

Genetics 1: Bacteria

Model Organisms

These include bacteriophages T4 and λ . T4 has been studied extensively and many basic concepts in molecular biology were derived from it. Some of phage λ 's properties were discussed by Dr. Poteete. For these lectures we will focus mostly on *Escherichia coli*. *Salmonella typhimurium* is a close relative (15% divergent DNA sequence) and basic principles apply to both.

Brief Basic Anatomy of *E. coli* (Voet², pp. 3, 4.)

4.7 Mb of sequenced DNA in a condensed nucleoid without a nuclear membrane

E. coli is haploid

Cytoplasm filled with polysomes

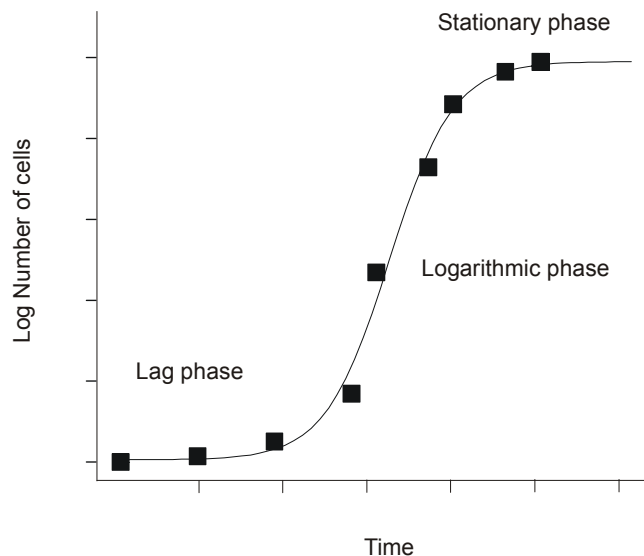
Transcription and translation are coupled

Inner and outer cell membranes (Gram negative)

Peptidoglycan cell wall

Flagella, pili

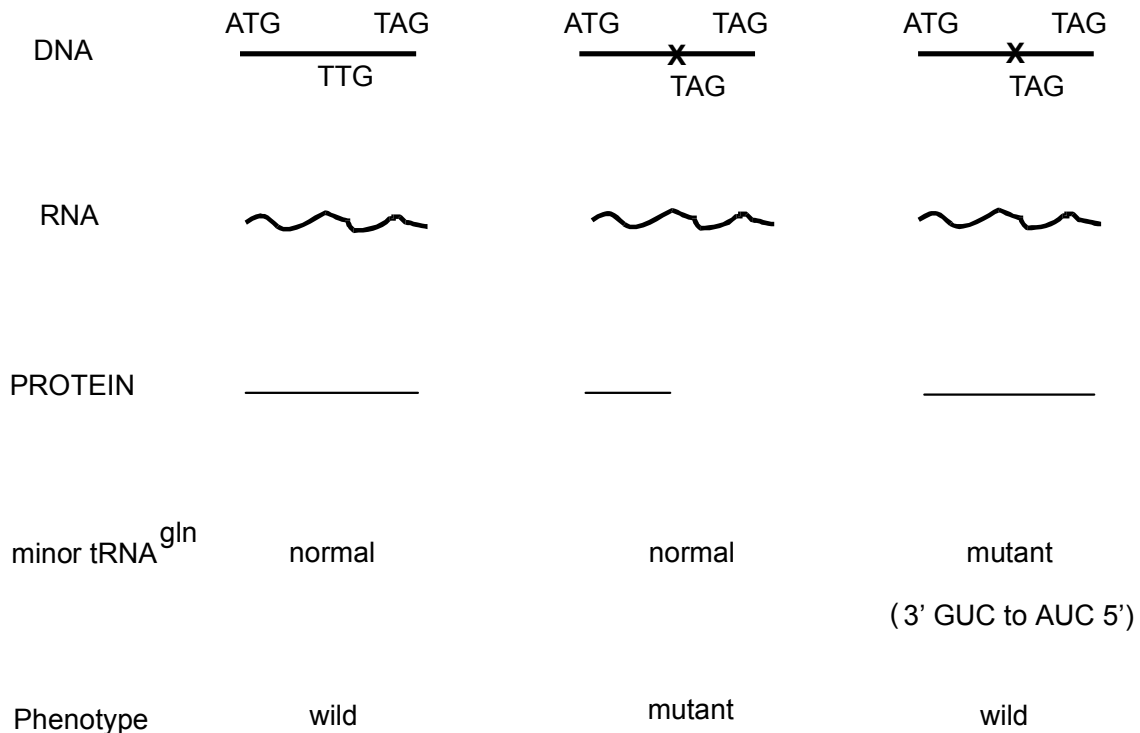
Growth of *E. coli* cultures



Cells are cultivated in either defined or rich media. Defined (or minimal) media contain a variety of salts and a buffer and usually glucose as the carbon source. Rich media are generally broths derived from meat products (e.g., L broth consists of tryptone, yeast extract and salt) with or without glucose. Regardless of the type of medium, a growth curve like that above is obtained. Experiments are generally carried out with logarithmic phase cells as these are a homogeneous population in contrast to lag or stationary phase cells.

Amber Mutations and Suppressors in E. coli

A typical ORF (open reading frame) in *E. coli* begins with an ATG and ends with a TAG. If a mutation occurs in a TTG codon in the ORF to TAG, an internal stop codon is generated. Consequently, a truncated peptide is made which we will assume cannot function catalytically. For example, if this ORF is the *lacZ* gene, then no active β -galactosidase is formed and the cell is phenotypically Lac⁻. We can isolate Lac⁺ derivatives from this mutant, which are either wildtype at the TAG codon (reverted to TTG) or are still mutant (TAG) and the latter are designated pseudorevertants. These pseudorevertants contain *amber* suppressor mutations (Su⁺) in genes coding for minor tRNAs which can now read TAG codons as “sense”. In the figure below, glutamine is inserted at the TAG. A variety of *amber* suppressors is available which allows for various amino acid replacements to be made in vivo.



In the figure above, the wildtype strain is Su⁻ and the strain with the mutated tRNA is Su⁺.

Conditional Lethal *Amber* Mutations in Bacteriophages

Wildtype T4 is treated with a mutagen and mutants are sought which form plaques on the Su^+ but not on the Su^- strain.

| | Plaques formed on <i>E. coli</i> strain | |
|---------------------|---|--------|
| | Su^+ | Su^- |
| T4 wildtype | + | + |
| T4 <i>am</i> mutant | + | - |

The above result indicates that the *amber* mutation is in an essential gene of T4. Therefore, by isolating a large number of *amber* mutants, it was possible to identify almost all T4 essential genes. In this case, the *amber* mutations are conditional lethal mutations.

Nonsense mutations at TAA (*ochre*) and TGA (*opal*) codons can also be generated.

Be aware that nonsense mutations in non-essential genes can occur and are not conditionally lethal. Such mutations are very useful to ensure that the reading frame of a particular ORF is correct and to produce stable fragments of a protein.

Isolation of *Escherichia coli* Mutants

I. Auxotrophic mutations

E. coli is able to grow in a simple medium of glucose plus salts ("minimal medium"). This property allows for the simple selection of auxotrophs, i.e., mutant cells which can't grow on this medium. For example, wildtype cells are mutagenized, enriched for auxotrophic mutants, and plated on minimal medium plus leucine. Replica plating is used to identify colonies able to grow on minimal plates supplemented with leucine but unable to grow on minimal medium.

| | Growth on minimal medium | |
|-------------------------|--------------------------|-----------------|
| | with leucine | without leucine |
| <i>E. coli</i> wildtype | + | + |
| <i>E. coli</i> Leu^- | + | - |

Isolation of hundreds of Leu^- mutants led to the eventual identification of six biosynthetic genes.

This type of genetic analysis was used for other amino acids, purines, pyrimidines, and vitamins. In short, any condition leading to auxotrophy can be analyzed genetically.

a. Sugar fermentation

In this case, the glucose is not present in minimal medium and is replaced by some other sugar, lactose, for example. In contrast to auxotrophy, however, Lac⁻ cells cannot grow on this medium but will grow on minimal medium with glucose. Lac⁺ cells on glycerol minimal medium containing the non-metabolizable lactose analogue, IPTG (isopropylthiogalactopyranoside), and X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside), form blue colonies whereas Lac⁻ bacteria produce white colonies.

Other indicator media are also used. For example, on McConkey medium Lac⁺ colonies are red and Lac⁻ are white. McConkey is a rich undefined medium and not a minimal medium.

b. Antibiotic resistance

The most commonly used include resistance to streptomycin (Str), rifampicin (Rif), and nalidixic acid (Nal). (Don't confuse these with plasmid encoded drug-resistance)

c. Phage resistance

d. Markers for DNA repair and recombination (sensitivity to ultraviolet light and other DNA damaging agents)

II. Transposon-induced mutations

The most common are Tn5 (kanamycin), Tn9 (chloramphenicol), and Tn10 (tetracycline resistance). Derivatives of Tn10 are available that are smaller (miniTn10), have different drug-resistances (e.g., miniTn10Kan), and, once inserted into the chromosome, can no longer transpose. The transposons are usually delivered into target cells on suicide vectors. After transposition has occurred the cells are pooled and then either screened or subjected to selection for a specific trait.

III. Temperature conditional lethal mutations

These come in two flavors- Ts (temperature sensitive) and Cs (cold-sensitive). For example, *dnaB*(Ts) mutants can grow at 30 C but not at 42 C. Cs mutants grow at 30 C but not 15 C.

Complementation

Once a large number of mutants have been collected, it's necessary to divide them into functional groups by complementation. The basic assumption is that mutations in the same gene (e.g., *lacY*) are not expected to complement each other but should complement mutations located in, for example, the *lacZ* gene. Complementation tests are often carried out in recombination-deficient cells. This test also requires that the cells are diploid for the region under study and single copy plasmids (such as the F (fertility) factor), or low copy plasmids derived from pSC101, are usually employed.

We will use the lactose system as an example. Remember that the enzymes that metabolize lactose are inducible in the wildtype to allow growth on lactose (Lac⁺).

| Genotype | Lac phenotype | Genotype | Lac phenotype |
|-------------------------------|---------------|--------------------|---------------|
| <i>lac-1/lac</i> ⁺ | + | <i>lac-1/lac-1</i> | - |
| <i>lac-2/lac</i> ⁺ | + | <i>lac-2/lac-2</i> | - |
| <i>lac-3/lac</i> ⁺ | + | <i>lac-3/lac-3</i> | - |
| <i>lac-1/lac-2</i> | + | | |
| <i>lac-1/lac-3</i> | + | | |
| <i>lac-2/lac-3</i> | + | | |

The Lac⁻ mutants fall into three groups by complementation corresponding to *lacZ*, *lacY* and *lacA* genes which encode β-galactosidase, permease and trans-acetylase respectively. All are recessive. This analysis was carried out on several hundred mutants. These experiments identified the structural genes.

Mutant strains that were constitutive were also obtained after mutagenesis. These were proposed to be regulatory genes. (What is their colony color?)

| Genotype | Lac phenotype | Genotype | Lac phenotype |
|--------------------|---------------|-------------------------------|-----------------|
| <i>lac-4/lac-4</i> | Constitutive | <i>lac-4/lac</i> ⁺ | Inducible (Ind) |
| <i>lac-5/lac-5</i> | Constitutive | <i>lac-5/lac</i> ⁺ | Constitutive |

The results indicate that the mutations are in different genes since one is dominant, and one is recessive.

Mutant strains were also isolated that do not induce β-galactosidase in the presence of IPTG.

| | | | |
|--------------------|--------------------------------------|-------------------------------|--------------------------------------|
| <i>lac-6/lac-6</i> | Lac ⁻ (Ind ⁻) | <i>lac-6/lac</i> ⁺ | Lac ⁻ (Ind ⁻) |
|--------------------|--------------------------------------|-------------------------------|--------------------------------------|

The results for *lac-6* indicate a dominant Lac phenotype.

The next step is to test the mutant alleles in *cis* and *trans*.

| | | | |
|---|--------------------------------------|---|--------------------------------------|
| $\frac{lac-4^+ lacZ^- lacY^-}{lac-4 lacZ^+ lacY^+}$ | Inducible | $\frac{lac-4^+ lacZ^+ lacY^+}{lac-4 lacZ^- lacY^-}$ | Inducible |
| $\frac{lac-5^+ lacZ^- lacY^-}{lac-5 lacZ^+ lacY^+}$ | Constitutive | $\frac{lac-5^+ lacZ^+ lacY^+}{lac-5 lacZ^- lacY^-}$ | Inducible |
| $\frac{lac-6^+ lacZ^- lacY^-}{lac-6 lacZ^+ lacY^+}$ | Lac ⁻ (Ind ⁻) | $\frac{lac-6^+ lacZ^+ lacY^+}{lac-6 lacZ^- lacY^-}$ | Lac ⁻ (Ind ⁻) |

Again, they are different because *lac-4* is recessive in both *cis* and *trans*; *lac-5* is *cis*-dominant; and *lac-6* is dominant in both *cis* and *trans* configurations.

| | | | |
|-------------------------------------|--------------|---------------------------------------|--------------|
| $\frac{lac-4 lac-5^+}{lac-6 lac-5}$ | Constitutive | $\frac{lac-4^+ lac-5^+}{lac-6 lac-5}$ | Constitutive |
|-------------------------------------|--------------|---------------------------------------|--------------|

Lac-5 is dominant even when *lac-6* is present.

What are the likely identities of *lac-4*, *lac-5* and *lac-6*?

An additional allele deserves mention. This showed a *trans*-dominant phenotype and could not be assigned to known structural genes. Further genetic and biochemical evidence, however, indicated that the mutation was in *lacI*. The *lac* repressor is a tetramer and in the wild type binds to the operator. The dominant allele (*lacI^d*) produced a repressor unable to bind to DNA. In a *lacI^d/lac⁺* diploid, mixed multimers are formed that are unable to bind DNA thus producing a dominant-negative effect. Most dominant-negative alleles work best in multicopy. This type of allele is widely used in eukaryotic systems to inactivate gene products.

Complementation tests using multicopy plasmids can lead to unexpected effects. What will be the phenotype of a wildtype *E. coli* carrying a multicopy *lacO⁺* gene?

Intragenic Complementation

A rare instance of intragenic complementation deserves mention. Mutations affecting the N-terminal 250 residues of β -galactosidase (Lac⁻) complement mutations in the rest of the gene. This is the basis of the widely used α -complementation test for gene cloning using the pUC (and other) plasmids. The bacterial host strains have the *lac* region on the chromosome deleted [$\Delta(lac-pro)$]. An F-prime episome is present that has a deletion ($\Delta M15$) in the operator-proximal part of the *lacZ* gene (Lac⁻) but the rest of the *lac* operon is present. The pUC plasmids encode the N-terminal part of the *lacZ* gene. Introduction of a pUC plasmid into the $\Delta(lac-pro)/F-lacZ\Delta M15$ strain results in a Lac⁺ phenotype because of α -complementation. Introduction of DNA fragments into the multiple cloning site of pUC leads to loss of the α -complementing peptide fragment (Lac⁻).

Mechanisms of Genetic Transfer

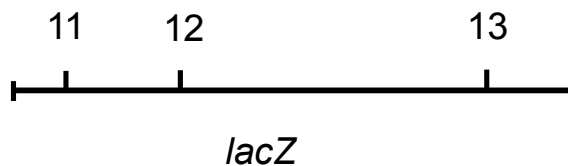
There are three general mechanisms for genetic transfer in bacteria: conjugation, transduction and transformation. All three modes of transfer lead to the formation of partial diploids where the recipient cell contains only a fragment of donor DNA. For homologous recombination to occur an even number of cross-overs is required.

The F (fertility) factor is a single-copy plasmid (about 100 kb) capable of self-transmission which, like phage lambda, can exist either autonomously in the cytoplasm or integrated into the host chromosome at many locations. F is the source of BAC (bacterial artificial chromosome) vectors. Bacteriophage P1 is also a plasmid that cannot integrate into the chromosome but can be used for transducing genes. It has a site specific-recombination system (*lox-cre*) that is widely used in eukaryotic molecular biology. It is also the source for PAC (phage artificial chromosome) vectors.

Genetic Mapping

I. Two-factor crosses

There will be more cross-overs between two alleles far apart than if they are close together. In the example below,

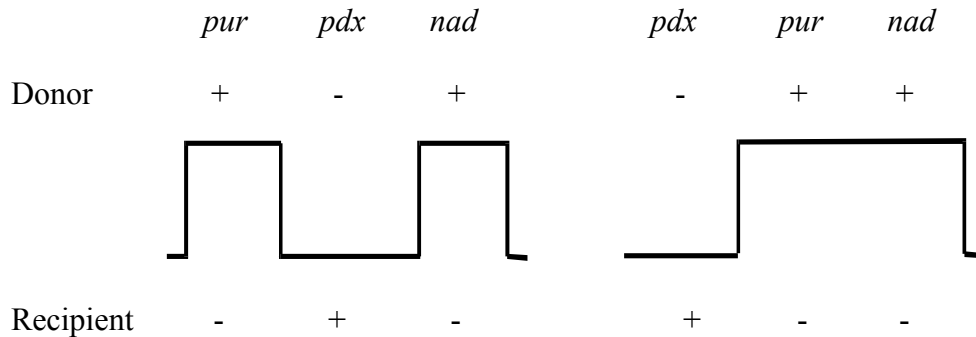


we would expect a higher frequency of Lac⁺ recombinants in crosses between *lacZ11* and *lacZ13* than between *lacZ11* and *lacZ12*. Similarly, the Lac⁺ frequency will be higher in the *lacZ12* x *lacZ13* cross than the *lacZ12* x *lacZ11* cross. If this is the case then the linear order *lacZ11-lacZ12-lacZ13* can be inferred.

Two-factor crosses are also used to map the order of genes. In practice, these crosses are useful but are limited by positive or negative interference and by variability in day-to-day experiments.

II. Three-factor crosses

These give an unambiguous order for alleles. For example, by two factor crosses (transduction) *pur* and *nad* are 26% linked and *pur* and *pdx*, 46%. This result leads to two possible arrangements: *pur-pdx-nad* or *pdx-pur-nad*. A cross using a *pur*⁺ *nad*⁺ *pdx*⁻ donor and a *pur*⁻ *nad*⁻ *pdx*⁺ recipient showed that the wildtype recombinant class was the least frequent (6%) while the frequency of the other recombinant classes was between 20-48%. This result indicates that the gene order is *pur-pdx-nad*.

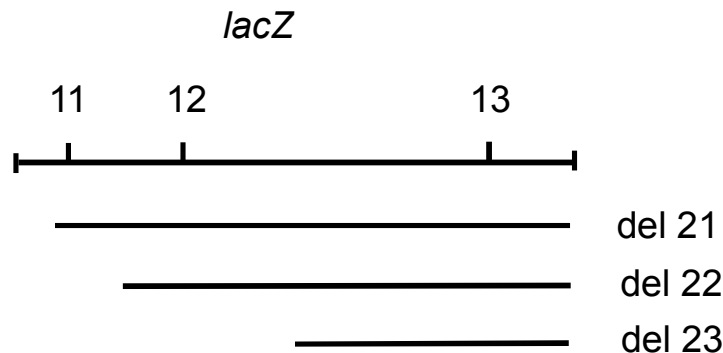


III. Deletion mapping

This is the most unequivocal method if the end points of the deletions are known.

| | Lac ⁺ recombinants | | |
|--------------------|-------------------------------|---------------|---------------|
| | <i>lacZ11</i> | <i>lacZ12</i> | <i>lacZ13</i> |
| <i>lacZ</i> del 21 | - | - | - |
| <i>lacZ</i> del 22 | + | - | - |
| <i>lacZ</i> del 23 | + | + | - |

The order is as shown.



Physical Maps

Restriction map with 6-base cutters

Restriction map with rare cutters

Restriction maps of entire bacterial chromosomes can be produced using pulse field gel electrophoresis. Shotgun sequencing is now the method of choice to make genomic maps of microorganisms. From the sequence, BLAST searches are used to identify proteins homologous to those in *E. coli* and other organisms. Open reading frames that show homology are presumed to be encoded by "housekeeping" genes, i.e., for DNA, RNA and protein synthesis as well as DNA recombination, repair, metabolism, motility, cell division, etc. Proteins that show no homologs may be specific for that bacterial species or genus.

Sequenced Genomes and Genomics

Sequenced genomes of bacteria can be found at

<http://www.ncbi.nlm.nih.gov/genomes/MICROBES/Complete.html>

At present, 145 microbial genomes have been completely sequenced. A big surprise is the extent of horizontal gene transfer among distantly related bacteria.

Proteomics

Proteins extracted from *E. coli* can be separated by 2D gel electrophoresis. Examination of mutant bacteria or those overproducing specific proteins can lead to their identification. See

<http://expasy.cbr.nrc.ca/ch2d/publi/ecoli.html>

Other techniques still being developed include identifying interacting partner proteins or proteins with a defined characteristic (e.g., those that can be phosphorylated or bind ATP).

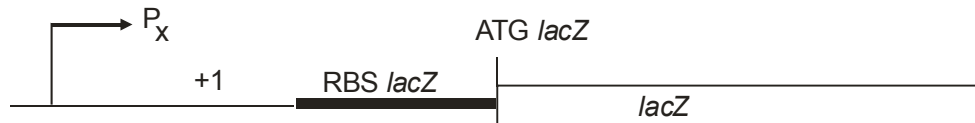
Gene Function and Regulation

- a. Deletion of gene
- b. Transcription and translational reporter fusions

In a translational fusion, the promoter, RBS and translation initiation codon (or more) of your favorite gene are fused to the eighth codon of *lacZ*. In a transcriptional fusion, the promoter is fused to the RBS and initiation codon of *lacZ*.

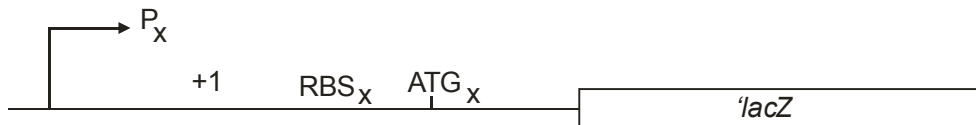
Other reporter genes can also be used (e.g., *lux*, *cat*, *gfp*, etc.). Green fluorescence protein (GFP) fusions are useful to locate proteins in the cell.

Transcriptional Fusion



Product - LacZ

Translational Fusion



Product - Protein X-'LacZ fusion

- c. Selection for exported proteins

PhoA is the gene that encodes alkaline phosphatase, an exported protein. Insertion of Tn_{phoA} into a gene for an exported protein leads to blue colonies on the appropriate medium.

- d. Oligonucleotide Arrays

Oligonucleotide arrays can be used to analyze transcriptional patterns of large numbers of genes in order to identify their function. This is essentially a northern analysis on a large scale where RNA extracted from cells is hybridized to genomic DNA oligonucleotide arrays. Changes in transcriptional patterns can be monitored under various conditions. Array data may, or may not, agree with gene fusion results. Post-transcriptional regulatory events are not monitored by arrays.

Reference

Snyder, L. and Champness, W. "Molecular Genetics of Bacteria", 2nd edition, 2003, ASM Press, Washington DC