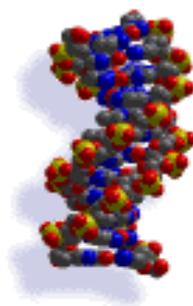




**Mend
The
Gap**



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

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



King Richard III of England
(1452–1485)



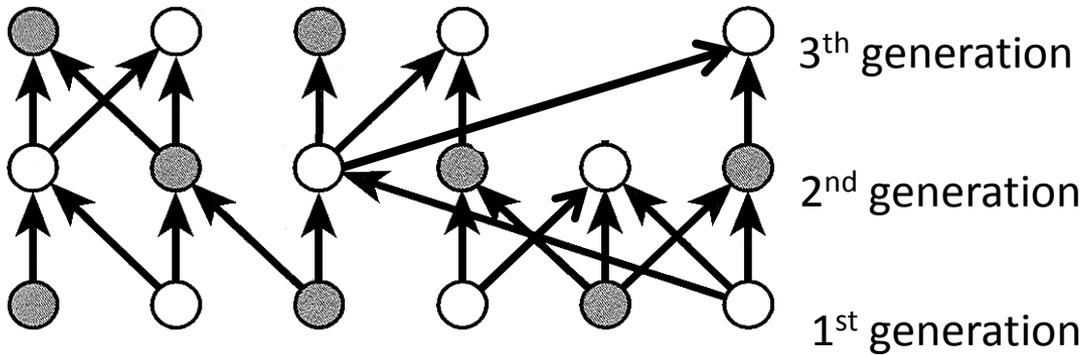
Vindija cave (Croatia)



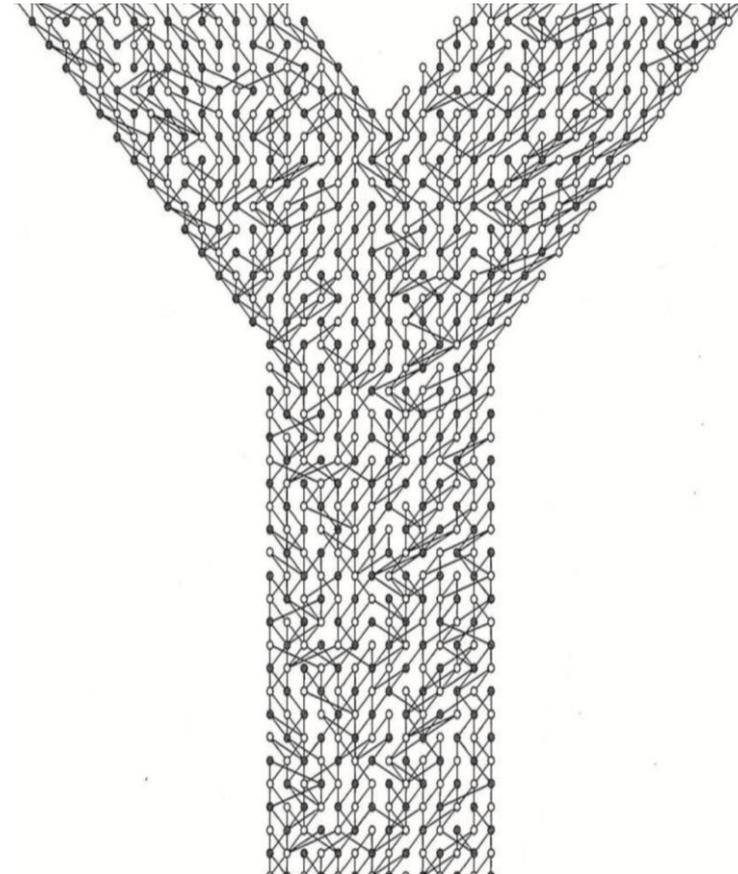
Oriental rat flea
(*Xenopsylla cheopis*)

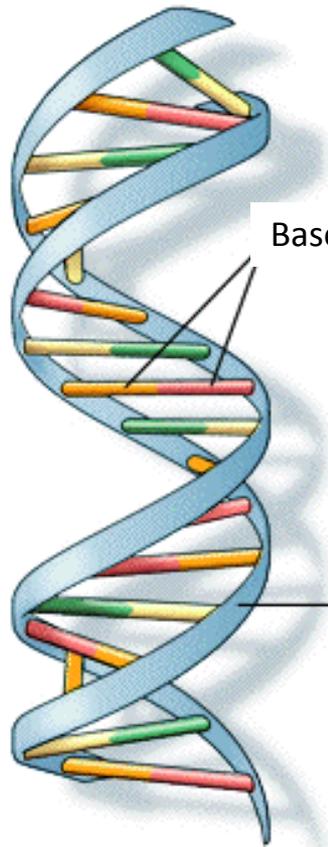
Why are DNA molecules so informative?

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!

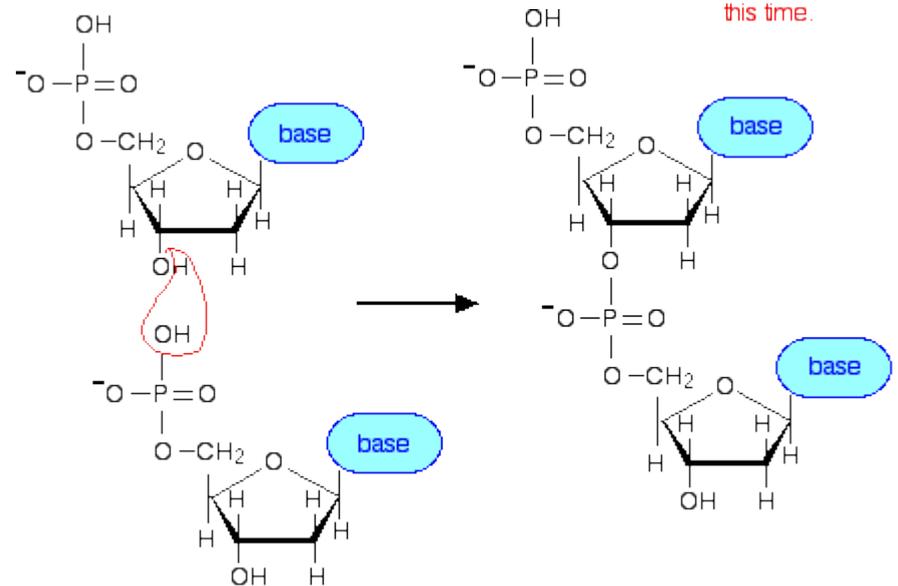
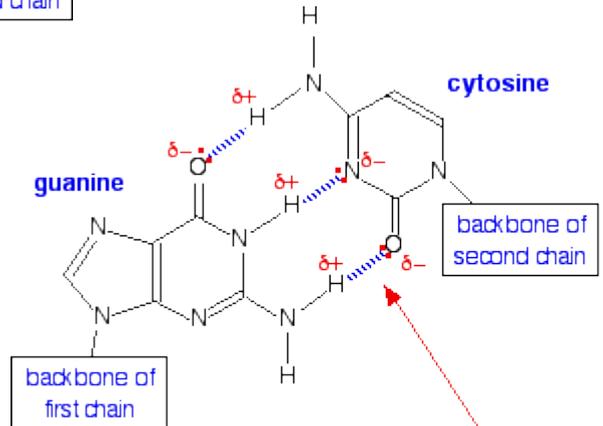
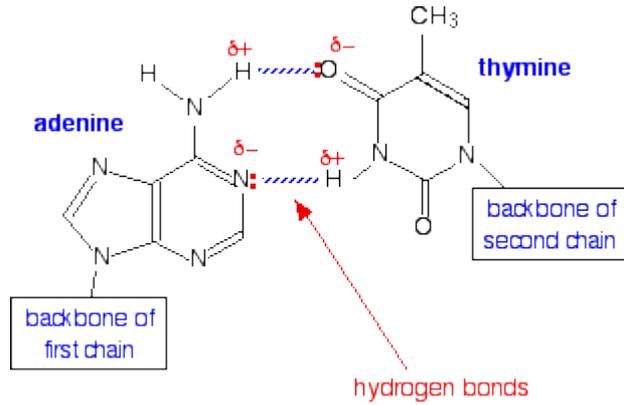




Base pairs

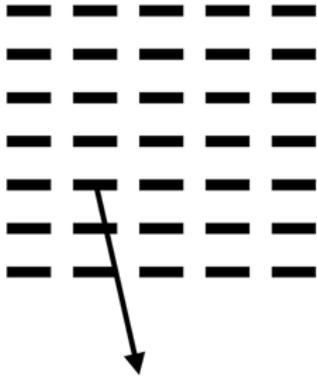


Sugar phosphate backbone



aggttatt ccaagttgc gcaatfaact ccacaagaaa ggacaaaaggaaactggag agccttgaag cattgaagaa aacttttaag aaactggatc atgatgtaaatatacatg
aaaggaatca aactattgag ggtctaagta accaagtagccatttacc aa gctcagctaa tggagtgta taggaggttg agttgttggga cgaacatcgatagaatagaacactgag
ttggaaagaa tcattgagga aatccattgaaagaatccag attcacaagg aacattacag aaagaaccaa ctcttgcaa tagaatgtgc aacaacacag tttcacagcg ggataca
cctatggcg atgggaggta atagtagtat gcaagaagct cactccatgt cttggcttcc tgataatgat caccagcaaa caatcttacc tggatgattcc agttttctc cccatagaga
gatggatggt tcgattcccc ttactcaag ctgcttcttt gagtctacga aaccagaaga tcagatatgc agcaaccgag gacaacagtt tgagcagtta gaacaacaag gaaacg
ttgggggta caacaacttg gagaggaata ttcatactct acaccgttg gtactacttt gggaatggaa gaagatcaag agaaaaagat aaaatctgaa atggaattga acaactt
caacagcaa cagcaacaac aacaacaaca acaacaagat cttcaatgt atgatccat ggctaataat aatgggtggct gctttcagat tctcatgat cagtccatgt ttgtcaat
catcatcat catcatcacc accatcatcaaaattgggtt ccagattcaa tgtttggta gacttttac aaccagcaac cgaactaactcaagacatgg ctcatgtatg tatgatgact g
tgaccaagg actagtttga gccattgcct acaatgtcta agataaatat atatatgtag acagcgagag agtgtgtgtacatgtttttg tatgtgtgag cttgtgtcat atactcaggg g
atttgtgta taagagatgt tctctaggaa ccgtatatta cctttaacaa gaatgataag tataatcaatataatattttt ttataggtg ttggccaata tattgtttga gacc agcctttct c
cggattgaa gacgttttt caagattcat caatcttc aaacaagaac gagagagtgc tctatacttt cctgaccaga atagacgccc agatattcaaaaacaagagt gtctactaag
aaactcaaga ctgaaaatga catcgtctt caagtcacga accctgcagc tatcaactct gacgtcgagg aacttgagca tgaagttgaggttacaac aacaactcca aatggcag
gaagaactaa gtgttatagt ttataattt ttaactcaa atcatgattg cttactatat aatatgagca ggagatacga accggatcca ataagattca cgaccatgga ggaatacгаа
agcaacttct cgacacteta acacatgttg tccaacggcg agaccatctc atgagcaacc atttatcttc gtatgaagca tctactatgc aaccaaacat tgggtggtcct tttgtaaatg
gggtggcta cctgaaaatg ggaccaatca aaccatttg tttgatgcat cggcccattc taaccaactc agagaattat catcggcaat gtatgaacca ttgttgcaag gaagtagc
agctcgaacaaaacaaca tgagtgaatg ccacgtgacg aaccacaatg gtgaaatgtt cctgagtgggctcaggcgt attcatctc ggctttatc gcttctatgc agcaacag
gaaggtgtg ggtcccagta tagaagagat gatgccagct cagcagagtg atattccggg agtgacggcggagacacaag ttgatcatga agtctccgac tatgaacaa aggtt
ctcagtag caataatgta ttatttttt atgtatacca ttttaaaaa tttattcata ttttttcca acaaccata aaaaagttaa ctttttttg ccatgagaag acgatgaaga gtcgtg
tttgatac tattctttt ggcaacacca ttttatagtt ttatgggcga aaagcaaatc aagttttatc ttcttaatt ttcagcctttct cggaaaaac tcggattgaa gacgttttt caag
aatcttc aaacaagaac gagagagtgc tctatacttt cctgaccaga atagacgccc agatattcaa acaaaagagt gtctactaag gattctgcag caactcaaga ctgaaaat
catcgtcttcaagtcacga accctgcagc tatcaactct gacgtcgagg aacttgagca tgaagttgtaggttacaac aacaactcca aatggcagaa gaagaactaa gtgttata
ttaactcaa atcatgattg cttactatat aatatgagca ggagatacga accggatccaataagattca cgaccatgga ggaatacгаа gttctgaga agcaacttct cgacactct
ccaacggcg agaccatctc atgagcaacc atttatcttc gtatgaagca tctactatgc aaccaaacat tgggtggtcct tttgtaaatg acgtcgttga aggttggcta cctgaaaatg
aaccatttg tttgatgcat cggcccattc taaccaactcagagaattat catcggcaat gtatgaacca ttgttgcaag gaagtagctc aagctcgaacaaaacaaca tgagtgaat
aaccacaatg gtgaaatgtt cctgagtgg gctcaggcgt attcatctc ggctttatc gcttctatgc agcaacagca tgaaggtgtg ggtcccagta tagaagagat gatgcca
cggaaaaac tcggattgaa gacgttttt caagattcat caatcttct aaacaagaac gagagagtgc tctatacttt cctgaccaga atagacgccc agatattcaaaaacaagag
gtctactaag gattctgcag caactcaaga ctgaaaatga catcgtctt caagtcacga accctgcagc tatcaactct gacgtcgagg aacttgagca agttttagaggttacaaca
atggcagaa gaagaactaa gtgttatagt ttataattt ttaactcaa atcatgattg cttactatat aatatgagca ggagatacga accggatcca ataagattca cgaccatgga
gttctgaga agcaacttct cgacacteta acacatgttg tccaacggcg agaccatctc atgagcaacc atttatcttc gtatgaagca tctactatgc aaccaaacat tgggtggtcct
acgtcgttga aggttggcta cctgaaaatg ggaccaatca aaccatttg tttgatgcat cggcccattc taaccaactc agagaattat catcggcaat gtatgaacca ttgttgca
gaagtagctc aagctcgaac aaaacaaca tgagtgaatg ccacgtgacg aaccacaatg gtgaaatgtt cctgagtg gctcaggcgt attcatctc ggctttatc gcttctatg
agcaacagca tgaaggtgtg ggtcccagta tagaagagat gatgccagct cagcagagtg atattccggg agtgacggcg agacacaag ttgatcatga agtctccgac tatga
gggttctca actcagtagc aataatgta ttatttttt atgtatacca ttttaaaaa tttattcata ttttttccaacaaccata aaaaagttaa ctttttttg ccatgagaag acgatga

1st generation of DNA sequencers
(started 1987)

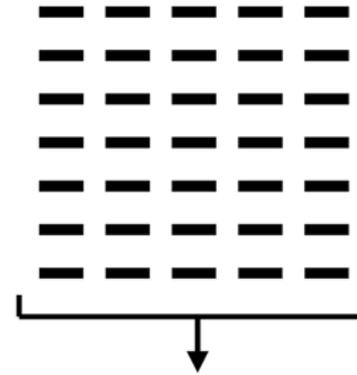


one fragment (ca 1000 bp long) per run



(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

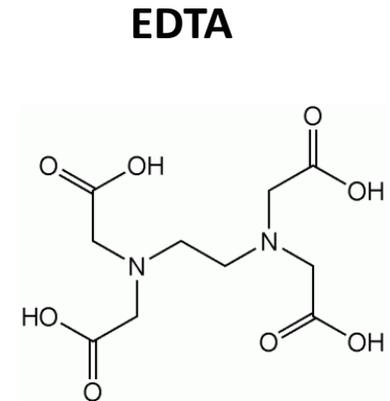
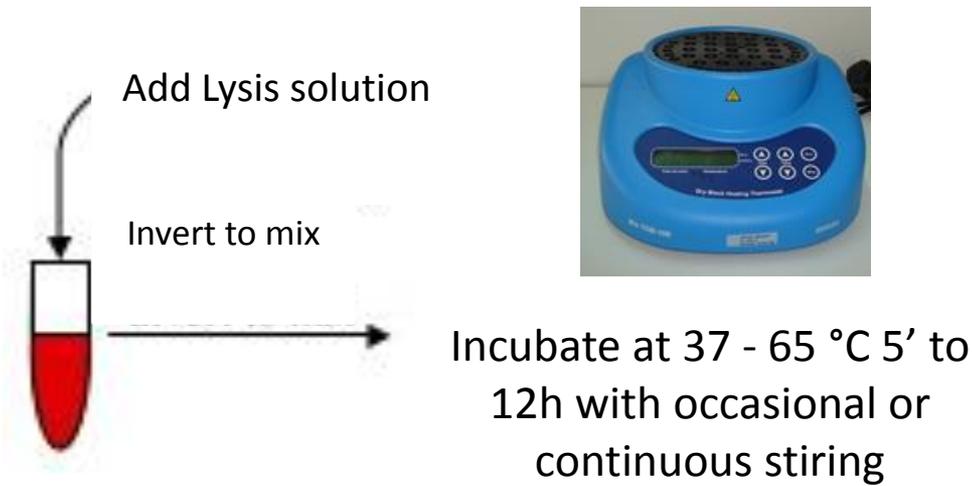
DNA isolation

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

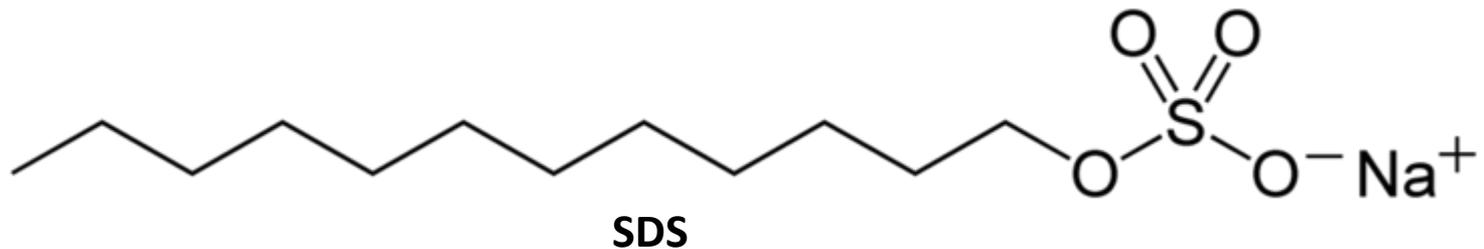
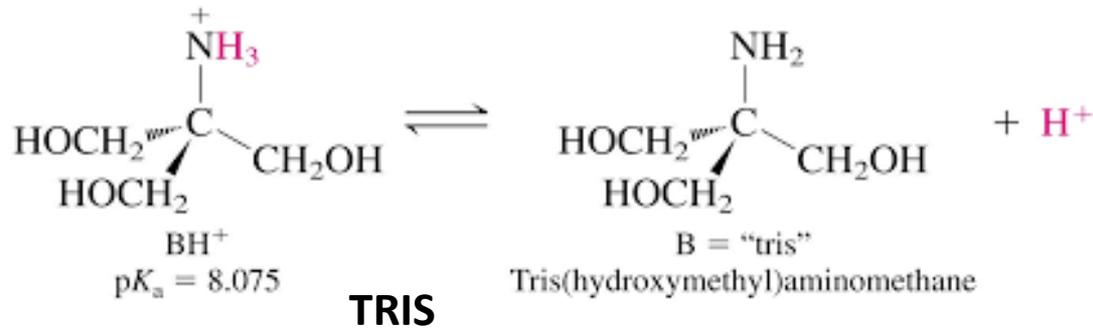


Archaeological sample need specific pretreatment because there is high possibility of **contamination** from microenvironment of the fossil, but also present day humans !!!



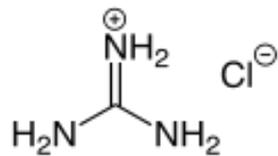
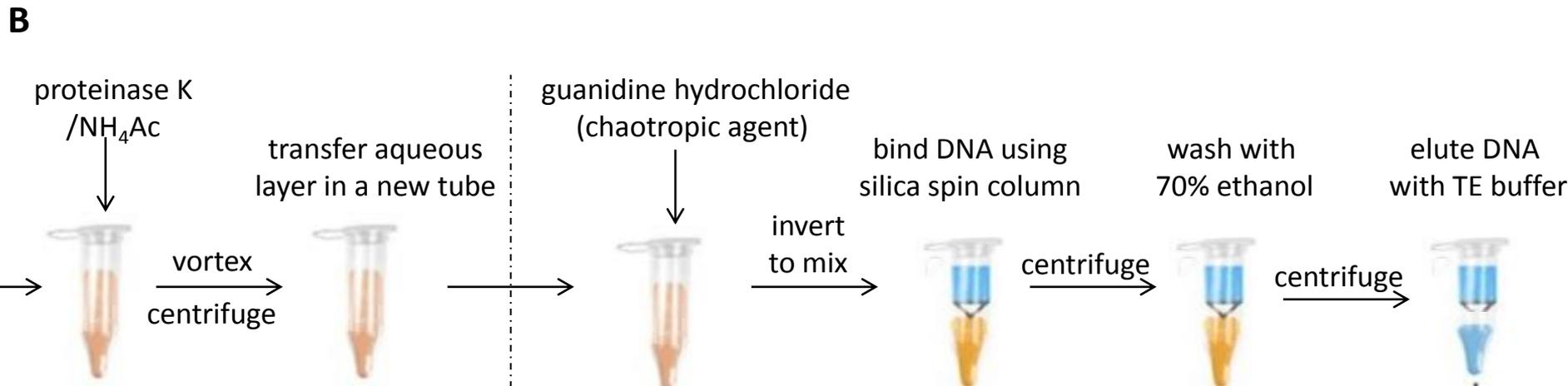
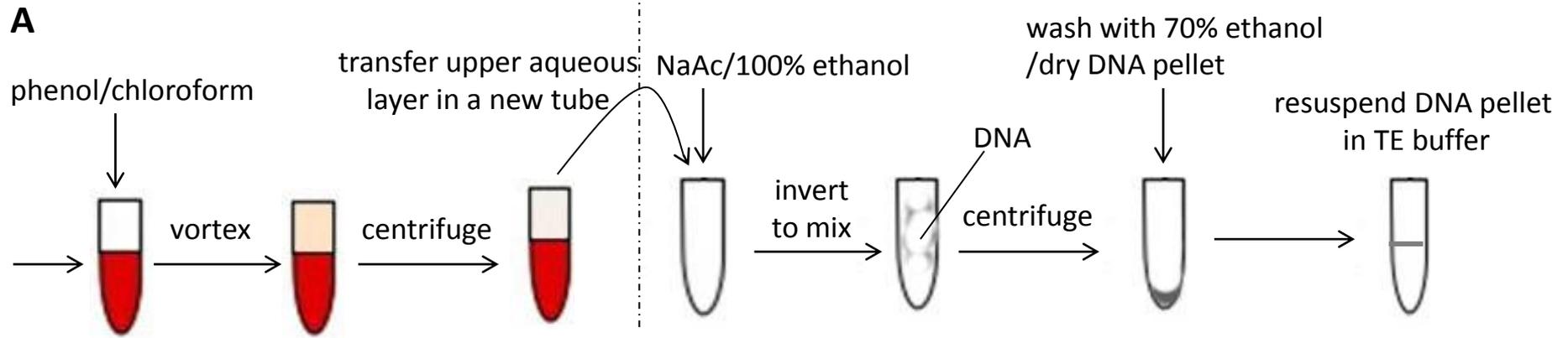


General solution of lysis buffer contains: TRIS-HCl, EDTA, SDS or CTAB



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

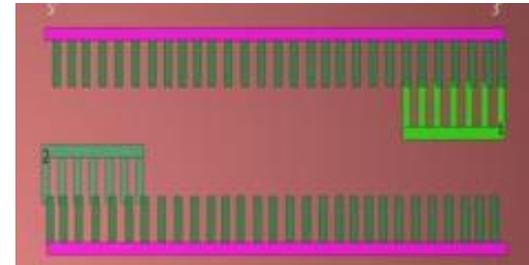
= amplifying a specific DNA fragment *in vitro* up to billion copies
(Kary B. Mullis - Nobel Prize in Chemistry 1993)



Thermal cycler

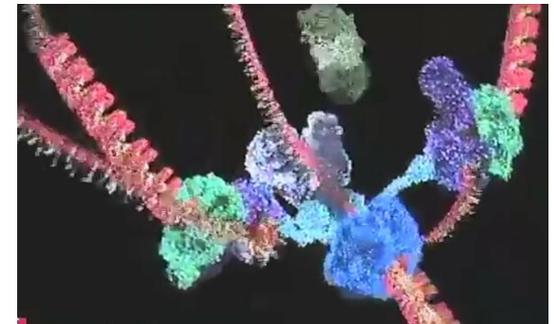
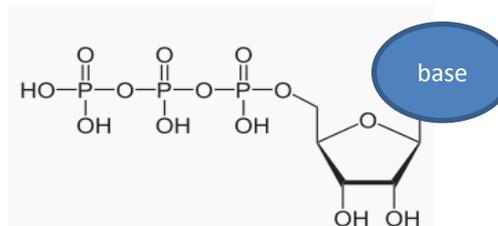


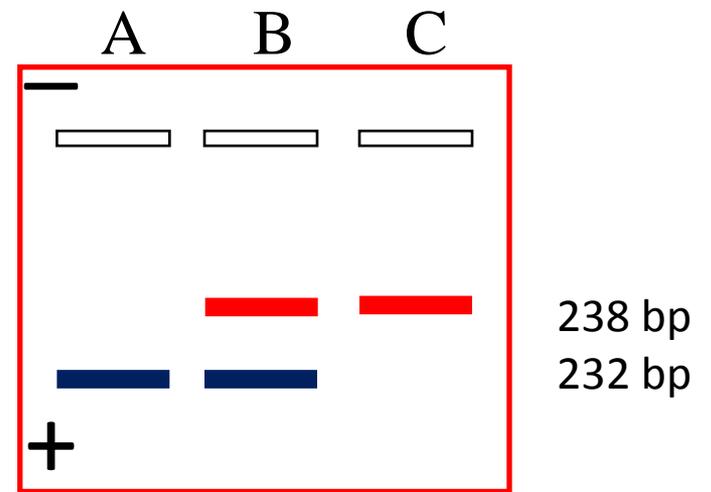
Thermal cycler device is able to specifically increase and decrease temperature



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



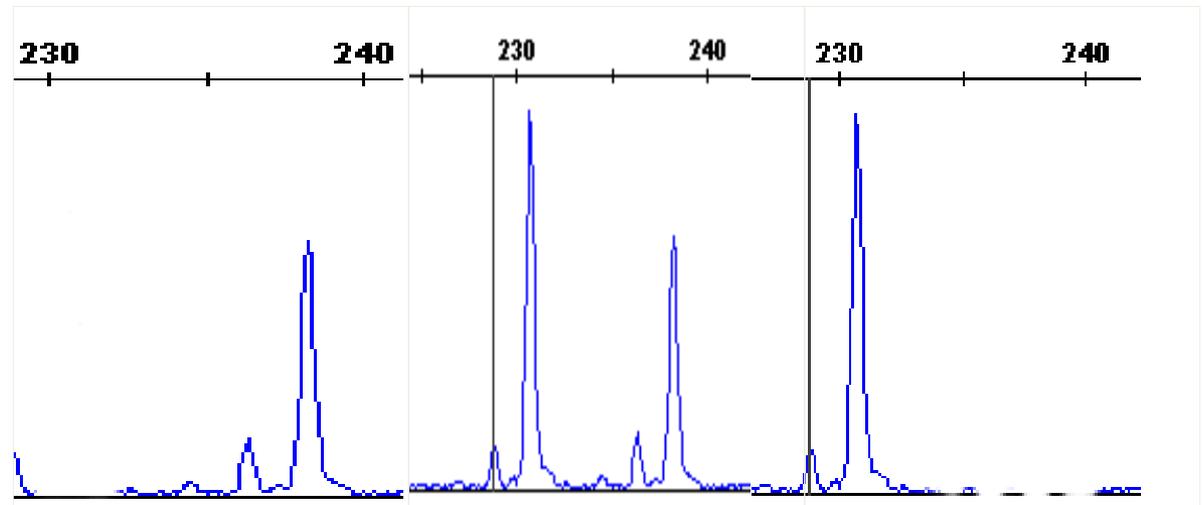


direct detection by electrophoresis (e.g. microsatellites)

→ PCR products



A B C

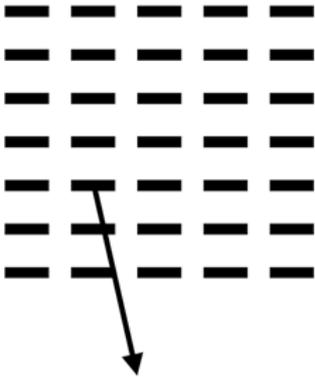


sequencing
(1st generation)



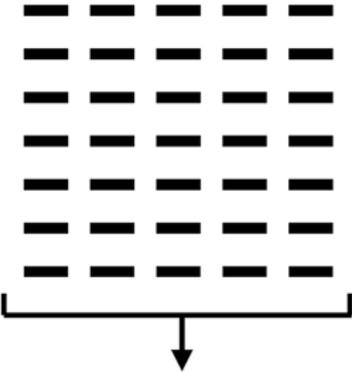
NGS sequencing

1st generation of DNA sequencers
(started 1987)



one fragment (ca 1000 bp long) per run

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

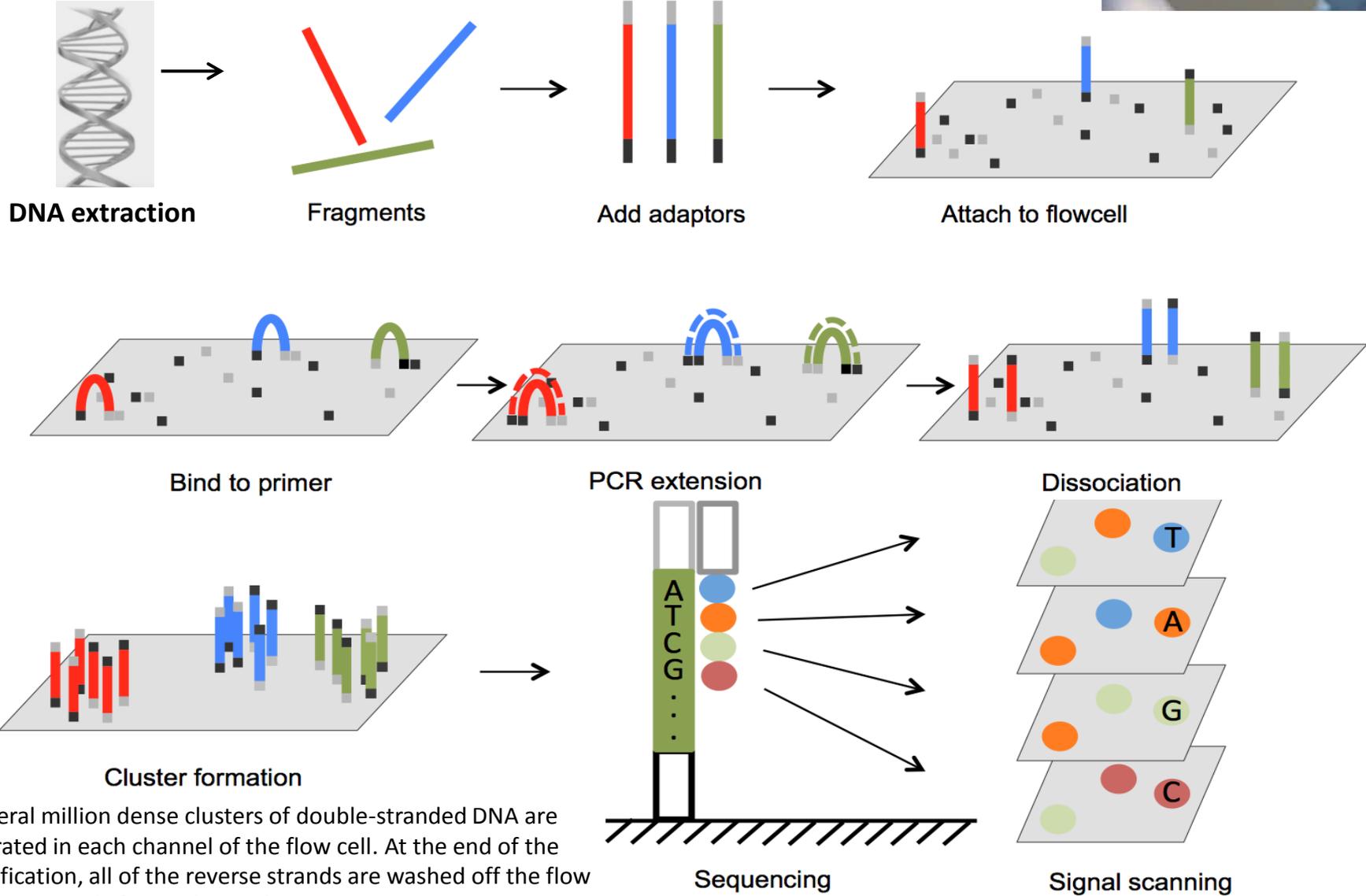


millions of fragments per run

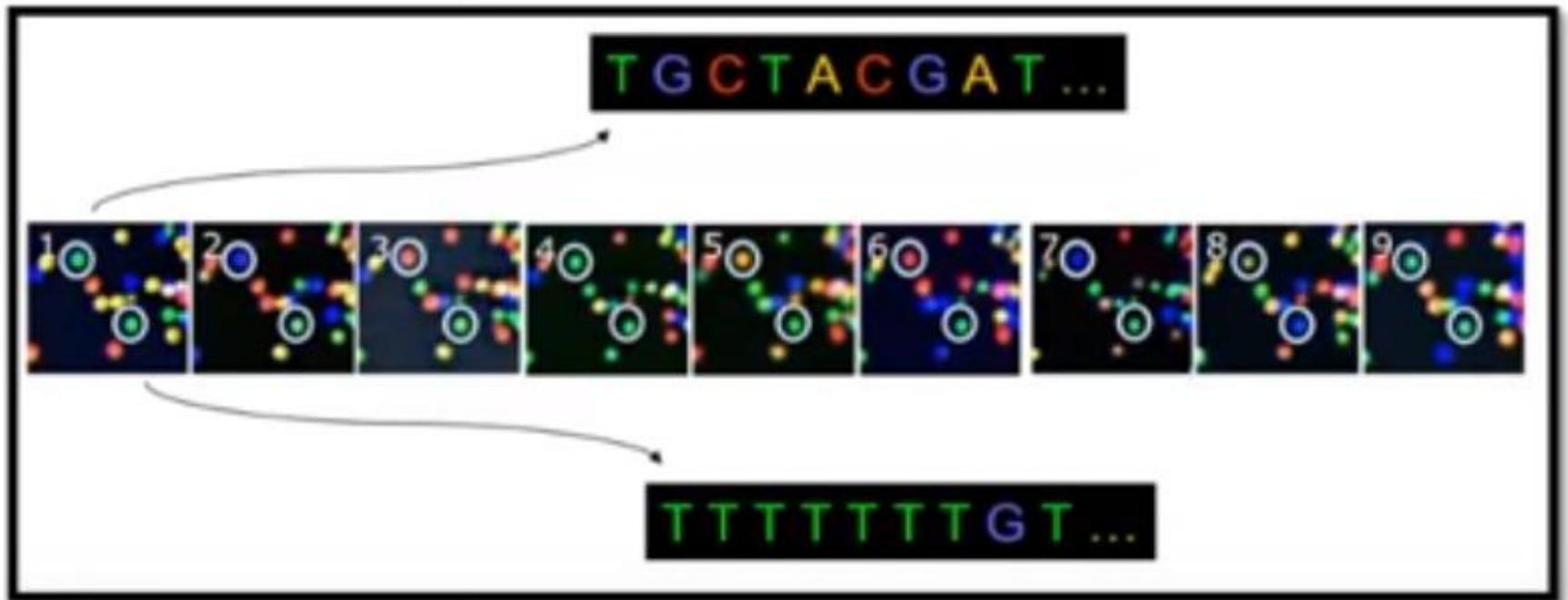
NGS = **N**ext **G**eneration **S**equencing

- 1) Sequencing by synthesis
- 2) Nanopore sequencing

Sequencing by synthesis



* Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell. At the end of the amplification, all of the reverse strands are washed off the flow cell, leaving only forward strands.



Illumina NGS sequencers (sequencing by synthesis)

MiSeq Series

MAX OUTPUT

15 Gb

MAX READ NUMBER

25 million

MAX READ LENGTH

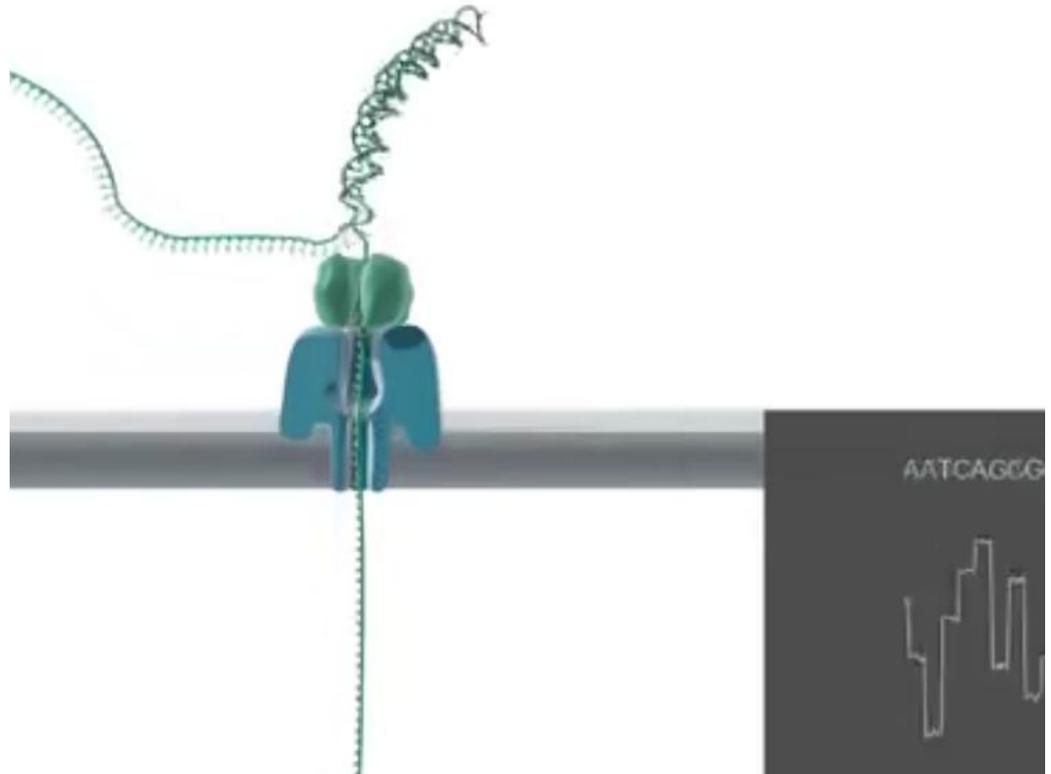
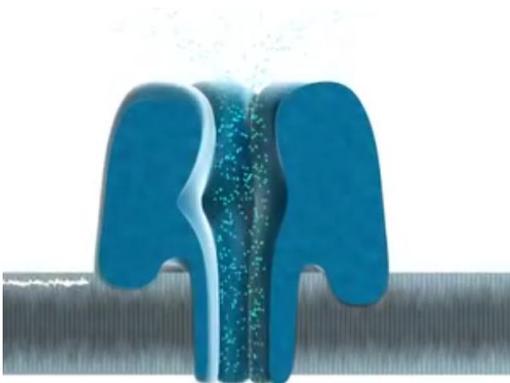
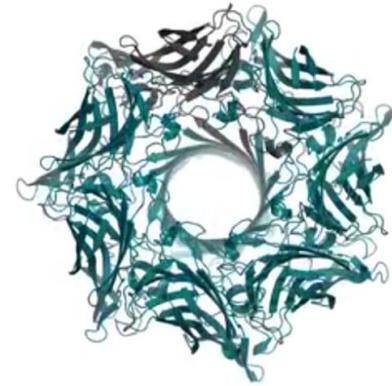
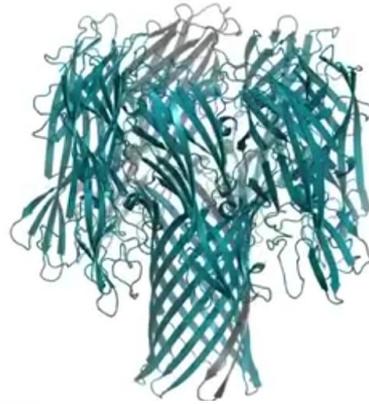
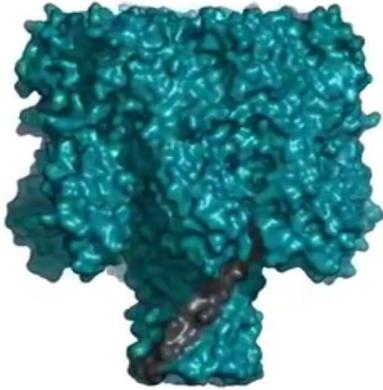
2x300 bp

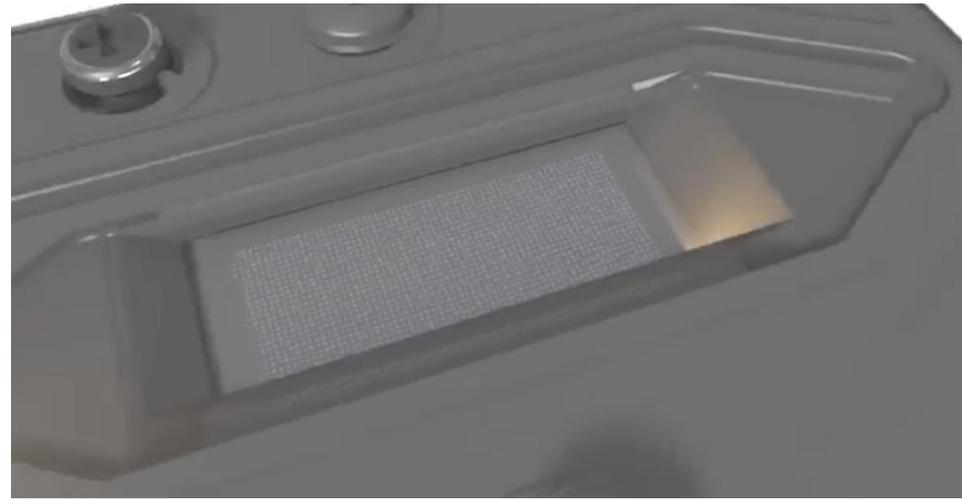
reference

```
CCTGGGAAGGGTCAGTGCAGTGAGGGCCCTGGGAAGGGTCAGTGC TGTGAGGGCCCTGGGAAGGGTCAGTGCAGTGAGGGCCCTG
CCCTGGGAAGGGTCAGTGCAGTGAGGGCCCTGG AGGGTCAGTGC TGTGAGGGCCCTGGGAAGGGTCA GAGTGAGGGCCCTG
ccccgggacgggtcagtgcgagtgagggccctggagggtcagtgctgtgagggccctgggaaggggtcagtgca gggggccctg
CCCTGGGAAGGGTCAGTGC AGTGAGGGCCCTGGGAAGGGTCAGTGC TGTGAGG GGAAGGGTCAGTGCAGTGAGGGCCCTG
ccctgggaaggggtcagtgcgagtgagggtcagggaaggggtcagtgctgtgagggccctgggaaggggt tgcgagtgagggccctg
CCCTGGGAAGGGTCAGTGCAGTGATG ccctgggaaggggtcagtgctgtgagggccctggga AGTGCGAGTGAGGGCCCTG
ccctgggaaggggtcagtgcgagtgaggg GAAAGGGTCAGTGC TGTGAGGGCACTGGGTAGGGT GAGTGAGGGCCCTG
ccctgggaaggggtcagtgcgagtgagggc gaaggggtcagtgctgtgagggccctgggaaggggtc AGTGAGGGCCCTG
cgtgggaaggggtcagtgcgagtgagggc gaaggggtcagtgctgtgagggccctgggaaggggtc agtgagggcactg
cgtgggaaggtgctcagttcgagtgagggtccct aaggggtcagtgctgtgagggccctgggaaggggtc AGGGCCCTG
ccctgggaaggggtcagtgcgagtgagggccctg aaggggtcagtgctgtgagggccctgggaaggggtca aagggcactg
ccctgggaaggggtcagtgcgagtgagggccctgggaa GTCAAGTGC TGTGAGGGCCCTGGGAAGGGTCAGTGC gggccctg
GGAAGGGTCAGTGCAGTGAGGGCCCTGGGAAGGGT TGTGAGGGCCCTGGGAAGGGTCAGTGCAGTGCAGT TGTG
CCCTGGGAAGGGTCAGTGCAGTGAGGGCCCTGGGAAGGGTC TGTgagggccctgggaaggggtcagtgcgagtgagggccctg
CC GAAAGGGTCAGTGCAGTGAGGGCCCTGGGAAGGGTC gtgagggccctgggaaggggtcagtgcgagtgaggg
ccc gaaggggtcagtgcgagtgagggccctgggaaggggtc TGAGGGCCCTGGGAAGGGTCAGTGCAGTGAGGGC
ccctg gaaggggtcagtgcgagtgagggccctgggaatgggtc GAGGGCCCTGGGAAGGGTCAGTGCAGTGCGGGCCCTG
ccgggggaaggggtcagtgcgagtgagggccctgggaaggggtcagtgcgcgctgggaaggggtcagtgcgagtgagggccctg
CCCTGGGAAGGGTCAGTGCAGT GAGGGCCCTGGGAAGGGTCAGTGC TGTGAGGGC ggaaggggtcagtgcgagtgagggccctg
ccctgggaa gatcagtgcgagtgagggccctgggaaggggtcagtgctgtgagggccctgggaaggggtcagtgcgagtgagggcc
ccctgggaa acagtgcgagtgagggccctgggaaggggtcagtgctgtgagggccctgggaaggggtca GGCCCTG
ccctgggaa cagtgcgagtgagggccctgggaaggggtcagtgctgt CCTGGGAAGGGTCAGTGGGAGTGAGGG
CCCTGGGAAGGGTCAGTGCAGT gggccattgggaaggggtcagtgctgtgagggccctgggaaggggtcagtgcgagtgagggccctg
ccctgggaaggggtcagtcgagtgagggccctgggaaggggtcagtgctgtga gaaggggtcagtgcgagtgagggccctg
ccctgggaagg cagtgcgagtgagggccctgggaaggggtcagtgctgt AAGGGTCAGTGCAGTGAGGGCCCTG
ccctgggaagg tgcgagtgagggccctgggaaggggtcagtgctgtg aaggggtcagtgcgagtgagggccctg
CCCTGGGAAGGGTTCAGTGCAGTGAG CCCTGGGAAGGGTCAGTGC TGTGAGGGCCCTGGGAAGGGTCAGTGCAGTGAGGGCCCTG
ccctgggaaggggtcagtcgagtgagggccctgggaaggggtcagggc TGTGGGGCCCTGGGAAGGGA GGCCCTG
```

NGS data

Nanopore Sequencing





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