CHAPTER 9

Molecular Analytical Techniques to be used in Taxonomy

All organisms are characterized by biological characteristics driven by their inherent genetic variations that enhance their fitness and survival to their living environment. Taxonomy describes and classifies organisms in respect of their unique morphological, genetic as well as behavioural characteristics. It gives a basic knowledge of the components of biodiversity, which is required for the decision making for effective conservation and sustainable use.

Awareness of the evolutionary history, taxonomic position and ages of divergence (phylogeny) of an organism is indispensable and molecular taxonomy and population genetics gives a precise information on species diversity by detecting DNA level variations and thereby powerfully contribute to taxonomic and biodiversity research (Hajibabaei et al., 2007). DNA sequence level analyses changed the perspectives of conventional taxonomic methods which is based on the external morphological and meristic features that has its limitations for an accurate conclusion. A whole specimen itself may exhibit significant intraspecific variations and little interspecific variations in their morphology. Egg and larval stage identification is complicated than adult. These issues can be clearly resolved by DNA based techniques. As a commercially important commodity in the world market, incorrect labelling of fish causes risks in the quality and threat of adulteration in edible fish products. Molecular taxonomic techniques make fish identification possible even after cutting and processing of fishes (Fomina et al., 2020).

The molecular techniques

Genetically controlled markers/Molecular markers are used to assess the genetic variation at DNA level (DNA markers) or through phenotypic expression that can be a protein (Protein markers). The emergence of molecular methods of species identification was only in the second half of 20th century. The very first technique used was based on the specific protein characterization using Electrophoresis (Isoelectric Focusing), capillary electrophoresis, HPLC and immunoassays (ELIZA). Rapid degradation, cross contaminations and differential expression in specific tissues etc are the limitations of using protein based techniques in a commercial system. The DNA based methods have been developed as an alternative only in the past two decades (Saritha et al., 2013). Compared to proteins, DNA is stable, have long life and found in all tissue types and secretions. Very small amount of sample is required for DNA extraction and DNA can be extracted even from processed, preserved and degraded samples. Also, DNA analysis is preferred due to larger variability of the genetic code. DNA markers are subdivided into Type I and Type II markers; Type I markers are associated with genes of known function and type II markers are associated with genes of unknown function.

Sajeela K A *ICAR-Central Marine Fisheries Research Institute, Kochi, Kerala*

The genetic markers detect interspecies and intra-species differences based on their rate evolution owing to the mutation and recombination. Intra-specific differences reveal stock composition and genetic relatedness within a species. Inter specific differences focuses on the delineation and phylogenetics. (Sunnucks, 2000). Depending upon question to be answered, suitable markers need to be selected for the respective species. If a specimen is suspected to be new, voucher specimen preservation for future reference is mandatory.

Non-specific DNA markers
Random markers are used when we target a segment of DNA of unknown function. The widely used methods of amplifying unknown regions are RAPD (Random Amplified Polymorphic DNA), Restriction fragment length polymorphism (RFLP) and AFLP (Amplified Fragment Length Polymorphism) DNA. These are simple methods that does not require any sophisticated equipment or prior sequence information of species.

The **RFLP** detects interspecific variations and generates species –specific bands profiles through Restriction digestion of DNA using one or more Restriction endonucleases. The fragments are visualized using conventional agarose gel electrophoresis. The RFLP profile of each species is the result of the unique genomic distribution of recognition sites and distance between different sites. Main disadvantage of RFLP is incomplete digestion and the addition or deletion of restriction sites as a result of intra-specific variations. Previous sample analysis detail is required for identifying the REN to be used.

RAPD is Random PCR amplification of DNA using short primers (9-10p long) of arbitrary nucleotide sequence. RAPD profiles are generated by the random PCR amplification of DNA segments using of usually 9 or 10 nucleotides long (Williams et al., 1990). RAPD randomly scan the genome. The primers anneal to different regions in the genome at low annealing temperatures and amplified between two nearby annealed primers in proper orientation. Specific banding pattern will be generated for each species in an electrophoretic gel as a result of difference in genomic binding location of primer binding sites. RAPD is also called Arbitrarily primed PCR or APPCR. It can be executed in a speedy and simple manner. Major disadvantages include inconsistency in the results and highly susceptible to DNA quality and quantity.

AFLP uses restriction enzymes to cut genomic DNA, followed by ligation of adaptors to the sticky ends of the fragments and then amplified using primers complementary to the adaptor and part of the restriction site fragments. After final amplification, selectively amplified fragments are separated by gel electrophoresis. AFLP is a combination of RFLP and RAPD for increased sensitivity, reproducibility and resolution. AFLP has high diversity index to develop a fingerprint of an organism. This technique has been used barely for fish species identification and mainly used in population genetics to determine slight differences within populations. The technique is laborious and costly as it requires expensive software packages for analysis.

Specific Nuclear DNA markers
Species-specific PCR (polymerase chain reaction). is the most common diagnostic method. Knowledge of nucleotide sequence of the gene is the prerequisite in this method. Species-specific primers are designed from the vast genomic sequences available and used for identification here. After amplification, the fragment visualized by electrophoresis and a positive result may give an idea about presence of a particular species, but a negative result

gives no information about the origin of the sample. Species-specific PCR can lead to false positive or false negative results, which require inclusion of reference samples in each analysis.

A modified version of conventional PCR is the **Real-time PCR** in which specific DNA sequence in a sample is amplified along with a fluorescent reporter molecule that enables detection and quantification of the accumulated product by a fluorescence detector. Real-time PCR is the most common technology to use for species identification. Post PCR stages like electrophoresis and staining can be eliminated and risks of contamination can be reduced significantly.

PCR-RFLP method has become more accurate for species detection. Specific gene primers used for amplification and the product is cut with corresponding restriction enzymes to generate smaller fragments and analysed by gel electrophoresis. The PCR-RFLP is a robust and easy method to use for identification of fish species. But optimization and analysis affect the reliability of results.

DNA Barcoding is introduced by Hebert et al. (2003) that involves biological identification through a 650bp mitochondrial cytochrome c oxidase I (COI) gene as a marker DNA sequence to create the global system for animal bio-identification. It is linked with Sanger sequencing that reads the bases in a small fragment of genome. DNA barcoding have been proposed as a fast, efficient, and inexpensive technique to catalogue all biodiversity. The sequence of the COI gene that amplified using universal primers and compared with the barcode library and the specimen is identified based on its closely matched sequence. The method of DNA barcoding is simple and less time consuming and the online barcoding libraries such as NCBI-BLAST and Barcode of Life Database (BOLD) helps in fast conclusions.

DNA microarray consists of small glass microscope slides, silicon chip or nylon membranes with many immobilized DNA fragments arranged in a standard pattern. A DNA microarray can be utilized as a medium for matching a reporter probe of known sequence against the DNA isolated from the target sample which is of unknown origin. Species-specific DNA sequences could be incorporated to a DNA microarray and this could be used for identification purposes. DNA extracted from a target sample should be labelled with a specific fluorescent molecule and hybridized to the microarray DNA. When the hybridization is positive a fluorescent signal is detected with appropriate fluorescence scanning/imaging equipment. Identification of hundreds thousands of species can be possible from PCR mixtures by DNA microarrays if species-specific probes are available. It is a cost effective and accurate method of species identification.

Microsatellites/Short tandem repeats (STRs)/simple sequence repeats (SSRs) are tandemly repeating sequences of 2-6 bp. These species specific hypervariable markers are used mainly for population genetic analysis.

Next generation sequencing (NGS) is similar in concepts with Sanger sequencing and differs with the sequencing volume. NGS efficiently generates millions of reads of short fragments results in sequencing hundreds to thousands of genes at one time. It has greater discovery power to detect novel or rare variants with deep sequencing.

DNA analysis are the commonly used method for fish species identification in recent times along with the conventional taxonomic procedures. PCR-based methods are the most promising

method that helps identifying different, even closely related fish species. Compared to all other methods, The DNA barcoding method with the use of NGS is the most promising but its high cost is the main disadvantage. Although DNA barcodes can significantly facilitate the process of species identification, comprehensive taxonomic analysis with several samples from the possible distribution ranges should be considered for validation of identification of a species, to avoid problems.

For further reading:

- Fomina, T. A., Kornienko, V. Y., & Minaev, M. Y. (2020). Methods Of Molecular Diagnostics For Fish Species Identification. Food systems, 3(3), 32-41.
- Hajibabaei, M., Singer, G. A., Hebert, P. D., & Hickey, D. A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. TRENDS in Genetics, 23(4), 167-172.
- Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(1512), 313–321. https://doi. org/10.1098/rspb.2002.2218
- Saritha .S, Amin, Adnan., Ramya V.C. Gowda and Naveen Kumar B.T. (2013). Molecular Techniques Used For Identifying Fish-A Review. Continental Journal of Agricultural Science. 7. 17-36. 10.5707/cjagricsci.