

Discuss the diversity of microbial genomes with special reference to genome size and their comparative analysis.

By year 2010, the sequencing of more than 1,600 bacterial genomes has been completed. Sequencing data indicate that bacterial genome sizes range from 500 to 10,000 thousand base pairs.

Average bacterial genome is only about 0.1% in size compared to the human genome, and have about 10% human gene content. One million base pair length contains on average about 10 genes in human, and about 500 to 10000 genes in bacteria. Bacterial genomes feature very small number of introns in bacterial genes, lesser average size of bacterial genes, and close proximity of genes to each other throughout bacterial genomes. Most bacterial genomes are single circular molecule and can also have additional plasmids, which are smaller and contain non-essential genes.

Bacterial genomes attain high mutagenic plasticity which is required for surviving in changing harsh environments. Many of the dynamic genomic events are similar across all the types of microorganisms. Some of the smallest, most stable, most deteriorated, most highly repeated and most highly recombined genomic sequences were described among the bacterial species. Massive gene amplifications (gene gain), deletions (gene loss), and recombination (genome wide rearrangement) events are shaping the bacterial genome. Acquisition of new traits may involve the acquiring of novel genes by horizontal gene transfer (HGT) or functional plasticity of existing genes. Gene gain and gene duplication events are equilibrated by gene loss on the scale of whole genome, and even with the on-going changes in the genome, its size remains approximately the same.

A single bacterial species can have many strains whose genetic composition differs. The differences prompted the introduction of two concepts describing the bacterial genome – core genome and pan-genome. Core genome represents the set of genes found across all strains of a pathogen. A pan-genome is the entire gene pool for the species, and includes genes that are present only in some strains. The pan-genome can contain dispensable or strain specific genes. Dispensable genes are those found in more than one strain of pathogen, but not in all strains. Strain specific genes are found only in a single strain.

Gene Loss and Genome Decay

Under conditions when a particular gene is no longer used (for example, as a result of moving to a different ecological niche and thus changing the lifestyle and metabolism) the loss of gene or genome decay may occur. Large number of pseudo genes was identified in some bacterial pathogen species. For example, leprosy causing pathogen [*Mycobacterium leprae*](#) contains nearly as many pseudo genes as functional genes. The whole-genome sequencing in combination with [comparative genomics](#) allows to detect the pseudogenes and the marks of genome decay.

Gene Gain and Gene Duplication

Horizontal (lateral) gene transfer (HGT) is considered to be one of the key forces driving gene gain. HGT refers to the gene transfer between organisms via a mechanism different from reproduction. Horizontal gene transfer may have been the cause for pathogen variations between [*Staphylococcus epidermidis*](#) and [*Staphylococcus aureus*](#). Comparative genetics helps to validate the presence and effect of this process.

There are several mechanisms behind the horizontal gene transfer. Transformation is an uptake and expression of genetic material from another species. In the course of another HGT process, transduction, the genetic material is transferred from one bacterial species to another with the help of viruses, such as bacteriophages. Bacteriophages are one of the most effective ways to transmit DNA between species and have great potential for genome diversification and the creation of new host interaction systems. For example, genes located in some prophages of secondary symbiotic bacteria may be involved in controlling the reproduction of host and defending it against eukaryotic parasites.

Genes can also be transferred via bacterial conjugation, the cell to cell contact resulting in exchange of genetic material.

Genome wide Rearrangement

Mobile genetic [insertion sequences](#) can play a role in genome rearrangement activities. Pathogens have been found to contain a large number of insertions and various repetitive segments of DNA. The combination of these two genetic elements is mediated by [homologous recombination](#). A number of pathogens, such as *Burkholderia mallei* and *Burkholderia pseudomallei*, have genome-wide rearrangements due to [insertion sequences](#) and repetitive DNA segments. Genome-wide rearrangements do contribute to the plasticity of bacterial genome.

Single Nucleotide Polymorphism

[Single nucleotide polymorphisms](#) (SNPs) are also a genomic variable that adds to the diversity of pathogen strains. Current efforts attempt to catalogue various SNPs in pathogen strains.