Genetics Lecture Notes

7.03 2005

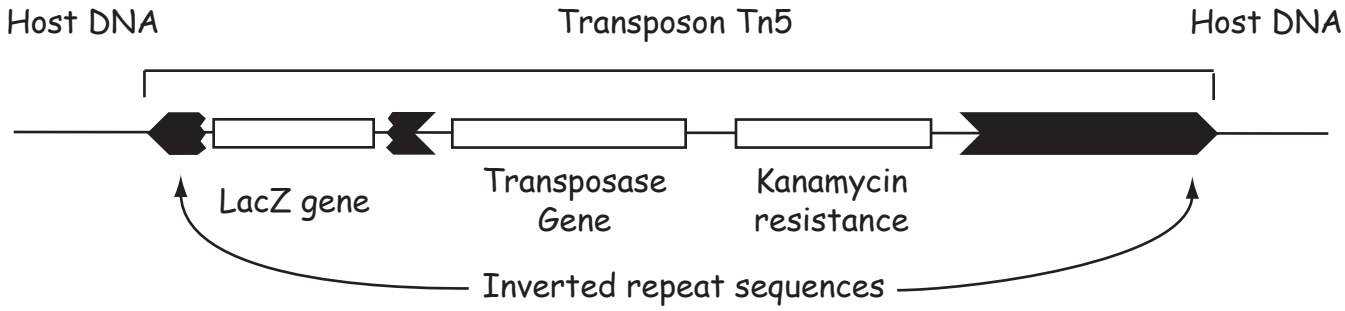
Lectures 13 – 16

Lecture 13

Transposable elements

Transposons are usually from 10^3 to 10^4 base pairs in length, depending on the transposon type. The key property of transposons is that a copy of the entire transposon sequence can at a low frequency become inserted at a new chromosomal site. The mechanism by which transposons insert into new sites differs from one kind of transposon to another, but the details are not important to understand how transposons can be used. It is worth contrasting the recombination events that occur during transposition to the homologous recombination events that we have considered in meiosis. In homologous recombination, crossovers occur between like sequences. While this type of recombination can generate new combinations of alleles the arrangement of genes is left undisturbed. In contrast, transposition involves recombination between unrelated sequences, namely the ends of the transposon and a site in the target sequence. Transposition therefore results in a new arrangement of genes along the chromosome.

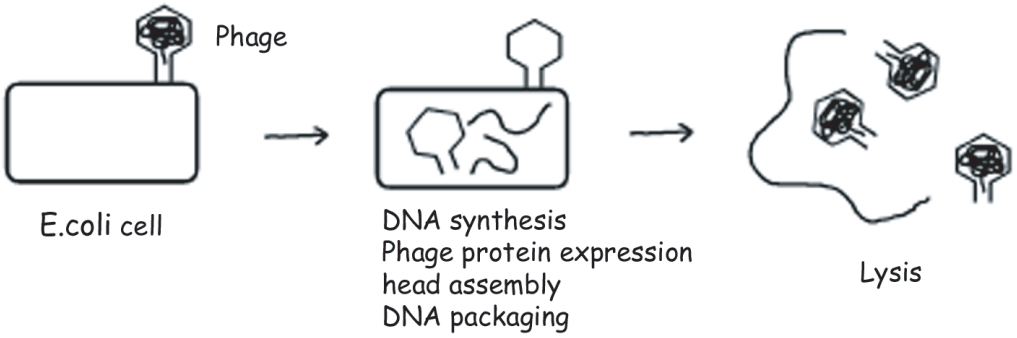
The generic structure of a transposon looks like this:



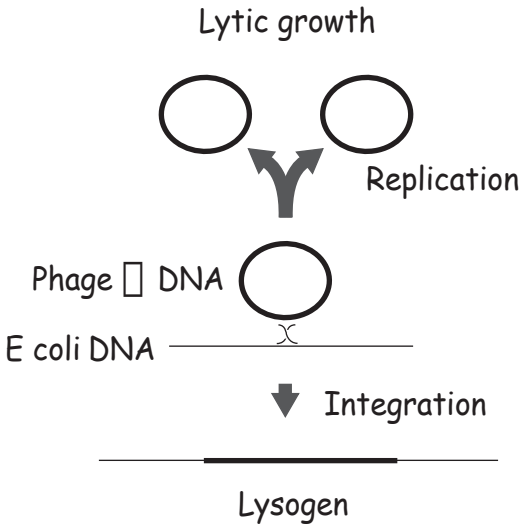
Transposon Element	Function
Transposase	An enzyme that cuts the target DNA more or less at random and splices the transposon ends to the target sequences, Other steps in transposition are performed by host enzymes.
Inverted Repeats	These sequences direct transposase to act at the ends of the transposon. Note that because the sequences are inverted, the two ends have identical sequence.
Selectable Marker(s)	Transposons are thought to have evolved by providing a selective advantage to the host cell. Many transposons carry genes that confer antibiotic resistance or some other benefit to the host.

The study of transposition mechanism and the biology of transposons is an interesting subject in genetics but for our current purposes we are going to concentrate on how transposons can be used for bacterial genetic analysis. For this purpose we will focus on the transposon Tn5 which can function in E. coli as well as a wide variety of other bacterial species. The selectable marker in Tn5 is a gene that confers resistance to the antibiotic kanamycin. Thus bacteria without Tn5 are sensitive to kanamycin (Kan^S), whereas bacteria that have Tn5 inserted into the chromosome are resistant to kanamycin (Kan^r).

One of the best ways to introduce a transposon into E. coli is by using a virus that infects bacteria known as a phage. For this purpose we will use a well studied type of phage known as □.



Once the DNA from phage □ enters a cell it circularizes and then can undergo two possible fates.



In the lytic growth mode phage DNA is replicated in preparation for packaging into new phage particles. Among the phage genes required for replication is the **P** gene.

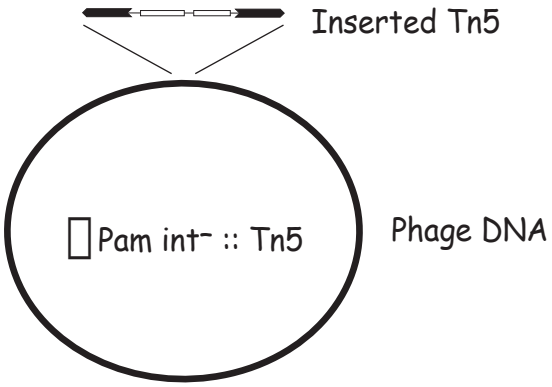
In the lysogenic mode phage DNA is integrated into the chromosome. This quiescent state the phage is replicated along with the chromosome in a state known as a lysogen. The phage gene required for integration is the **Int** gene.

To introduce random insertions of Tn5 into the E. coli chromosome we will start with Tn5 carried on a special λ phage vector: λ Pam int⁻::Tn5.

Pam designates an amber (nonsense) mutation in the λ phage P gene. When λ Pam phage infect E. coli with an amber suppressor (Su^+) the phage multiply normally, which allows propagation of these mutant phage. But when λ Pam phage infect a nonsuppressing host (Su^-) the phage DNA cannot replicate.

int– is a mutation in the λ integrase gene. Phage with this mutation can not integrate into the host chromosome to make a stable prophage.

::Tn5 designates that the λ phage carries an inserted copy of Tn5.



When λ Pam int–::Tn5 infects a wild type (Su^- $Kan^{s)}$ E. coli host, the phage DNA cannot replicate (Pam) nor can it integrate (int^-) thus the only way for the E. coli to become Kan^r is for Tn5 to transpose from the λ DNA to some location on the E. coli chromosome. This type of transposition is an inherently rare process and will occur in about one out of 10^5 phage-infected E. coli cells.

This is how a transposon mutagenesis can be done:

1) Infect $2x10^9$ wild-type E. coli cells with λ Pam int^-::Tn5 so that each cell receives at least one phage chromosome.

2) Select for Kanr by plating on medium that contains kanamycin. There should be a total of about $2x10^4$ Kan^r colonies. Each of these should have Tn5 inserted into a different site on the E. coli chromosome.

The genes of E. coli are densely spaced along the chromosome and about half of the Tn5 insertions will lie in one gene or another. There are 4,200 genes in E. coli so our collection of $2x10^4$ random Tn5 insertions will likely contain at least one insertion in each gene. (Note that insertions in genes that are essential for E. coli growth such as the genes for RNA polymerase or ribosomal subunits will not be recovered because these insertion mutants will not form colonies on the kanamycin plates).

Let's say that we are interested in the E. coli genes that are involved in synthesis of histidine. To find insertion mutants that can not synthesize histidine (His^-) we could screen amongst our collection of $2x10^4$ random Tn5 insertions to find those that are His^-. The easiest way to do this would be to plate out the collection of insertions at a density of 200 colonies per plate (100 plates total). Each of these master plates would then be replica plated (first by transfer to a sterile piece of velvet) to a plate that contains histidine and also to a plate that lacks histidine. His^- insertion mutants would be identified as colonies that can not grow on the plates that lack histidine. Note that the same collection of random Tn5 insertions can be screened multiple times to find interesting mutations with different phenotypes.

3) Identify His^- Tn5 insertion mutants by replica plating to find colonies that specifically can not grow on plates that don't contain histidine.

Once we have a set of His^- insertion mutations (in the present example, one might expect to find 10-20 different His– mutants), the affected gene(s) can be identified by the simple fact that they will be "tagged" by the inserted Tn5 sequences. The easiest way to identify the site of insertion is by performing a special PCR amplification of the DNA fragment that corresponds to the novel junction between Tn5 and the bacterial chromosomal sequences. Ordinarily PCR reactions are carried out using two DNA primers, each of which corresponding to an end of the sequence to be amplified. When we want to amplify a junction fragment we can use as one of the primers a sequence that lies near the end of Tn5 but we won't yet know the relevant chromosomal sequence to allow the other primer to be designed. There are several tricks that can be used to circumvent this problem, which are too complicated to describe here. Suffice it to say that there are ways that the junction fragment can be amplified by PCR using only sequences defined by the Tn5 portion of the junction fragment.

4) Use the known sequence of the end of Tn5 to PCR amplify a fragment that spans the junction between the end of Tn5 and the E. coli chromosomal site that was the target for insertion. DNA sequencing of the amplified junction fragments will give the identity of the target sequences. Since we know the DNA sequence of the entire E. coli chromosome, the gene that was the target for Tn5 insertion can be identified unambiguously.

5) The DNA sequence of the junction fragments will identify all of the genes that have been inactivated to give the His^- phenotype.

The procedure just outlined can be used to isolate and characterize a wide variety of useful mutations. A major limitation of this method is that as stated earlier, transposon mutations usually completely disrupt the target gene and therefore lead to a complete inactivation of the gene product. Often we will want to work with point mutations (such as temperature sensitive mutations or nonsense mutations). In the next lecture we will see how transposons can also be used to facilitate analysis and manipulation of point mutations.

Lecture 14

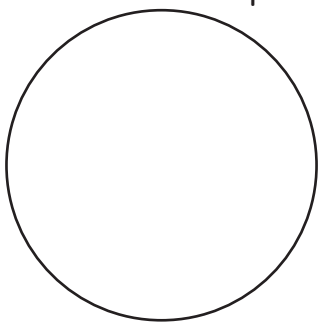
Gene Manipulation in Bacteria

There is no meiosis in bacteria so special techniques have been worked out for manipulating genes in bacteria so that mapping experiments, strain construction, and complementation tests can be done.

First, we need a way of getting chromosomal DNA from one cell into another. There are several ways to do this. All of the methods have in common the use of special extra chromosomal elements for mobilizing chromosomal genes; the methods differ according to which extra chromosomal element is used.

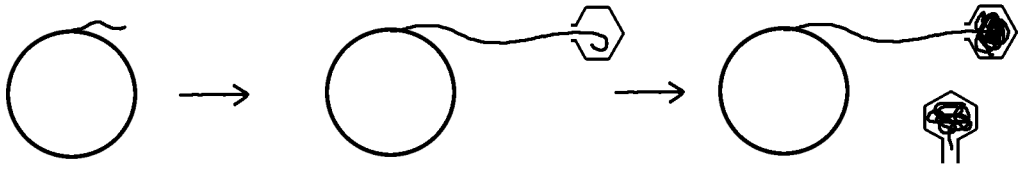
We will consider a method that uses phage and is known as Transduction

E. coli chromosome
4.6×10^6 base pairs

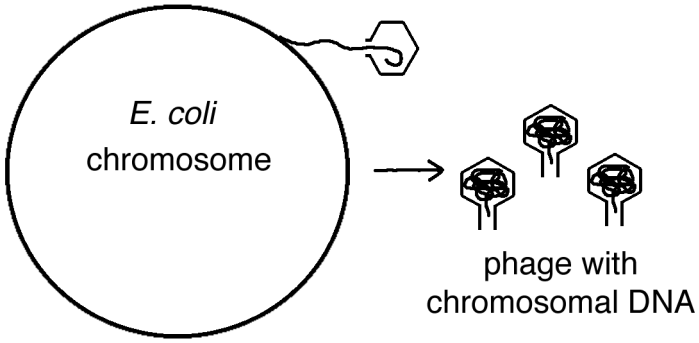


Phage P1 chromosome
10^5 base pairs

After infection of E. coli, the phage DNA is replicated by a mechanism known as a "rolling circle" and the phage is packagedinto phage particles one headfull at a time:



1/300 phage mistakenly packages E. coli chromosome DNA instead of phage DNA.

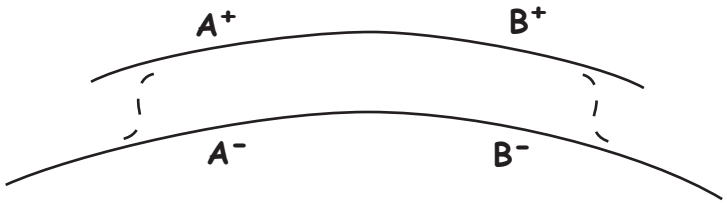


Each phage particle will package about 1/50 of the E. coli chromosome. By combining probabilities we see that about 1/15,000 phage will carry a particular E. coli gene.

A basic transduction experiment to measure the linkage between markers A and B is done as follows:

(1) Grow P1 on A^+B^+
(2) Infect A^-B^-
(3) Select for A^+ and then screen for B^+

The idea is that we are looking for the rare cases where some chromosomal DNA carrying gene A is moved into the recipient. To find these recombinants, we select for A^+. Then we screen for B^+ to see how often gene B comes along with gene A.

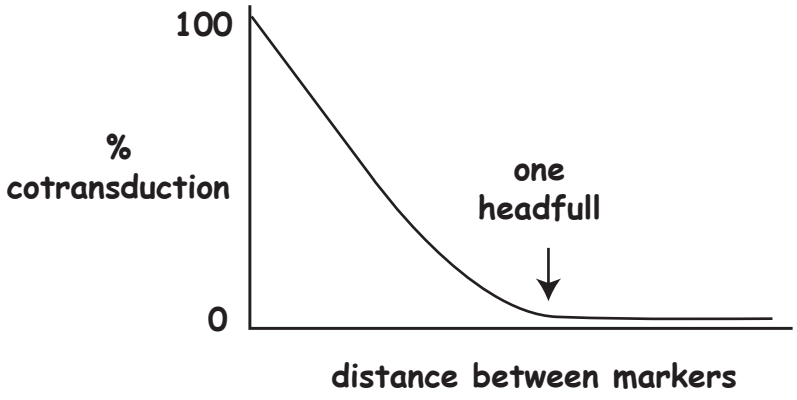


The measured frequency of cotransduction of B with A gives a measure of distance according to the following rules:

• If distance between A and B is greater than one headfull (10^5 bp) then there will be no cotransduction.

• If A and B are very close together then there will be 100% cotransduction.

• Cotransduction frequency is an inverse measure of distance.



The experiment just described is the bacterial equivalent of a 2-factor cross and will give us relative distances between genes.

We can also do a 3-factor cross to determine gene order.

(1) Grow P1 on $A^+B^+C^+$
(2) Infect $A^-B^-C^-$
(3) Select for A^+ and then screen for B^+ and/or C^+

Genotypes

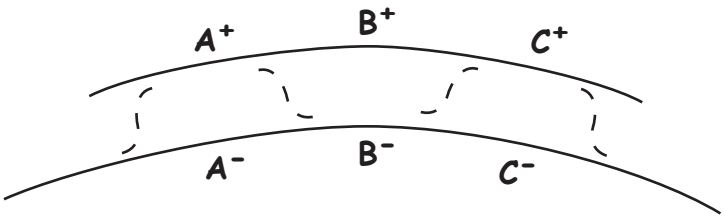
$A^+B^+C^+$ 2 crossovers ⟵ (A to C distance)

$A^+B^+C^-$ 2 crossovers ⟵ (A to B distance)

$A^+B^-C^-$ 2 crossovers

$A^+B^-C^+$ 4 crossovers (very rare)

(Note that there are only four possible genotypes because we select A^+)



A limitation of transduction experiments is the need for a good selectable marker. Tn5 insertions provide a way to extend the utility of transduction for mapping and strain construction.

For example, let's say that we have isolated a new mutation in the MotA gene. MotA is a component of the bacterial flagellar moter and $MotA^-$ mutants are nonmotile, a phenotype easily detected by the inability of $MotA^-$ colonies to "swarm" outward on soft agar plates. Imagine that we want to map the MotA– mutation or to move this mutation into an E. coli strain with a new genetic background. Clearly direct transduction of $MotA^-$ would not be possible since we have no way to select for rare (1/15,000) transductants with the nonmotile $MotA^-$ phenotype. One solution would be to use a nearby marker for which we can select to move $MotA^-$ by its cotransduction with the selectable marker. Unfortunately, good selectable markers are not common and we are unlikely to have a good selectable marker placed within cotransduction distance of $MotA^-$ readily available. A powerful alternative approach would be to isolate a random Tn5 insertion that is close to $MotA^-$ and to use the Kan^r trait conferred by Tn5 as the selectable marker for cotransduction.

The steps for finding a linked Tn5 insertion are as follows:

1) Start with a collection of random Tn5 insertions into wild type E. coli (the isolation of such a collection was described in last lecture). Grow phage P1 on the mixture of 2×10^4 different Tn5 insertion mutants. Note that this donor strain is $MotA^+$.

2) Use the resulting P1 phage to infect a $MotA^-$ recipient strain. Select for transduction of the Tn5 insertions by selecting for growth of the transductants on kanamycin plates. Screen for cotransduction of $MotA^+$ by testing each of the Kan^r transductants for motility on soft agar. The desired cotransductant will be Kan^r and will be motile. Given that one P1 phage headfull corresponds to about 1/50 of the E. coli chromosome, about 1 in 500 Tn5 insertions will be close enough to the MotA gene to show 90% cotransduction. Thus if we test about 10^3 Kan^r transductants for motility, we are likely to find at least one that has cotransduced the $MotA^+$ marker.

3) Once a Tn5 (Kan^r) $MotA^+$ transductant has been identified, grow P1 on Tn5 (Kanr) $MotA^+$.

4) Use the P1 phage from step 3) to infect a $MotA^-$ recipient strain. Select for transduction of the Tn5 insertions by selecting for growth of the transductants on kanamycin plates.

Test the resulting Kan^r transductants for their motility. The transductants that have cotransduced the $MotA^+$ marker will be motile, whereas the transductants still contain the $MotA^-$ allele will be nonmotile. The fraction of the total transductants that are motile will give the distance between $MotA^-$ and the Tn5 insertion as a cotransduction frequency.

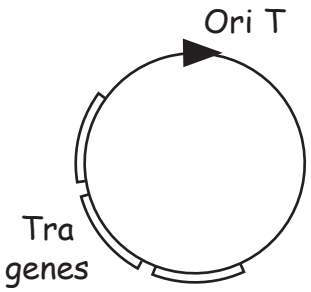
A Tn5 (Kanr) $MotA^-$ transductant isolated in step 4) can then be used to transduce the $MotA^-$ marker into a new recipient strain by cotransduction with Tn5. Note that if we had isolated a second $MotA^-$ mutant, transduction into this strain would amount to a 3-factor cross and would provide a way to determine the order of the two different $MotA^-$ alleles.

Lecture 15

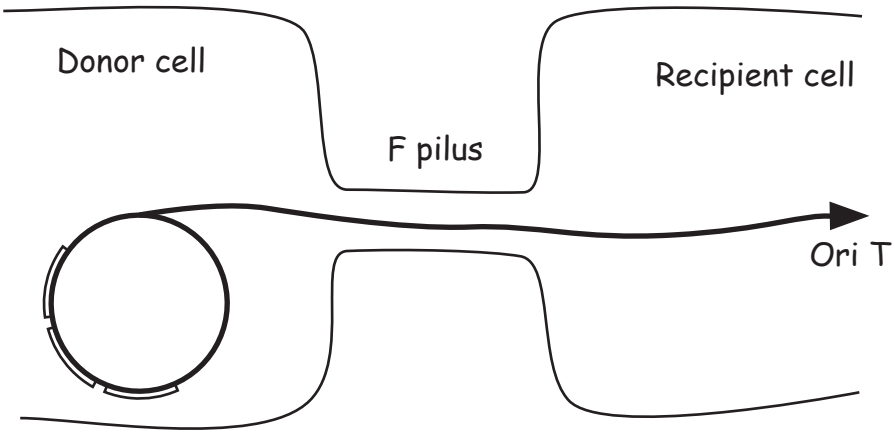
Plasmid conjugation in Bacteria

In order to perform tests for dominance or for complementation in bacteria we need a way to make the bacteria diploid for part of the chromosome. To do this we need to consider a different extrachromosomal element:

The **F** plasmid
(length 10^5 base pairs)



There are some special terms to describe the state of F in a cell: F^- refers to a strain without any form of F, whereas F^+ refers to a strain with an F plasmid.



F is very efficient at transferring itself from an F^+ cell to an F^- cell. After culturing F^+ and F^- cells together about 1/10 of the F^- cells will become F^+.

The property that makes F useful for genetic manipulation is that at low frequency the plasmid will integrate into chromosome. This occurs because F carries insertion sequences that are also present at multiple locations on the chromosome. Crossing over between insertion sequences on F and on the chromosome gives integration.

Hfr: a strain with F integrated into the chromosome that will give efficient transfer of some chromosomal markers.

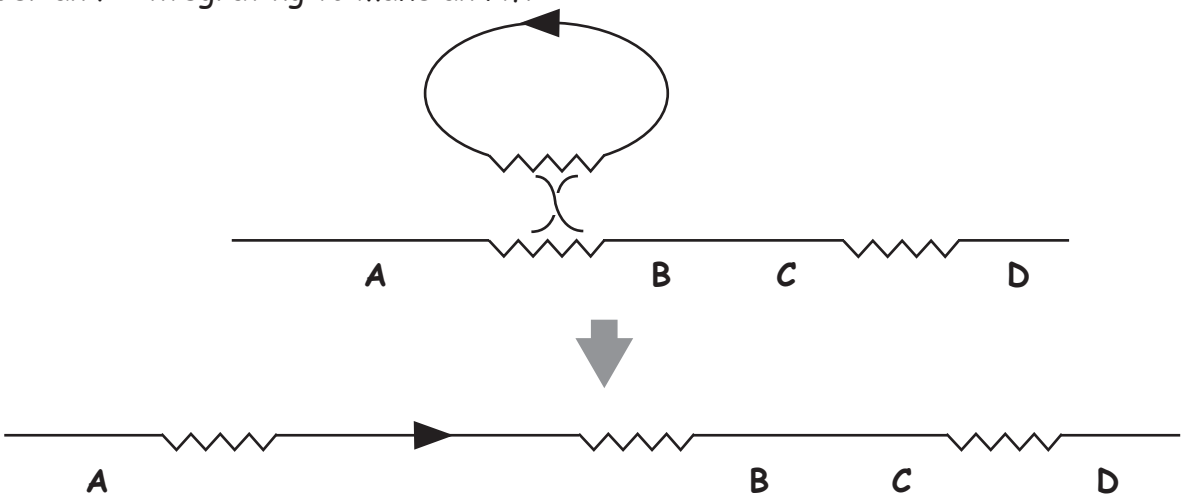
F+ plasmid: 1) Transfers itself at a frequency of 0.1

2) Does not transfer chromosomal markers

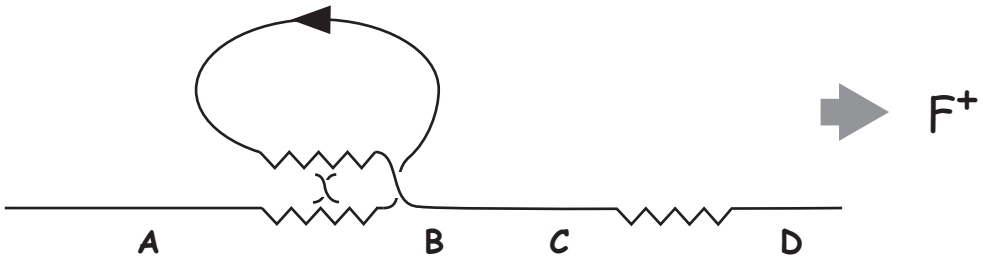
Hfr 1) Transfers some chromosomal markers efficiently

2) Other markers transferred inefficiently - Gradient of transfer
(It takes about 100 minutes to transfer the entire chromosome)

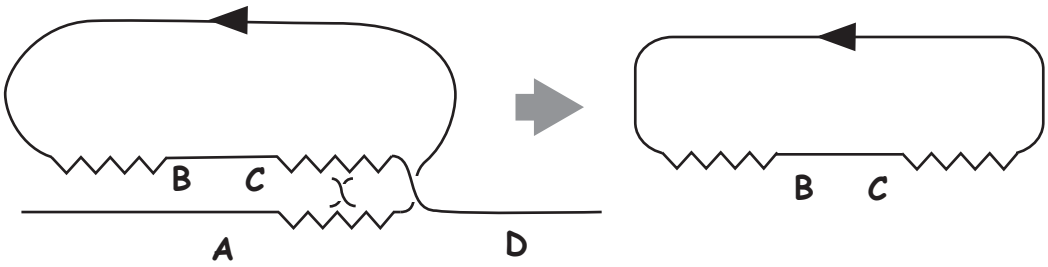
Consider an F^+ integrating to make an Hfr:



This process can be reversed to go back to the F^+ state:



The recombination can occur at a different position to give an F plasmid that carries a part of the chromosome. This form of F is called an F'.



F's are usually isolated by selection for early transfer of a marker that is transferred late in the Hfr. In the example above the F' could have been isolated from a population of Hfrs by selecting for early transfer of either B or C.

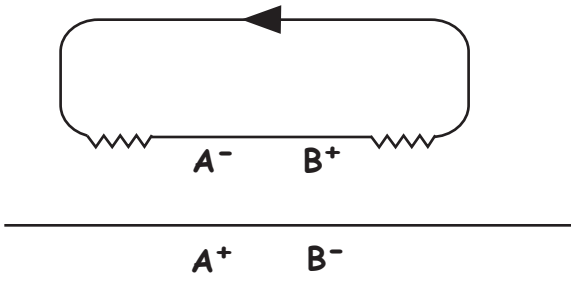
F' 1) Very efficient transfer of markers carried on F'.
These can be markers that were transferred very late in the Hfr from which the F' was derived.

2) No transfer of chromosomal markers not on F'.

F's can be used to perform genetic tests of function because a cell containing a F' will be diploid for the region of the chromosome carried on F. This is known as a merodiploid. For example, if we isolated a new Lac– mutation we could use an F' Lac^+ to determine whether the Lac^- mutation is dominant or recessive.

	Growth on lactose	
Lac^+	+	
Lac^-	–	
Lac^- / F' Lac^+	+	(Lac^- is recessive)

It is also possible to test for functional complementation of two linked mutations. Consider two mutations, A^- and B^-, that are close together and have the same phenotype. We can introduce an F' carrying A^- into a strain with a B^- mutation. If the merodiploid has a wild type phenotype then we know that the mutations complement and are therefore in different genes.



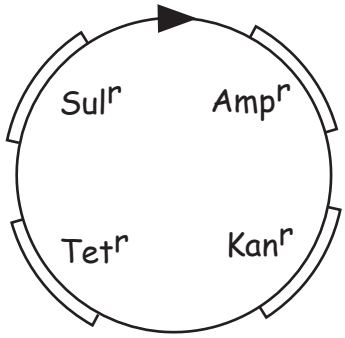
Lecture 16

Gene Cloning

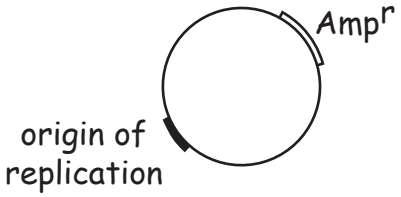
F is one of many bacterial plasmids, most of which are also transmissible from one cell to another.

R factors - This type of plasmid was discovered in Japan in early 1950's . They came from hospital patients that were infected with bacteria that were resistant to several different antibiotics. This was surprising since antibiotics work by very different mechanisms.

For example, resistance to ampicillin, kanamycin, tetracycline, and sulfonamide could be conferred at once on transfer of a given R factor. In fact, most of the antibiotic resistance genes are actually in transposons that are carried on the R factor.



Modern cloning vectors are stripped down versions of R factors. They usually carry one or two drug resistance genes and an origin of replication.



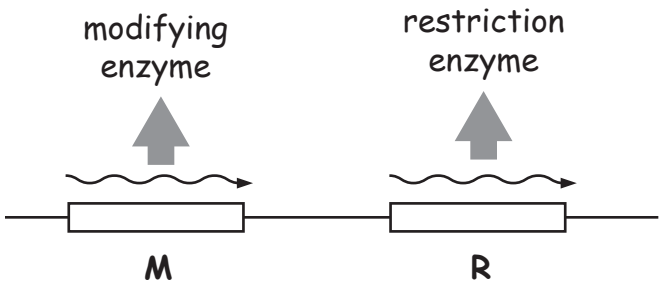
Cloning involves the use of enzymes in vitro to make plasmids carrying pieces of the chromosome. One of the important tools is a set of enzymes that can cleave DNA at specific sites. These enzymes are known as **Restriction Enzymes**. They were discovered in the following way:

	E. coli C	E. coli K
□ (grown on C)	10^8/ml	10^3/ml
□ (grown on K)	10^8/ml	10^8/ml

This phenomenon known as "host restriction" behaves like a genetic change that reverses at a high frequency. The explanation is that E. coli K makes enzyme that cleaves □ DNA. The K strain doesn't destroy its own chromosome because it also makes an enzyme that modifies the cleavage site.

The phage that grow on K have by rare chance escaped cleavage long enough to be modified.

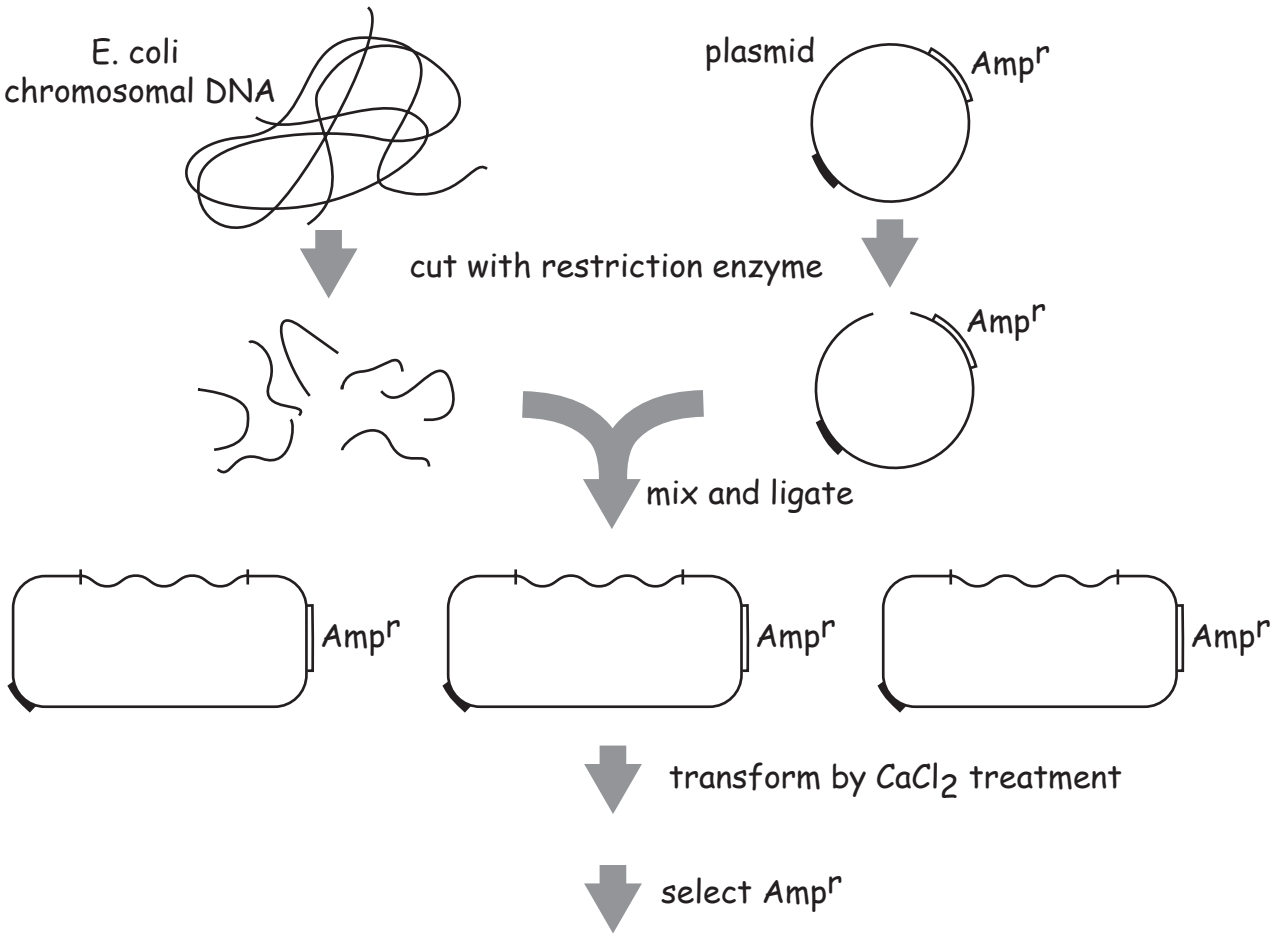
The genes for restriction enzymes usually come in pairs with the gene for the restriction enzyme (R) next to the gene for the enzyme that modifies the same sequence (M).



Mutants the have a mutated version of the restriction enzyme but a wild type version of the modifying enzyme (R^- M^+) are useful because they do not show host restriction but phage grown on these strains are resistant to host restriction. It is a useful exercise to think about why a strain with a mutated modifying enzyme but a wild type restriction enzyme (R^+ M^-) would be inviable.

A large number of these enzymes have been isolated from different bacterial species. Most of the enzymes recognize palindromic DNA sequences of 4 or 6 base pairs. Restriction enzymes can be used to cut chromosomal DNA into fragments. These fragments can be ligated into plasmid DNA that has been cut at a single site. This procedure takes advantage of the fact that the DNA ends that remain after cleavage with a restriction enzyme will base pair with other ends cut with the same enzyme. The collection of a large number of random chromosomal fragments carried in plasmids is known as a Library

Generation of a library yields a a very large collection of plasmids each with a different chromosomal insert.



Cloning by Complementation

Say we wanted to clone the Lac operon. First a library would be made from DNA from a Lac^+ E. coli strain. This library would then be used to transform a Lac^- strain. Transformants would first be selected by Ampr. The resistant colonies would then be screened for the ability to grow on lactose (Lac^+). These clones should contain plasmids carrying a functional Lac operon.

How many clones would we need to screen? Each plasmid carries about 5×10^3 bp of chromosomal DNA. The chromosome is 5×10^6 base pairs so the entire genome will be covered if several thousand clones are screened.

All sorts of genes from E. coli have been cloned by looking for DNA fragments that can restore function to a mutant. It is also possible to find genes from other bacteria. The following is a dramatic example of a cloning experiment to find an important protein for a pathogenic bacterium.

Yersinia is the bacillus that causes bubonic plague, a disease that killed 100 million people in the 6th century A.D.. One reason that Yersinia is such a deadly pathogen is that it escapes the immune system by multiplying within cells. The problem was to find the Yersinia genes that enable the bacterial cells to invade human cells. To do this an assay was needed.

A test for bacterial invasion consists of a layer of mammalian tissue culture cells. The bacteria are allowed to settle onto the cells for awhile, then the bacteria that have not entered the cells are killed with the antibiotic gentamicin, which can not cross the membrane of tissue culture cells. The bacteria that have entered cells escape gentamicin and can be recovered from the inside of the cells after the cells are lysed with detergent.

E. coli normally can not invade cells. The gene for invasion was found by transforming E. coli with a library of Yersinia DNA and then selecting for E. coli that had invaded cells. A single gene was found that encodes a surface protein known as invasin.