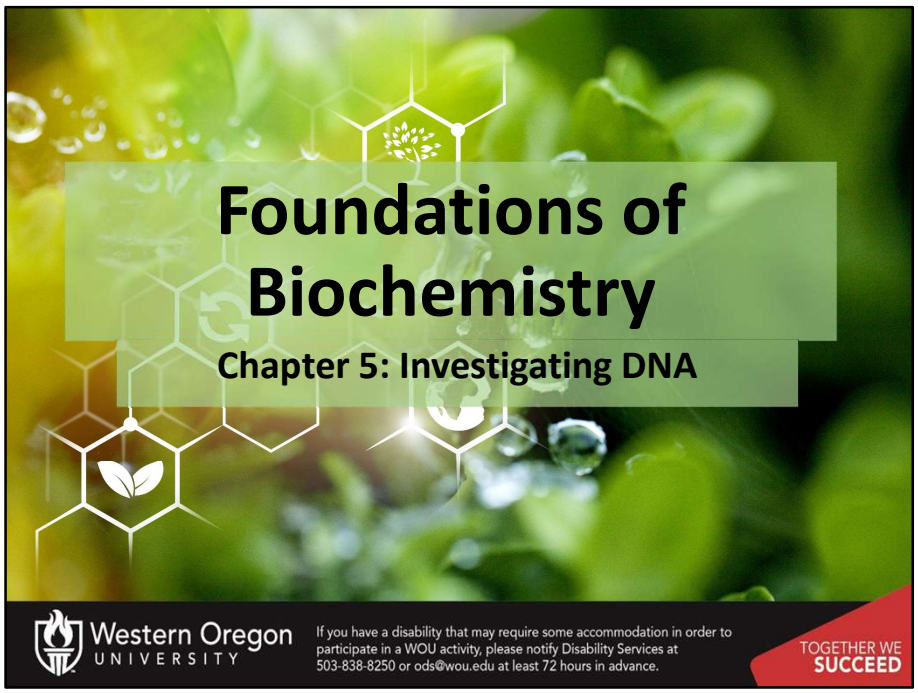
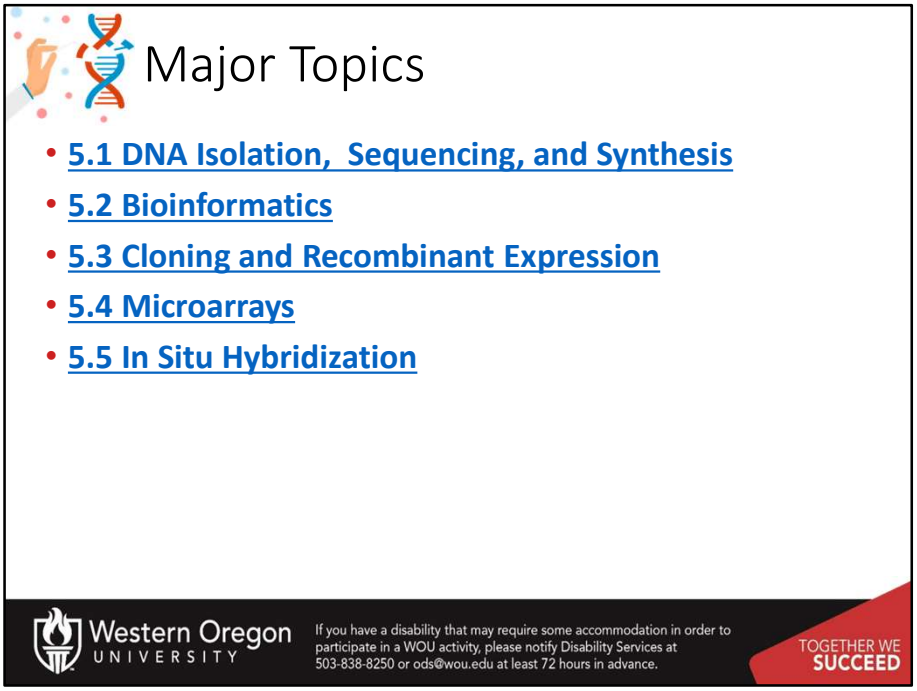
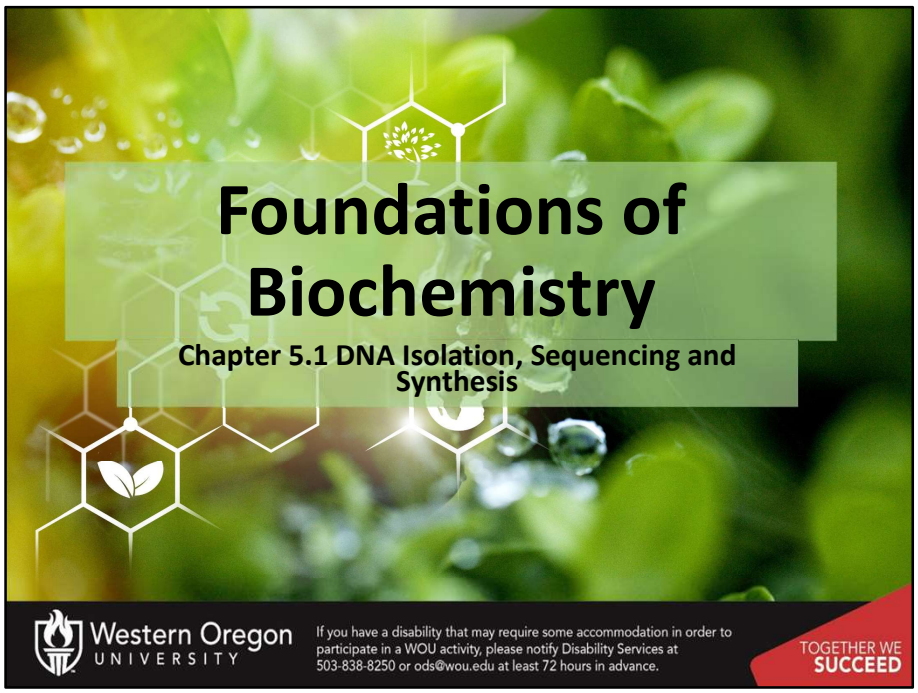


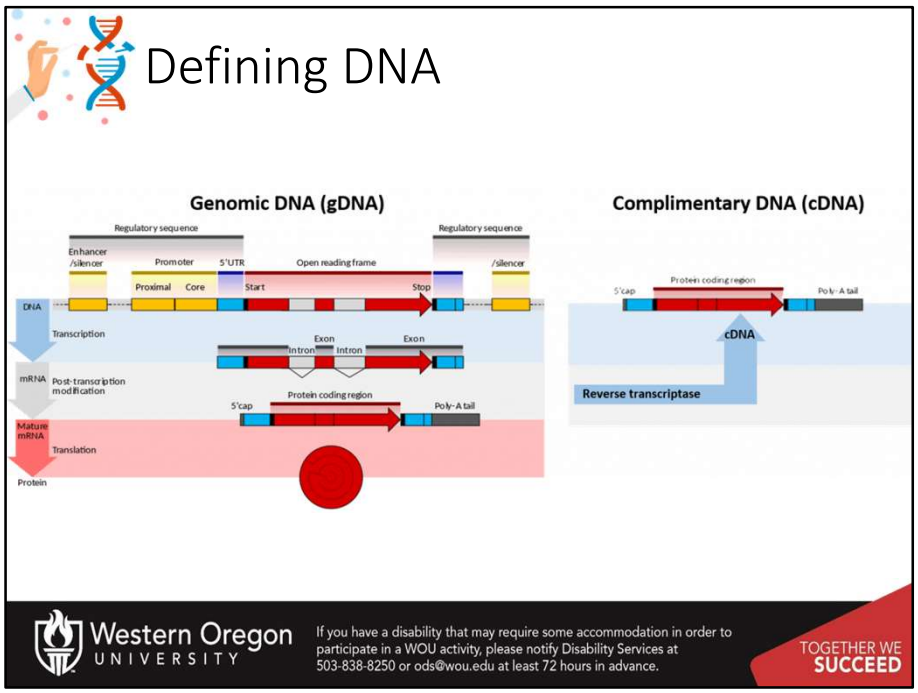
In this chapter, we will be focusing on techniques used in the laboratory to study DNA and RNA.



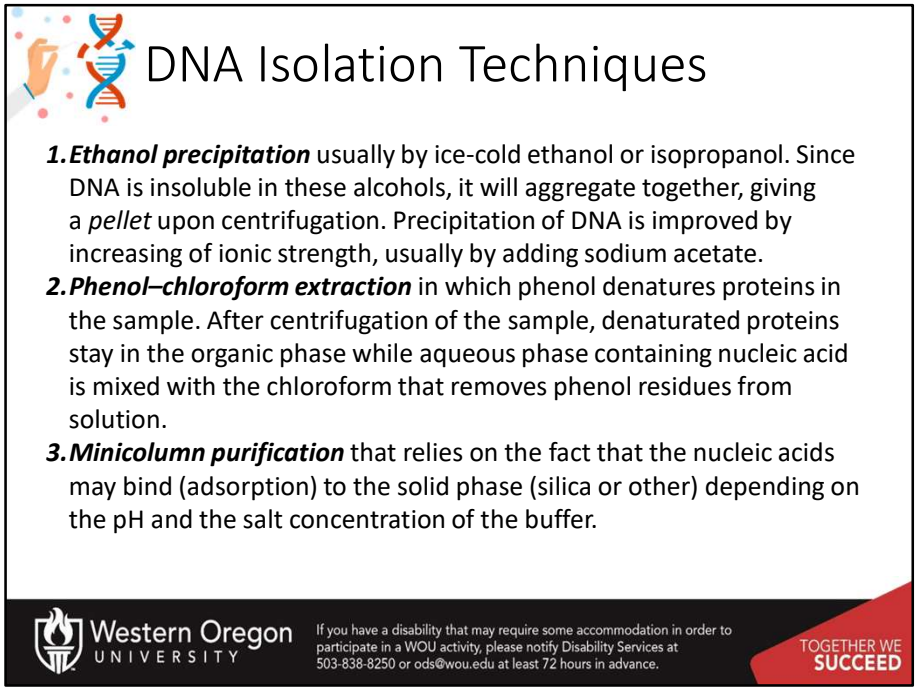
You will become familiar with techniques used to isolate and purify DNA, how DNA is sequenced, and synthesized, some of the major bioinformatics tools used to analyze genome sequences, the fundamentals of DNA cloning and recombinant protein expression, microarrays and in situ hybridization techniques.



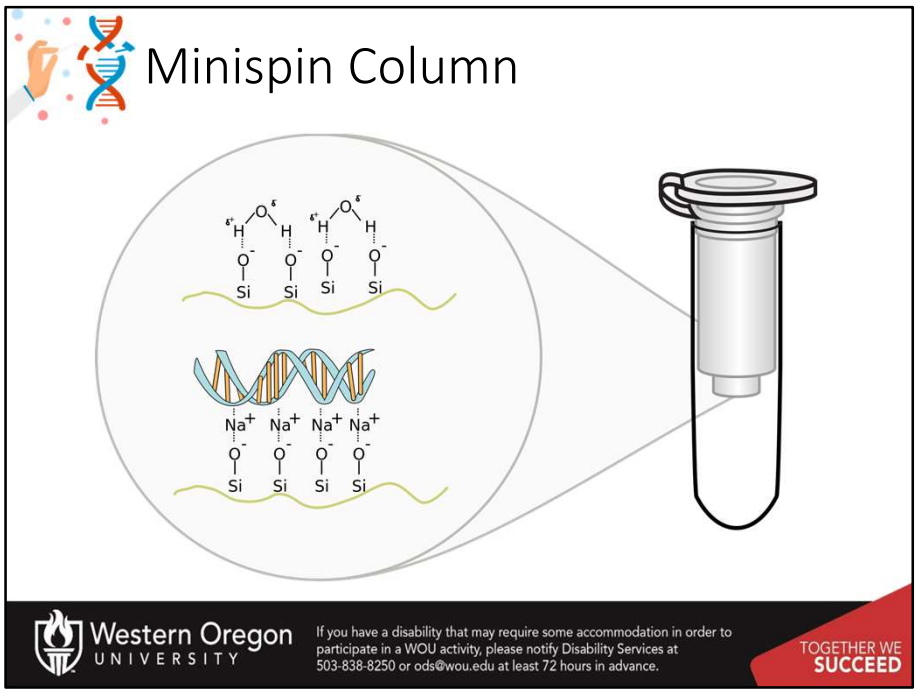
In this first section, we will cover DNA isolation techniques, sequencing and synthesis.



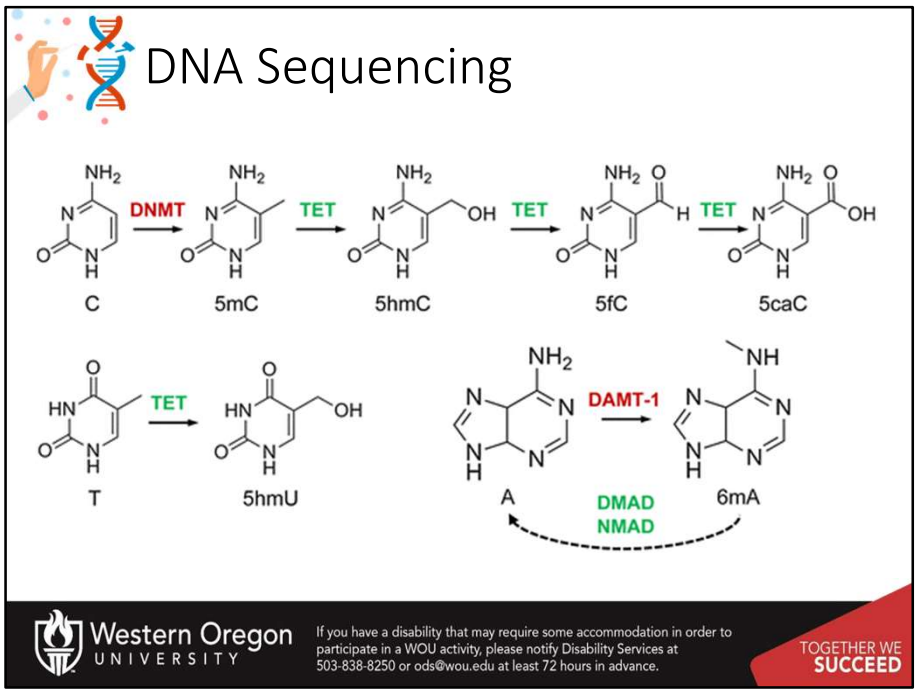
Genomic deoxyribonucleic acid (gDNA) is chromosomal DNA, in contrast to extra-chromosomal DNA such as that found in the mitochondria of mammals or plasmid structures in bacteria. Plasmids will be discussed in more detail when we learn about cloning and recombinant expression. The techniques used to isolate DNA are different, if you are planning to isolate genomic DNA or something much smaller, such as plasmid DNA. Furthermore, gDNA from eukaryotic cells will contain all of the intron sequences within it. Thus, it is often quite difficult to work with directly. It is more common to want to use complimentary DNA, or cDNA, especially if you are planning on doing a cloning experiment. cDNA is DNA synthesized from a single-stranded RNA (e.g., messenger RNA (mRNA) or microRNA) template in a reaction catalyzed by the enzyme reverse transcriptase. Reverse transcriptase is an enzyme found in retroviruses such as HIV, that have RNA as their core genetic material. Upon entering the host cell the RNA is reverse transcribed to produce a copy of cDNA that can then integrate into the hosts genomic DNA. In biotechnology, reverse transcriptase is often used to create cDNA from the mRNA expressed in specific cells or tissues. In this way, the eukaryotic genes can be cloned without any introns housed in the structure.



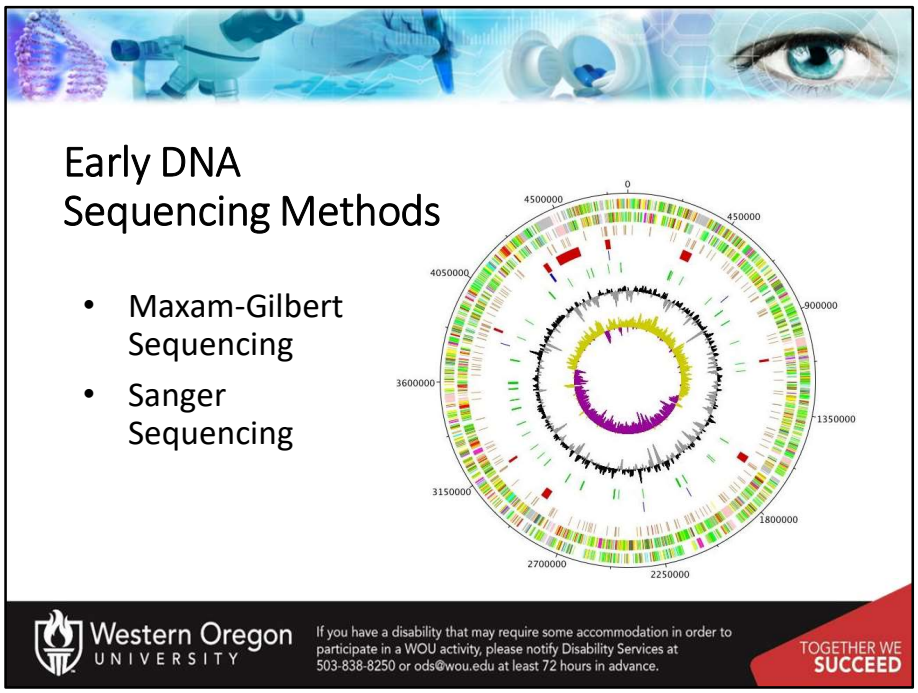
Depending on one's needs, there are three major ways of isolating both DNA and RNA, depending on what you want to accomplish. They are Ethanol precipitation, Phenol-Chloroform Extraction, and Minicolumn purification. Note that special care must be taken when working with RNA. It is much more labile than DNA and will often degrade very easily. Thus, it often requires working with samples on ice and requires the use of RNAse inhibitors, to prevent degradation. Large nucleic acid molecules are insoluble in ethanol (especially at a concentration of 70%. Precipitation is increased by the addition of sodium acetate to help increase the ionic strength of the solution. It is often used when cleaning DNA samples in between reaction steps. Especially, if the buffer needs to be changed from one reaction to the next. It is usually not sufficient to purify DNA from cellular extract, although crude extracts can be prepared this way. Phenol-chloroform is a very efficient liquid-liquid separation technique used to purify genomic DNA and RNA samples. In fact, DNA and RNA can be separated from each other by using different pHs of the aqueous layer. However, these solvents are quite toxic and have given way to minicolumn purification techniques which are often much faster, safer, and can produce highly purified samples.



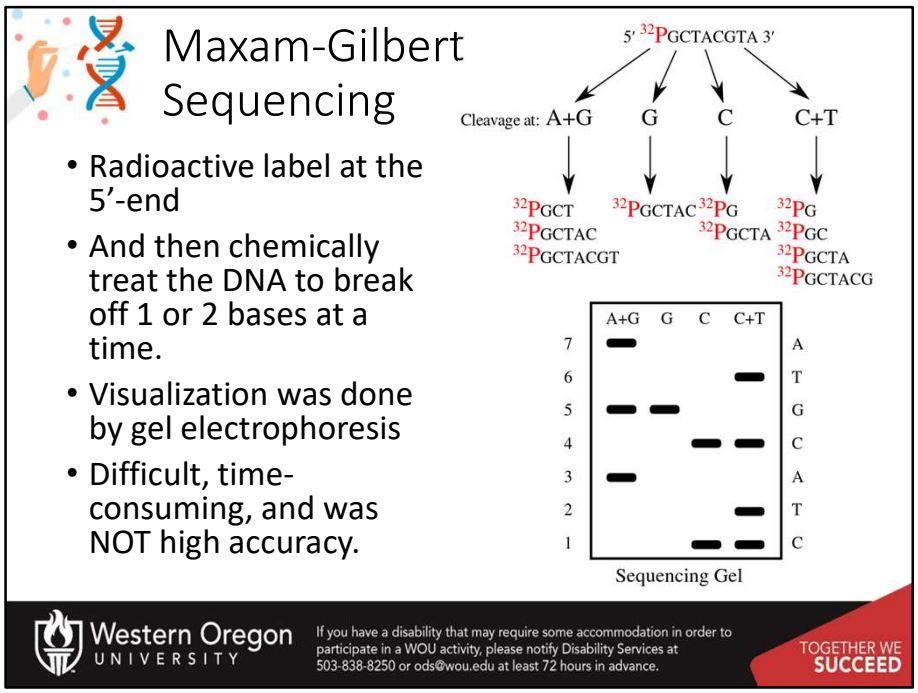
Spin column-based nucleic acid purification is a solid phase extraction method to quickly purify nucleic acids. This method relies on the fact that nucleic acid will bind to the solid phase of silica under certain conditions and then released when those conditions are altered. For binding, a buffer solution is added to the DNA lysate along with ethanol or isopropanol. This forms the binding solution. The binding solution is transferred to a spin column and the column is put in a centrifuge. The centrifuge forces the binding solution through a silica gel membrane that is inside the spin column. If the pH and salt concentration of the binding solution are optimal, the nucleic acid will bind to the silica gel membrane as the solution passes through. To wash non-specific cellular components from the column, the flow-through is removed and a wash buffer is added to the column. The column is put in a centrifuge again, forcing the wash buffer through the membrane. This removes any remaining impurities from the membrane, leaving only the nucleic acid bound to the silica gel. To elute, the wash buffer is removed and a low salt elution buffer (or simply water) is added to the column. The column is put in a centrifuge again, forcing the elution buffer through the membrane. The elution buffer displaces the nucleic acid from the column allowing it to be collected in the flow through.



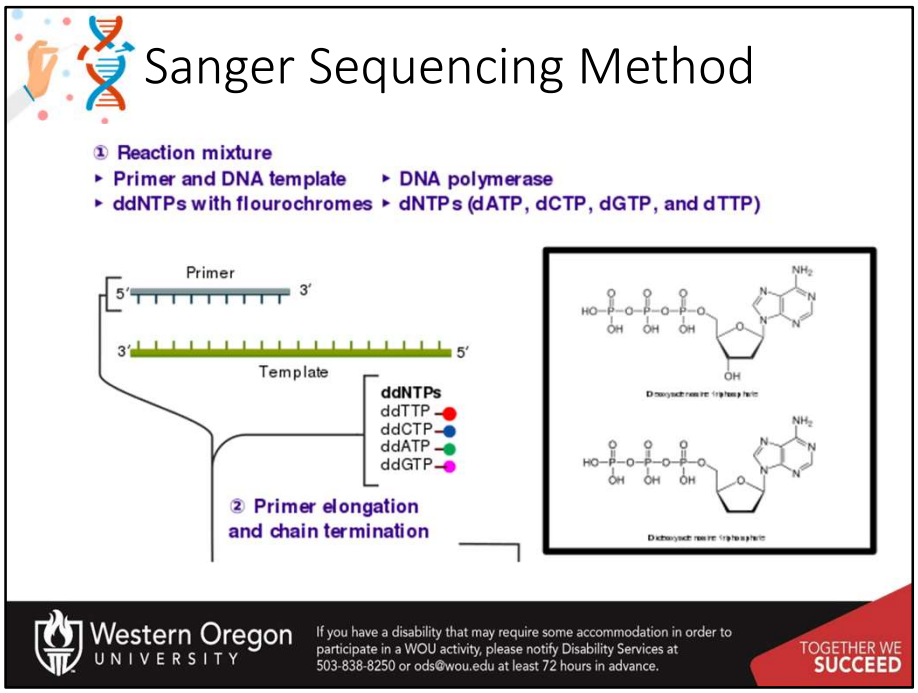
One factor that has made the development of sequencing techniques difficult is that the 4 nucleotide bases that are found in DNA can have chemical modifications associated with them, sucha as methylation, hydroxylation, and even oxidation to form carboxylic acids. These modifications make detection more challenging that simply recognizing four different base options..



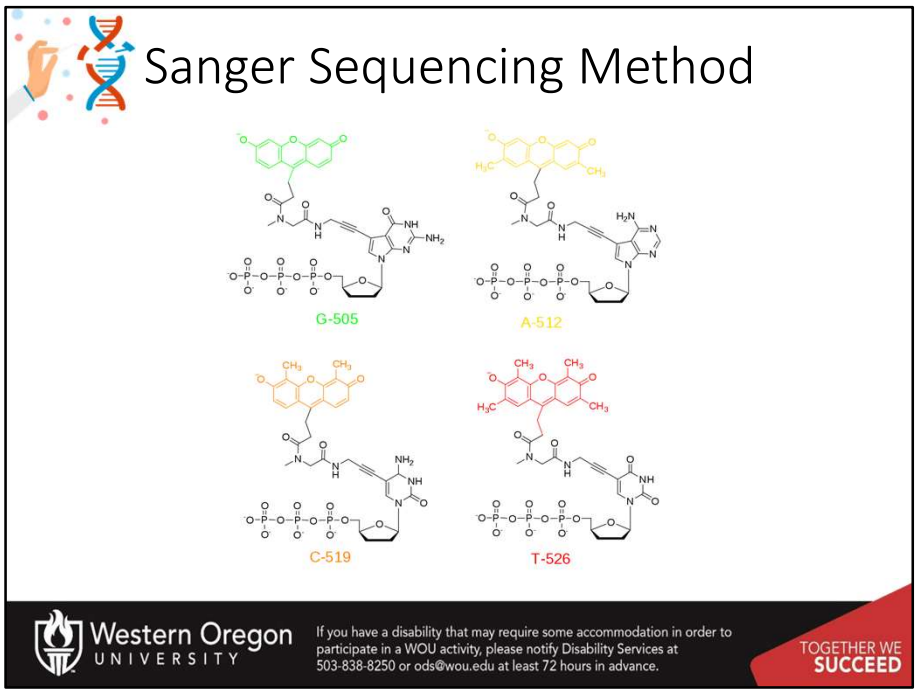
However, DNA sequencing techniques have been incredibly refined and advanced over the last 35 years, such that the sequencing of genomes is now commonplace and extremely quick! In fact, you can even sequence a genome from as little as a single cell! We will take a short walkthrough of the history of sequencing. Early DNA sequencing methods included Maxam-Gilbert Sequencing and Sanger Sequencing.



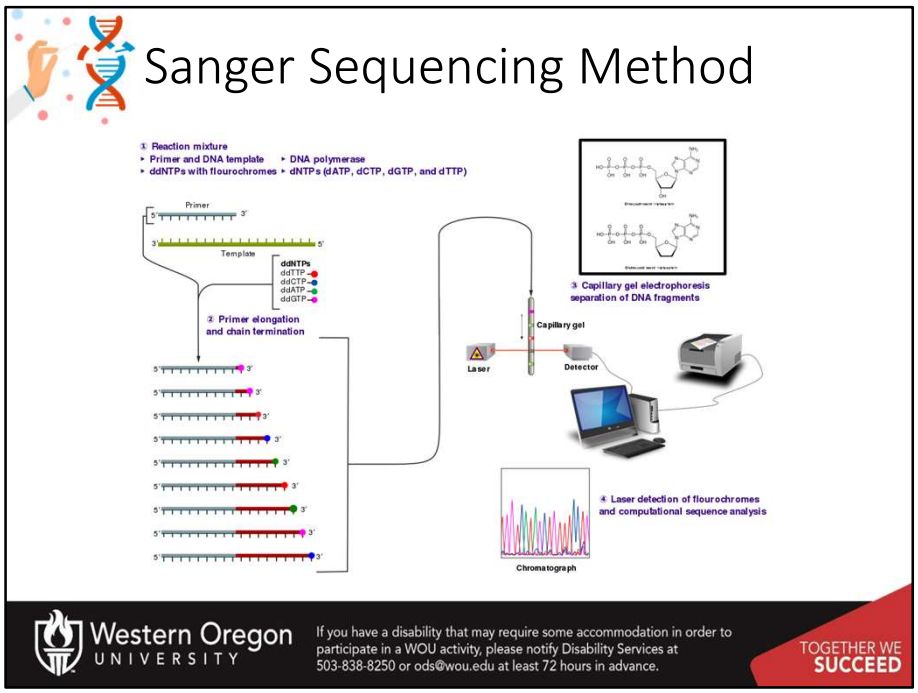
Maxam-Gilbert Sequencing was the first technique developed and used radioactive labels on the DNA, followed by chemical degradation. It was time consuming and not highly accurate. Thus, it was quickly replaced with the more efficient Sanger Sequencing Method.



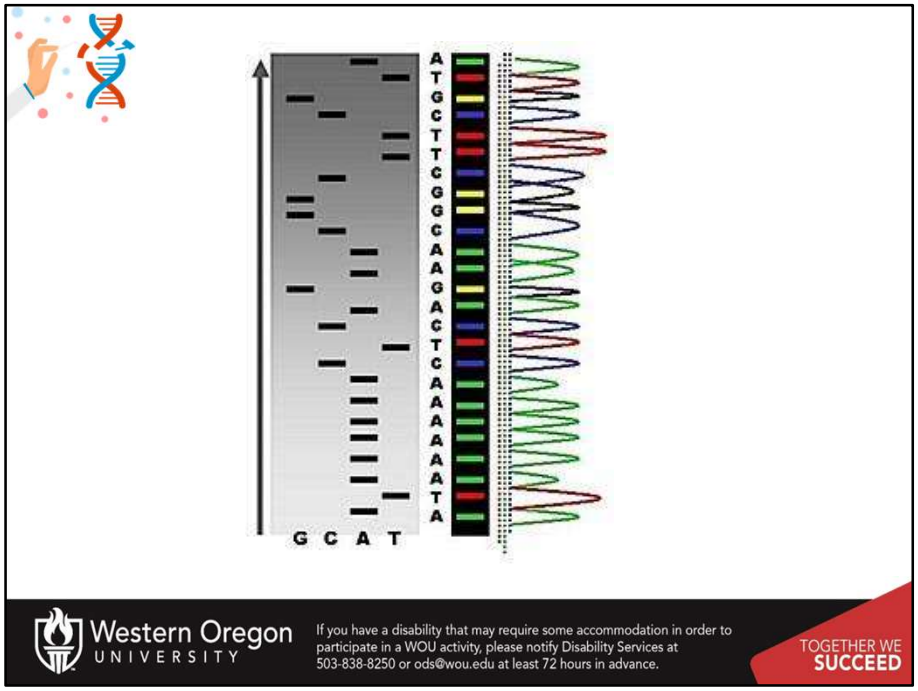
Sanger sequencing is also known as chain termination sequencing. In this method, the DNA of interest is used as a template. It is separated or denatured into single strand and a small primer is used to start a sequencing reaction with a DNA polymerase enzyme and the four standard nucleotide triphosphates (A, G, C, T) in the reaction mixture. Fluorescently labeled dideoxy nucleotide triphosphates are also added in a small concentration compared with the regular nucleotides. The dideoxy nucleotides do not contain a 3'OH. Thus, when they are added during the polymerization reaction, synthesis of DNA will terminate.



This figure shows the different chromophores that can be added onto the dideoxy nucleotide triphosphates. Each one is a different color, which aids in the detection of the product.



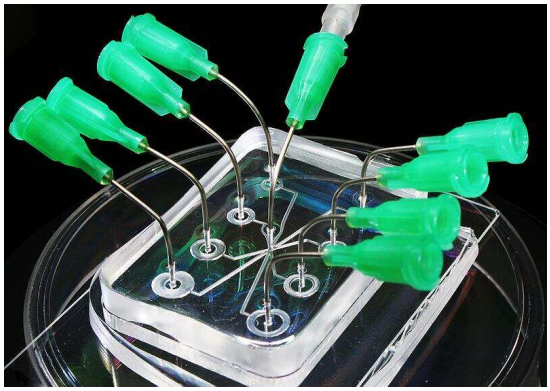
Because the dideoxy bases are added in small concentration in comparison to the other four regular bases, overtime, the mixture will have many different sized termination fragments, whenever that dideoxy base is incorporated. This mixture of products can then be separated using capillary gel electrophoresis. The samples are then passed by a detector when they elute from the column and the fluorescent peaks are measured. This gives a sequencing chromatogram, that is representative of the complementary sequence to the template DNA strand used in the reaction.



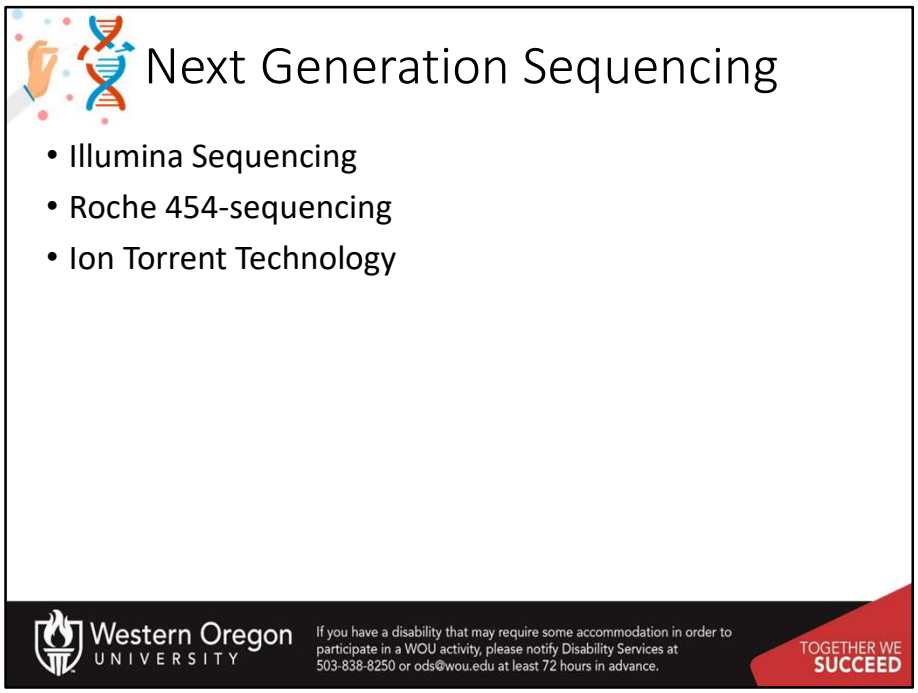
Very early Sanger sequencing used radioactively labeled dideoxy nucleotides and the termination products were visualized on a PAGE gel, as shown here. Fluorescent labels have made this technique much safer and user friendly. It is now automated and produces chromatograms that are easily read, according to their color. This was the type of sequencing used to sequence the first human genome in 2003 at a cost of $2.7 billion dollars. Today newer sequencing techniques have increased sequencing speeds and have reduced costs, such that sequencing a human genome costs roughly $1,000 (this is close to being usable for patients receiving medical treatments and could provide much more tailored medical care in the near future). These newer DNA sequencing methods are called Next Generation Sequencing and use microfluidic sequencing techniques.

Microfluidics

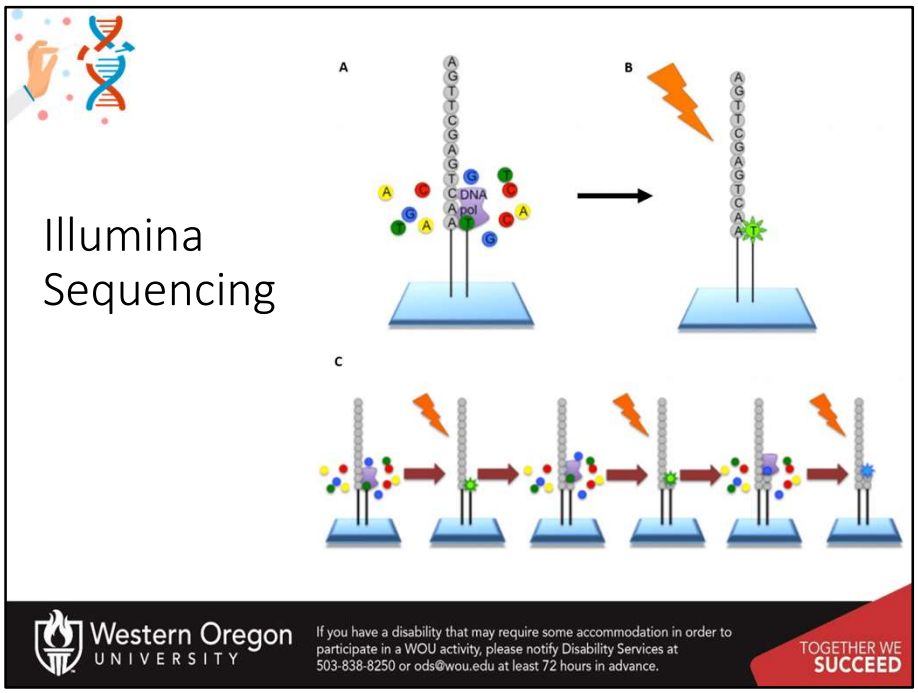
- Lab-on-a-chip Technologies



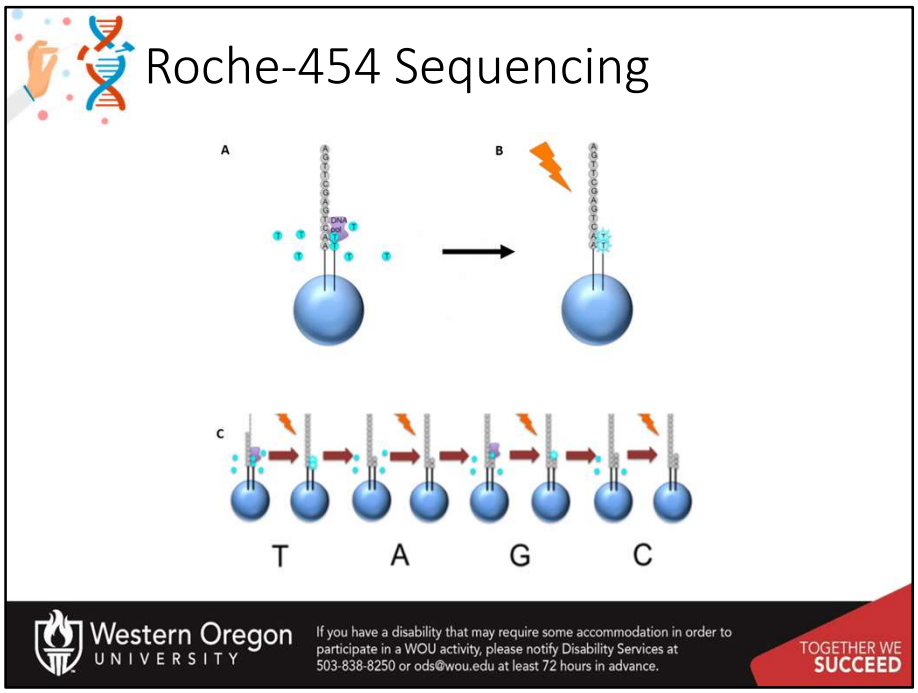
The advent of microfluidic technologies have really enabled the improvement of sequencing technologies. Essentially sequencing reactions can be run in small sample channels within the device that are about the size of a human hair. Microfluidics were first done using Sanger sequencing technologies.



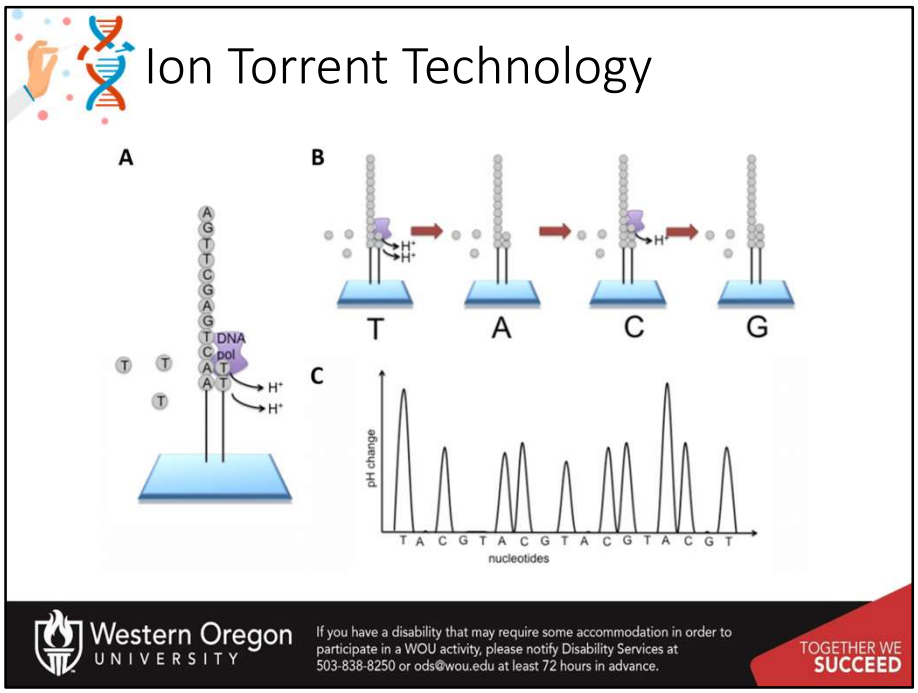
Currently there are three major Next Generation Sequencing techniques in use (although there are others as well.) Here we will focus on illumina sequencing, Roche 454-sequencing , and ion torrent sequencing technologies.



In Illumina sequencing the genome is cleaved into 100 – 150 bp reads, where individual small fragments of DNA are bound within tiny wells inside a tiny microchip. The wells with the DNA fragments are flooded with nucleotides and DNA polymerase. These nucleotides are fluorescently labelled with each color corresponding to a specific base. The reactions also have a terminator present, so that only one base is added at a time. An image is taken of the slide. In each reaction location, there will be a fluorescent signal indicating that a specific base that has been added. The data is recorded and the slide is then prepared for the next cycle. In preparation, the terminators are removed, which will allow the next base to be added, and the fluorescent signal is cleaved, preventing the fluorescent signal from contaminating the next image. The process is repeated, adding one nucleotide at a time (G, A, T, or C) and imaging in between. All of the sequence reads will be the same length as single bases are added at each cycle. All of the fragments will be the same length as only 1 base is added at a time.

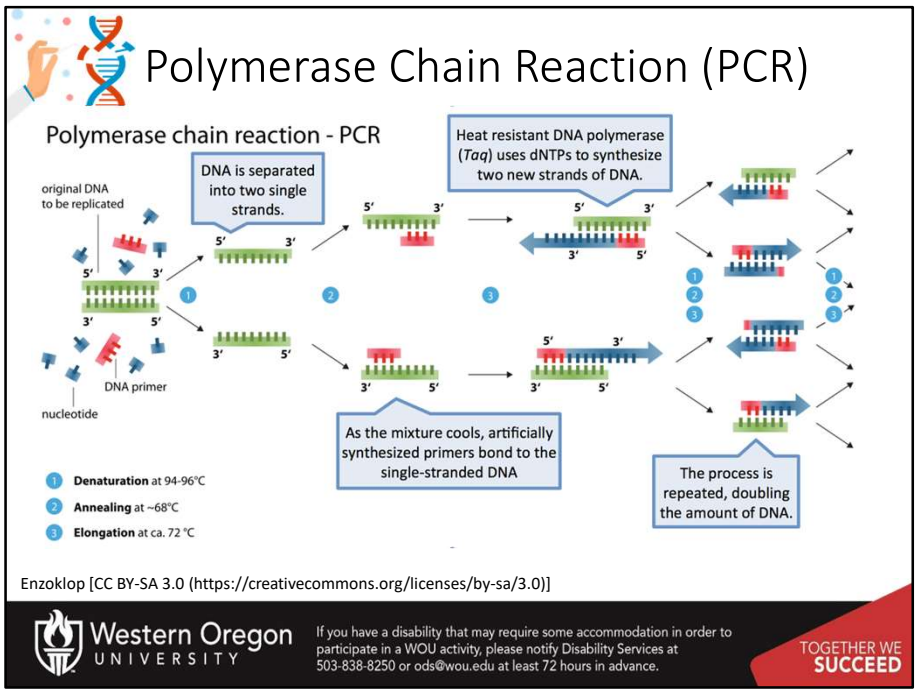


Roche 454-Sequencing is similar to the Illumina process but can sequence much longer reads. Like Illumina, it does this by sequencing multiple reads at once by reading optical signals as bases are added. As in Illumina, the DNA or RNA is fragmented into shorter reads, in this case up to 1kb (1,000bp), rather that 150. Generic adaptors are added to the ends and these are annealed to beads, one DNA fragment per bead. The fragments are then amplified by PCR using adaptor-specifc primers. Each bead is then placed in a single well of a slide. So each well will contain a single bead, covered in many PCR copies of a single sequence. The wells also contain DNA polymerase and sequencing buffers. Once the PCR product is attached to the bead, the slide is flooded with one of the four NTP species. Where this nucleotide is next in the sequence, it is added to the sequence read. If that single base repeats, then more will be added. So if we flood with Guanine bases, and the next in the sequence is G, one G will be added, however if the next part of the sequence is GGGG, then four Gs will be added. The addition of each nucleotide releases a light signal. These locations of signals are detected and used to determine which beads the nucleotides are added to. The NTP mix is washed away. The next NTP mix is now added and the process repeated, cycling through the four NTPs. All of the sequence reads from 454 sequencing will be different lengths, because different numbers of bases will be added with each cycle.



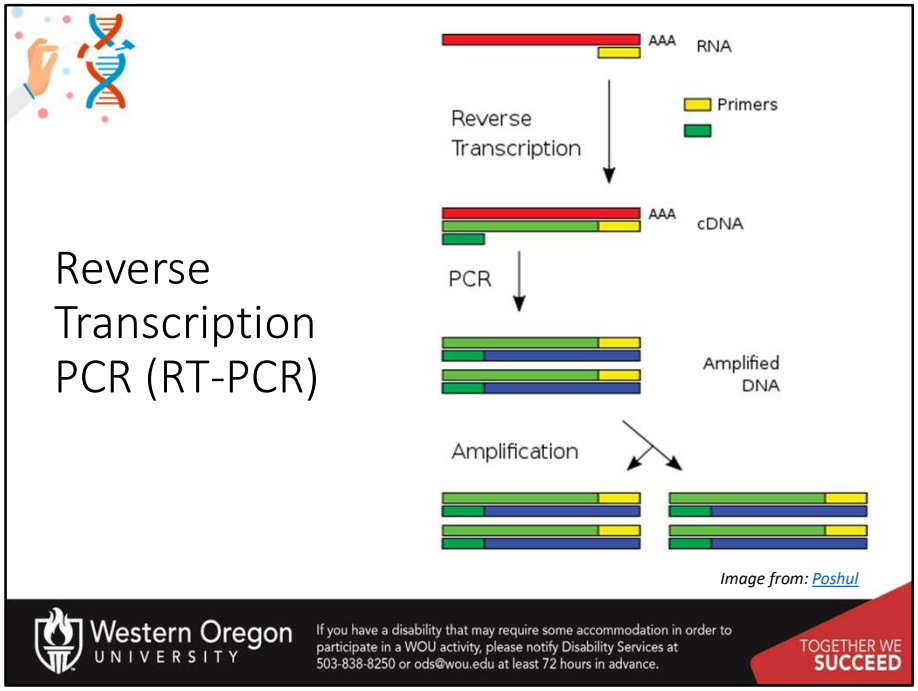
All of the technologies thus far, have required either radioactive or fluorescent labels to read the incorporation of a base into the molecule. This requires a detector meaning that the machine is quite large and cannot be used in a mobile way. Ion Torrent Technologies are changing this possibility, making large scale DNA sequencing in the field possible. This system detects electrical signals on a semiconductor chip, rather than optically reading dye-labeled nucleotides. This is possible as the addition of a dNTP to the DNA polymer causes the release of an H^+ ion

Because these devices are only detecting the current change, they can be very small! Even hand sized USB devices. However, they still require quite a bit of memory to run, as one small chip can collect reads of 30,000 bases. These devices are also not as accurate as the other methods, but multiple read throughs over the same sequence region help to reduce errors. As someone who grew up using radioactive Sanger sequencing methods that could only read about 300 bases per read, this type of technology absolutely blows my mind!

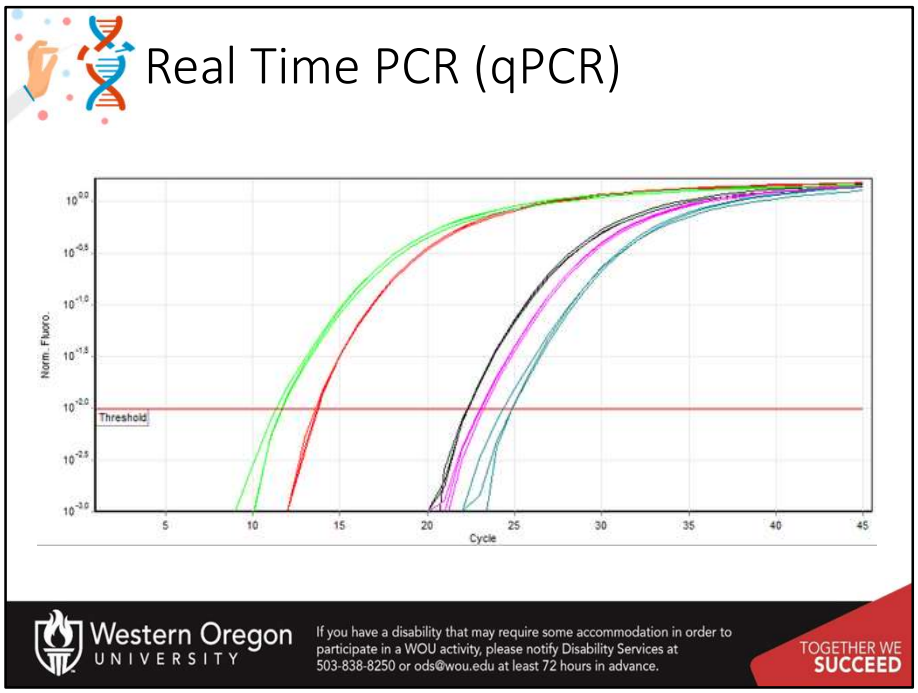


The one thing that NGS all has in common, is the use of PCR to generate the sequencing reads. So let's take a look at Polymerase Chain Reaction (PCR) in a little more detail. PCR is one of the major ways we have to synthesize DNA molecules, leading to more than a billion-fold amplification of specific gene sequences. The main components of a PCR reaction are a template, primers, the nucleotide triphosphate building blocks, and the DNA polymerase enzyme. The ***DNA template*** contains the specific region that you wish to amplify, such as the DNA extracted from a piece of hair for example. ***Primers***, or oligonucleotides, are short strands of single-stranded DNA complementary to the 3' end of each target region. Both a forward and a reverse primer are required, one for each complementary strand of DNA. ***DNA polymerase*** is the enzyme that carries out DNA replication. Thermostable analogues of DNA polymerase I, such as Taq polymerase, which was originally found in a bacterium that grows in hot springs, is a common choice due to its resistance to the heating and cooling cycles necessary for PCR. This process involves cycling through 3 sequential rounds of temperature dependent reactions: DNA melting (denaturation), annealing and enzyme-driven DNA replication (elongation). ***Denaturation*** begins by heating the reaction to about 95°C, disrupting the hydrogen bonds that hold the two strands of template DNA together. Next, the reaction is reduced to around 50 to 65°C, depending on the physicochemical variables of the primers, enabling ***annealing*** of complementary base pairs. The primers, which are added to the solution in excess, bind to the beginning of the 3' end of each template strand and prevent re-hybridization of the

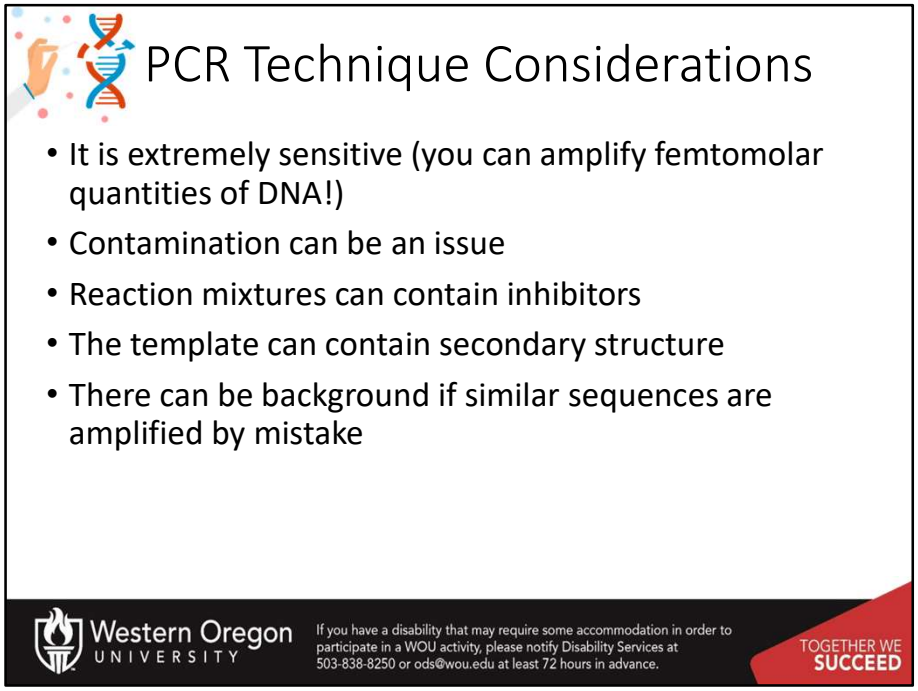
template strand with itself. Lastly, enzyme-driven DNA replication, or ***elongation***, begins by setting the reaction temperature to the amount which optimizes the activity of DNA polymerase, which is around 75 to 80^{o}C. At this point, DNA polymerase, which needs double-stranded DNA to begin replication, synthesizes a new DNA strand by assembling free-nucleotides in solution in the 3' to 5' direction to produce 2 full sets of complementary strands. The newly synthesized DNA is now identical to the template strand and will be used as such in the progressive PCR cycles. Each round of PCR will exponentially increase the amount of DNA present and result in the replication of the desired region, even if the template strands are much longer or represent a mixture of the entire genome.



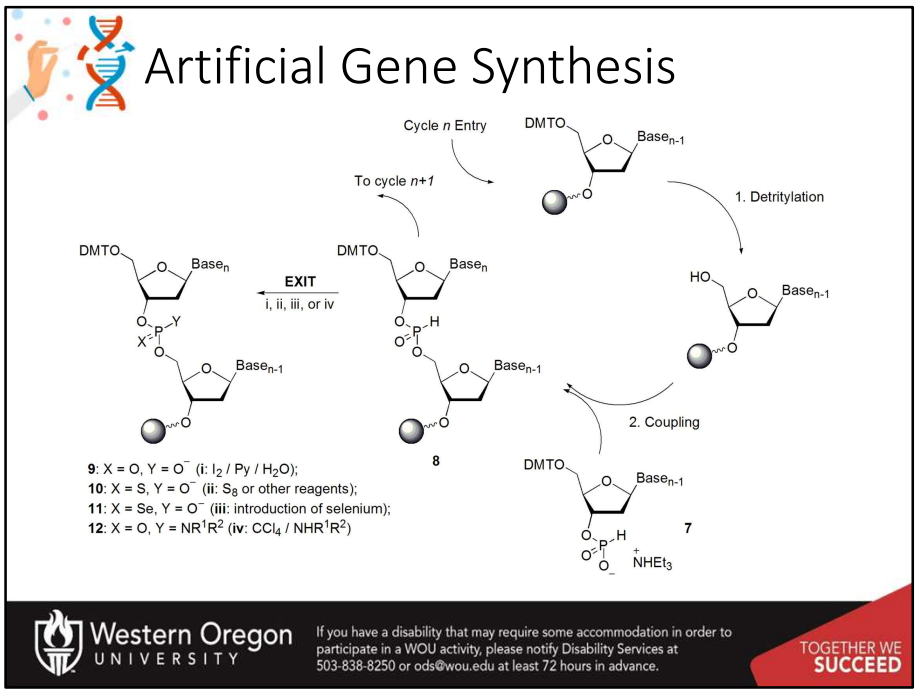
Since its advent, PCR technology has been creatively expanded upon, and ***reverse-transcription PCR (RT-PCR)*** is one of the most important advances. Real-time PCR is frequently confused with reverse-transcription PCR, but they are separate techniques. In RT-PCR, the DNA amplified is derived from mRNA by using reverse-transcriptase enzymes, to produce a cDNA copy of the gene. Using primers sequences for genes of interest, traditional PCR methods can be used with the cDNA to study the expression of genes qualitatively. It also allows the amplification of eukaryotic gene sequences that do not contain introns. This is really useful if you want to express a protein of interest.



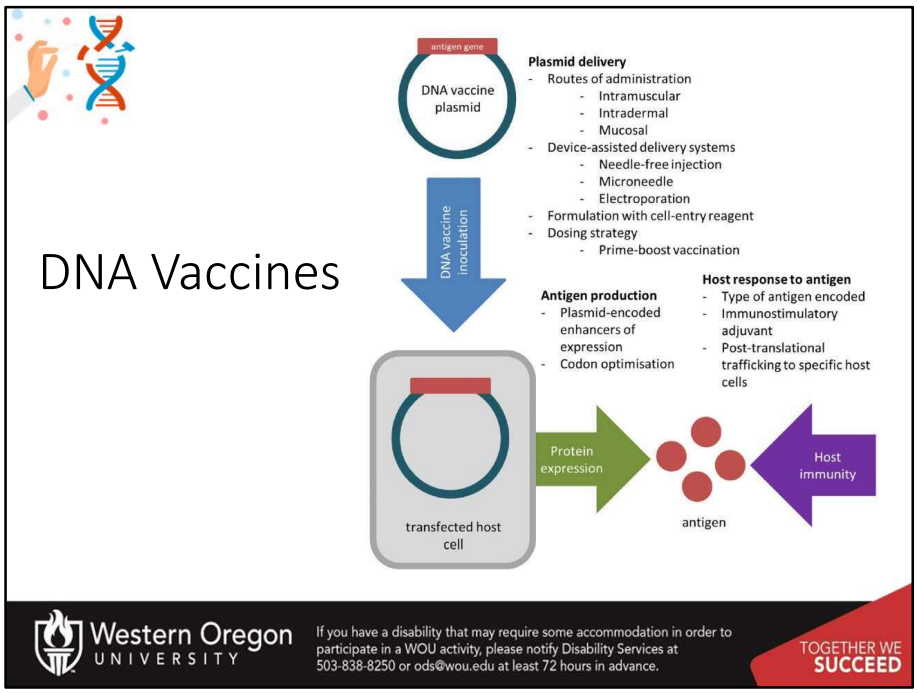
Real time PCR is also known as quantitative PCR. It monitors the amplification of a targeted DNA molecule during the PCR (i.e., in real time), not at its end, as in conventional PCR. Real-time PCR can be used quantitatively. It is often used with RT-PCR to measure the amount of mRNA present in a sample and gain an understanding of how much gene transcription is going on within the system. This is great for comparing samples that have been treated differently, such as a control sample, and a sample exposed to UV-light.



Sensitivity can be both a good thing and a bad thing in PCR. You can detect extremely small quantities of DNA that might be present, which is great if you are trying to catch a killer, or identify a genetic disease, but great care must be taken in sample preparation, as contamination can be a real issue. PCR is also a technique that often has to be trouble shooted to get it to work well. If the template DNA is not pure enough it may be contaminated with inhibitors and block the polymerase activity. Furthermore, the template DNA may have too much secondary structure for the primer to bind well or if there are similar sequences within the genome, other sequences may also be amplified creating background that might make it difficult for you to purify your product. However, even with these challenges, it remains one of the most commonly used techniques in molecular biology and biochemistry labs.



Artificial gene synthesis, sometimes known as **DNA printing** is a method in synthetic biology that is used to create artificial genes in the laboratory. Based on solid-phase DNA synthesis, it differs from molecular cloning and polymerase chain reaction (PCR) in that it does not have to begin with preexisting DNA sequences. Therefore, it is possible to make a completely synthetic double-stranded DNA molecule with no apparent limits on either nucleotide sequence or size. This type of synthesis is most commonly used to make the small single-stranded DNA primers that are then used in PCR reactions. It is similar to what we saw for protein synthesis, in that, a nucleoside is crosslinked to a bead support that will be held tethered within a column. Note that the 5'-OH position of the sugar has a protecting group on it to keep it from reacting within the test tube during the crosslinking process. The column is rinsed and the protective group is removed. The next base is then added, and you can see it is weird, compared to a normal nucleotide triphosphate used in DNA synthesis in vivo. In solid phase synthesis a monophosphate is used, only , the phosphate is linked at the 3' position of the incoming nucleotide and will react with the 5'-OH of the sugar that is tethered to the bead support. (this is essentially backwards from DNA synthesis in vivo, that always synthesizes DNA from the 3'-OH position). This process is repeated to add additional bases to the molecule. The final product will be deprotected and cleaved from the bead matrix.



DNA vaccines are able to avoid many issues associated with egg-based vaccine production by generating viral proteins within host cells. To create a DNA vaccine, an antigen-encoding gene is cloned into a non-replicative expression plasmid, which is delivered to the host by traditional vaccination routes. Host cells which take up the plasmid express the vaccine antigen which can be presented to immune cells via the major histocompatibility complex (MHC) pathways. CD4+ T helper cell activation following MHC class II presentation of secreted DNA vaccine protein is critical for the production of antigen-specific antibodies. These can be safer than traditional vaccines that may have live, attenuated viruses and pose more of an infection risk. However, their efficacy, may be less than traditional vaccines. In the next few sections, we will look first at bioinformatics and then at procedures that allow the cloning and expression of specific genes.