GENETIC STOCK CHARACTERIZATION OF FISH USING MOLECULAR MARKERS

A. Gopalakrishnan

Director
ICAR- Central Marine Fisheries Research Institute

32

Introduction

Accurate Identification of genetic resources is necessary for detecting new species and varieties for products of commercial value. Fish, as a group, apart from their economic value from a biodiversity viewpoint, have the highest species diversity among all vertebrate taxa. They exhibit enormous diversity in size, shape, biology and in the habitats they occupy. In terms of habitat diversity, fishes live in almost all conceivable aquatic habitats, ranging from Antarctic waters to desert springs. Of the 62,305 species of vertebrates recognized world over, 34,090 (nearly 52%) are valid fish species; a great majority of them (97 %) are bony fishes and the remaining (3 %) are cartilaginous (sharks and rays) and jawless fishes (lampreys and hagfishes). Further, on an average, 300 new fish species are described each year, and global surveys indicate that there could well be at least 5,000 species more to be discovered.

Loss of biodiversity is one of the greatest challenges facing modern society. This environmental crisis is increasingly evidenced by the loss or deterioration of genetic resources and habitats, as well as recent attempts to highlight and address the issue at the highest international levels. Appropriate conservation efforts for protection of the natural biological wealth warrant right attention for their sustainable utilization and for posterity. Public concern for biodiversity conservation has risen in the last 50 years and led to national and international policies, legislation, and actions to conserve biodiversity, notably the Convention on Biological Diversity (CBD). To conserve and sustainably utilize the bioresources of the country and for maintaining sovereignty over them, several nations enacted the Biological Diversity Act (BDA). This encompasses guide-lines to address a wide range of issues related to the utilization of bioresources and information within the country as well as by other countries.

Management of Fish Genetic Resources

The objective of management (documentation + conservation + sustainable utilization) of species and their habitats is to maintain the genetic identity and integrity of the species in their natural habitat as well as a genetically sustainable fishery. Hence, documentation of genetic variation and diversity is of vital significance to evolve conservation strategies with long-term impact. Genetic resources can be viewed as genetic differences at three hierarchical levels of organization, viz., species, populations and individuals. At the highest level, species consist of '**populations'** or **'genetic stocks'** that are reproductively isolated from populations of other species. Each species harbours a unique set of genetic material and therefore, conservation, may aim at a specific species, which requires sound knowledge about its biology, biogeography and within species (inter-populational) level genetic diversity. At the population level of organization, the identification of **discrete genetic breeding units** (usually called a '**stock**' in fisheries biology; this is roughly equivalent to a **'population'** or **'genetic stock'** to a geneticist) has been a major theme in fisheries research. The definition of a stock can vary, as the motivations of fishery managers may be influenced by political, economic or biological mandates. **Finally, the largest store of genetic variability in most species exists as genetic differences among individuals within a population**. Hence, the goal of pre-serving genetic variability in a population coincides with the goal of maintaining large ecologically sound natural populations. A fundamental need is to define distinct entities that range from individuals to species to ecosystems and beyond.

Population/Genetic Stock Identification (GSI)

Assessment of genetic variability is important for the management of wild genetic resources of fish. Most species are composed of **populations**, also called **genetic stocks**, between which limited gene flow occurs. These populations maintain their genetic makeup or characteristics distinct from other populations of the same species because of genetic variation within the species. This differentiation depends upon forces such as migration, mutation, selection, and genetic drift, which act on the species/population during its evolution. If such units are overfished, it is unlikely that population sizes will recover because of migration, and hence a collapse of the fishery may occur. Therefore, with the loss of a genetic stock, a species also loses the animals that are adapted to a particular habitat through evolution. Moreover, interbreeding of non-native fish stocks/species with a different make up tends to reduce the genetic variation that naturally exists between genetic stocks. In other words, different natural genetic identities available for a species in different habitats are lost.

A fundamental problem for fisheries management is the identification of populations/stock of a species and this idea has been brought together with the definition of stock for management. The term stock has been used in various management contexts with little or no genetic content. Several approaches have been advocated to solve this problem.

- Ihssen *et al*. (1981) defines a stock as '*an intra-specific group of randomly mating individuals with temporal or spatial integrity*'.
- Larkin (1972) *defined a stock as 'a population of organisms which share a common gene pool, is sufficiently discrete to warrant consideration as a self-perpetuating system which can be managed'.*
- In fishery management, *a unit of stock is normally regarded as a group of fish exploited in a specific area or by a specific method*.

If fishery managers are to include genetic considerations in their decisions, they will need information on the biological differences between discrete local groups of a species and they will need to understand the genetic and ecological processes that influence discreteness. Thus, the implementation of management strategies based on molecular genetic data can have indirect benefits for population biodiversity, as the main objective of such management plans is to avoid population crashes, which in turn benefits the maintenance of population genetic diversity.

Molecular Genetic Markers: The primary objective of the **genetic stock identification (GSI)** in fish is to assess the distribution and pattern of genetic variability at intra-as well as inter-specific population levels. The first priority for such research is identification of appropriate **molecular genetic markers** to assess genetic diversity. Fish stock identification was initially based solely upon morphological and meristic differences. Because these characters can be influenced by the environment, their variations may not have a genetic basis, and hence do not necessarily provide information on genetic and evolutionary relationships. In the 1950s, dissatisfaction with performance of phenotypic methods for stock identification encouraged early exploration of genetic markers. The markers developed have spurred development of statistical algorithms and revolutionized the analytical power necessary to explore genetic diversity among populations. Methods that take advantage of naturally occurring genetic markers have attracted a good deal of attention because application of physical tags is very labour intensive, and biological markers, such as scale patterns, can vary dramatically from year to year. The first GSI methods using **soluble proteins** and **gene products** such as **allozymes** (enzymes at cellular level) for estimating the contributions of two or more salmon stocks to a mixed harvest were developed in the late 1970s. Since then, the rapidly expanding availability of highly variable genetic markers and refinements in statistical analyses have considerably increased the ability to analyze the stock structure of different fish species; but this has also led to the genetic 'marker wars' among fish geneticists during the past several decades. For many years, allozymes were the universal workhorse genetic makers, and they made many valuable contributions to basic and applied conservation and management. Around 1980, the first applications of **mitochondrial DNA (mtDNA)**

analysis to natural populations were published, and gradually, it replaced allozymes and provided answers to key management questions regarding stock structure. The development of DNA amplification using the **polymerase chain reaction (PCR)** technique has opened up possibility of examining genetic changes in populations over the past 100-years or more even using archive material. In PCR reaction, a DNA sequence can be amplified many thousand folds to provide sufficient product for restriction analysis or direct sequencing. Once appropriate primers are available, large number of individuals can be assayed quickly thus facilitating large population screening for variability. Portions of the mtDNA such as, the ATPase 6 and 8 and hypervariable trans-membrane segments of cytochrome *b* (Cyt*b*) that evolve exceptionally rapidly have been used for high-resolution analysis of genetic stock structure in fish. Although mtDNA has indeed provided a wealth of new insights, it is not a solution and has some limitations with respect to fishery management (*e.g.,* it is maternally inherited, so provides information only about female migration or gene flow, and it is only a single marker and hence has much less power than a full suite of nuclear markers).

In the 1990s, **microsatellites** (**Short Tandem Repeats**— **STRs** or **Simple Sequence Repeats—SSRs**) muscled aside mtDNA and these highly variable **co-dominant markers** have provided greatly increased power and opened up exciting new opportunities (e.g. parentage analysis and individual assignments) that were generally not feasible with allozymes or mtDNA. Microsatellites are repeated DNA sequences having a unit length of 2-6 base pairs tandemly repeated minimum 6 times usually; maximum several times at each locus. They are found in all prokaryote and eukaryote genomes investigated to date. Individual alleles at a locus differ in the number of tandem repeats of the unit sequence owing to gain or loss of one or more repeats and they as such can be differentiated by electrophoresis according to their size.

There are four types of microsatellites

1. Perfect: Perfect tandem repeat sequences.

2. Imperfect: Tandem repeat sequences with intervening sequences.

3. Compound: More than one kind of repeats, adjacent ones.

4. Complex: More than one kind of repeats, with intermediary sequences.

Based on the number of base pairs in a repeat unit, microsatellites can be again classified into *mono* (e.g. C or A), *di* (e.g. CA), *tri* (e.g. CCA), *tetra* (e.g. GATA), *penta* (e.g. CGATA) and *hexa* (e.g. ATGGCA) repeat unit microsatellites. Microsatellites that are used in stock identification studies typically contain di- (AC)n, tri-(ACC)n, or tetra-nucleotide (GATA)n repeats. The most common ones are dinucleotide repeats. Tetra-nucleotide microsatellites are gradually replacing dinucleotide loci as the preferred genetic marker for stock analysis. Microsatellite loci are abundant in all eukaryote genomes and it has been estimated that there are from 10^3 to 10^5 microsatellite loci dispersed at 7 to 10^{10} base pair (bp) intervals or one locus at every 100-300 kilobase pair (kbp) intervals in the eukaryotic genome. Fish genomes may contain more microsatellite loci than most other invertebrate and vertebrate taxa. Mapping studies suggest more or less even distributions of microsatellites throughout genomes, although they are somewhat rarer within coding sequences.

Several features of STR render them invaluable for examining fish population structure. Microsatellites are codominant in nature and inherited in Mendelian fashion, revealing polymorphic amplification products from all individuals in a population. They contain information, which are directly related to the effective number of alleles at each locus. PCR for microsatellites can be automated for identifying simple sequences repeat polymorphism. Small amount of samples of blood or alcohol preserved tissue is adequate for analyzing them. Because they are highly variable in nature, abundant variants are ensured for characterization of populations. However, sample size in excess of 50 may be required to represent the genotype frequencies. The microsatellites are non-coding and therefore variations are independent of natural selection. These properties make microsatellites ideal genetic markers for defining population genetic diversity and distance measures. Because most STR loci are unlinked and inherited independently, the greater the number of loci screened, the greater the likelihood of selecting loci that reveal significant allelic frequency differences among populations and more statistical power is gained in quantifying the extent of genetic differentiation among populations. Additionally, analysis of a larger number of loci may provide a more accurate picture of the evolutionary history of the genetic stocks.

Analysis of microsatellite polymorphisms is a PCR-based approach in which oligonucleotide primers are designed based on unique single-copy sequences flanking the microsatellite repeats. DNA extracted from tissue samples are subjected to PCR reactions. PCR primer pairs are selected such that PCR products are of small molecular size (usually <350bp), providing relative ease in amplification from low-quality DNAs and also allowing for distinguishing small differences in the molecular size of alleles among individuals by using polyacrylamide-gel electrophoresis or automated DNA sequencers. Ideally, each individual shows a single (homozygote) or two-band (heterozygote) DNA pattern, with one band inherited from each parent. Polymorphic alleles at a locus are usually characterized by their molecular sizes. For dinucleotide repeats, these will differ by two base units. Based on the STR allele frequency data, powerful statistical tests are employed to arrive at a decision whether the genetic stocks of a species are significantly different from one another.

However, the field now seems poised to shift towards another type of marker, **single-nucleotide polymorphisms (SNPs)**. Like allozymes, SNPs are generally diallelic, so each marker has less power than a single microsatellite locus. They occur in vast numbers throughout the genome; therefore, eventually large overall increases in power are possible.

Furthermore, once developed, SNPs can be assayed more reliably and cheaply than microsatellites, which could be a considerable advantage in large-scale fishery management applications. However, development of sufficient numbers of SNP markers will be neither easy nor cheap, and analytical issues such as minimizing ascertainment bias remain to be resolved. Despite growing competition from new genotyping and sequencing techniques and latest class of markers, the use of the versatile and cost-effective microsatellites continues to increase, boosted by successive technical advances. Next-generation sequencing (NGS) technologies and the rise of commercial services allow the identification of large numbers of microsatellite loci at reduced cost in non-model species. As a result, more stringent selection of loci is possible, thereby further enhancing multiplex quality and efficiency. Numerous examples also exist where microsatellite analysis is used for fish population analysis and management of Pacific salmon (Fisheries and Oceans Canada website: http://www.pac.dfo-mpo.gc.ca/science/facilities-installations/pbs-sbp/mgl-lgm/proj/ index eng. htm online.) and also for cod where microsatellites have even been used as evidence in a court cases against a fishermen claiming a false origin of his catch. **Use of 20-25 polymorphic microsatellite loci (preferably tetra-nucleotide repeats) and 70–100 individuals from each population has become the standard and scientifically accepted protocol for population genetic analysis of fish along with information on biology and morphometry (*TRUSS*) data. Sequence information of mitochondrial complete ATPase 6/8 and Cytb genes of at least 20 individuals per population are also often generated along with this**.

Genetic Stock Structure in fish:

Distinct population structure has been observed in many fish species across the world indicating that propagation-assisted restoration programmes must be stock-specific to replenish declining populations. Generally, between populations of marine and freshwater species, marked differences exist in the level of genetic differentiation and genetic diversity, with marine species generally exhibiting lower levels of inter-population differentiation and greater genetic diversity. This is mainly due to the higher effective population sizes and/or higher inter-population migration rates in marine environments compared with freshwater. In addition, marine fishes and invertebrates are generally broadcast spawners and hence have large potential for movement between areas by larval drift in currents. In addition, adults of many species are capable of making long distance migrations. Early genetic studies of commercially important marine fishes using allozymes and proteins indicated that they generally had moderate levels of gene diversity and little population subdivision, often covering over several hundred kilometers. However, unexpected fine-scale population sub-structuring and deep genetic lineages have been observed in recent studies with high-resolution markers in many fishes which calls for further in-depth integrated approaches of molecular genetics with life-history traits. This will prove whether the variability is due to isolation or adaptations to particular marine habitats or as a result of non-genetic factors such as large reproductive variation among families. Regular monitoring of populations is also essential to enable a distinction between normal population-size fluctuations and those severe enough to warrant conservation measures.

The greatest genetic threats in the marine ecosystem are the extinction of genetically unique subpopulations and loss of genetic diversity primarily through overfishing and climate change. Illegal unreported and unregulated (IUU) fishing also contributes to this condition, and thus poses a severe threat to marine ecosystems. Controlling for compliance and enforcing fishing regulations is hampered by difficulties in identifying the geographical origin of fish and fish products, at point of landing and further down the food supply chain. Presently, there are no validated genetic methods for identifying the geographical origin of marine fish and investigate commercial fraud. '**FishPop-Trace**'(https://fishpoptrace.jrc.ec.europa.eu/web/fishpop trace/) is an international project, funded by the European Union (EU) framework programme (FP7), aiming to generate forensically validated reference panels of SNP markers for geographical origin assignment in four commercially important fish species, cod (*Gadus morhua*), hake (*Merluccius merluccius*), herring (*Clupea harengus*) and common sole (*Solea solea*). SNP markers are selected these are subsequently genotyped across populations to provide high resolution data to analyze genetic variation. These markers are validated to be used as tags for traceability and enforcement applications leading to a reduction in IUU fishing and conservation of remaining marine resources.

For a successful stocking programme such as sea ranching of endangered seahorse or sacred chank, genetic structure of the original wild population must be determined before any new fish are released into the waters. This information can be used to develop hatchery guidelines for breeding fish for stocking purposes. By ensuring that the stocked population is having the same genetic make-up as the wild population, re-integration of the stocked fish will likely be more successful and deviations from the original genetic structure will be minimal.

Integrating Population Genetics Data into Marine Fisheries Management

Maintaining the maximum level of genetic variations in fish stocks is vital for the preservation of genetic resources. Therefore, excessive loss of genetic variability should be avoided for sustainable management of resources. Application of molecular marker techniques to a number of species has shown that these methods can provide information on genetic stock structure that can be of direct management relevance. However, such information has not always been incorporated into fishery management and policy decisions in several countries. The complex problem requires agreement among scientists, governmental organizations and policy makers to define and implement policies on the sustainable management of these natural resources. Numerous factors (as mentioned below) have contributed to the imperfect integration of genetic data into management of aquatic species.

The fish stock assessment teams generally include quantitative fishery biologists and statisticians. In appropriate situations, the teams should be expanded to include geneticists as well as field biologists. It is always better that fish geneticists fully understand the complexities of the management process so that genetic information can be packaged in the most effective manner, and importance of GSI can be portrayed effectively for the policy makers. Also the managers involved in monitoring of fishery resources should acknowledge that GSI can provide valuable management information. Scientists, managers and policymakers could work together more effectively to foster productive dialogue to link statutory definitions and management or conservation goals.

It is difficult to develop an ideal sampling design for a genetic study without understanding the details of the life history of the target species and physical processes in the aquatic ecosystem. Genetic data can be integrated with other types of biological and oceanographical information. The sampling design of genetic studies does not always match the geographical regions to which management controls are applied. This can rarely result in discrepancy between biological and genetic management units. Implementing GSI over a broad geographical area requires extensive efforts to collect baseline data for populations from different coasts and to standardize laboratory procedures so that comparable data can be obtained by different laboratories. This requires funds, broad collaboration among laboratories and a willingness to share unpublished data.

Most fish geneticists are unfortunately, not exposed to the techniques of statistical model and decision analysis that form the basis for modern stock assessment science. Equally, managers and assessment biologists similarly would benefit from a greater literacy regarding the genetic principles that can profoundly affect the aquatic living resources for which they share stewardship responsibility. Therefore, it might be necessary to develop brief **integrated training courses** to equip geneticists and managers to work on assessment teams.

The purpose of stock assessment in fisheries is to provide timely and appropriate scientific advice on fisheries management for sustained production. Though there are few multi-species models, the assessments are almost mostly conducted for single species, whereas in reality, stocks are influenced by multi-species interactions. In addition, gears mostly harvest many species at a time, leading to difficulty in implementation of the management measures derived from single species stock assessment. Due to the lack of adequate and efficient models for multi-species interactions, stock assessments will generally continue to be based on single species models. Although the main approach in population genetic studies of natural populations still involves collecting individuals from two or more geographical locations and considering them as putative populations, *landscape genetics/seascape genetics*—the study of spatial genetic patterns in continuously distributed species—is rapidly evolving and the methods are beginning to be applied especially to marine species as well. These studies are expected to provide important insights into biological processes leading to effective multi-species stock assessment and management of marine ecosystems. However, considerable dialogue between geneticists, stock assessment scientists and managers, as well as creative thinking on both sides are required to develop effective ways to integrate these insights into fisheries management.

In conclusion, fish genetic stock diversity conservation requires preservation of as much variation as possible at all taxonomic levels and concerted efforts by integrating capture, culture fisheries and environmental programmes using latest technological innovations. The genetic tools will provide innovative means in the future and are an assuring approach for food security of the world and in reducing the fishing pressure on natural resources. Genetic data need to be integrated with other types of biological and oceanographical information for understanding the details of the life history of the target species and physical processes in the marine ecosystem. Although better monitoring of biodiversity, better assessment of risk and a more strategic approach to conserving biodiversity are all essential components to successful risk management, an equally important need is the open dialogue among geneticists, quantitative fishery biologists, statisticians, conservationists and planners that would help sustainable management of stocks of the world's amazingly rich assemblage of fishes.

Suggested Reading

Ayyappan S., Jena J. K. and Gopalakrishnan A., 2014. Molecular Tools for Sustainable Management of Aquatic Germplasm Resources of India. *Agricultural Research* 3(1): 1-21.

Ihssen P.E., Booke H.E., Casselman J.M., McGlade J.M., Payne N.R., Utter F.M., 1981. Stock identification: materials and methods. *Can J. Fish. Aquat. Sci.,* 38:1838–1855.

Larkin P.A., 1972. The stock concept and management of Pacific salmon. *In*: Simon R.C., Larkin P. A. (Eds) The stock concept in Pacific salmon, H.R. MacMillan Lectures in fisheries. University of British Columbia, Vancouver, p 231.