


# How to Perform Bioinformatic Analysis for Microbial Sequencing Data

The goal of all sequencing techniques is to provide possible identities for genes, map the entire genome of an organism, and to discover novel genes or even novel organisms. Downstream processing of sequencing results is essential to achieve these goals. Bioinformatics combines the principles of biological and computer science, aiming to collect, store, analyze, and provide access to biological data such as nucleic acid and amino acid sequences. Using bioinformatics, a large amount of information on molecular biology can be organized to provide standardized data and used by researchers worldwide to design tools for the analysis of these data, and to give an accurate and meaningful interpretation of the information.




Bioinformatics has been used for microbial biotechnology in ways such as computationally analyzing raw sequences, identify particular regions in a sequenced genome, infer on possible phenotypes, and understand the higher level of functions of these molecules. Bioinformatic databases aim to store the information that can be accessible to other scientists in order to come up with a universal and standardized understanding of these phenomena. Databases have been established to collect, annotate, and provide certain biological data that is needed to properly characterize particular sequences.

In order to perform bioinformatic analysis, it is important to familiarize oneself with the databases. Some databases contain links to other databases, depending on what you will need for your analysis. NCBI and KEGG are the most commonly used databases for bioinformatic analysis. There are also other specialized databases that may allow for more specific and accurate analysis.

**NCBI**

NCBI provides data retrieval systems and computational resources for the analysis of sequences for prokaryotic and eukaryotic organisms. Information on sequences, structures, functions, related journals and books, taxonomical characteristics, among others, are available on the website to cater to the different needs of researchers.

**KEGG**

KEGG is a collection of databases including information on genomes, biological pathways, diseases, drugs, and chemical substances, used for bioinformatics research and education, including data analysis in genomics, metagenomics, metabolomics and other omics studies, modeling and simulation in systems biology, and translational research in drug development.

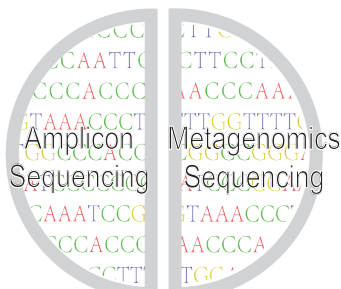








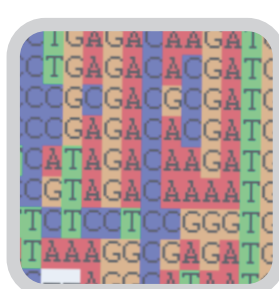
Sequencing can be one of the following: marker gene sequences, metagenome, meta-transcriptome, metabolome, and whole genome sequences. Bioinformatic analysis by computational search and sequence alignment can be used to identify homologous sequences available in the databases, annotation of novel, and established genes and their functions, use of mathematical modeling techniques to identify patterns, features, and functions, and a combination of these techniques to fully characterize the gene, and the organism of interest.



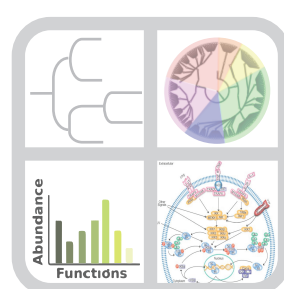
Raw Data



Homology Search



Alignment

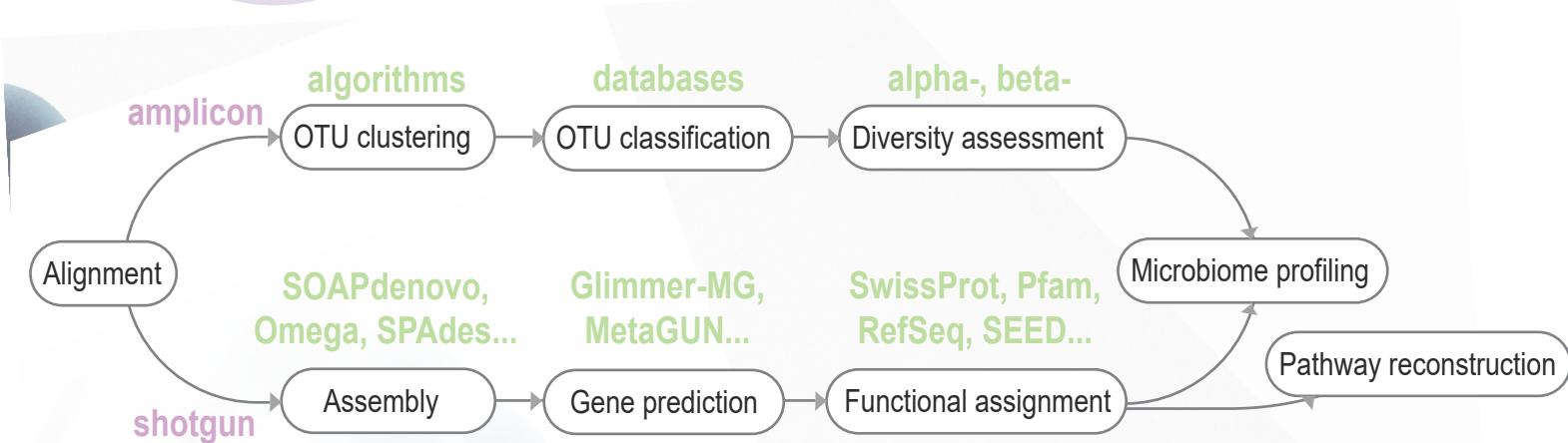


In-depth Analysis

Sequence alignment is one of the basic bioinformatics techniques used to identify sequences. Raw sequences collected from samples are aligned with available sequences in a particular database using established algorithms that measure the similarity and homology of the sequences. This is commonly done for samples of known organisms to identify mutations and divergence of the sequence in order to establish phylogenetic relationships.

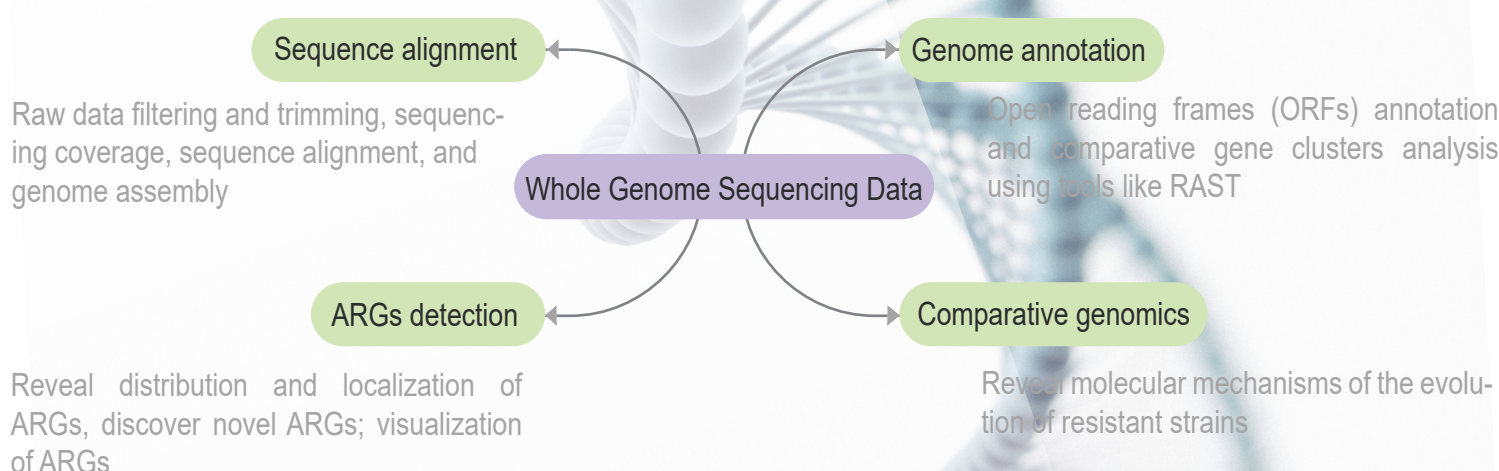


Different types of raw data can interpret microbiome distinctly, and there are a wide range of bioinformatic analysis arithmetic that aim to understand the behavior of microbial genes, their function, evolution, and relationships with other microbial groups. The approaches to be used in bioinformatic analysis depend on the information desired by the researcher and the purpose of the analysis. Ultimately, determining the goal of the research is essential to decide on the approaches of the sequence analysis.



## ..... Antibiotic Resistance Genes Analysis

In-depth analysis can deal with all kinds of issues, for instance, antibiotic resistance of microbes. This feature encoded in antibiotic resistance genes (ARGs), either intrinsic or acquired, indicates the antibiotic susceptibility of any given microbe. Compared to traditional microorganism screening methods, ARGs screening (by WGS) and bioinformatic analysis provide a high-throughput and much more efficient and accurate method for ARG analysis, which can easily identify and locate more than 1,000 types of ARGs, identify novel ARGs, and reveal the molecular mechanism underlying ARGs transfer. ARG research can be applied in many fields such as diagnostic tests, new antibiotics development, management and modification of existing antibiotics, and revealing the factors that promote the emergence and resistance of pathogenic microbes.



CD Genomics is a biotech enterprise using functional genomics BT+IT (high-throughput gene sequencing technology, biological information technology, high-performance biological cloud computing technology) technology. We provide customers with comprehensive data analysis services with regard to next-generation and third-generation high-throughput sequencing outputs. We are committed to providing quality services to customers through data mining, development, and standardized data analysis processes, as well as continuous and satisfactory assistance to customers in the fields of life science research, pharmaceutical research, and biotechnology. To learn more about CD Genomics – Bioinformatics Analysis, please feel free to contact us.