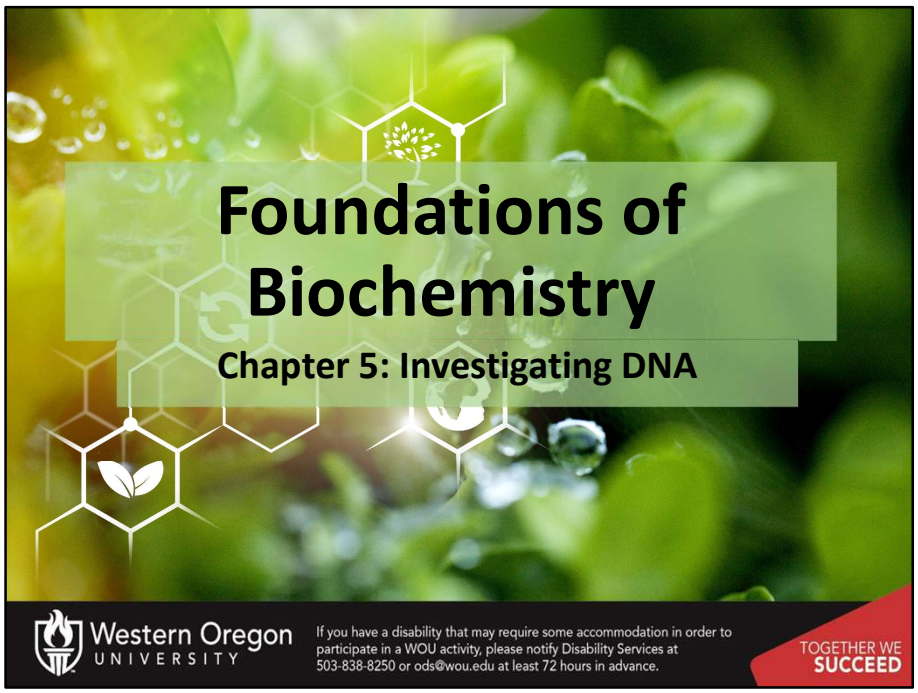
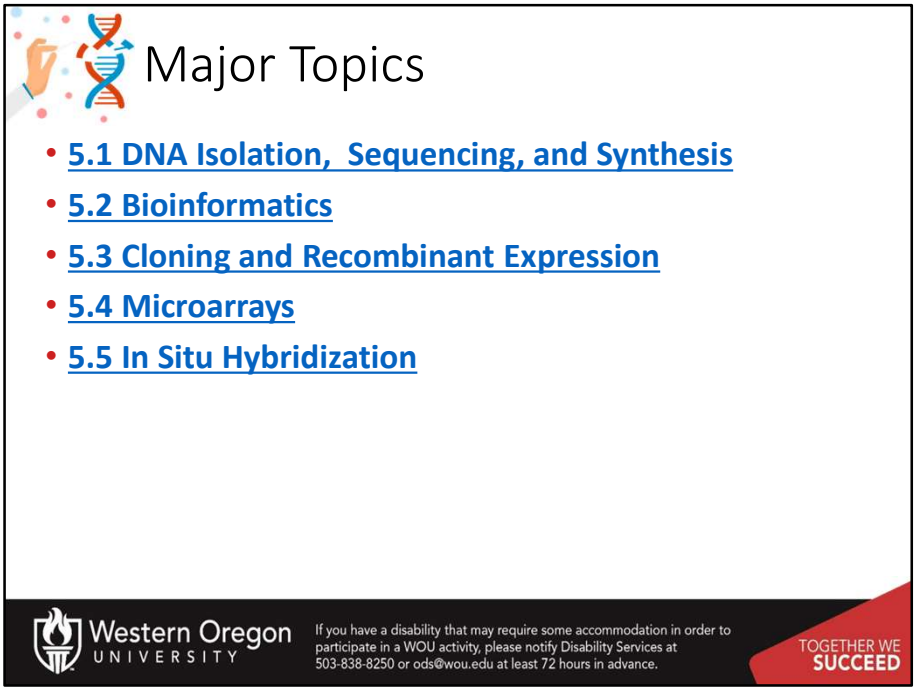
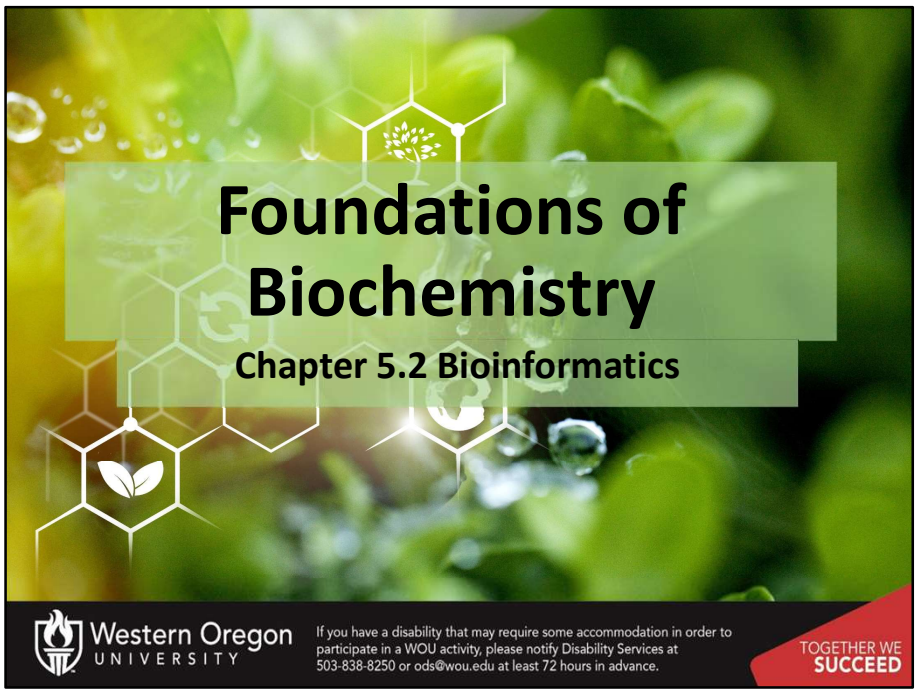



Welcome back! We will be continuing our focus on techniques used in the laboratory to study DNA and RNA.

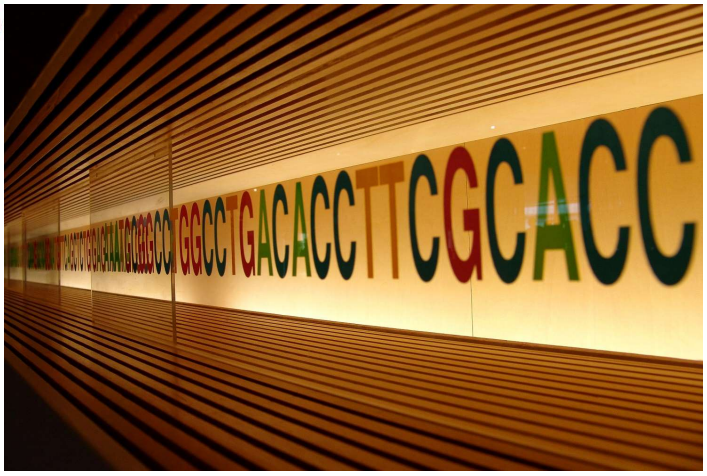



The last two lectures on Chapter 5 will focus on sections 5.2 through 5.5. This lecture will focus on 5.2 bioinformatics and the first part of section 5.3 on cloning and recombinant expression.




In this lecture we are continuing our exploration of DNA laboratory techniques with a focus on the basics of bioinformatics and DNA cloning techniques.

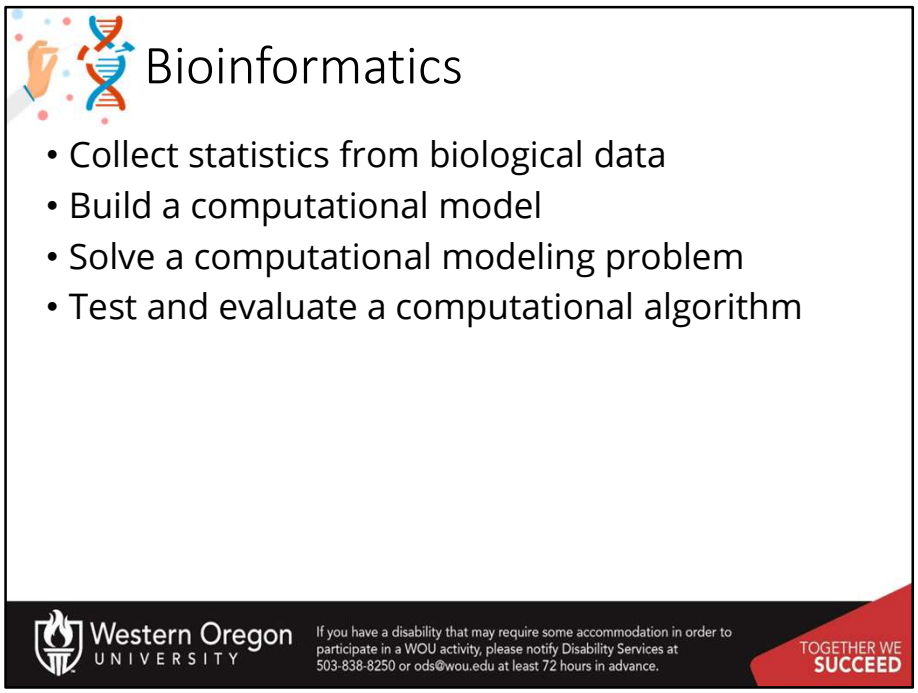
## Genome Sequencing Projects



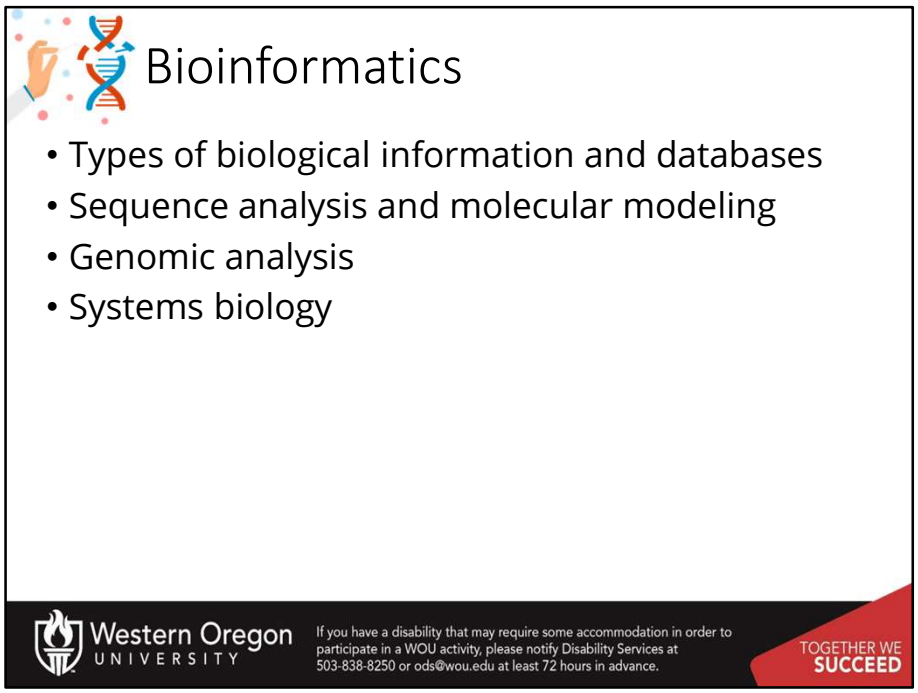

**DNA**—Photo by Flickr user MIKI Yoshihito



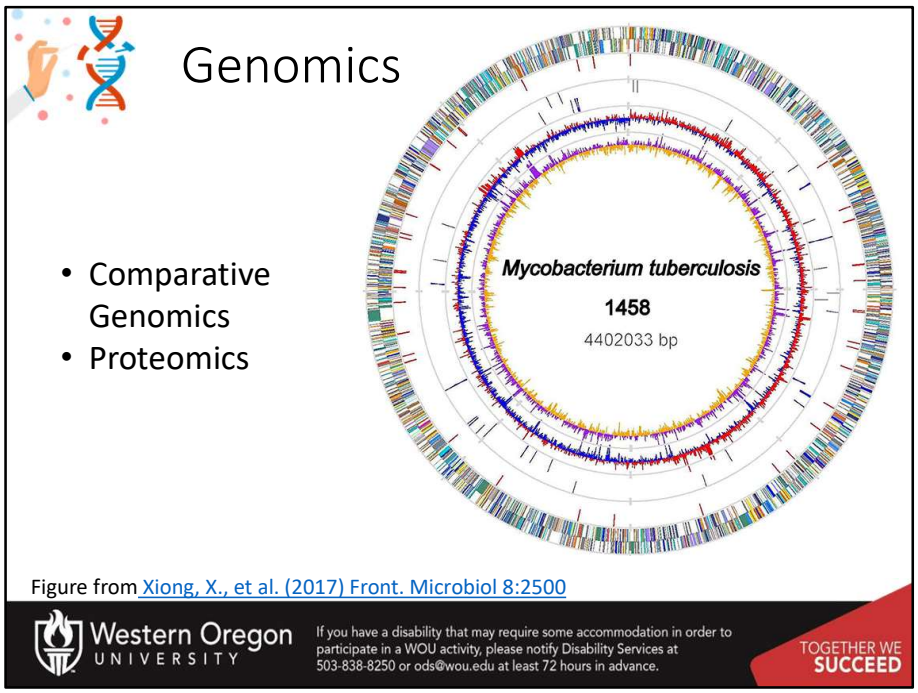
With the advent of cheaper Next Generation Sequencing technologies that we explored in a previous section, there has been a huge amount of sequencing data from all kinds of different organisms that has flooded our databases. Thus, we have needed the development of computer tools and resources to help us analyze the data and make predictions based on comparisons with sequences that have known functions.




Bioinformatics usually involves the following steps:




And addresses the following aspects:




A lot of resources and tools have been developed for analyzing genome sequences. Genomics tools have been developed to analyze genome sequences for potential or predicted open reading frames (ORFs) and often rely on homology with known sequences for identification. This has given rise to comparative genomics, where complete genomes from similar organisms are compared. This can be especially useful in identifying virulence factors in bacteria and viruses, especially when one related strain shows high toxicity while another does not. And proteomics is specialized to study the proteins expression patterns within different locations in a single organism or between different but related organisms.

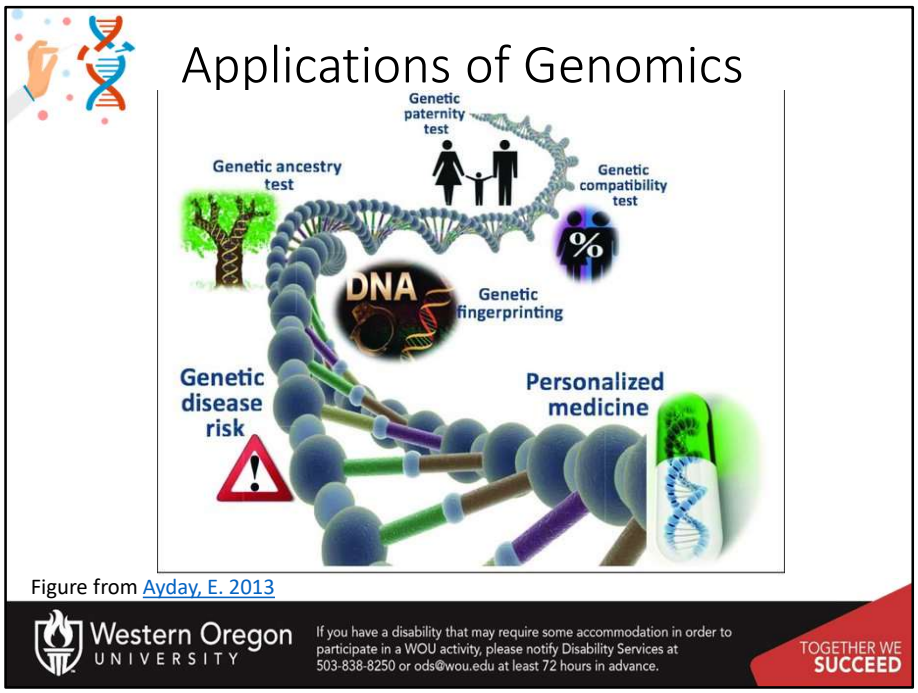
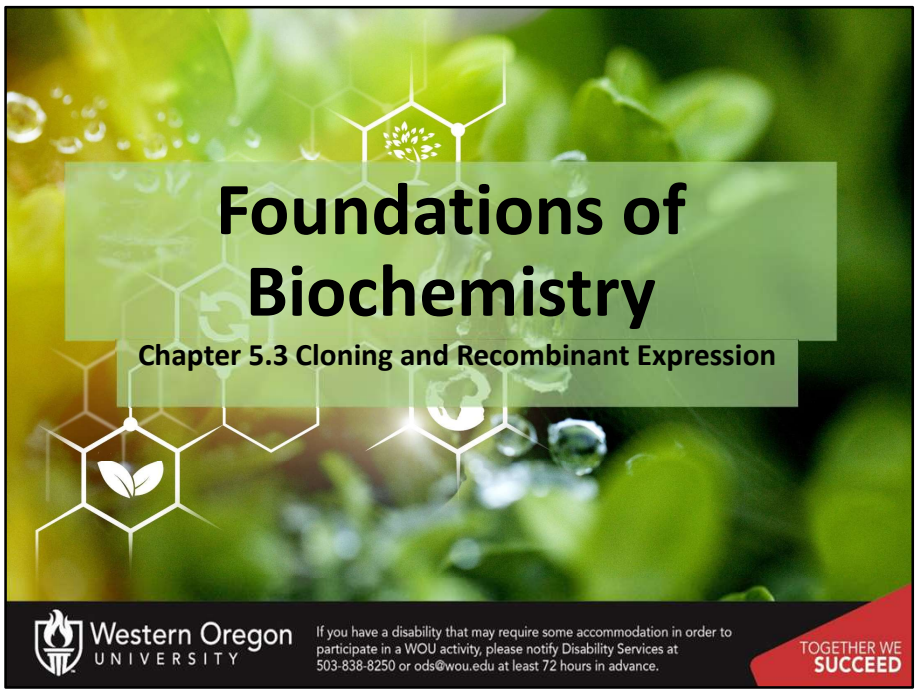
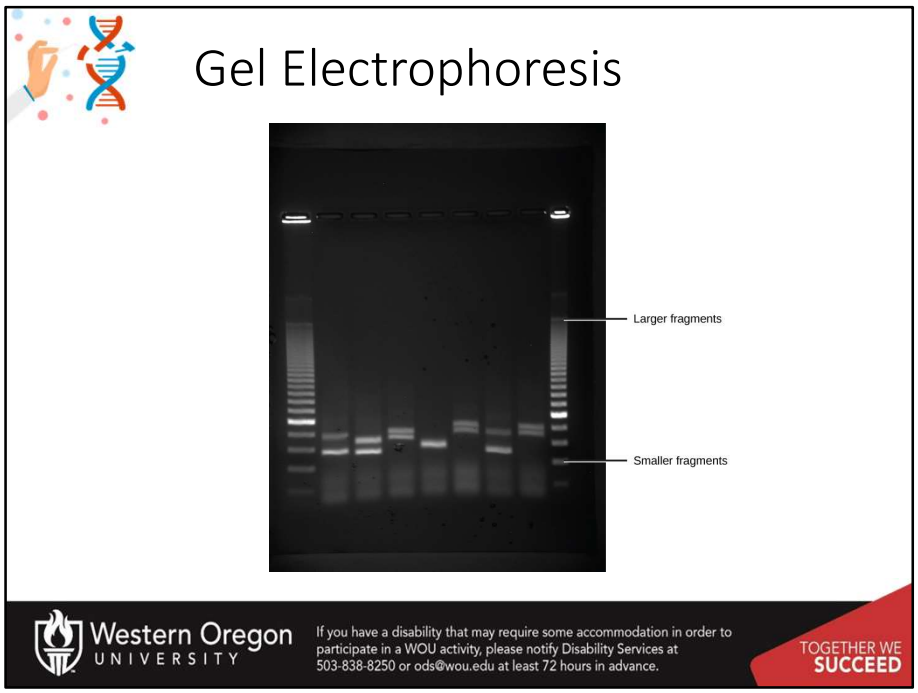



Figure from Ayday, E. 2013

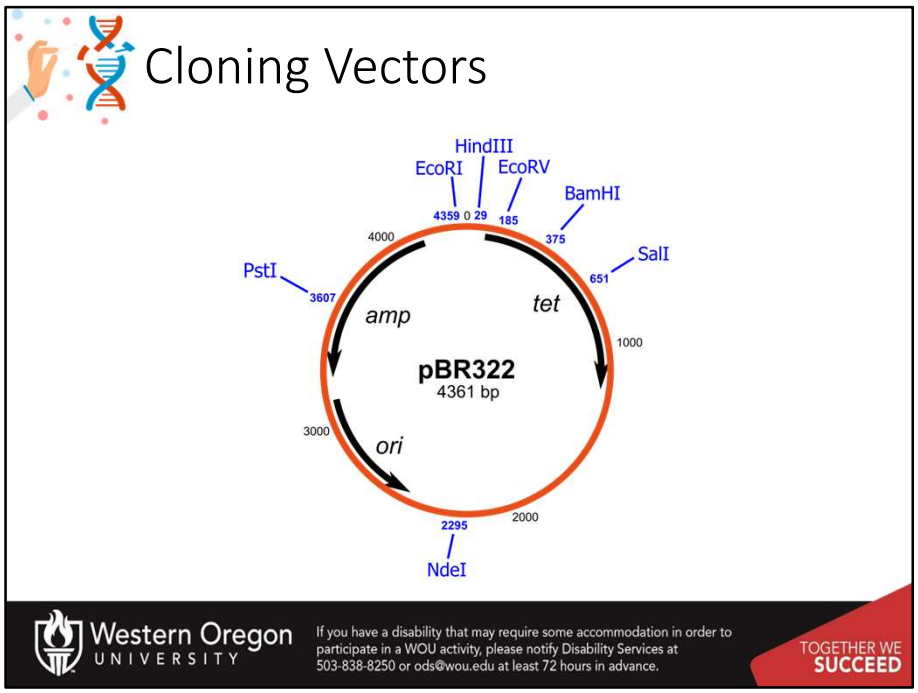
There are many applications for genome analysis. These include areas such as drug discovery, personalized medicine (especially in areas such as cancer research and diagnositcs), forensic science and ancestry identification, biodefense, and the development of new fields such as nutrigenomics. Genome analysis can provide the starting place for many interesting lines of research. In the next section, we will investigate gene cloning and protein expression techniques.



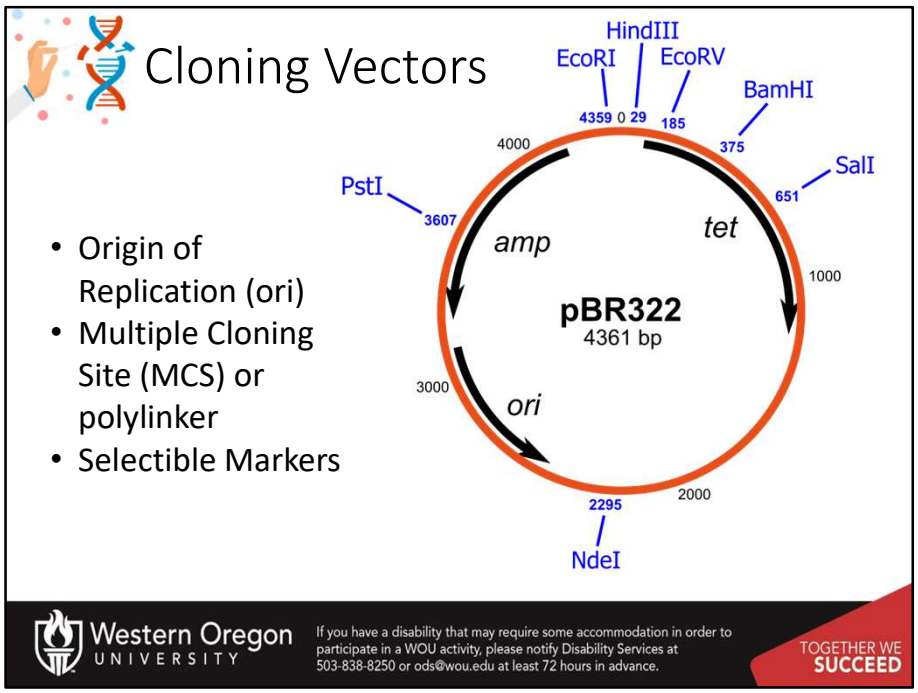

Let’s get started learning about techniques used to clone genes and express proteins.



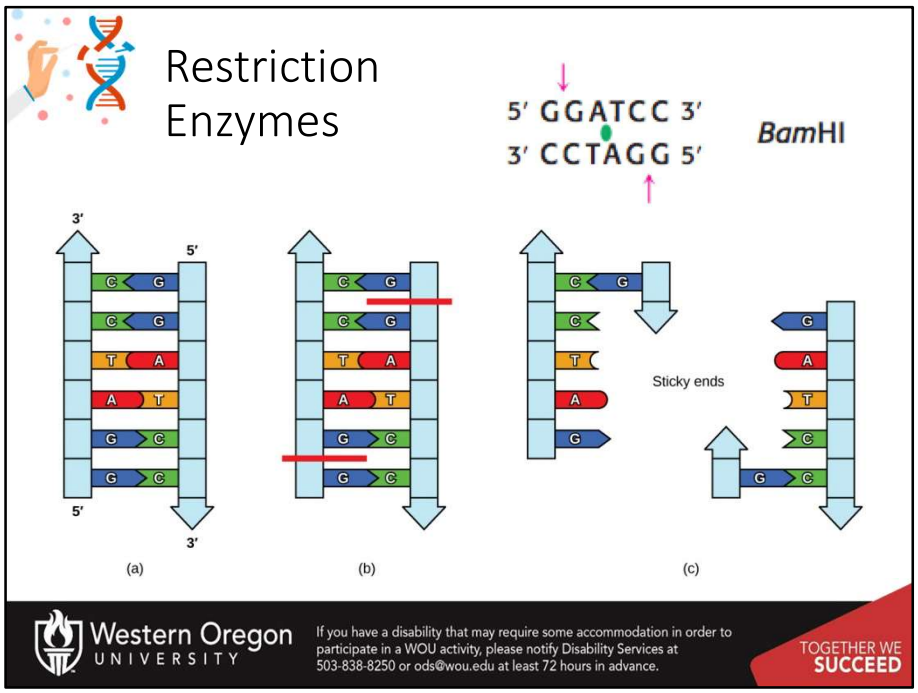

As we saw with protein analysis, gel electrophoresis is an important way of visualizing, isolating and purifying DNA. Typically agarose gels are used to analyze DNA fragments. As with protein gels, DNA gels are run through an electric current (towards the positive lead as the DNA is quite negatively charged due to the phosphate groups), based on its size. Smaller fragments can pass more quickly through the gel and will move farther away from the wells than larger fragments. This is especially useful for visualizing PCR products and DNA plasmids for gene cloning.



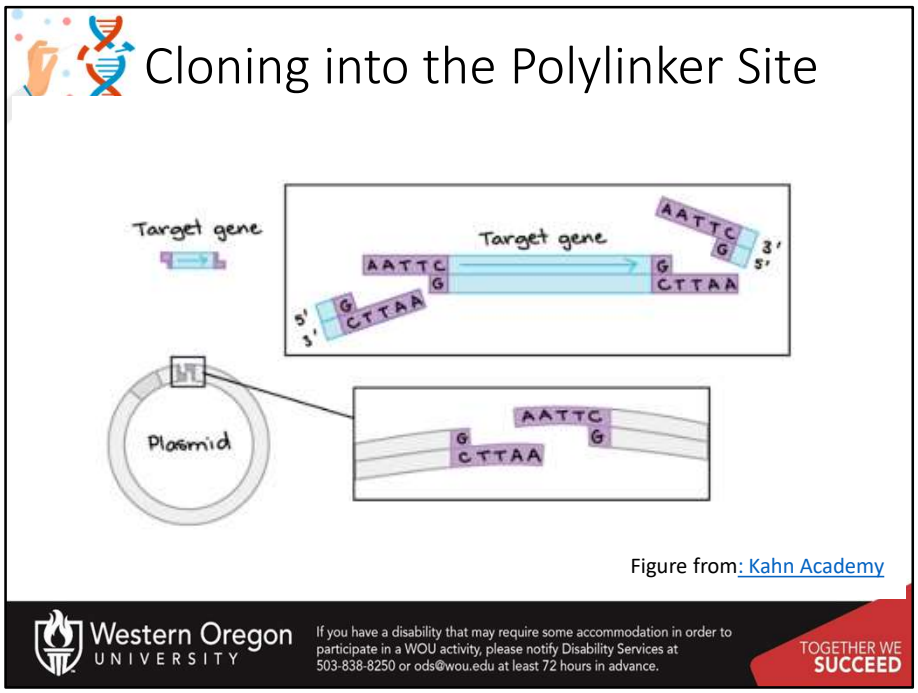

Once a gene of interest has been identified RT-PCR is used to amplify a copy of cDNA (recall that this type of DNA will not contain any intron sequences making gene expression easier). Regular PCR can be used for prokaryotic genes, as they do not normally contain intron sequences. Once the gene has been amplified, it needs to be cloned into an expression vector. This will allow the gene to be replicated within a host species, and also allow a host species to express the gene of interest and produce the desired protein.



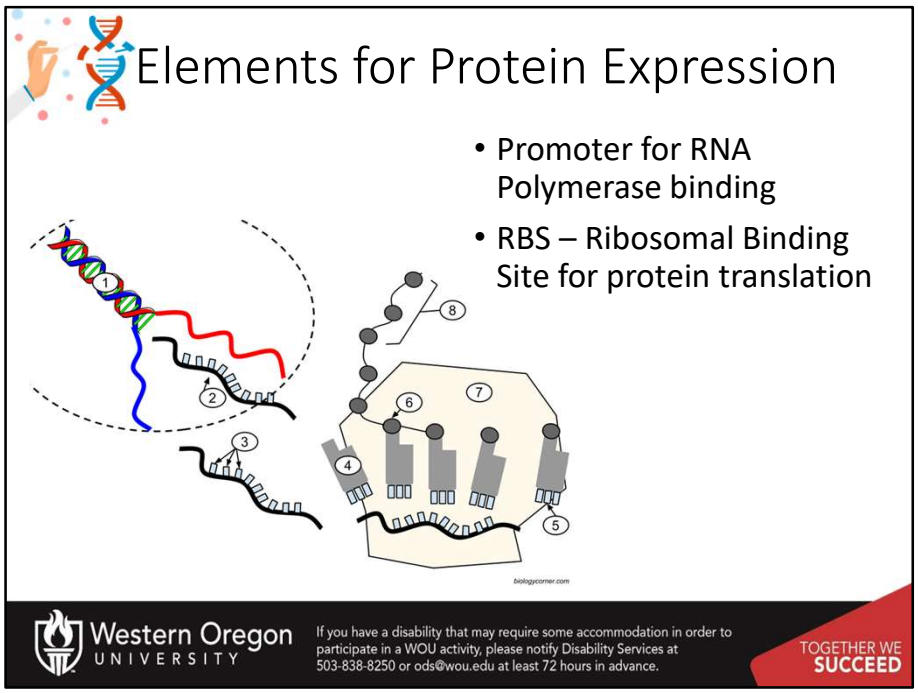

Cloning vectors typically contain specific features that allow them to replicate and be selected for within the host. This includes the origin of replication (ori) which allows the vector to be recognized by the host DNA replication machinery and replicated within the cell. Without an ori, the vector would be lost from the host following replication. In addition to ori's, cloning vectors usually contain a selectable marker that allow host cells that contain the vector to be isolated away from cells that do not contain the vector. In bacterial systems, antibiotic resistance genes are usually used for this purpose. Once your gene of interest is cloned in and the vector is transformed into the host, the host can be grown on the antibiotic of choice. Only host cells containing the vector and expressing the antibiotic resistance gene will then survive. Those colonies can then be used to either express the protein or be a reservoir for collecting and reisolating copies of the cloning vector. Cloning vectors also contain a polylinker region or multiple cloning site that contains sequences of DNA recognized by restriction enzymes. These restriction sites are noted in blue for this vector. This type of restriction enzyme system is highly useful for the cloning procedure and we will take a look at this in greater detail.



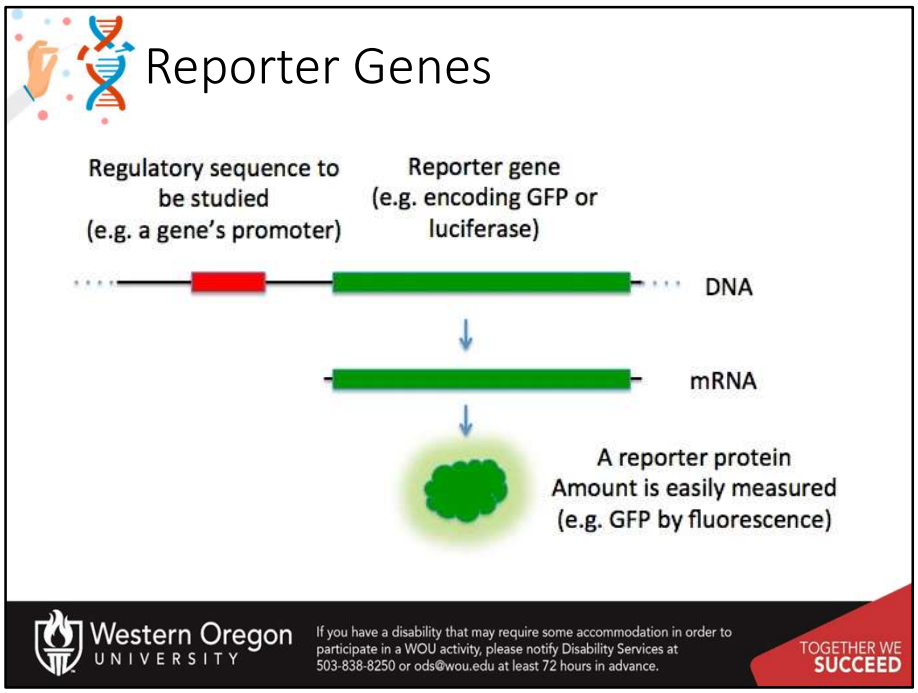

The restriction enzymes most widely used in cloning recognize specific sequences within the DNA called palindromes. The one pictured above is called BamHI, and it's palindrome is 5'-GGATCC-3'. You can see that when you read either the forward or reverse strand in the 5' to 3' direction, that the sequence is the same…always GGATCC. This is what is meant by a palindrome. Some restriction enzymes, such as BamHI do not cut the restriction site right in the middle, instead the cut sites are offset on either strand, creating a single stranded DNA overhand on both of the ends of the fragments. In the case of BamHI, it cuts the palindrome after the first G on both strands. These overhanging sites are known as sticky ends, and provide an important resource for aiding in the cloning process. Restriction enzymes and their sequences are named from the organisms from which they were derived. BamHI is from ***B****acillus* ***am****yloliquefaciens*, strain H, and was the 1st restriction enzyme isolated from this strain. Hence the name BamHI.



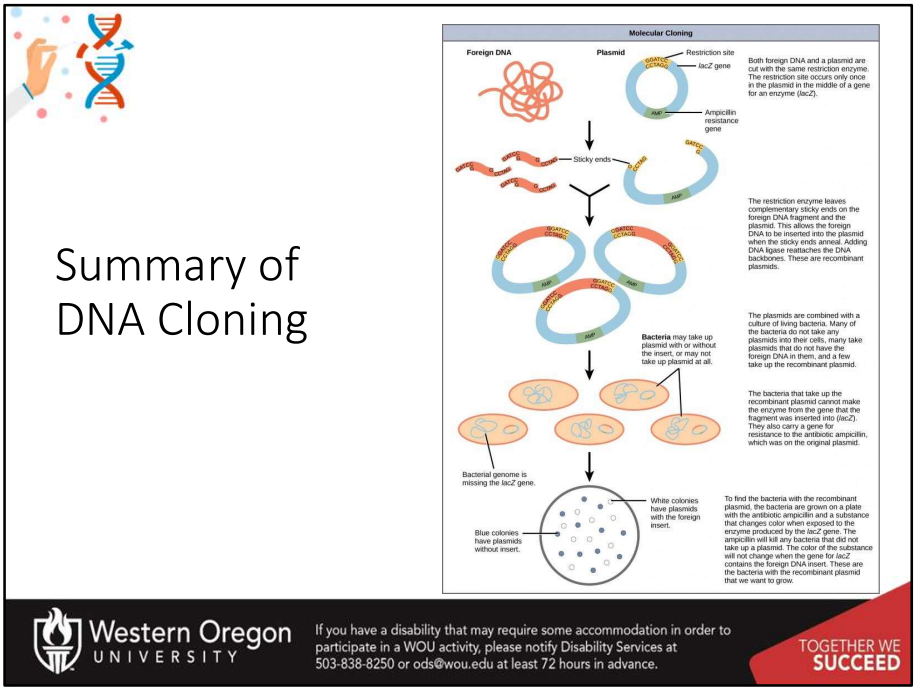

Restriction enzyme systems can be so useful for cloning, because PCR primers can be engineered to contain the correct restriction site at each end of the sequence, so that both the PCR product and the vector can be cut with the restriction enzyme. The enzyme is then removed from each of the DNA fragments (by ethanol precipitation!) and the two fragments can be mixed together. In some cases the sticky overhangs of the insert will line up with the sticky overhangs of the vector, causing the gene of interest to 'stick' to the vector. An enzyme called DNA ligase can then be added to the mixture to seal in the backbone and form the covalent bonds of the sugar phosphate backbone, effectively permanently attaching the insert to the vector at the specific restriction enzyme location. Because the cloning occurs at a known location within the vector, very detailed maps can be made of the cloning vector that allow its easy recognition on during agarose gel electrophoresis.



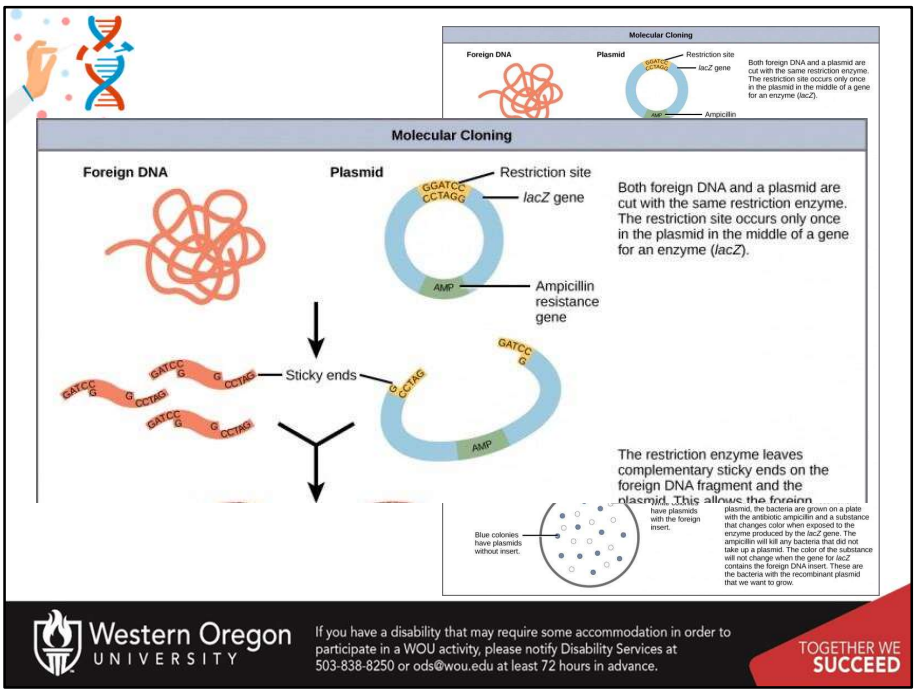

If you want to express your gene of interest and then isolate or study the protein encoded by it, then the vector needs to contain some expressional elements as well. It will need a promoter in the DNA sequence that will be recognized by the RNA polymerase so that transcription can occur, generating a mRNA molecule. The resulting mRNA must also have a ribosomal binding site, so that it is recognized by the ribosome and can be translated into the protein sequence. If these are not present within the cloning vector, these sequences can often be engineered into the PCR primer sequence. Although many commercial vectors have been made and optimized to have these features.



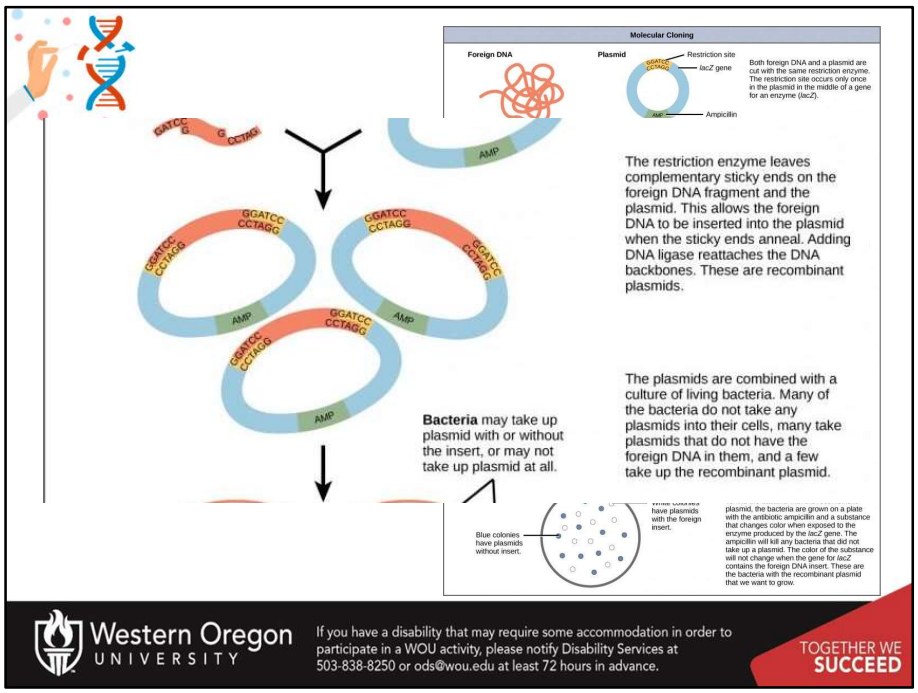

Sometimes you may want to study a promoter region and see what types of enhancer proteins or transcription factors may bind to the region. In this case, you may want to use a reporter gene system to assess the biological activity. In this type of system the promoter is region of interest and the area of the vector that will vary. The downstream gene will typically encode for a protein that is easy to assay or visualize such as the green fluorescent protein or an enzyme, such as LacZ that can cleave a dye molecule providing a colorimetric method of detection.



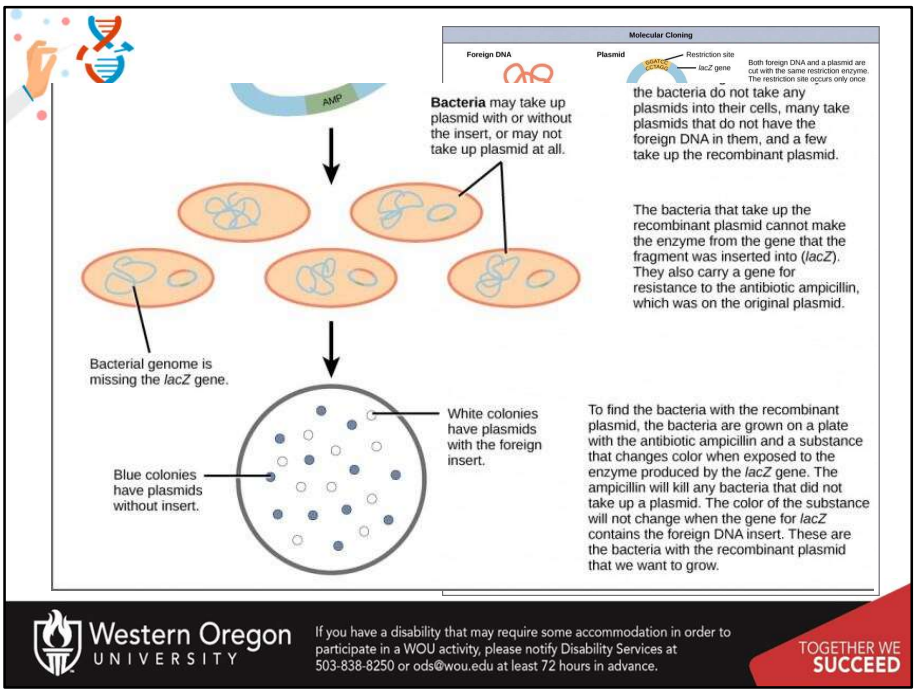

Here is a summary chart of a typical DNA cloning experiments



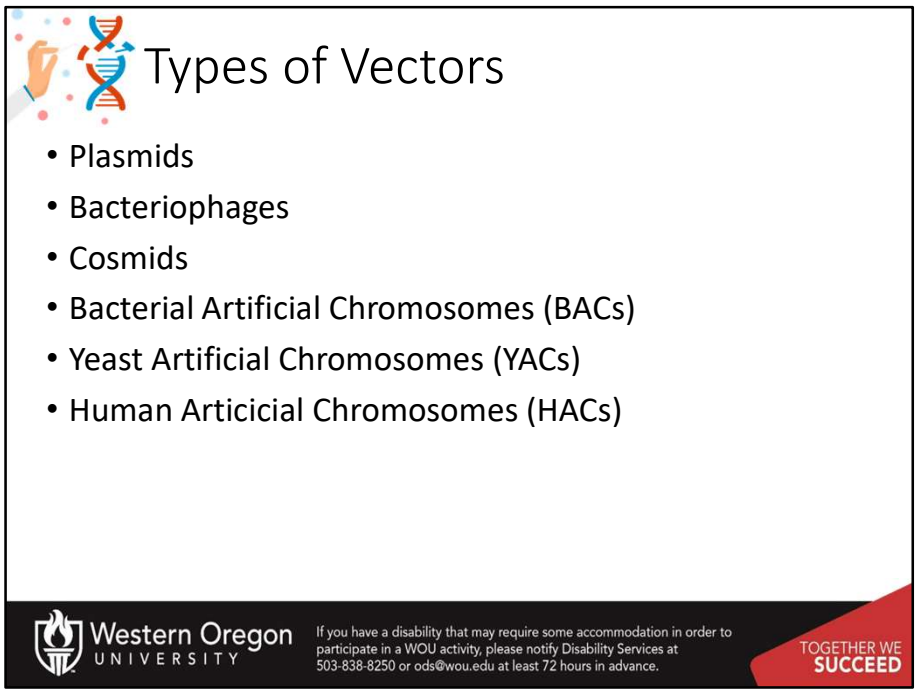

In the first part, the genomic DNA from your organism of interest is isolated. This is noted as the foreign DNA above. This DNA serves as the template within a PCR reaction to amplify a specific gene region. Both the PCR product and the vector that it will be clones into (shown in blue) are then cleaved, in separate reactions, using appropriate restriction enzymes.



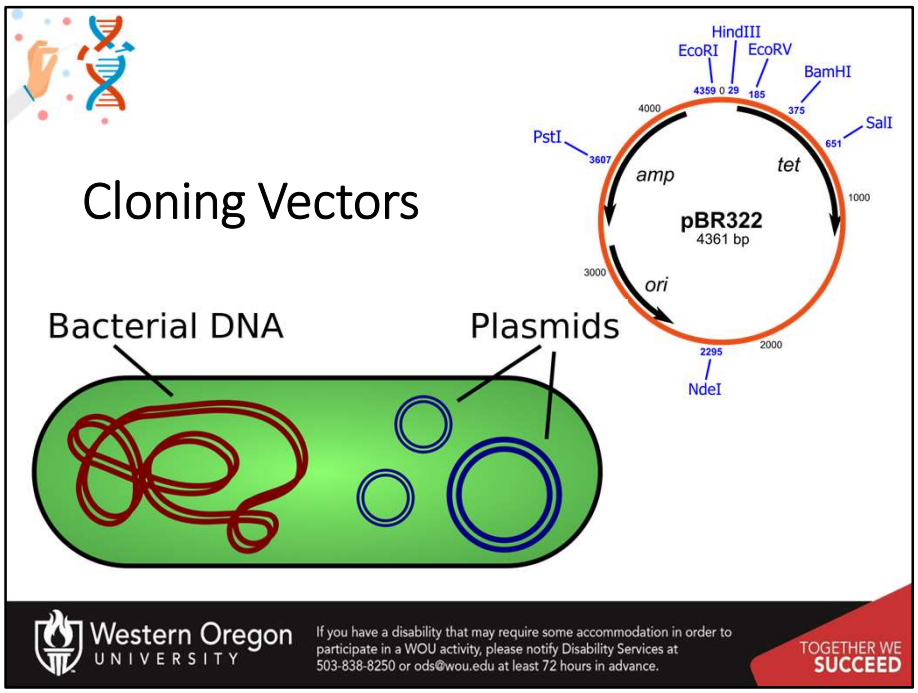

The restriction enzymes and buffer are removed from the system using ethanol precipitation of the DNA. The PCR and vector samples are then reconstituted in the ligation buffer and the DNA ligase enzyme is added to a mixture of the PCR product and vector. This causes the fusing of the vector and sealing of the backbone, such that the vector now contains the PCR insert.



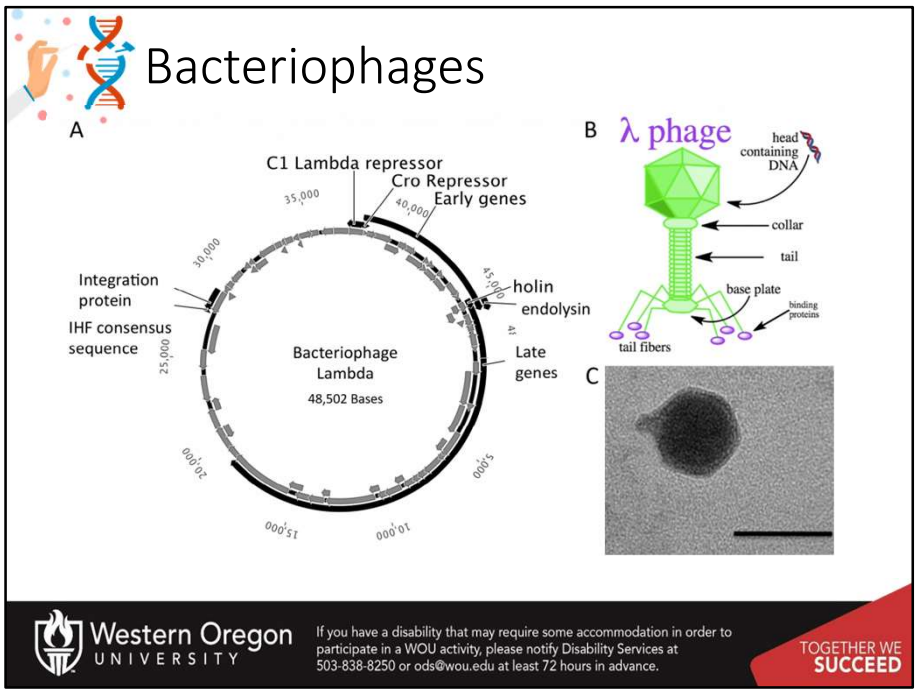

The ligated vector is then transformed into a bacterial host species. Blue/white selection is a specialized system set up in some vectors. In these systems the lacZ gene is housed within the vector across the multiple cloning site and the bacterial host strain is deficient and does not have the lacZ gene. The lacZ enzyme (beta-galactosidase) can cleave a dye molecule and create a blue color within the bacterial host. During the cloning process, the gene of interest disrupts the lacZ gene, so that it is no longer functional. Thus, when bacteria are transformed with the ligation mixture, three outcomes are possible. First bacteria may not take up the vector. These bacteria are removed from the system by growing the bacteria on selective, antibiotic-containing media. Thus, only bacteria that receive the plasmid vector can survive. The resulting colonies will also be grown on the dye precursor, such that if the beta-galactosidase enzyme is present, the colonies will turn blue. Thus, if bacteria that have received vectors with the gene of interest housed within them (destroying the lacZ gene), they will remain white, while colonies that turn blue, will house the vector that has not taken up the PCR insert and retains an intact lacZ gene. The white colonies can then be selected for further study.



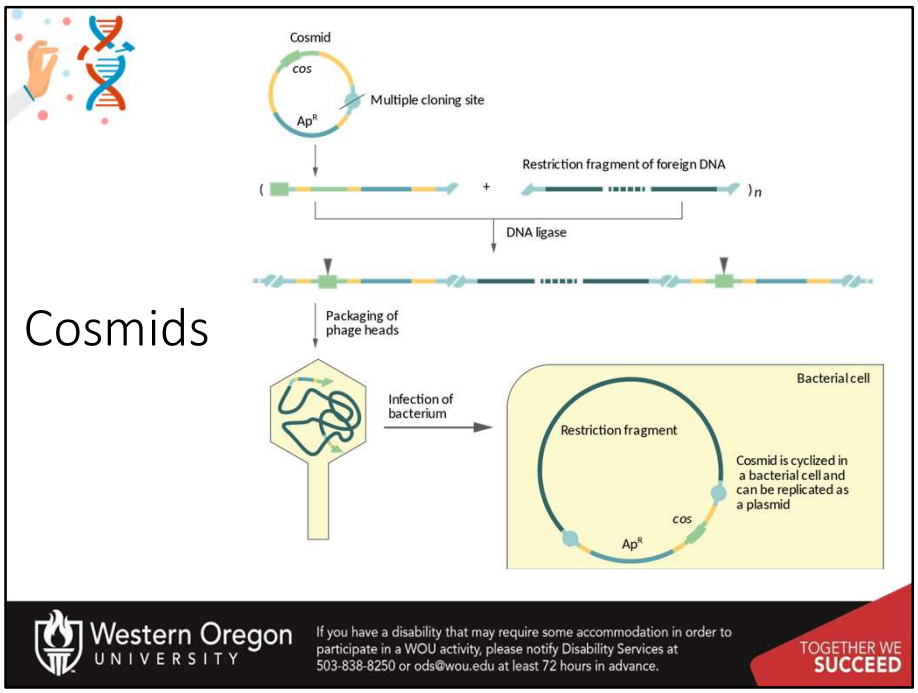

There are many different types of vectors that can be used to carry DNA information to a host. The one shown on the previous slide was a plasmid



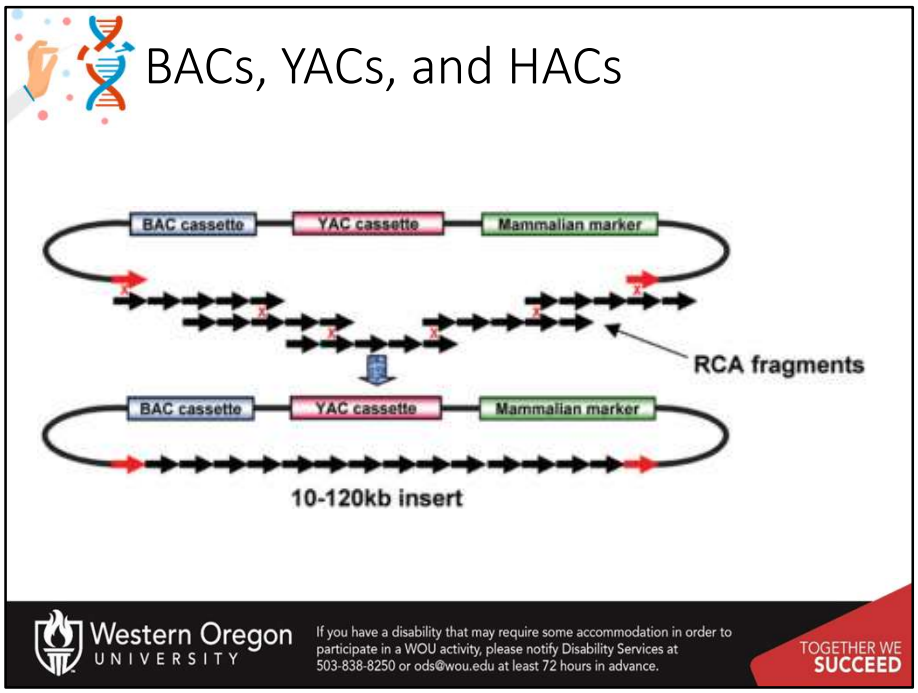

A **plasmid** is a small, extrachromosomal DNA molecule within a cell that is physically separated from chromosomal DNA and can replicate independently. They are most commonly found as small circular, double-stranded DNA molecules in bacteria; however, plasmids are sometimes present in archaea and eukaryotic organisms. In nature, plasmids often carry genes that benefit the survival of the organism and confer selective advantage such as antibiotic resistance. While chromosomes are large and contain all the essential genetic information for living under normal conditions, plasmids are usually very small and contain only additional genes that may be useful in certain situations or conditions. We have artificially done extensive development of plasmids to serve as vectors in biochemistry and molecular biology. They are probably one of the most widely used tools within the lab. They are usually suitable for carrying 1 kb to 5 kb of DNA inserts.



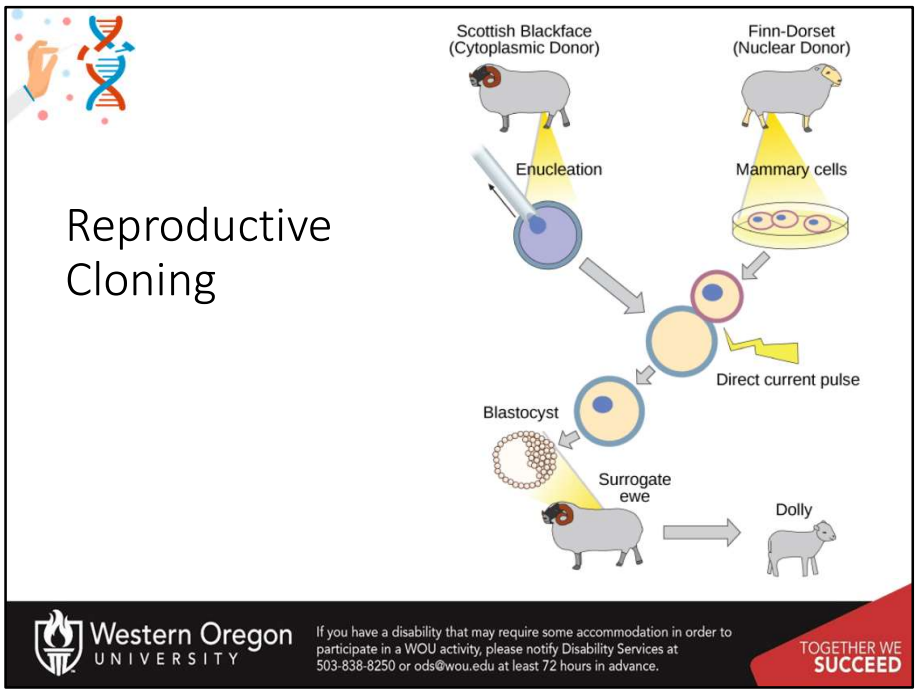

Bacteriophages are viruses that infect bacteria and achaea. They are among the most common and diverse entities in the biosphere. Bacteriophages are ubiquitous viruses, found wherever bacteria exist. It is estimated there are more than $10^{31}$ bacteriophages on the planet, more than every other organism on Earth. They can carry quite a bit more DNA than a standard plasmid. For the lambda phage shown here, an upper limit of about 53 kb of DNA can be reached. They also have a lower minimal requirement for DNA length as well, and are not as suitable for small insert sizes. Interestingly, viruses are also proving to be useful vectors for delivering genetic information into humans as well. Currently, adenoviruses are being used to deliver vaccines for more virulent viral infections, such as Ebola, and in some gene therapy strategies.



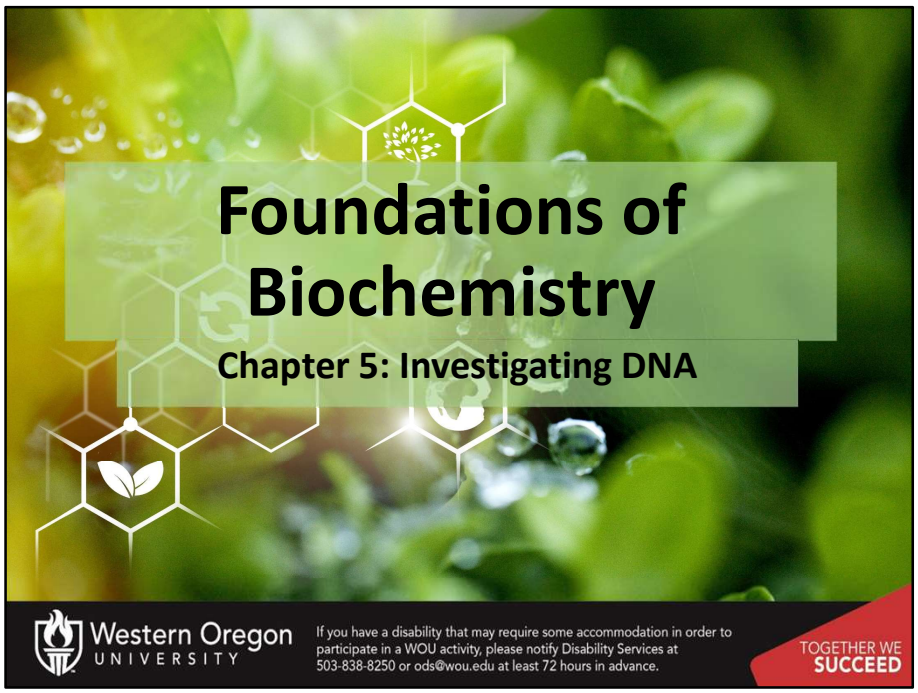

Cosmids are plasmids that incorporate a segment of bacteriophage λ DNA that has the cohesive end sites (*cos*) which contains elements required for packaging DNA into λ particles. It is normally used to clone large DNA fragments between 28 and 45 Kb. Once the phage particles have infected the host, the DNA recircularizes and forms a replicating plasmid within the bacteria.



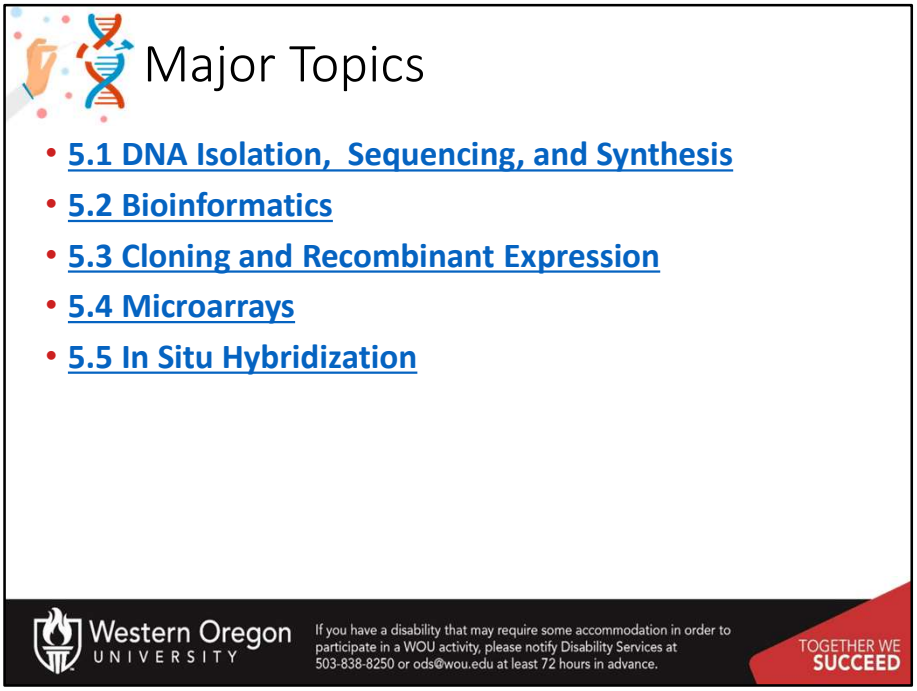

Artificial chromosome structures have also been developed in many host species, and can hold large DNA fragments, up to 120 kb or more in some cases. The one shown here contains elements from both BAC and YAC systems, allowing growth in both species.



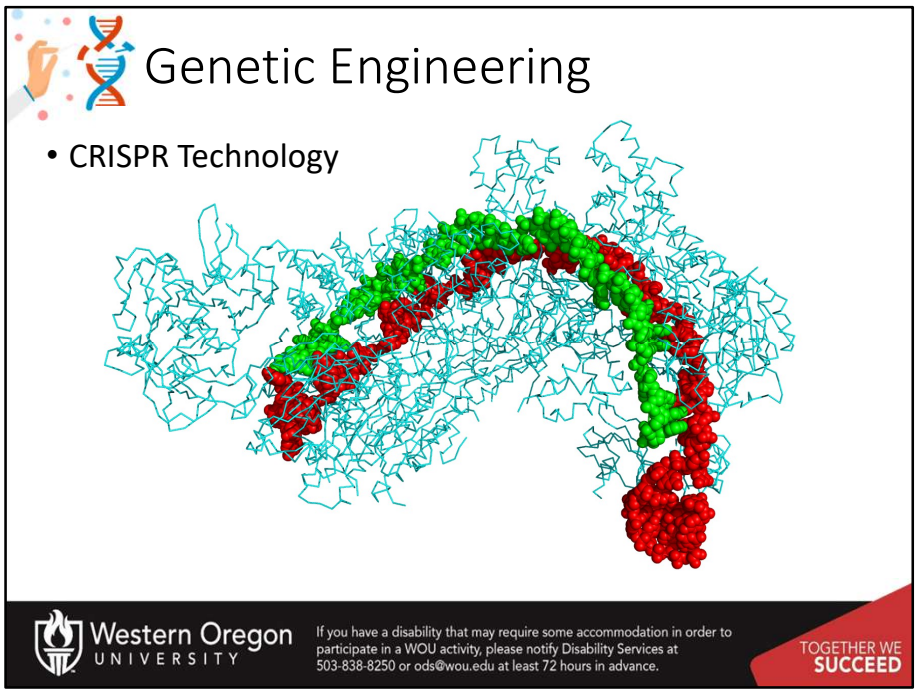

Reproductive cloning is a method used to make a clone or an **identical copy of an entire multicellular organism**. Most multicellular organisms undergo reproduction by sexual means, which involves the contribution of DNA from two individuals (parents), making it impossible to generate an identical copy or a clone of either parent. Recent advances in biotechnology have made it possible to reproductively clone mammals in the laboratory. **Dolly the sheep was the first agricultural animal to be cloned.** To create Dolly, the nucleus was removed from a donor egg cell. The enucleated egg was placed next to the other cell, then they were shocked to fuse. They were shocked again to start division. The cells were allowed to divide for several days until an early embryonic stage was reached, before being implanted in a surrogate mother.



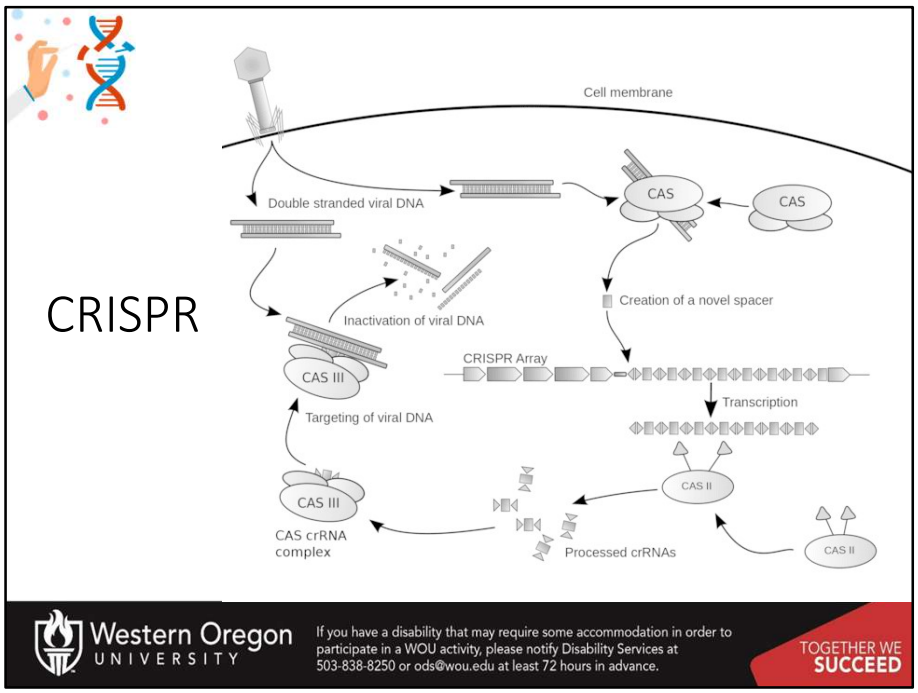

Hello, and welcome back. This is the final lecture of Chapter 5.



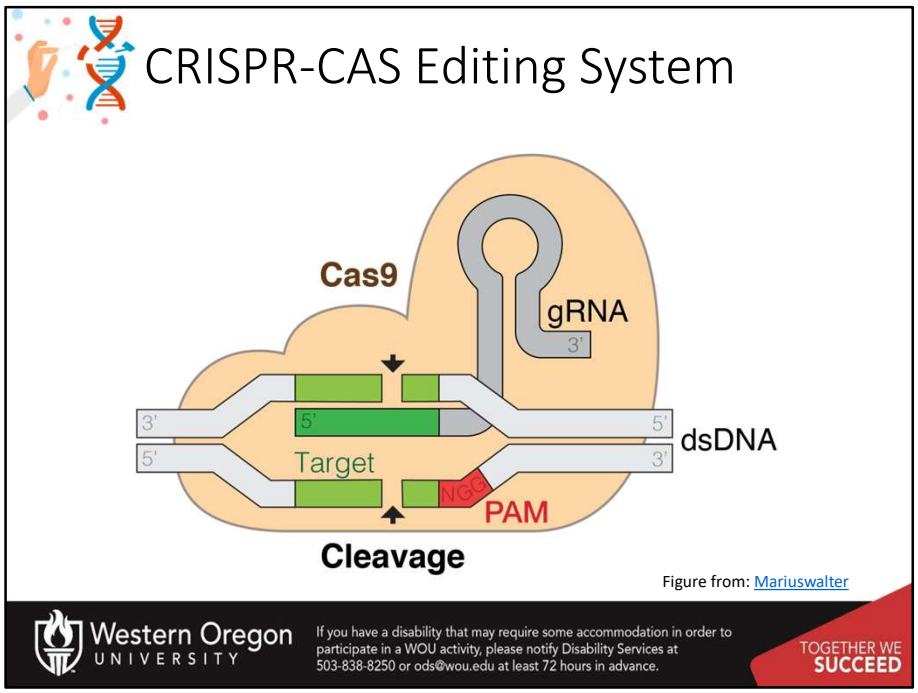

In this lecture, we will finish out section 5.3 with a discussion on CRISPR/Cas technlologies and end with sections 5.4 and 5.5 on microarray and in situ hybridization techniques.



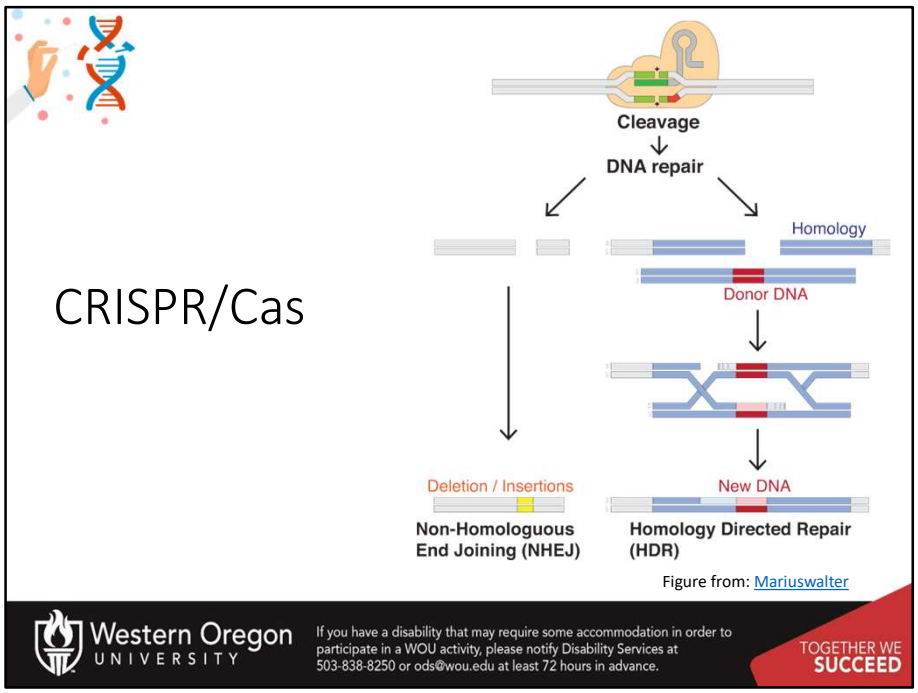

**CRISPR** stands for ***clustered regularly interspaced short palindromic repeats*** and represents a family of DNA sequences found within the genomes of prokaryotic organisms such as bacteria and archaea. These sequences are derived from DNA fragments of bacteriophages that have previously infected the prokaryote and are used to detect and destroy DNA from similar phages during subsequent infections. Hence these sequences play a key role in the antiviral defense system of prokaryotes.



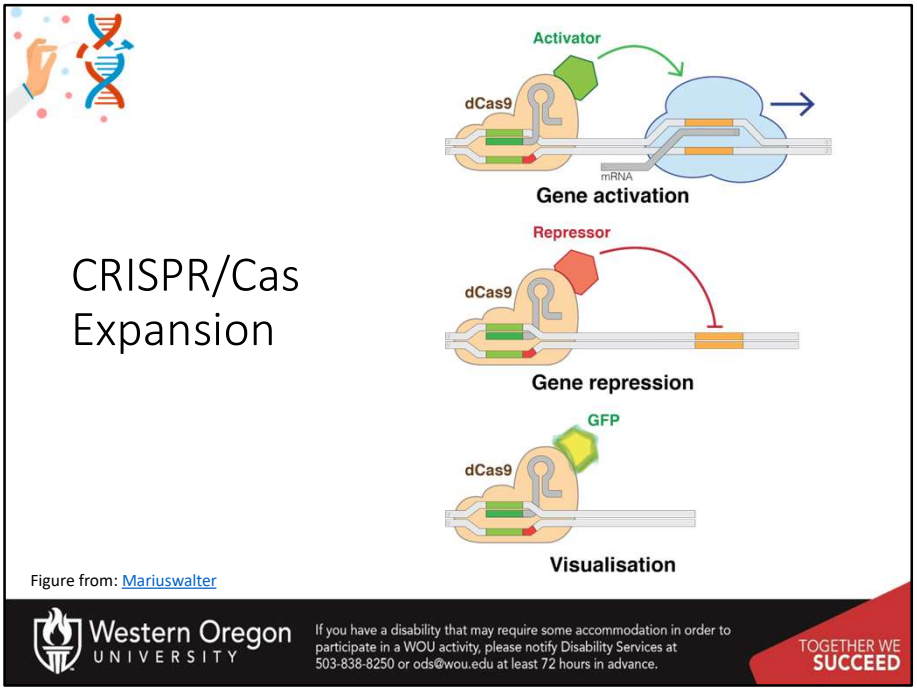

Here is a diagram showing how CRISPR works to create adaptive immunity within the bacteria during the immune response to a viral infection. The CRISPR array contains some CRISPR Associate genes (CAS genes) which encode proteins that will be used to recognize foreign DNA as it enters the cell. These CAS proteins will recognize the foreign DNA and cut small fragments from it. You'll notice that the CRISPR array also contains an AT-rich leader sequence followed by short repeats that are separated by unique spacers. The unique spacer regions come from these small viral fragments that are collected by the CAS proteins and are inserted between the clustered palindromic repeats. Small CRISPR RNAs (crRNAs) are then transcribed from the repeat region and become associated with the second type of CAS protein complex. This CAS protein complex contains helicase and restriction enzyme subunits. When it associates with the crRNAs, this complex can be targeted to bind with the infecting viral genome, where it will unwind the genome and cut it up into small, fragments that are no longer infectious. The viral fragments then remain within the host genome and serve as an adaptive memory to past infections. Causing faster immune response the next time the bacteria is exposed to a related bacteriophage.




We have adopted parts of this bacterial immune system to make a gene editing tool that can be used in living cells to inactivate or insert genes of interest. This can be used in human cells, as well, and holds great potential for the treatment of genetic diseases and disorders. Essentially the CAS9 protein is used as a targeting and editing system. There will be a guide portion of DNA that will help target the CAS9 protein in the genome, where it will then mediate the cleavage of the DNA molecule. This can cause the inactivation of a gene sequence if needed, or it can be used also insert a new fragment of DNA as well.



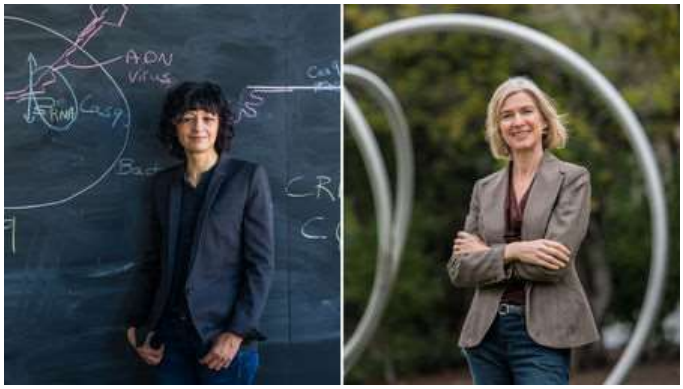

In the case of inactivation, the cleaved DNA is repaired using host repair processes such as Non-homologous end joining (NHEJ) repair, whereas, if a new gene is needed to be introduced into the genomic DNA, then a donor DNA that has homology in the region of the cleaves site, will enable the incorporation of the new DNA, through homology directed repair processes. Note that later in the term, we will come back to DNA repair processes and go through these mechanisms in more detail. But for now, know that these mechanisms are possible and can be used to alter the genetic make up of organisms, holding such great potential for advances in medical science, but also bring about ethical concerns regarding its use for more cosmetic and/or skill enhancing genetic alterations.




CRISPR/Cas systems are also finding versatility in regulating gene expression as well. The Cas protein nuclease activity can be inactivated and the protein can be coupled with transcription factor activators and repressors to influence gene expression in a more transient way. The guide system can target the transcription factor to a specific region of the DNA where it will temporarily affect gene transcription within the neighboring region. This has a lot of potential for medicine, but is also a very powerful tool for research, enabling our ability to gain a deeper understanding of biochemical regulation and cellular function.

# 2020 Nobel Prize in Chemistry

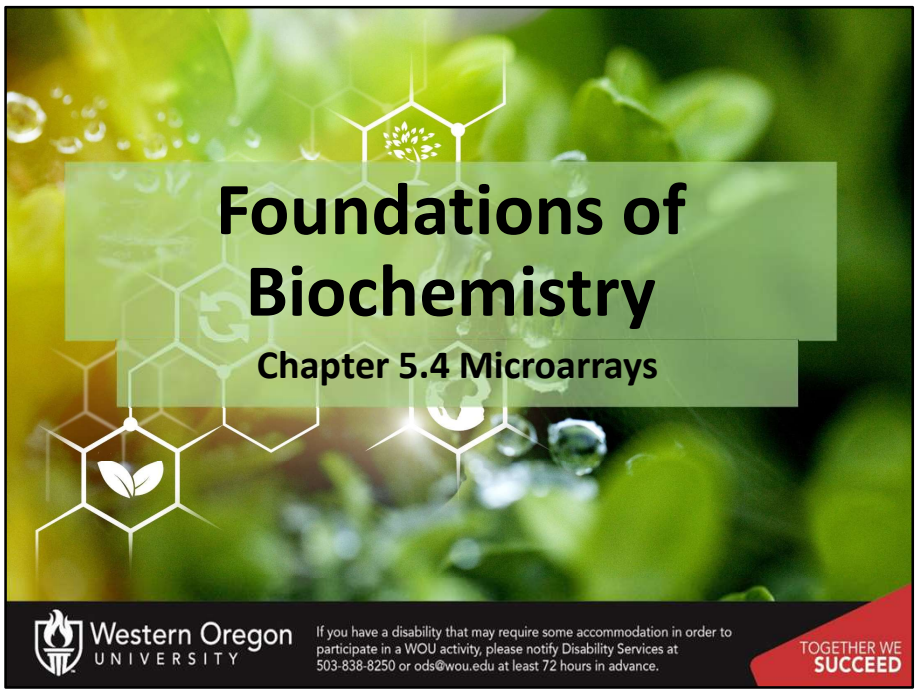
- Emmanuelle Charpentier and Jennifer A. Doudna



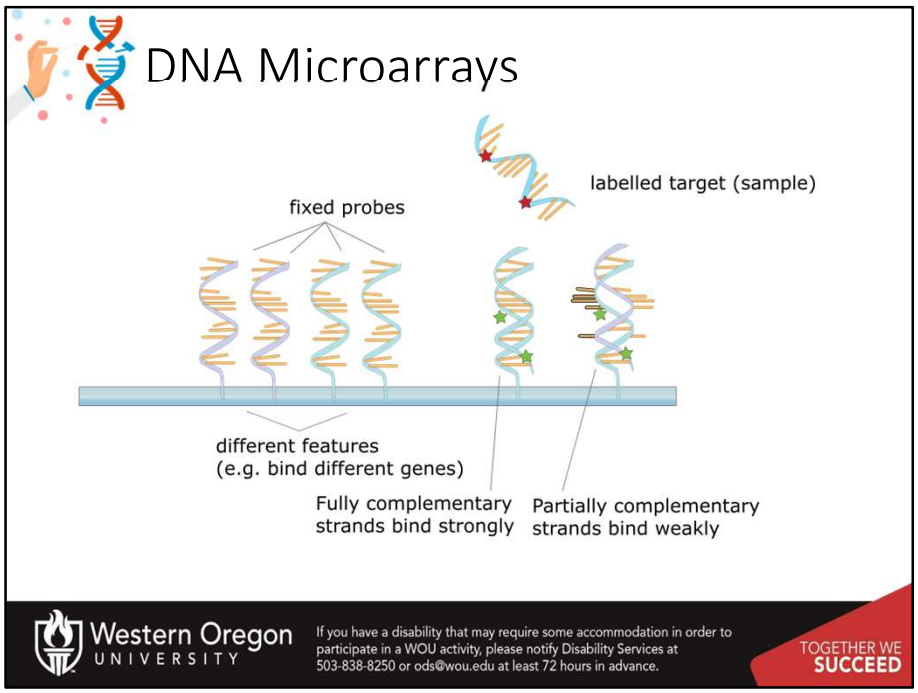
Western Oregon UNIVERSITY

If you have a disability that may require some accommodation in order to participate in a WOU activity, please notify Disability Services at 503-838-8250 or ods@wou.edu at least 72 hours in advance.

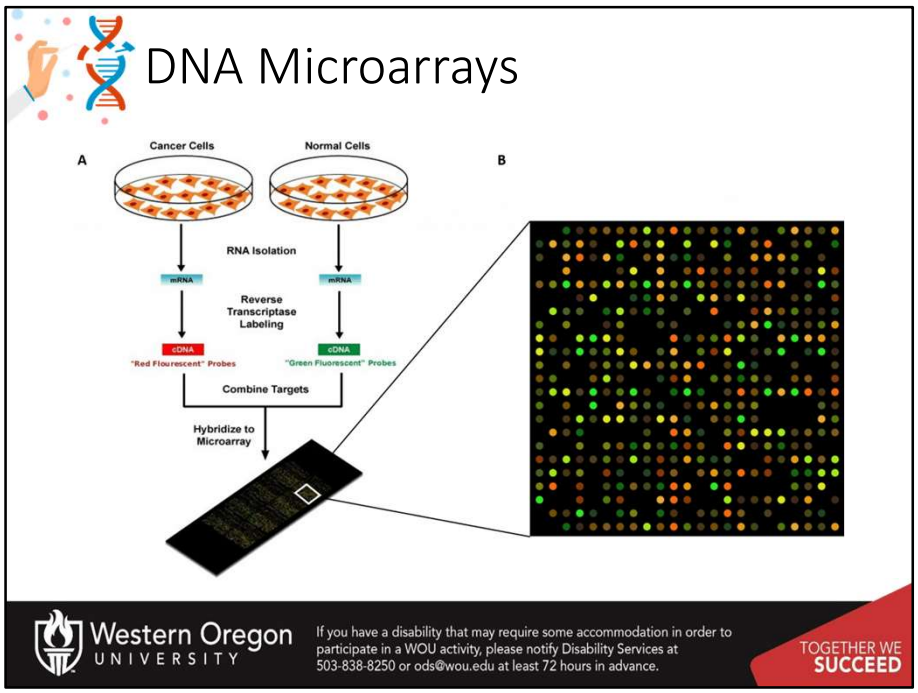
TOGETHER WE SUCCEED



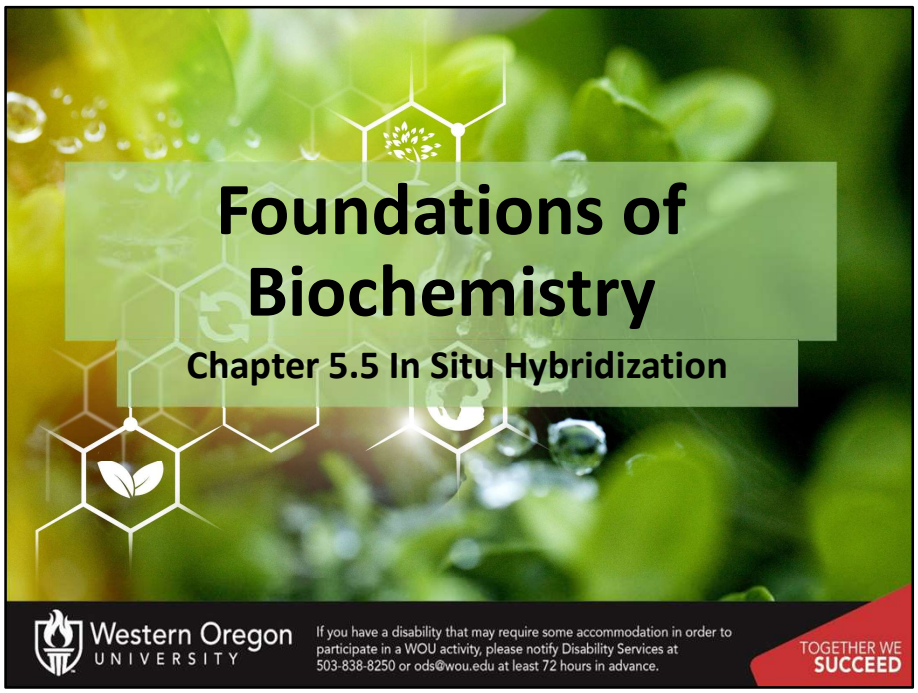

In this section, we will learn more about DNA microarrays and their use in biochemical research.



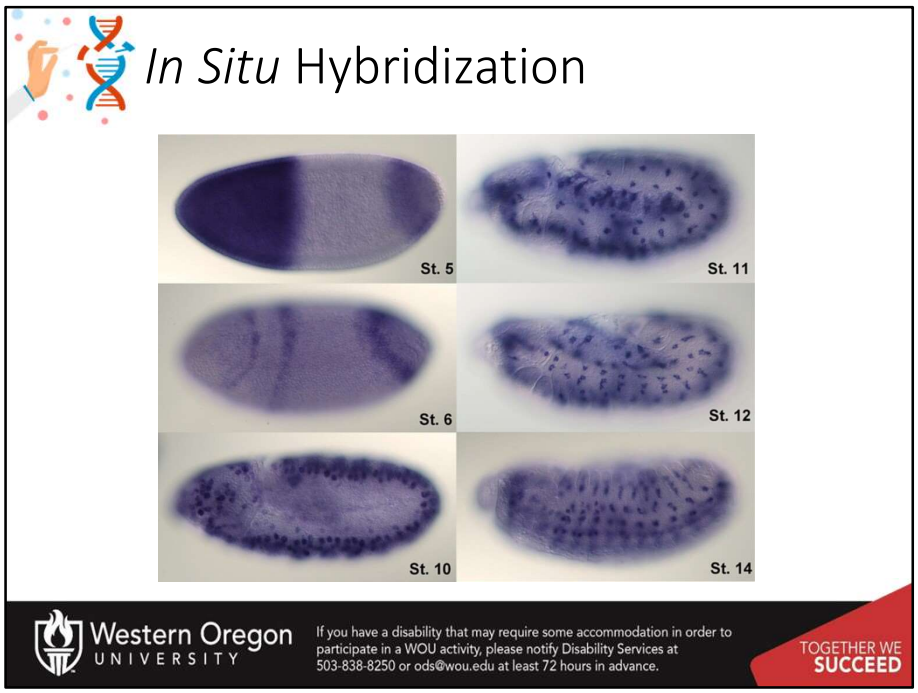

So what exactly is a DNA microarray? A microarray is a small biochip where a collection of microscopic single-stranded DNA 'spots' are attached to a solid surface (ie they are covalently crosslinked). Scientists use DNA microarrays to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome. Each DNA spot contains picomoles ($10^{-12}$ moles) of a specific DNA sequence, known collectively as *probes.* Target sequences can then be hybridized with the probe sequences to yield a signal. Signals are usually generated by a fluorophore marker that is attached to the target sequences.




Here is an example of what can be done with microarray technology. This experiment is a dual-color microarray, where mRNA is isolated from two different cell samples from the same patient, representing cancer cells and normal cell populations. Each sample set are labeled with a different fluorescent probe (in this case, the cancer cells in red and the normal cells in green). Thus, geen will indicate gene expression at that loci in normal cells, whereas red indicates gene expression in cancer cells. Yellow indicates gene expression in both samples.




In this section, we will learn more about in situ hybridization with nucleic acid probes.




In situ hybridization using nucleic acids, seeks to evaluate the expression and localization of specific mRNA within an organism, an organ, or a cell type, depending on the mode of analysis. The key techniques currently in use include *in situ* hybridization to mRNA with oligonucleotide and RNA probes (both radio-labeled and hapten-labeled), analysis with light and electron microscopes, whole mount *in situ* hybridization, double detection of RNAs and RNA plus protein, and fluorescent *in situ* hybridization to detect chromosomal sequences. For hybridization histochemistry, sample cells and tissues are usually treated to fix the target transcripts in place and to increase access of the probe. As noted above, the probe is either a labeled complementary DNA or, now most commonly, a complementary RNA (riboprobe). The probe hybridizes to the target sequence at elevated temperature, and then the excess probe is washed away (after prior hydrolysis using RNase in the case of unhybridized, excess RNA probe). Solution parameters such as temperature, salt, and/or detergent concentration can be manipulated to remove any non-identical interactions (i.e., only exact sequence matches will remain bound). Then, the probe that was labeled with either radio-, fluorescent- or antigen-labeled bases (e.g., digoxigenin) is localized and quantified in the tissue using either autoradiography, fluorescence microscopy, or immunohistochemistry, respectively.