Genetics
& Cell biology

◉ Sheet ○ Slides

Number

7

Done by:

Bana Anwar Shawareb

Corrected by:

Yazan Sbeih Hijazeen

Doctor

Ma’moun

DNA sequencing and PCR

Resources: [The cell: a molecular approach by Cooper and Hausman]

Pages: (124-129), (159-162), (166-177)

In this lecture we will be discussing 2 techniques, **DNA sequencing** and **PCR (Polymerase Chain Reaction)**. These techniques made dramatic changes to the field of molecular biology and caused great advancements.

1. DNA Sequencing :

- ### Definition:

DNA sequencing is basically knowing the order of bases. In other words, we have a certain gene and we want to know how the bases are arranged, e.g. (A then C then G etc.) and by that we can identify the gene.

By this technique, we can know the code of DNA (just like using the 0-1 code in computer programming, in molecular biology we have the 4 bases: Adenine (A), Thymine (T), Cytosine (C), and Guanine (G)).

In 1989, Scientists started identifying the human genome. Based on the complexity of the physiology of our body, they expected to find 100,000 genes, but actually they found that they were approximately 25,000 genes.

-Human (Homo sapiens) ⟶ 3.2 billion base pairs/ 25,000 genes

- ### Importance of DNA sequencing :

1-Identification of genes and their localization

2- Identification of protein structure and function

3-Identification of DNA mutations

4-Genetic variation among individuals in health and disease

5-Evolutionary conservation among organisms

* What was the first organism to have its genome sequenced?

1-Viruses and prokaryotes were studied first

2- Human mitochondrial DNA

3-The first eukaryotic genome sequenced was that of yeast, Saccharomyces cerevisiae.

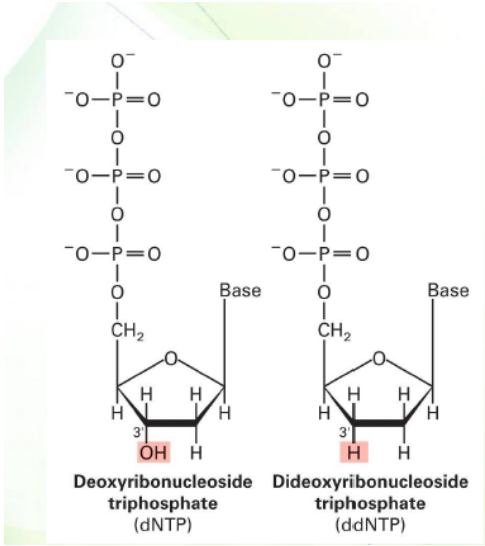
4- The genome of a multicellular organism, the nematode Caenorhabditis elegans.

* Determination of the base sequence in the human genome was initiated in 1990 and completed in May 2006 via the Human Genome Project.

(0-10 minutes)

- ## Method of DNA sequencing:

-The nucleotide used in DNA sequencing is a **Dideoxynucleotide** (missing 2 Hydroxyl groups on carbons 2 and 3). The importance of having an OH on carbon 3 is that any nucleotide that is going to be added will be added on this location. So, if this Hydroxyl group is missing (like in the case of dideoxynucleotides), the reaction will stop because there will not be further elongation in the DNA strand.



-Remember that the DNA sequencing reaction has millions of DNA fragments. So, we are not working on one sequence only.

The five components of a DNA sequencing reaction:

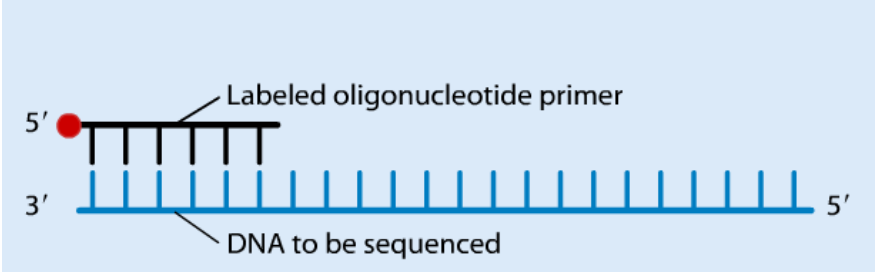
1-The DNA fragment that we want to sequence (DNA template)

2-Deoxynucleotides (A, T, C, G): the substrates of this reaction

3-DNA polymerase: for DNA synthesis

4-Primer: Because DNA polymerase can't perform its function without a primer (cannot do de novo synthesis)

5-Dideoxynucleotides: for DNA termination



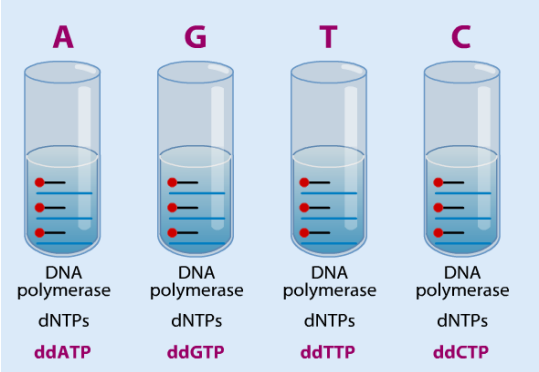
Note: The sequence of the primer should be complementary to the location of DNA to be sequenced.

A) The old method (using Radioactivity):

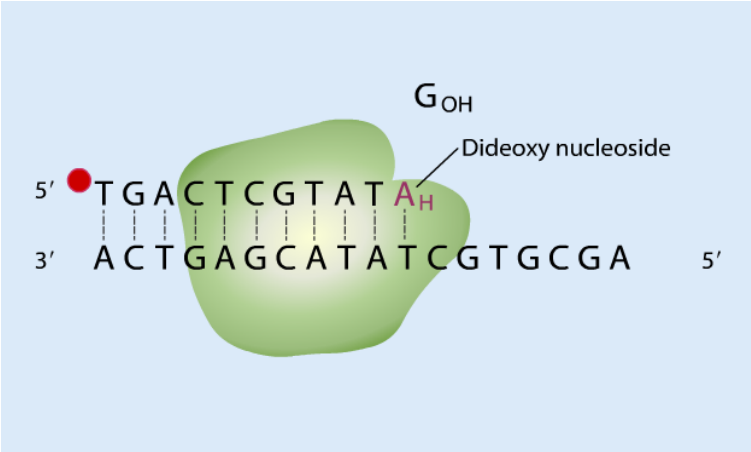
*Increasing the temperature to 95 °C to get a single strand.

*DNA synthesis is initiated from a primer that has been labeled with a radioisotope.

*Four reactions are run, each reaction contains: DNA template, DNA polymerase, Primer, and the deoxynucleotides. Then we add different types of dideoxynucleotides to each reaction (dideoxy A, G, T, C).



*DNA polymerase begins the process of synthesis, a series of labeled DNA molecules are generated, each terminated by the dideoxynucleotide in each reaction.



➡ Let's take the reaction with a **dideoxynucleotide A** as an example:

DNA polymerase starts synthesis (reading and placing complementary bases) until it reaches a Thymine (T). In this case DNA polymerase has 2 options: 1) Putting a deoxynucleotide A and therefore allowing further addition of nucleotides or 2) Putting a dideoxynucleotide A and stopping the elongation of DNA (stopping DNA synthesis) .Thus, forming a fragment. **This is a matter of probability**.

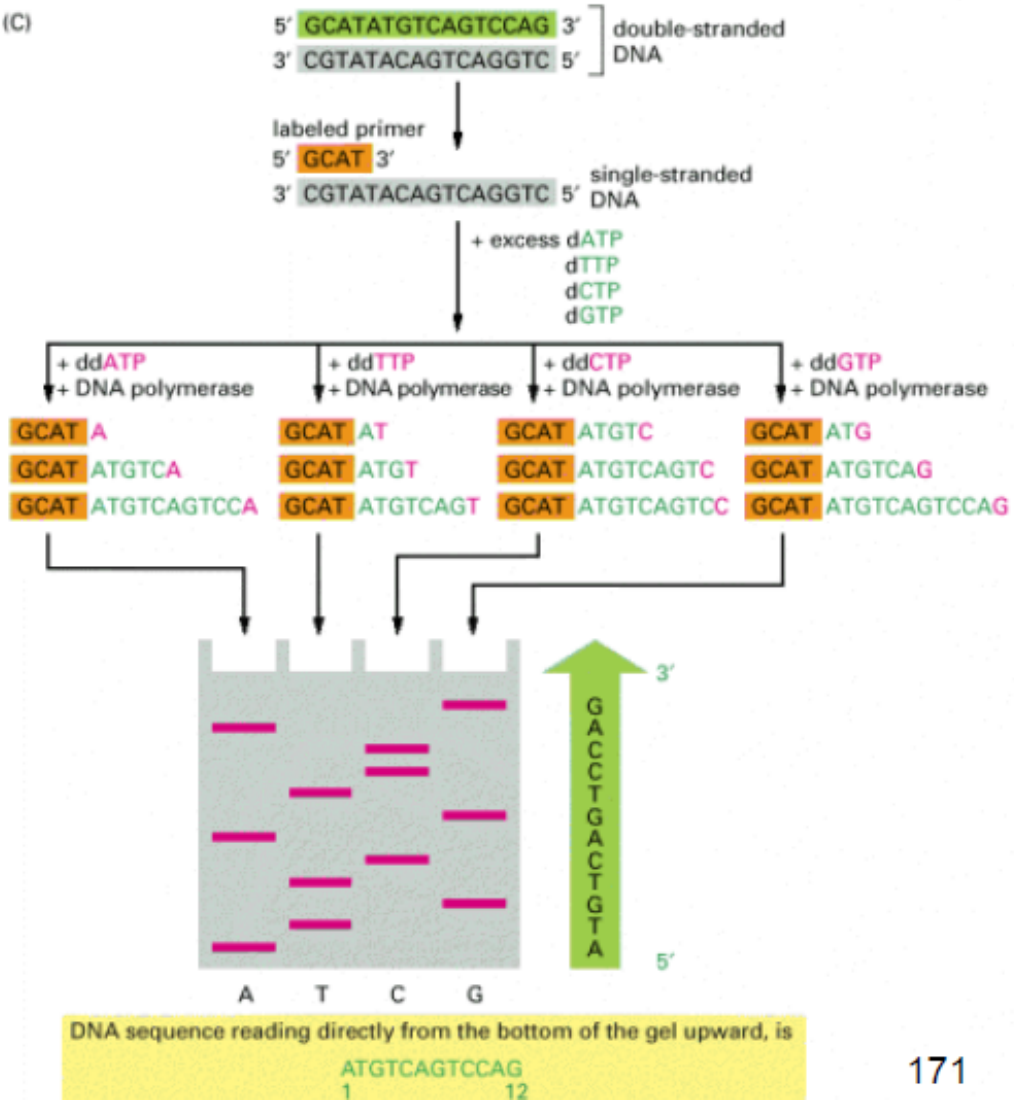
(10-20 minutes)

*So, eventually, we will have multiple fragments with different lengths (according to the location of termination of synthesis by ddNTPs)

*Next, these DNA fragments are separated according to size by gel electrophoresis and detected by exposure of the gel to X-ray film (each reaction is put in a separate lane). Electrophoresis allows the separation of fragments based on the difference of one nucleotide only (high resolution).

[Acrylamide gel is used in electrophoresis]

*Remember that DNA synthesis begins from 5’ to 3’, so when we read the DNA strand, we start with 5' (shortest fragment) to 3' (longest fragment). However, if we were asked to give the sequence of the **template**, we start from 3' to 5' (antiparallel).



- <u>Disadvantages of the old method (Radioactivity):</u>

a-Time Consuming.

b-We have to use four different lanes for the four reactions.

c-This method is performed manually.

d-Dangerous and not healthy.

So, scientists came up with an easy, safe method in which we use fluorescence...

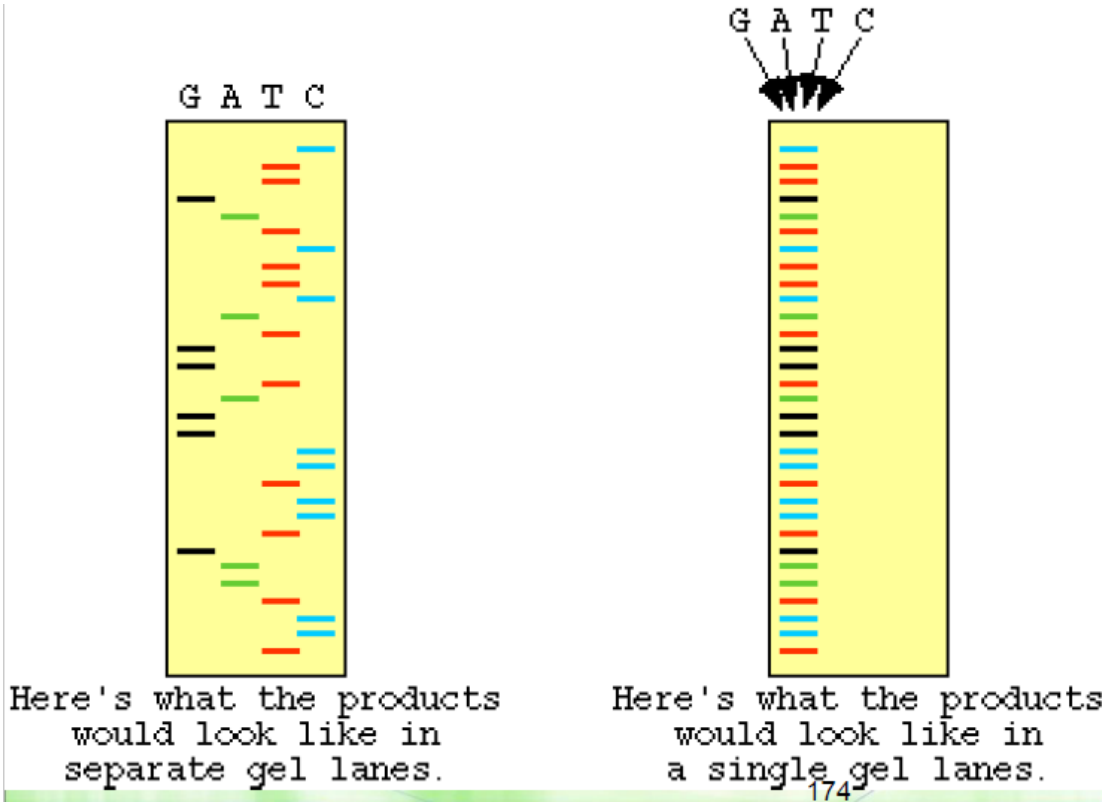
B) The new method (Fluorescence-based DNA sequencing):

*This method is automated; the reading is done by a laser (not manually as in radioactivity).

*In this method, only one reaction is needed (one tube is used), and the four dideoxynucleotides are labeled with fluorescent tags and put in this tube. (Note that in this method the primers are not labeled, but the dideoxynucleotides are).

*Different dideoxynucleotides give us different colors.

*The computer starts reading and recognizing the fragments (according to their colors) and eventually gives us the DNA sequence.



Our cells are diploid, so we have two copies of every gene (allele) and genetic variation can be found between maternal and paternal DNA (they are 99.9% similar but there is some variation).

-So if we take the similar maternal and paternal DNA and apply the old method (radioactivity), a single band will appear.

-If the two were different, we will have 2 bands on the same location;

Homozygous (normal) ---> One band

Heterozygous (normal) ---> Two bands (Two different fragments with the same length).

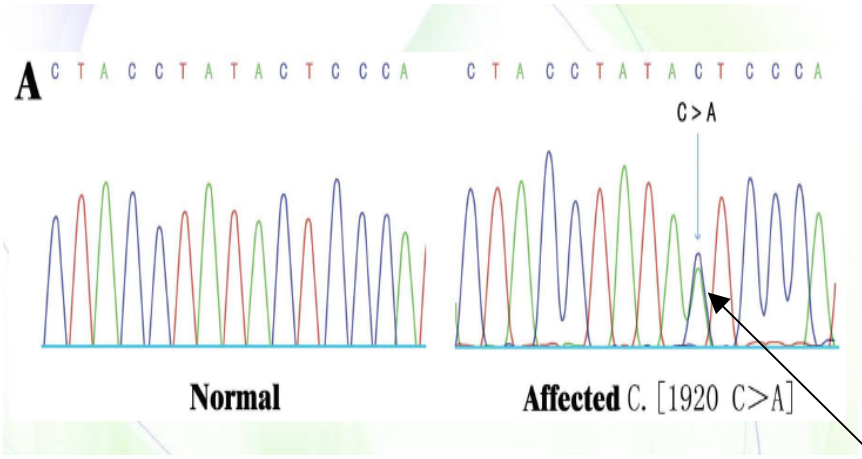
Mutation in both alleles ---> Two bands (to detect this mutation, the mutated individual must be compared with normal individual's results).

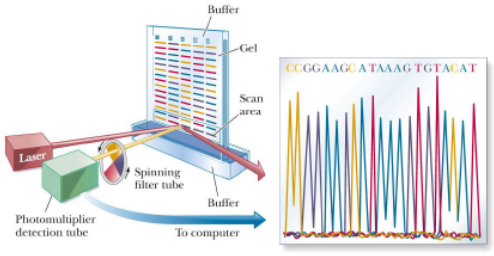
-If we use fluorescence-based sequencing, the computer will show the bands as peaks in different colors.

So, if a person has a mutation in both alleles (homozygous) ---> One peak (different from the normal individual; different color).

Heterozygous ---> Two signals traveling at the same time ---> Two different peaks (overlapping). We call this "**Double Peak**".

(20-35 minutes)





Double peak

2- PCR (Polymerase Chain Reaction) :

- Is a method for amplification of a specific region on a DNA sequence devised by Kary Mullis (by amplification we mean having many copies of one DNA sequences).

-It is a chain reaction (repeated many times).

-The PCR method is extremely sensitive; it can detect a single DNA molecule in a sample.

***Components of the PCR reaction:**

1) DNA template

2) DNA polymerase (heat-stable; obtained from microorganisms whose natural habitat is hot springs. For example, the widely used Taq DNA polymerase is obtained from a thermophilic bacterium, Thermus aquaticus, and is thermostable up to 94°C, and works optimally at 70-75°C).

3) Deoxynucleotides: A, T, G, C (the substrates of this reaction)

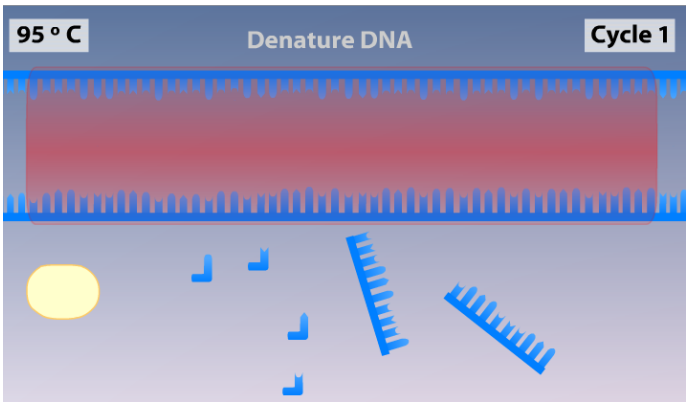
4) Primers: in this reaction we need 2 primers flanking on each side of the template (these primers should be specific for the target sequence and which are often about 15-25 nucleotides long).

- **PCR Cycle:**
  *Each PCR cycle consists of three steps

Cycle 1

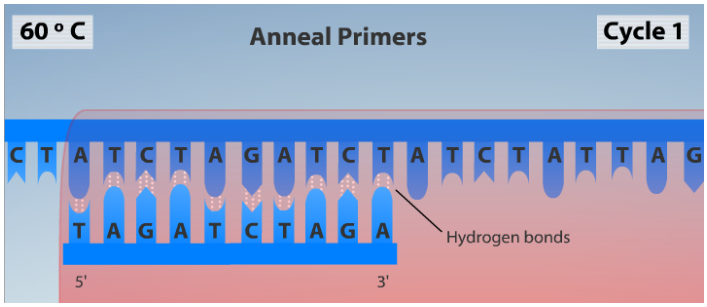
STEP 1: Denaturation:

The temperature is raised to 95 °C. This allows the separation of the two DNA strands (The hydrogen bonds that hold together the two polynucleotides of the double helix are broken, so the target DNA becomes denatured into single-stranded molecules).



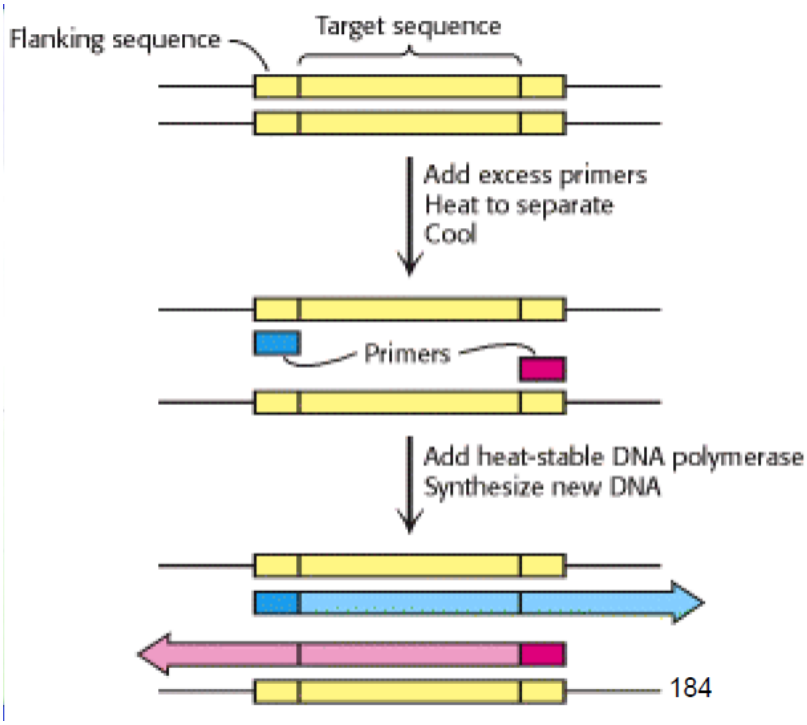
STEP 2: Annealing:

The temperature is decreased to 50-60°C . In this step, the primers anneal to the DNA template.



STEP 3: DNA synthesis:

Again, the temperature is raised to 72°C . In this step, elongation of the DNA occurs



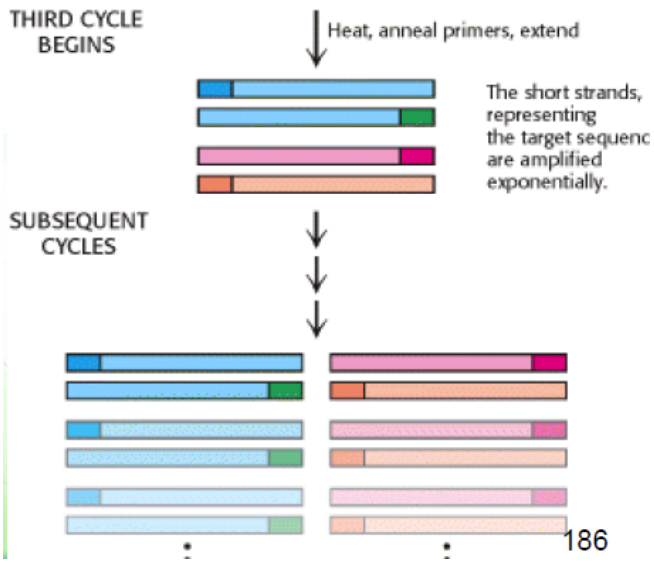
Cycle 2

The same 3 steps in cycle 1 are repeated.

Cycle 3

The same steps are repeated. By the end of this cycle, the target DNA begins amplification.

***These steps are repeated 25-30 times (25-30 cycles) to get over 250 million short products derived from each starting molecule.**



NOTE: -The products of each cycle serve as DNA templates for the next cycle, hence the term polymerase "chain reaction"

-Every cycle doubles the amount of DNA.

-Amount of DNA short products = $\mathbf{2^n}$, where n is the number of cycles.

- **Detection of the DNA fragment:**

Using Agarose gel electrophoresis and by staining.

NOTE: Staining → Adding color

Labeling: adding specific materials to release signals (like fluorescent tags)

- **Importance of Primers :**

Primers allow specific binding to the target DNA region. In other words, the specificity of amplification depends on the specificity of the primers to not recognize and bind to sequences other than the intended target DNA sequence.

*If annealing is done by one of the primers only, (one is attached and the other is not) ----> No amplification

*If the primers perform non specific binding (bind to two regions), amplification of both regions will occur; therefore we will get 2 fragments with different lengths. This is called **Non-specific Amplification.**

⇨ **How do we specify the region of amplification (amplification of a particular region)?**

Many factors contribute to this specific amplification, including the annealing temperature. If we decrease the temperature to 55°C for example, non-specific binding will occur. On the other hand, increasing the temperature will result in minimization (only specific binding will occur).

*The temperatures of DNA denaturation (step 1) and DNA synthesis (step 3) are constants.

⇨ Human DNA polymerase will be denatured at very high temperatures (in step 1); therefore the DNA polymerase used here is extracted from thermophilic bacteria called Thermus Aquaticus.

- **Advantages of this technique:**

1-Easy, fast, sensitive, robust

2-Discovery of gene families

3-Disease diagnosis

- **Disadvantages of this technique:**

1-Primers must be known
2-Contamination
3-Product length is limited (usually <5 Kb)
4-Accuracy is an issue
5-Not quantitative

- Uses: **Forensic medicine:**

-PCR amplification of multiple genes is being used to establish paternity and criminal cases. Although, an individual DNA profile is highly distinctive because many genetic loci are highly variable within a population.

(35-51 minutes)

**Please refer to the animations on the website for better understanding.

The End