

Prokaryotic genomes and genetics

I. Prokaryotic Genomes

A. Genome type

1. Circular chromosomes

Prokaryotic DNA is composed of double-stranded anti-parallel DNA strand. Most prokaryotes have circular chromosomes. The genome is supercoiled with one negative turn every 200 bp-twists, which aids in compaction of the DNA inside the cell.

2. Linear chromosomes

Some prokaryotes have linear chromosomes. Inverted repeats at the end of the linear chromosome function like telomeres to ensure fidelity of DNA replication to the end of the chromosome. An example of a species with a linear chromosome is *Streptomyces*.

3. Nucleoid architecture

Bacterial chromosomes have (at least) four histone-like proteins that are involved in maintaining the architecture and compactions of the chromosome. HU, IHF, FIS and H-NS. The bacterial chromosome and associated proteins is called the nucleoid.

B. Genome structure and regulatory aspects

1. Transcription and translation are "coupled" in prokaryotes

Because the nucleoid does not have a membrane surrounding it, as soon as mRNA is transcribed from the DNA, ribosomes attach and begin translating the mRNA into protein. Thus, transcription and translation are "coupled" in prokaryotes.

2. Operons

Many transcriptional units encode more than one gene, which is termed an OPERON. Genes with related functions are often located together in an operon. An operon is a group of genes that has a single promoter site (site where RNA polymerase binds and transcribes mRNA) and is transcribed as a single polycistronic mRNA molecule. Translation of genes within the polycistronic mRNA begins at the first AUG and proceeds to transcription termination signals. Ribosomes may or may not detach from the mRNA between cistrons. A basic operon is composed of a promoter, transcriptional start point, the leader sequence, ribosome binding site and transcription termination signal.

A comparison of operon structure between completed bacterial genome sequences has been very useful for functional predictions.

3. Regulon

A regulon is a set of operons that may be scattered across the chromosome but that are all controlled by the same regulatory protein. These operons encode proteins associated with a cellular function or pathway.

4. Modulon/Stimulon

The regulation of a seemingly unrelated number of operons that affect global responses to stress, nutritional parameters and heat shock. Regulatory proteins that simultaneously regulate the expression of a number of operons and regulons that results in global regulation and the change of state within the cell. Mutations in global regulators cause pleiotropic effects.

Modulons can be put into four categories:

1. nutrient and energy supply modulon
2. stress response modulon
3. cellular differentiation modulon
4. aggregation/cell-cell interactions modulon

It is thought that approximately 50 global regulatory systems exist in a typical prokaryotic cell.

C. Plasmids

1. Types of plasmids

Sizes of plasmids can range from a few kilobase pairs to hundreds of kbp and can be linear or circular. Broad host range plasmids can replicate in diverse bacteria (encode own replication machinery). Narrow host range plasmids can replicate only in related bacteria or single species (require host factors for replication).

a. small circular plasmids that have a high copy number in the cell. These plasmids do not promote their own transfer into other cells. High copy number ensures that each daughter cell upon division will receive plasmids.

b. Large plasmids that have low copy number in the cell and promote their own transfer to other cells via conjugation.

2. Plasmid stability and copy number

Dependent upon:

a. Plasmid integrity: presence of repeated sequences in plasmids can result in loss or gain of sequences because of recombination events.

b. Partitioning and systems: Mandatory for low copy number plasmids. Partition systems have proteins (trans-acting) that bind to specific sites (cis-acting) on the plasmid. Proteins are believed to mediate attachment of the plasmid to specific sites in the cell membrane which promote partitioning.

3. Traits commonly found on plasmids

a. Antibiotic resistance

b. Colicins and bacteriocins: antimicrobial compounds

- c. Virulence determinants-toxins carried by integrated bacteriophage "pathogenicity islands" Table 5.1
- d. Virulence determinants for plant pathogens
- e. Metabolic activities, utilization of carbon sources, sucrose, lactose, urea, etc. degradation of toxic compounds, toluene, aromatic compounds (2,4-D).
- f. Nitrogen fixation.
- g. Some plasmids do not confer any obvious phenotype to their host and are therefore termed "cryptic" plasmids.

D. Transposons and insertion sequences

Types of movable elements

a. Insertion sequences (IS) elements: Ubiquitous and identified in most bacterial genera, although number and type vary between species and strain to strain.

Structure: IS elements range in size from 750-1550 base pairs. Central region of IS element encodes a transposase which is an enzyme that is required for movement of the IS element. At the ends of the IS elements are inverted repeats (IR), which can vary in size between IS elements

b. Transposons

Most important distinction between IS elements and transposons is that transposons carry a marker that is detectable (i.e. antibiotic resistance gene).

1. **Class I transposons** have IS elements at their ends. Class I transposons use a non-replicative mechanism for transposition. Transposon completely excised from donor site and inserted into target site. Cut and paste type of replication. Examples are Tn5/TN10.
2. **Class II transposons** have inverted sequences at their ends, which are not IS elements. Class II transposons move by replicative transposition: A copy of the transposon is integrated at a new site.

II. Genetic analysis

A. Conjugal transfer

The unidirectional transfer of either plasmid or chromosomal DNA from one cell to another by cell-cell contact. DNA transfer is not reciprocal, therefore the two cells that participate are referred to as "donor" and "recipient".

Mechanism of conjugation.

1. Conjugation is dependent upon the presence of a plasmid (integrated or not) or a transposon that carries the genetic information (*tra* genes) required for transfer.
2. Integration of a plasmid or transposon into the bacterial chromosome allows transfer of chromosomally located bacterial genes.

3. Donor cell carries gene encoding the pili, which make contact with receptors on the surface of the recipient cell. The pili contract, thus bringing the two cells together.
4. Transfer of DNA is initiated by a protein that makes a single strand nick in donor DNA at the origin of transfer (*oriT*). The *oriT* is characterized by containing repeated DNA elements believed to be important in DNA binding.
5. A helicase (*traI*) unwinds the DNA and ss DNA is transferred to the recipient cell beginning at the 5' end. DNA is replaced in the donor cell by replication machinery by a rolling circle type of mechanism. DNA replication in recipient cell replicates the plasmid (if only plasmid transferred).
6. This process can also mobilize plasmids that do not encode the transfer machinery, but contain an *oriT* site.
7. In chromosome transfer, the plasmid has integrated at specific sites of the host bacterial chromosome. Nicking at *oriT* initiates transfer. Chromosomal DNA must recombine with host chromosome to be maintained.

Conjugation in gram+ bacteria

1. Conjugal transfer also occurs in gram+ bacteria, but does not require a pilus. Induction of conjugation requires secretion of sex pheromones that results in cell aggregation and induction of conjugative transfer machinery.

Three factor crosses

1. Three factor crosses are used to define gene order.
2. Cis-trans test: defines genes by complementation. Independent genetic units are called cistrons and includes regulatory sequences. Regulatory sequences are cis-acting elements; they exert their effects only on cistrons that are located next to them. Trans elements act on regions separate from themselves. Encode diffusible factors (proteins or RNA) that can act at a distance.
3. Allelism test: Complementation requires formation of a partial diploid
4. Cross-feeding: accumulate a metabolite which is released into the medium. Categorize mutants into groups and arrange them in order of a biochemical pathway.

B. Transduction: Genetics of bacteriophages

1. Types of phages

Lytic. Enter bacterial cell and cause lysis

Lysogenic. Enter cell and either cause lysis or are integrated into the bacterial chromosome. Phage genes are repressed until activated, usually by stress. Cell are resistant to further infection.

2. Generalized transduction

Fragments of bacterial chromosome are incorporated into phage heads and are transduced into new host cells. Not all bacteriophages are capable of transduction. The host cell DNA must be degraded to an appropriate size to be packed in the phage head and the specificity of the packaging process must be low. The transduced DNA, if it is a plasmid, can replicate autonomously in the host, but if it is host DNA, it must recombine with the new host chromosome to be maintained.

3. Specialized transduction

When lysogeny breaks down and the phage enters the lytic cycle, the excision of the phage from the genome is sometimes imprecise, resulting in the incorporation of flanking host DNA into the phage head. These flanking DNA markers will be transduced at a very high frequency.

C. Transformation

1. Natural competence

Occurs, for example in *Bacillus subtilis*, *Haemophilus influenzae*, *Streptococcus pneumoniae* *Neisseria gonorrhoeae*. Likely to occur at a low level in most bacterial species. Mechanism of transformation varies between species.

a. Competence usually occurs at a specific stage of growth, typically late log phase, just before cells enter stationary phase. Competence is associated with nutrient depletion and accumulation of secreted competence factors, which act via two-component regulatory systems.

b. DNA is actively taken up from outside the cell. In *B. subtilis* and *S. pneumoniae*, dsDNA binds to receptors on the cell surface. DNA is fragmented by double-stranded cleavage. In *B. subtilis* and *S. pneumoniae*, DNA uptake is not species specific. In *H. influenzae*, DNA uptake is species-specific and is dependent upon an 11bp sequence. Vesicle-like membrane extrusions called transformosomes actively take up DNA and internalize it into the cell.

c. The incoming DNA must recombine with host chromosome for maintenance and replication. Recombination into the host chromosome is dependent upon RecA, which is involved in homologous recombination.

2. Artificially induced competence.

a. In many bacteria, naturally-occurring competence and DNA uptake do not occur or occur at very low levels. An increase in "competence" can be accomplished by treating bacterial cells with chemicals or electrical pulses that results in membrane and cell wall loosening that facilitates dsDNA uptake. Artificially transformed cells take up circular dsDNA.

IV. GENOMICS

The full genomic sequence of approximately 254 bacterial and archeal species have been determined (~650 partial or in progress). Comparative genomic analysis has become a powerful tool to decipher the function of genes encoded by various genomes. Sequencing technology has progressed to the point that the DNA sequence of an entire bacterial genome can be determined in just a few days. The functional annotation of such a genome, however, will take much longer. Most initial annotation is "in silico" and is accomplished by determining protein sequence similarities to proteins in other organisms.

A. List of genomes

1. **List or prokaryotic genomes** that have been or are currently being sequenced. Most of these genome sequences are available in publicly accessible databases (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>).

B. Functional annotation of microbial genomes

30-40% of ORFs in various genomes are of unknown function. Of these, approximately half show similarities to ORFs in other organisms. Approximately 20-25% with no match in the genome database.

1. Definition of terms

- A. **Homolog**: Genes that are descended from a common ancestor. All b-Globins NOTE: AMINO ACID SEQUENCE SIMILARITY DOES NOT ENSURE IDENTICAL FUNCTION OR HOMOLOGY.
- B. **Orthologs**: genes in different species that evolved from a common ancestor. Orthologs are predicted to retain a common function throughout evolution. Diverged from each other after speciation events (Human and Chimp B-globin. A complete list of orthologs is a prerequisite for any meaningful comparison of genome organization.
- C. **Paralogs**: genes created by gene duplication within a genome and evolve new functions. Different globins. At close phylogenetic distances, similarity is a good approximation of orthologous function. At large phylogenetic distances, may confuse and blur orthologous genes because of paralogous genes and gene loss.

D. **Xenolog:** Homologous genes that have diverged from each other after a horizontal gene transfer event (antibiotic resistance genes in bacteria).