A Little Bit about *Escherichia coli*

E. coli eats glucose.

But *E. coli* can also eat lactose, if need be.

To do so, it uses an enzyme called beta-galactosidase which breaks lactose down into glucose and galactose. (Then it eats the glucose.)

If *E. coli* is cultured on a glucose tray, then the bacterium produces none of the beta-galactosidase enzyme. But if it is cultured on a lactose tray, in the absence of glucose, then the bacterium starts producing beta-galactosidase within a few minutes.

How?

Beta-galactosidase

Beta-galactosidase is an enzyme, which is a protein, which means that *E. coli* has some gene which codes for that protein.

Apparently, sometimes that gene is producing its protein product (expressed), and sometimes it is not.

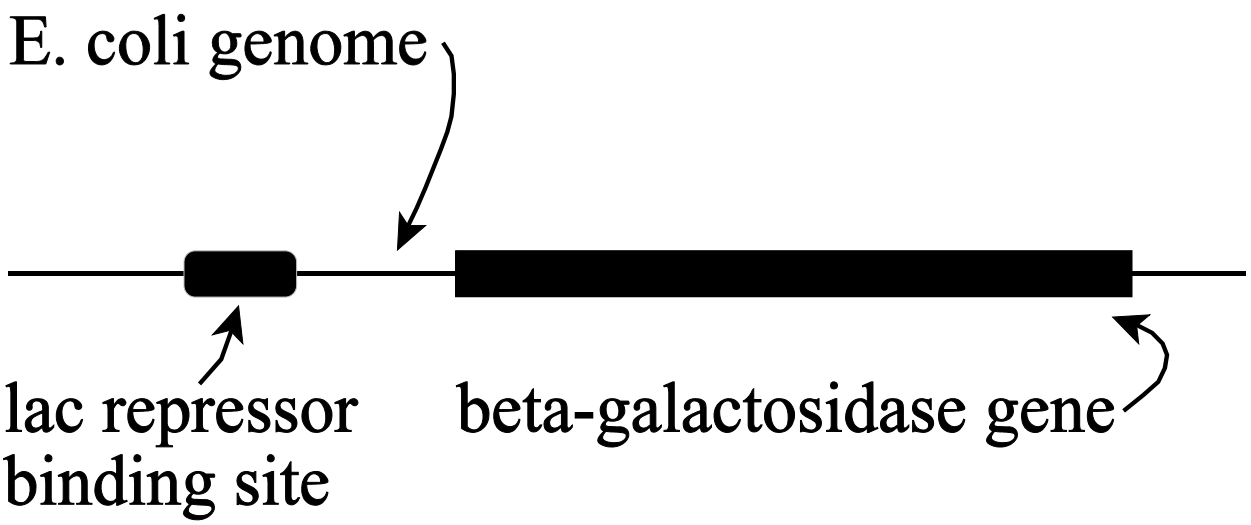
What mechanism controls whether this gene is expressed or note?

2. A protein called *lac repressor*

3. A locus near the beta-galactosidase gene to which lac repressor binds, and which prevents transcription from starting at the beta-galactosidcase gene

4. A rule for when beta-galactosidase is or is not present at that locus

The Beta-galactosidase Repressor



When lac repressor attaches at its binding site, expression of the beta-galactosidase gene is suppressed. Lac repressor is designed perfectly to attach to this locus.

When lactose is present, lac repressor becomes chemically unable to stay attached at the binding site, and gene expression occurs.

This is an example of a genetic "on/off" switch.

Enhancers and Repressors

Genetically, we are more than 99% similar to a chimpanzee. Most of the difference between us and them comes not from the protein products of the genes, but the signaling which takes place during embryonic development, turning genes on and off at the right times.

For example, polydactyly arises when the "make an appendage" gene is not suppressed at the right time, or is enhanced when it should not be.

The point...

What is important for our purposes today is the question of how to discover these binding sites, and how to know which proteins bind to them.

Ultimately, we wish to understand the structure of gene expression networks.

The Neutral Theory of Evolution

Mutations occur within genomes at some constant rate, without regard to where the mutations are taking place.

Sometimes the mutation kills the creature, sometimes it does not matter, and sometimes it makes the creature different, and perhaps better.

Natural Selection is then responsible for deciding which creatures live.

Mutations within a Gene

If the mutation takes place within a gene, it may or may not affect the organism.

Synonymous substitutions: Substitutions which do not affect the protein product.

Non-synonymous substitutions: Substitutions which do affect the protein product.

	Second Base				
First Base	U	C	A	G	Last Base
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Intergenic Mutations

What about mutations between the genes, or within the introns? These are *non-coding* regions, and have no protein product.

These regions typically show a higher degree of mutation, and thus a lower degree of similarity. Portions of this part of the genome which *do* show a high degree of similarity are therefore assumed to be functionally important, and probably play a role in gene expression.

This technique, of looking for regions highly conserved between two or more genomes, in order to find functionally important regions, is called *phylogenetic footprinting*.

Exercises — Phylogenetic Footprinting

Presentation Problems:

1. Suppose that two genomes are sufficiently divergent that, when aligned, it may be assumed each base pair has a 3/4 probability of being identical in the two sequences, and a 1/4 probability of being different. Given a section of the DNA alignment of length 8,
   a. What is the probability that all base pairs in that section are the same?
   b. What is the probability that there will be exactly one mutation in that section?

2. Genome rearrangement question: The two matrices which follow are alignments of two sequences according to the Smith-Waterman algorithm. The first matrix has both sequences in the "forward" direction. The second has sequence 2 (the one down the left column) reversed.

   Your job is to find out how to split the first sequence into a few pieces (the fewer, the better) so that they can be rearranged to make the second sequence in the same way that the mouse X chromosome could be rearranged to make the human X chromosome (see the first slide in your notes).

   Finally, once you have identified the pieces (there are about 5 of them) find the optimal sequence of reversals to transform the first sequence into the second.

(See Rearrangement workshop for the tables for this problem.)