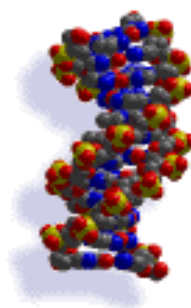




Mend  
The  
Gap



# DNA Analysis - Basic Tools and Techniques



Zlatko Liber

University of Zagreb, Faculty of Science, Zagreb, Croatia

Centre of Excellence for Biodiversity and Molecular Plant Breeding, Zagreb, Croatia

E-mail: [zlatko.liber@biol.pmf.hr](mailto:zlatko.liber@biol.pmf.hr)

# Why are archaeologists interested in studying ancient DNA (aDNA)?



King Richard III of England  
(1452–1485)



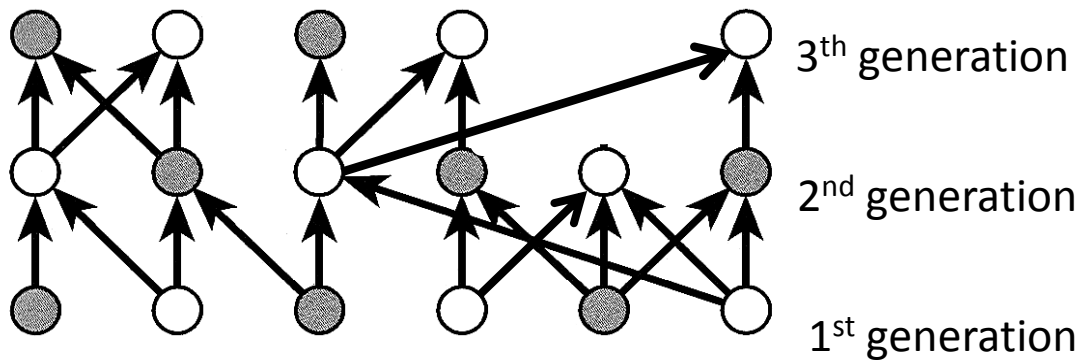
Vindija cave (Croatia)



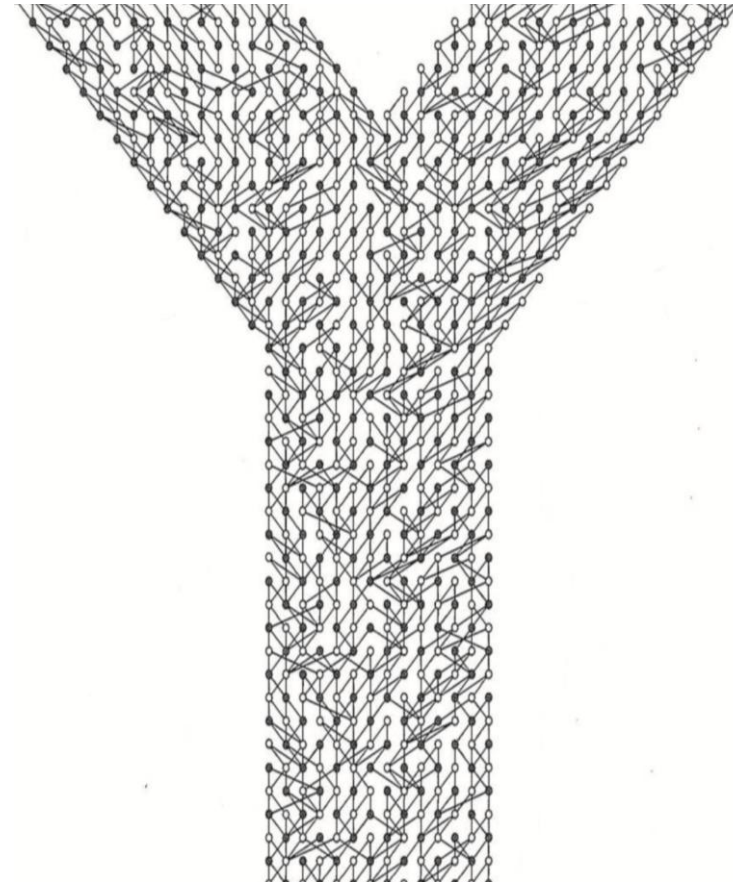
Oriental rat flea  
(*Xenopsylla cheopis*)

# Why are DNA molecules so informative?

**DeoxyriboNucleic Acid (DNA)** is a molecule that carries all instructions used in the growth, development, functioning and reproduction in all known living beings.

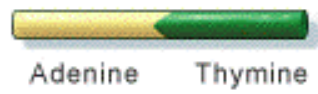


**DNA is a time traveller!**

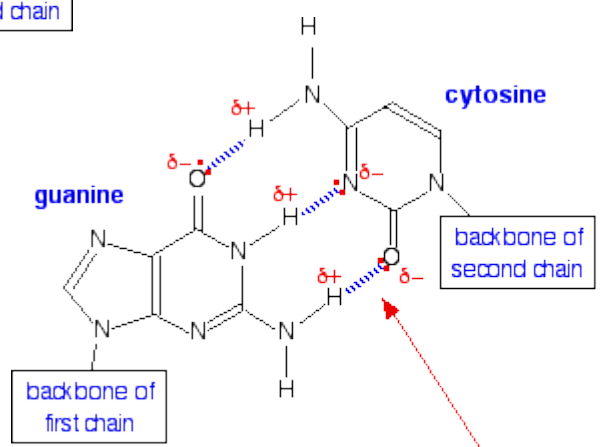
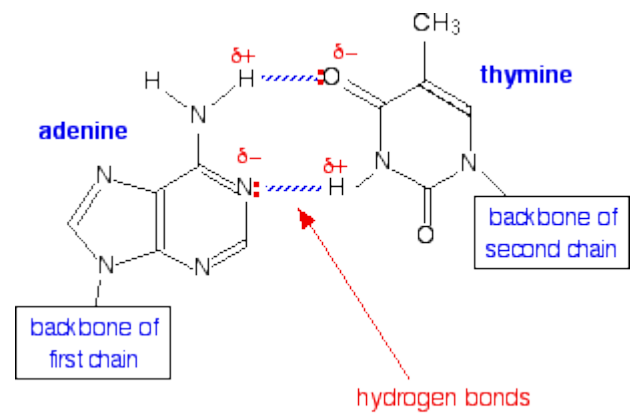




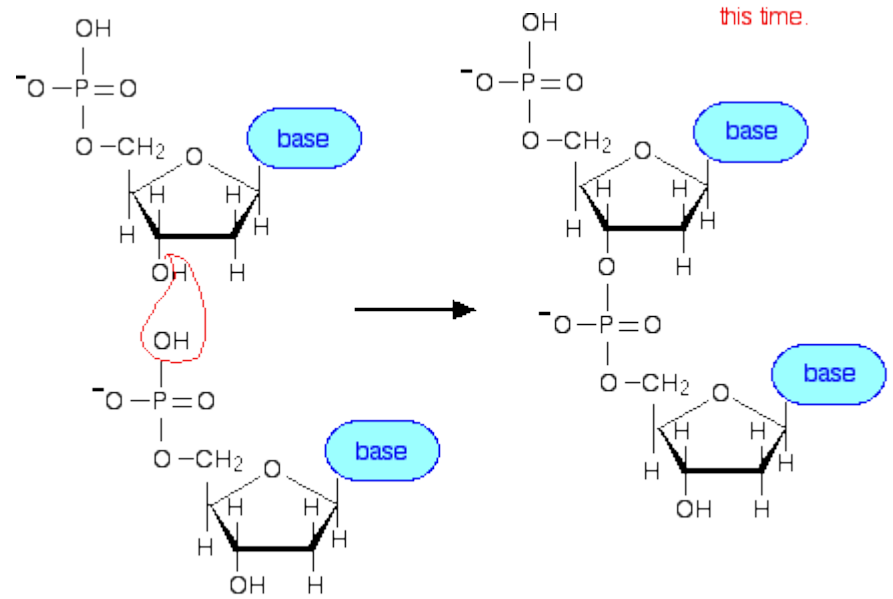
Base pairs



Sugar phosphate backbone



Notice 3 hydrogen bonds this time.

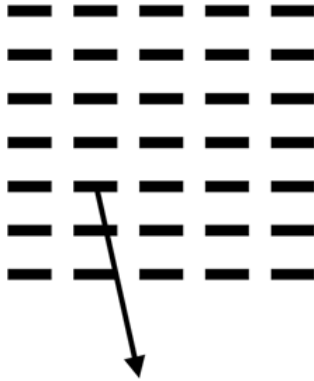




agaggttat tccaagt tgc gcaattaac ccacaagaaa ggacaaaaaggaaactggag agccttgaag cattgaagaa aacttttaag aaactggatc atgatgtaaatatacatg  
aaaggaatca aactattgag ggtctaagta accaagtagccatttacc aa gctcagctaa tggagtgta taggaggttg agttgttgga cgaacatcgatagaatagaaacactgag  
ttggagaa tcataggga aatccattgaaagaatccag attcacaagg aacattacag aaagaaccaa ctcttgccaa tagaatgtgc aacaacacag ttccacagcg ggataca  
cctatggcg atgggaggta atagtagtat gcaagaagct cactccatgt cttggcttcc tgataatgat caccagcaaa caatcttacc tgggtattcc agttttctc cccatagaga  
gatggatggg tcatteccg ttactcaag ctgcttctt gagtctacga aaccagaaga tcagatatgc agcaacccgg gacaacagtt tgagcagtta gaacaacaag gaaacg  
ttgggggta caacaacttg gagaggaata ttcatatct acaccgttg gtactactt gggaatggaa gaagatcaag agaaaaagat aaaatctgaa atggaattga acaactt  
caacagcaa cagcaacaac aacaacaaca acaacaagat cttcaatgt atgatcccat ggctaataat aatgggtggct gctttcagat tctcatgat cagtccatgt ttgtcaatg  
catcatcat catcatcacc accatcatcaaaattgggtt ccagattcaa tgtttggta gacttctac aaccagcaac cgaactaactcaagacatgg ctcatgtatg tatgatgact g  
tgaccaagg actagtttga gccattgcct acaatgtcta agataaatat atatatgtag acagcgagag agtgtgtgtacatgtttttg tatgtgtgag cttgtgtcat atactcaggg g  
atttgtgta taagagatgt tcttaggaa ccgtatatta cctttaacaa gaatgataag tatatcaatatattaatgtt ttataggtg ttggccaata tattgtttga gacc agcctttct c  
cggattgaa gacgttttt caagattcat caatcttc aaacaagaac gagagagtgc tctatactt cctgaccaga atagacgccc agatattcaaaacaaagagt gtctactaag  
aaactcaaga ctgaaaatga catcgtctt caagtcacga accctgcagc tatcaactct gacgtcgagg aacttgagca tgaagttgaggttacaac aacaactcca aatggcag  
gaagaactaa gtgttatagt ttataattt tttaactcaa atcatgattg cttactatat aatatgagca ggagatacga accggatcca ataagattca cgaccatgga ggaatacga  
gcaacttct cgacactcta acacatgttg tccaacggcg agaccatctc atgagcaacc atttatctc gtatgaagca tctactatgc aaccaaacat tgggtggtcct ttgtaaatg  
gggtggcta cctgaaaatg ggaccaatca aaccatttg ttgatgcat cggcccatc taaccaactc agagaattat catcggcaat gtatgaacca ttgttgcaag gaagtagc  
agctcgaacaaaacaaca tgagtgaatg ccacgtgacg aaccacaatg gtgaaatgtt ccctgagtgggctcaggcgt attcatctc ggccttattc gcttctatgc agcaacag  
gaaggtgtg ggtccagta tagaagagat gatgccagct cagcagagtg atattccggg agtgacggcgagacacaag ttgatcatga agtctccgac tatgaaacaa aggtt  
ctcagtag caataatgta ttattttt atgtatacca ttttaaaaa ttattcata ttttttcca acaaccata aaaaagtga cttttttt ccatgagaag acgatgaaga gtcgtgt  
ttgtatac tattctttt ggcaacacca ttatatgtt ttatgggcga aaagcaaatc aagttttatc ttcttaatt ttcagcctttct ccggaaaaac tcggattgaa gacgttttt caag  
aatcttc aaacaagaac gagagagtgc tctatactt cctgaccaga atagacgccc agatattcaa aacaaagagt gtctactaag gattctgcag caactcaaga ctgaaaat  
catcgtcttcaagtcacga accctgcagc tatcaactct gacgtcgagg aacttgagca tgaagttgtaggttacaac aacaactcca aatggcagaa gaagaactaa gtgttata  
tttaactcaa atcatgattg cttactatat aatatgagca ggagatacga accggatccaataagattca cgaccatgga ggaatacga gtttctgaga agcaacttct cgacactct  
ccaacggcg agaccatctc atgagcaacc atttatctc gtatgaagca tctactatgc aaccaaacat tgggtggtcct ttgtaaatg acgtcgttga aggttggcta cctgaaaatg  
aaccatttg ttgatgcat cggcccatc taaccaactcagagaattat catcggcaat gtatgaacca ttgttgcaag gaagtagctc aagctcgaacaaaacaaca tgagtgaat  
aaccacaatg gtgaaatgtt ccctgagtgg gctcaggcgt attcatctc ggccttattc gcttctatgc agcaacagca tgaaggtgtg ggtccagta tagaagagat gatgcca  
ccggaaaaac tcggattgaa gacgttttt caagattcat caatcttct aaacaagaac gagagagtgc tctatactt cctgaccaga atagacgccc agatattcaaaacaaagag  
gtctactaag gattctgcag caactcaaga ctgaaaatga catcgtctt caagtcacga accctgcagc tatcaactct gacgtcgagg aacttgagca agtttgtaggttacaaca  
atggcagaa gaagaactaa gtgttatagt ttataattt tttaactcaa atcatgattg cttactatat aatatgagca ggagatacga accggatcca ataagattca cgaccatgga  
gttctgaga agcaacttct cgacactcta acacatgttg tccaacggcg agaccatctc atgagcaacc atttatctc gtatgaagca tctactatgc aaccaaacat tgggtggtcct  
acgtcgttga aggttggcta cctgaaaatg ggaccaatca aaccatttg ttgatgcat cggcccatc taaccaactc agagaattat catcggcaat gtatgaacca ttgttgca  
gaagtagctc aagctcgaac aaaacaaca tgagtgaatg ccacgtgacg aaccacaatg gtgaaatgtt ccctgagtg gctcaggcgt attcatctc ggccttattc gcttctatg  
gcaacagca tgaaggtgtg ggtccagta tagaagagat gatgccagct cagcagagtg atattccggg agtgacggcg agacacaag ttgatcatga agtctccgac tatga  
gggttctca actcagtagc aataatgta ttattttt atgtatacca ttttaaaaa ttattcata ttttttccaacaaccata aaaaagtga cttttttt ccatgagaag acgatga  
gagagagagat gatgccagct cagcagagtg atattccggg agtgacggcg agacacaag ttgatcatga agtctccgac tatga



1<sup>st</sup> generation of DNA sequencers  
(started 1987)

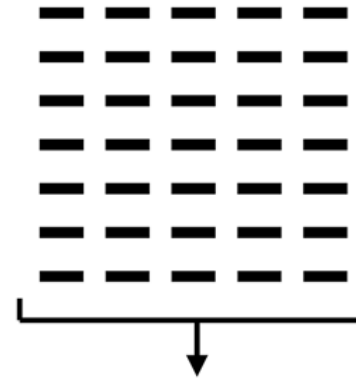


one fragment (ca 1000 bp long) per run



(human genome project /1990 – 2003/)

2nd and 3rd generations of DNA sequencers  
(last ten years)



millions of fragments per run



MinION

# How to investigate DNA?

1. DNA isolation
2. PCR methods
3. Sequencing (1st, 2nd or 3th generation)
4. Bioinformatics

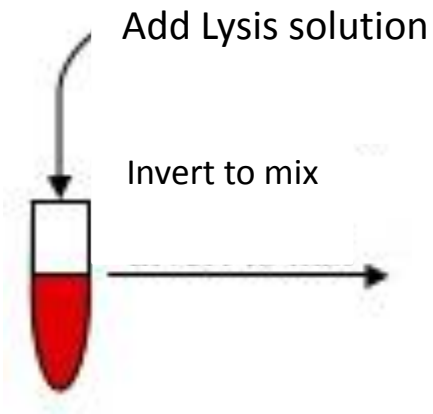
## DNA isolation

- 1) Cell lysis
- 2) Removing proteins and some other compounds
- 3) DNA purification



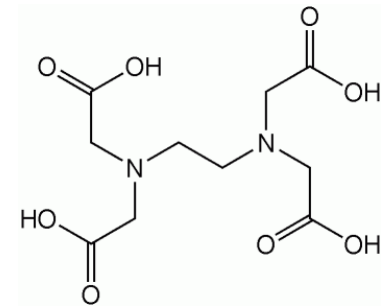
**Archaeological sample** need specific pretreatment because there is high possibility of **contamination** from microenvironment of the fossil, but also present day humans !!!



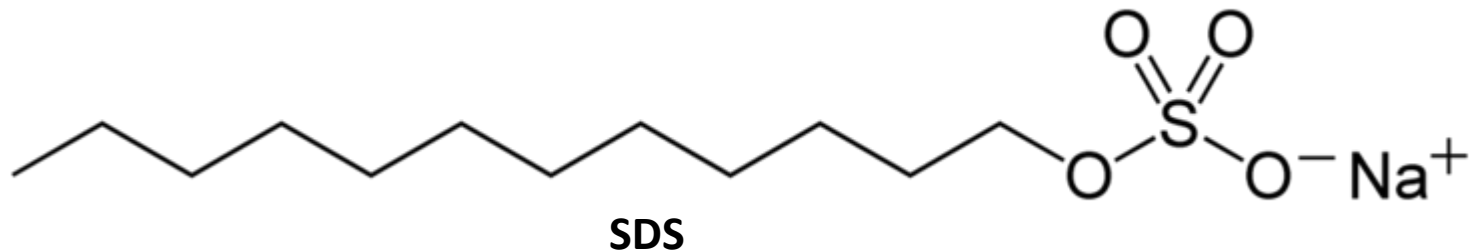
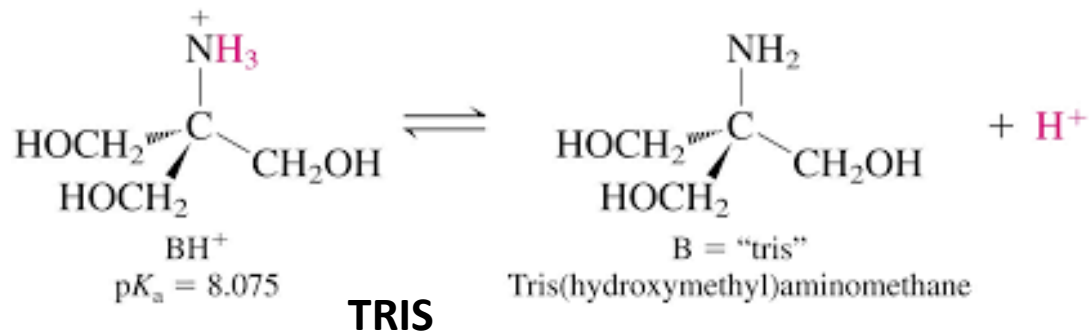


Incubate at 37 - 65 °C 5' to  
12h with occasional or  
continuous stirring

## EDTA



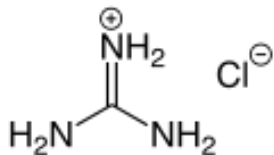
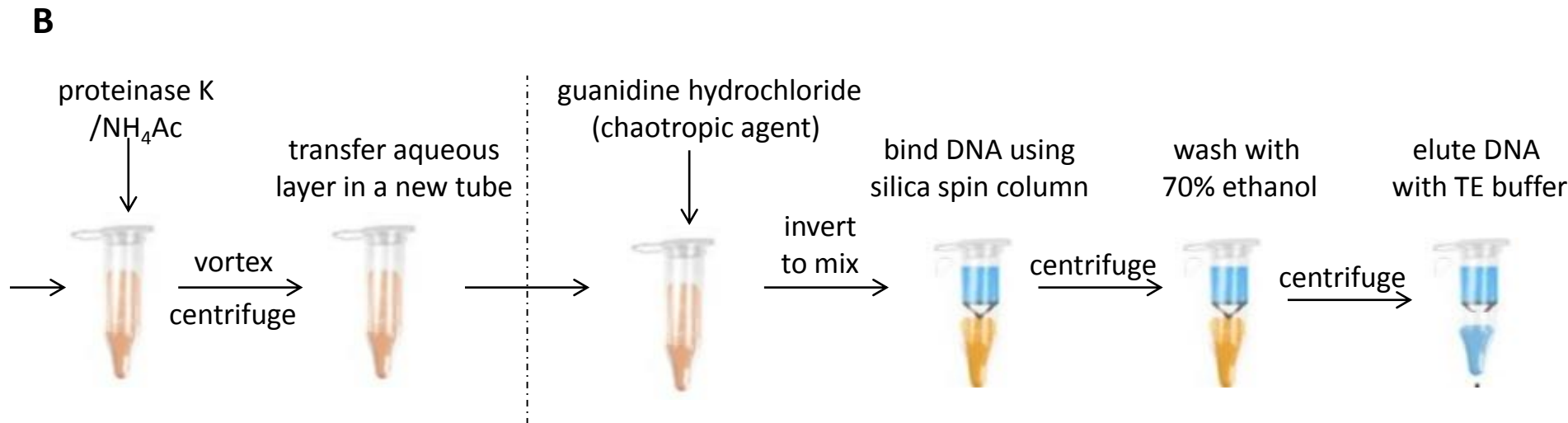
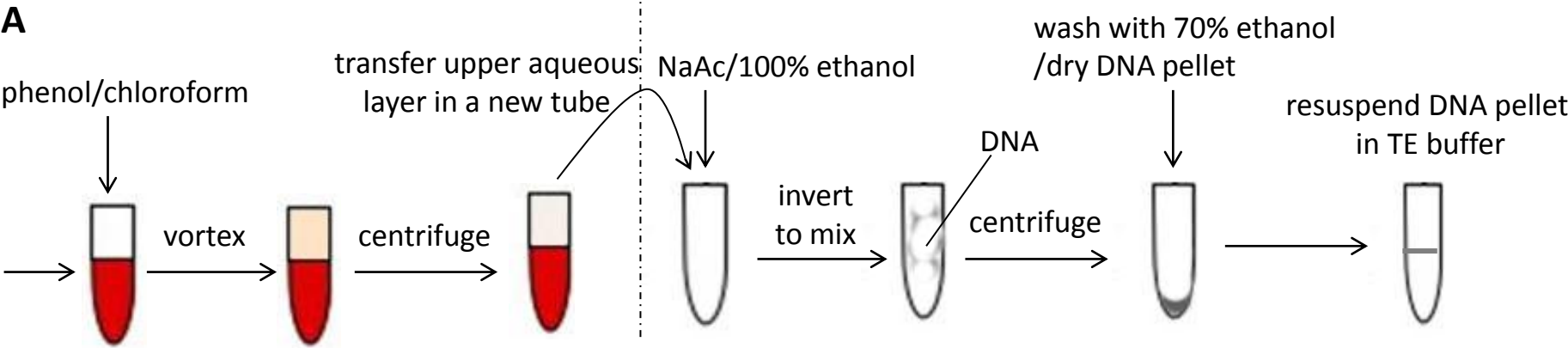
General solution of lysis buffer contains: TRIS-HCl, EDTA, SDS or CTAB





## 2) Removing proteins and some other compounds

### 3) DNA purification



Chaotropic agent - disrupt the hydrogen bonding network between water molecules

- DNA sample from two ~40,000-year-old Austrian cave bears was a mixture of bacterial, fungal, plant and other DNAs. Less than 6% of the recovered DNA was determined to be of cave bear (J. P. Noonan et al. 2005: Genomic sequencing of Pleistocene cave bears. Science 309, 597–599. )

The precautions usually applied in ancient DNA laboratory against DNA contamination:

- 1) complete separation of ancient laboratory and their rooms from other laboratories
- 2) direct delivery of all equipment and reagents to the laboratory
- 3) positive pressure generated with filtered air that excludes particles larger than  $0.2\ \mu\text{m}$
- 4) UV irradiation and bleach treatment of all surfaces
- 5) bone surface was removed prior to extraction



# PCR methods

## Polymerase Chain Reaction

= amplifying a specific DNA fragment *in vitro* up to billion copies  
(Kary B. Mullis - Nobel Prize in Chemistry 1993)



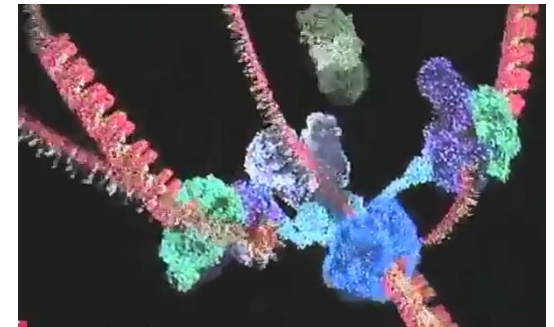
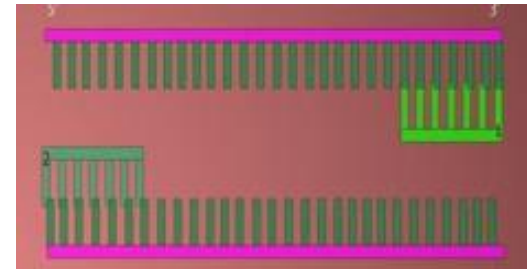
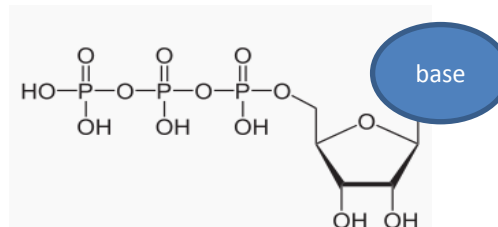
Thermal cycler



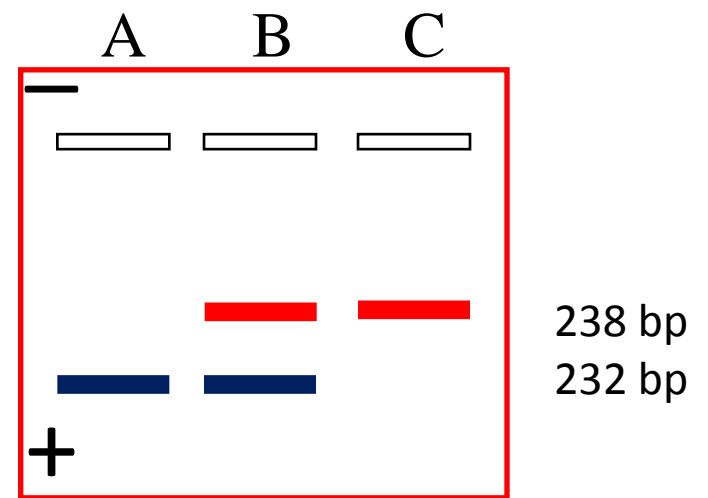
Thermal cycler device is able to specifically increase and decrease temperature

## PCR ingredients:

- 1) DNA
- 2) PCR buffer with  $MgCl_2$
- 3) Taq [DNA](#) polymerase
- 4) dNTP mix (A, C, G, T)
- 5) Primer 1
- 6) Primer 2





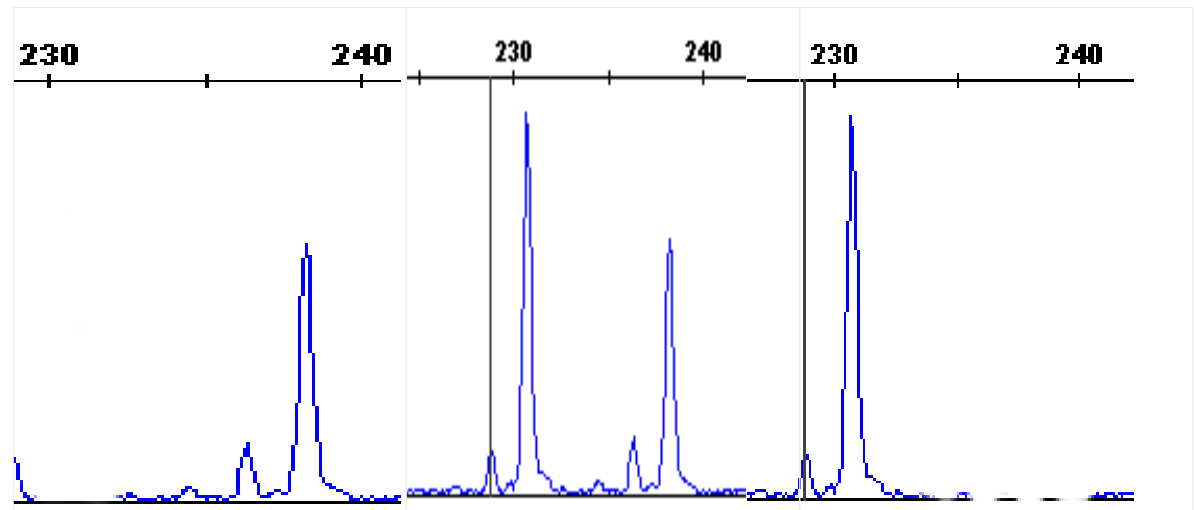


direct detection by electrophoresis (e.g. microsatellites )

→ PCR products

A B C

sequencing  
(1<sup>st</sup> generation)

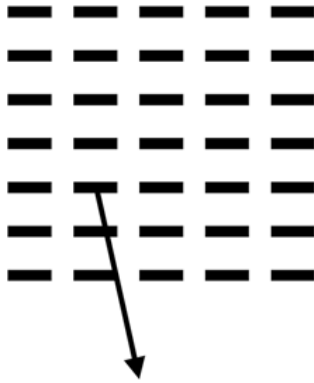




NNN 10 20 30 40 50 60 70 80 90 100 110 120  
 A G C G C A N C A G C T C G G C C C A T G C G C C G C G G A A T T C G A T T C T A A G G C C T T G C T A G C A G A A G C A T T T C A A G G G A A T T T A G A A A T A A T G C A G A T G T G A T A A T T A C A A C A C T A C A A G G T  
 130 140 150 160 170 180 190 200 210 220 230 240 250  
 A A C A T A A G A C T T T T T C C T G T A T C T T G G A T G G G C T A A C C A T C T T T T T C A C A A G A C A A T  
 260 270 280 290 300 310 320 330 340 350 360 370  
 C A C C A T G T C T T T G A A A G A T A T G A T G C T T T T C A A A C C C T A T G T T T T G T A A T C A T T C T A A T C C T A T G C A A C C C A T G T G C T C A T T A C C A T C T A T A A T G T G A T T A G A G A A A A A T A T G A T T T G A A C A A C A  
 380 390 400 410 420 430 440 450 460 470 480 490 500  
 G A G A A A T C A A G A A A G T A G G A C A A A A C T A T G A C A T G T T A A T T T C T T A A G T A G C A C C A T C T A A C A C A A T T T T C T T T G T T A C T T T A C T T T A C A A T T A C A T A A G T T A G A T G T T C A A T T C G A A T T T A A T  
 510 520 530 540 550 560 570 580 590 600 610 620  
 A T G A A C T T G A A A T C A T G C T T C T T G C T T T A A T T G T C A A T C T A T C A A A T G T A C C A T C C C A A A T T T C A T C A C A A T A T T T T A T C T T G A C T T G A G T T T T G A C T A A A A T A A T C A T A A A T A T T T C T T T C C G

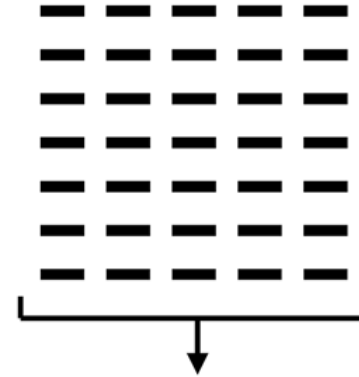
## NGS sequencing

1<sup>st</sup> generation of DNA sequencers  
(started 1987)



one fragment (ca 1000 bp long) per run

2<sup>nd</sup> and 3<sup>rd</sup> generations of DNA sequencers  
(last ten years)



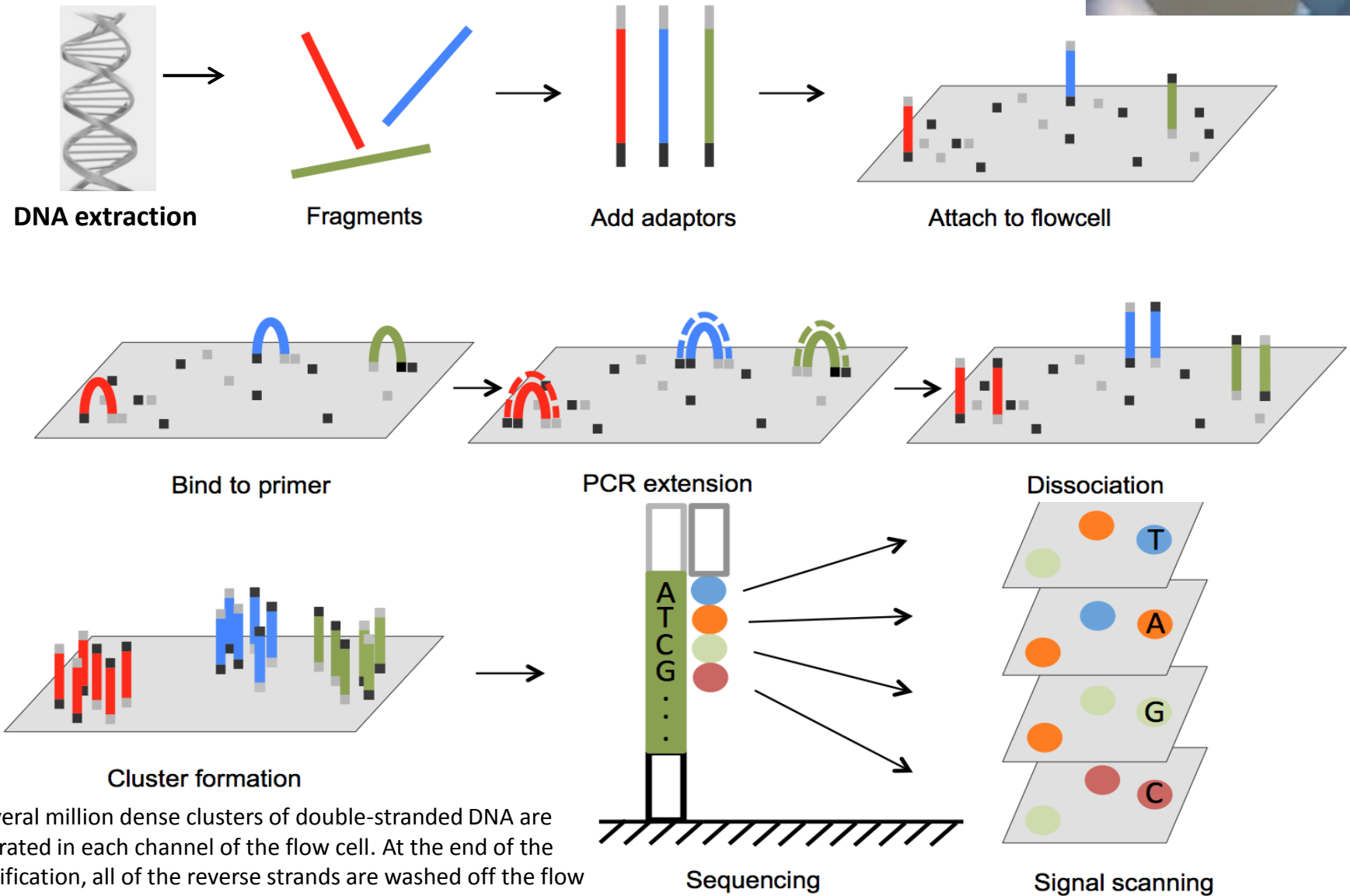
millions of fragments per run

NGS = **N**ext **G**eneration **S**equencing

- 1) Sequencing by synthesis
- 2) Nanopore sequencing



# Sequencing by synthesis



TGCTACGAT...



TTTTTTTGT...



## Illumina NGS sequencers (sequencing by synthesis)

### MiSeq Series

MAX OUTPUT

**15 Gb**

---

MAX READ NUMBER

**25 million**

---

MAX READ LENGTH

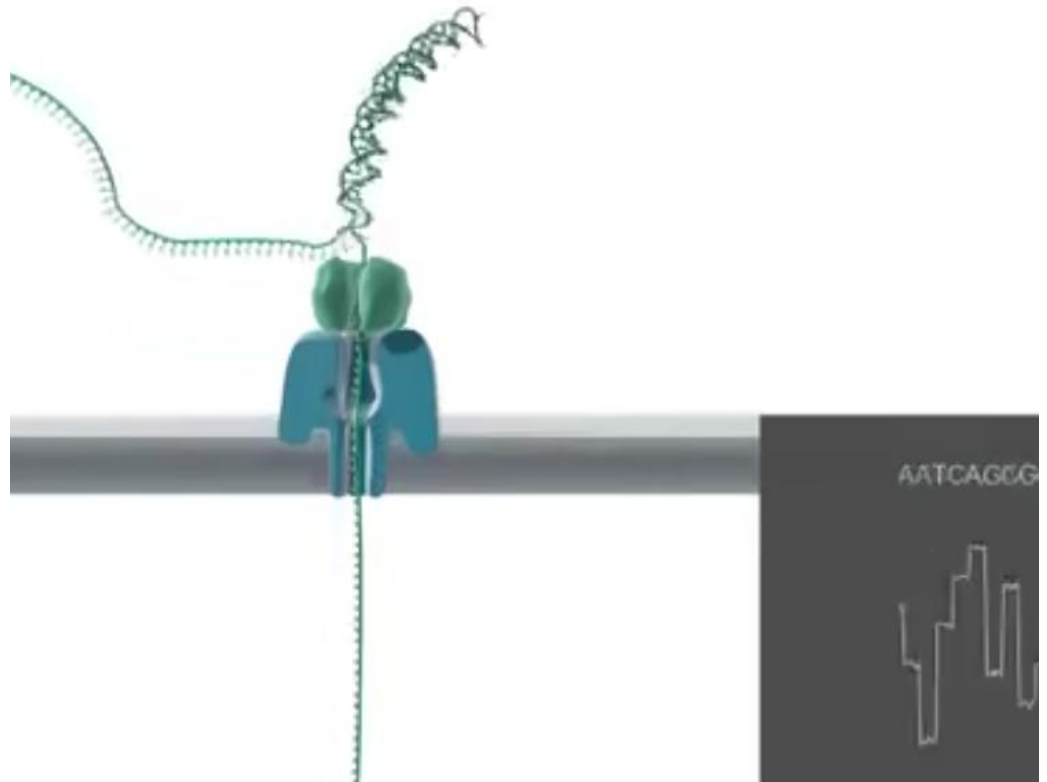
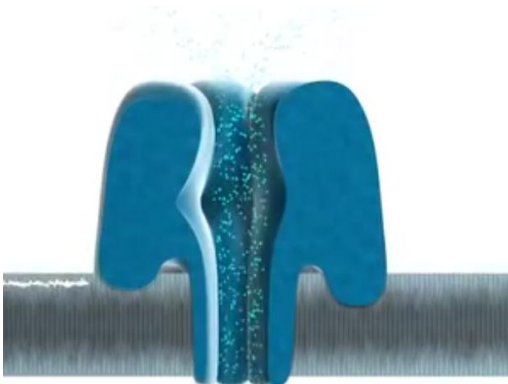
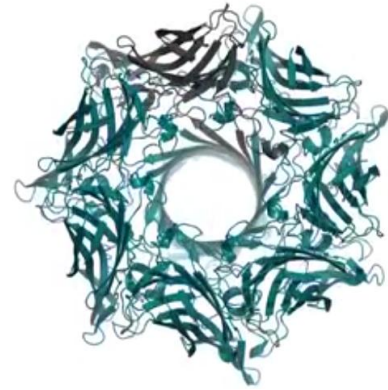
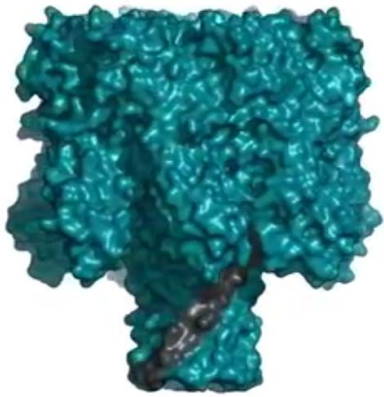
**2x300 bp**

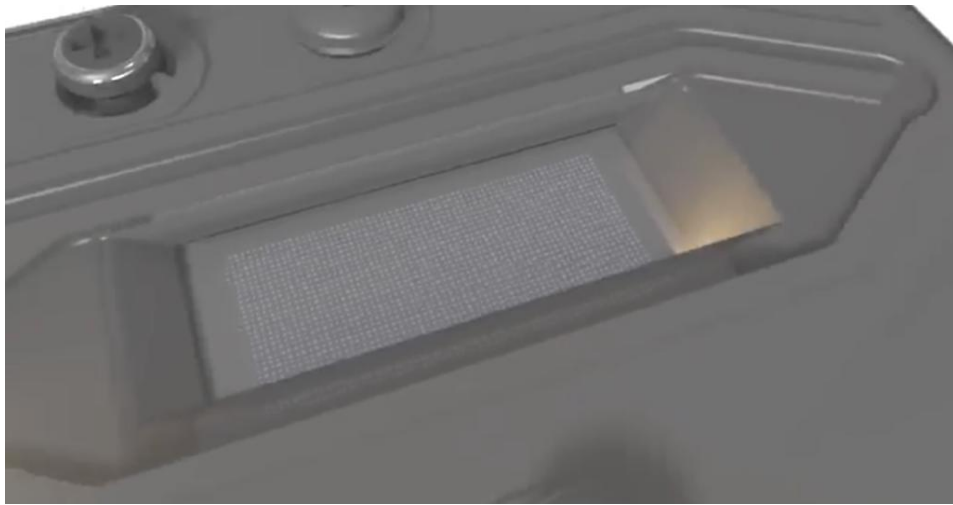
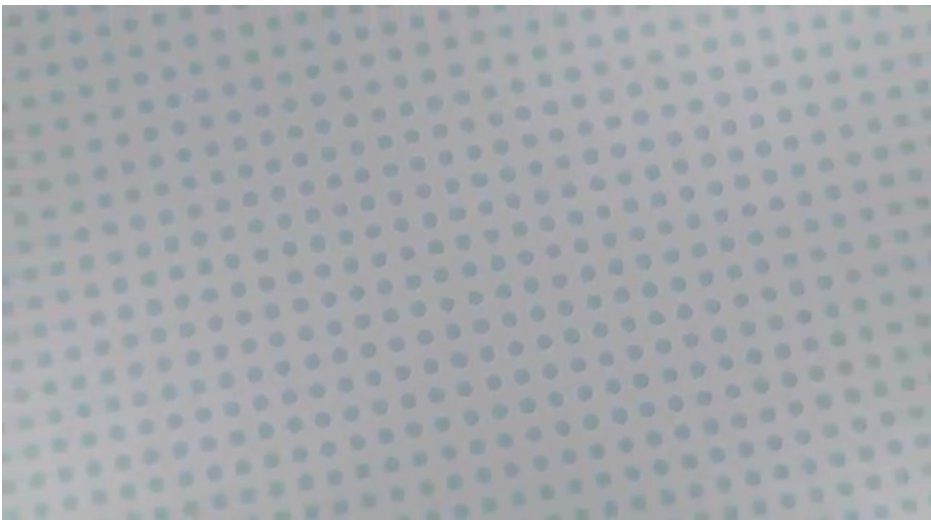






# Nanopore Sequencing





- yield : 21 Gb, read number 2.2 millions, longest read: 200Kb, run time 48h,, weight 87g