# INTRODUCTION TO DNA SEQUENCING

## (sanger & next generation)

18th June, 2021

Dr. Obishakin Emmanuel Tumininu (DVM, PhD)
Head, Biotech center, NVRI, Vom

AFRICA CENTRE OF EXCELLENCE IN PHYTOMEDICINE RESEARCH

# My opening words

1. Not your fault that you do not know
2. Be encouraged to seek for knowledge
3. It may be expensive and difficult but try to follow it
4. It is rewarding, put you on global knowledge stream !

# Goals of this lecture

1. Overview of DNA, RNA and PCR basics
2. To understand the principles of sequencing
3. Types of sequencing and their advantages
4. Introduction to sequence analysis

# What to do to learn

1. Be attentive
2. Read extra, browse, videos, make google your friend
3. Keep learning, we are all students
4. Practice, practice and practice !

# Review notes on the nucleic acids

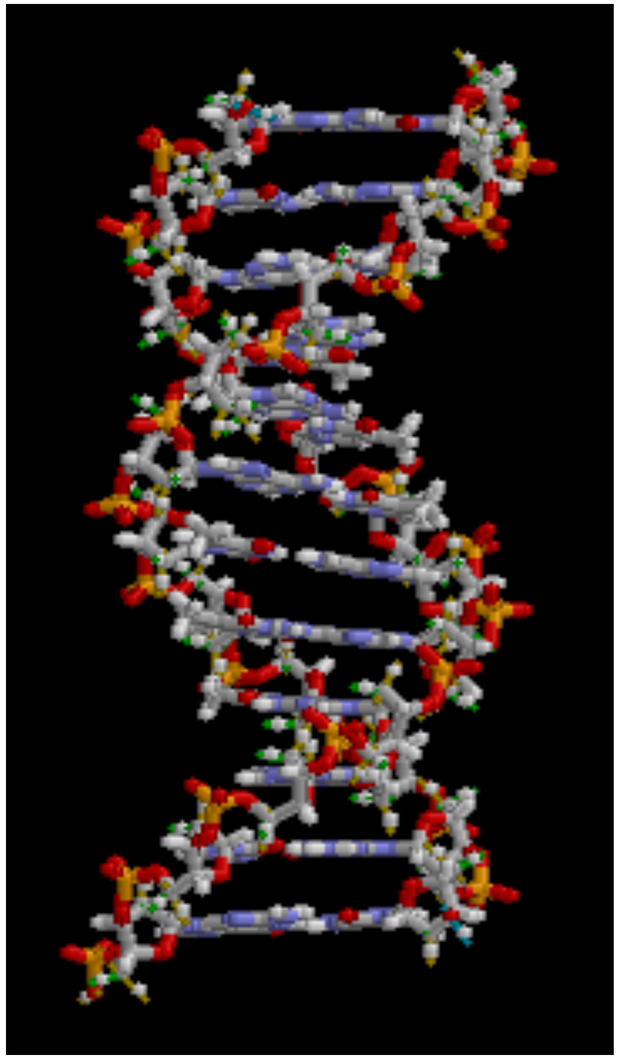
1. The Deoxyribonucleic acid

2. The Ribonucleic acid

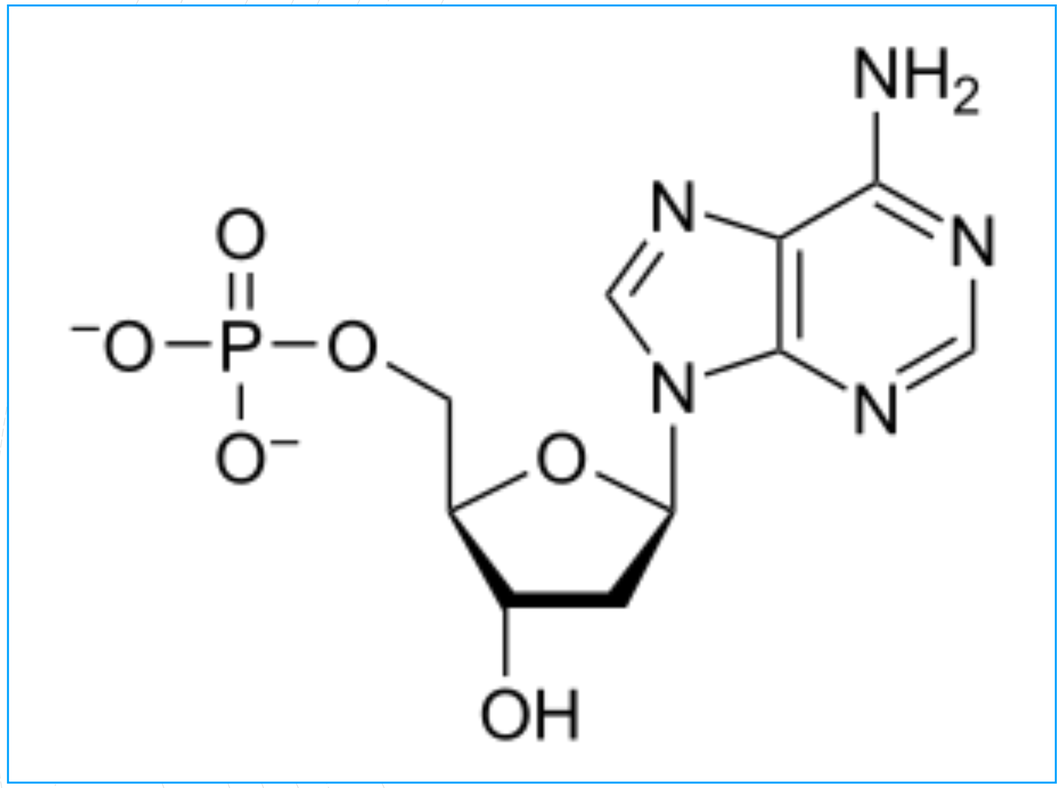
# What is DNA

- **Deoxyribonucleic acid** ( **DNA**) is a complex molecule
- composed of two polynucleotide chains
- The chains coil around each other to form a double helix
- It contains genetic instructions for the development, functioning, growth and reproduction of all known organisms and many viruses.

# What is DNA



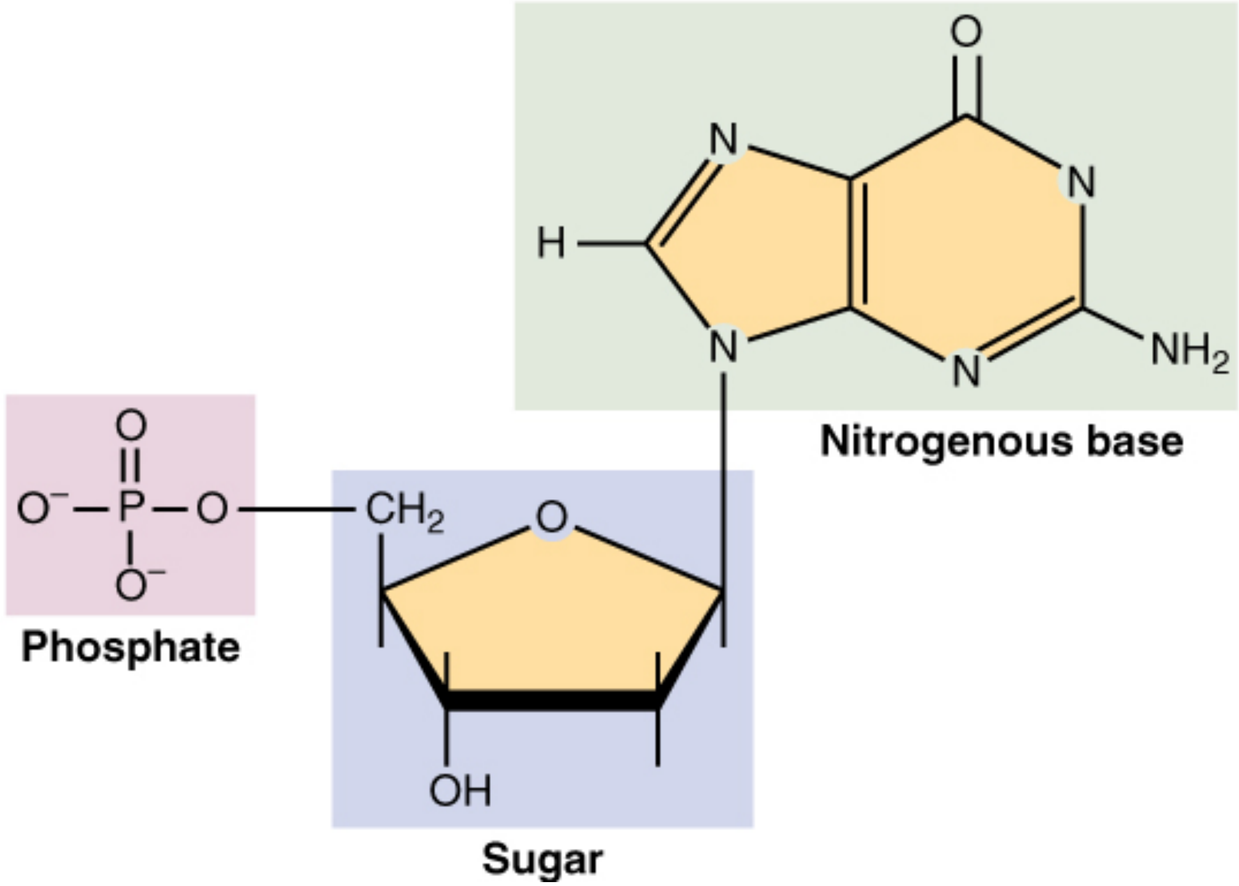
# Components of DNA




- Monomeric units called **Nucleotides**

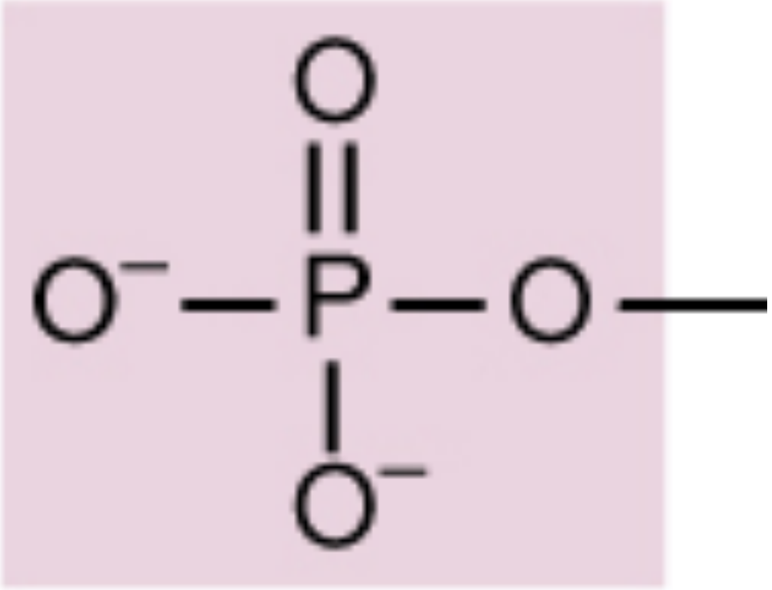
# Components of a nucleotide

- A phosphate group
- A deoxyribose sugar
- A nitrogenous base



# Phosphate group

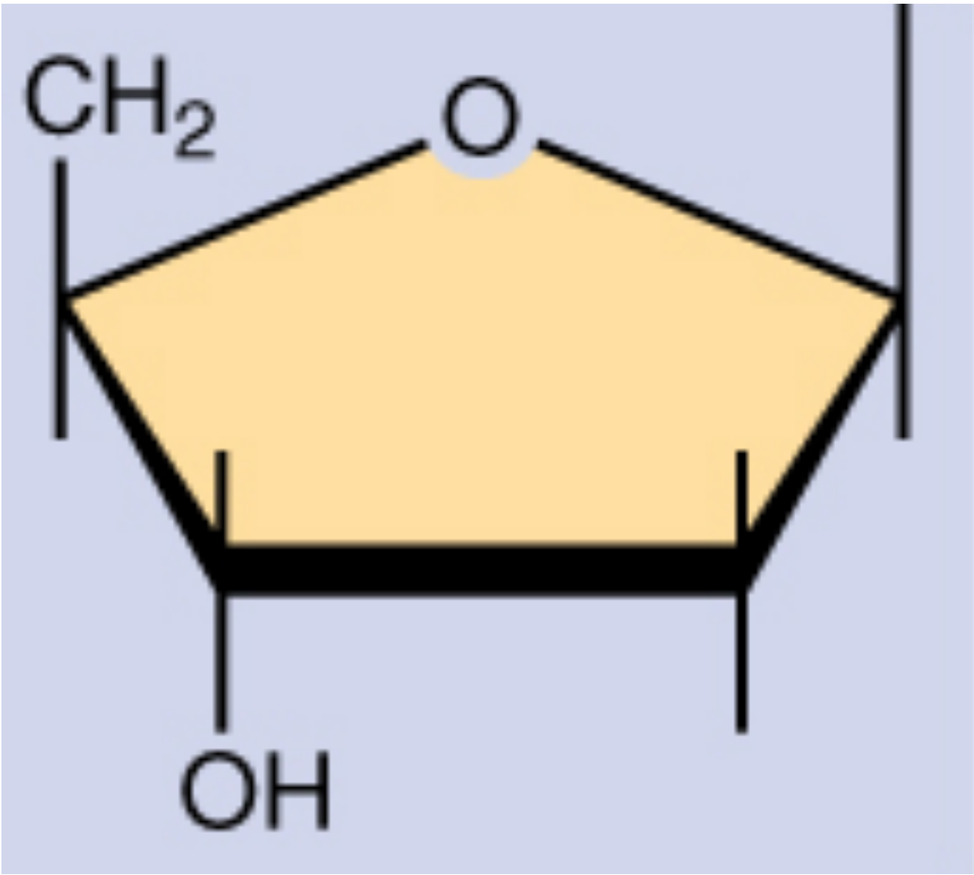
- Strongly negatively charged (anion)
- Gives the DNA its overall negative charge
- Makes the DNA acidic
- Phosphorus “expanded octet”
- can make 5 covalent bonds



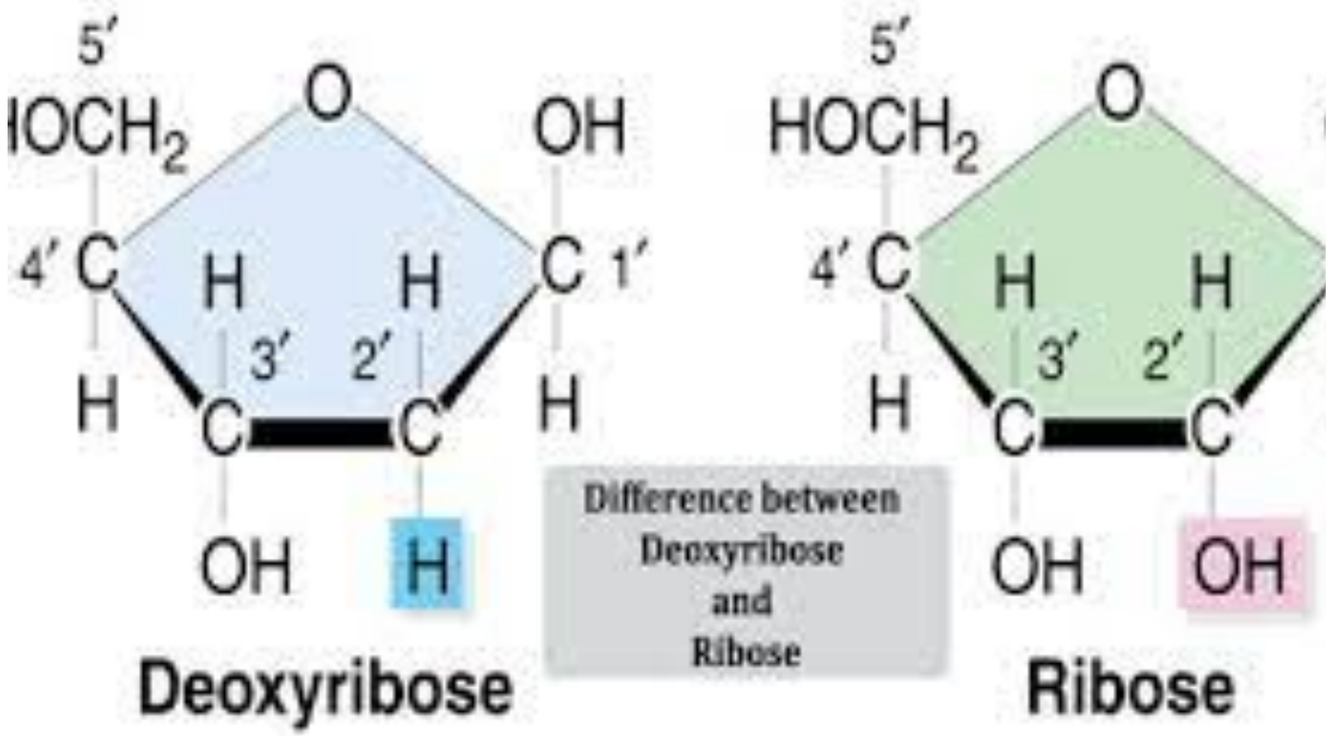

Phosphate

# Components of a nucleotide

- A ribose monosaccharide
- 5 carbon
- 2-deoxyribose sugar
- Ribose with 1 less oxygen

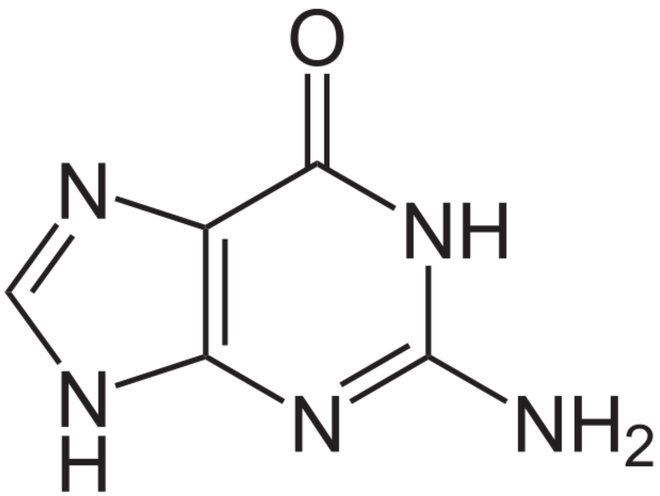


# Ribose vs Deoxyribose

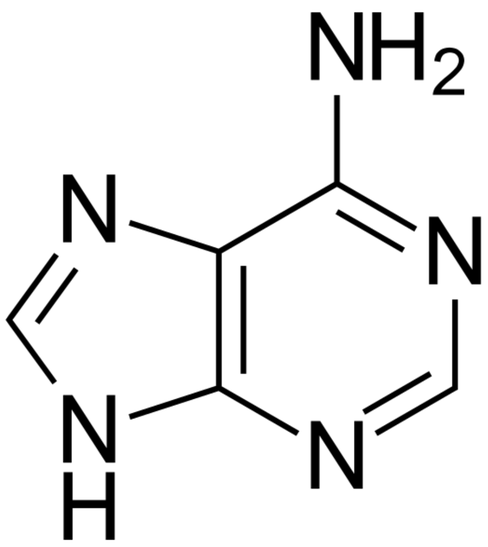



- A ribose monosaccharide
- 5 carbon
- 2-deoxyribose sugar
- Ribose with 1 less oxygen

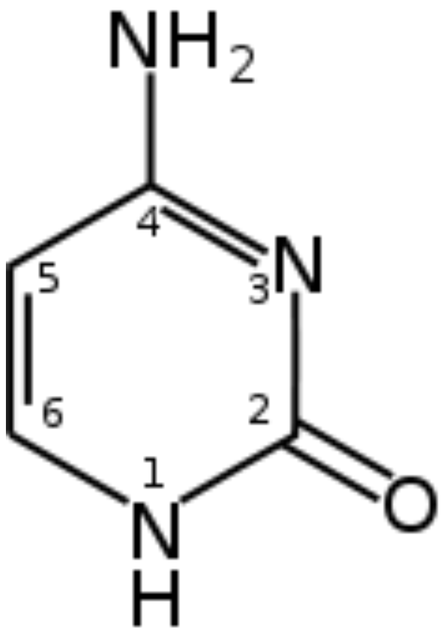
# Nitrogenous bases



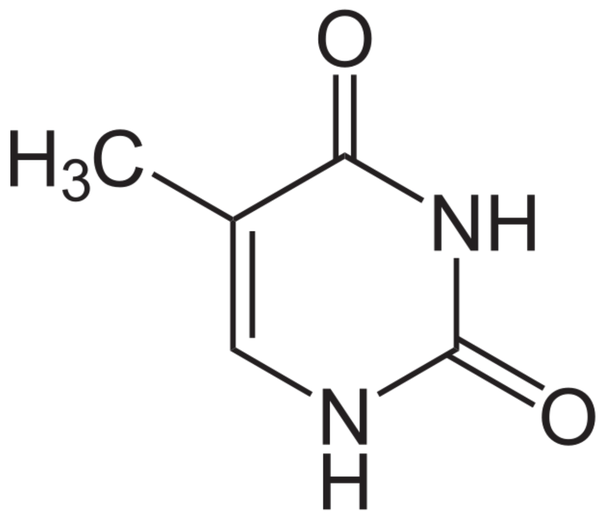

Guanine




Adenine




Cytosine



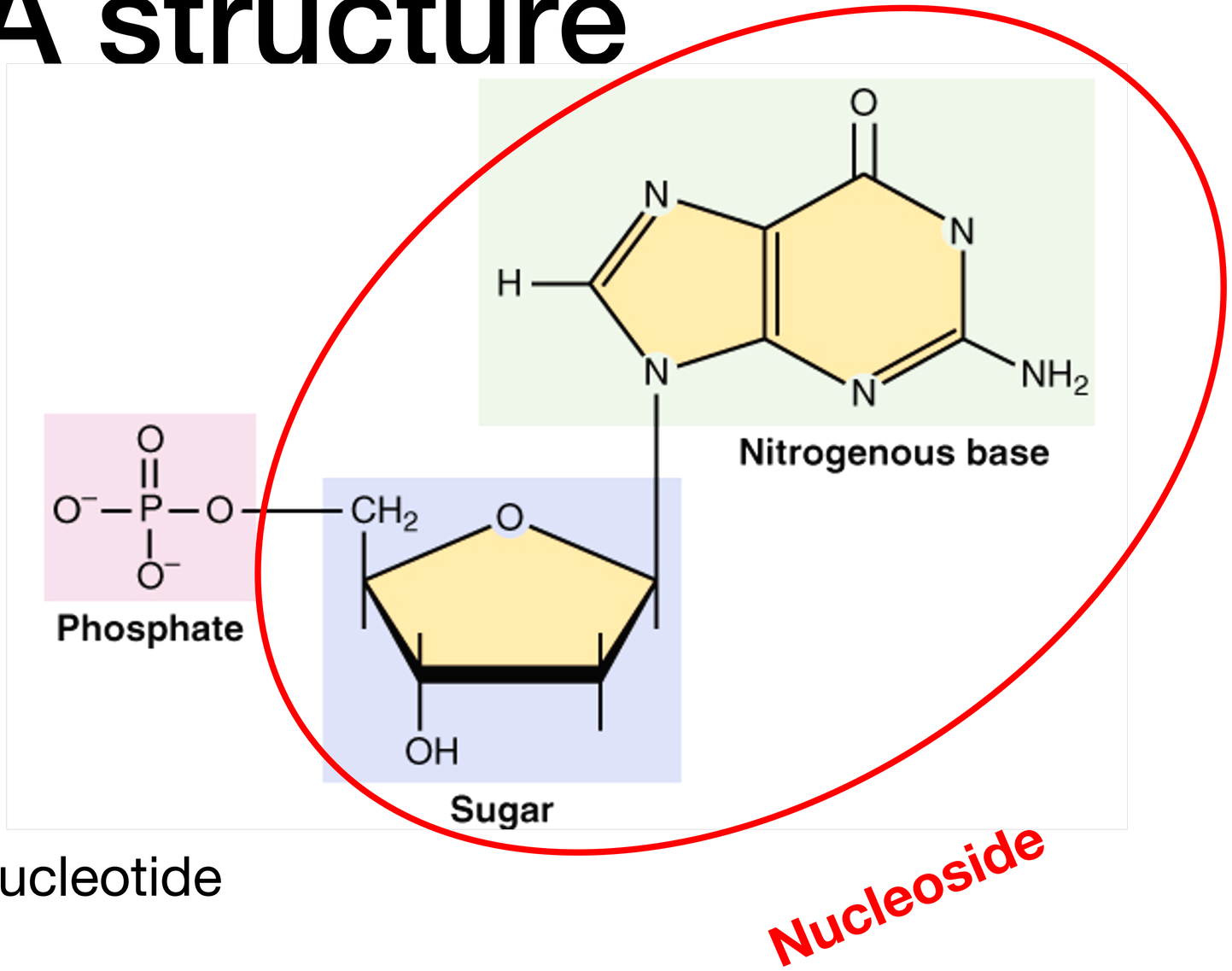

Thymine

**Purines**

**Pyrimidines**

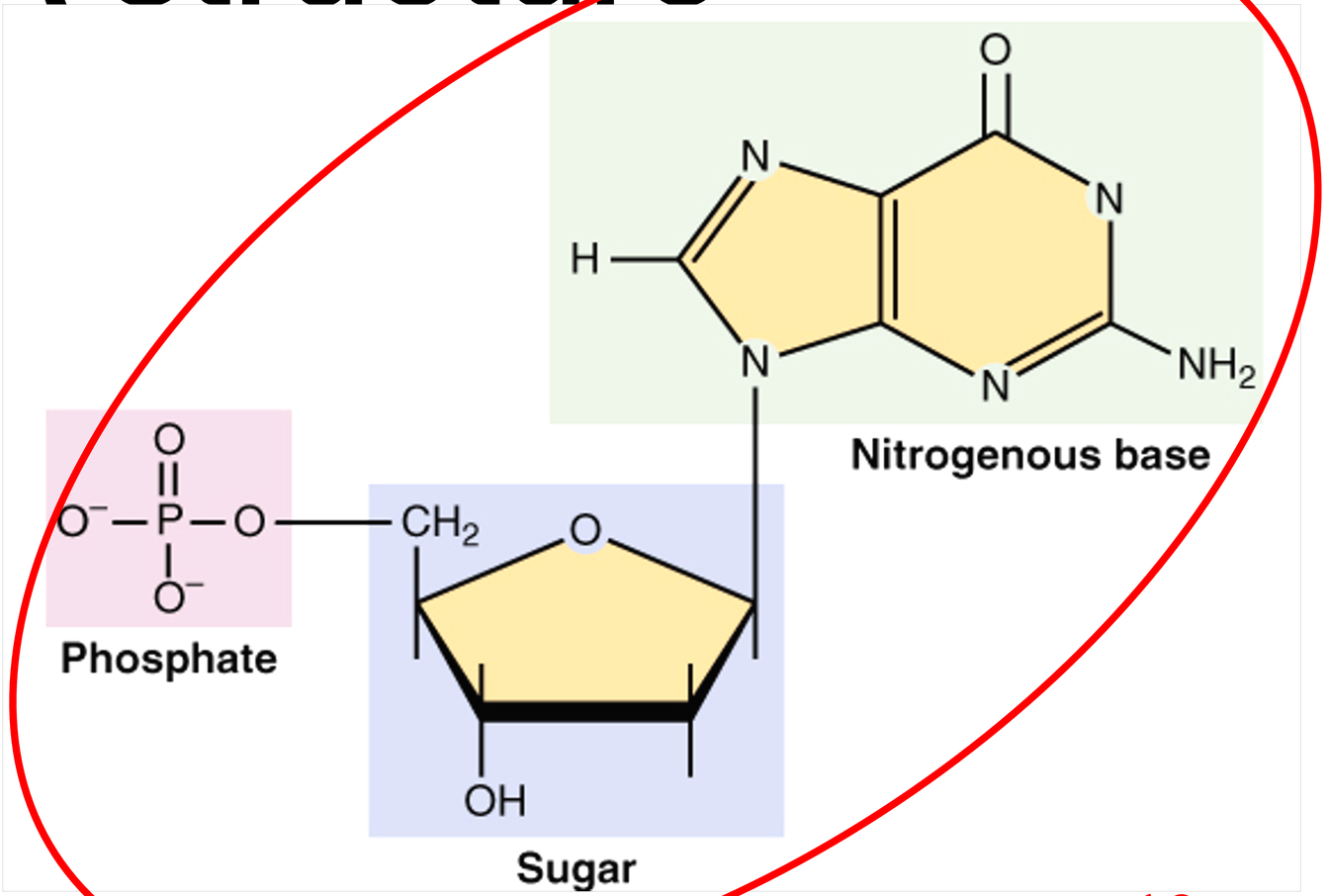
# DNA structure

- A phosphate group
- A deoxyribose sugar
- A nitrogenous base
- Sugar +base=nucleoside
- Sugar+ base + phosphate= nucleotide

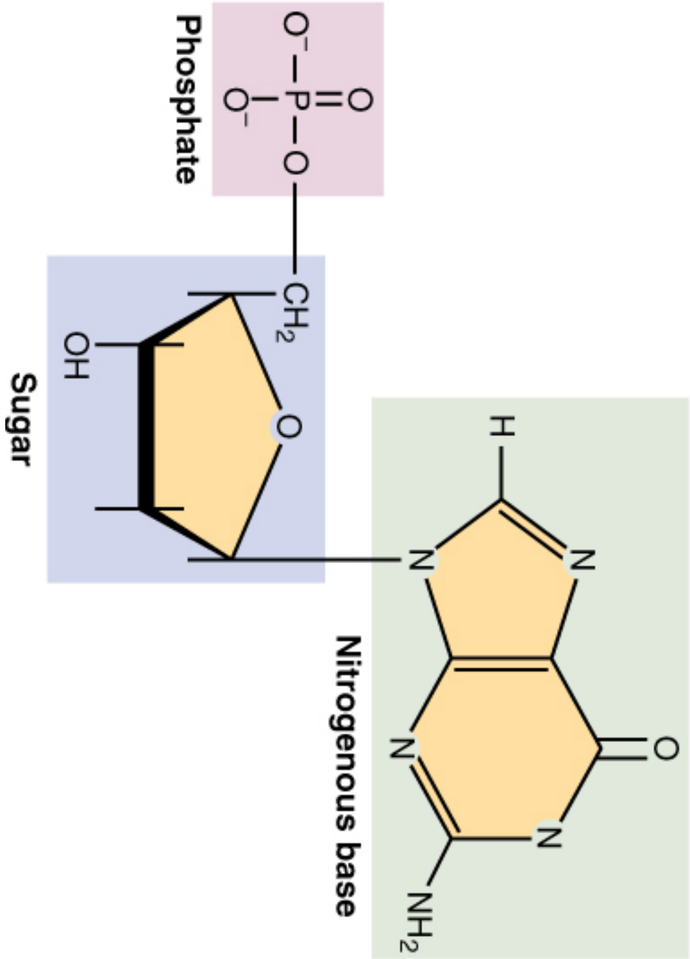


# DNA structure

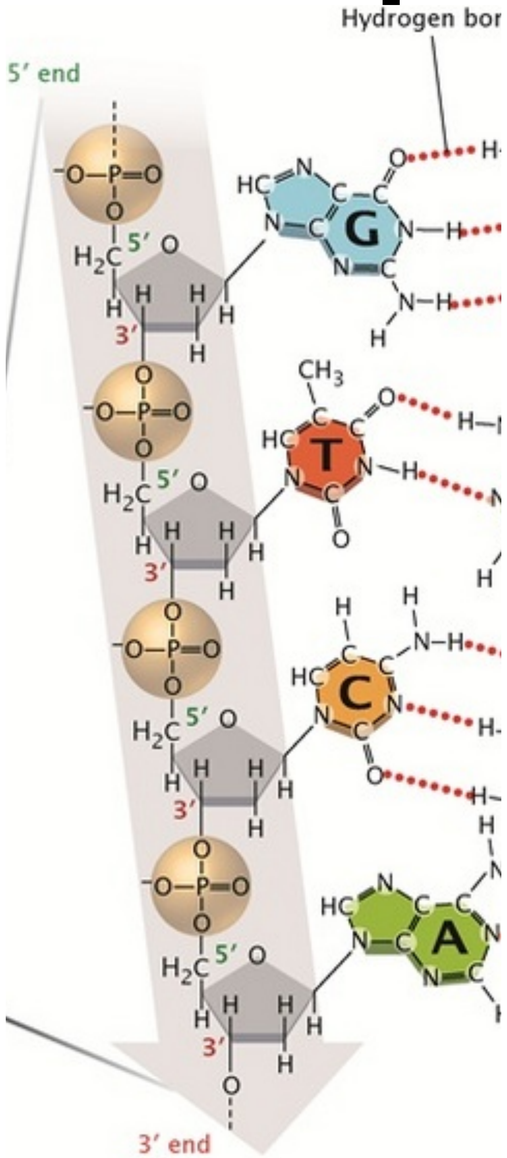
- A phosphate group
- A deoxyribose sugar
- A nitrogenous base
- Sugar +base=nucleoside
- Sugar+ base + phosphate= nucleotide



# DNA structure

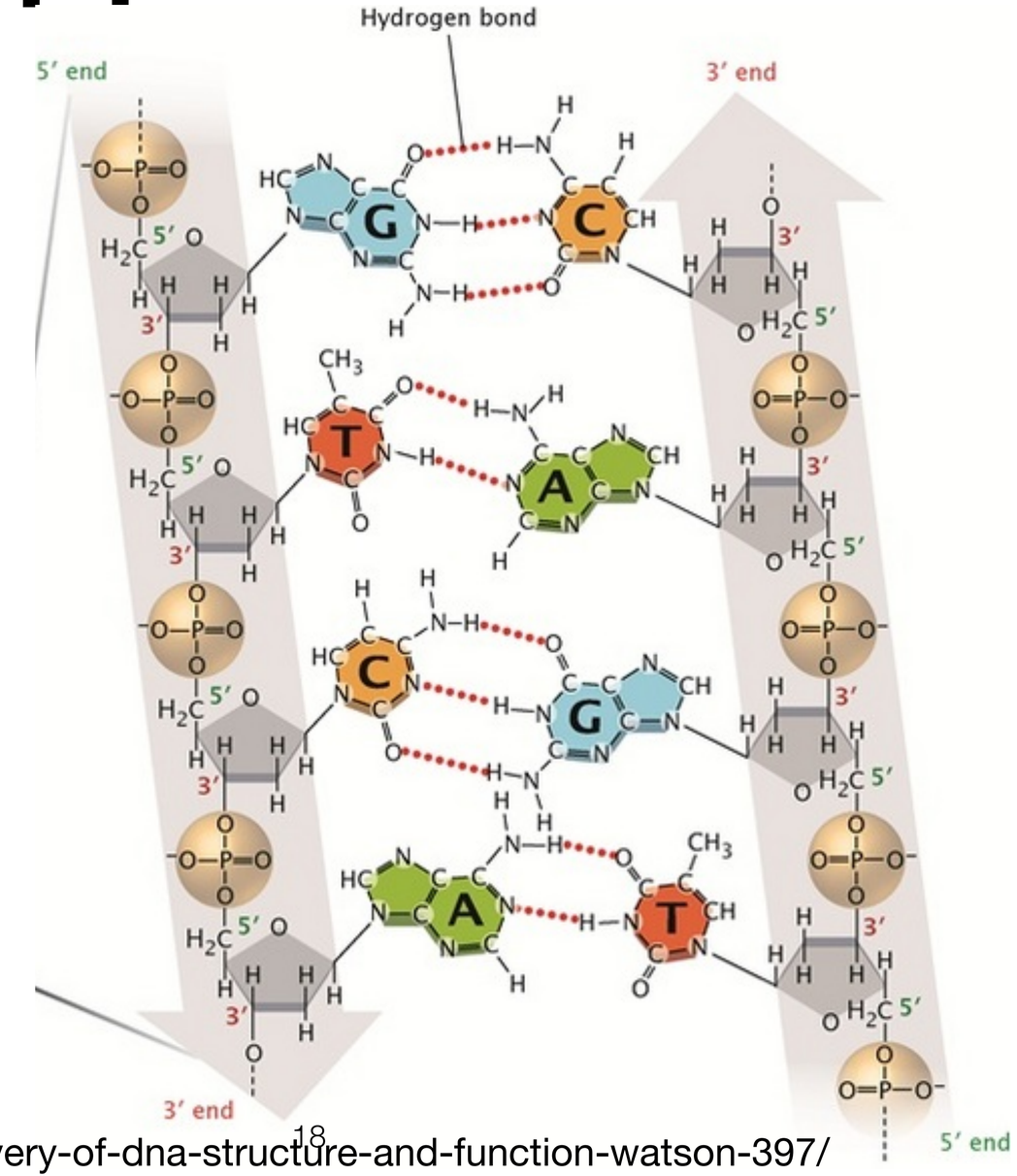


# 5 ‘prime’ vs 3 ‘prime’ ends



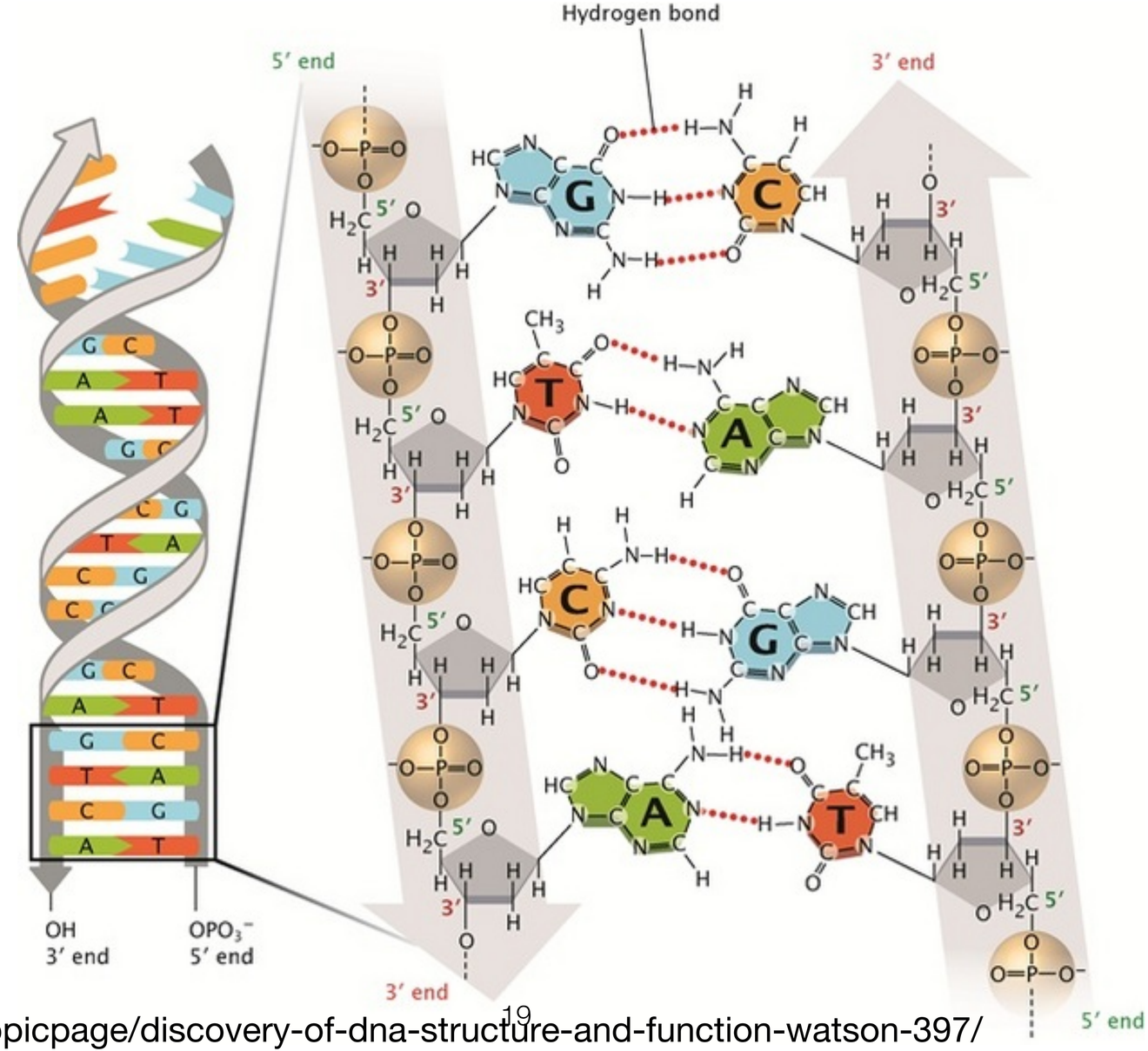

https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397/

# 2 strands, opposite directions




https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397/

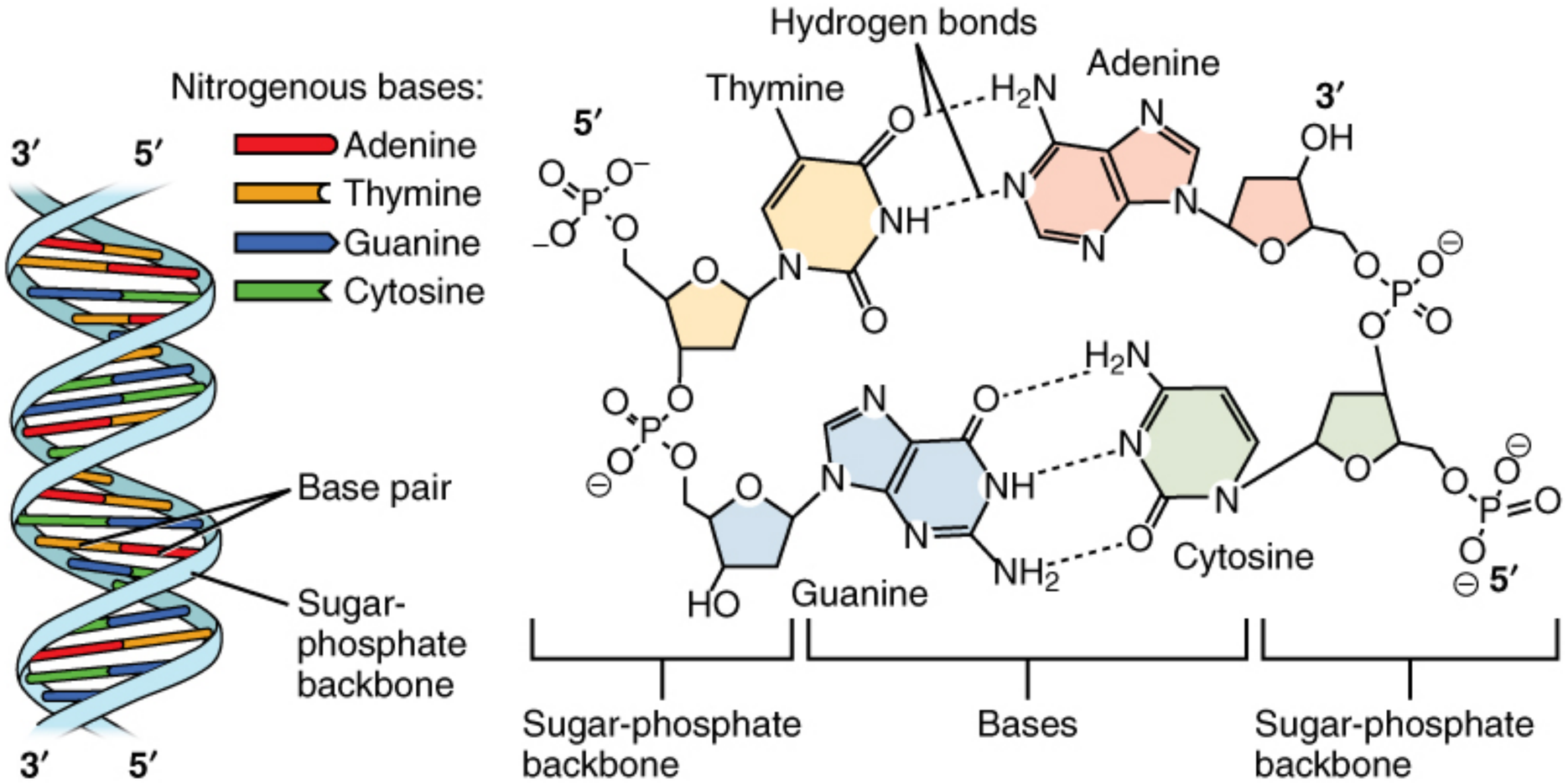
# Double helix



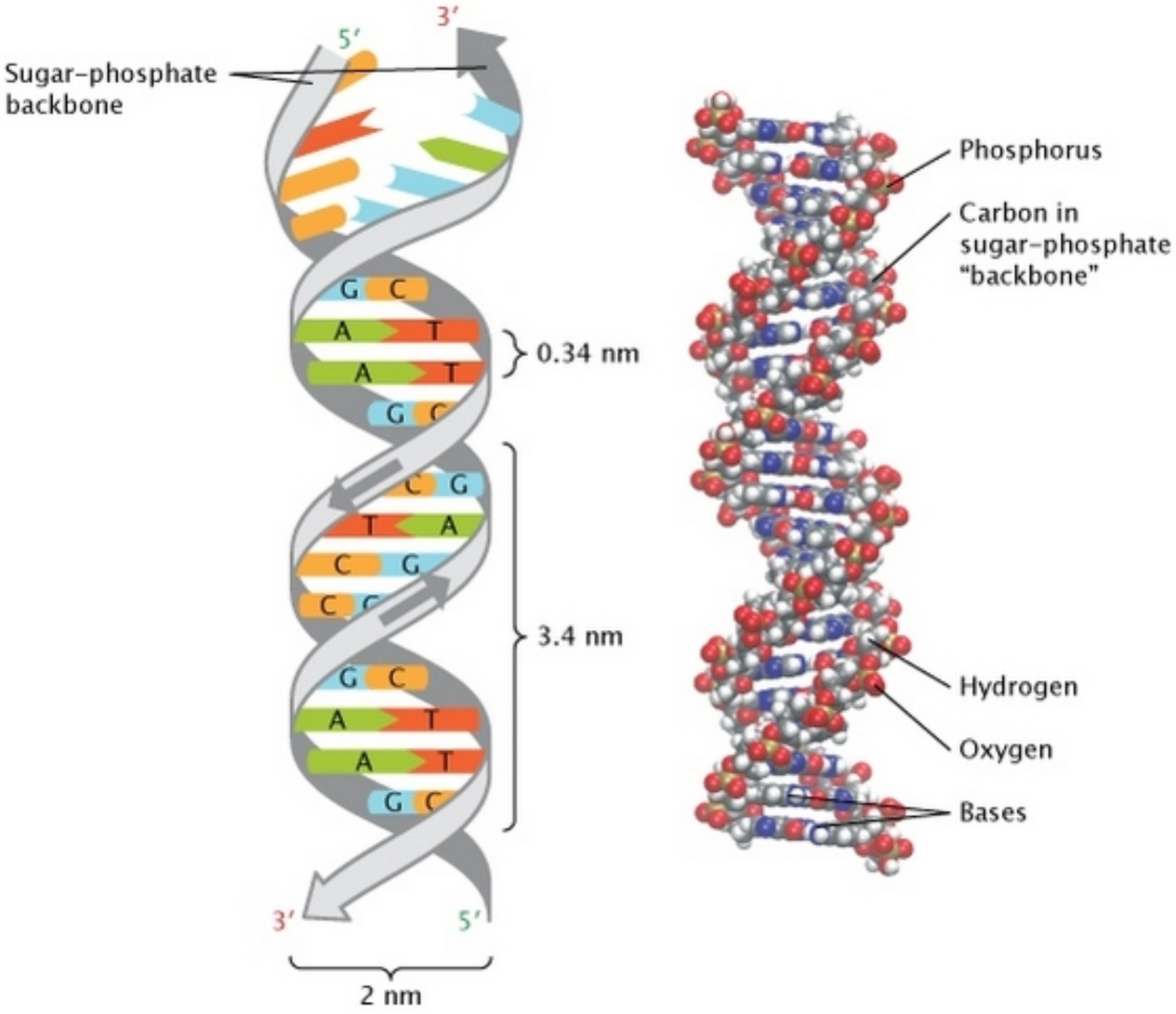

James watson
and Francis Crick

https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397/

# Double helix



# Double helix-major and minor groves

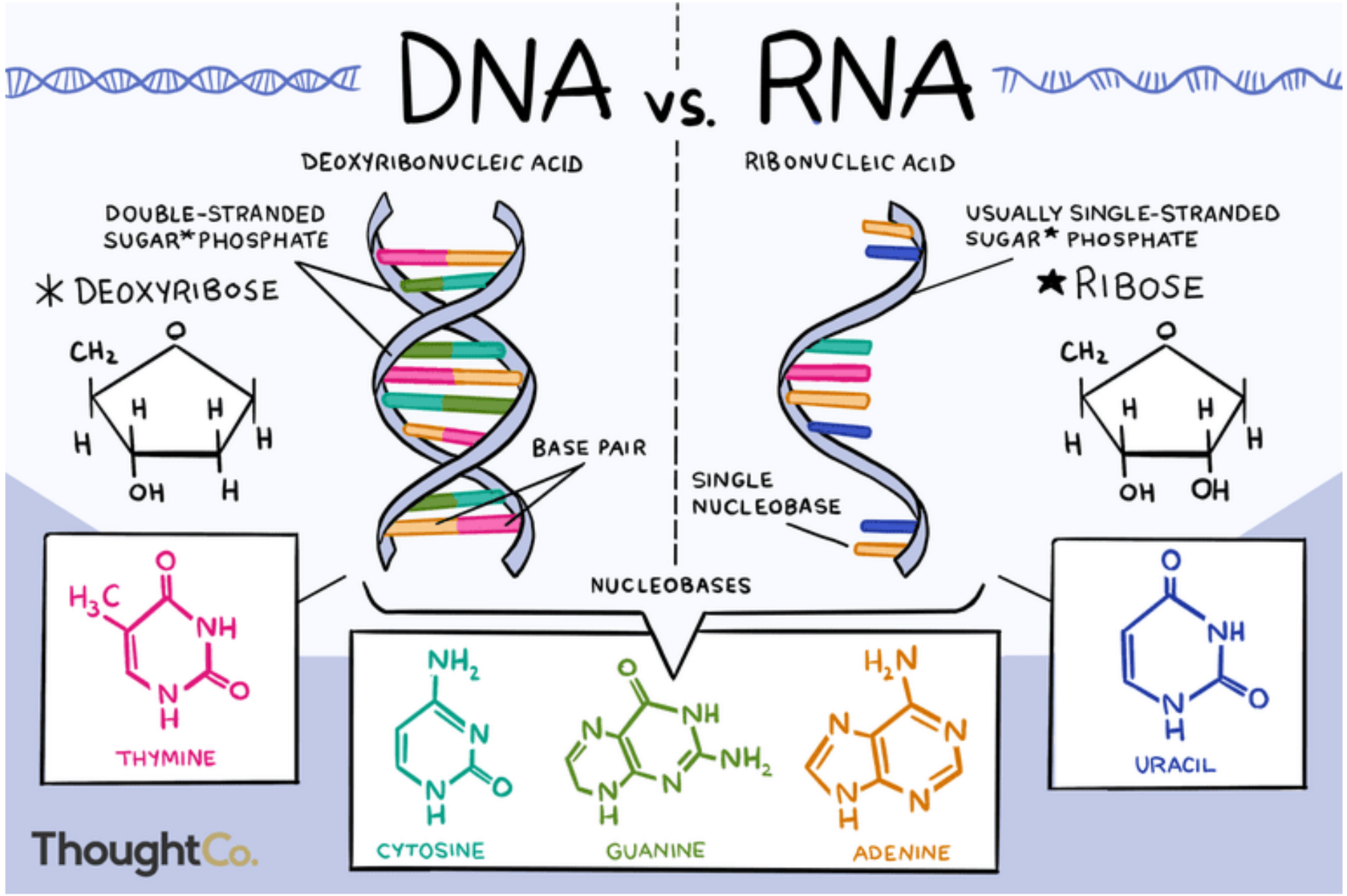



James watson
and Francis Crick

https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397/

# Ribonucleic acid (RNA)

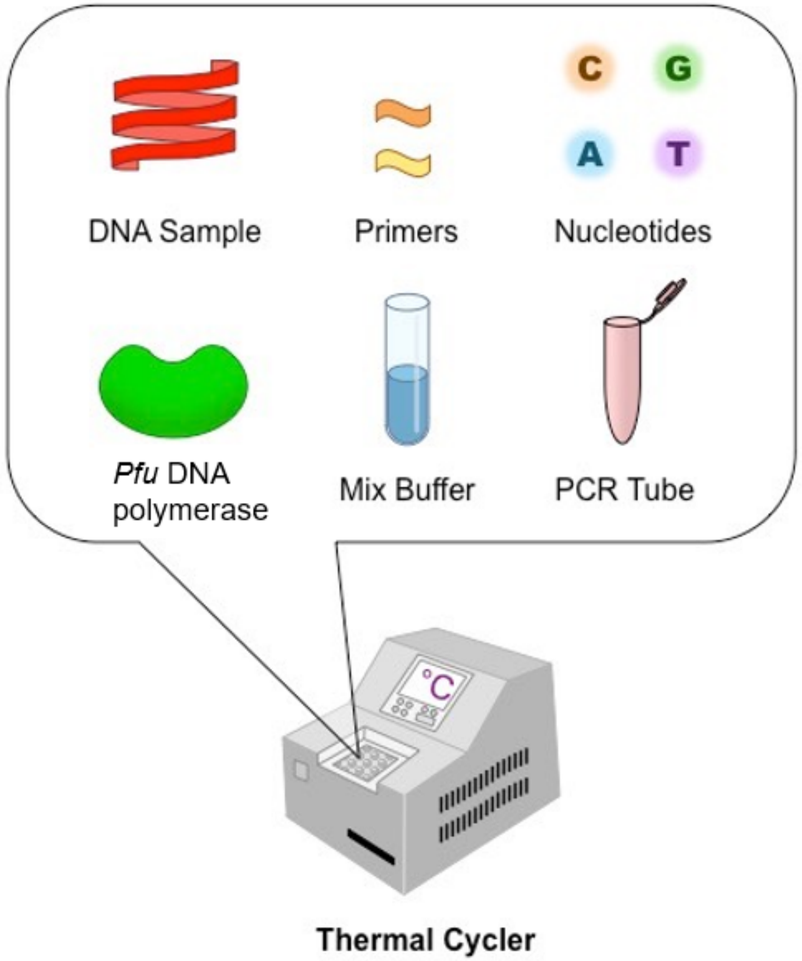
# DNA vs RNA




https://www.thoughtco.com/dna-versus-rna-608191

# Polymerase Chain Reaction (PCR)

- ❑ **Molecular biology technique**
- ❑ **Amplification of gene fragments**
- ❑ **Components of a PCR reaction**
  - ***Template DNA***
  - ***Primers***
  - ***Deoxy Nucleoside Triphosphates (dNTPs)***
  - ***Thermostable DNA polymerase***



# Polymerase Chain Reaction (PCR)

❑ **Components of a PCR reaction**

***1. Template DNA***

- ***Contains the gene sequence information amplified by PCR;***
  - ***e.g. genomic DNA (gDNA)***

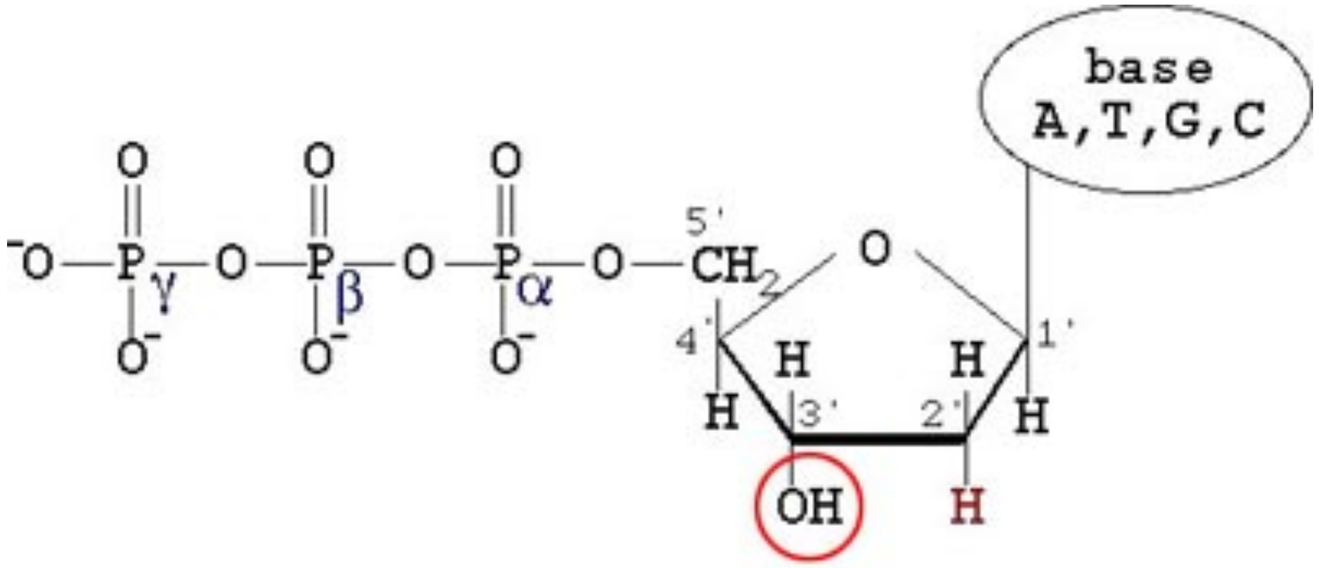
***2. Primers***

- ***Single-stranded DNA sequences;***
  - ***18-24 bp in length;***
- ***Anneal with the template DNA to amplify specific gene fragment;***
  - ***"Universal" primers used for 16S rRNA gene sequence amplification.***

# Polymerase Chain Reaction (PCR)

❑ **Components of a PCR reaction**

***3. Deoxy nucleotide triphosphates (dNTPs)***
***(dATP, dTTP, dGTP, dCTP)***



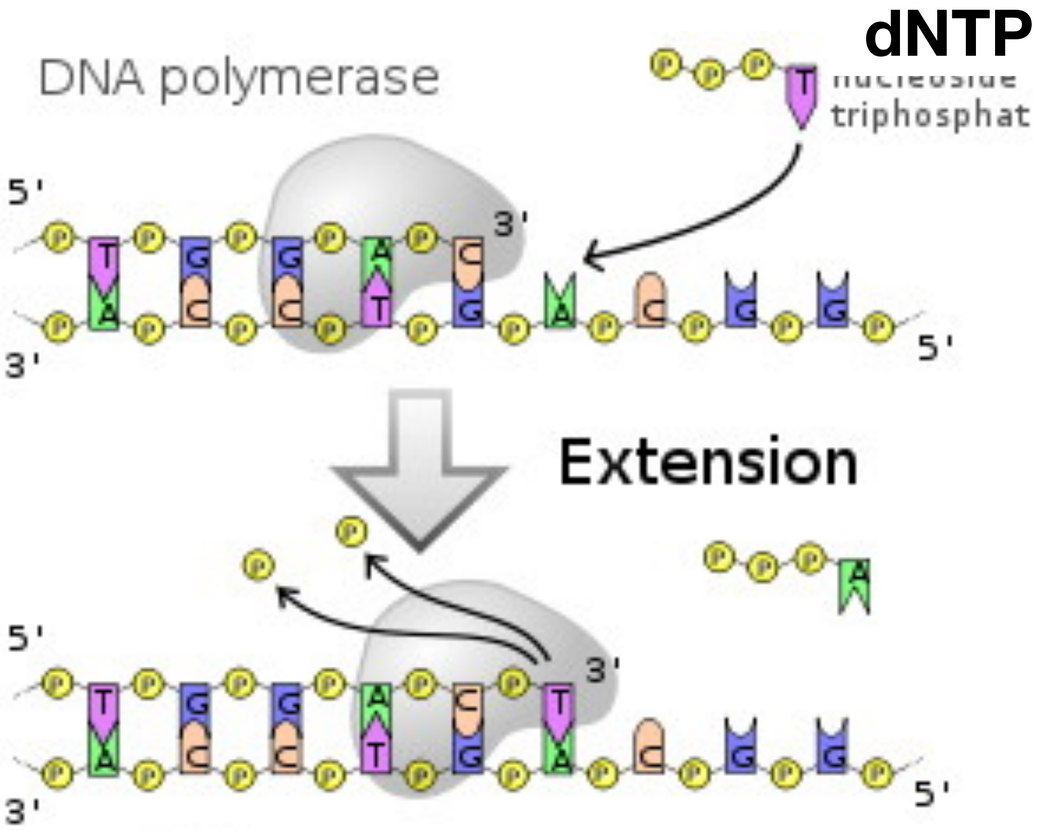

dNTP
deoxyribonucleotide triphosphate

# Polymerase Chain Reaction (PCR)

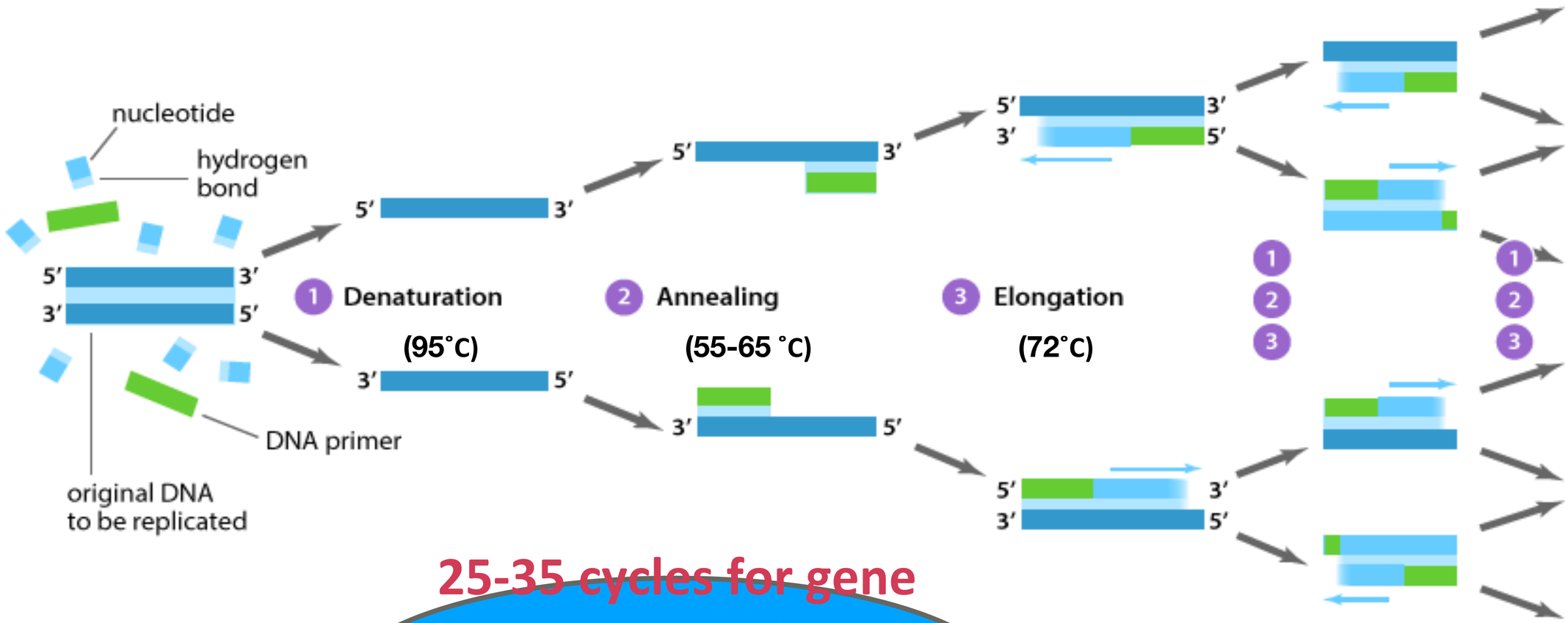
❑ **Components of a PCR reaction**

***4. Thermostable DNA polymerase***

- ***Extends the annealed primer by adding complementary dNTPs;***
- ***e.g. Pfu DNA polymerase purified from Pyrococcus furiosus.***



# Basic steps in a PCR reaction




25-35 cycles for gene fragment amplification

# QUESTIONS?

# THE PRINCIPLES OF SEQUENCING

Sequencing:

to determine the primary structure of an unbranched biopolymer. Sequencing results in a linear depiction known as a sequence which summarizes the atomic-level structure of the sequenced molecule.

# THE TYPES OF SEQUENCING

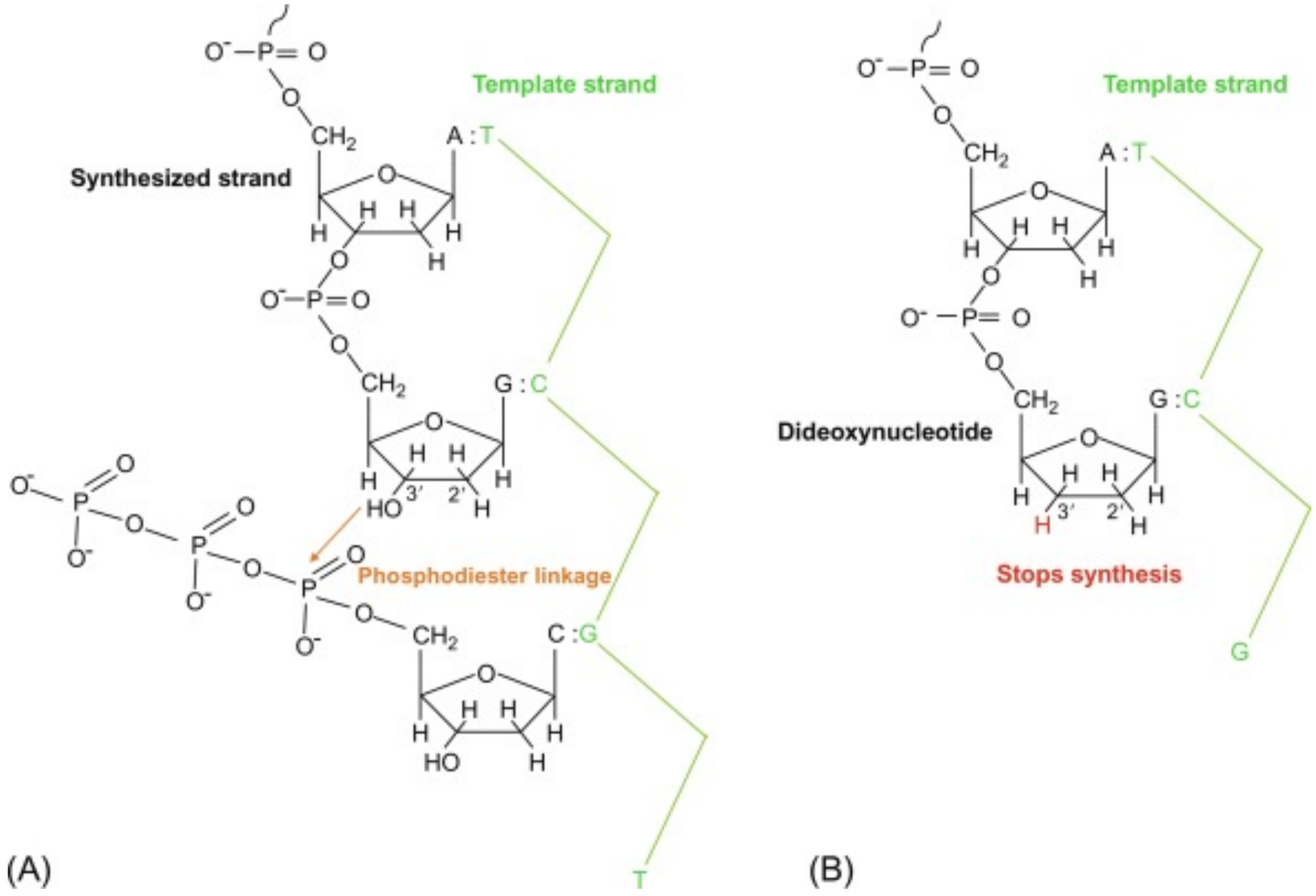
1. SANGERS Sequencing (chain termination sequencing, di-deoxynucleotide sequencing)
2. Next generation sequencing

# Sangers sequencing

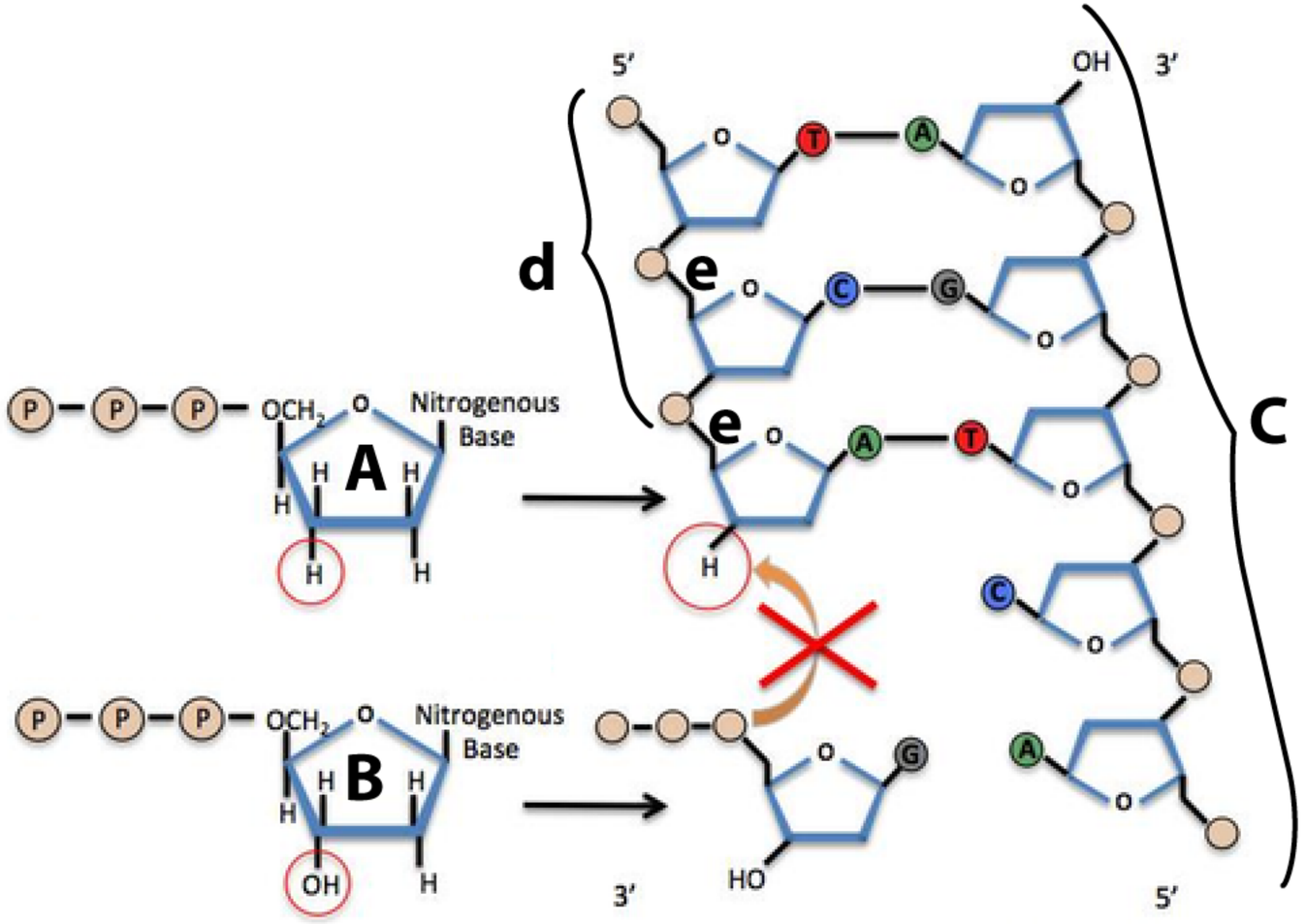
It is a method of DNA sequencing based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymeraseduring in vitro DNA replication.

Developed by Frederick Sanger and colleagues in 1977

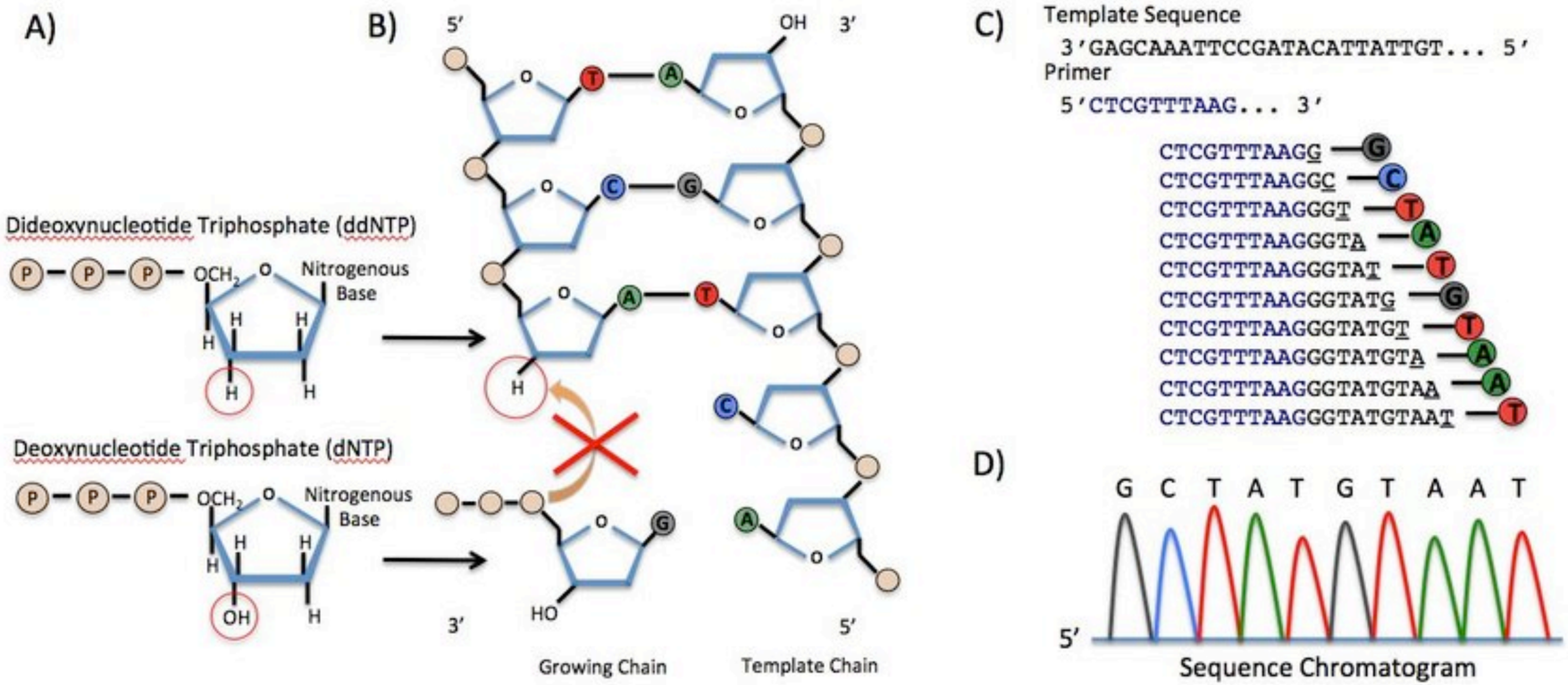
# Sangers sequencing principle



# Sangers sequencing principle

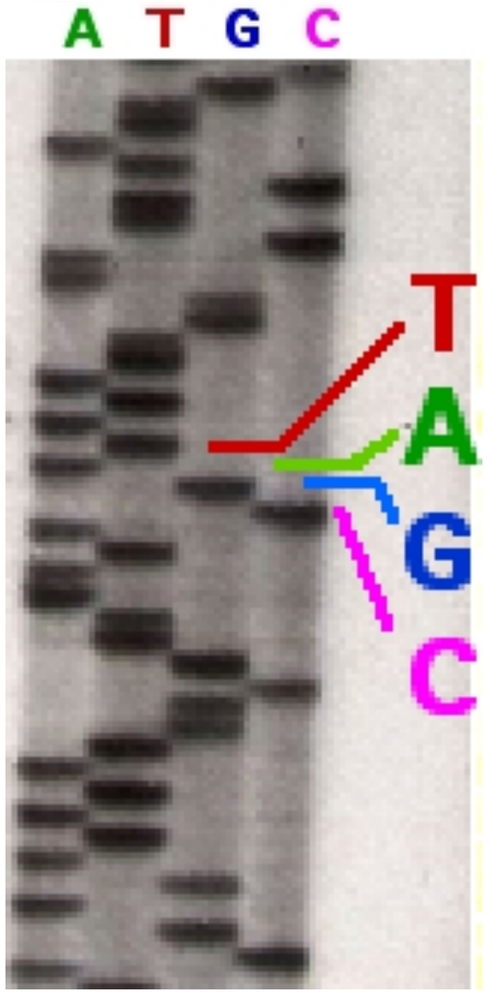


# Sangers sequencing principle




https://www.researchgate.net/publication/303565848_The_Use_of_Genetic_Sequencing_Technologies_to_Determine_HIV-1_Viral_Tropism_and_to_Evaluate_the_Effects_of_Maraviroc_on_Patient_Viral_Populations/figures?lo=1

# Sangers sequencing

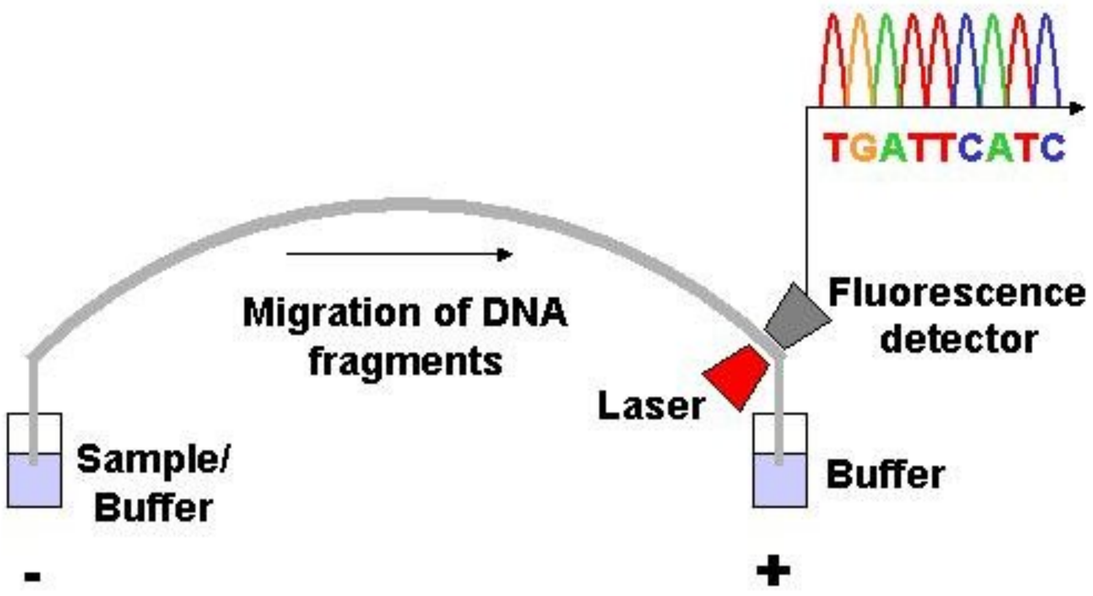



https://en.wikipedia.org/wiki/Sanger_sequencing#/media/File:Sanger-sequencing.svg

# Sangers sequencing

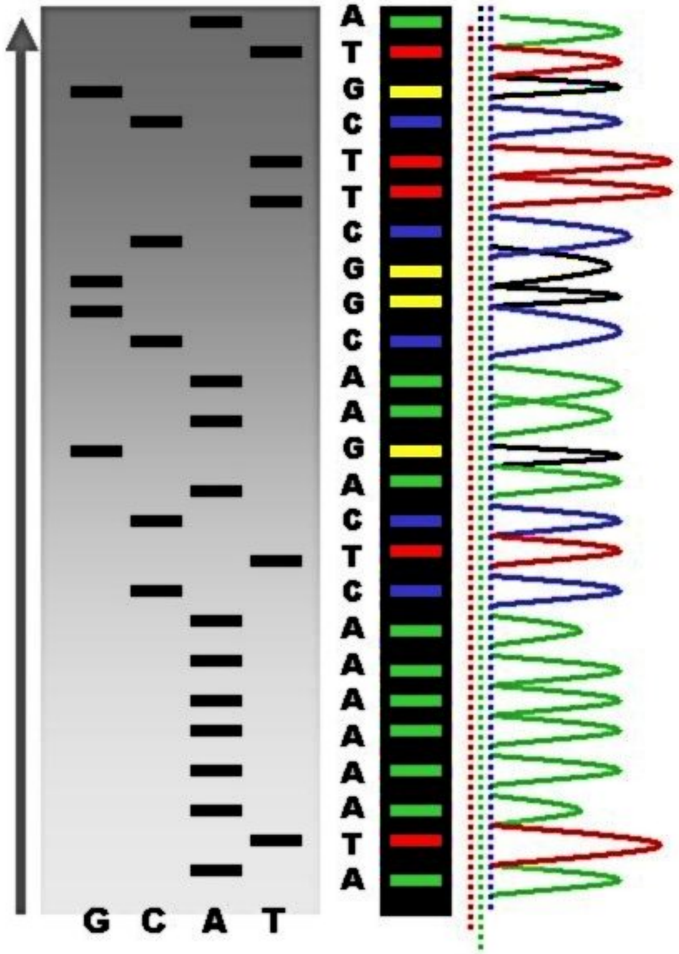
https://letstalkscience.ca/educational-resources/backgrounders/sanger-sequencing

# Sangers sequencing



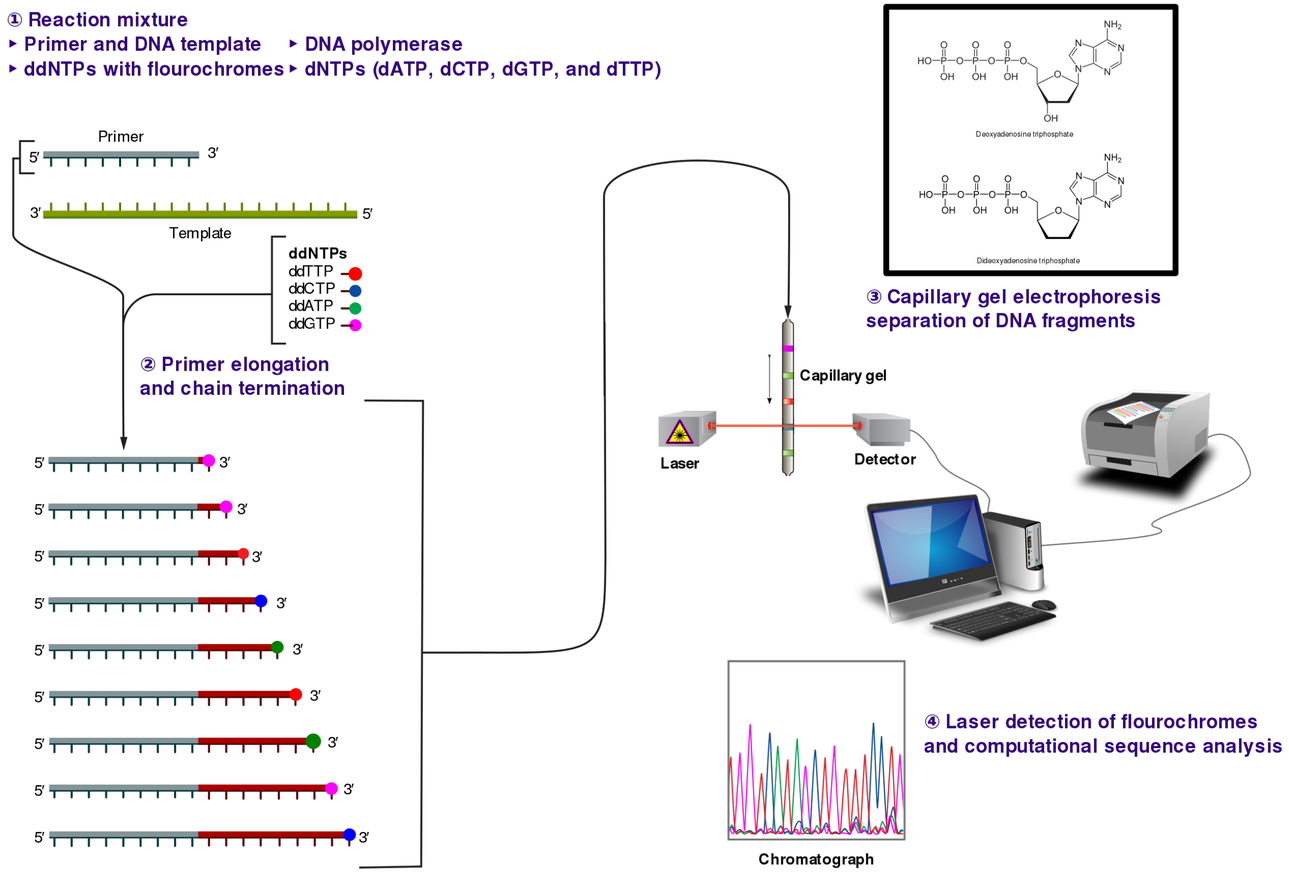

https://en.wikipedia.org/wiki/Sanger_sequencing#/media/File:Sanger-sequencing.svg

# Sangers sequencing



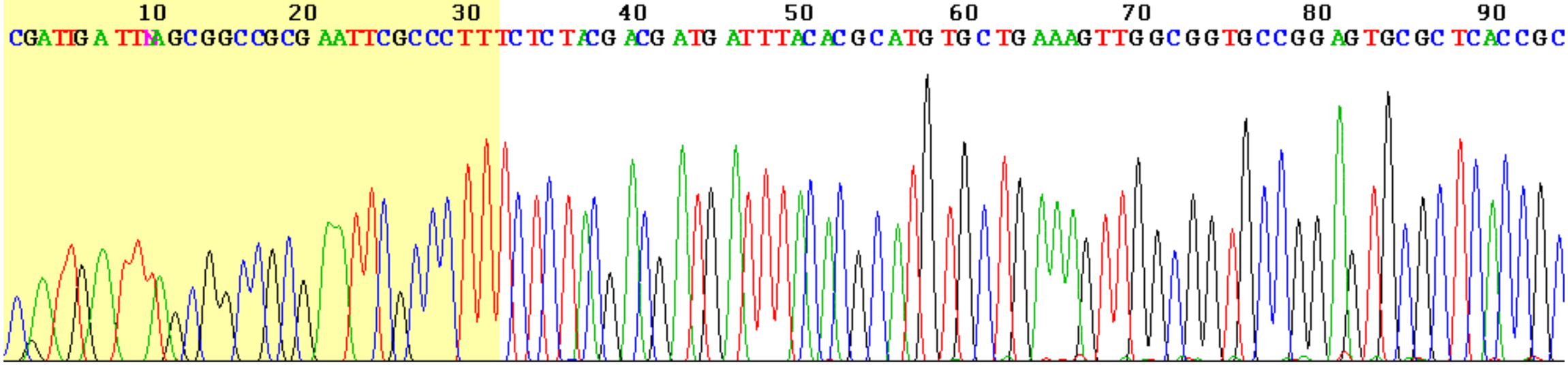

https://en.wikipedia.org/wiki/Sanger_sequencing#/media/File:Sanger-sequencing.svg

# Sangers sequencing




https://en.wikipedia.org/wiki/Sanger_sequencing#/media/File:Sanger-sequencing.svg

# Sangers sequencing




https://en.wikipedia.org/wiki/Sanger_sequencing#/media/File:Sanger-sequencing.svg

# Next generation sequencing (NGS)

1. The generation of many millions of short reads in parallel
2. The speed up of sequencing the process compared to the first generation
3. The low cost of sequencing and
4. The sequencing output is directly detected without the need for electrophoresis.

# Next generation sequencing (NGS)

| Platform | Instrument | Reads per run | Avg Read length (pb) | Read Type | Error Type | Error Rate (%) | Data Generated per run (Gb) | Year |
|---|---|---|---|---|---|---|---|---|
| **First Generation** | | | | | | | | |
| ABI Sanger | 3730xl | 96 | 400 – 900* | SE | NA | 0.3 | 0.00069 to 0.0021 | 2002 |
| **Second Generation** | | | | | | | | |
| 454 | GS20 | 200 | 100 | SE, PE | indel | 1 | 0.02 | 2005 |
| 454 | GS FLX | 400 | 250 | SE, PE | indel | 1 | 0.1 | 2007 |
| 454 | GS FLX Titanium | 1 M | 450 | SE, PE | indel | 1 | 0.45 | 2009 |
| 454 | GS FLX Titanium+ | 1 M | 700 | SE, PE | indel | 1 | 0.7 | 2011 |
| 454 | GS Junior | 100 | 400 | SE, PE | indel | 1 | 0.04 | 2010 |
| 454 | GS Junior+ | 100 | 700 | SE, PE | indel | 1 | 0.07 | 2014 |
| Illumina | MiniSeq | 25M (maximum) | 150 | SE, PE | mismatch | 1 | 7.5 (maximum) | 2013 |
| Illumina | MiSeq | 25M (maximum) | 300 | SE, PE | mismatch | 0.1 | 15 (maximum) | 2011 |
| Illumina | NextSeq | 400M (maximum) | 150 | SE, PE | mismatch | 1 | 120 (maximum) | 2014 |
| Illumina | HiSeq | 5B (maximum) | 150 | SE, PE | mismatch | 0.1 | 1.5Tb (maximum) | 2012 |
| Illumina | HiSeq X | 6B (maximum) | 150 | SE, PE | mismatch | 0.1 | 1.8Tb (maximum) | 2014 |
| SOLiD | 5500 W | 3B | 75 | SE | mismatch | ~0.1 | 160 | 2011 |
| SOLiD | 5500xl W | 6B | 75 | SE | mismatch | ~0.1 | 320 | 2013 |
| Ion Torrent | PGM 314 chip v2 | 400.000-550.000 | 400 | SE | indel | 1 | 0.06 to 0.1 | 2011 |
| Ion Torrent | PGM 316 chip v2 | 2M - 3M | 200 | SE | indel | 1 | 0.6 to 1 | 2011 |
| Ion Torrent | PGM 318 chip v2 | 4M - 5.5M | 400 | SE | indel | 1 | 1.2 to 2 | 2013 |

https://www.longdom.org/open-access/generations-of-sequencing-technologies-from-first-to-next-generation-0974-8369-1000395.pdf

# Next generation sequencing (NGS)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Third Generation** | | | | | | | | |
| PacBio | RS C1 | 432 | 1300 | SE | indel | 15 | 0.54 | 2011 |
| PacBio | RS C2 | 432 | 2500 | SE | indel | 15 | 0.5 to 1 | 2012 |
| PacBio | RS C2 XL | 432 | 4300 | SE | indel | 15 | 0.5 to 1 | 2012 |
| PacBio | RS II C2 XL | 564 | 4600 | SE | indel | 15 | 0.5 to 1 | 2013 |
| PacBio | RS II P5 C3 | 528 | 8500 | SE | indel | 13 | 0.5 to 1 | 2014 |
| PacBio | RS II P6 C4 | 660 | 13500 | SE | indel | 12 | 0.5 to 1 | 2014 |
| PacBio | Sequel | 350 | 10000 | SE | NA | NA | 7 | 2016 |
| Oxford Nanopore | MinION Mk | 100 | 9545 | 1D,2D | indel/mismatch | 12 | 1.5 | 2015 |
| Oxford Nanopore | PromethION | NA | 9846 | 1D,2D | NA | NA | 2Tb to 4Tb | 2016 |
| *depending on run module; NA: Not available; SE: Single End; PE: Paired End; M: Million; B: Billion; Gb: Gigabytes; Tb: Terabytes | | | | | | | | |

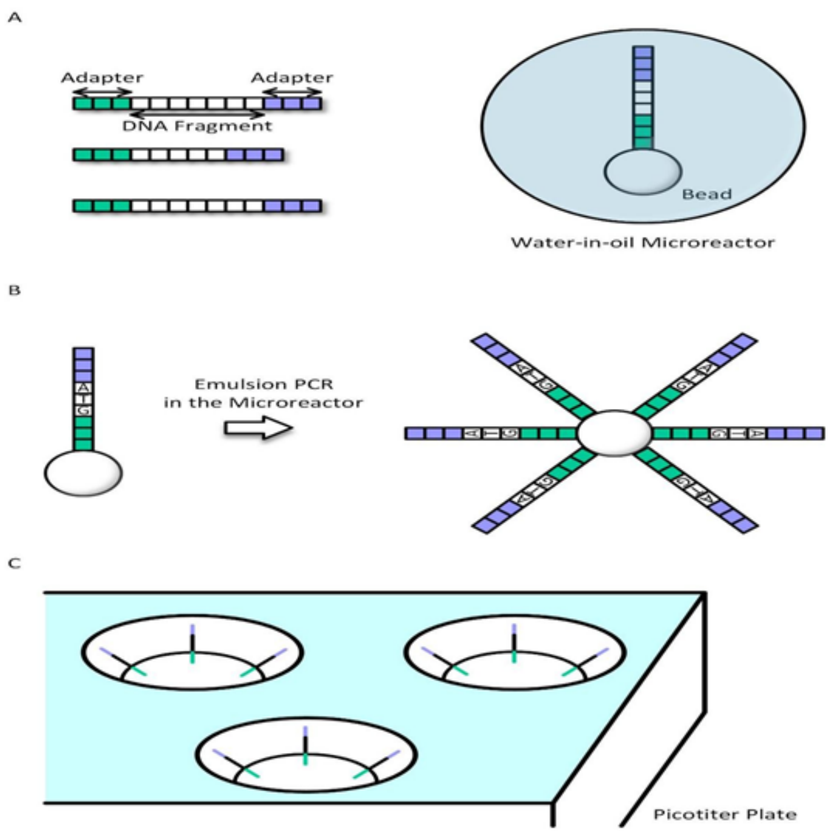
https://www.longdom.org/open-access/generations-of-sequencing-technologies-from-first-to-next-generation-0974-8369-1000395.pdf

# Next generation sequencing (NGS)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Third Generation** | | | | | | | | |
| PacBio | RS C1 | 432 | 1300 | SE | indel | 15 | 0.54 | 2011 |
| PacBio | RS C2 | 432 | 2500 | SE | indel | 15 | 0.5 to 1 | 2012 |
| PacBio | RS C2 XL | 432 | 4300 | SE | indel | 15 | 0.5 to 1 | 2012 |
| PacBio | RS II C2 XL | 564 | 4600 | SE | indel | 15 | 0.5 to 1 | 2013 |
| PacBio | RS II P5 C3 | 528 | 8500 | SE | indel | 13 | 0.5 to 1 | 2014 |
| PacBio | RS II P6 C4 | 660 | 13500 | SE | indel | 12 | 0.5 to 1 | 2014 |
| PacBio | Sequel | 350 | 10000 | SE | NA | NA | 7 | 2016 |
| Oxford Nanopore | MinION Mk | 100 | 9545 | 1D,2D | indel/mismatch | 12 | 1.5 | 2015 |
| Oxford Nanopore | PromethION | NA | 9846 | 1D,2D | NA | NA | 2Tb to 4Tb | 2016 |
| *depending on run module; NA: Not available; SE: Single End; PE: Paired End; M: Million; B: Billion; Gb: Gigabytes; Tb: Terabytes | | | | | | | | |

https://www.longdom.org/open-access/generations-of-sequencing-technologies-from-first-to-next-generation-0974-8369-1000395.pdf

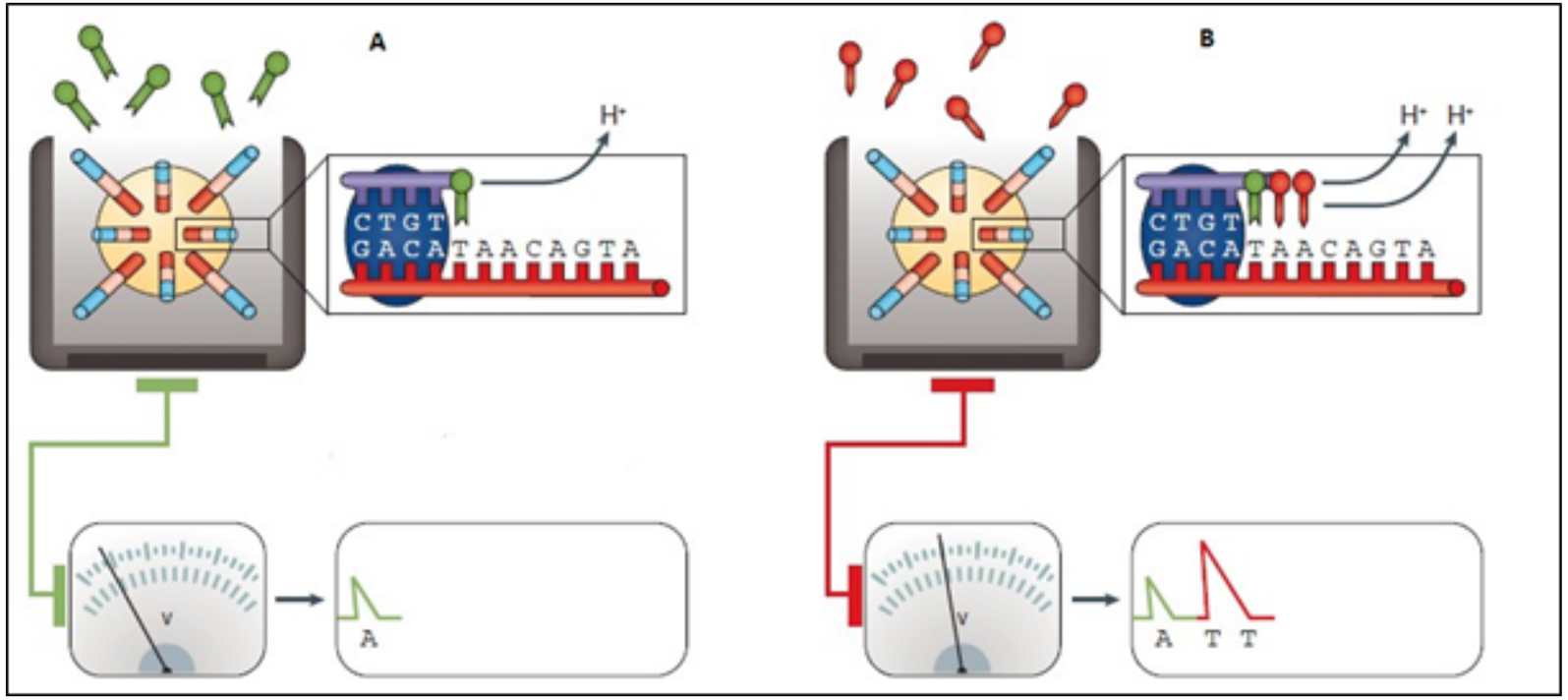
# Next generation sequencing (NGS)




Roche/454 sequencing technology uses pyrosequencing methods using pyrosequencing technique which is based on the detection of pyrophosphate released after each nucleotide incorporation in the new synthetic DNA strand

https://www.longdom.org/open-access/generations-of-sequencing-technologies-from-first-to-next-generation-0974-8369-1000395.pdf

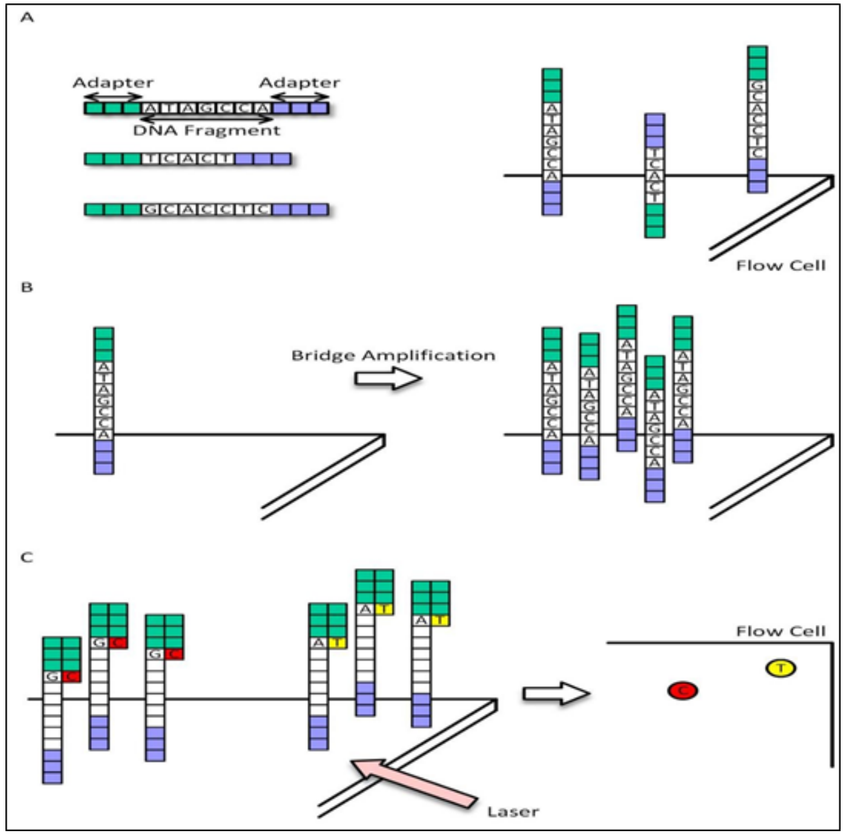
# Next generation sequencing (NGS)




Ion torrent sequencing technology based on the detection of the hydrogen ion released during the sequencing process

https://www.longdom.org/open-access/generations-of-sequencing-technologies-from-first-to-next-generation-0974-8369-1000395.pdf

# Next generation sequencing (NGS)

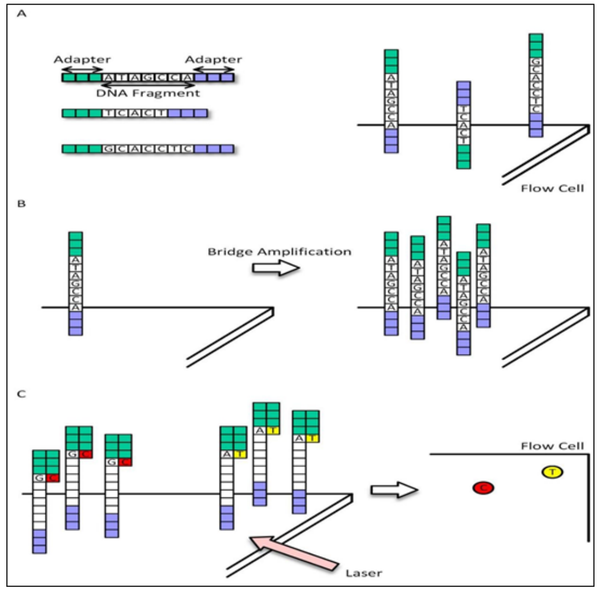



**Illumina sequencing technology uses adapters, it is the most common method**

https://www.longdom.org/open-access/generations-of-sequencing-technologies-from-first-to-next-generation-0974-8369-1000395.pdf

# Next generation sequencing (NGS)

1. DNA samples are randomly fragmented into sequences
2. Adapters are ligated to both ends of each sequence.
3. Adapters are fixed themselves to the respective complementary adapters,
4. Adapters are hooked on a slide with many variants of adapters (complementary) placed on a solid plate .



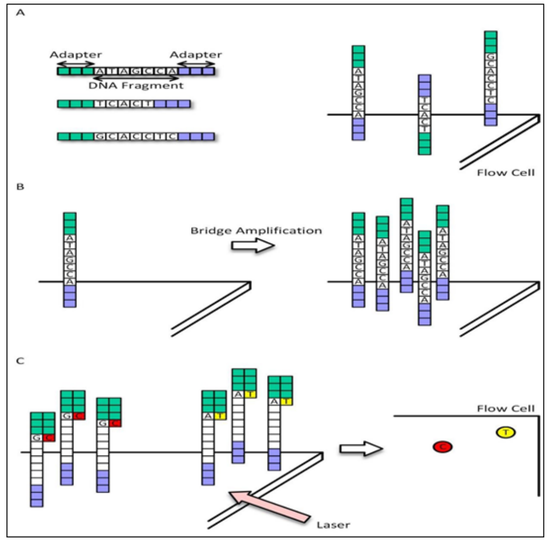

https://www.longdom.org/open-access/generations-of-sequencing-technologies-from-first-to-next-generation-0974-8369-1000395.pdf

# Next generation sequencing (NGS)

5. During the second step, each attached sequence to the solid plate is amplified by “PCR bridge amplification” that creates several identical copies of each sequence;

6. Set of sequences made from the same original sequence is a cluster.

7. Each cluster contains approximately one million copies of the same original sequence




https://www.longdom.org/open-access/generations-of-sequencing-technologies-from-first-to-next-generation-0974-8369-1000395.pdf

# Next generation sequencing (NGS)

8. The last step is to determine each nucleotide in the sequences, Illumina uses the sequencing by synthesis approach that employs reversible terminators in which the four modified nucleotides, sequencing primers and DNA polymerases are added as a mix, and the primers are hybridized to the sequences

9. Then, polymerases are used to extend the primers using the modified nucleotides. Each type of nucleotide is labeled with a fluorescent specific in order for each type to be unique.

https://www.longdom.org/open-access/generations-of-sequencing-technologies-from-first-to-next-generation-0974-8369-1000395.pdf

# Next generation sequencing (NGS)

10. The nucleotides have an inactive 3’-hydroxyl group which ensures that only one nucleotide is incorporated. Clusters are excited by laser for emitting a light signal specific to each nucleotide, which will be detected by a coupled-charge device (CCD) camera and Computer programs will translate these signals into a nucleotide sequence

11. The process continues with the elimination of the terminator with the fluorescent label and the starting of a new cycle with a new incorporation

https://www.longdom.org/open-access/generations-of-sequencing-technologies-from-first-to-next-generation-0974-8369-1000395.pdf

# NGS Video

**https://www.illumina.com/science/technology/next-generation-sequencing/beginners.html**

**https://www.youtube.com/watch?v=CZeN-IgjYCo**

# Questions?

# Sequencing data analysis and Phylogenetic analysis

**Softwares**

MEGA : https://www.megasoftware.net/download_form
BioEdit : http://www.mbio.ncsu.edu/BioEdit/bioedit.html

GENEIOUS: https://www.geneious.com

e.t.c

# Sequencing data analysis and Phylogenetic analysis

1. First import your sequence into a software e.g bioedit or Geneious
2. Crosscheck the sequence chromogram to be sure that the peaks are correct
3. Edit sequence by trimming, correcting and save as FASTA file
4. Download other similar sequences from the gene bank (NCBI)
5. Align the sequences on MEGA or Bioedit
6. Compose a phylogeny on bioedith or MEGA
7. Observe and read your phylogeny

# Sequencing data analysis and Phylogenetic analysis

Download your softwares and start with practical.

Good luck.