

7.016 Recitation 12 – Fall 2018

(Note: The recitation summary should NOT be regarded as the substitute for lectures)

(This material is **COPYRIGHT protected**.)

Summary of Lecture 17 (10/17) and 18 (10/22):

PCR: is used to make multiple copies of a chosen piece of DNA. PCR is the process of doing about 30 rounds of DNA replication in a row. For PCR, in a test tube you combine the original DNA, nucleotides, a pair of DNA primers and **thermostable** DNA polymerase (isolated from the *Thermus aquaticus* bacteria found in hot springs). *This polymerase lacks the proof reading ability unlike the regular DNA polymerase.* Then the reaction mix is placed in a machine that performs multiple rounds of replication cycling through three temperatures. The first temperature (>90°C approximately) denatures the double-stranded DNA molecule, the second (50-60°C approximately) allows the primers to base-pair with the template, and the third (70°C approximately) allows the DNA polymerase to generate the newly synthesized strand 5'→ 3' starting from the 3'OH of the primer.

You can also use PCR to identify the recombinant clone, from the ligation mix, that contains the gene of your interest and this is often called cloning by PCR. It is important to note this can be used to select a gene that has not been conventionally cloned already.

<https://www.youtube.com/watch?v=iQsu3Kz9NYo>

DNA sequencing (NextGen sequencing): DNA sequencing is the technique by which the nucleotide sequence of DNA is determined. DNA sequencing is the process of doing DNA replication in the presence of all four nucleotides (dNTPs), and a small percentage of “dideoxynucleotides (2', 3' di-deoxynucleotides or ddNTPs) that act as chain terminators and do not allow for replication to continue any further. For DNA sequencing, in a test tube you put the original template DNA, the four deoxynucleotides, DNA polymerase, a primer, and then a small concentration of “ddA” (colored red), “ddT” (colored blue), “ddC” (colored yellow), and “ddG” (colored green). Most of the time, the deoxynucleotides (dNTPs) will be added and the template will be extended normally. Occasionally, a ddC (for example) will be put in across from a G in the template. This will stop replication and label the fragment yellow. You allow the reaction to proceed and then load all of the replication products onto a DNA sequencing gel that separates them by size. The sequence can then be read from the result of that gel (in earlier years) or by a machine now.

<https://www.youtube.com/watch?v=3M0PyxFPwkQ>

<https://www.youtube.com/watch?v=vK-HIMaitnE>

<https://www.youtube.com/watch?v=jFCD8Q6qSTM>

Genome sequencing: Genomes are sequenced by randomly breaking them into pieces, which are sequenced by machines. A computer then analyzes those sequences and reassembles them into one long continuous piece (for a bacterium with a single circular chromosome) or several linear pieces (for a eukaryote with multiple chromosomes). Once the genome is sequenced, programs can predict where the genes are in the genome by looking for the signatures of genes, like start codons and stop codons, long open reading frames, promoters, and splice sites (if applicable). Each gene in the genome can then be compared to all other known sequences using different programs such as BLAST. If a new gene shows a large amount of homology to a previously studied gene, the most likely possibility is that the new gene encodes a protein with a similar function to the protein encoded by the previously studied gene.

Now you can get your genome sequenced by companies such as 23 and me.

<https://www.23andme.com/?myg=true>

Single nucleotide polymorphisms (SNPs) and SNP microarrays: In order to understand the genetic bases of different diseases, biologists can use rapid genotyping technologies to create haplotype maps, which are used to identify the single nucleotide polymorphisms (SNPs) that are linked to genes involved in diseases. A set of SNPs that are present on a segment of chromosome are usually inherited as a unit. This linked piece of chromosome is called a haplotype. SNPs can be anywhere in the genome - in between the coding regions of the gene, in introns or in intergenic regions between two genes. SNP defined haplotypes can be used as markers to predict disease-associated phenotypes of an individual. These associations have revealed particular haplotypes that are associated with modestly increased risks for breast cancers, diabetes, arthritis, obesity and coronary heart disease.

<https://learn.genetics.utah.edu/content/precision/snips/>

Gene editing by Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR): CRISPR systems are adaptable immune mechanisms used by many bacteria to protect themselves from foreign nucleic acids, such as those of viruses.

Here you introduce a guide RNA (gRNA) that is complementary to and can hybridize with the target gene of interest. You can introduce the gRNA into a cell or an embryo. The target sequence has a Cas-9 nuclease site. Once the gRNA hybridizes with the target DNA, the Cas-9 nuclease makes a double stranded break in the target DNA. These breaks can be repaired by...

-Non-homologous end joining (**NDHJ**), imprecise repair of this double stranded break results in addition or deletion of nucleotides within the target gene thus changing the reading frame.

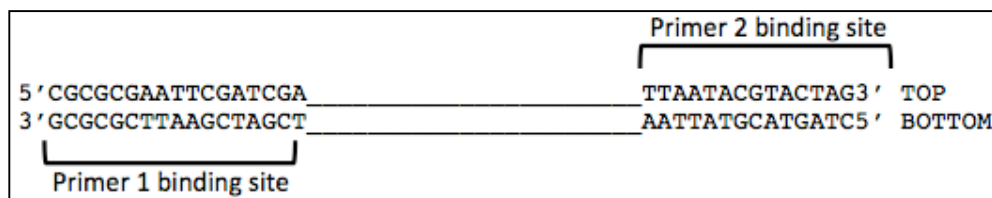
- Homologous directed repair (**HDR**) where the targeted sequence replaced the mutated sequence.

<https://www.youtube.com/watch?v=2pp17E4E-O8>

<https://www.youtube.com/watch?v=0dRT7slyGhs>

Questions

1. You are interested in making many copies of a specific DNA sequence. The sequence to be amplified is flanked by regions whose sequence is given below:



Primer 1

Set A: 5' TCGATCGAATTC 3'

Set B: 5' GCTTAAGCTAGC 3'

Set C: 5' GAATTCGATCGA 3'

AND

AND

AND

Primer 2

5' TAATACGTACTA 3'

5' GATCATGCATAA 3'

5' CTAGTACGTATT 3'

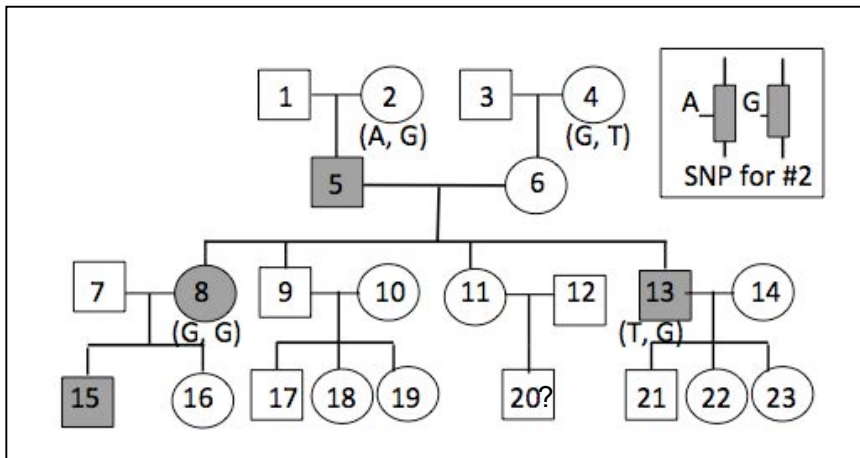
Select the primer pair from above that you can use to PCR amplify the gene: **Set 1/ Set B/ Set C?**

2. You decide to sequence the following piece of DNA.

5'-...GTATAGCCTCCGT...-3'
 3'-...CATATCGGAGGCA...-5'

- a) The primer you use for sequencing is 5'-ACG-3'. **Note:** real primers are usually 18-22 nts long). You set up a DNA sequencing reaction with this DNA, this primer, DNA polymerase, dATP, dTTP, dGTP, dCTP, and ddATP. How many different products would form from this reaction?
- b) How many nucleotides long would each different product of this reaction be?
- c) Write the sequence of DNA that you read. Label the 5' and 3' ends of the molecule.

3. The following human pedigree shows the inheritance of a RARE disease. **Note:** Individuals marrying into the family (7, 10, 12, 14) do not have the disease-associated allele. Assume that no other mutation arises within the pedigree. Also assume complete penetrance. The inset shows the location of the SNPs on the genes associated with the disease for Individual 2. Use the symbol X^D , X^d , D or d where appropriate. In each case, use the letter "D" to represent the allele associated with the dominant phenotype and "d" to represent the allele associated with the



a) What is the **most likely** mode of inheritance of this disease?

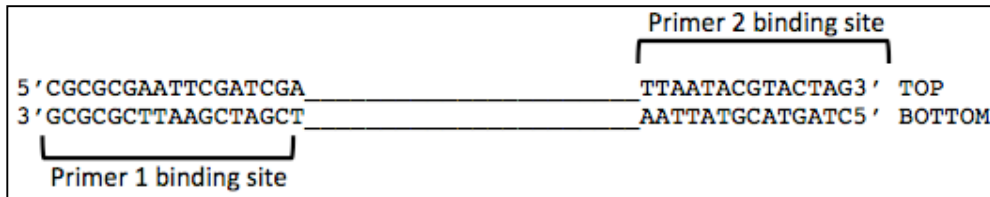
b) The disease shown by the pedigree above is caused by a mutation in Gene D that encodes Protein D. You identify a SNP that is **tightly linked** to Gene D and may be used **as a marker for the disease**. The alleles (A, G, T, C) of this SNP for some individuals are given in the pedigree above.

- i. Identify the SNP that is tightly linked with the mutant allele of Gene D.
 - ii. Write the SNP genotypes of the following **Individual 5**: _____
- c) What is probability that #20 will be affected? _____

4. You are given a sequence, which encodes a protein that you would like to knock out to study how it contributes to a cell. **Explain** how you can use the CRISPR- Cas-9 technology to achieve this.

The key

1. You are interested in making many copies of a specific DNA sequence. The sequence to be amplified is flanked by regions whose sequence is given below:

**Primer 1****Primer 2****Set A:** 5' TCGATCGAATTC 3'

AND

5' TAATACGTACTA 3'

Set B: 5' GCTTAAGCTAGC 3'

AND

5' GATCATGCATAA 3'

Set C: 5' GAATTCGATCGA 3'

AND

5' CTAGTACGTATT 3'

Select the primer pair from above that you can use to PCR amplify the gene: **Set 1/ Set B/ Set C?**

Note: You need a pair of primer to PCR amplify both strands of DNA. The strands should bind to region flanking the sequence of your interest. The binding of primer should be antiparallel to the template strand. It is the 3'-OH end of the primer that gets elongated by Taq DNA polymerase.

Set A: won't work since although primer 1 may be the primer for the bottom strand, primer 2 is not the primer or the top strand.

Set B: primer 1 is binding to the top strand 3'->5' instead of 5'->3', so it will not work.

2. You decide to sequence the following piece of DNA.

5'...GTATAGCCTCCGT...-3'

3'...CATATCGGAGGCA...-5'

a) The primer you use for sequencing is 5'-ACG-3'. **Note:** real primers are usually 18-22 nts long). You set up a DNA sequencing reaction with this DNA, this primer, DNA polymerase, dATP, dTTP, dGTP, dCTP, and ddATP. How many different products would form from this reaction?

Since you are using ddATP in the reaction mixture halts whenever there is an incorporation of a ddATP instead of dATP. Since there are 3A's in the sequence (excluding the A that is a part of the primer) and therefore you may have 3 reaction products of different lengths. There is always a probability that you may have a reaction product where on dATP is incorporated and NO ddATP is incorporated at all. Combining these two together you may have 4 reaction products (5b, 10b, 12b and 13bases long)

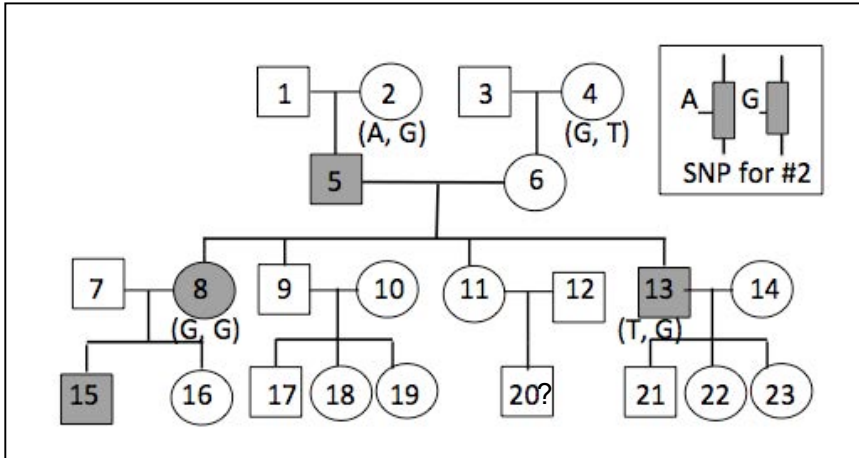
b) How many nucleotides long would each different product of this reaction be?

They would be 5b, 10b, 12b and 13bases long.

c) Write the sequence of DNA that you read. Label the 5' and 3' ends of the molecule.

5'...GAGGCTATAC...-3' (the first three bases are the primers)

3. The following human pedigree shows the inheritance of a RARE disease. **Note:** *Individuals marrying into the family (7, 10, 12, 14) do not have the disease-associated allele. Assume that no other mutation arises within the pedigree. Also assume complete penetrance. The inset shows the location of the SNPs on the genes associated with the disease for Individual 2. Use the symbol X^D , X^d , D or d where appropriate. In each case, use the letter "D" to represent the allele associated with the dominant phenotype and "d" to represent the allele associated with the*



a) What is the **most likely** mode of inheritance of this disease?

X-linked recessive

b) The disease shown by the pedigree above is caused by a mutation in Gene D that encodes Protein D. You identify a SNP that is **tightly linked** to Gene D and may be used as a marker for the

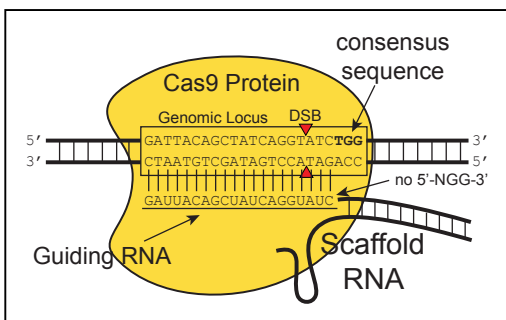
disease. The alleles (A, G, T, C) of this SNP for some individuals are given in the pedigree above.

i. Identify the SNP that is tightly linked with the mutant allele of Gene D: G

ii. Write the SNP genotypes of the following Individual 5: G or $X^G Y$

c) What is probability that #20 will be affected? 50% since #11 is a carrier

4. You are given a sequence, which encodes a protein that you would like to knock out to study how it contributes to a cell. Explain how you can use the CRISPR- Cas-9 technology to achieve this.



You will design a 20-bp guide RNA that targets the Cas9 endonuclease to a genomic target locus by hybridization with unwound DNA. Following recognition of the complementary strand, Cas9 creates a double stranded break (DSB, indicated with arrowheads) that the cell can repair in an error-prone manner that introduces insertions or deletions. The guide RNA requires a 5'-NGG-3' consensus sequence at the 3' end of the genomic target strand (where N could be any nucleotide). This is shown in the schematic to the left.

MIT OpenCourseWare
<https://ocw.mit.edu/>

7.016 Introductory Biology
Fall 2018

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.