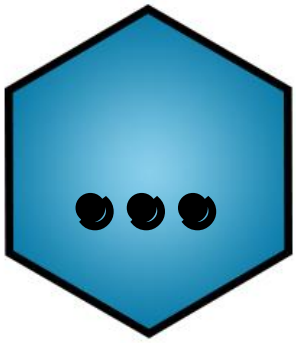


Faculty of medicine – JU2018

◉ Sheet ○ Slides

DONE BY

Reham Badayneh

CONTRIBUTED IN THE SCIENTIFIC CORRECTION

Alia Abbadi

CONTRIBUTED IN THE GRAMMATICAL CORRECTION

Alia Abbadi

DOCTOR

Mamoun Ahram

- We have talked about the old techniques of DNA sequencing, which are really slow (we can do these techniques for 1,2,3…10 samples but to finish the sequencing of the human genome it would take a lot of time)
- With the development of technology (computer science, Nano technology, informatics... etc.) we are now able to sequence whole human genome for a person in (24) hours instead of 15 years .

Now we will talk about this new DNA sequencing technique which is called

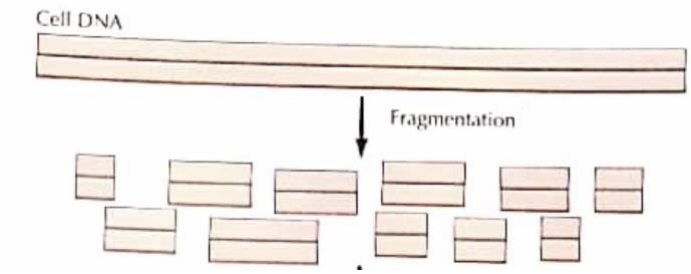
Next-generation sequencing (NGS)

-It is the fastest / best technique so far.

Note: please do watch this video for better visualizing of the process:
https://www.youtube.com/watch?v=shoje_9IYWc
And this one JUST to have real look at NGS:
https://www.youtube.com/watch?v=tuD-ST5B3QA&feature=youtu.be
NOTE: details in the videos are not required.

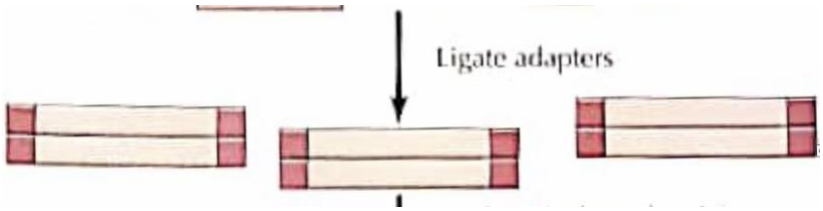
The process:

1- We **take the cellular DNA** (let's say it's about 300 billion pairs) → then it is **fragmented** into smaller pieces of <u>different</u> sizes (every piece will have 300-500 base pairs, as a result, we can calculate how many DNA fragments we can generate from the 300 billion base pairs, here it's about thousands to millions fragments of DNA will be generated).
<u>Remember: when we take a human DNA, we don't take only one DNA molecule we take thousands and hundreds of human DNA</u>.

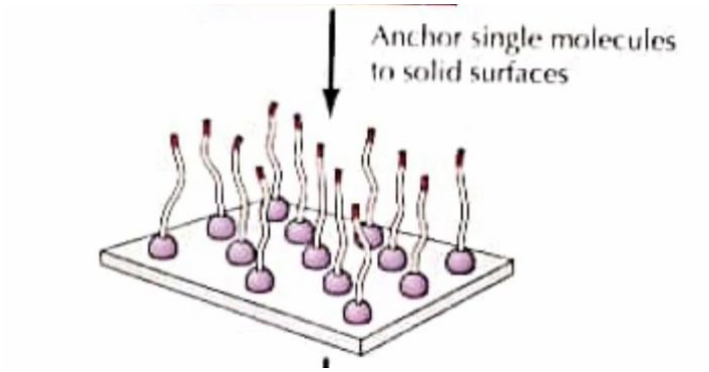


2- We take (isolate) each fragment and we **add (ligate) DNA adapters** at *both* ends of each fragment.

* *What is the adapter*? It's an oligonucleotide added on both ends of DNA fragments with a special sequence that **doesn't exist** in the human genome.



3- DNA fragments are then **attached to a solid surface** (this solid surface is very small and on which millions of fragments are attached)



4- Each DNA fragment is **amplified** like PCR – (polymerase chain reaction) using DNA polymerase.

- Note: DNA polymerase is an enzyme that needs:

A- **Primer:** to start synthesizing DNA since it can't initiate DNA synthesis by itself (the primer is complementary to the **adaptor**; thus it anneals - hybridizes- to it)

** All the fragments have the same adapter → All complementary strands need the same primer.*

B- **Substrate**: nucleotides

5- DNA polymerase starts to **add nucleotides** to the primer, BUT NOTICE: the idea here is that the nucleotides used in this method have 2 characteristics

1- Each one of the nucleotides gives a *special color* or *signal* as soon as it's added to (incorporated into) the DNA strand, *let's say for example T*

gives a green signal, A gives a red signal, G gives a yellow signal, C gives a blue signal.

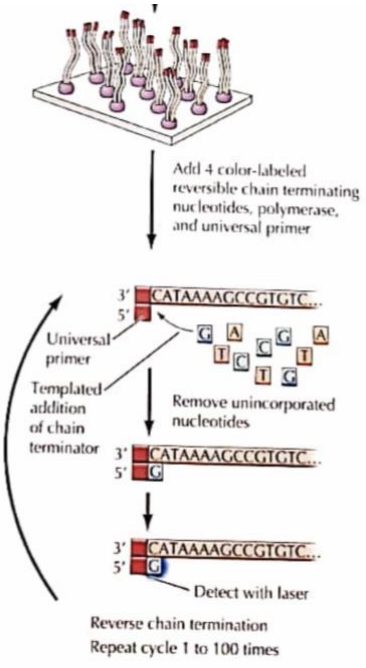
2- The nucleotides have **terminating ends**, which means that when one nucleotide is added another nucleotide cannot be added (attached) to the previously added one; because the previous nucleotide is modified in a way that ***does not*** allow DNA polymerase to add another nucleotide after it (on its 3' end)

➔ We have a laser & a detector, once a nucleotide is added, a signal comes out, the detector reads that signal (It reads the million signals coming from these million DNA fragments)

6- Unincorporated nucleotides **are removed**, and new nucleotides are added.

7- A **chemical modification** takes place on the previously added nucleotide allowing another nucleotide to be added by DNA polymerase, a signal comes out from the newly attached nucleotide and synthesis is stopped, and again there is a laser and a detector that reads the signals. THEN AGAIN, DNA polymerase cannot add another nucleotide unless the newly added one is chemically modified and then we can add another nucleotide, a signal comes out, the detector reads the signal …. And the cycle is repeated until **we read the sequence** of each one of these fragments (we have millions of DNA fragments thus, we will have millions of DNA sequences)

RECAP: **DNA fragmentation → Ligation of adaptors → Attachment to solid surface → Amplification of DNA fragments → Addition of first nucleotide to the primer by DNA pol → DNA synthesis is stopped and the added nucleotide gives a certain signal that is detected → chemical modifications of the previous nucleotide allow addition of another one → emits a signal to be detected → addition is resumed by chemical modifications and so on



8- Then comes **bioinformatics**: remember that we have millions of DNA fragments and these fragments can overlap allowing us to get the sequence of the desired DNA, so what bioinformatics do is that it reads all fragments and puts them together, into one large sequence of DNA.

- Sequencing of these millions fragments is done in about 24 hours and then the computer uses bioinformatics for analysis and giving the final results which takes a long time.

How detection works:

All fragments give out a signal per cycle at the same time and the detector reads it to produce a DNA sequence.

☺ *To understand let's check this example*

Look at figure below: Each box represents signals detected in a cycle and each colored circle represents the signal emitted from one DNA fragment. Now let's have a look at circle number 1.

In cycle 1: it produces a yellow signal (which represents G)

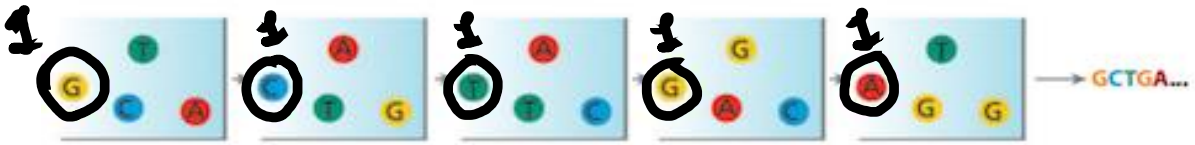
In cycle 2: it produces a blue signal (which represents C)

In cycle 3: it produces a green signal (which represents T)

In cycle 4: it produces a yellow signal (which represents G)

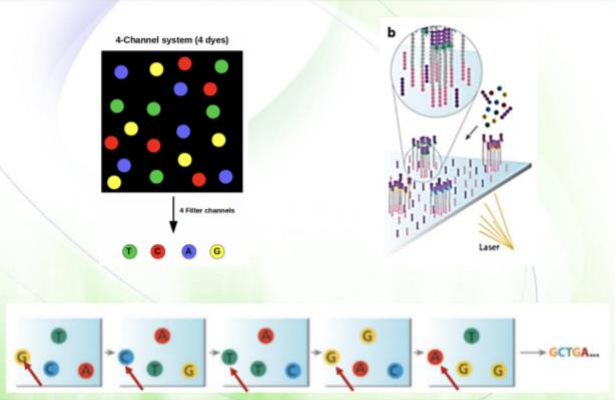
In cycle 5: it produces a red signal (which represents A)

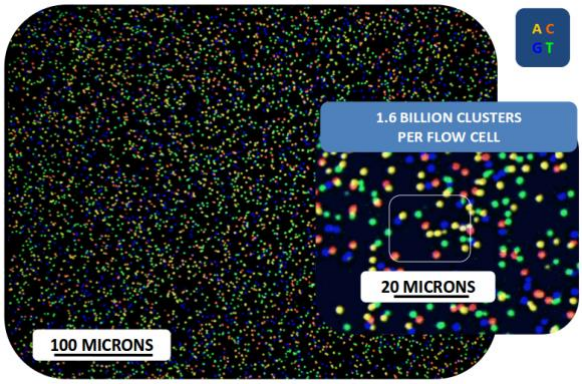
So after 5 cycles for example, the DNA sequence detected from this dot (circle) which represents one DNA fragment is GCTGA…



And this is true for the other DNA fragments.

Check the following pictures of NGS:





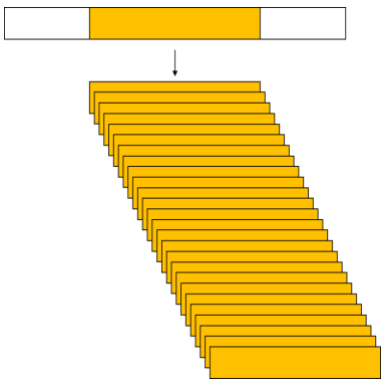
Polymerase chain reaction (PCR)

- This technique has revolutionized medicine and biology.
- It is a chemical ***chain*** reaction meaning that it is composed of sequential steps.
- It is a chemical reaction that is based on a polymerase.

  ☺ That's why it is called *polymerase* chain reaction
- It is a reaction that allows the DNA from a selected region of a genome to be amplified a billionfold, effectively "purifying" this DNA away from the remainder of the genome.

  For example, if I want to study a certain gene from the DNA (only this particular gene on this particular region on the DNA) or let's say that we have Dinosaur's DNA (we can get little amount of Dinosaur's DNA if any) or in forensic medicine we have a dead person that is for example dead for 10 years and we want to identify who is this person, so we extract very little amount of his DNA. So, in order to study these DNA molecules, we need to **amplify them** since we can't deal with this little amount of DNA. SO, WE USE PCR 😍

- PCR is **extremely sensitive**; it can detect a single DNA molecule in a sample. It also can give us multiple copies of a certain desired region of DNA. E.g. When we have a large DNA molecule and we are interested in studying only this small region (COLORED IN YELLOW).



❖ *Components of PCR (What do we need for the reaction)*

1- The **DNA template** (we start with it)

2- A **suitably heat-stable DNA polymerase** (the enzyme) which is obtained from microorganisms whose natural habitats is hot springs like the widely used **Taq DNA polymerase** which is obtained from a thermophilic bacterium, **THERMUS AQUATICUS** (it is thermostable up to 95°C)

Taq: **T** from Thermus referring to temperature

Aq from aquaticus referring to water

3- All **four deoxyribonucleoside triphosphate** (dNTPs) [the substrate for the enzyme]

4- A **pair of primers** (DNA polymerase cannot begin synthesis unless there is a primer) {the 15-25 nucleotides-long primers should surround the target sequence (*from both sides, one on each strand*)}

**** Why do we need 2 primers?*** ⍰

2 primers are needed to copy both DNA strands because the first step of PCR is actually to separate the 2 DNA strands, and since the goal of the technique is to produce more copies of DNA, we need both strands.

We have a primer attached to each strand on the **3' end:**

- Primers in PCR ***can be DNA or RNA primers***. DNA polymerase doesn't care whether it is RNA or DNA primer. But DNA primers are favored and most commonly used because they are more heat-stable. In our cells it is RNA primers just because the enzyme used to synthesize the primer (primase) adds ribonucleotides not deoxyribonucleotides.

❖ *The process* ☺

please refer to this video for full understanding: ***https://youtu.be/iQsu3Kz9NYo***

- We said that PCR is a chain reaction; meaning that the same set of reactions will be repeated over and over again, and each set of reactions is considered as one cycle (the reactions in cycle one will be repeated in cycle two and so on)
- We are interested as we said in *a certain region* on the DNA not all the DNA sequence.
- We must have all our components ready for the reaction:

  The DNA template, DNA polymerase, pair of primers, all four nucleotides.

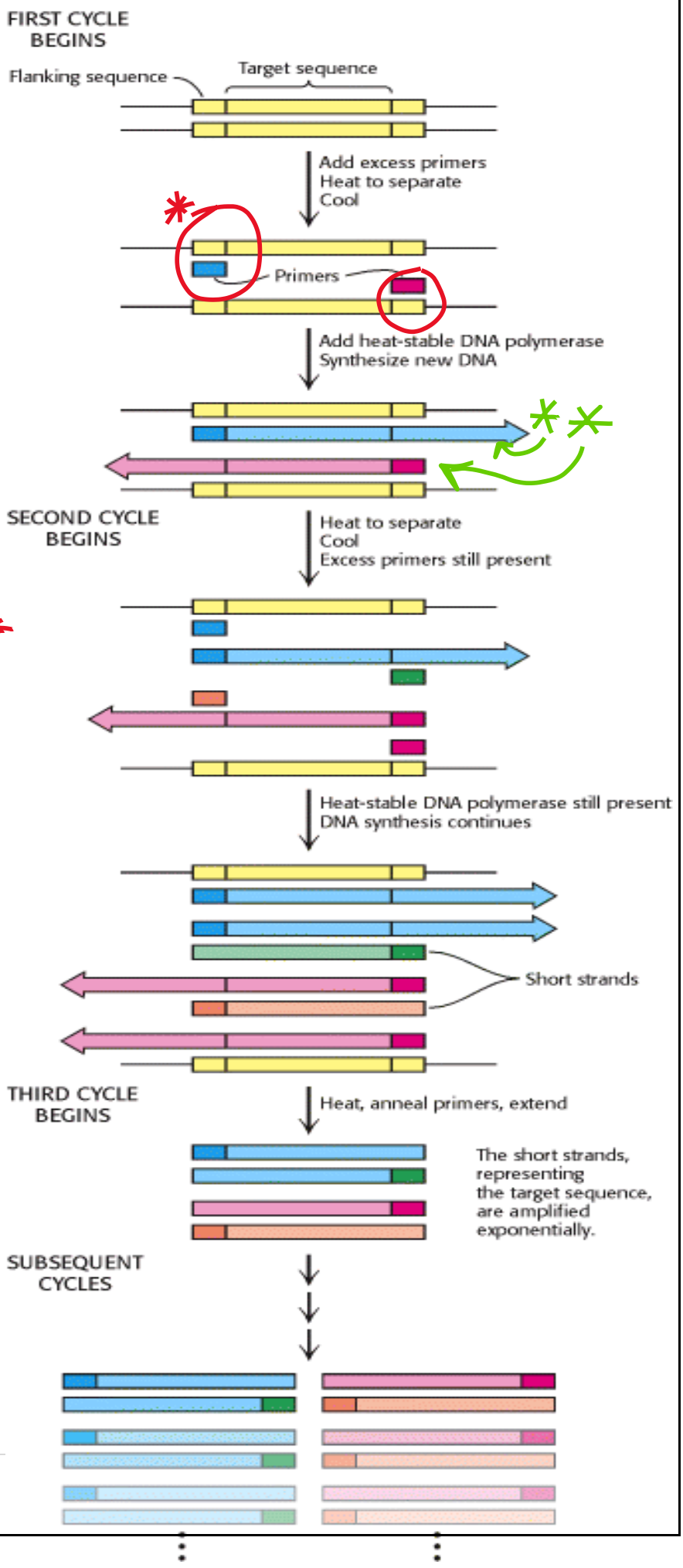
STEPS: ***Note: make sure to look at the image as you read the steps***

1- **The first step** in the first cycle is ***DENATURATION*** which starts at **95°C** [DENATURATING TEMPERATURE]; where the DNA template is denatured into two single-stranded molecules.

2- **The second step** in the first cycle is ***ANNEALING***: the temperature is lowered to almost **60°C** [ANNEALING TEMPERATURE] (ranges from **50°C to 70°C**); where primers anneal (bind) to the ends of the region that I want to amplify on both strands {one primer for each strand}

3- **The third step** in the first cycle is ***DNA SYNTHESIS*** the temperature rises to (**72°C**); which is the optimal temperature for the DNA polymerase (this is where it works at best). [Synthesis of the DNA occurs on both strands].

4- NOW, we have these DNA products [PCR products] but that's not what we want, we *only* want a certain region,



so, we move on to cycle number 2 repeating the same steps. Then, we move on to cycle number 3 again repeating the same steps. *NOW*, we will have the desired product (**which are copies of the DNA region that we want**), but we still need *more* copies of it, so the cycle is repeated until we reach about (**20-30 cycles**), by then we will get a large number of copies. (After 30 cycles, there will be over 250 million short products derived from the *starting molecule*!) these copies can be used for multiple things, which will be discussed soon.

Notes:

- **20-30** cycles are required for DNA amplification, and each cycle *doubles* the amount of DNA.
- It's called "chain reaction" because the products of each cycle serve as DNA templates for the next one, plus the same steps are repeated in all cycles.
- ***Now let's think about something; PCR cycles occur at very high temperatures (95°C, 60°C, 72°C). How can DNA polymerase handle this high temperatures and not get denatured?***

The DNA polymerases we use in this case are ***heat-stable DNA polymerases*** which can handle the high temperatures without being denatured. Remember: we said that for example *Taq DNA pol* is thermostable up to 95°C and its optimal temperature is (72°C).

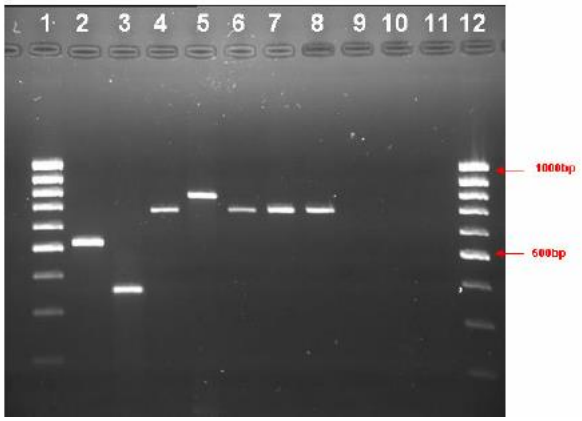
- ***The PCR device used takes only one hour to complete the reaction nowadays; whereas long time ago it took more than 16 hours to complete the reaction; this depends on the metal used. How??***

The trick is ***how fast*** this metal absorbs and loses heat. Nowadays, they use special metals that gain and loose heat rapidly. So, temperature rises up to 95°C and then drops to 60°C and then goes up again to 72°C *really fast*.

Detection of DNA fragments:

☺ after doing multiple different PCR reactions producing different DNA fragments, these DNA fragments can be **visualized as a discrete band** of special size *by agarose gel electrophoresis*.

☺ Each PCR reaction produces lots of copies of DNA fragments with *exactly the same* size represented as a single band. (check bands from 2-8)

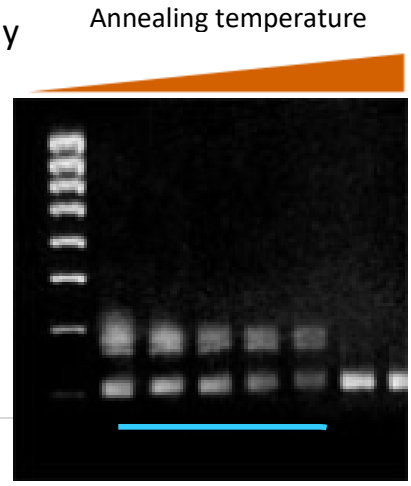


This means that PCR is specific (every reaction produces just one single band), but what determines this specificity?

- The specificity of amplification depends on the **specificity of the primers** to not recognize and bind to sequences *other* than the intended target DNA sequences. (where the primer anneals)

☺ Annealing temperature is important in making the primers specific. HOW??

Primers can anneal to DNA strand either specifically (perfect annealing or hybridization) or non-specifically (imperfect annealing or hybridization) depending on the annealing temperature; so we see here in the picture, ***at lower temperatures*** (about 50°C) we have 2 products (2 bands) this reflects that **imperfect hybridization** occurred by



binding to a DNA sequence *other than the target one*, now by **increasing** the temperature up to almost (60°C) [*which is the temperature that primers anneal perfectly and imperfect annealing doesn't take place*] gradually we will have a single PCR product. But, increasing the temperature more than a certain limit that will cause the primers to be denatured.

Note: we need both primers to anneal perfectly, because if one of them anneals perfectly and the other imperfectly it is useless because the DNA molecule or the particular region won't be amplified from just 1 strand.

❖ *USES OF PCR*

1- **Discovery of gene families**, by using nonspecific annealing which results in nonspecific amplification. How?

Remember in nonspecific amplification, we're getting multiple products from ***the same primers***, which means that these products *including the specific one* have ***similar DNA sequences***!

Example: if we're identifying the gene of histone (H2A) and design primers for that gene but we keep getting a nonspecific product which is histone (H2B) because these genes share some similar DNA sequences, this way we recognized gene families!

*Note: nonspecific annealing occurs at relatively high temperatures, just not high enough as in specific annealing.

2- **Disease diagnosis**

By amplifying a region and sequencing it, we can recognize if there is a mutation.

3- **Paternity and criminal cases.**

An individual DNA profile is highly distinctive because many genetic loci are highly variable within a population. *Example*: remember VNTRs? If we amplify the region that contains VNTRs, we'll get different products with different individuals depending on ***the length of the VNTRs*** in that specific region.

4- **Viral and bacterial load: the quantity of virus in a given volume. How? By quantitative PCR.**

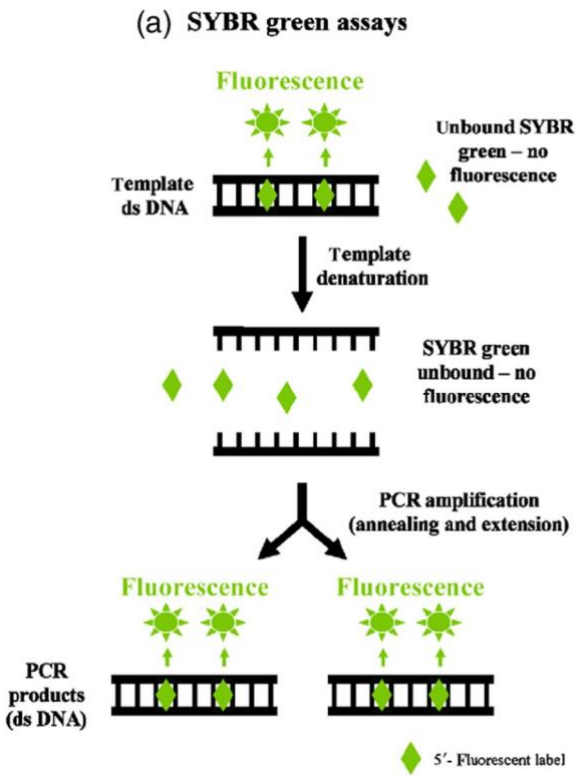
Check this scenario: we have 3 people suffering from flu, the doctor suspected that they suffer from avian flu, so he performed PCR on the DNA of these 3 people and added a specific primer that would anneal to the DNA of the virus that causes avian flu. The results were positive for 2 of these people, meaning that they DO suffer from that flu, but the third one DOES NOT, because the target viral DNA sequence wasn't present in his cells.
BUT notice: this way we were just able to know if the virus is there or not, we didn't know which person has more of the virus in his body. Quantitative PCR can help in this situation.

Quantitative PCR

Useful video: https://youtu.be/EaGH1eKfvC0

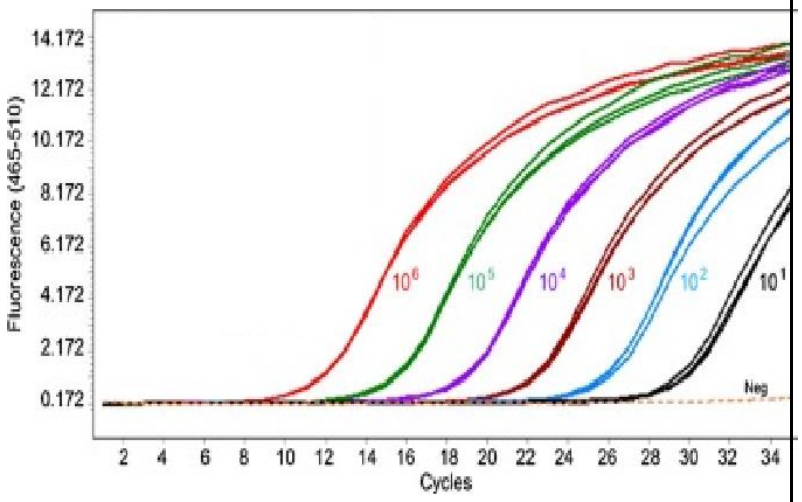
📝 Components needed for the reaction:

1- DNA polymerase
2- DNA template
3- DNA primers
4- All four deoxyribonucleotides (dNTP's)
5- What is special about quantitative PCR is that we add a small molecule called **(SYBR green)** that binds to double stranded DNA and fluoresces only when bound. *Let's say* that we have **1** double stranded DNA molecule, SYBR GREEN binds & gives a signal. If we have **2** DNA molecules, more SYBR GREEN will bind to the molecule producing stronger signals. So, the strength of the signal that is generated from SYBR GREEN molecules reflects ***how much DNA molecules is in our sample*** {the more DNA molecules there is, the stronger the signal produced by SYBR GREEN}
*This also means that more amplification → stronger signal



NOTE: ***The PCR instrument has a detection limit,*** *it can't detect 1 or 2 molecules, for example, let's say that it can detect DNA if there is (10,000) DNA molecules or more, so a sample containing less than 10,000 DNA molecules cannot be detected, that's why we amplify and repeat the cycles multiple times until the signal can be detected.*

Now, let's look at the following graph:



- On the X-axis is the number of PCR cycles done
- On the Y-axis is the intensity of the signal detected
- We have 6 samples
  - **1-** Sample number 1 (red): the signal is detected on cycle number 10
  - **2-** Sample number 2 (green): the signal is detected on cycle number 14
  - **3-** Sample number 6 (**the black**): the signal is detected on cycle number 30

➔ This means that the sample that contains higher starting material (MOST DNA) is detected first, i.e. it needed a *smaller number* of cycles and amplification to give a detectable signal.

Sample 1 (most DNA) is detected first on cycle 10

Sample 6 (least DNA) is detected later on cycle 30

Remember that with every PCR cycle the amount of DNA is doubled. And again, in **sample 1** the signal is detected on the 10^{th} cycle because *now* this sample has got *enough* molecules that give a detectable signal. In **sample 6** the signal is detected on the 30^{th} cycle because on *that cycle* it reached the detection limit. So, sample 1 has more DNA molecules because it reached the detection limit first by needing a lower number of cycles to do so.

Another example for clarification: look at the following table; assuming that we have 2 samples, and the PCR instrument requires at least 10,000 DNA molecules for them to be detected.

Sample number	**Original number of DNA molecules in the sample**	**Number of DNA molecules After cycle #1**	**Number of DNA molecules After cycle #2**	**Number of DNA molecules After cycle #3**	**Number of DNA molecules After cycle #4**
1	10	20	40	80	160
2	1000	2000	4000	8000	16000

AND NOW after the 4th cycle, a signal is detected from sample 2 but no signal is detected from sample 1; because sample 2 has reached 10,000 molecules, but sample 1 requires more cycles! This shows that sample 2 has got more starting material than sample 1.

- So quantitative PCR gives us:

1- **A relative quantitation of amount of DNA in a sample (which sample has more DNA)**
2- **It can also give an exact measure. For example It can show how much viral DNA molecules exist in the sample (10, 50..) but this won't be discussed now ☺**

Back to our example, using the quantitative PCR, we can know which person has more of the avian flu virus and requires immediate treatment.

End of sheet 29

good luck