8.5C: Phylogenetic Analysis

The molecular approach to microbial phylogenetic analysis revolutionized our thinking about evolution in the microbial world.

Learning Objectives

- Outline the approaches to perform phylogenetic analysis

Key Points

- The purpose of phylogenetic analysis is to understand the past evolutionary path of organisms. Due to technological innovation in modern molecular biology and the rapid advancement in computational science, accurate inference of the phylogeny of a gene or organism seems possible in the near future.

- The developing technology of nucleic acid sequencing, together with the recognition that sequences of building blocks in informational macromolecules can be used as ‘molecular clocks’ that contain historical information, led to the development of the three-domain model (Archaea – Bacteria – Eucaryota).

- As more genome sequences become available, scientists have found that determining these relationships is complicated by the prevalence of lateral gene transfer among archaea and bacteria.

- Even using improved DNA-based identification methods, the total number of bacterial species is not known and cannot even be estimated with any certainty. Currently, there are a little less than 9,300 known species of prokaryotes.

Key Terms

- **Lateral gene transfer**: Horizontal gene transfer (HGT), also lateral gene transfer (LGT) or transposition refers to the
transfer of genetic material between organisms other than vertical gene transfer. Vertical transfer occurs when there is gene exchange from the parental generation to the offspring. LGT is then a mechanism of gene exchange that happens independently of reproduction.

- **microbial phylogenetics**: The study of the evolutionary relatedness among various groups of microorganisms.

Microbial phylogenetics is the study of the evolutionary relatedness among various groups of microorganisms. The molecular approach to microbial phylogenetic analysis revolutionized our thinking about evolution in the microbial world. The purpose of phylogenetic analysis is to understand the past evolutionary path of organisms. Even though we will never know for certain the true phylogeny of any organism, phylogenetic analysis provides best assumptions, thereby providing a framework for various disciplines in microbiology. Due to the technological innovation of modern molecular biology and the rapid advancement in computational science, accurate inference of the phylogeny of a gene or organism seems possible in the near future.

Gene sequences can be used to reconstruct the bacterial phylogeny. These studies indicate that bacteria diverged first from the archaeal/eukaryotic lineage. The term "bacteria" was traditionally applied to all microscopic, single-cell prokaryotes. However, molecular systematics showed prokaryotic life to consist of two separate domains, originally called Eubacteria and Archaebacteria, but now called Bacteria and Archaea that evolved independently from an ancient common ancestor. The archaea and eukaryotes are more closely related to each other than to the bacteria. Due to the relatively recent introduction of molecular systematics and a rapid increase in the number of genome sequences that are available, bacterial classification remains a changing and expanding field. For example, a few biologists argue that the Archaea and Eukaryotes evolved from Gram-positive bacteria.

While morphological or metabolic differences allowed the identification and classification of bacterial strains, it was unclear whether these differences represented variation between distinct species or between strains of the same species. This uncertainty was due to the lack of distinctive structures in most bacteria, as well as lateral gene transfer between unrelated species. The developing technology of nucleic acid sequencing, together with the recognition that sequences of building blocks in informational macromolecules can be used as ‘molecular clocks’ that contain historical information, led to the development of the three-domain model (Archaea – Bacteria – Eucaryota) in the late 1970’s, primarily based on small subunit ribosomal RNA sequence comparisons pioneered by Carl Woese and George Fox.

Figure: **Evolutionary tree showing the common ancestry of all three domains of life**: A highly resolved Tree Of Life, based on completely sequenced genomes. Bacteria are colored blue, eukaryotes red, and archaea green. Relative positions of some phyla are shown around the tree.
As more genome sequences become available, scientists have found that determining these relationships is complicated by the prevalence of lateral gene transfer (LGT) among archaea and bacteria. Due to lateral gene transfer, some closely related bacteria can have very different morphologies and metabolisms. To overcome this uncertainty, modern bacterial classification emphasizes molecular systematics, using genetic techniques such as guanine cytosine ratio determination, genome-genome hybridization, as well as sequencing genes that have not undergone extensive lateral gene transfer, such as the rRNA gene.

As with bacterial classification, identification of microorganisms is increasingly using molecular methods. Diagnostics using such DNA-based tools, such as polymerase chain reaction, are increasingly popular due to their specificity and speed, compared to culture-based methods. However, even using these improved methods, the total number of bacterial species is not known and cannot even be estimated with any certainty. Following present classification, there are a little less than 9,300 known species of prokaryotes, which includes bacteria and archaea. But attempts to estimate the true level of bacterial diversity have ranged from $10^7$ to $10^9$ total species – and even these diverse estimates may be off by many orders of magnitude.

There are four steps in general phylogenetic analysis of molecular sequences: (i) selection of a suitable molecule or molecules (phylogenetic marker), (ii) acquisition of molecular sequences, (iii) multiple sequence alignment (MSA), and (iv) phylogenetic treeing and evaluation.

Multilocus sequence analysis (MLSA) represents the novel standard in microbial molecular systematics. In this context, MLSA is implemented in a relatively straightforward way, consisting essentially in the concatenation of several sequence partitions for the same set of organisms, resulting in a “supermatrix” which is used to infer a phylogeny by means of distance-matrix or optimality criterion-based methods. This approach is expected to have an increased resolving power due to the large number of characters analyzed and a lower sensitivity to the impact of conflicting signals (i.e. phylogenetic incongruence) that result from eventual horizontal gene transfer events. The strategies used to deal with multiple partitions can be grouped in three broad categories: the total evidence, separate analysis, and combination approaches. The concatenation approach that dominates MLSAs in the microbial molecular systematics literature is known to systematists working with plants and animals as the “total molecular evidence” approach. It has been used to solve difficult phylogenetic questions such as the relationships among the major groups of cetaceans, that of microsporidia and fungi, or the phylogeny of major plant lineages. The total molecular evidence approach has been criticized because by directly concatenating all available sequence alignments. The evidence of conflicting phylogenetic signals in the different data partitions is lost along with the possibility to uncover the evolutionary processes that gave rise to such contradictory signals.