the features responsible for the production of the proteins of the nine known genes of the organism, including initiation and termination sites for the proteins and RNAs. Two pairs of genes are coded by the same region of DNA using different reading frames.

strand DNA of Φ X has the same sequence as the mRNA and, in certain conditions, will bind ribosomes so that a protected fragment can be isolated and sequenced. Only one major site was found. By comparison with the amino acid sequence data it was found that this ribosome binding site sequence coded for the initiation of the gene G protein¹⁵ (positions 2,362–2,413).

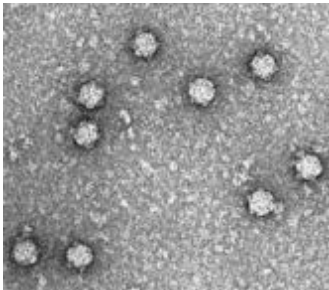
At this stage sequencing techniques using primed synthesis with DNA polymerase were being developed¹⁶ and Schott¹⁷ synthesised a decanucleotide with a sequence complementary to part of the ribosome binding site. This was used to prime into

Nature, 1977.

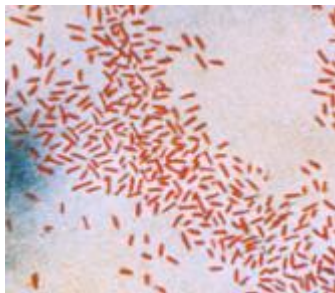
- iste godine (1977.) – metoda sekvenciranja po Sangeru ("*chain termination technology*")



Hutchison, *Nucleic Acids Res.* 2007 .



ΦX174



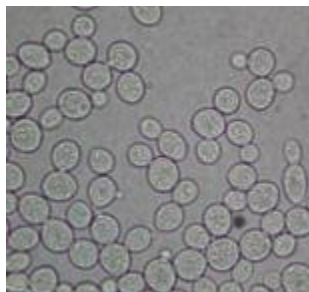
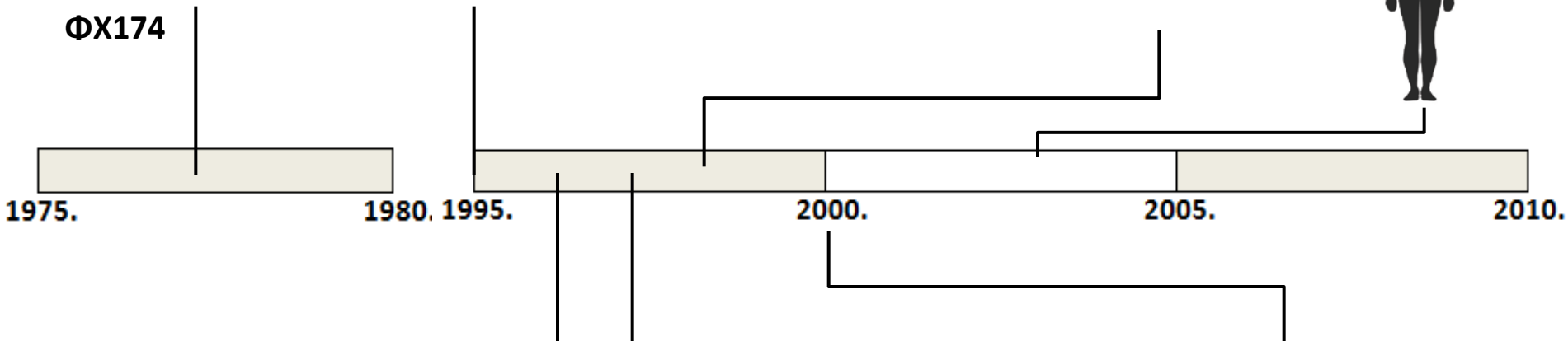
Haemophilus influenzae



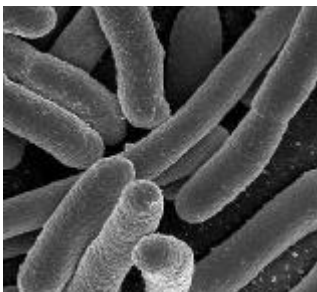
Mycoplasma genitalium



Caenorhabditis elegans



Saccharomyces cerevisiae



Escherichia coli



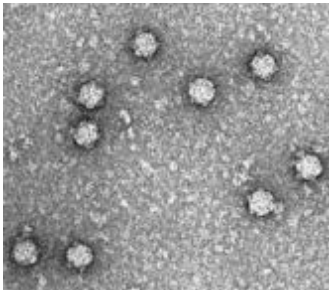
Arabidopsis thaliana



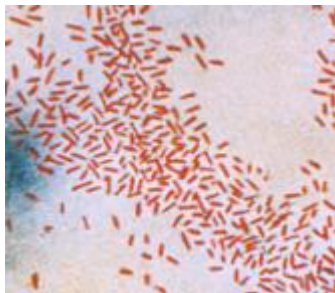
Drosophila melanogaster

- *Human Genome Project*
- 15 godina, 5-10 milijardi dolara
(2003., 3 milijarde dolara)





ΦX174



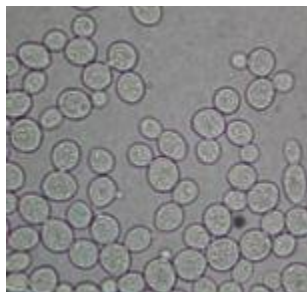
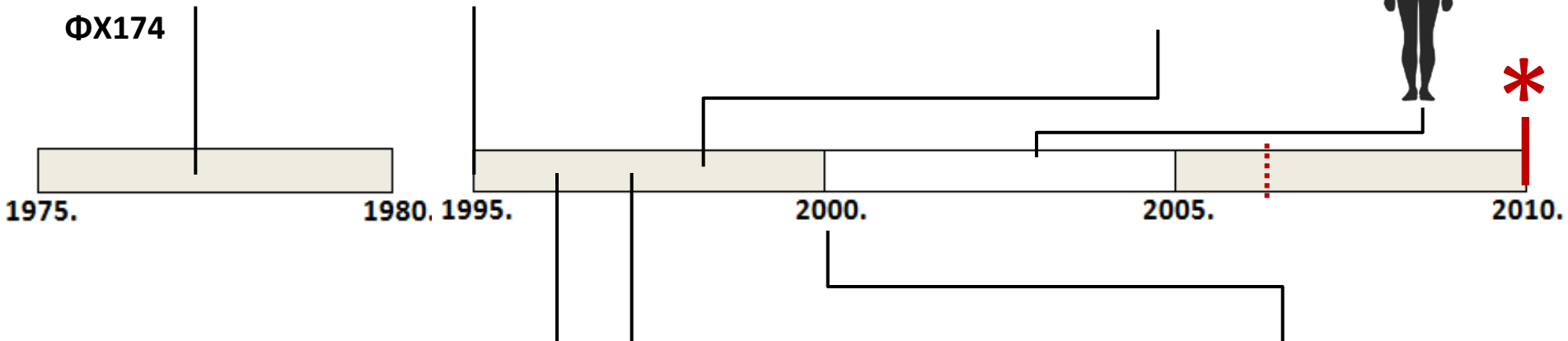
Haemophilus influenzae



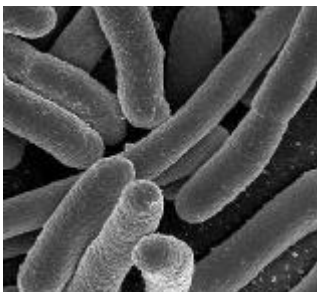
Mycoplasma genitalium



Caenorhabditis elegans



Saccharomyces cerevisiae



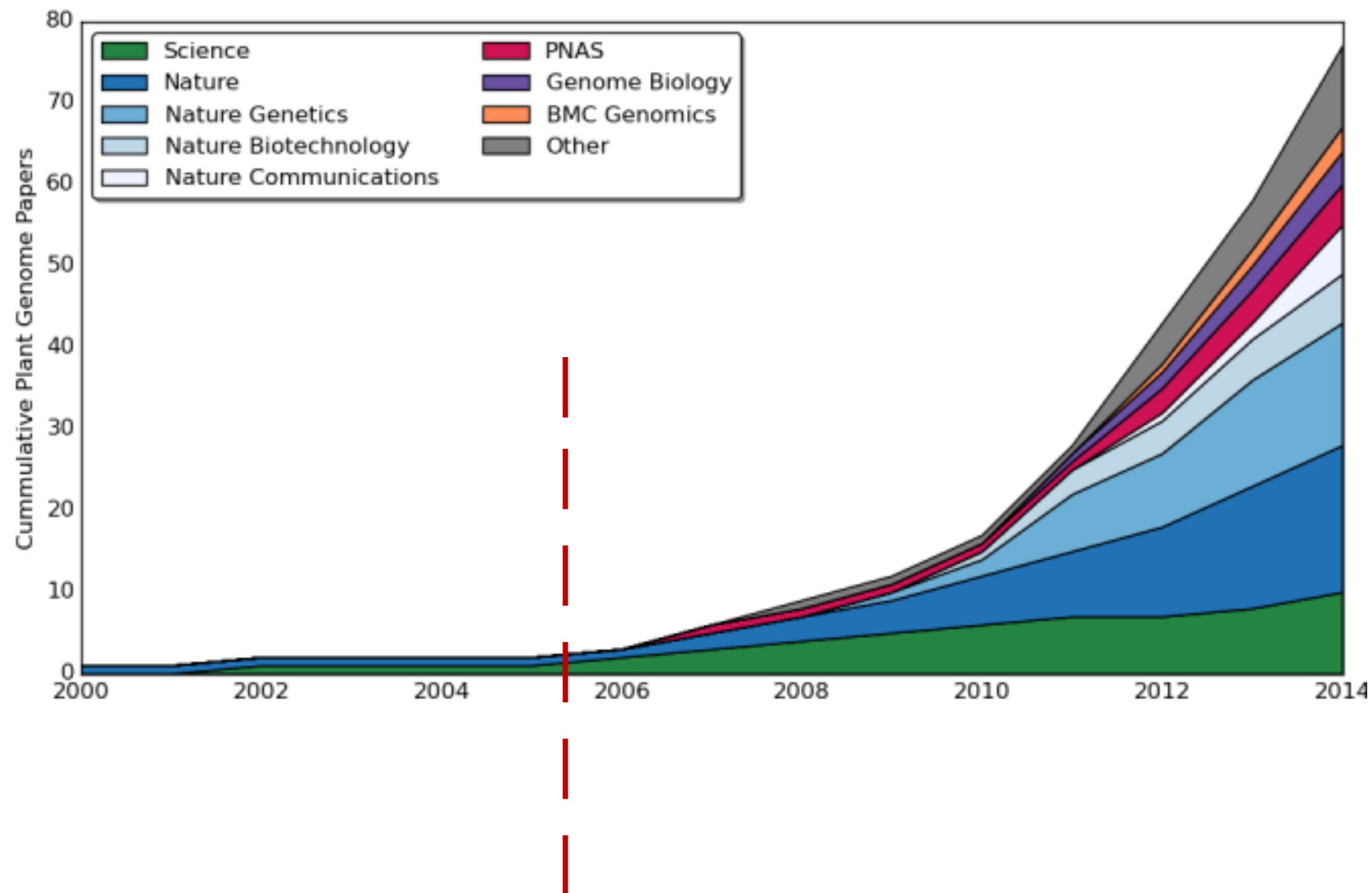
Escherichia coli



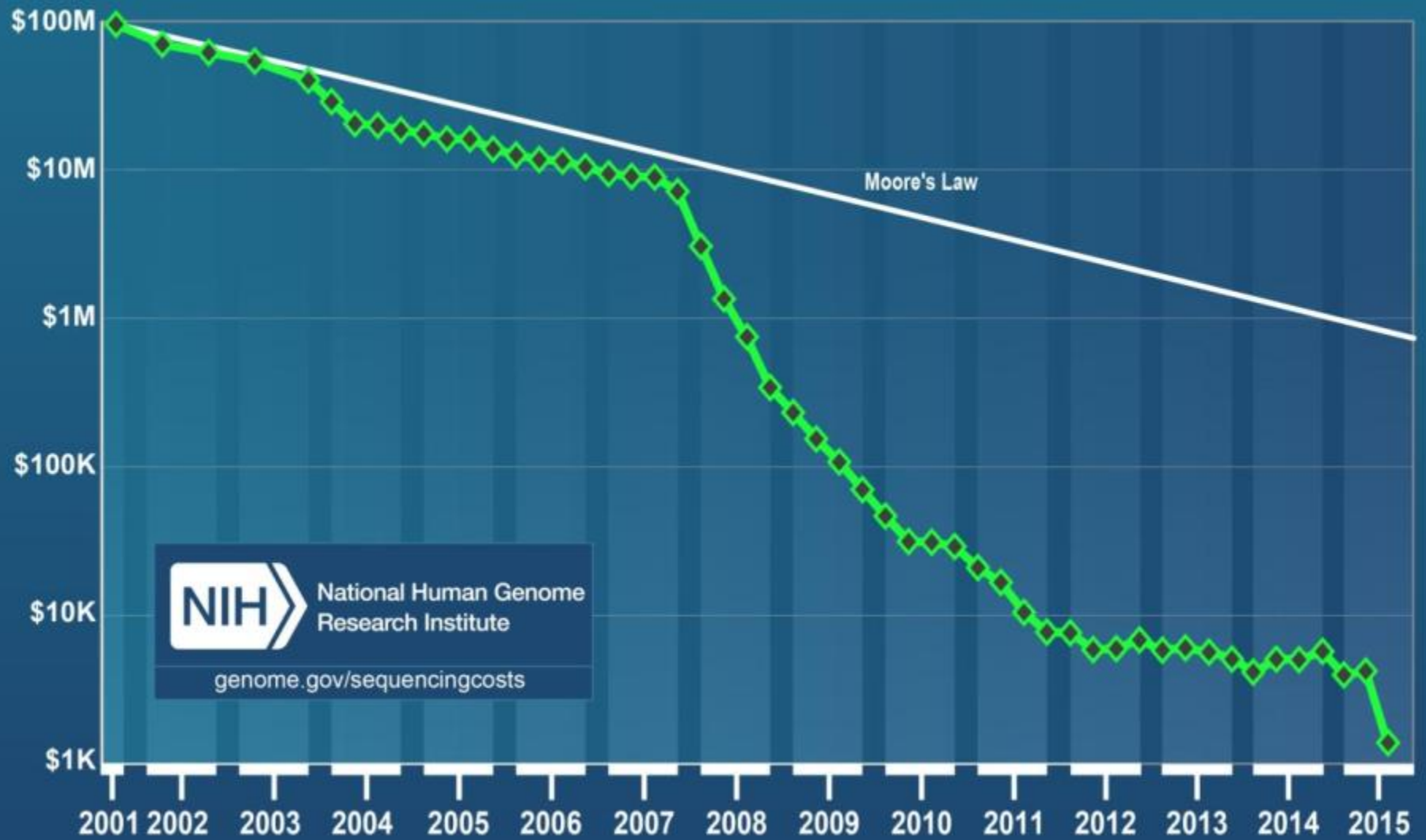
Arabidopsis thaliana



Drosophila melanogaster



Cost per Genome



Cost per Genome



Sekvenciranje sljedeće generacije

- skup tehnologija koje mogu čitati DNA puno brže i jeftinije od tradicionalnog Sangerovog sekvenciranja – čitanje puno odsječaka odjednom!

	Sanger	NGS (454)	NGS SOLiD	NGS Latest?
Read length	400-900	700	2x50	2x150
Accuracy	99.9%	98%	99.9%	99%
Reads / run	1	1 million	1.4 billion	6 billion
Time / run	20min	24h	1 week	3 days
Cost / bp	\$2400	\$10	\$ 0.13	\$0.01?

DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage ϕ X174)

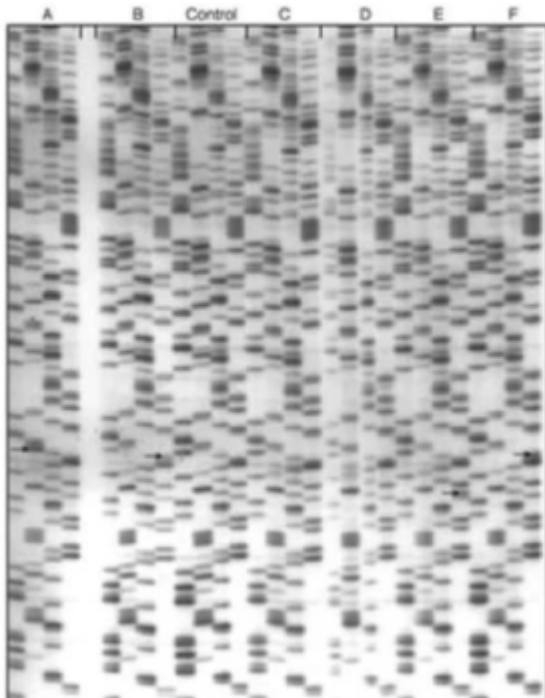
F. SANGER, S. NICKLEN, AND A. R. COULSON

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

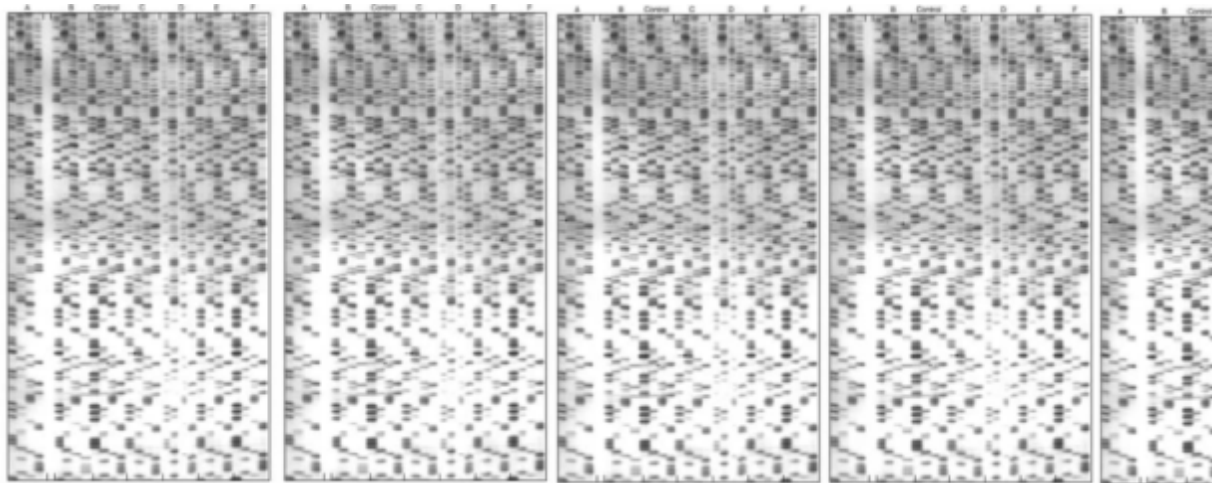
Contributed by F. Sanger, October 3, 1977

ABSTRACT A new method for determining nucleotide sequences in DNA is described. It is similar to the “plus and minus” method [Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* 94, 441–448] but makes use of the 2',3'-dideoxy and arabinonucleoside analogues of the normal deoxynucleoside triphosphates, which act as specific chain-terminating inhibitors of DNA polymerase. The technique has been applied to the DNA of bacteriophage ϕ X174 and is more rapid and more accurate than either the plus or the minus method.

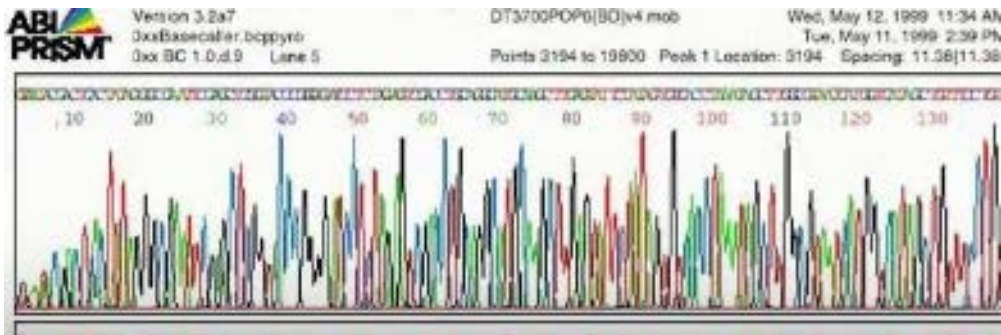




- TCAAGAGAGTGAG...
- duljina odsječka oko 150 pb
- 3 dana za odsječak od 200 pb!



3 milijarde pb / (200 pb x 8 sekvenci/gel) = 1,9 milijuna gelova



- 300pb/reakciji, 4 reakcije/smjena
- 3 smjene/dan = **3600pb/dan**/uređaj
- povećanje od 10x u 5 godina!



ABI SOLiD



454 FLX



Illumina Genome Analyzer

npr. 454 GS može sekvencirati **1 milijardu pb/dan**!

- James Watson – 1. ljudska DNA sekvencirana pomoću NGS tehnologije
- 24,5 milijardi baza genomske DNA sekvence
- 3,6 milijuna varijanti

Jim Watson	Human Genome Project
454 Life Sciences, A Roche Company	Sanger
2 months, 3 instruments	10-13 years
<\$1 million \$250,000 with Titanium	\$100 million - \$2.7 billion
7.4x coverage	7.5x coverage
250 bp read length 400bp with Titanium	500-800 bp read length

Sequencers	454 GS FLX (Roche)	HiSeq 2000 (Illumina)	SOLiDv4 (Applied Biosystems)	Ion torrent (Life Technologies)
Methods	Pyrosequencing	Sequencing by synthesis	Sequencing by ligation	Ion semiconductor
Read length	700 bp	50–250 bp	35–50 bp	400 bp
Accuracy*	$Q > 30$	$20 < Q < 30$	$Q > 30$	Q20
Reads per run	1 million	Up to 3 billion	1.2–1.4 billion	Up to 80 million
Time per run	24 h	1–10 days	1–2 weeks	2 h
Cost per 1 million bases	\$10	\$0.05 to \$0.15	\$0.13	\$1
Advantages	Read length	High throughput	Low cost per base	Less expensive equipment
	Fast		Accuracy	Fast
Disadvantages	Runs expensive	Expensive	Slower method	Homopolymer errors
	Homopolymer errors	High concentrations of DNA	Palindromic sequences errors	
	Low throughput	Short reads	Short read	

*The values of accuracy have been converted in a *Q* score value (Ewing and Green, 1998) and refer to the optimal experimental conditions for each NGS platform. *Q* score is the measure of base calling accuracy (Ewing and Green, 1998). Low *Q* values (*Q*10) can lead to increase false-positive variant calls.

Morini i sur., Front. Genet., 2015



PACIFIC
BIOSCIENCES™



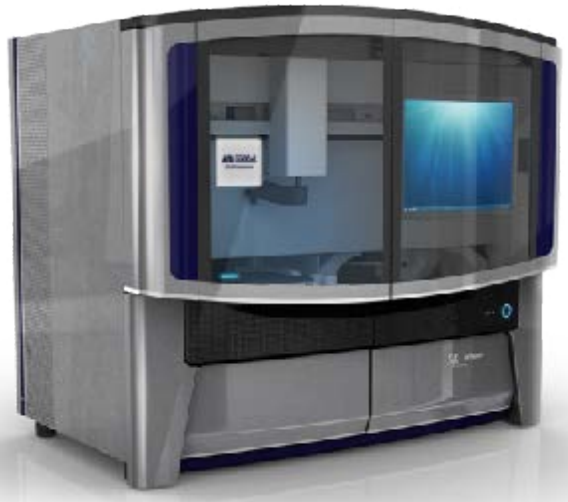
Ion Torrent (Life)



GS FLX 454 (Roche)



HiSeq 2000 (Illumina)



ABI 5500xl (AB)



**MinION (Oxford
nanopore)**



PacBio RS II (Pacific Biosci.)

medicina
krojena
prema
pojedinu

genetske
bolesti

klinička
dijagnostika

Upotreba
Sangerove
metode:

rutinska upotreba
(npr. klinička dijagnostika,
znanstvena istraživanja...)

**potvrda NGS
podataka!**

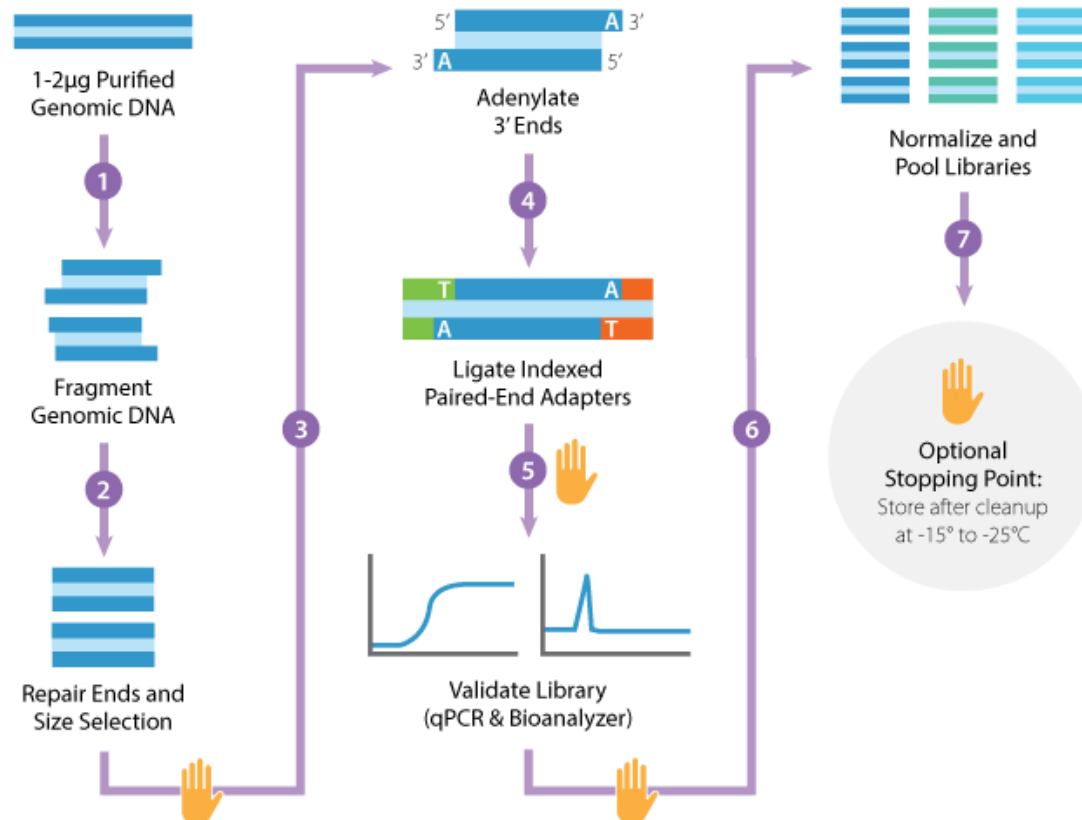
NGS – zajedničko

- priprema uzoraka – tzv. „knjižnice”
(engl. “*library*”)
- umnažanje
- sirovi podaci

1 Library Preparation

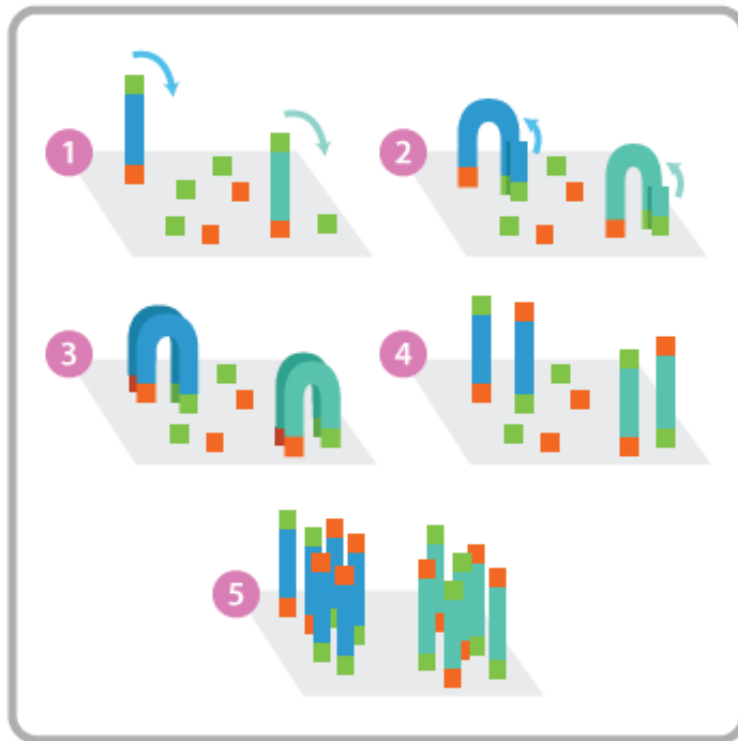


TruSeq PCR-free Library Preparation Kit

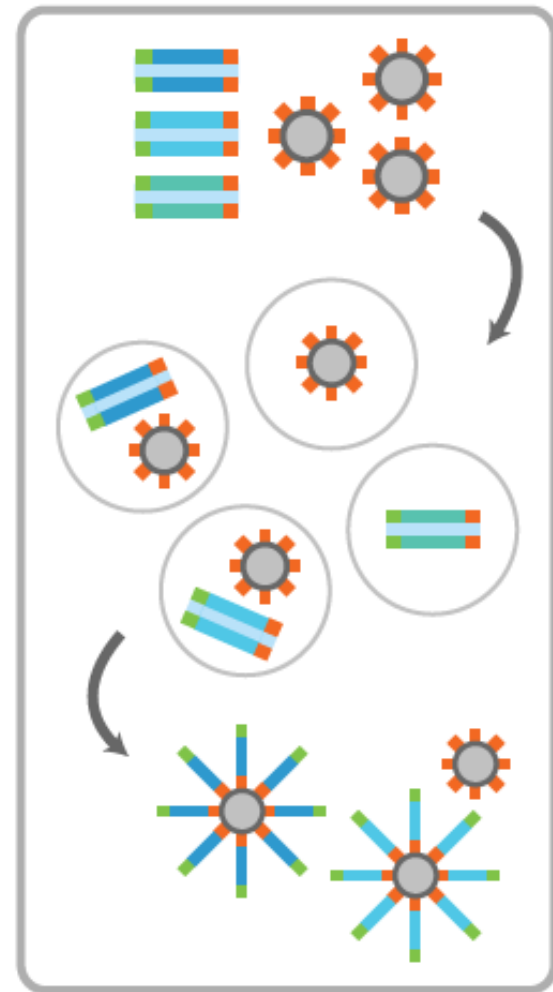


2 Clonal Amplification

Bridge PCR



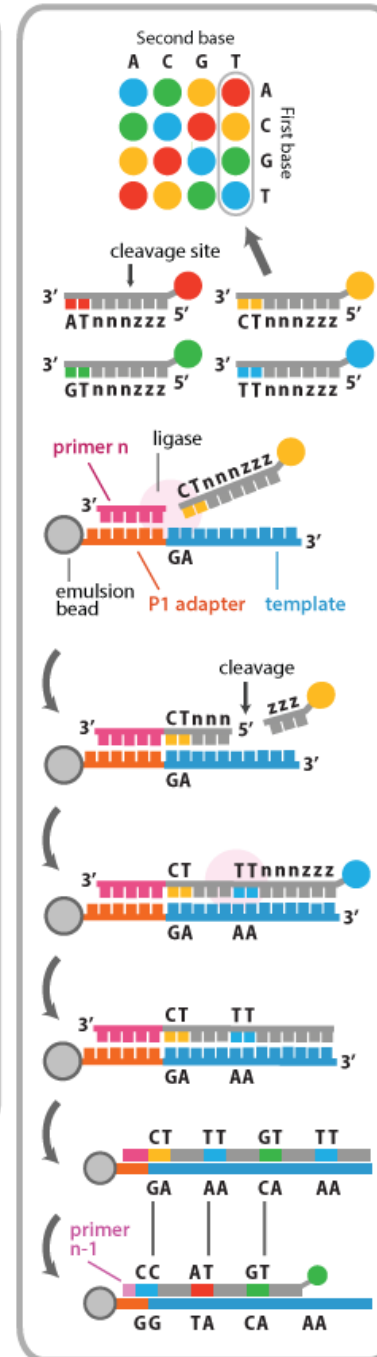
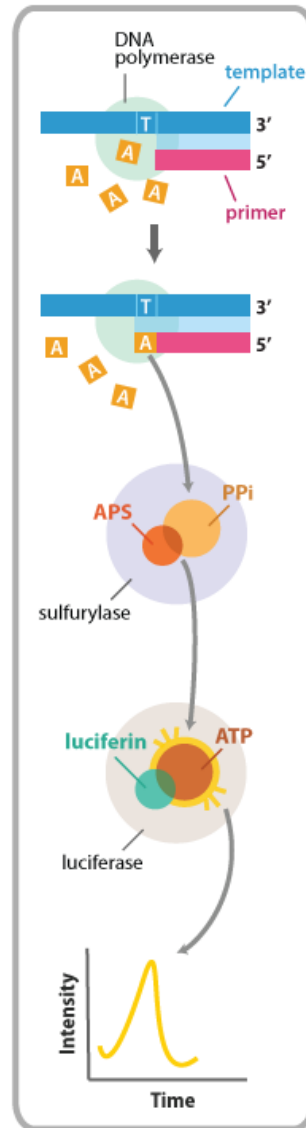
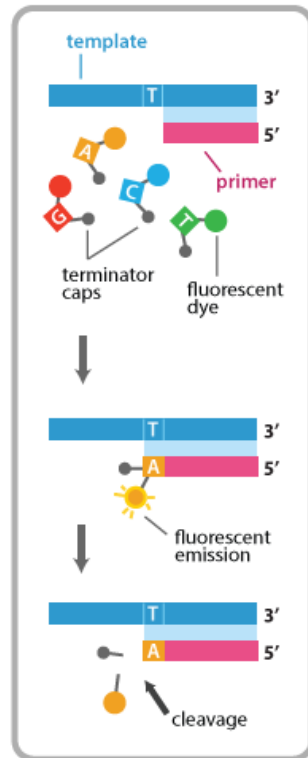
Emulsion PCR



3 Cyclic Array Sequencing

!!!

Sequencing by Ligation



Sequencers	454 GS FLX (Roche)	HiSeq 2000 (Illumina)	SOLiDv4 (Applied Biosystems)	Ion torrent (Life Technologies)
Methods	Pyrosequencing	Sequencing by synthesis	Sequencing by ligation	Ion semiconductor
Read length	700 bp	50–250 bp	35–50 bp	400 bp
Accuracy*	$Q > 30$	$20 < Q < 30$	$Q > 30$	Q20
Reads per run	1 million	Up to 3 billion	1.2–1.4 billion	Up to 80 million
Time per run	24 h	1–10 days	1–2 weeks	2 h
Cost per 1 million bases	\$10	\$0.05 to \$0.15	\$0.13	\$1
Advantages	Read length	High throughput	Low cost per base	Less expensive equipment
	Fast		Accuracy	Fast
Disadvantages	Runs expensive	Expensive	Slower method	Homopolymer errors
	Homopolymer errors	High concentrations of DNA	Palindromic sequences errors	
	Low throughput	Short reads	Short read	

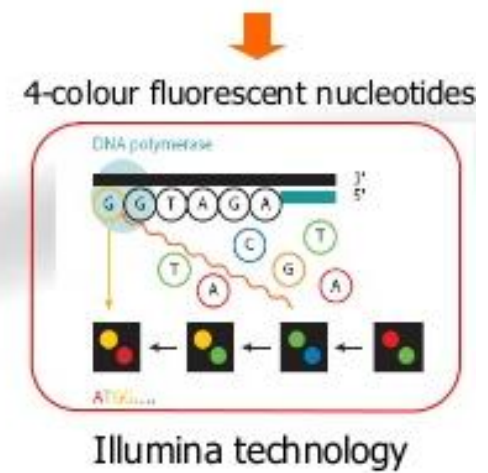
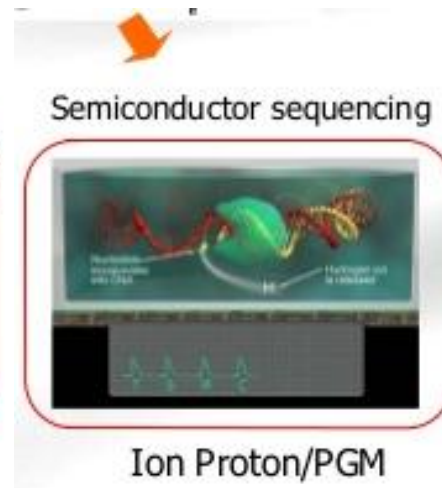
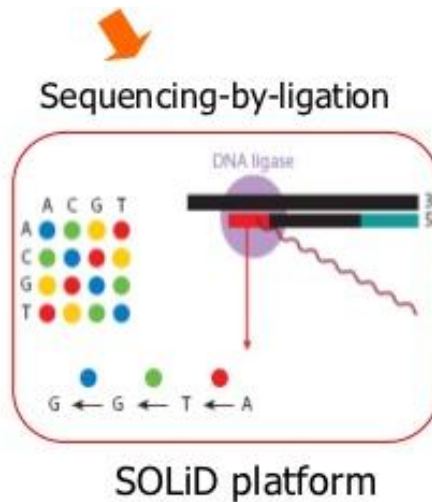
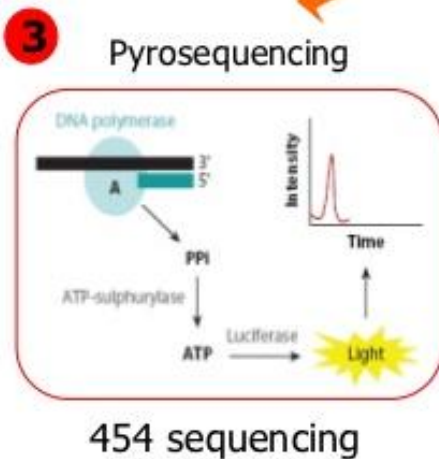
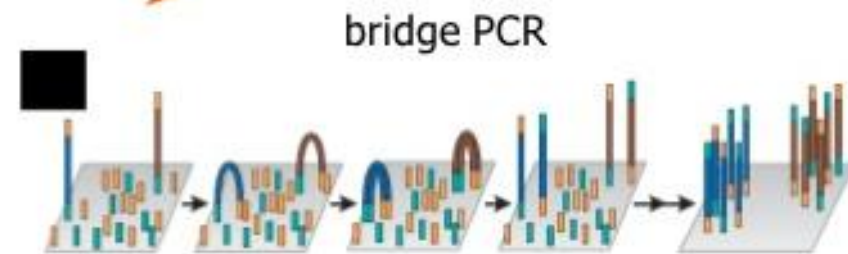
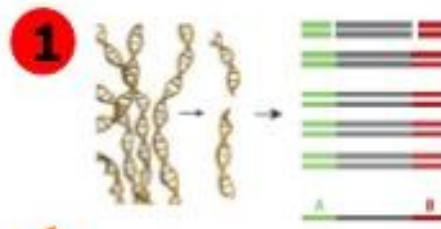
*The values of accuracy have been converted in a *Q* score value (Ewing and Green, 1998) and refer to the optimal experimental conditions for each NGS platform. *Q* score is the measure of base calling accuracy (Ewing and Green, 1998). Low *Q* values (*Q*10) can lead to increase false-positive variant calls.

Morini i sur., Front. Genet., 2015

NGS – različito

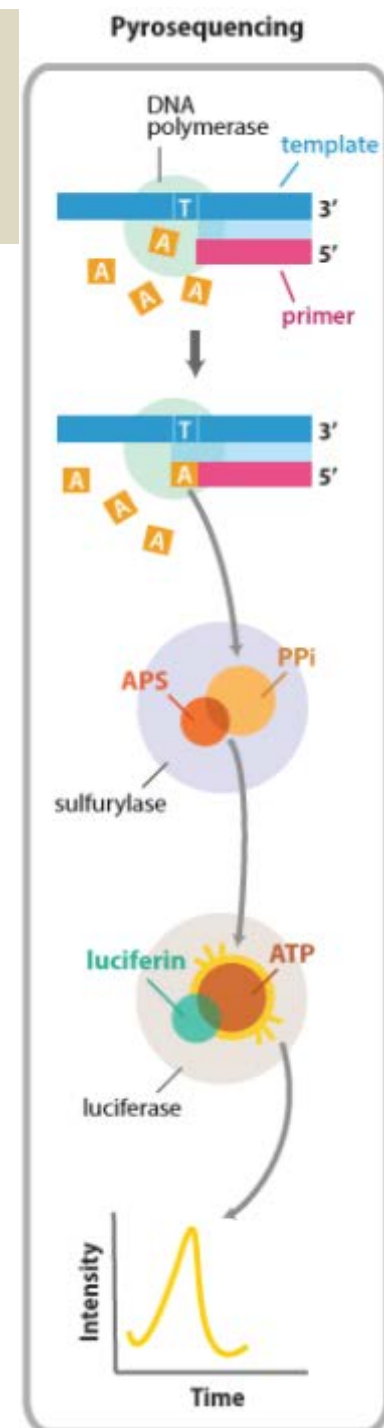
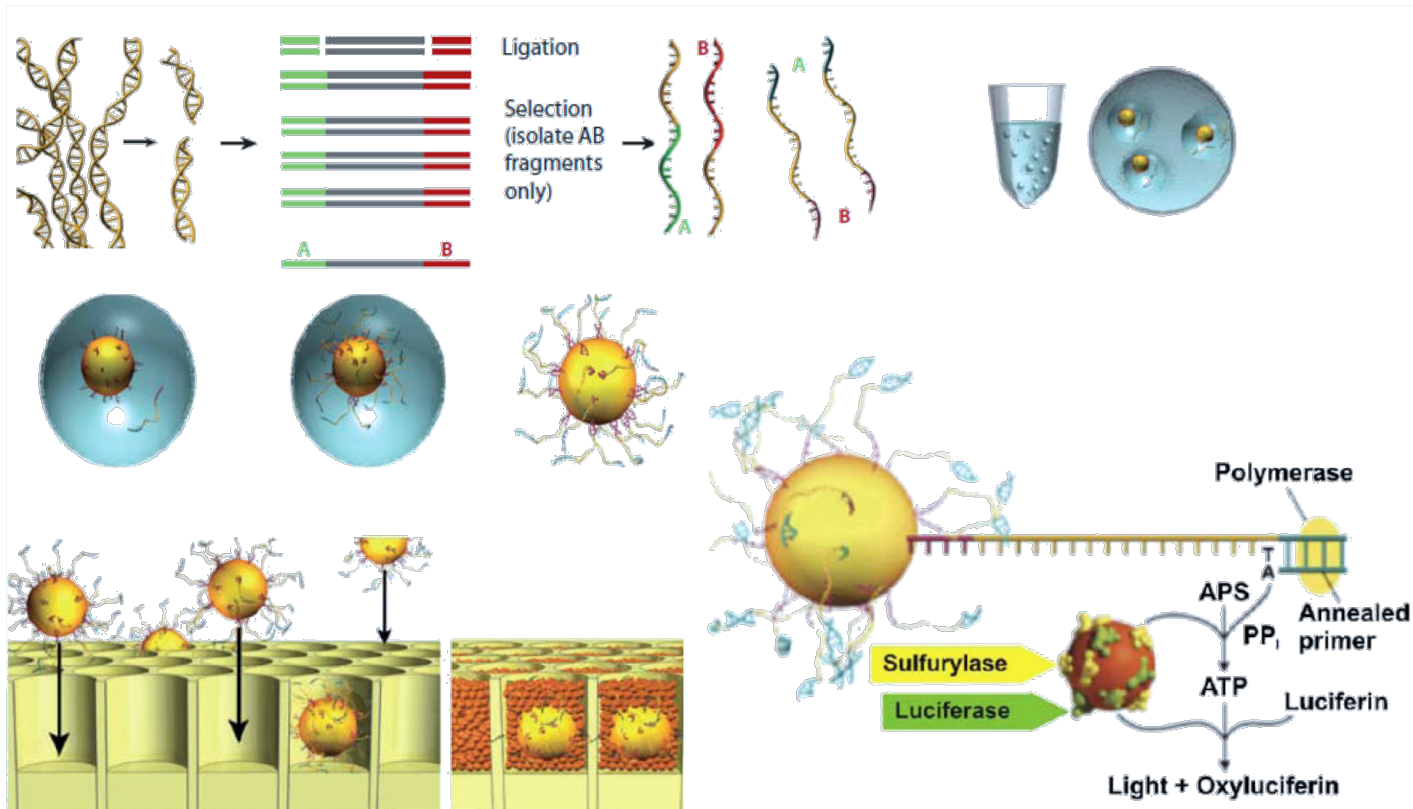
- pirosekvenciranje (engl. *pyrosequencing*)
- sekvenciranje sintezom (engl. *sequencing by synthesis*)
- sekvenciranje vezivanjem (engl. *sequencing by ligation*)
- (engl. *ion semiconductor sequencing*)

- 1 Library preparation
- 2 Clonal amplification
- 3 Cyclic array sequencing



pirosekvenciranje

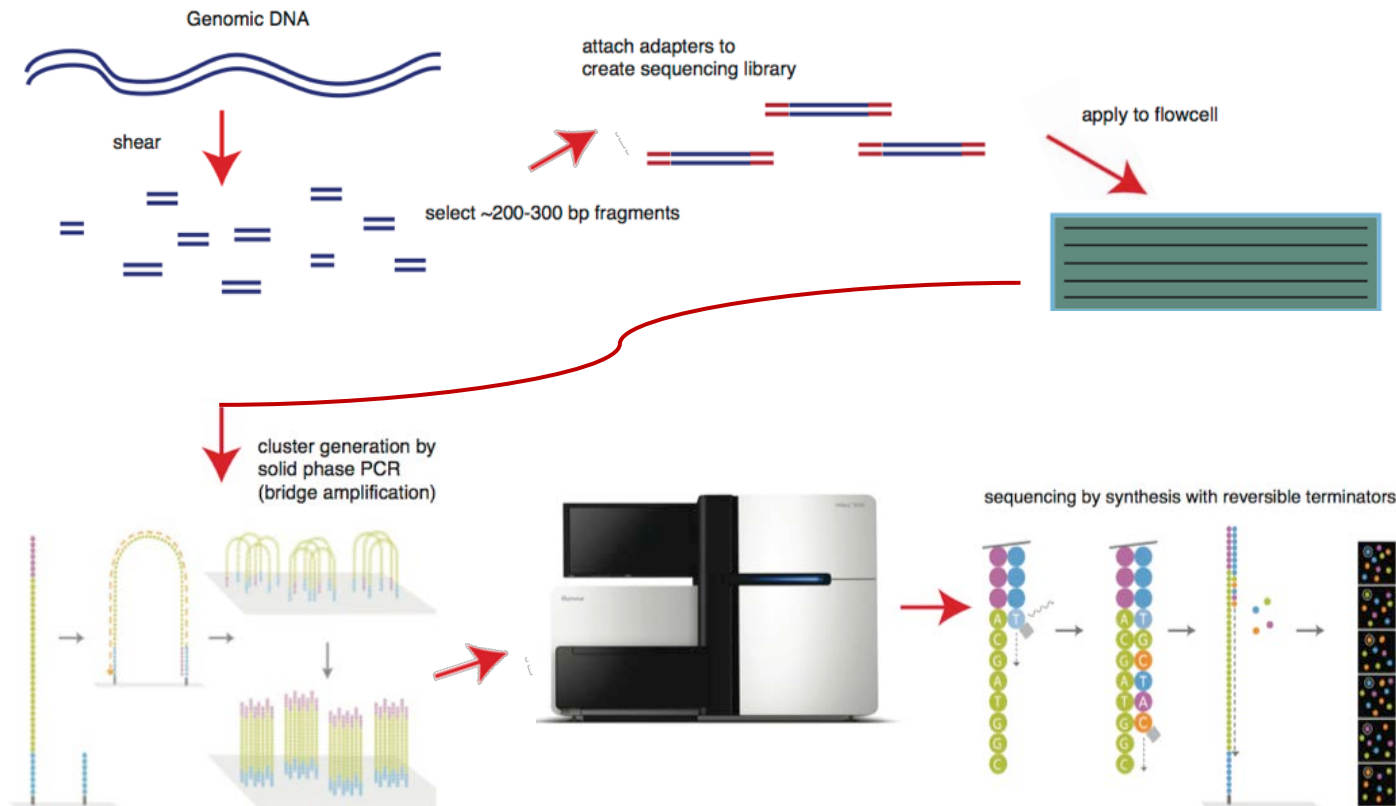
(engl. *pyrosequencing*)



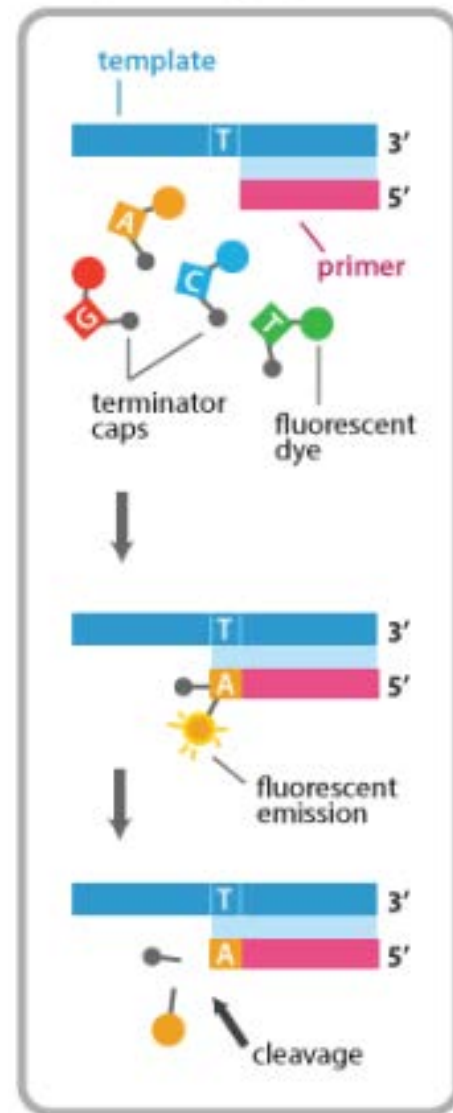
- visok trošak reagensa, visoka stopa pogreške, kod nizova od 6 i više pb (homopolimeri)

sekvenciranje sintezom

(engl. sequencing by synthesis)



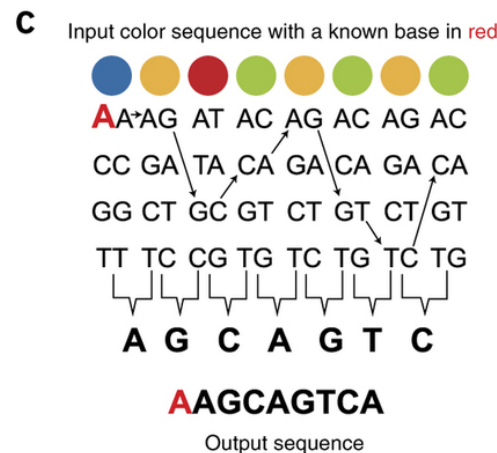
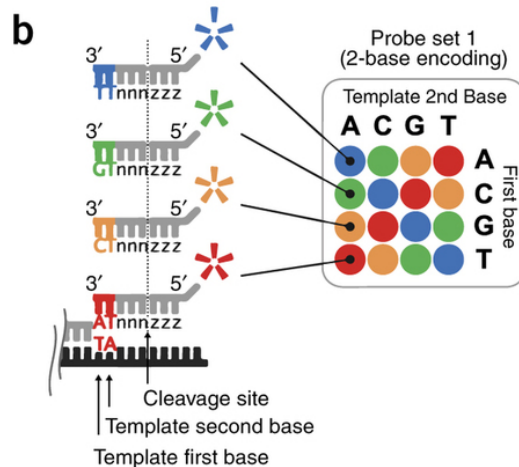
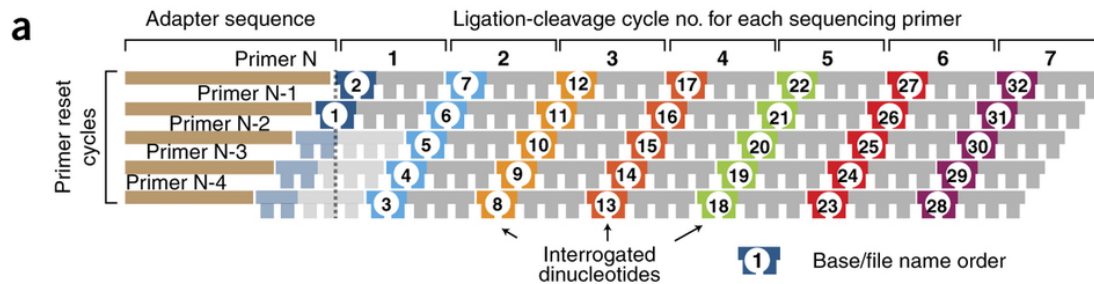
Sequencing by Synthesis



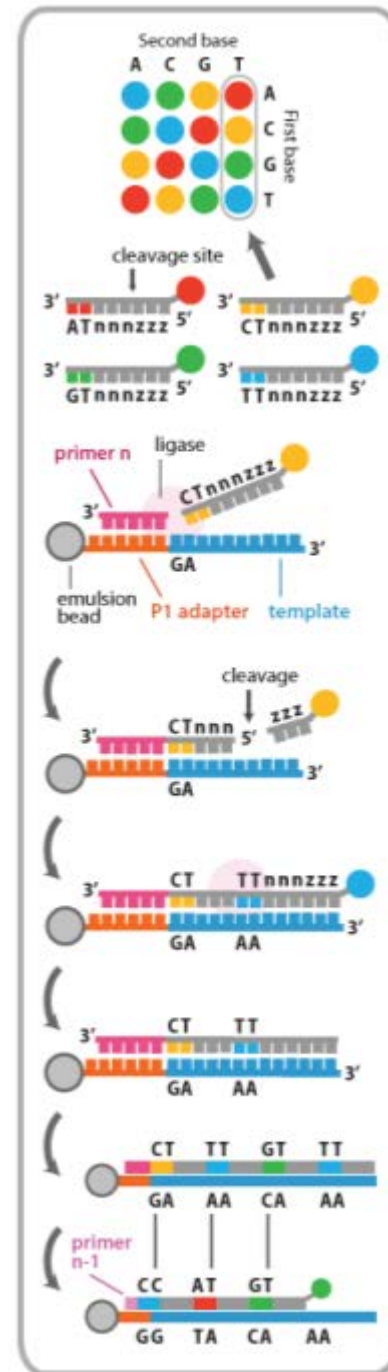
- kako reakcija napreduje, javlja se sve više grešaka – nepotpuno uklanjanje fluor. signala (veći pozadinski šum)

sekvenciranje vezivanjem

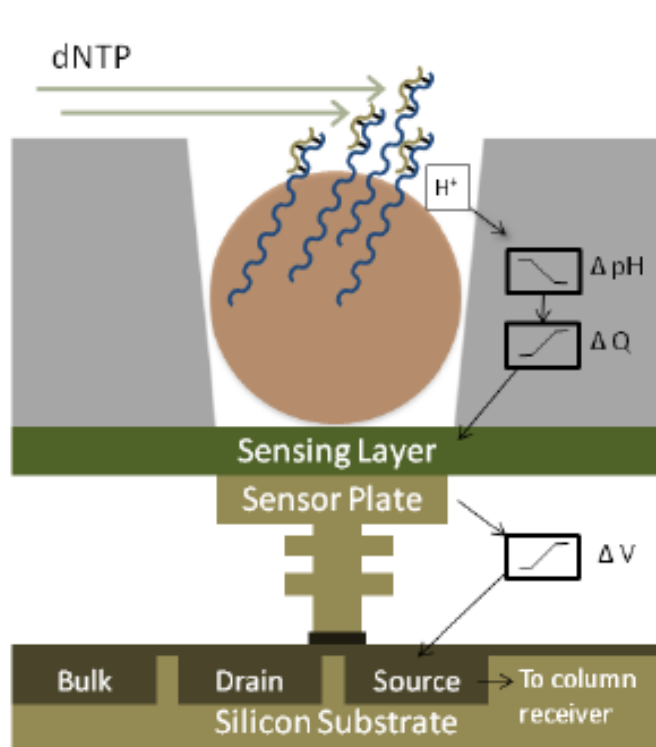
(engl. *sequencing by ligation*)



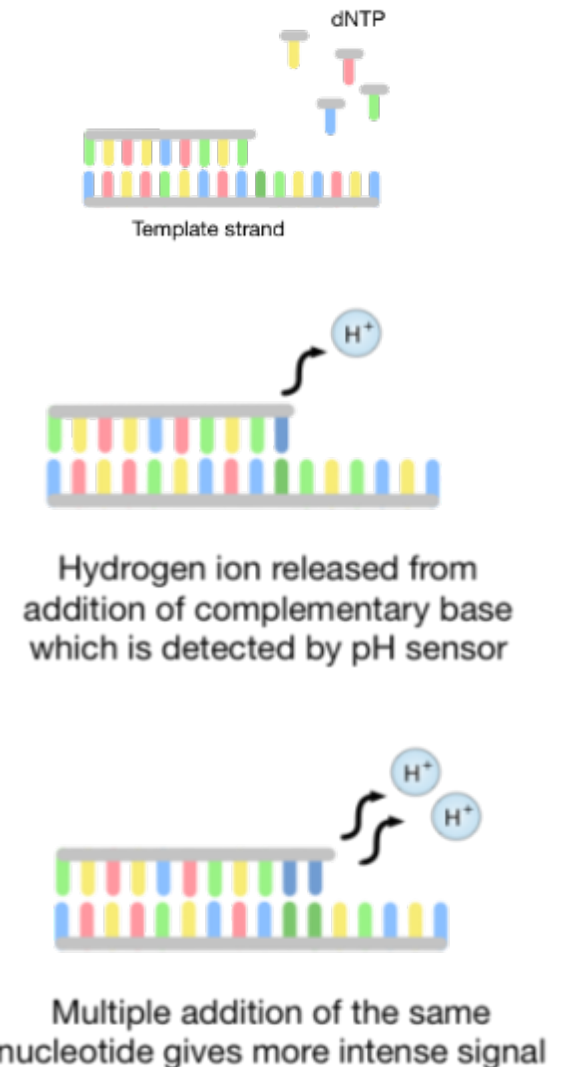
- vrlo kratki odsječki



(engl. *ion semiconductor sequencing*)

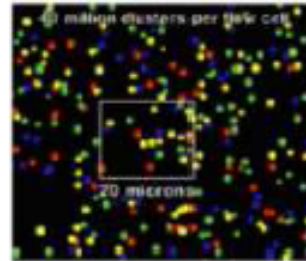


- visoka stopa pogreške kod homopolimernih sljedova (manji trošak i brže)



čovjek (3,300,000,000), miš (2,800,000,000), *Arabidopsis thaliana* (135,000,000) i *E. coli* (4,639,221) 30x

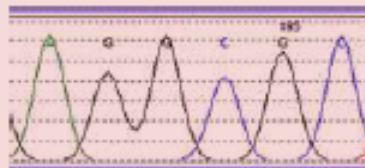
<i>Coverage of genome per run</i>				
pirosekvenciranje (engl. <i>pyrosequencing</i>)	0	0	5	151
sekvenciranje sintezom (engl. <i>sequencing by synthesis</i>)	455	536	11k	323k
sekvenciranje vezivanjem (engl. <i>sequencing by ligation</i>)	97	114	2k	69k
(engl. <i>ion semiconductor sequencing</i>)	3	4	74	2k

[illegible]

Sequencing

Image processing

Reads generation



Read alignment

Variant Detection

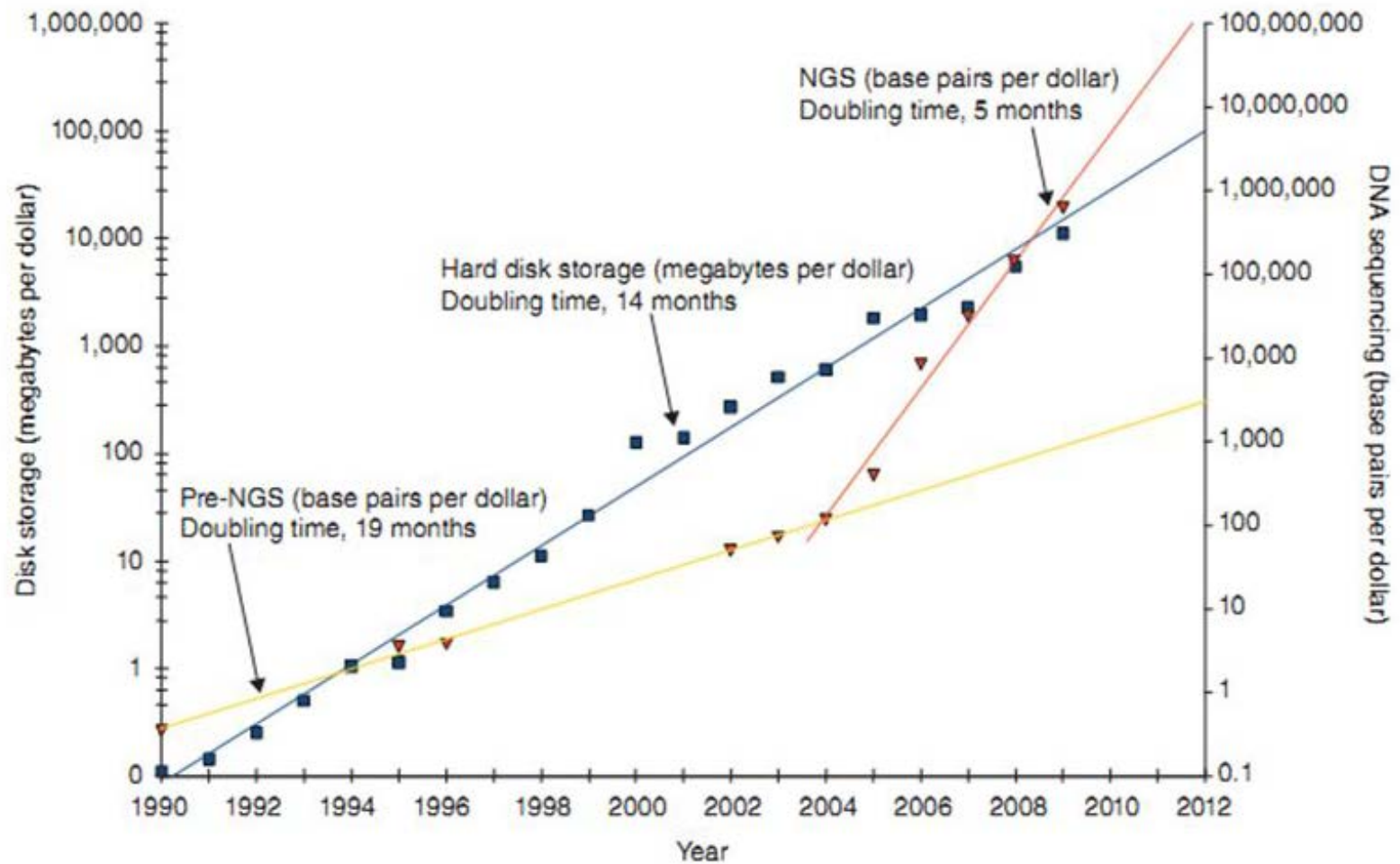
Variant prioritization

BIOINFORMATIKA

FASTQ format

sekvenca + kvaliteta

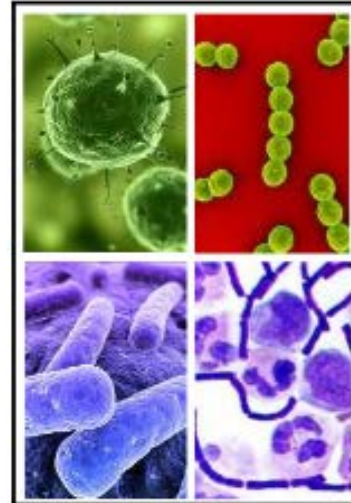
```
@SRR350953.5 MENDEL_0047_FC62MN8AAXX:1:1:1646:938 length=152
NTCTTTTTCCTTCTTTTGCCAACTTCAGCTAAATAGGAGCTACACTGATTAGGCAGAACTTGATTAACAGGGCTTAA
GGTAACCTTGTTGTAGGCCGTTTTGTAGCACTCAAAGCAATTGGTACCTCAACTGCAAAAGTCCTTGGCCC
+SRR350953.5 MENDEL_0047_FC62MN8AAXX:1:1:1646:938 length=152
+50000222C@@@@@22:::8888898989:::~::~<<<:<<<<<:<<<::<<:::~::~<<<<:<:<<<IIIIIIGFEE
GGGGGGGIIEIGDGBGGGGGGDDIIGIIEGIGG>GGGGGGDGGGGGIHIIIBIIIGIIHHIIIGII
@SRR350953.6 MENDEL_0047_FC62MN8AAXX:1:1:1686:935 length=152
NATTTTACTAGTTTATTCTAGAACAGAGCATAACTACTATTCAATAAACGTATGAAGCACTACTCACCTCCATTAACAT
GACGTTTTTCCCTAATCTGATGGGTCATTATGACCAGAGTATTGCCGCGGTGGAAATGGAGGTGAGTAGTG
+SRR350953.6 MENDEL_0047_FC62MN8AAXX:1:1:1686:935 length=152
#---+83355@@@CC@C22@@C@@CC@@C@@@CC@@@@@@@@@@@@@C?
C22@@C@:::~::~@@@@@@C@@@@@@@@@CIGIHIIIDGIGIIIIHHIIHGHHIIHHIFIIIIHHIIIIIBIIIEIFGIIFG
FIBGDGGGGGGFIGDIFGADGAE
@SRR350953.7 MENDEL_0047_FC62MN8AAXX:1:1:1724:932 length=152
NTGTGATAGGCTTTGTCCATTCTGGAACTCAATATTACTTGCAGAGTCCTCAAAGGTAATTTTTGCTATTGCCAATATTCC
TCAGAGGAAAAAAGATAACAATACTATGTTTTATCTAAATTAGCATTAGAAAAAAAAATCTTTCATTAGGTGT
+SRR350953.7 MENDEL_0047_FC62MN8AAXX:1:1:1724:932 length=152
#. , ' )-2--/@@@@@@@@@@@<:<<:
778789979888889:::~::~99999<<::<:::~::~<<<<<@@@@@:::~::~IHIGIGGGGGGDGGDGGDDDIHHIIHHIIII8
GGGGGIHHIIIIIGIIGIBIGIIIIIEIHGGFIHHIIIIIIIGIIFIG
```

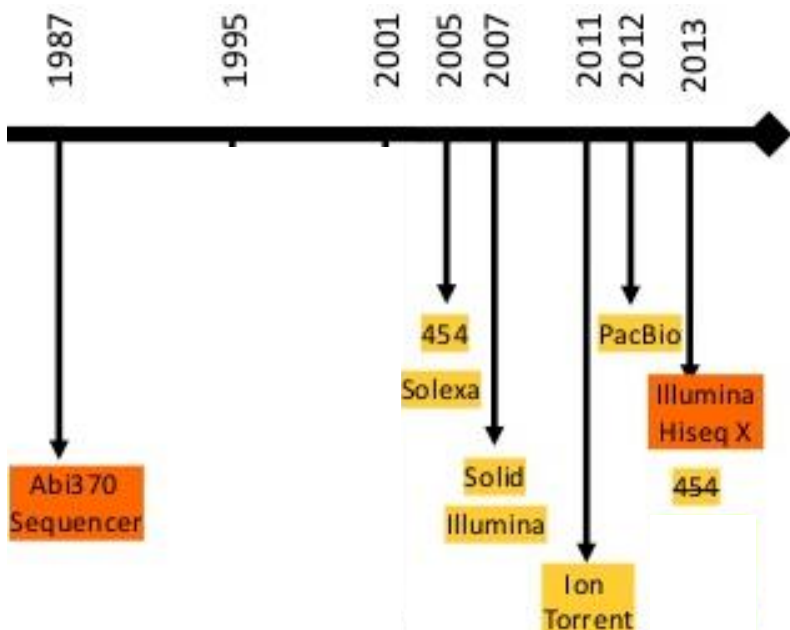
Stein, *Gen. Biol.* 2010.

NGS – primjena

- *Whole Genome Sequencing* (WGS)
- *Exome Sequencing* (Exome-Seq)
- *RNA Sequencing* (RNA-Seq)
- *Methylation Sequencing* (Methyl-Seq)



- personalizirana medicina – krojena prema pojedincu
- “*cancer genomics*” – što pretvara normalnu stanicu u stanicu raka?
- epidemiologija/forenzika – epidemije, zločini
- metagenomika – mi smo domaćini brojnim mikrobima!
- poboljšanje usjeva – sekvenciranje važnih usjeva
-



PRVA generacija

- dugi odsjecci visoke kvalitete
- *"low throughput"*
- skupo

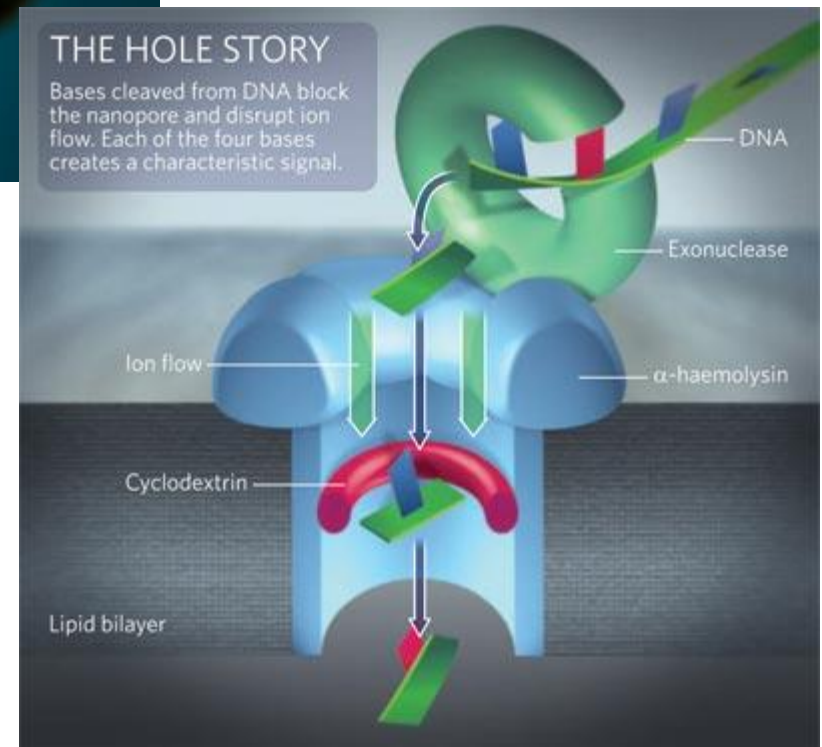
DRUGA generacija

- kratki odsjecci
- *"high throughput"* (masivno paralelno sekvenciranje)
- manji troškovi
- 454/Roche, Solexa/Illumina, SOLiD, IonTorrent/Life Tech.

TREĆA generacija

- molekula
- lagana priprema uzoraka
- Oxford Nanopore, Pacific Biosciences

Method	Read length	Accuracy (single read not consensus)	Reads per run	Time per run	Cost per 1 million bases (in US\$)	Advantages	Disadvantages
Single-molecule real-time sequencing (Pacific Biosciences)	10,000 bp to 15,000 bp avg (14,000 bp N50); maximum read length >40,000 bases ^{[61][62][63]}	87% single-read accuracy ^[64]	50,000 per SMRT cell, or 500–1000 megabases ^{[65][66]}	30 minutes to 4 hours ^[67]	\$0.13–\$0.60	Longest read length. Fast. Detects 4mC, 5mC, 6mA. ^[68]	Moderate throughput. Equipment can be very expensive.
Ion semiconductor (Ion Torrent sequencing)			up to 80 million	2 hours	\$1		Higher error rate
Pyrosequencing (454)	700 bp	99.9%	1 million	24 hours	\$10		More expensive than other methods
Sequencing by synthesis (Illumina)	50 to 300 bp		up to 1 billion (TruSeq and)	1 to 11 days, depending upon sequencer and specified read length ^[69]	\$0.05		Requires more space
Sequencing by ligation (SOLiD sequencing)	50+35 or 50+50 bp	99.9%			\$13		Not as widely used as other methods
Chain termination (Sanger sequencing)	400		N/A	3 hours	\$2400		Not practical for large-scale sequencing projects. This method also requires the time consuming step of plasmid cloning or PCR.



Budućnost (je već ovdje).

Rapid Short-Read Sequencing and Aneuploidy Detection Using MinION Nanopore Technology

Shan Wei, Zev Williams

GENETICS January 5, 2016 vol. 202 no. 1 37-44 (Early Online January 1, 2016);

DOI: 10.1534/genetics.115.182311

Article

Figures & Data

Supplemental

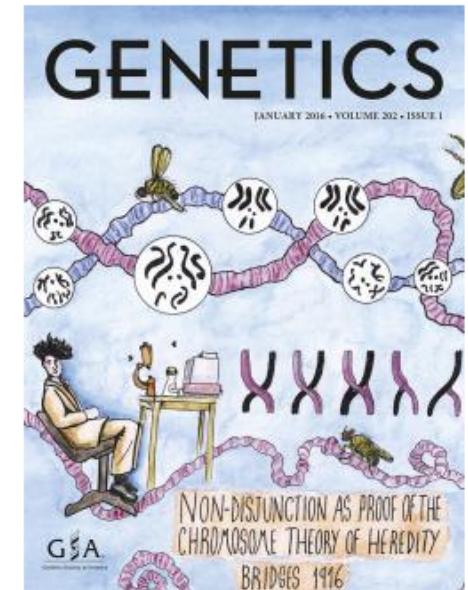
Info & Metrics



Abstract

MinION is a memory stick-sized nanopore-based sequencer designed primarily for single-molecule sequencing of long DNA fragments (>6 kb). We developed a library preparation and data-analysis method to enable rapid real-time sequencing of short DNA fragments (<1 kb) that resulted in the sequencing of 500 reads in 3 min and 40,000–80,000 reads in 2–4 hr at a rate of 30 nt/sec. We then demonstrated the clinical applicability of this approach by performing successful aneuploidy detection in prenatal and miscarriage samples with sequencing in <4 hr. This method broadens the application of nanopore-based single-molecule sequencing and makes it a promising and versatile tool for rapid clinical and research applications.

Volume 202 Issue 1, January 2016



A koliko to sve košta...?

Next-Gen Sequencer	Machine Cost	Cost per run	Minimum Throughput	Sequencing Run Time	Cost Per Mb
Illumina MiSeq	\$125,000	\$750	1500 Mb (2 x 150 Bases)	27 Hours	\$0.5
454 GS Junior	\$108,000	\$1,100	35 Mb (400 Bases)	8 Hours	\$31
Ion Torrent PGM - 314 Chip	\$80,490	\$225	10Mb (100 Bases)	3 Hours	\$22.5
Ion Torrent PGM - 316 Chip	\$80,490	\$425	100Mb	3 Hours	\$4.25
Ion Torrent PGM - 318 Chip	\$80,490	\$625	1000Mb	3 Hours	\$0.63

Samo primjer, cijene na upit!

Kupnja uređaja – na što paziti!

Primjer (illumina).

Define Your Applications.

What types of NGS experiments do you plan to perform and which applications will you use most? Consider your application's needs for these three areas:

- ☐ Throughput per run.
- ☐ Read length.
- ☐ Paired-end sequencing.

Understand the Equipment Options.

Buying equipment is a commitment; consider your needs now and for the future.

- ☐ Desktop Low-Throughput Sequencer.
- ☐ Desktop High-Throughput Sequencer.
- ☐ High-Throughput / High-Volume Sequencer.

Think About Budget Holistically.

There are many factors to consider beyond the initial capital expenditure.

- ☐ Operational Expenses: What's the cost per sample? Factor in consumables and library prep kits.
- ☐ Hands-On Labor: Consider the element of efficiency. The more time a lab tech must spend to carry out a given sequencing experiment, the less time that person has available for other important projects.
- ☐ Ancillary Equipment: What add-ons are required? Determine whether additional equipment will be needed to get the job done.
- ☐ Data Storage: On-site or in the cloud? Inquire about the costs and learn about the benefits and drawbacks of each solution.

Envision Your New Workflow and Informatics.

- ☐ Seek out any and all opportunities to save time and ensure accuracy.
- ☐ Sample and Library Preparation: How many days will it take to create the libraries? How much of that time is “hands-on”? Does the equipment vendor offer a library solution? Do the sample prep solutions support a broad range of applications?
- ☐ Data Analysis: Explore what types of analysis tools and protocols are available to users of a given platform.

Address the Big Issue of Quality.

If you're trying to put together a jigsaw puzzle and you can't make out the design on some of the pieces, you're going to have a difficult time achieving your goal. The same is true of DNA-driven research.

- ☐ Q-Score: Quality scores measure the probability that a base is called incorrectly. Thus, a higher quality score indicates a smaller probability of error.
- ☐ Instrument Operation: Address the homopolymer issue.
- ☐ Workflow Design: Can you plug and play?
- ☐ Underlying Technology: Different instruments use different underlying sequencing platforms. Make sure the technology is well proven and scalable to your future needs.

Know the Terms.

Download the full Buyer's Guide for a glossary of the most frequently used terms. If, during the buying process, you run into a word or phrase you don't know, ask a field application specialist or company representative for clarification.

Consider Reputation, Community and Support.

Ensure there's an established base of users and responsive service support. Consider the following items.

- ☐ Reputation: Investigate published research and ask existing users about ease of workflow.
- ☐ Community: Look for an established group of users who specialize in your key applications.
- ☐ Support: Seek out a well-defined and personalized onboarding process.

Kraj!

Hvala na pozornosti... 😊