

## **Genomic study of fish pathogens using softwares**

Shri.V.Radhakrishnan Nai  
Scientist  
ICAR-CIFT, Cochin  
[vrnair08@gmail.com](mailto:vrnair08@gmail.com)

### **INTRODUCTION**

Micro-organism weighs over 60 % of living beings in weight though they are invisible with naked eyes. Most of them are either sub-microscopic or microscopic in nature. They are nature friendly for the sustenance of the human ecology. However, a few species are of pathogenic in nature to animals and human beings. Fisheries, a highly dynamic eco-system are no exception with the effect of pathogenic microorganism viz. fish pathogens. It is also noteworthy that in India, fisheries is a bio-economic system that contributes over 1% of GDP and over 5% of agricultural GDP.

In India, intensive fish farming often leads to disease problems resulting in serious economic losses. Bacterial pathogens are majorly the root cause of fish diseases and they include *Salmonella*, *E.Coli*, *Campylobacter*, *Staphylococcus spp.*, *Enterococcus spp.*, *Aeromonas hydrophila*, *Vibrio species* and *Edwardsiella tarda*. It includes other pathogens of bacteria such as *Streptococcus* and *Pseudomonas*. These pathogens cause many foodborne illnesses and infections such as tetanus, typhoid fever and diphtheria. Microorganisms express their pathogenicity by means of their virulence. The determinants of virulence of a pathogen are any of its genetic or biochemical or structural features that enable it to produce disease in a host. In bacterial host mediated pathogenesis, (e.g., tuberculosis), tissue damage results from the toxic mediators released by lymphoid cells rather than from bacterial toxins. This emphasizes the need to understand the pathogenesis and pathology of diseases in fisheries environment by identifying the pathogens.

### **MICROBIOLOGY**

Microbiology is the study of microscopic organisms. They are unicellular (single cell) or multicellular (cell colony). Microbiology includes sub-disciplines like virology, mycology, parasitology and bacteriology. Microbiologists carry out biochemical analysis for checking and identification of microorganisms. However, in the modern age Microbiologists often rely on molecular approaches by extraction or detection of nucleic acid, either DNA or RNA sequences to vindicate their stance.

**Antonie van Leeuwenhoek** is considered a father of microbiology as he observed microscopic organisms in 1676, using simple microscopes. Microbiology has been developed as a scientific discipline in the 19th century through the microbial studies carried out by **Louis Pasteur**. The main challenge faced by micro biologist is the size and number of microbes present in a unit space. It is difficult to assess the influence of microbes owing to its size of infinitesimally small dimension and infinitely large number of organism present in the host. However high-end computing and software facilities and computer aided instrumentations have mitigated the problems faced by microbiologists to a greater extent.

## **BIOTECHNOLOGY**

Biotechnology is an applied science to use living systems to develop or make products. In other words, it is the technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use. Exploitation of biological processes for industrial and other purposes through genetic manipulation of microorganisms is also emphasized in this field of science. For thousands of years now, humankind has been using biotechnology in the field of agriculture, food science and medicine. The term is believed to have been coined in 1919 by Hungarian agricultural engineer **Károly Ereky**. In the late 20th and early 21st centuries, biotechnology has expanded to include new and diverse sciences such as genomics, recombinant gene techniques, applied immunology, and development of pharmaceutical therapies and diagnostic tests.

While some fear having association of some microbes with various human illnesses, major chunk of microbes are responsible for numerous beneficial processes such as industrial fermentation (e.g. the production of alcohol, vinegar and dairy products), antibiotic production and as vehicles for cloning in more complex organisms such as plants. Scientists have also exploited their knowledge of microbes to produce biotechnologically important enzymes such as *Taq* polymerase, reporter genes for use in other genetic systems and novel molecular biology techniques such as the yeast two-hybrid system.

A variety of biopolymers, such as polysaccharides, polyesters, and polyamides, are produced by microorganisms. Microorganisms are used for the biotechnological production of biopolymers with tailored properties suitable for high-value medical application such as tissue engineering and drug delivery. Some polyester materials are used for fabrication of fishing gears.

Microorganisms are beneficial for microbial biodegradation or bioremediation of domestic, agricultural and industrial wastes and subsurface pollution in soils, sediments and marine environments. The ability of each microorganism to degrade toxic waste depends on the nature of each contaminant. Since sites typically have multiple pollutant types, the most effective approach to microbial biodegradation is to use a mixture of bacterial and fungal species and strains, each specific to the biodegradation of one or more types of contaminants.

The monstrous genera *Thiomargarita* and *Epulopiscium* in which some species of bacteria that measure over 600 to 700  $\mu\text{m}$  in length or diameter and are visible to the naked eye. However, large bacteria are rare in nature, Most of the bacteria size around 0.4 and 2  $\mu\text{m}$  in diameter and 0.5 and 5  $\mu\text{m}$  in length. It is all the more important to see that bacteria are boring, at least in morphological sense. Table1 gives the size of some of the bacteria.

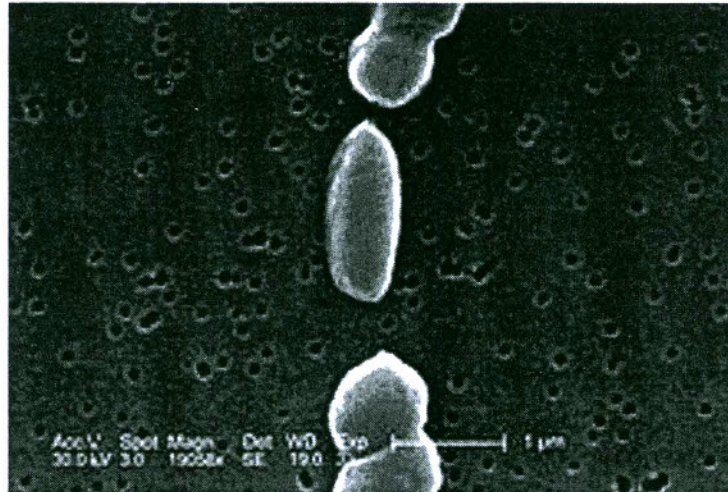
**Table 1: Dimension of size of bacteria**

SL.No	name	Dimension of size ( $\mu\text{m}$ =micron)
1.	<i>Clostridium spp</i>	0.3-2 x 1.5-20 microns
2.	<i>Aeromonas hydrophila</i>	0.3-1.0 x 1.0-3.5 $\mu\text{m}$
3.	<i>Vibrio spp.</i>	0.5 x 1.5-3.0 $\mu\text{m}$

- |    |                   |                             |
|----|-------------------|-----------------------------|
| 4. | <i>E.coli</i>     | 1-3 x 0.4-0.7 $\mu\text{m}$ |
| 5. | <i>Salmonella</i> | 0.7-1.5 x 2-5 $\mu\text{m}$ |

Bacteria usually are microns in diameter. Only general shape and major morphological features are visible in light microscope. In general, bacteria range from 0.2-10 microns. A scanning electronmicroscopic (SE) microgram of *Vibrio parahaemolyticus* is given in fig.1

Figure 1: SE Microgram of *Vibrio parahaemolyticus*



(Image source [https://en.wikipedia.org/wiki/Vibrio\\_parahaemolyticus#/media/](https://en.wikipedia.org/wiki/Vibrio_parahaemolyticus#/media/) dt.06/06/2018)

The excessively large number of microorganism present in a very small unit space and the infinitesimally small dimension of the microorganisms by nature are the challenges faced by microbiology as a scientific discipline. The digital era of computer science mitigates this problem to a greater extent by facilitating data collection from laboratory using high-end computer aided instruments, data processing and analysis for generation of information with the help of high speed and high precision computing facilities. Huge volume of genomic data is by far a challenging task for development, processing and management of databases either. Thanks to the development in the field of informatics and computing in the new era, available to cope up with the situation considerably.

#### **IMPORTANT PATHOGENS AND LIKELY PROBLEMS WITH FISH/FISHERY PRODUCTS**

Bacteria in food may cause illness in humans by infection or intoxication. Examples of some of the types of bacteria that may be found in seafood that cause foodborne illness by infection are *Vibrio*, *Salmonella*, *Shigella*, and *Listeria*. From 1973 to 2006, *Vibrio* species accounted for 38% of the outbreaks associated with seafood and 54% of the illnesses. *Salmonella* and *Shigella* each were associated with about 10% of the reported illnesses, and *Listeria monocytogenes* approximately 1%. Foodborne intoxications occur when patients consume pre-formed toxins that are produced by certain types of bacteria when they grow and multiply in the food. *Clostridium botulinum* can produce a potent neurotoxin during growth under anaerobic conditions (absence of oxygen) usually associated with vacuum packed, improperly canned, or fermented products. *C. botulinum* toxin was associated with almost a fourth

of the seafood related outbreaks from 1973 to 2006 and caused 152 illnesses and 38% of all hospitalizations. Bacteria such as *Staphylococcus aureus* can produce enterotoxins that cause foodborne illness, but less than 5% of the seafood associated outbreaks and illnesses were associated with this pathogen over the past three decades. Preventing the growth of these bacterial pathogens is important to prevent infection or intoxication when seafood is eaten.

### **FISH DISEASES**

Many diseases are caused in fishes due to pathogens which create serious concerns to fish farmers. Some of the fish pathogens causing diseases are *Aeromonas hydrophyla*, *A. salmonicida*, *Pseudomonas fluorescens*, *P. putrefaciens*, *Flexibacter columnaris*, *Edwardsiella tarda*, *Vibrio alginolyticus* and *V. parahaemolyticus*.

It is highly imperative to identify the pathogens either by biochemical method or by molecular method through the study of genomic sequence to tackle the causative reasons for the diseases prevailing in fishery environment as well as in fishery products borne diseases. Though biochemical methods are proven methods to identify the micro-organisms, molecular methods through the study of genomic sequences of the pathogenic cells is of great application with high precision in the digital era. Many a time it can be used as a method of corroboration for conventional detection techniques.

### **GENOMICS**

Genomics is an interdisciplinary field of science focusing on the structure, function, evolution, of genomes. A genome is the complete set of DNA of an organism, including all of its genes. Unlike in genetics, where the study of individual genes and its roles in inheritance is emphasized, genomics aims at the collective characterization and quantification of genes, which direct the production of proteins with the assistance of enzymes and messenger molecules. In turn, proteins make up body structures such as organs and tissues as well as control chemical reactions carry signals between cells. Genomics also involves the sequencing and analysis of genomes through uses of high throughput DNA sequencing and bioinformatics to assemble and analyze the function and structure of entire genomes. Advances in genomics have triggered a revolution in microbial research and systems biology to facilitate understanding of even the most complex biological systems like human brain.

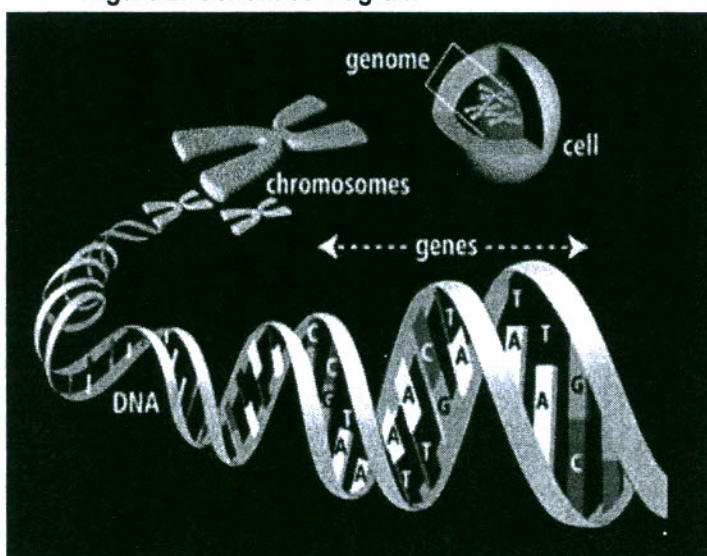
### **BIOINFORMATICS**

Bioinformatics is an interdisciplinary field of science that develops methods and software tools for understanding biological data. As an interdisciplinary field, bioinformatics combines computer science, statistics, mathematics to analyze and interpret biological data. Bioinformatics has been used for in-silico analysis (Computer based analysis) of biological data using mathematical and statistical techniques.

Bioinformatics is used as an umbrella term for the biological studies that use database management and computer programming, fondly known as informatics as part of their methodology. Common uses of bioinformatics include the identification of candidate genes whereby identify a species under microbiological wet lab study. Often, such identification is made with the aim of better understanding the genetic basis of diseases, unique adaptations, desirable properties (esp. in microorganism species), or differences between populations. In a less formal way, bioinformatics also tries to understand the organizational principles within nucleic acid and protein sequences. This will supplement and corroborate biochemical and other conventional techniques to strengthen the science of microbiology.

The genomic analysis and interpretation of data is based on **Central Dogma of Molecular Biology** which states that "The gene region of DNA in the nucleus of the cell is copied (transcribed) in to the RNA and RNA travels to protein production sites and is translated in to protein is the Central Dogma of Molecular Biology". This underlines the fact that DNA and the embedded genes are responsible for morphological characteristics and manifestation of response to every stimulus of living organisms. Genomic approaches have opened up new vistas for increasing the quality and their by productivity of biological systems. During the last decade, omics (field of study in biology ending in -omics, such as genomics, proteomics or metabolomics ) has witnessed an information explosion. Omics databases contain huge amount of information that are not amenable to traditional analytical approaches. In a multi-disciplinary area with a blend of biology, mathematics and computing science that can be used to derive biological insights from various omics data. It is an application of computing technology along with informatics for the management of biological information. A diagrammatic representation of genome is shown in fig.2.

**Figure 2: Genomics diagram**



(Image source.<http://www.differencebetween.info/difference-between-gene-and-genome> dt.30-07-2017)

The field of biotechnology has evolved such that the most important task now involves the analysis and interpretation of various types of data, including nucleotide and amino acids sequences, protein domains and protein structures.

### **DEOXYRIBONUCLEIC ACID (DNA)**

**Deoxyribonucleic acid (DNA)** is a thread-like chain of nucleotides carrying the genetic instructions used in the growth, development, functioning and reproduction of all known living organisms . DNA and ribonucleic acid (RNA) are nucleic acids; alongside proteins, lipids and complex carbohydrates (polysaccharides), they are one of the four major types of macromolecules that are essential for all known forms of life. Most DNA molecules consist of two biopolymer strands coiled around each other to form a double helix.

The two DNA strands are called polynucleotides since they are composed of simpler monomer units called nucleotides. Each nucleotide is composed of one of four nitrogen-containing nucleobases (cytosine [C], guanine [G], adenine [A] or thymine [T]), a sugar called deoxyribose,

and a phosphate group. The nucleotides are joined to one another in a chain by covalent bonds between the sugar of one nucleotide and the phosphate of the next, resulting in an alternating sugar-phosphate backbone. The nitrogenous bases of the two separate polynucleotide strands are bound together, according to base pairing rules (A with T and C with G), with hydrogen bonds to make double-stranded DNA.

The complementary nitrogenous bases are divided into two groups, pyrimidines and purines. In a DNA molecule, the pyrimidines are thymine and cytosine; the purines are adenine and guanine.

DNA stores biological information. The DNA backbone is resistant to cleavage, and both strands of the double-stranded structure store the same biological information. This information is replicated as and when the two strands separate. A large part of DNA (more than 98% for humans) is non-coding, meaning that these sections do not serve as patterns for protein sequences.

The two strands of DNA run in opposite directions to each other and are thus antiparallel. Attached to each sugar is one of four types of nucleobases. It is the sequence of these four nucleobases along the backbone that encodes biological information. RNA strands are created using DNA strands as a template in a process called transcription. Under the genetic code, these RNA strands are translated to specify the sequence of amino acids within proteins in a process called translation.

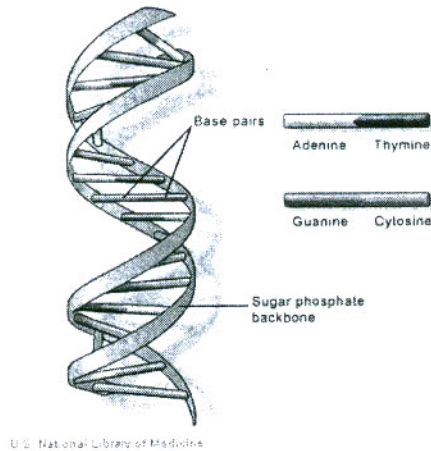
Within eukaryotic cells, DNA is organized into long structures called chromosomes. During cell division these chromosomes are duplicated in the process of DNA replication, providing each cell its own complete set of chromosomes. Eukaryotic organisms store most of their DNA inside the cell nucleus and some of their DNA in organelles, such as mitochondria or chloroplasts. In contrast prokaryotes (bacteria and archaea) store their DNA only in the cytoplasm. Within the eukaryotic chromosomes, chromatin proteins such as histones compact and organize DNA. These compact structures guide the interactions between DNA and other proteins, helping control which parts of the DNA are transcribed.

DNA was first isolated by Friedrich Miescher in 1869. Its molecular structure was first identified by James Watson and Francis Crick at the Cavendish Laboratory within the University of Cambridge in 1953, whose model-building efforts were guided by X-ray diffraction data acquired by Raymond Gosling, who was a post-graduate student of Rosalind Franklin. Anything a cell could possibly want is stored in its DNA. When a cell wants to build a protein, it finds the appropriate piece of DNA, makes a copy of it (called RNA), and uses the instructions in the copy to make the protein. In living organisms, DNA does not usually exist as a single molecule, but instead as a pair of molecules that are held tightly together. These two long strands entwine like vines, in the shape of a double helix. The nucleotide contains both a segment of the backbone of the molecule (which holds the chain together) and a nucleobase (which interacts with the other DNA strand in the helix). A nucleobase linked to a sugar is called a nucleoside and a base linked to a sugar and one or more phosphate groups is called a nucleotide. A polymer comprising multiple linked nucleotides (as in DNA) is called a polynucleotide.

The backbone of the DNA strand is made from alternating phosphate and sugar residues. The sugar in DNA is 2-deoxyribose, which is a pentose (five-carbon) sugar. The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar rings, which are known as the 3' and 5' carbons, the prime symbol being used to distinguish these carbon atoms from those of the base to which the deoxyribose forms a glycosidic bond. When imagining

DNA, each phosphoryl is normally considered to "belong" to the nucleotide whose 5' carbon forms a bond therewith. Any DNA strand therefore normally has one end at which there is a phosphoryl attached to the 5' carbon of a ribose (the 5' phosphoryl) and another end at which there is a free hydroxyl attached to the 3' carbon of a ribose (the 3' hydroxyl). The orientation of the 3' and 5' carbons along the sugar-phosphate backbone confers directionality (sometimes called polarity) to each DNA strand. In a double helix, the direction of the nucleotides in one strand is opposite to their direction in the other strand: the strands are antiparallel.

**Figure 3: DNA double helix structure**

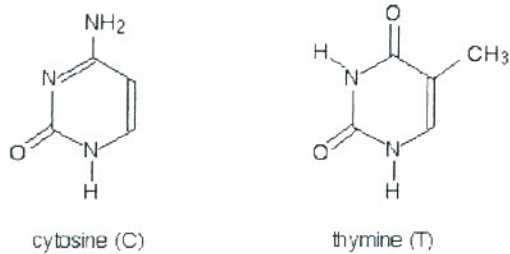


The four bases found in DNA are adenine (A), cytosine (C), guanine (G) and thymine (T). These four bases are attached to the sugar-phosphate to form the complete nucleotide, as shown for adenosine monophosphate. Adenine pairs with thymine and guanine pairs with cytosine. It was represented by A-T base pairs and G-C base pair

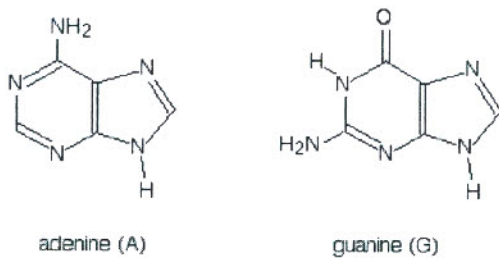
Proteins are the 'machinery' of a cell. They can perform many functions like transportation, structural support, movement and metabolism. Proteins are made from amino acids. There are twenty different amino acids that are used to build millions of different protein molecules. The principle of bioinformatics is that these molecules can be studied by using computers to analyze the DNA, RNA, and amino acid sequences from which they are created. Because there are so many different molecules, the best way we have of understanding how the entire system works is to use bioinformatics.

### **Base pairing**

In a DNA double helix, each type of nucleobase on one strand bonds with just one type of nucleobase on the other strand. This is called complementary base pairing. Here, purines form hydrogen bonds to pyrimidines, with adenine bonding only to thymine in two hydrogen bonds, and cytosine bonding only to guanine in three hydrogen bonds. This arrangement of two nucleotides binding together across the double helix is called a Watson-Crick base pair.



**Figure 4: Chemical Structure of nucleobase**



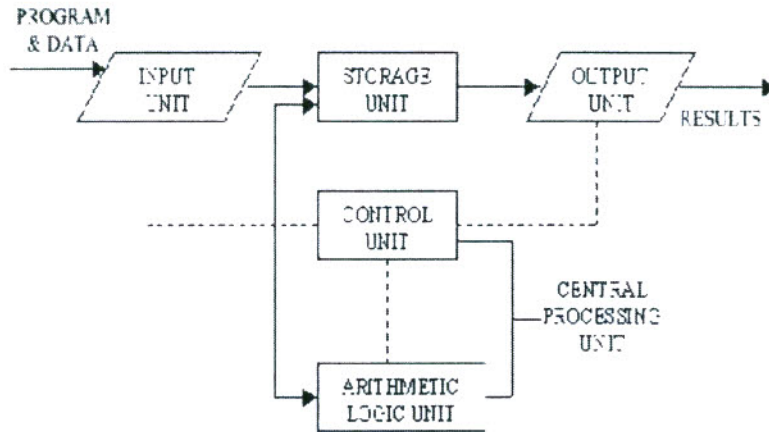
Traditionally molecular biology research was carried out entirely at the experimental lab but the huge increase in the scale of data being produced in this genomic era has necessitated incorporating computer in to research process. Sequence generation and its subsequent storage, interpretation and analysis are solely computer dependent. However, the molecular biology of an organism is a very complex issue with research being carried out at molecular level. The first challenge facing the bioinformatics community today is the intelligent and efficient storage of this massive data by providing reliable access to this data.

### **COMPUTER SYSTEMS**

A computer system allows users to input, manipulate and store data. It includes hardware's like processor, monitor, keyboard, mouse and other peripheral components along with software like operating system and other system and application programmes. All of these components also can be integrated into all-in-one units, such as desktop or laptop computers.

Very high speed and repetition of processing of data, high precision accuracy of result derived after data processing and its capacity storage area are the forte of computer systems which could be harnessed by the concerned users. Though the numerical data collected are however large or small it be, the processing with very high speed and generation of result with desired accuracy can be achieved through suitable instruction given to the computer system called computer programming. Logical flow of computer system is given in Figure 5.

**Figure 5: Logical flow of computer system**



On these count, data processing of voluminous microbiological data with generation of result with high accuracy is the order of the day with high-end speed computing system armed with versatile system and bioinformatics softwares and 4GL programming languages like Java and Perl.

### **BIOINFORMATICS AND AVAILABLE WEB BASED GENOMIC DATABASES**

Bioinformatics is the study of informatics processes in biotic systems. National Centre for Biotechnology Information (NCBI) defines "bioinformatics as the field of science, in which biology, computer science and information technology merge in to a single discipline. There are three sub discipline within bioinformatics: the development of new algorithms and statistics with which to assess relationships among members of large data sets; the analysis and interpretation of various types of data including nucleotides and amino acid sequences, protein domain and protein structures and the development and implementation of the tools that enable efficient access and management of different types of information. At the beginning of genomic revolution, the main concern was creation and maintenance of databases to store biological information such as nucleotide and amino acid sequences which are produced tremendously due to revolutionary developments in the fields of informatics and microbiology. This databases could be used to access existing data and to submit new and revised data to NCBI, (<http://www.ncbi.nlm.nih.gov/>)

### **BIOLOGICAL DATABASES**

Biological databases are huge databases mostly sequence data generating from major genome sequencing projects all over the world. The information about DNA, protein and functions of protein must be stored in an intelligent fashion so as to solve problems quickly by the available information stored in the data bank in the clouds, many of which are accessible by concerned user on internet. Some of the web based databases are available in table2

**Table 2: STANDARD WEB BASED GENOMIC DATABASES**

SI.No.	NAME OF THE DATABASE	DESCRIPTION
1.	PDB (Protein Data Bank)	Databank contains Protein Structures
2.	Swiss-Prot	Databank containing protein sequence and their functions



2. **Automated Colony Counter** Estimate No. of bacteria in a sample
3. **Electronic Colony Counter** Counting the colonies of bacteria in a petri-dish
4. **Magnetic Stirrer** Dissolving chemical substance effectively
5. **Sonicator** Rupture cells using high frequency waves
6. **Vortex Mixer** Used for thorough mixing of liquids in test tubes
7. **Laminar Flow Chamber** Used for aseptic transfer of sterilized materials, as well as for inoculation of microbes
8. **Electronic Cell Counter** Used to directly count the number of bacteria in a given liquid sample
9. **Microscopes** Used for visual observation of morphology, motility, staining and fluorescent reactions of bacteria
10. **Spectrophotometer:** Measuring the differences in color intensities of solutions.
11. **Automatic Bacteria Identification System** Used for automatic computer-assisted identification of bacteria
12. **PCR Thermo cycler** Used to amplify segments of DNA via the polymerase chain reaction (PCR)
13. **Ultra-centrifuge** Precipitating large biological molecules from solution or separate them by their different rates of sedimentation.
14. **Gas Chromatography (GC)** Used for separating and analyzing compounds that can be vaporized without decomposition
15. **High Performance Liquid Chromatography (HPLC)** Technique used to separate, identify, and quantify each component in a mixture.

16. **Thin Layer Chromatography (TLC)** Technique used to separate non-volatile mixtures.
- 17 **Paper Chromatography** A simple technique of separating constituents in a sample solution using a chromatography paper

### **GENOMIC SOFTWARES**

Online soft wares and genomic databases as mentioned above has revolutionized genomic analysis and made genomics a separate branch of science in this digital era. Identification of microorganism especially fish pathogens using analysis of genomic sequences using bioinformatics tools available with NCBI give more teeth for analysis of genomic data. We can customize software kits to identify the fish pathogens using NCBI freeware available (**Softwares were discussed in detail in practical classes**) . Now many software tools are available free software and paid software.

### **CONCLUSION**

Most critical task of bioinformatics involves the finding of genes in the DNA sequences of various organisms, developing methods to predict the structure and functions of the newly discovered proteins and structural RNA sequences, clustering protein sequences in to families of related sequences, development of protein models, aligning similar proteins and generating phylogenetic trees to examine the evolutionary relationships. The sequencing of the genomes of microbes should have enormous benefits for the biological systems including human health in general and fishery eco-system in particular. Computational analysis of this sequence data generated by genome sequencing is critically important. Bioinformatics tools can be used to search for the gene within this genome to understand their functions with the help of high-end computing facilities.

Microorganism enjoys key position in the sustenance of fish farming eco system. Though most of the microorganism is environment friendly there is some pathogenic microbe in aquatic system which affects fish health and in turn human health. Microbes play an important role in the degradation of fish products, thus better knowledge of the microbiological conditions throughout the fish production chain may help to optimize fish product quality and fishery resource utilization. Under these circumstances, a regular monitoring of fish health and the quality of fish products in perspective of microbiology with bioinformatics analytic tools is of paramount importance for the food security need of our country.

### **References:**

- Crick F. (1970), CENTRAL Dogma of Molecular Biology, *Nature*, 227, 561-563
- Cervera, M.T., Plomion, C. and C. Malpica (2000) Molecular markers and genome mapping in woody plants. In: Molecular biology of woody plants. Forestry Sciences, Volume 64 (S. M. Jain and S. C. Minocha, eds), pp. 375-394. Kluwer Academic Publishers, The Netherlands.
- Devey, M.E., M.M. Sewell, T.L. Uren, and D.B. Neale (1999) Comparative mapping in loblolly pine and radiata pine using RFLP and microsatellite markers. *Theoretical and Applied Genetics* 93:656-662.
- Gion, J. M., Rech, Ph., Grima-Pettenati, J., Verhaegen, D., and C. Plomion (2000) Mapping candidate genes in *Eucalyptus* with emphasis on lignification genes. *Molecular Breeding* 6: 441-449.

Glaubitz, J. and G.F. Moran (2000) Genetic tools: the use of biochemical and molecular markers. In: Forest conservation genetics: Principles and practice (Young, A., D. Boshier and T. Boyle, eds), pp. 39-59. CABI Publishing, United Kingdom.

<https://en.wikipedia.org/wiki/DNA> dated 06-05-2018

<http://www.ncbi.nlm.nih.gov/books/nkb21101>