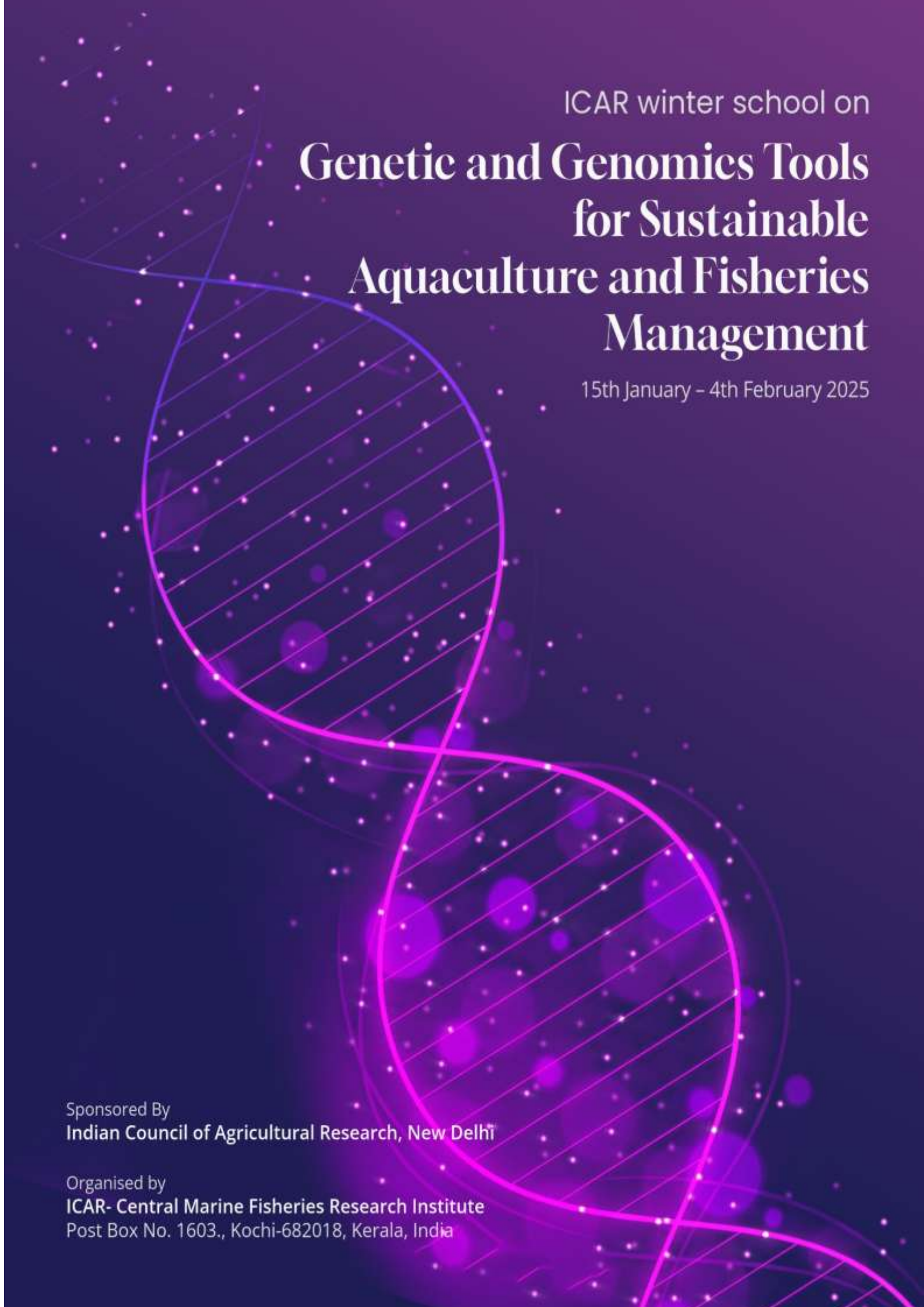
ICAR winter school on

# Genetic and Genomics Tools for Sustainable Aquaculture and Fisheries Management

15th January – 4th February 2025



Sponsored By
**Indian Council of Agricultural Research, New Delhi**

Organised by
**ICAR- Central Marine Fisheries Research Institute**
Post Box No. 1603., Kochi-682018, Kerala, India

# Training Manual

ICAR Sponsored Winter School on Genetics and Genomic Tools for Sustainable Aquaculture and Fisheries Management

15th January to 4th February 2025.

CMFRI Training Manual Series No. 75/2026

ICAR – Central Marine Fisheries Research Institute

| | | |
|---|---|---|
| Publisher | : | Dr. Grinson George<br>Director<br>ICAR – Central Marine Fisheries Research Institute<br>Ernakulam North P.O., Pin – 682018, Kochi, Kerala. |
| Editor | : | Dr. Sandhya Sukumaran |
| Editorial Assistance | : | Ramitha P. A.<br>Vysakh V. G. |
| Course Director | : | Dr. Sandhya Sukumaran<br>Principal Scientist<br>Marine Biotechnology Fish Nutrition Health Division<br>ICAR-CMFRI, Kochi |
| Course Co-ordinators | : | Dr. Sumithra T. G., Scientist, MBFNHD<br>Dr. Chandrasekar S., Scientist, MBFNHD<br>Dr. Jeena N. S., Senior Scientist, MBFNHD |
| Cover Design | : | Abilash P.R. |

For citing this book.

Sukumaran, S.(ed) (2025). Genetics and Genomic Tools for Sustainable Aquaculture and Fisheries Management. Training Manual. ICAR- Central Marine Fisheries Research Institute, Kochi, India

For citing a Chapter

V. V. R. Suresh (2025). Mariculture in India: challenges and prospects. In: Sukumaran, S.(ed) (2025). Genetics and Genomic Tools for Sustainable Aquaculture and Fisheries Management. Training Manual. ICAR- Central Marine Fisheries Research Institute, Kochi, India



ISBN:

भारतीय कृषि अनुसंधान परिषद
Indian Council of Agricultural Research
केन्द्रीय समुद्री मात्स्यिकी अनुसंधान संस्थान
Central Marine Fisheries Research Institute
(कृषि अनुसंधान एवं शिक्षा विभाग, कृषि एवं किसान कल्याण मंत्रालय, भारत सरकार)
(Department of Agricultural Research and Education, Ministry of Agriculture and Farmers' Welfare, Govt. of India)
पोस्ट बोक्स सं. 1603, एरणाकुलम नोर्थ पी.ओ., कोच्ची - 682 018, केरल, भारत
Post Box No. 1603, Ernakulam North P.O., Kochi - 682 018, Kerala, India
Phone: 91 484 2394867/2391407, Fax: 91 484 2394909/2396685, Email: director.cmfri@icar.org.in, Web: www.cmfri.org.in



# Foreword

The progress in molecular biology and biotechnology over the past few years has been immense, driven primarily by the ongoing revolution in high-throughput DNA sequencing technologies. Marine biology and fishery science have significantly benefited from these modern innovations, dramatically accelerating our understanding of complex adaptations, functional genomics, and population dynamics in aquatic organisms. Classic fish genetics has rapidly shifted toward advanced genomic platforms incorporating high-resolution data from neutral and non-neutral markers. Modern breakthroughs in Next-Generation Sequencing (NGS), transcriptomics, and environmental DNA (eDNA) metabarcoding now offer unprecedented capabilities for real-time monitoring of marine fish stocks, precise species characterisation, and mapping the microbial environments essential for fish health. These rapid, highly cost-effective tools present exciting new frontiers for ecological research, marine biodiversity conservation, and the sustainable scaling of aquaculture productivity through molecularly guided management strategies.

The ICAR-Central Marine Fisheries Research Institute (CMFRI) has made significant contributions to marine biotechnology research in the country and played a pivotal role in the development of the marine fisheries sector. The ICAR Winter School on **"Genetic and Genomics tools for Sustainable Aquaculture and Fisheries Management"** conducted at ICAR-CMFRI, Kochi, from **15th January to 04th February 2025,** is specially designed to provide hands-on exposure to various cutting-edge applications of molecular tools in fisheries resource management, molecular taxonomy, functional genomics, nutrigenomics, and metagenomics. I hope this training manual and compendium of lectures will be extremely useful for the participants to effectively utilise these advanced genomic tools in their respective research areas. On behalf of ICAR-CMFRI, I warmly welcome all the participants from various institutions and wish them great success in their future endeavours. I am certain that this training program will foster new knowledge, long-term collaborations, and enduring professional friendships.

**Grinson George**
Director

ICAR-CMFRI

# PREFACE

Knowledge about DNA, the fundamental building block of life, has contributed significantly to the progress of society as a whole by enabling vital interventions to decipher, sustain, and regulate genetic functions across medicine, agriculture, veterinary sciences, and fisheries management. While the foundational principles of the genetic code have been understood for decades, the rapid evolution of DNA sequencing technologies - progressing from first-generation methods to high-throughput Next-Generation Sequencing (NGS) platforms has fundamentally transformed modern biological research, making molecular information indispensable to successful fishery management programs and sustainable aquaculture ventures. With this in view, an ICAR Winter School was organized on **"Genetic and Genomics Tools for Sustainable Aquaculture and Fisheries Management"** at the Marine Biotechnology, Fish Nutrition and Health Division of ICAR-Central Marine Fisheries Research Institute (CMFRI), Kochi, from 15th January to 4th February 2025, with prestigious funding support from the Indian Council of Agricultural Research, New Delhi.

Most of the critical aspects of genetics and genomics that are highly relevant to marine biology and tropical fisheries are included in this training manual, highlighting core themes such as molecular barcoding for marine biodiversity characterization, population genetics for stock identification, functional genomics, population genomics, transcriptomics, and NGS workflows. In addition to these primary areas, advanced modules addressing environmental DNA (eDNA)-based metabarcoding for fish stock estimation, functional applications like nutrigenomics, and metagenomic tools for fish health management are comprehensively covered, alongside detailed laboratory protocols, practical demonstrations, and hands-on bioinformatics software exercises. I hope that this compendium of lectures and technical protocols will prove to be an invaluable asset to scientists, faculty members, and researchers who wish to integrate and apply these state-of-the-art molecular technologies within their respective fields of research.

**Sandhya Sukumaran**

Course Director

# Contents

## Metagenomics and its prospects in fisheries

# Overview of the Marine Biotechnology, Fish Nutrition & Health Division at ICAR-CMFRI: Advanced biotechnological tools for sustainable aquaculture and fisheries management

Kajal Chakraborty*

Marine Biotechnology Fish Nutrition and Health Division, ICAR-CMFRI, Kochi-18 ,*kajal.chakraborty@icar.org.in

**Abstract**

The Marine Biotechnology, Fish Nutrition & Health division at ICAR-CMFRI spearheads advanced research in fish genetics, genomics, nutrition, health, and marine bioprospecting, emphasizing sustainable aquaculture and resource utilization. Its innovations, including genome studies, stress-tolerance gene identification, pathogen diagnostics, and novel aquafeeds like insect-protein-based formulations, cater to societal needs for health, nutrition, and eco-friendly solutions. Leveraging biotechnological tools, the division explores marine biodiversity for bioactive compounds with therapeutic and commercial potential, producing nutraceuticals like Cadalmin™ and pioneering technologies such as in vitro fish flesh production. This chapter highlights how these advances address global demands for sustainable food systems, biomedicines, cosmeceuticals, and renewable resources while safeguarding marine ecosystems.

**An introduction to the Marine Biotechnology, Fish Nutrition & Health Division at ICAR-CMFRI**

The Marine Biotechnology, Fish Nutrition & Health division at CMFRI focuses on cutting-edge research in various aspects of fish genetics, genomics, molecular biology, fish physiology, nutrition, health, biochemistry, and bioprospecting. Ongoing research projects cover a wide range of topics including the development of diagnostic tools for fish diseases, exploration of marine microalgae for stress tolerance, and the formulation of nutraceuticals from marine organisms. Major achievements include the identification of stress tolerance genes from microalgae for potential use in transgenic studies, the development of formulated feed for marine ornamental fishes, and the establishment of sensitive diagnostic methods for detecting viral diseases in finfish populations. Additionally, the division has made significant contributions to understanding the presence of Perkinsus beihaiensis in Indian oyster populations and the development of novel nutraceuticals from green mussels and marine

macroalgae. These advancements underscore the division's commitment to leveraging biotechnology for the improvement of fish health, nutrition, and sustainable aquaculture practices.

**Thrust Areas of Research**

- Stock structure identification using genetic and genomic tools to formulate effective management strategies for sustainably managing important marine resources, and using advanced biotechnological tools, such as genetic engineering, omics, and gene editing to improve desired mariculture traits, and to support genetic improvement programs of prioritized maricultured species.
- Develop nutritionally balanced feed formulations and species-specific larval feed using functional feed additives, feed attractants, and exploring alternative and novel feed ingredients to reduce dependency on fishmeal and fish oil in aqua feed, thus ensuring the long-term sustainability of aquaculture.
- Develop high-value bioactive products, biomaterials, nutraceuticals, marine biomanufacturing, and cosmeceuticals from seaweeds, marine microalgae, marine finfish and shellfish, marine microbes, extremophiles, and other marine organisms with potential therapeutic applications, and for the advancement of marine biotechnology.
- Developing disease diagnostic tools, prophylactics, and vaccines to strengthen preparedness against existing and emerging diseases in mariculture, pathogen profiling and surveillance efforts to monitor and prevent disease outbreaks in mariculture systems and marine environment, and to enhance the overall health and sustainability of mariculture practices and marine fishery resources.
- Developing *in vitro* fish flesh production systems.

**Major Achievements of Marine Biotechnology, Fish Nutrition & Health Division at ICAR-CMFRI**

Fish Genetics and Genomics

- Whole genome and transcriptome assembly of the Indian oil sardine, *Sardinella longiceps* and the Asian green mussel, *Perna viridis*

- Genetic stock structure of Indian oil sardine, Indian mackerel, Indian anchovy, ribbon fish, carangids, threadfin breams, scalloped hammerheads, oceanic whitetips, spadenose sharks, Spanish mackerels, king mackerels, lobsters, and squids delineated using advanced biotechnological tools.
- Highly differentiated SNPs detected in the Indian oil sardine, Sardinella longiceps using genomic tools which can serve as geographic tags.
- The complete mitogenome of the Indian oil sardine, gold stripe sardinella, Randall's thread fin bream, Indian scad, mud spiny lobster, black clam, pearl spot and ribbon fish has been characterised which provided valuable insights into taxonomy and adaptation of OXPHOS genes in these important species.
- Characterized Whole mitogenome of the edible oyster *Crassostrea madrasensis*.
- The transcriptome profiling of the metamorphosing flat fish, Indian Halibut, *Psettodes erumei* to understand their cellular, biological, and functional modifications to their efficient and successful production

Nutrition

- Commercialisation and release of of aquafeed- BSF Insect Protein-Based Fish Feed Technology licensed (non-exclusive) to M/S Amala Ecoclean Pvt Ltd. for commercial production and marketing.
- Cadalmin™ BSF Pro and inauguration of the Black soldier fly larvae based bioconversion unit
- Novel fish feed Cadalmin™Microfin for Mariculture fish species
- Application of nutrigenomics for improved growth and immune functions of larval stages in targeted marine fin fishes to precisely predict the critical nutritional requirements of cobia and Silver pompano larvae.
- Development of suitable feeds/feed supplements/ feeding strategies through the identified nutritional gaps for enhancing the growth, survival and quality of targeted fish species.
- Cadalmin ™ Silvergrow, a novel Silver Pompano grow out feed was developed and commercialized.

- The full-length transcriptome profiling of Silver Pompano, *Trachinotus blochii* (Lacepède, 1801) and Cobia (*Rachycentron canadum*) were carried out to aid in nutritional studies.
- Indigenous Re-circulatory Aquaculture System (i-RAS) developed for fish nutrition research.
- CMFRI's Cadalmin$^{TM}$Varsha series of freshwater ornamental fish feeds were trialled in commercial extruders and refined for commercial production.
- Pilot scale black soldier fly (BSF) rearing and research facility was established. Replacing dietary fish meal by using black soldier fly larval meal (BSFLM) was studied for silver pompano, *blochii*, and revealed that up to 90% of the fish meal can be replaced using BSFLM without adverse effects on the growth and feed utilization.
- A protocol was developed to produce cotton seed protein concentrate with yield of 24-31% of the raw material.

Marine Bioprospecting

- Developed bioactive compounds from marine organisms, including seaweeds, microalgae, mollusks, microorganisms, sponges, echinoderms, etc. for use against various lifestyle diseases.
- Developed and commercialized nutraceuticals fr use against various lifestyle and metabolic disorders, including Cadalmin$^{TM}$Antidiabetic extract for use against type-2 disbetes (ADe-marketed as Algamin Diabet-ease by Pioneer Pharmaceuticals), Cadalmin$^{TM}$LivCure extract for use against non-alcoholic fatty liver disease/NFALD (marketed as Green Rex by Eminiotech), Cadalmin$^{TM}$ Green Algal extract (GAe-marketed as Algamin Arthrit-ease by Pioneer Pharmaceuticals), Cadalmin$^{TM}$Green-Mussel extract for use against arthritis (marketed as AFDC Musseltone by AFDC Ltd), Anti-hypothyroidism-extract for use against hypothyroid disorders (marketed by VLCC), Cadalmin$^{TM}$ Antiosteoporotic (AOe) and Immuno-boost-extract (IBe) marketed by Chazah (Pluviago) Pharmaceuticals, Cadalmin$^{TM}$ Antihypertensive-extract for use against hypertension (AHe-marketed as Algamin Tension-ease by Pioneer Pharmaceuticals), Cadalmin$^{TM}$Immunalgin extract (IMe) to improve immunity against delta variant of SARS

CoV-2 virus and Antihypercholesterolemic-extract (ACe) for use against dyslipidemia to Pioneer Pharmaceuticals.

- Developed Beta-Chitosan from cuttlebone of cephalopods with bioactive potential against hydroxymethyl glutaryl coenzyme-A reductase, dipeptidyl peptidase-4, angiotensin-2 converting enzyme, and 5-lipooxygenase and a green method of pigment extraction from shrimp shell wastes
- Developed database of the nutritional qualities of marine finfish, mollusks and custaceans in various seasons and years and development of database of the maternal fish consumption and prevention of low birth weight in Kerala, West Bengal and Lakswadweep Island

In-vitro cell culture technology

- Development of a model iridescent prototype of lustrous marine pearl through *in-vitro* mantle epithelial cell culture technology.
- Development of a cost effective medium for the growth and proliferation of *in-vitro* cultured mantle epithelial cells.
- Development of a semi-solid substrate for the biomineralization of *in-vitro* cultured mantle epithelial cells onto bead nucleus.

Fish Health

- Developed disease diagnostic kits {Rapid Duplex PCR Kit of CMFRI for Early Detection of White Spot Syndrome Virus (WSSV) of Shrimp; Betanodadetect-kit and method for detecting beta noda virus infection; Perkdetect - a LAMP based diagnostic for Perkinsus sp; LAMP-based diagnostic against WSSV, and a recombinant vaccine against Vibrio alginolyticus; Multiplex-PCR technology for the simultaneous detection of major aquaculture pathogens) besides investigation on the virulence factors and genetic profiling of *parahaemolyticus* isolated from bivalves, and abundance and AMR profiling of bivalve-associated *V. parahemolyticus* and other pathogens.
- Unravelling larval microbiome and immunome of targeted fin fishes for improved mariculture production to decipher the microbiome and immunome profiles during early

life stages of the most commercially important marine fish species of India under Dr. EG Silas Centre of Excellence and Innovations in Marine Fish Microbiome and Nutrigenomics.

- Documented and characterized several new species of myxosporean parasites and other parasites and pathogens of mariculture significance.
- Technology for degradation of unprocessed marine crustacean shell wastes using microbes
- Development of better microbial management and immunomodulation strategies/protocols to address the poor growth and survival rates during the larval stages of the fish, in vitro efficacy characteristics and genetic determinants of resistance of marine/estuarine aquaculture pathogens against aquaculture antimicrobials.
- Screening of wild and farmed bivalves for the OIE listed pathogens, and efficacy evaluation of antibiotics against various pathogens in cultured fish to document the susceptibility to antibiotics.

**Advanced biotechnological tools in fisheries research**

Advanced biotechnological tools hold transformative potential for sustainable utilization of marine resources. These tools enable biodiversity documentation, selective breeding, and nutrigenomic interventions, significantly enhancing mariculture productivity. Genomic technologies contribute to pollution biomonitoring, species traceability, and stock management, while cutting-edge advancements like CRISPR-Cas9 and in vitro fish meat production promise revolutionary impacts. Marine bioprospecting targets bioactive compounds for pharmaceuticals, nutraceuticals, and renewable biofuels, with value-added products derived from algal co-products. Despite covering 70% of Earth's surface, marine biotechnology is still underdeveloped, hindered by high costs and ecological risks from overexploitation. This article explores progress across diverse marine biotechnology domains, including transgenics, bioremediation, metagenomics, and marine biomaterials.

**Bioprocess technologies**

Bioprocess technologies like enzyme-assisted hydrolysis have revolutionized the extraction of valuable marine constituents, utilizing up to 50% of by-products generated annually from 140 million tons of fish and shellfish. These advancements allow the identification of survival-stimulating compounds in marine environments. Environmentally friendly extraction methods,

such as supercritical fluid extraction (SFE) and subcritical water extraction (SWE), have emerged as sustainable alternatives to chemical extraction, particularly in fish oil production.

**Marine metagenomic and transgenic approaches**

Recombinant DNA technologies, including metabolic engineering and transgenic algae development, enable the expression of biosynthetic enzymes in alternative hosts. These techniques improve yields and novel metabolite production by regulating enzyme specificity. Nuclear transformation methods for red, green, and brown algae have shown promise for enhancing algal biomass efficiency.

**Industrial applications of marine bioresources**

Marine phytobiomass offers potential in agriculture as fertilizers, feed, and biostimulants. Seaweed and seagrass extracts have long been valued for improving soil fertility and plant growth, while marine protein hydrolysates provide stimulatory effects. Functional compounds from marine organisms, such as antioxidants, pigments, and prebiotics, are extensively utilized in food, pharmaceutical, and cosmetic industries. Marine-derived long-chain omega-3 fatty acids, chitin, and glucosamine demonstrate significant biomedical and industrial applications.

**Bioremediation**

Marine microbes play a pivotal role in bioremediation, tackling pollutants like oil spills, toxic chemicals, and sewage. Innovations include genetically engineered bacterial consortia for oil degradation and plasmid-aided breeding for enhanced pollutant breakdown. Transcriptome analyses aid in developing biosensors and biomarkers, while recombinant technologies facilitate heavy metal decontamination in marine environments.

**Transgenic fish and mariculture**

Transgenic fish offer enhanced growth rates, feed efficiency, and disease resistance. Techniques such as antisense and ribozyme technologies have improved resistance to viral infections like Hematopoietic Necrosis Virus (HNV). Gene manipulation enables faster growth, cold tolerance, and production of biomedical agents. Advanced molecular biology sheds light on disease mechanisms, enabling the development of robust mariculture stocks. Transgenic techniques have enhanced fish growth rates, cold tolerance, and disease resistance. For

instance, AquAdvantage™ salmon, the first genetically engineered fish approved for commercial use, demonstrates the potential of transgenics in sustainable mariculture.

**Marine biotechnology in renewable energy**

Marine microalgae, with their high biomass and rapid growth, are being developed for biodiesel production. These biofuels offer environmental benefits, including low emissions and biodegradability, making them a promising alternative to conventional fuels.

**In vitro fish meat production**

In vitro fish meat, a product of cellular mariculture, presents a sustainable alternative to traditional aquaculture. This process involves culturing fish muscle tissues under controlled conditions, providing eco-friendly and ethical solutions to meet growing protein demands.

**Marine nutraceuticals**

Compounds like fucoidan, carrageenan, and marine antioxidants show promise in treating chronic diseases. For example, ICAR-CMFRI's contributions to nutraceutical development underscore the immense therapeutic potential of marine resources. Seaweeds are gaining recognition for their health benefits, including anti-inflammatory, antimicrobial, and antioxidant properties. Polysaccharides such as agar, alginate, and carrageenan form the backbone of the phycocolloid industry and have diverse biomedical and nutraceutical applications. From traditional uses like cod liver oil to modern marine-derived pharmaceuticals, the ocean remains a vital source of therapeutic compounds. Compounds like fucoidan, carrageenan, and marine antioxidants show promise in treating chronic diseases. For example, ICAR-CMFRI's contributions to nutraceutical development underscore the immense therapeutic potential of marine resources.

M**arine biotechnological conservation**

Genomic tools are crucial for conserving marine biodiversity, characterizing threatened species, and enhancing stock assessments. Cryopreservation techniques and next-generation sequencing support sustainable mariculture practices and germplasm preservation.

## Conclusions

Advanced marine biotechnological tools unlock the vast potential of oceanic resources, aligning with sustainable development goals. Genomic, bioprospecting, and conservation methods enhance productivity, environmental monitoring, and functional food development. Despite challenges in cultivation and commercialization, continued research will bridge the gap between marine biotechnological discoveries and their applications, ultimately benefiting humanity.

# Harnessing Biotechnological Tools for Marine Fisheries Resource Management and Sustainable Aquaculture

Sandhya Sukumaran

Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018

sandhya.sukumaran@icar.org.in

## Introduction

The field of biotechnology has made considerable progress over the last few decades, and the potential applications of biotechnology tools in the fisheries and aquaculture sector are huge. The ocean's wealth must be preserved for long-term sustainability, and biotechnology offers promising opportunities for sustainable exploitation of marine resources. Fish are a vital source of protein for millions of people, making responsible fisheries management an absolute necessity to prevent a progressive loss of genetic diversity. Aquaculture plays an equally important role in national fish production, and the contribution will be strengthened by innovations in biotechnology for culture fisheries. Recognised as the fastest growing food sector in the world, aquaculture faces challenges which can be addressed by introducing biotechnology tools.

## Genetic And Genomic Techniques for Ensuring the Sustainability of Capture Fisheries

### Genetic stock structure assessment

The information regarding the stock structure of fish is crucial for the development of fisheries management plans that support the sustainable management of fish resources. Stocks are defined as temporally or spatially separated units with specific biological properties. When a population contains different sub-populations or populations, they shall be managed as separate entities. The sustainability of these stocks should be assessed on a case-by-case basis to develop effective strategies for sustainable fisheries management. In cases of overfishing, it is important that fishing strategies are carefully evaluated and the necessary regulatory measures are enforced. Genetic and genomic technologies are increasingly relied upon to provide insights into the structure and boundaries of populations. Traditional morphological methods are often cumbersome and may be imprecise.

Mitochondrial genes, including cytochrome c oxidase 1, the control region (D-loop), cytochrome b, ATPase 6/8, and NADH genes, have been widely employed to gain insights into stock structure. The mitochondrial genome is characterized as haploid, double-stranded, and maternally inherited, which leads to a reduction in the effective population size of mitochondrial DNA. A decrease in effective population size enhances genetic differentiation between gene pools, rendering mitochondrial DNA an appealing marker for investigating subpopulation structure. Effective amplification can be performed with a small quantity of tissue, as it is present in numerous copies within the cell, depending on the type of cell. Nuclear DNA markers, such as RAPD or Random Amplified Polymorphic DNA, utilize a random primer for the amplification of genomic DNA, producing multiple bands that can be compared across different populations. However, due to the absence of repeatability and reproducibility, this method is no longer utilized. Microsatellite markers, which consist of regions in the genome with repeat units of 1-6 base pairs in length, have become the preferred choice for population genetic studies because of their abundance in the genome, ease of detection and amplification, and specificity. Their co-dominant nature allows for Mendelian inheritance, enabling effective differentiation between homozygous and heterozygous individuals. They also display a higher evolutionary rate, making them suitable for detecting low genetic stock structure, and their non-coding nature guarantees that variations are independent of natural selection. Single Nucleotide Polymorphisms (SNPs) are point mutations that are found in abundance throughout the genome. SNPs are now the preferred markers for many studies aimed at identifying population-level variations.

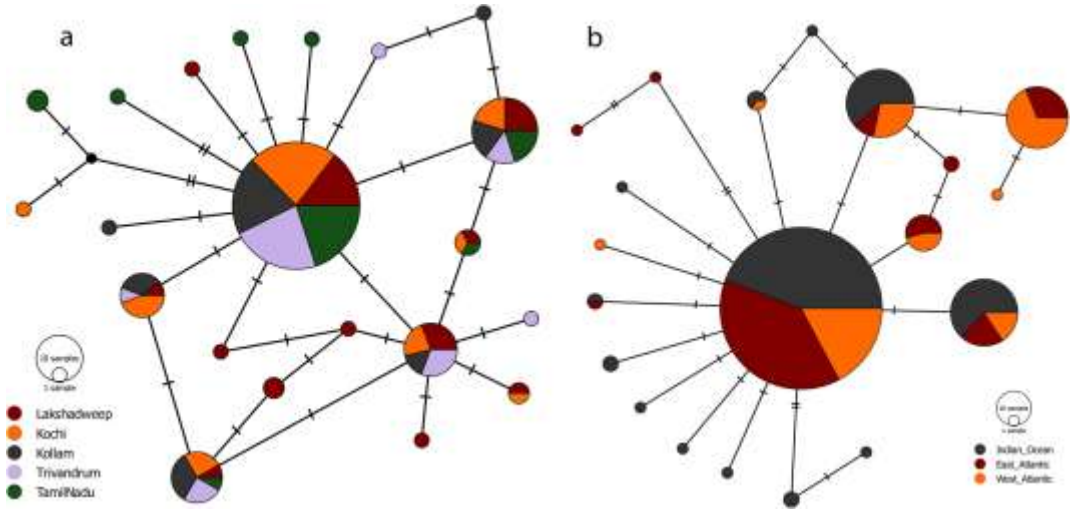



**Figure 1:** Haplotype network diagram constructed using control region sequences of *Carcharhinulongimanus*

**Molecular taxonomy & DNA Barcoding**

Tropical seas are abundant in biodiversity, and it is imperative to protect this diversity. The first step in ensuring effective protection is the documentation of this diversity. Moreover, information on the spatio-temporal distribution patterns of marine biodiversity and community structure is essential for conservation efforts. Accurate cataloguing of threatened or endangered species should also be conducted, for which molecular techniques such as barcoding can be utilized. Molecular tools can effectively identify and characterize various strains, stocks, and hybrids.

The Mitochondrial Cytochrome C Oxidase 1 gene is recognized as the universal barcode for species identification, and this approach has become increasingly popular. The sequences of thousands of organisms are available through the NCBI and GenBank data repositories, enabling users to verify the identity of specific species sequences via NCBI's BLAST search.

Another online platform that facilitates sequence sharing among researchers is the Barcode of Life Data System (BoLD). The Central Marine Fisheries Research Institute (CMFRI) has compiled a barcode database for many significant fin and shellfish species, such as tunas, whale sharks, sardines, oysters, mussels, and cuttlefishes. DNA barcoding is also instrumental in diagnosing the presence of invasive species and their proliferation within ecosystems, which aids in effective quarantine and eradication efforts. Moreover, DNA barcodes can be effectively employed to identify predator-prey interactions by monitoring prey species, recognizing the presence of pest or pathogen species in the ecosystem, creating species-specific markers, and conducting various related research initiatives. The identification of eggs and larvae from numerous species can be achieved in advance through the use of DNA barcodes. In the context of mussel mariculture, molecular markers can be utilized to determine the timing and location of spat fall.

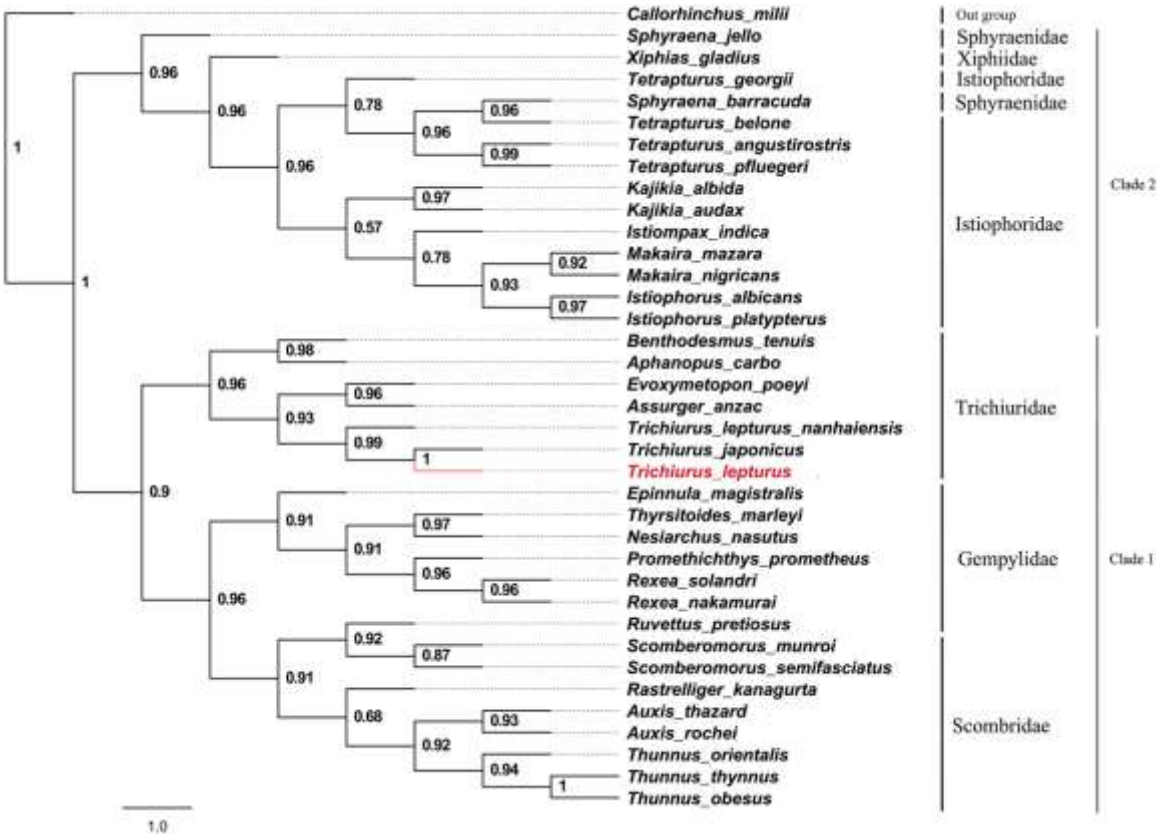



**Figure 2**: A phylogenetic tree constructed using mitogenomic sequences in *Trichiurus lepturus* showing the relationships.

## Aquaculture Biotechnology

Aquaculture is of paramount importance for supplementing marine fishery production, and the long-term sustainability of fishery management can only be assured through aquaculture efforts. The marine fisheries sector is encountering various challenges due to climate change and overfishing, necessitating the enhancement of aquaculture productivity through advanced genomic tools. Aquaculture activities must be conducted sustainably, maintaining the fragile equilibrium of the ecosystem. The aquaculture sector's efficiency can be improved by utilizing superior germplasm resources, with aquaculture biotechnology playing a significant role in this regard. Sustainable biotechnological innovations can be employed for induced breeding, disease diagnosis and prevention, aquaculture nutrition through the use of nutrigenomic tools, and the genetic enhancement of farmed species through selective breeding.

The germ plasm's quality can be improved through the utilization of quantitative genetic methodologies. Selective breeding can generate better broodstock for various commercially vital fish species. A suitable selective breeding approach can be established with adequate knowledge of the heritability of the selected trait, the relationships between genotypic and phenotypic features, heterosis, and genotype-environment interactions. The CMFRI has created improved strains of *Artemia franciscana* with modified naupliar size based on quantitative genetics principles. At the Central Institute of Freshwater Aquaculture in Bhubaneswar, selective breeding of Rohu, *Labeo rohita* has been performed, leading to the development of the Jayanti Rohu variety, which shows enhanced growth rates. In India, mariculture is on the rise, particularly with species like Cobia, *Rachycentron canadum*, and Pompano, *Trachinotus blochii*. To achieve sustainable production, the next phase could involve selective breeding and genetic advancements.

The enhancement of crops or livestock can be accomplished by selecting suitable inbreeding or crossbreeding strategies. By merging inbreeding with crossbreeding, it is feasible to generate superior offspring that display hybrid vigor. A variety of inbreeding and crossbreeding experiments can yield hybrid vigor. Chromosomal engineering can be effectively applied to produce individuals of higher quality, which exhibit increased growth and reproductive success. Techniques such as androgenesis, gynogenesis, and ploidy manipulation have been utilized to create individuals with enhanced growth and vigor in aquaculture. In androgenesis, the reproductive process is adjusted so that only paternal genetic material is inherited. This results in the production of viable YY supermales when the male is heterogametic, and successful instances of YY supermale production have been recorded in cyprinids, cichlids, and salmonids. In gynogenesis, progeny inherit only maternal genetic material, enabling the formation of all-female populations. The use of androgenesis or gynogenesis is beneficial when either males or females exhibit superior growth rates and performance.

Ploidy manipulation is an alternative approach to enhance growth rates, utilizing techniques such as triploidy or polyploidy, which are particularly useful in shellfish due to their ease of maintenance. Successful induction of triploidy has been documented in species such as *C. gigas*, *C. virginica*, *Saccostrea glomerata*, and *Ostrea edulis*. The generation of all-male or all-female populations can be achieved through hormonal manipulation of sex and reproduction. Hormonal analogues that are synthetically produced are available in the market and can be

employed for this manipulation. Induced breeding of freshwater carps has been conducted by introducing a hormonal extract containing GnRH, which is a critical regulator and initiator of the reproductive cascade in vertebrates. Currently, synthetic hormones like ovaprim and ovatide are available for use in induced breeding. Cryopreservation of gametes and embryos in cold environments is another significant intervention in aquaculture, aimed at overcoming seasonal barriers to reproduction. This technique offers a superior method for preserving fish gametes, which can be utilized for controlled reproduction, thereby reducing reliance on wild-collected seeds. Cryopreservation is also beneficial for sequential hermaphrodites such as seabass or grouper (protandrous/protogynous), as sourcing males or females from the wild for controlled hatchery production is quite difficult. While sperm cryopreservation is standardized and widely practiced in fish, the cryopreservation of ova or embryos has yet to be standardized due to various inherent issues associated with fish eggs and embryos.

Superior genetic stock can be produced with the help of genetic engineering tools. Transgenic tools can be efficiently employed for the production of fish with a faster growth rate, improve environmental tolerance and disease resistance. Antifreeze protein genes, growth hormone genes, or fluorescent protein genes can be inserted into the genome of the desired species of fish, which subsequently get expressed in the progeny. Preliminary success has been reported in India in developing gene transfer technology in zebra fish, medaka, and Indian catfish. Production of genetically modified zebra fish (Glofish), which can produce fluorescent pigments red, green, and yellow, has been successful, and it is a very popular household aquarium pet.

In marker assisted selection (MAS) technique, prospective breeders are chosen based on genotypes with the help of molecular markers. Marker assisted selection programmes can be carried out using molecular markers like allozymes, RFLP, RAPD, AFLP, microsatellite, SNPs, ESTs and mitochondrial DNA. Marker assisted selection is useful for identification of genetic releatednes, diversity, determination of pedigree, genetic tagging, tracking of family and population lines and identification of strains. Identifying markers linked to quantitative trait loci or QTL also could be done with Marker Assisted Selection techniques.

**Feed biotechnology and nutrition**

Appropriate feed is essential to the success of any aquaculture endeavor, so feed biotechnology and nutrition are crucial. In nutritional research, successful biotechnological interventions can

enhance the performance of candidate fishes. Enzymes can be added to formulated feeds to increase the availability of nutrients. These enzymes should have a long shelf life and be able to tolerate changes in physico-chemical conditions, such as elevated temperatures. When added to feeds, the enzyme phytase aids in the breakdown of indigestible phytic acid in plant-based nutrient sources, facilitating the release of digestible phosphorus.

Fish that can be raised can have their disease resistance increased by adding probiotic bacteria to their diet. By competitively excluding harmful bacteria, probiotics, which are live microorganisms, can be added to diets and provide the host with some sort of health benefit. Enzymes that speed up food digestion can also be released by these bacteria. *Saccharomyces cerevisiae*, *Bifidobacterium bifidium*, *Lactobacillus acidophilus*, *Lactobacillus bulgaricus*, *Lactobacillus planetarium*, and *Aspergillus oryzae* are frequently added to probiotic products.

Prebiotics allow the growth of gut microflora by providing food for probiotic organisms that are resistant to attack by endogenous enzymes. Prebiotics have a long shelf life and can tolerate high pelletizing temperatures in the feed. Using genetically modified microorganisms to supplement dietary amino acids can also improve the quality of feed. Adding vital amino acids like lysine and methionine to feed can also enhance its quality. Another recent advancement in feed biotechnology is the use of nucleotides as feed additives to boost the expression of desired characteristics like growth or disease resistance. By examining how nutrition affects an organism at the molecular level, functional genomics principles can be applied to nutrition research.

**Fish Health**

To guarantee the sustainability and financial success of aquaculture endeavors, health management is crucial. Traditional illness diagnosis techniques take a long time and have poor sensitivity, speed, and specificity. Making informed judgments on the best course of treatment depends heavily on accurate and effective illness diagnosis. Finfish and shellfish aquaculture are susceptible to a variety of bacterial and fungal illnesses, with white spot virus disease of shrimp (*P. monodon*) being one of the most dangerous. Research organizations such as CIBA and CMFRI have created white spot virus early detection and prevention tools.

Another biotechnological intervention that will aid in managing and preventing disease is the injection of DNA vaccines. DNA vaccines are made up of an infectious organism's DNA that is injected into a host and subsequently expressed in the host.

By introducing lytic bacteriophages, pathogenic bacterial infections can be treated by phage therapy. Phage viruses help eliminate bacteria, and they must not come into contact with other helpful bacteria or the surrounding tissue. Since the virus can spread quickly, it is best to provide a single, little dosage. In aquatic environments, phage treatment has not yet gained traction.

RNA-guided control of gene expression patterns in eukaryotic cells is known as RNA interference (RNAi). The expression of genes with complementary sequences to short chains of double-stranded ribonucleic acid (dsRNA) in the cell may be disrupted. RNA interference is a form of post-transcriptional gene silencing in which dsRNA attaches to a particular mRNA, causing the homologous endogenous transcript to degrade and lowering gene activity. For the treatment of viral illnesses under culture settings, RNA interference (RNAi) is a potentially useful and significant technology.

Recombinant technology is useful for creating genetically modified creatures with different genetic makeup. To create transgenic fish that are resistant to illness, several studies are being conducted.

**Marine Bioprospecting**

Investigating novel sources of chemicals, microbes, genes, and other useful marine products is known as marine bioprospecting. Utilizing biotechnological technologies, these biological resources may be sustainably harnessed to support local communities' socioeconomic growth. Pharmaceuticals, cryoprotectants, cosmaceuticals, and neutraceuticals may be derived from marine creatures. It is possible to extract a variety of new medications from marine creatures that might replace antibiotics. Marine microorganisms and invertebrates can produce secondary metabolites that have anti-inflammatory and anti-cancer properties.

**Contributions of ICAR-CMFRI To Marine Biotechnological Research**

The utilization of biotechnology within the fisheries and aquaculture sectors holds significant promise and is currently in a developmental stage in India. Research conducted at CMFRI has

greatly advanced the application of biotechnology in the management of fisheries resources and aquaculture practices. CMFRI has undertaken groundbreaking research to comprehend the genetic variability and stock structure of various species, including the Indian oil sardine *Sardinella longiceps*, Indian mackerel *Rastrelliger kanagurta*, and Indian anchovy *Stolephorus indicus*, Scalloped hammerhead shark, *Sphyrna lewini,* oceanic whitetip shark, *Carcharhinus longimanus*, spade nose shark, *Scoliodon laticaudus*, Indian Scad *Decapterus russelli* and Randall's threadfin bream, *Nemipterus randalli* utilizing mitochondrial and microsatellite markers. The complete mitogenome of several fish species, such as the Indian oil sardine *Sardinella longiceps*, Goldstripe sardine *Sardinella gibbosa*, pearl spot *Etroplus suratensis*, ribbon fish *Trichiurus lepturus*, Indian Scad *Decapterus russelli* and Randall's threadfin bream, *Nemipterus randalli* has been thoroughly characterized. Research into the climatic adaptations of the Indian oil sardine has been conducted through whole mitogenome scans, which have identified the existence of locally adapted populations. Ongoing research aims to further explore genomic adaptations to climate change in the Indian oil sardine (*Sardinella longiceps*). Biotechnological advancements in aquaculture have resulted in the creation of formulated feeds, such as Varna, for the rearing of ornamental fish, as well as the commercialization of various nutraceuticals derived from marine organisms. CMFRI has developed several diagnostic kits to identify various diseases affecting farmed marine finfish and shellfish. Triploid oysters represent another innovative biotechnological achievement by CMFRI. Additionally, numerous probiotics have been developed by CMFRI.

**Conclusion**

India, with a coastline of about 7500km, an Exclusive Economic Zone of about 2 million square kilometers, and vast areas of fresh and brackishwater resources, has immense potential to harness the valuable aquatic resources for a sustainable blue economy. The blue economy aims to tap ocean resources sustainably, and biotechnological innovations play a major role in contributing to the blue economy. Biotechnology is a multidisciplinary science that requires integration of biological, chemical, engineering, and material sciences. A holistic approach involving the government and private sector for the application of biotechnological tools in emerging areas of fisheries science will bring about rapid advancements which will be beneficial for the society as a whole.

**Key References**

Askari, Gh., Shabani, A., Miandare, H.K. 2013. Application of molecular markers in fisheries and aquaculture. Scientific Journal of Animal Science 2(4): 82-88.

Davis, G.P. and Hetzel, D.J.S.2000. Integrating molecular genetic technology with traditional approaches for genetic improvement in aquaculture species. *Aquaculture Research*, 31, 3-10.

Hallerman, E.M. 2006. Use of molecular tools for research and improvement of aquaculture stocks. *The Israeli Journal of Aquaculture – Bamidgeh* 58(4): 286-296.

Hernandez-Urcera, J., Vera, M., Magadan, S., Pino-Querido, A., Cal, R.M. and Martinez, P. 2012. Development and validation of a molecular tool for assessing triploidy in turbot (*Scophthalmus maximus*). Aquaculture 330-333: 179-184.

Ken Overturf. 2009. Molecular research in Aquaculture. Wiley-Blackwell .

Okumus, I. and Ciftci, Y. 2003. Fish population genetics and molecular markers:II-Molecular markers and their applications in fisheries and aquaculture. *Turkish Journal of Fisheries and Aquaculture* 3: 51-79.

Presti, R.L., Lisa, C. and Di Stasio, L.2009. Molecular genetics in aquaculture. *Italian Journal of Animal Sciences*. 8: 299-313.

Zheng, J.Y., Zhuang, W., Yi, Y.T., Wu, G., Gong, J. and Shao, H.B. 2013. Developmentally utilizing molecular biological techniques into aquaculture. *Reviews in fisheries science*. http://dx.doi.org./10.1080/10641260903477499.

# Mariculture in India: challenges and prospects

Dr. V. V. R. Suresh*
Mariculture Division, ICAR-CMFRI, Kochi- 682 018, India,
*sureshvr5644@gmail.com

By 2050, the global population is projected to reach 9.7 billion, necessitating a 50% increase in food production to ensure food security (FAO, 2017). Meeting this ambitious target presents formidable challenges, particularly amidst intensifying constraints on land and freshwater resources, the pervasive impacts of climate change, and geopolitical uncertainties. Against this backdrop, aquaculture has emerged as the most rapidly expanding sector in global food production, achieving an impressive annual growth rate of over 6% over the past two decades. Within aquaculture, mariculture—farming marine organisms—has established itself as the fastest-growing subsector with remarkable growth potential. In 2020, mariculture contributed approximately 33.0 million tonnes of food fish, accounting for 27% of the global aquaculture production of food fish. Including seaweed production, the total mariculture output surged to 68.1 million tonnes, constituting a substantial 55.6% of the world's aquaculture production that year. This underscores mariculture's pivotal role in addressing global food and nutritional security, particularly as marine resources are increasingly harnessed for sustainable production. With its burgeoning seafood demand, India is uniquely poised to capitalize on the growth potential of mariculture. Reliance solely on capture fisheries is insufficient to meet the rising domestic and international demand, highlighting the need for concerted efforts to expand mariculture. The National Policy on Marine Fisheries (NPMF, 2017) emphasizes the critical role of mariculture in augmenting fish production from coastal and offshore waters. In response, the Indian government is advancing a multi-faceted strategy to address the sector's institutional, technological, and commercial needs. As the global spotlight on blue economy initiatives intensifies, India's commitment to scaling up mariculture aligns with broader efforts to ensure sustainable ocean-based development, enhance livelihoods, and achieve long-term food security and blue economy aspirations.

**Mariculture genesis in India**

Mariculture, as defined by the FAO (1997), involves cultivating marine organisms in seawater environments such as ocean enclosures or seawater-filled tanks and ponds. It plays a critical role in supplementing land-based economies, conserving marine capture fisheries, and bolstering global food and nutrient supplies (Schubel and Thompson, 2019). In India, the rising seafood demand cannot be met solely through capture fisheries. Mariculture development in India began in the 1970s with ICAR-CMFRI's initiatives in Mandapam and Tuticorin, focusing on seaweed and bivalve culture, followed by advancements in shrimp farming. However Focused mariculture activities began in 2008, leveraging the available area for mariculture, including 19 million hectares of territorial waters and 202 million hectares of Exclusive Economic Zone (EEZ). Indian mariculture has an estimated production potential of ≈5 million tonnes per annum, against the current annual production of 0.124 million tonnes. including 0.012 million tons of food fish, 0.04 million tons of shellfish, 0.072 million tons of seaweed, and 20 species of ornamental fish (Gopalakrishnan et al., 2023).

The sector has expanded to include the cultivation of finfishes (e.g., cobia, pompano, sea bass), shellfish (e.g., mussels, oysters, shrimp), ornamental fishes, marine pearls, and seaweed. These technologies hold the key to unlocking the immense potential of India's 8,118 km coastline and 202 million hectares of EEZ. Gopalakrishnan et al. (2023), have highlighted vast untapped opportunities for blue economic growth through Mariculture in India. Harnessing this potential requires a strategic mission addressing species prioritization, hatchery technology development, grow-out scaling, site identification, and enabling policies. A comprehensive approach is essential to sustainably grow mariculture and advance India's blue economy.

**Prioritization of mariculture species**

The strategic selection of target species for mariculture and the development of hatchery technologies represent pivotal elements in advancing the blue economy. This selection process necessitated a comprehensive evaluation based on factors such as market demand, ecological sustainability, biological attributes, and economic viability. In 2017, a meticulous initiative led to the thoughtful prioritization of 76 diverse species for mariculture development. This selection comprised 23 finfish species, 7 molluscs, 6 crustaceans, and 31 ornamental species, supplemented by 4 region-specific species and 5 species of conservation interest (Ranjan et al., 2017). Such strategic prioritization underscored a nuanced approach aimed at fostering sustainable and diversified mariculture practices.

**Captive breeding and seed production**

The standardization and demonstration of breeding and seed production technologies have significantly advanced the utilization of India's marine resources. These efforts span finfish, shellfish, and seaweed, focusing on optimizing environmental conditions for enhanced reproduction and genetic diversity, as well as ensuring healthy juvenile cultivation for grow-out facilities. These innovations promote sustainable marine resource management and the commercialization of mariculture, bolstering India's blue economy.

Captive breeding technologies are established for over 27 species, including finfish, molluscs, shellfish, and live-feed organisms. Initiatives such as hatcheries, seed banks, rearing units, and SPF/SPR/genetically improved brood banks are underway, alongside a planned seed certification system to ensure quality. CMFRI has set up a National Brood-bank Facility for Cobia and Silver Pompano and leads the All India Network Projects on Mariculture and Ornamental Fishes to address sectoral challenges. Year-round seed production is now possible for seven marine finfish species—Cobia, Silver pompano, Indian pompano, Orange-spotted grouper, Pink ear sea bream, John's snapper, and Vermiculated spinefoot—as well as 20 high-value ornamental species, including five crossbreeds.

**Mariculture systems**

Prominent mariculture systems include sea cage farming, Integrated Multitrophic Aquaculture (IMTA), Recirculating Aquaculture System (RAS), culture of seaweeds, and bivalve farming. These methods emphasize sustainable cultivation, efficient resource use, and environmental conservation, showcasing a dynamic approach to the growth of the mariculture sector.

**Sea Cage Farming**

Sea cage farming, introduced in India in 2007, marks a significant advance in intensive finfish production. Notable progress has been made in cage design, mooring systems, and species selection. Collaborative initiatives with Fisher Cooperatives have popularized this practice. ICAR-CMFRI has developed comprehensive guidelines and best practices, including *Good Aquaculture Practices for Sea Cage Farming* (Sekar et al., 2022), marine finfish hatchery protocols (Ranjan et al., 2022), and farming practices for mussels (Mohamed et al., 2019) and seaweed (Johnson et al., 2023). Indigenous 6-meter diameter GI and HDPE cages introduced by CMFRI, yielding 2–3 tonnes of fish per cycle, have been widely adopted. Sea cage farming is thriving in Maharashtra, Tamil Nadu, Kerala, Karnataka, and Odisha, revolutionizing coastal aquaculture (Johnson et al., 2023).

**Forming of Bivalves.**

Mussel cultivation, particularly of *Perna viridis*, has expanded across Kerala, Karnataka, Goa, and Maharashtra, leveraging techniques like rack, raft, and long-line culture. North Kerala's Padanna estuary, involving 2,000 farmers, contributes 75% of India's green mussel production. With an annual output of 10,000 tonnes, farming benefits nearly 6,000 women-led self-help groups (Gopalakrishnan et al., 2019).

**Forming of Edible oysters.**

The farming of edible oysters (*Crassostrea madrasensis*) has advanced through technological developments in captive breeding and seed production. Thriving in shallow estuaries, bays, and backwaters, oyster farming promotes sustainability and enhances coastal livelihoods, becoming a vital part of India's mariculture sector.

**Lobster Farming**

Low-intensity lobster farming, focusing on *Panulirus polyphagus* and *P. homarus homarus*, shows potential. With a 4-month culture period, lobsters grow from under 50g to around 200g, producing up to 300 kg per 6-meter cage. Sustainable practices ensure optimized production, supporting this emerging sector.

**Sea weed Farming**

Seaweed farming provides sustainable livelihood opportunities with low capital requirements. SHGs in Tamil Nadu employ 3.7 x 3.7 m rafts for 45-day cycles, producing 240 kg per cycle. Despite industrial demand of 0.26 million tonnes (wet weight), current production stands at 0.072 million tonnes in 2023, against an estimated potential of 9.58 million tonnes. CMFRI's efforts to enhance micropropagation techniques, processing, and marketing aim to bridge this gap.

**Challenges**

India's mariculture sector faces several critical challenges impeding its growth. A major obstacle is the policy vacuum, with unclear and fragmented regulatory frameworks causing delays, licensing issues, and inefficient resource allocation. The absence of designated mariculture zones exacerbates conflicts over land and nearshore water use. State-level policies with clear guidelines for zoning, leasing, and financial incentives are crucial for sustainable growth.

Technological adoption remains limited, with traditional practices dominating over advanced systems like offshore cages and automated monitoring, leading to suboptimal productivity.

Inadequate availability of high-quality seeds and broodstock further hinders consistent production and strains natural populations. Coastal land competition from industrial activities, urbanization, and tourism, coupled with pollution, poses environmental challenges.

Social and economic barriers, including limited training for coastal communities and restricted access to credit and insurance for small-scale farmers, worsen the stagnation. Climate change compounds these issues with rising sea temperatures, ocean acidification, and cyclones threatening infrastructure and productivity. Addressing these hurdles is imperative to unlock the sector's potential.

**Prospects**

Despite these challenges, India possesses enormous potential to develop a thriving mariculture industry. The country’s rich marine biodiversity allows for the cultivation of a wide variety of species, ranging from finfish like cobia and seabass to shellfish such as pearl oysters and shrimp, as well as edible seaweeds. The global demand for seafood, particularly for high-value exports, positions India as a potential leader in sustainable marine farming.

Technological advancements are paving the way for the future of mariculture in India. Systems like recirculating aquaculture and integrated multitrophic aquaculture (IMTA) are promising environmentally sustainable and economically viable methods of farming. The use of digital technologies, such as IoT and AI-based monitoring systems, can further enhance operational efficiency and yield predictability.

Government initiatives, such as the Blue Revolution and the Pradhan Mantri Matsya Sampada Yojana (PMMSY), provide financial and policy support for the sector. These programs aim to create infrastructure, facilitate research, and promote capacity building among fishers and farmers. Coupled with growing consumer awareness of the nutritional benefits of seafood, these efforts can stimulate domestic demand while simultaneously catering to the burgeoning global market.

Mariculture also has significant potential as a means of generating employment and enhancing livelihoods, especially for coastal communities. Women’s participation in seaweed farming, pearl culture, and ornamental fish farming is increasing, offering new economic opportunities. By engaging local communities and leveraging traditional knowledge, the sector can foster inclusive growth and reduce dependency on capture fisheries.

**Role of genetics and genomics**

Genetics and genomics play a pivotal role in enhancing mariculture by enabling precise selective breeding for desirable traits such as growth rate and disease resistance through genetic markers. These technologies also help identify genes related to immunity, facilitating the development of disease-resistant species and reducing reliance on antibiotics. By pinpointing genes for better growth and feed conversion, genetics improves farm productivity, while also providing insights into species adaptation to various environmental conditions, supporting sustainable breeding. Additionally, genomics helps preserve genetic diversity in breeding programs, aiding stock enhancement and restoration efforts. The study of gene-nutrition interactions (nutrigenomics) allows for tailored feeding strategies that improve growth and health efficiency. Genetics also contributes to breeding for superior meat quality, enhancing the market value of farmed species. Furthermore, genomic research helps minimize environmental impacts, preventing genetic pollution and promoting ecosystem compatibility in aquaculture systems. Lastly, genomics aids in precision aquaculture, where real-time health and environmental data optimize management and sustainability.

**Way Forward**

To unlock mariculture's potential in India, a policy-driven approach is vital. A comprehensive national policy should address zoning, leasing, and environmental safeguards while simplifying regulatory frameworks to support both small- and large-scale operations. Research and development must focus on seed production, disease management, and species diversification, with collaborative efforts between institutions like CMFRI and private enterprises to drive innovation with application of genetic tools.

Infrastructure development, including hatcheries, mariculture parks, and industrial farming zones, should be prioritized. Clear policies supporting large-scale mariculture and ensuring sustainability will attract private investment. Additionally, capacity building through training programs is essential for both coastal communities and industrial farmers to adopt modern practices.

Access to credit, insurance, and fiscal incentives will support small-scale entrepreneurs and large-scale ventures. Emphasizing sustainable practices, such as Integrated Multitrophic Aquaculture (IMTA), and integrating digital tools, green technologies, and automation will enhance efficiency and reduce ecological impacts, ensuring long-term sector growth.

## Conclusion

Mariculture in India stands at a crossroads of challenges and opportunities. While issues like technology gaps, environmental conflicts, and regulatory inefficiencies persist, the sector's vast untapped potential cannot be ignored. With strategic interventions in policy, technology, and community engagement, mariculture can transform into a cornerstone of India's blue economy. It promises not just economic prosperity but also a pathway toward ecological sustainability and food security, aligning with the nation's vision for a resilient and inclusive future.

## References


Business Standard. 2023. CMFRI identifies 146 sites along the Indian coastline for sea cage farming. https://www.business-standard.com/article/economy-policy/cmfri-identifies-146-sites-along-indian-coastline-for-sea-cage-farming-123020700952_1.html

FAO. 2017. The future of food and agriculture: Trends and challenges. Food and Agriculture Organization of the United Nations, Rome.

FAO. 1997. Mariculture Development: A New Challenge. FAO Fisheries Circular No. 928 FIIU/C928. FAO., Rome. http://www.fao.org/docrep/w6602e/w6602e00.htm

Johnson, B., D. Divu, Reeta Jayasankar, L. Ranjith, Gyanaranjan Dash, et al. 2020. Preliminary estimates of potential areas for seaweed farming along the Indian coast. *Marine Fisheries Information Service Technical & Extension Series* No. 246: 14-28.

Johnson, B., Tamilmani, G., Divu, D., Mojjada, S. K., Megarajan, S. et al. 2023. *Good Management Practices in Seaweed Farming.* CMFRI Special Publication (148), 30p.

Gopalakrishnan, A., Kirubagaran, R., George John, Ponniah, A.G., Gopakumar, G., Sunilkumar Mohamed K., Krishnan, Imelda Joseph, Boby Ignatius, Abdul Nazar, A.K., Jayakumar, R., Raju, M.S., Sreepada, R.A., Shinoj, P. and Rajesh N. (2019). Draft National Mariculture Policy 2019 (NMP2019); Report of the Committee constituted by the National Fisheries Development Board (NFDB), Ministry of Fisheries, Animal Husbandry & Dairying, Govt. of India. CMFRI Marine Fisheries Policy Series No.17/2020, 22p.

Gopalakrishnan, A., Boby Ignatius and VVR Suresh. 2023. Mariculture Development in India: Status and Way Forward. *Indian J. Plant Genet. Resour.* 35(3): 317–321.

Mohamed K. S., P. K. Asokan, G. Sasikumar, P. Kaladharan, V. Kripa, et al. 2019. Guidance for Good Mussel Farming Practices in India based on a Case study from Kerala. CMFRI Mar. Fish. Policy Series 10, 64p.

NFDB. 2018. Guidelines for Sea Cage Farming in India. National Fisheries Development Board, Hyderabad. 22p.

NPMF, National Policy on Marine Fisheries (2017) Ministry of Agriculture and Farmers Welfare (Department of Animal Husbandry, Dairying and Fisheries), Government of India, The Gazette of India Notification, Extraordinary, Part 1, Section 1, No. 121, New Delhi, the 1st May, 2017, F. No. 21001/05/2014-FY (Ind) Vol. V.

Ranjan, R., Muktha, M., Ghosh, S., Gopalakrishnan, A., Gopakumar, G., Joseph, I. (Eds.). 2017. Prioritised species for mariculture India. ICAR-Central Marine Fisheries Research Institute, Kochi, 450p.

Ranjan, R., Megarajan, S., Xavier, B., Avadhanula Ravi K, Ghosh, S., Gopalakrishnan, A. 2022. Good Aquaculture Practices for Marine Finfish Hatchery. CMFRI Special publication No. 142, 34p.

Schubel, J. R., Thompson, K. 2019. Farming the Sea: The only way to meet humanity's future food needs. GeoHealth, 3, 238-244. https://doi.org/10.1029/2019GH000204

Sekar Megarajan, Ritesh Ranjan, Biji Xavier, Biswajit Dash, Shubhadeep Ghosh, A. Gopalakrishnan. 2022. Good Aquaculture Practices (GAP) in Sea cage farming of Indian pompano and Orange spotted grouper. ICAR-Central Marine Fisheries Research Institute Special Publication No. 143, 72p.

# ICAR-NBFGR in conservation of fish genetic resources

T. T. Ajith Kumar,

Centre for Peninsular Aquatic Genetic Resources, ICAR - National Bureau of Fish Genetic Resources, Kochi, Kerala – 682018, ttajith87@gmail.com

## Introduction

India possesses rich aquatic biodiversity across different ecosystems. Of the 32,042 finfish species reported globally, about 9.2% are distributed in Indian waters. Apart from finfish resources, nearly 2934 species of crustaceans, 5070 species of molluscs and 765 species of echinoderms also contribute to India's rich aquatic genetic resources. Natural aquatic resources are important as the majority of the food source is still come from the wild due to low domestication levels. Management of fisheries resources draws parallel to wildlife and forestry, besides agriculture. In addition, being a source of food, aquatic germplasm resources are also an important source of various products of commercial value and sustain other related trades such as ornamental organisms, bioactive compounds etc. The challenge is to secure the Intellectual Property Rights related to aquatic germplasm, so that the country can maintain its stake in its natural wealth and its potential benefits (ICAR-NBFGR, 2016). Cataloging and devising conservation measures of the aquatic genetic resources of the country is essential for sustainable utilization of these resources for prosperity.

## Database development

The Institute has developed databases on diverse ecosystems and a few online databases on genomic resources of the fish species in India. A database on the fish diversity of India has been developed which contains information about 2953 finfish species. The database has been developed on freshwater fishes of the northeast and western Ghats, and fish diversity checklists for eight states and three ecosystems (Western Ghats, Gulf of Mannar and Vembanad Lake). Four online databases, namely, Genome, FisOmics, AqGRISI and NRFC are developed for the fish germplasm resources of the country.

## Documentation and cataloging of fish genetic resources

Explorations of various river basins viz., Indus, Ganga, Ghagghar, Brahmaputra, Mahanadi, Krishna, Godavari and their tributaries in addition to the rivers in the Peninsular region were

carried out and accessions were collected for taxonomic validation, stock identification, and repository. Tissue bank has been established with over 13,000 organisms. Discovered more than fifty new species during explorations of ecologically diverse habitats in India during the last two decades, in collaboration with other partner organizations.

**Generation of genomic resources**

Species-specific molecular signatures of over 600 Indian marine and freshwater fin and shellfish species were generated to resolve taxonomic ambiguity and for accurate documentation of species diversity. Population genetic structure of 26 fin and shellfish species studied across the native distribution range in Indian waters, which would help in management strategies, stock-specific conservation, and river ranching programs.

Under the CRP Genomics, the institute is a lead centre, a high-quality draft genome assembly of *Tenualos ilisha* has been generated and seven novel gene variants are identified. Differential expressing genes in various salinity levels were studied and identified with hub genes, which are highly associated with salinity tolerance with osmoregulatory function in the high-value fish species.

Draft whole-genome sequencing of two commercially important fish species, *Labeo rohita,* and *Clarias magur* were completed in collaborative mode. Complete mitochondrial DNA has been sequenced in eight fish species for ascertaining the evolutionary significance of the species.

**Establishment of fish germplasm resource centers for conservation and sustainable utilization**

**Marine Resources**

An initiative of ICAR-NBFGR at Agatti Island of Lakshadweep on ornamental invertebrates, breeding and dissemination through community aquaculture for improving the income of native women inhabitants. A facility has been established for marine ornamentals at Agatti island, Lakshadweep, which is hand-holding community aquaculture units, maintained by the women of Agatti Island to raise marine ornamentals to marketable size for the generation of alternate income sources. This is a unique venture, where science and societal development is taken up hand in hand through the use of indigenous organisms. The program was initiated in 2018 as a DBT-funded research project and the scope of the project was expanded under the

NBFGR-TSP fund and local trainees (77 women) had a month's duration of training to develop their capacity in growing these organisms. To take this initiative to the next level, ICAR-NBFGR has established a new facility, named “Incubation Centre”. This will be a place for “learning with working” for native women, mass-producing the species for dissemination, and also aggregating the marketable shrimps from beneficiaries, by SHG. To move further in marketing, ICAR-NBFGR, and M/S VGP Marine Kingdom, Chennai signed an agreement to procure the shrimps/fish for trade/exports.

In a similar line, another initiative of a live germplasm resource center for ornamental clownfish species was established at Airoli, Mumbai in collaboration with the Mangrove Foundation, Government of Maharashtra, where 10 clownfish species are being maintained (Fig. 2). A master hatchery facility is established for producing the seed, which could be given to the trained beneficiaries to rear them to the marketable size as a measure of livelihood development. Ten indigenous clown fishes are currently being produced and used in community aquaculture program. Captive propagation and providing technical know-how to the local community to rear the species to marketable size will provide additional income to the beneficiaries of coastal and island regions of the country, in addition to sustainable livelihood development. A similar initiative has been started recently for the beneficiaries belonging Pitchavaram region, Tamil Nadu, where the clown fishes are the candidate species for the initiated activity (Fig. 3).

**Freshwater Resources**

The institute has established a germplasm resource center in Guwahati in collaboration with Guwahati University, Assam for northeastern fish resources. Additionally, several fish species are maintained in the institute's Germplasm Resource Centre in Lucknow, including *Labeo rohita*, *Catla catla*, *Labeo rajasthanicus*, *Mystus cavasius*, *Garra gotyla*, *Labeo dero*, *Sperata seenghala*, *Labeo bata*, *Mastacembelus armatus*, and *Rita rita*. A newly discovered fish species from river Krishna, *Pangasius silasi*, is being maintained in Nagarjuna Sagar dam, Telangana. Furthermore, a live Germplasm Resource Centre in Kochi, Kerala, supports the conservation of endemic fishes of the Western Ghats. This facility preserves multi-species broodstock and contributes to genetic characterization, on-farm evaluations, and capacity-building programs for farmers and fisheries officials in the Peninsular region, promoting sustainable aquaculture and biodiversity conservation.

**Breeding Protocols for Endangered and Critically Endangered fishes**

- ICAR-NBFGR has made significant strides in developing breeding protocols for several endangered, critically endangered, and rare fish species. These efforts are essential for conservation, breeding, and restocking native habitats to protect biodiversity. *Horabagrus nigricollaris*: Developed the captive breeding technique for this endangered catfish and transferred technology to the farmers, enabling hatchery-based conservation, besides ranching was done in its native habitat.
- *Hemibagrus punctatus*: Established breeding and rearing protocols for the critically endangered catfish. The captive-bred fish were released into the Shivanasamudra Sanctuary in Kerala, contributing to the recovery of this threatened species in its natural habitat.*Ompok malabaricus:* Successfully developed captive breeding and rearing protocol for the rare and endemic catfish of Western Ghats.
- Rare ornamental barbs: Breeding technique for five rare ornamental barb species, including three endangered species *Dawkinsia rubrotinctus* (Three-Spot Barb), *Dawkinsia arulius* (Aral Barb), and *Dawkinsia tambraparniei* (Tambraparni Barb) and two vulnerable species, *Pethia setnai* and *Pethia nigripinna* were developed. Additionally, successful induced breeding of the zodiac loach, *Mesonoemacheilus triangularis* and a prized ornamental species native to Kerala's fast-flowing rivers was also carried out.

The breeding protocols developed for these endangered and rare species by ICAR-NBFGR play a critical role in the conservation of India's freshwater and ornamental fish biodiversity. By enabling hatchery-based breeding, restocking efforts, and technology transfer to local farmers, ICAR-NBFGR is ensuring that these species can be sustainably managed, reducing the risk of extinction and supporting both ecological and economic goals.

**Mass-Scale Seed Production for Endemic and Conservation-Critical Species**

Further, the institute has achieved significant milestones in developing breeding and larval rearing protocols for several endemic and conservation-critical fish species, enabling large-scale seed production. These efforts contribute to both biodiversity conservation and sustainable aquaculture practices. Thebreeding and larval rearing protocols for endemic catfishes, including *Clarias dussumieri* (Near Threatened), *Horabagrus brachysoma* (Vulnerable), and the minor carp *Labeo dussumieri* was standardized. The advancements in

breeding and seed production protocols for these species not only support conservation efforts for endemic and vulnerable fish but also create opportunities for sustainable aquaculture.
Key innovations include:

**Transforming Aquaculture with *Clarias dussumieri* in Biofloc System**

Trials with *Clarias dussumieri* in biofloc systems demonstrated its potential as a high-performing native species for aquaculture. Biofloc technology, which utilizes microbial communities to maintain water quality and recycle nutrients, enhances productivity while reducing environmental impact. The species has emerged as a promising indigenous species for Indian aquaculture, offering a sustainable alternative to exotic species like tilapia. Its adaptability to biofloc systems underscores its potential for eco-friendly aquaculture practices.

**Cryopreservation of fish sperm and its utility in conservation**

ICAR-NBFGR standardized species-specific cryopreservation protocols of spermatozoa for around 30 fish species for supporting captive breeding and restocking of fish species. Sperm cryopreservation protocols have been developed for around 30 commercially important and few endangered fish species in the country. The initiative addresses the issues of insufficient milt production or variation in the maturity of two sexes for captive breeding of cultivable species. Further, NBFGR is conducting cross-breeding experiments across the country using the cryopreserved milt of Indian Major Carps from their native habitat *i.e.,* River Ganga for enhancing the genetic diversity of the brooders and young ones, thereby improving the growth and reproductive traits.

**Cell line repository**

The National Repository of Fish Cell Lines (NRFC) was established at the institute premises at Lucknow with funding support by the Department of Biotechnology, Government of India, which serves as a National Referral Centre of fish cell lines for facilitating fish cell lines requirement for research purposes. About 50 fish cell lines developed from 24 fish species are being maintained. The facility accepts fish cell lines from various research institutions, maintains the cell lines using cryopreservation, and distributes the same on a need basis to the researchers in the country.

**Exotics and Fish health management**

National Surveillance Programme for Aquatic Animal Diseases (NSPAAD) was implemented with the ICAR-NBFGR as a lead institute with the support of DAHDF and NFDB. The success has become visible within a short span of six years from the inception of the program, initiated in 14 States, and has been expanded to cover 19 States and 2 Union Territories with 31 collaborating centers. The program has been helping in providing evidence-based inputs to DoF regarding aquatic animal health issues, the introduction of exotics, and trade and science-based inputs to queries from the trading countries.

National Strategic Plan for 'Aquatic Exotics and Quarantine Guidelines' is developed to regulate the introduction of exotics in the country. Species-specific guidelines for the import of tilapia, sutchi catfish, ornamental fish and white shrimp were developed. Data on the distribution of alien fish species in various drainages were generated and the impact of 14 species was quantified using a developed risk assessment process and protocols (FIST).

Bibliographic databases on fish pathogens and diseases, aquatic species introduction in the country, Indian fish pathologist's directory and alien fishes and quarantine information system have been developed. Molecular techniques for detection of OIE-listed pathogens such as *Aphanomyces invadans*, koi herpes virus, yellowhead virus, and taura syndrome virus were developed. Monoclonal antibodies were produced against serum immunoglobulins of *Labeo rohita* and *Channa striatus.*

**Conclusion**

To conclude, the ICAR-NBFGR is working with multipronged strategy which harmonizes germplasm centres as source of molecular characterized and evaluated broodstock, *ex situ* conservation technologies to support genetic exchange and also help *in situ* conservation in its native habitat of distribution. This envisages capacity building of the farmers, work on their awareness and understand the impact of seed quality on socio-economics of the farmers.

**Acknowledgements**

Author is extending gratitude to the Director, ICAR - NBFGR for his support and encouragements.

# Population genomic approaches for fisheries management and conservation

Sandhya Sukumaran
Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India - 682018
sandhya.sukumaran@icar.org.in

The vast amount of data that genomic tools can produce is driving the growth of genomic research in aquaculture and fisheries. Compared to mammals, which have 4,629 species, and birds, which have 9,946 species, fish are thought to be the most successful and diverse vertebrates on Earth, with about 28,600 species. Fish exhibit a variety of life histories, reproductive patterns, and prey-predator relationships in habitats that range from below freezing to over 40°C and from freshwater to hypersaline waters. They are regarded as great models for genomic studies on selection, adaptation, and evolution because of their great diversity and divergence.

**Population genetics**

Population genetic tools are well-established because they use neutral markers, such as mitochondrial DNA genes and microsatellites, to infer population characteristics. The tools of population genetics have made it possible to draw conclusions about genetic drift, gene flow, and inbreeding with a fair degree of accuracy, which has greatly advanced our understanding of the evolutionary processes crucial to the formation of population structure in a variety of oceanic habitats. The primary drawback of population genetics inferences, which were based on a small number of well-characterized neutral genes, is that they do not reveal information about adaptive mechanisms functioning at the genome level.

**Population genomics**

Data from numerous genome-wide markers are utilized in population genomics, unlike population genetics, which relies on a limited number of markers. Both genic and non-genic region information can be produced using population genomic tools, allowing for the differentiation of effects caused by selection, mutation, and recombination from those resulting from genetic drift, inbreeding, and gene flow. When only neutral markers are employed,

insights into adaptive evolution are absent, as adaptive divergence is crucial for local adaptation. Local adaptation ensures optimal fitness for organisms adjusting to specific environments or habitats, which drives the structuring of subpopulations in fish. Understanding locally adapted populations is vital for proposing conservation strategies aimed at preserving genetic diversity over ecological time scales. Climate-related habitat changes and the movements of marine fish can be anticipated if we possess data on environment-induced microevolution. Predicting genome-level selection footprints associated with specific traits is challenging, as multiple genes contribute to the characteristics of an adaptive trait. In this context, population genomics becomes increasingly significant, as genomic variations in numerous functional genes related to adaptation and selection across populations in both space and time can be identified using genomic tools. Various techniques, such as Expressed Sequence Tags Sequencing (EST), transcriptome sequencing, and whole genome sequencing, will assist in deriving crucial information regarding functional gene modifications associated with environmental changes.

**Approaches in population genomics**

One of the best ways to comprehend variations in real time is to examine changes in the expression of functional genes that play significant roles in adaptation; this method is also referred to as the candidate gene approach. For these investigations, functional genes with a known role and impact on an adaptive trait may be chosen. It is possible to select genes that are associated with changes in fish structure or with physiological processes and the resulting phenotypic variations. Information about functional gene variations can be obtained using allele frequency-based tests or sequence data.

The genome scan approach examines areas or loci that show significant structuring as a result of genetic hitchhiking. Studying these loci will reveal information about the selective forces operating on the genome because balancing selection or directional selection will result in variable patterns on such loci. This approach could make use of a wide range of genetic markers found throughout the genome, such as microsatellites, SNPs, AFLPs, or functional markers like genes connected to ESTs. Genes associated with ESTs may be found through transcriptome sequencing. Therefore, loci under selection can be found using genome scan techniques.

QTL mapping is based on data from linkage maps, which are created using a dense set of markers that provide hints about the genetic basis that determines phenotypic traits. This is accomplished through controlled laboratory family crosses. It would be less expensive to

compare QTL data between closely related species. The best candidates for QTL-related experiments will be aquaculture-friendly species because QTL investigations necessitate controlled experimental conditions. The information about linkage disequilibrium between numerous genetic markers that are connected to phenotypic variations forms the basis of association mapping. Admixture mapping is used to map admixture zones in the wild, and it aids in detecting natural admixture events like intra-specific hybrid zones. Seascape or landscape genomics aims to link genetic markers with geographic data in order to gain insights into geographic patterns of genetic differentiation. In sea scape genomics, both neutral and non-neutral or functional gene markers can be used to infer information about selection based on the degree of functional gene divergence at each geographic location. If there is any genetic divergence in non-neutral markers, it indicates populations that have adapted locally.

To comprehend the role of important functional genes in regulating essential physiological processes, information about differential gene expression patterns is crucial. Because transcription profiling and annotation are so simple, transcriptomic techniques, such as mass scale RNA sequencing, are very helpful in understanding variations in gene expression patterns in real time and can be applied to all species. While real-time PCR measures the gene expression of one or a few candidate genes, microarrays can measure the differential expression of hundreds or thousands of genes at once.

By tagging the reaction mixture with fluorescent dyes, real-time PCR measures the product in real time. It is based on the same principles as normal PCR. Each PCR cycle can be used to measure fluorescence, which is a measure of the expression levels of a particular gene. The main drawback with real-time PCR is that only very few genes within a genome can be studied using this method. But microarrays enable the investigation of the gene expression of thousands of genes simultaneously. Microarrays consist of thousands of single-stranded DNA spots attached to surfaces like glass slides, with each spot corresponding to a gene. Gene expression can be measured by washing fluorescently tagged cDNA over the slides so that complementary strands hybridize on to the array and fluoresce. The intensity of each spot will be proportional to the expression of that particular gene. All these investigations require controlled experimental conditions to relate gene expression data with adaptive variation.

**Genomics research in marine fishes - present status**

The enormous diversity and heterogeneity of marine fishes make them excellent models for fish genomics research. Marine fishes also exhibit high levels of gene flow mainly due to the

pelagic larval phase and consequent dispersal of many pelagic and demersal resources. Population genetic structure studies on many fish groups like clupeids (eg: herring, anchovy and sardine), scombrids (eg: mackerel, tuna and bonito), pleuronectids (eg: plaices, soles and flounders), and gadids (eg: cod, hake and haddock) have been carried out for defining management units. The outcome of these genetic investigations revealed a lack of structuring with high levels of intra-population diversity values. In addition, the effective population size of many marine fishes is very high, which also contributes to substantial gene flow and allele homogenization. In spite of all these factors, the presence of locally adapted populations with substantial divergence in morphological and physiological traits has been indicated in marine fishes.

Candidate functional genes have been studied in wild populations of many fishes. Pantophysin (Pan I) in Atlantic cod, Ectodysplasin (EDA) gene in stickle back and Lactate dehydrogenase B (Ldh-B) in killifish has been well characterized and studied. Haemoglobin polymorphism investigations in Atlantic cod revealed clear genetic differentiation across populations. Several insertion/deletion polymorphisms have been detected in Heat shock cognate gene Hsc70 in European flounder (*Platichthys flesus*). Genome scans in three-spined sticklebacks provided insights into their adaptation to fresh and marine environments. Divergence in different ecotypes has been studied using EST-based microsatellites and EDA gene linked indels. Outlier loci subjected to selection has been identified in Atlantic cod using mass scale genome scans. Adaptive population divergence corresponding to different geographic locations has been detected in many marine fishes.

Whole genome sequencing of 10000 vertebrates was aimed at by the project Genome 10K (Genome 10K community of scientists, 2009), and genome characterization of many of the fishes are progressing. The whole genome of cartilaginous elephant fish, *Carcharhinus milii,* which is a member of the Holocephali, has been assembled. Other endangered and threatened aquatic organisms being sequenced by Genome 10K and Beijing Genomics Institute (BGI) are *Acipenser sinensis* and *Hippocampus comes*. Recently whole genome of *Hippocampus comes* has been published.

Genomic signatures associated with migratory and resident ecotypes of endangered sock eye salmon *Oncorhynchus nerka* were studied using 2600 SNPs derived from Restriction site

associated DNA sequencing which provided vital clues to adaptive genetic variation between ecotypes. Putative neutral and divergent adaptive markers were used to determine population differentiation in an anadromous Pacific smelt using RAD sequencing. The study provided signatures of divergent selective pressures in differing freshwater and marine environments between regional eulachon populations. This information was helpful in defining management units. Cryptic population differentiation was studied in the Baltic Sea herring *Clupea harengus* using a total of 4756 divergent SNPs derived through RAD sequencing approach and they could detect high degree of differentiation. RAD sequencing has also been employed for understanding demographic history, quantifying population connectivity and detection of loci subject to local adaptation.

In the Indian context, the ICAR-Central Marine Fisheries Research Institute has developed whole-genome, transcriptome, and mitogenome data for the following species.

| Species | Resource | Genome Size / Assembly Size | Scaffold N50 | Assembly Completeness / Quality | Protein-coding Genes | Other Statistics / Remarks |
|---|---|---|---|---|---|---|
| Indian oil sardine (*Sardinella longiceps*) (Sukumaran et al. 2023) | Whole genome assembly | 1.077 Gb | 31.86 Mb | BUSCO completeness 93.5% (Actinopterygii dataset) | 46,316 | Repeat content 23.24%; |
| Asian green mussel (*Perna viridis*) (Sukumaran et al. 2024) | Chromosome-level genome assembly | 723.49 Mb | 49.74 Mb | 99% of the assembly is anchored to 15 chromosomes | 49,654 | 634 tumour-associated genes; 408 viral-carcinogenicity-associated genes |
| Indian oil sardine (*Sardinella longiceps*) (Sukumaran et al. 2023) | Transcriptome | 95,426 transcripts | 1937bp (contig N50) | 72% coverage of expressed genes in the genome | 34,525 | Expressed gene dataset generated |
| Asian green mussel (*Perna viridis*) (Vysakh et al. 2025) | Tissue-specific (gill) transcriptome | 438,842 transcripts | 1,958 bp | BUSCO completeness 94.5% against Mollusca database | 19,464 | Multiple tissue transcriptomes available |
| Indian brown mussel (*Perna* | Transcriptome generated from 4 different | 157,203 transcripts | 1725bp | --- | 28,692 | De novo transcriptome assembly |

| | | | | | | |
|---|---|---|---|---|---|---|
| *indica*) (Mary S et al. 2021) | tissues( gill, foot, abductor, mantle) | | | | | |
| Charru mussel (*Mytella strigata*) (Vysakh et al. 2025) | Transcriptome from (hepatopancreas, gill, mantle, foot, and adductor muscle) | 401,754 transcripts generated | 1578 bp | BUSCO completeness 98.11% against Metazoan database | 60,362 | De novo transcriptome assembly |
| Yellowfoot clam (*Paphia malabarica*) (Sukumaran et al. 2026) | Transcriptome (adductor muscle, mantle, foot and gill) | 274,880 transcripts | 1605bp | BUSCO completeness 97.8% against Metazoan database | 55,980 | First comprehensive genomic resource. |
| Grey bamboo shark (*Chiloscyllium griseum*) (P.Harshan et.al, 2024) | Transcriptome (heart, spleen, brain, kidney and liver) | 482,871 transcripts | 1569 bp | BUSCO completeness 91.5% against vertebrate database | 70,647 | De novo transcriptome assembly, Foundation for comparative genomic studies. |
| Indian squid (*Uroteuthis duvaucelii*) (Krishnan N et.al, 2024) | Transcriptome (brain, eye, gill, heart and gonads) | 188,143 transcripts | 1359 bp | BUSCO completeness 71.71% against Mollusca database | 26,230 | De novo transcriptome assembly of cephalopod species. |
| Cobia (*Rachycentron canadum*) (Ebeneezar, Sanal, et al,2023) | Transcriptome from different life stages (larval, Juvenile, adult) | 35,661 isoforms | 2984 | BUSCO completeness 71.94% against vertebrate database | 38,243 | First full length transcriptome generated |
| *Sardinella longiceps* (Sebastian et al. 2018) | Mitogenome | 16,613bp | -- | -- | 13 protein-coding genes 22 tRNA genes 2 rRNA genes. | First time characterized. |
| Goldstripe Sardinella *Sardinella gibbosa* (Sebastian et al.2018) | Mitogenome | 16658 bp | -- | -- | 13 protein-coding genes 22 tRNA genes 2 rRNA genes | First time characterized. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | 1 non coding region. | |
| pearl spot *Etroplus suratensis* (Sebastian et al. 2019), | Mitogenome | 16456 bp | | | 13 protein-coding genes<br>22 tRNA genes<br>2 rRNA genes<br>1 non coding region | First time characterized. |
| Randall's threadfin bream *Nemipterus randalli* (Raj et al.2024), | Mitogenome | 16642 bp | | | 13 protein-coding genes<br>22 tRNA genes<br>2 rRNA genes<br>1 non coding region | First time characterized. |
| Indian scad *Decapterus russelli (*Jose et al.2022*)*, | Mitogenome | 16,542 bp | | | 13 protein-coding genes<br>22 tRNA genes<br>2 rRNA genes<br>1 non coding region | First time characterized. |
| Ribbon fish *Trichiurus lepturus* (Mukundan et al. 2020), | Mitogenome | 16,840 bp | | | 13 protein-coding genes *(including ATP8)*<br>22 tRNA genes<br>2 rRNA genes | First time characterized. |
| Black clam *Villorita cyprinoides* (Rahuman et al. 2020), | Mitogenome | 15,880 bp | | | 13 protein-coding genes<br>22 tRNA genes | First complete mitochondrial genome using next-generation sequencing. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | 2 rRNA genes | |
| mud spiny lobster *Panulirus polyphagus* (Jeena et al. 2024), | Mitogenome | 15,685 – 15705 bp | | | 13 protein-coding genes (PCGs) 22 (tRNA) genes 2 r RNA genes 1Control Region (CR)/D-loop ) | The mitochondrial genome from three different locations specifies regional variation among them. |
| *Sthenoteuthis oualaniensis* (Jeena et al. 2023). | Mitogenome | 20318bp | | | 8 protein-coding genes (PCGs) 23 tRNAs 2 rRNAs 2 non-coding regions | First complete mitochondrial genome characterised. . |

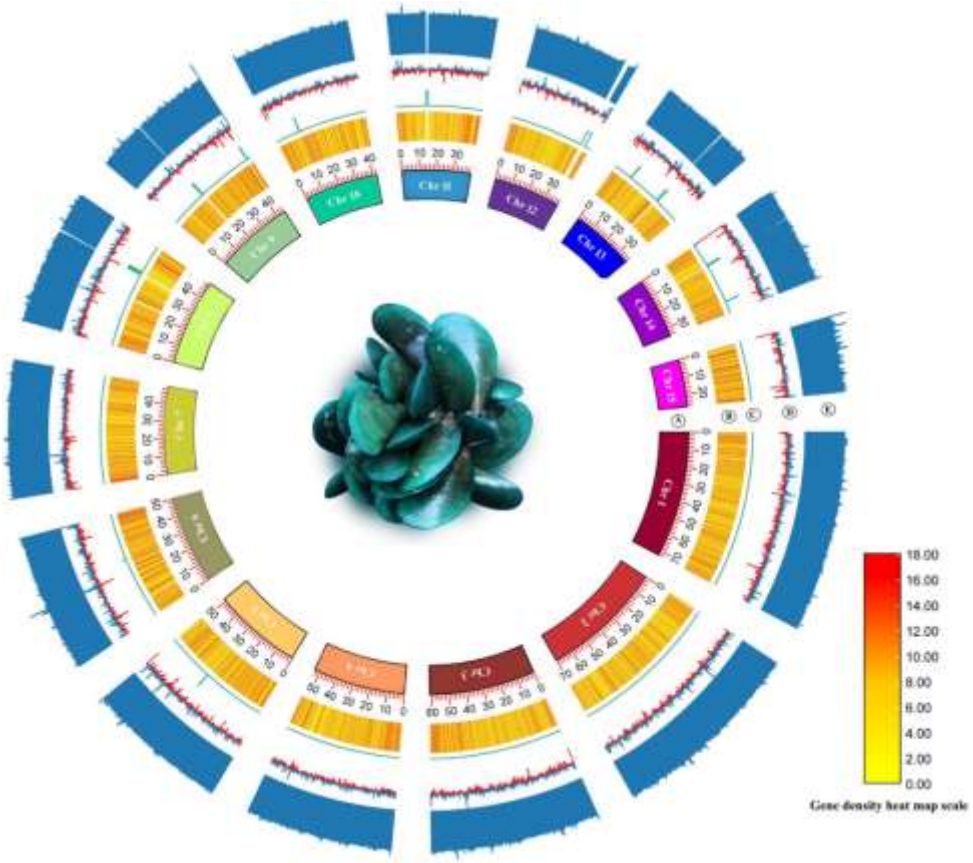



**Figure 1:** Circos plot of the *P. viridis* genome assembly. (A) 15 chromosomes (B) Gene density heat map (C) N - ratio, (D) GC skew (E) the distribution of GC content

## Genomic and mitogenomic investigations in marine fishes to delineate population structure and climatic impacts in the Indian context

Population genomic studies have been conducted in *Sardinella longiceps*, an Indian oil sardine, using ddRAD sequencing to provide information on adaptive genetic variation in response to diverse habitats in the Indian Ocean (Sebastian et al. 2021). Based on single-nucleotide polymorphism analysis, the Gulf of Oman samples differed significantly from those from the Indian Ocean. A correlation between SNP alleles and the environment revealed candidate loci correlated with environmental variables such as annual sea surface temperature, chlorophyll and dissolved oxygen concentrations, which may represent genomic regions that are different due to local adaptations. Modelling of larval dispersal along the south-western coast of India has shown a high rate of dispersal of the larval species. The analysis concluded that the two main subpopulations (Oman Gulf and Indian Ocean) need to be managed regionally to maintain genetic diversity, which is crucial for climate resilience.

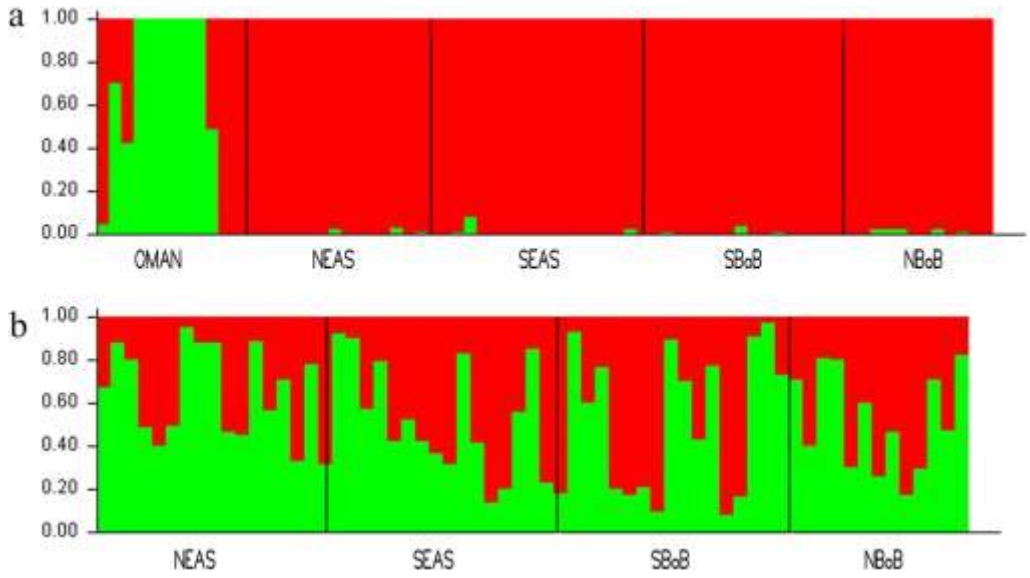



**Figure 2:** Results of Structure analysis undertaken in the Indian oil sardine, Sardinella longiceps using SNPs generated using ddRAD sequencing.

Mitogenomic studies of the Indian oil sardine, *Sardinella longiceps*, have been carried out to understand the developmental consequences of diverse environmental conditions on the mitogenome by analyzing the selection patterns of mitochondrial OXPHOS genes in sardine collected from various eco-regions in the Indian Ocean (Sebastian et al. 2020). The gene regulation of the OXPHOS system is a key factor in meeting the energy requirements of pelagic marine species with higher metabolic rates. The positive and diversifying patterns were more common in the fish of the South East Arabian Sea (SEAS), followed by the Northern Arabian Sea (NAS), and were rare in the populations of the Bay of Bengal (BoB). Due to the selection pressure to survive in a highly variable ocean environment with seasonal hypoxia, variable SST, and food availability, SEAS fish have shown an accelerated rate of substitution (Sebastian et al. 2020).

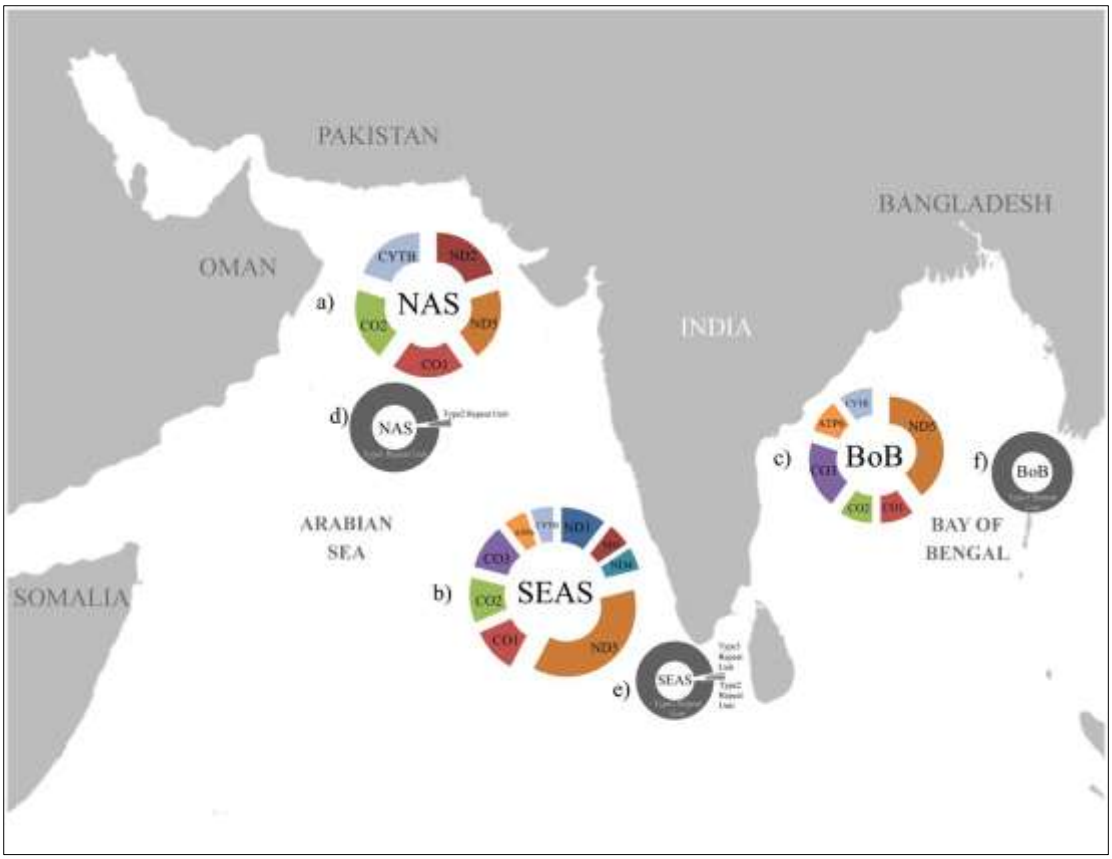



**Figure 3:** Graphical representation of the geographical distribution of positively selected sites and Control region repeat unit types in the mitogenome of *S. longiceps* in the 3 eco-regions of the Indian Ocean, (a) Frequency of positively selected sites in NAS, (b) Frequency of positively selected sites in SEAS, (c) Frequency of positively selected sites in BoB, (d) Frequency of haplotype with Type 1 repeat unit in NAS, (e) Frequency of haplotype with Type 2 repeat unit in SEAS and (f) Frequency of haplotype with repeat unit Type 3 in BoB.

The climatic vulnerability and adaptive capacity of the Clupeids of the world oceans have been studied by assessing the signals of diversifying selection in 70 clupeid species belonging to five families (Sebastian et al. 2022). Clupeids adapted to freshwater, brackish, and marine waters from temperate to tropical regions have been analyzed to understand the potential signs of diversification. Signals of positive and diversifying selection have been detected in mitochondrial protein-coding genes, suggesting that mitochondrial OXPHOS machinery is a key adaptation mechanism. Mitogenomic plasticity, and thus increased resilience of the metabolic machinery, would contribute to the evolutionary success and abundance of clupeoid fish. These findings are of great importance in the current climate change scenario, where only resilient populations will survive and clupeoids have a chance (Sebastian et al. 2022).

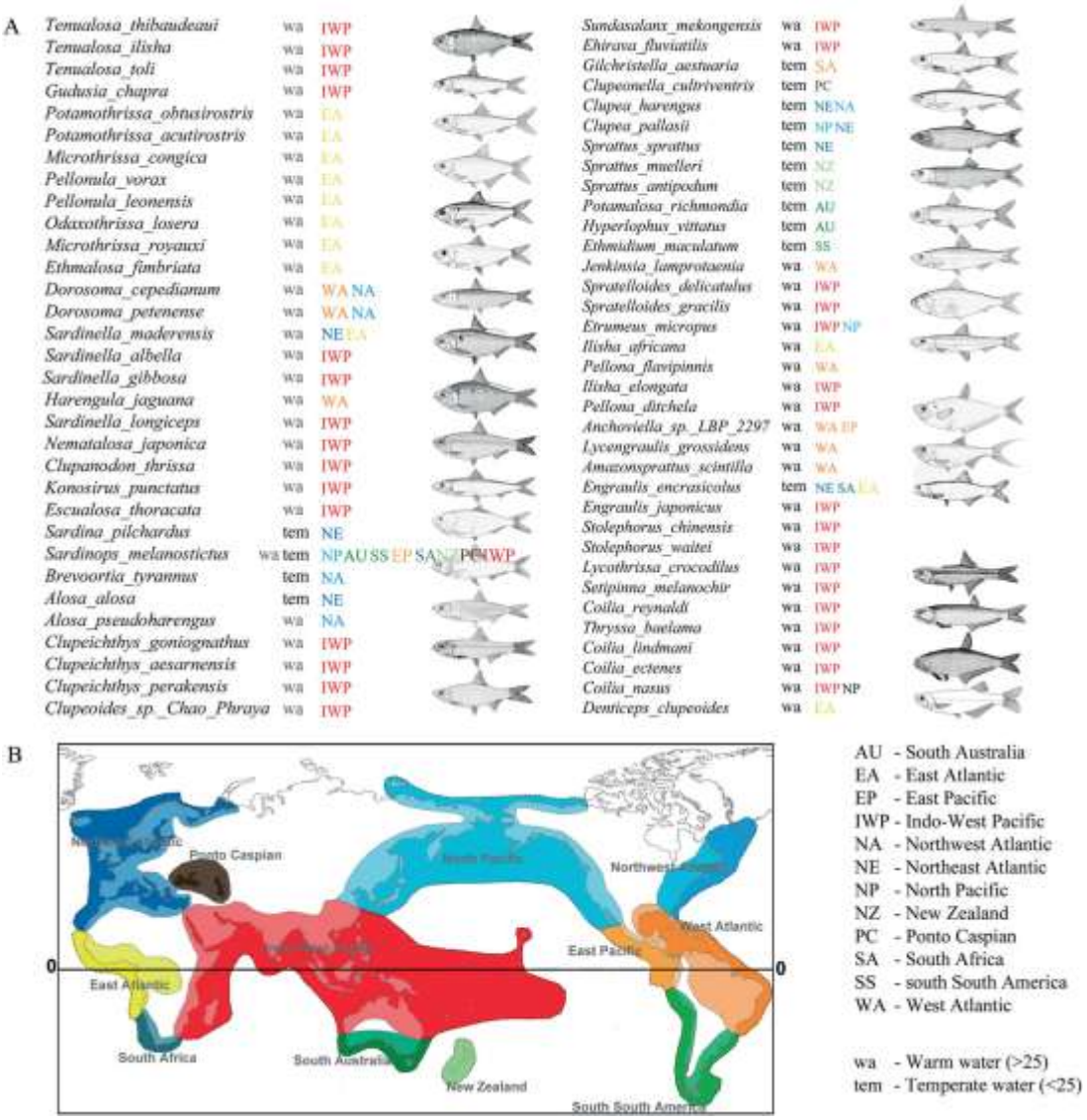



**Figure 4**: Geographical distribution pattern of the Clupeoids. (A) 70 Clupeoids used for the study (AU South Australia, EA East Atlantic, EP East Pacific, IWP Indo-West Pacific, NA North west Atlantic, NE North East Atlantic, NP North Pacific, NZ New Zealand, PC Ponto Caspian, SA South Africa, SS South South America, WAWest Atlantic) and (B) their biogeographical distribution. The spatial distribution data of the species were taken from the FishBase database (www.fishbase.in) and Lavoue et al. 2014**.**

The selection patterns of *E. suratensis* OXPHOS genes were examined by comparing OXPHOS genes from 105 fish species collected from various ecological zones in India. The signs of positive and diversifying selection observed in mitogenomes have been correlated with the characteristics of the habitat ( Sebastian et al. 2021). The observed habitat-specific mutational signals are due to adaptation to these environmentally diverse habitats, which produce genotypic and phenotypic variants with specific metabolic and bioenergetic requirements (Sebastian et al. 2021)

**Future of population genomics**

Fishes face many physical, chemical, and biological challenges in the dynamic environment in which they live. Surviving in these conditions requires alterations in phenotypic characteristics, and the major driver for these variations is natural selection occurring at varying degrees. Genome investigations provide insights regarding micro-evolutionary patterns, and the advent of next-generation sequencing technologies has revolutionized these approaches. We are witnessing an era of genomics and the floodgates for genomic technologies being opened now; a revolution in fish genomic research is imminent, with more and more fish genomes being sequenced.

**Key references**

Allendorf, F.W., Hohenlohe, P.A. and Luikart, G. 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics 11: 697-709.

Cossins, A.R. and Crawford, D.L. 2005. Fish as models for environmental genomics. Nature Reviews Genetics 6:324-340.

Crollius, H.R. and Weissenbach, J. 2005. Fish genomics and biology. Genome Research 15: 1675-1682.

Ding, F., Chu, W., Cui, P., Tao, M., Zhou, R., Zhao, F., Hu, S. and Zhang, J. 2011. EST-Based identification of genes expressed in skeletal muscle of the Mandarin Fish (*Siniperca chuatsi*). Genomics, Proteomics and Bioinformatics 9(1-2): 30-36.

Guo, B., DeFaveri, J., Sotelo, G., Nair, A. and Merila, J. 2015. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. BMC Biology 13: 19

Hansen, M.M.,Olivieri, I., Waller, D.M., Nielsen, E.E. and The GeM working group. 2012. Monitoring adaptive genetic responses to environmental change. Molecular Ecology 21: 1311-1329.

Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M.I., Ogden, R., Limborg, M.T., Cariani, A., Maes, G.E., Diopere, E., Carvalho, G.R. and Nielsen, E.E. 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Molecular Ecology Resources 11(Suppl.1), 123-136.

Hemmer-Hansen, J., Therkildsen, N.O. and Pujolar, J.M. 2014. Population genomics of marine fishes: Next – Generation prospects and challenges. Biological Bulletin 227: 117-132.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A. and Cresko, W.A. 2010. Population genomics of parallel adaptation in Threespine stickleback using sequenced RAD tags. PLoS Genetics 6(2): 1-23.

Nielsen, E.E., Hemmer-Hansen, J., Larsen, P.F. and Bekkevold,D. 2009. Population genomics of marine fishes: identifying adaptive variation in space and time. Molecular Ecology 18: 3128-3150.

Quere, N., Guinand, B., Kuhl, H., Reinhardt, R., Bonhomme, F. and Desmarais, E. 2010. Genomic sequences and genetic differentiation at associated tandem repeat markers in growth hormone, somatolactin and insulin-like growth factor-1 genes of the sea bass, *Dicentrarchus labrax*. Aquatic Living Resources 23: 285-296.

Zhu, M. and Zhao, S. 2007. Candidate gene identification approach: progress and challenges. International Journal of Biological Sciences 3(7): 420-427.

Sebastian W, Sandhya Sukumaran, Abdul Azeez S, Muraleedharan KR, Dinesh Kumar P K, Zacharia PU and Gopalakrishnan A (**2021**) *Genomic investigations provide insights into the mechanisms of resilience to heterogeneous habitats of the Indian Ocean in a pelagic fish.* Nature Scientific Reports. pp. 1-16.

Sebastian W., Sandhya Sukumaran, Zacharia PU, Muraleedharan KR, Dinesh Kumar P K and Gopalakrishnan A (2020) *Signals of selection in the mitogenome provide insights into adaptation mechanisms in heterogeneous habitats in a widely distributed pelagic fish.* Nature Scientific Reports, 10 (9081). pp. 1-14.

# Progress in Molecular Marker Development and Their Applications in Biological Investigations

Sandhya Sukumaran

Marine Biotechnology Fish Nutrition and Health Division, ICAR- Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India- 682018 sandhya.sukumaran@icar.org.in

## Introduction

Recognizing the genetic variation present within and across species is vital for inferring the conservation requirements of various species. The study of genetic diversity and variation can be conducted using molecular markers such as allozymes, mitochondrial DNA markers, microsatellites, single nucleotide polymorphisms, and many other cutting-edge tools. Genetic variation contributes to differentiation at the levels of population, species, or higher taxonomic groups, which can be effectively delineated with molecular markers.

Molecular markers can be classified into Type I and Type II categories. Type I markers are those linked to genes with established functions, while Type II markers are associated with genes whose functions remain unknown. Allozyme markers fall under the Type I category since the proteins they encode are linked to certain known functions. In contrast, microsatellites and other neutral markers are classified as Type II markers, as they are generally not connected to any recognized function.

## Allozymes

The allozyme method is recognized as one of the leading approaches for the analysis of genetic variation and stock structure in fish. Allozymes function as codominant markers that are expressed in heterozygous individuals following Mendelian inheritance patterns. Allozyme analysis provides critical information regarding single locus genetic variation among populations, thus making it particularly suitable for genetic stock structure studies. The allozyme analysis process involves the extraction of allozymes from tissues according to standard protocols. After extraction, allozyme variation can be detected through electrophoresis on either acrylamide or cellulose acetate gels. Homozygous individuals will present a single band, while heterozygous individuals will exhibit two bands. The method's simplicity, low cost, and minimal need for specialized equipment have made allozymes the preferred molecular

markers for genetic research. Any soluble proteins can be utilized for allozyme analysis, allowing for the simultaneous screening of multiple loci. However, the allozyme method does have limitations, including the requirement for large tissue samples (which means larvae cannot be studied), the invasive nature of the sampling process, and the inability to detect point mutations, as many do not lead to amino acid changes. Despite these limitations, allozyme analysis has been widely used in fisheries science for drawing conclusions about fish systematics, population genetic structure, conservation genetics, mixed stock analysis, and forensic investigations.

**Mitochondrial DNA**

Mitochondrial DNA is located in the mitochondria of the cells and hence is termed as non-nuclear DNA. It is inherited maternally with a haploid genome and entire genome undergoes transcription as one single unit. Mitochondrial DNA is not subjected to recombination and hence they are homologous markers. Mitochondrial DNA is selectively neutral and occurs in multiple copies in each cell. As they are physically separate from the rest of cell's DNA, it is easy to isolate them from any tissue samples. The effective population size of mitochondrial DNA is smaller than nuclear DNA due to their maternal inheritance which makes them sensitive to population bottlenecks and hybridizations. There are several direct and indirect methods to understand the variations in nucleotide sequence of mtDNA. Restriction Fragment Length Polymorphism analysis involves digesting the whole mtDNA with restriction endonucleases and visualizing the fragments on a gel. DNA sequencing of small fragments of mitochondrial genes and analyzing the sequences is another method. The simplicity and accuracy of these methods made mitochondrial DNA a marker of choice for many investigations.

Multiple sets of universal primers have been formulated from conserved sequence regions of mitochondrial DNA, which are commonly used. Inter-specific comparisons are conducted using gene regions that evolve slowly, while intra-specific comparisons are made with gene regions that evolve quickly. The mitochondrial control region is rapidly evolving and is utilized for population comparisons. Cytochrome b and NaDH dehydrogenase genes (ND1, ND2) are also widely used. The conserved nature of the cytochrome c oxidase 1 gene has established it as a universal barcode for differentiating species at the species level. DNA barcodes are roughly 650 base pairs long, and this mitochondrial DNA region has been employed for cataloguing

the biodiversity of various organisms. Mitochondrial genes have found extensive applications in taxonomy, evolutionary biology, and population genetics research.

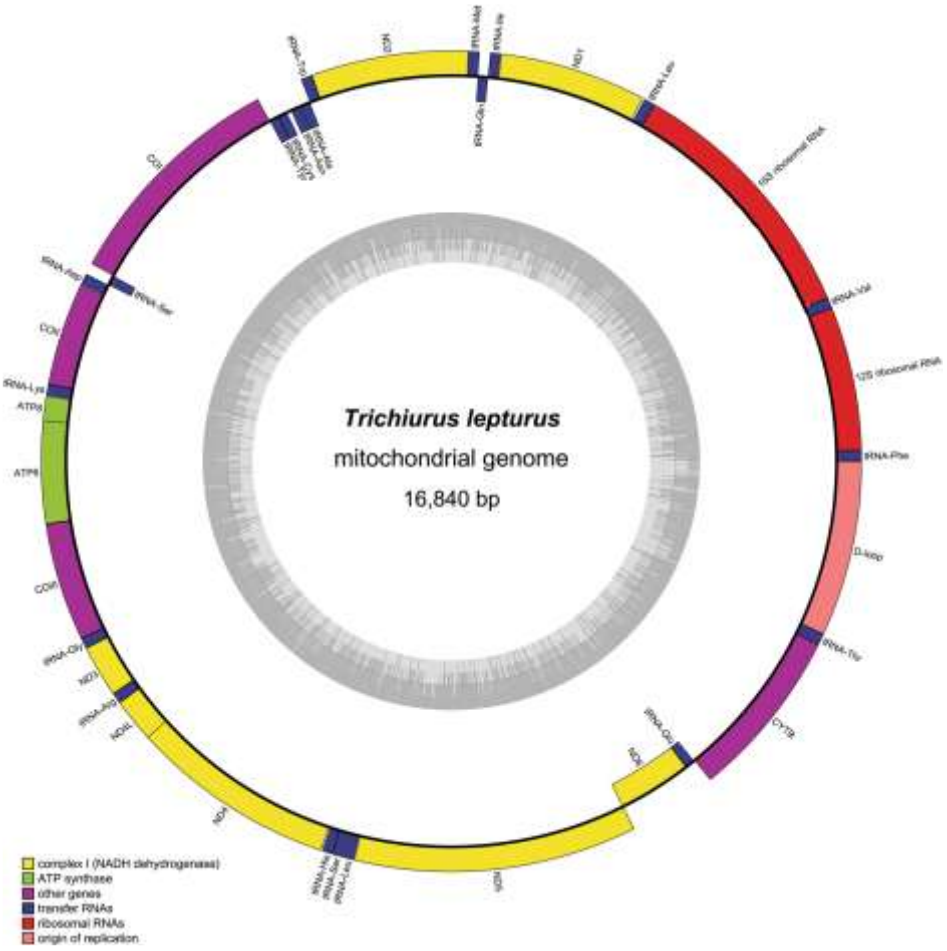



**Figure 1: T**he completed mitogenome of the Ribbon fish, *Trichiurus lepturus.*

**Nuclear DNA markers – Random markers**

Arbitrary nuclear markers are designed to target DNA fragments whose functions remain unknown. The techniques RAPD (Random Amplified Polymorphic DNA), ISSR (Inter-Simple Sequence Repeats), and AFLP (Amplified Fragment Length Polymorphism) are widely recognized methods for amplifying unknown DNA regions. In RAPD, an arbitrary primer is employed to amplify unknown loci. This method is noted for its rapidity, affordability, and high levels of polymorphism, and it does not require any prior understanding of the genetic constitution of the organism for the application of RAPD markers. However, a significant limitation of RAPD markers is their inconsistency and lack of reproducibility across laboratories, resulting in a multitude of products being generated. It is classified as a dominant

marker, as it does not allow for the differentiation between homozygous and heterozygous loci. Additionally, RAPD patterns are particularly sensitive to slight changes in amplification conditions.

Inter-simple sequence repeats (ISSRs) represent sections of the genome that are flanked by microsatellite sequences. The amplification of these regions using PCR primers yields several products, which can be utilized as a dominant multilocus marker system for investigating genetic diversity. ISSR markers, similar to RAPD markers, are straightforward to use, economical, and less methodologically intensive, rendering them ideal for beginners. They are also applicable to organisms that lack comprehensive genetic data. Nevertheless, like RAPD markers, the dominant inheritance pattern and random nature contribute to a decrease in accuracy.

Amplified Fragment Length Polymorphism (AFLP) markers merge the advantages of RFLP and RAPD methodologies. A pair of restriction enzymes is utilized to digest the complete genomic DNA. Moreover, double-stranded nucleotide adapters are ligated to the ends of the DNA fragments, serving as binding sites for primers for PCR amplification. Primers that are complementary to the adaptor and restriction site sequences, with additional nucleotide sequences at the 3'-end, are employed to amplify the ligated fragments. The detection of DNA fragments is performed on polyacrylamide gels to identify polymorphisms.

**Nuclear DNA markers – Specific markers**

Variable Number of Tandem Repeats are segments of DNA that are repeated in tandem from tens to thousands of times within nuclear genomes. The number of tandem repeats can differ across various loci and among individuals. There are two primary categories of highly polymorphic and repetitive DNAs present in individuals: minisatellite and microsatellite DNA. Genetic loci characterized by repeat lengths ranging from 9 to 65 base pairs are referred to as minisatellite DNA, while those with repeat lengths of 2 to 8 base pairs are classified as microsatellite DNA. Microsatellites are more prevalent in the genome compared to minisatellites. These loci are extensively utilized in the population genetics of both vertebrates and invertebrates. Minisatellites can be categorized into multilocus and single locus types, primarily employed for parentage analysis. However, they are not particularly advantageous for population genetic studies due to the intricate mutation processes they undergo. The majority of research on minisatellites has concentrated on single locus minisatellite probes,

which have proven effective in identifying genetic variations among populations. Additionally, minisatellites have found widespread application in forensics, genetic identity testing, parentage analysis, and in confirming gynogenesis and androgenesis, among other processes.

**Microsatellites**

Microsatellites are repetitive DNA sequences that occur multiple times at various locations within an organism's DNA. Due to their high variability, microsatellites serve as effective markers. They appear once every 10kbp, in contrast to minisatellites, which are found every 1500kbp in fish, making microsatellites particularly valuable for genome mapping and studies of genetic stock structure.Given their high variability, non-coding nature, and classification as selectively neutral, the fundamental premise for utilizing microsatellites is that the divergence in sequences correlates with the duration since separation. Microsatellites are inherited according to Mendelian principles and function as codominant markers, evolving at a rate of 10-3 to 10-4 mutations per generation.

Because of their high polymorphism, microsatellites have gained a lot of popularity. Cross-amplification, or using primers from closely related species, lowers the cost of finding microsatellite sequences in a different species. Polyacrylamide gel electrophoresis (PAGE) is used to visualize microsatellite loci after they are amplified from genomic DNA using certain primers in a PCR machine. Size polymorphisms are analyzed by automated genotyping utilizing labeled primers in genotyping equipment. Many loci may be tested thanks to automated genotyping, which improves speed, accuracy, and precision. The existence of stutter bands and null alleles are significant limitations when employing microsatellite markers. Mutations in the primer binding sites of the microsatellite locus result in the formation of null alleles. Null alleles decrease accuracy, particularly when paternity or related studies are being performed, hence loci displaying them should be eliminated from study. Stutter bands, which vary by 1-2 bp, are caused by insufficient denaturation of amplification products or slipped strands that hinder PCR. Since these loci do not exhibit appreciable stuttering, the issue can be resolved by amplifying tri-nucleotide and tetra-nucleotide repetitions. In fisheries research, microsatellite markers are crucial, particularly for phylogenetic studies, phylogeography, genetic stock structure, biodiversity conservation, stocking effects, and hybridization. Additionally, it may be used to map genomes, identify people forensically, and comprehend kinship and behavioral patterns.

**Single Nucleotide Polymorphisms**

Single nucleotide polymorphisms (SNPs) are the most prevalent polymorphisms in the coding and non-coding regions of the genome that can be found using PCR, microchip arrays, or fluorescence technology. SNPs are markers of the next generation and can be used for a variety of applications, such as genomic investigations, the effects of climate change, and the detection of diseases. Single nucleotide substitutions, such as transitions, transversions, insertions, or deletions, result in the genome.

**DNA microarrays or DNA chips**

DNA microarrays consist of several immobilized DNA fragments organized in a regular pattern on tiny glass microscope slides, silicon chips, or nylon membranes. They serve as a medium for comparing DNA extracted from an unidentified target sample with published probes for recognized sequences. For identification reasons, species-specific DNA sequences can be added to a DNA microarray. In order to detect a fluorescent signal with the proper fluorescence scanning or imaging equipment when the hybridization is affirmative, DNA extracted from target samples should be tagged with certain fluorescent molecules and hybridized to the microarray DNA.

**Expressed Sequence Tags (ESTs)**

ESTs are single pass sequences that may be produced by randomly sequencing clones of cDNA. Through expression analysis, ESTs are highly helpful for identifying genes and comprehending their expression. These markers can be used to quickly and accurately investigate the genes expressed in certain tissue types under particular physiological circumstances or developmental phases. Using cDNA microarrays, differentially expressed genes may be systematically identified. Linkage mapping greatly benefits from the use of ESTs.

**Key references**

Hallerman, E. M. (2003). Population genetics: principles and applications for fisheries scientists.

Beaumont, A. R., & Hoare, K. (2003). *Biotechnology and genetics in fisheries and aquaculture*. Blackwell publishing company.

Kucuktas, H., & Liu, Z. (2007). Allozyme and mitochondrial DNA markers. *Aquaculture genome technologies*

Zhu, Z. Y., Lin, G., Lo, L. C., Xu, Y. X., Feng, F., Chou, R., & Yue, G. H. (2006). Genetic analyses of Asian seabass stocks using novel polymorphic microsatellites. *Aquaculture*, *256*(1-4), 167-173

O'Reilly, P. T., & Kozfkay, C. C. (2014). Use of microsatellite data and pedigree information in the genetic management of two long-term salmon conservation programs. *Reviews in Fish Biology and Fisheries*, *24*(3), 819-848.

Chakona, A. J. (2022). Advancements in Fish Phylogenetics: From Molecular Markers to Next-Generation Sequencing. *FishTaxa-Journal of Fish Taxonomy*, *26*, 45-54

Ibrahim, S., Yue, C., Yang, C., Yiting, R., Xie, B., Yang, Q., ... & Deng, Y. (2025). A review of genomics application in aquaculture: Status, challenges, and priorities for future direction. *Annals of Animal Science*, *26*(3), 695-715..

.

,

# Molecular phylogeny and Systematics

Jeena N S* and Summaya Rahuman

Marine biotechnology fish nutrition and health division, , ICAR-CMFRI, Kochi. * jeenacmfri@gmail.com

Classification is a fundamental aspect of all scientific disciplines, particularly in biology. Taxonomy, a core component of systematics, deals with the processes of classification, identification, and nomenclature. Traditional taxonomy, often referred to as the Linnaean system, relies on the morphological and anatomical characteristics of organisms to classify and identify them. The underlying principle is that the more traits organisms share, the closer their evolutionary relationship is presumed to be. Key features of the Linnaean system that remain influential today include its hierarchical classification framework and the binomial nomenclature system. Hierarchical classification organizes living organisms into a structured system of ranks, reflecting their perceived evolutionary relationships. The modern taxonomic system comprises eight primary levels, each representing a specific degree of relatedness.

**Phylogenetics and Evolutionary Relationships**

Phylogenetic studies play a crucial role in understanding the diversity of life, as a good understanding into evolutionary history inform effective conservation strategies and biodiversity planning. The primary goal of most phylogenetic research is to reconstruct evolutionary trees that depict the relationships among organisms. Molecular phylogenetics, a subfield of phylogeny, focuses on analyzing genetic differences, particularly in DNA sequences, to infer evolutionary histories and construct phylogenetic trees. This approach falls under the broader field of molecular systematics, which also integrates molecular data into taxonomy and biogeography.

Systematics focuses on understanding the relationships among living beings within the Tree of Life. Over time, the concept of "relationship" in systematics has evolved, it is now widely understood as the genealogical connections shaped by evolutionary descent. DNA provides more detailed phylogenetic information than proteins and is the primary tool in molecular phylogenetics for studying evolutionary relationships. Currently, phylogenetic systematics are now widely accepted as the standard framework for understanding evolutionary relationships.

## Fish Taxonomy and Phylogenetic Advances

Globally, approximately 37,106 valid fish species have been documented (Fricke et al., 2025), with India contributing around 3,157 species, of which 1,545 are exclusively marine (NBFGR, 2020). The first clear phylogenetic classification of fishes was introduced by G. Nelson in 1969, accompanied by a detailed discussion of phylogenetic systematics principles. The advent of genomics in the early 21st century facilitated the availability of larger nuclear gene datasets, enabling the construction of multilocus phylogenetic trees. Advances in PCR and Sanger sequencing technologies further accelerated the generation of multi-gene (near to 20 gene fragments) phylogenies, which significantly enhanced our understanding of fish evolution and relatedness. The use of molecular data to explain higher-order phylogenetic relationships among fishes began in the 1990s. The analysis with complete mitochondrial genomes started in 1999, revealed unexpected evolutionary connections among fish lineages.

In recent years, next-generation sequencing (NGS) technologies have revolutionized phylogenetic research by identifying thousands of genetic markers in a single reaction (e.g., target enrichment). These advancements allow for the analysis of hundreds to thousands of gene fragments, addressing complex phylogenetic questions with unprecedented precision. Numerous laboratories are now engaged in generating extensive DNA sequence data for genetic diversity and phylogenetic studies. Efforts to compile genome-scale databases for large-scale phylogenomic analyses of fishes are actively ongoing (Betancur-R et al., 2017)

## Understanding Phylogenetic Trees

A phylogenetic tree, also known as an evolutionary tree, is a graphical representation of the evolutionary relationships among species. It provides a visual framework to understand how different species or sequences that are derived from them related through shared ancestry.

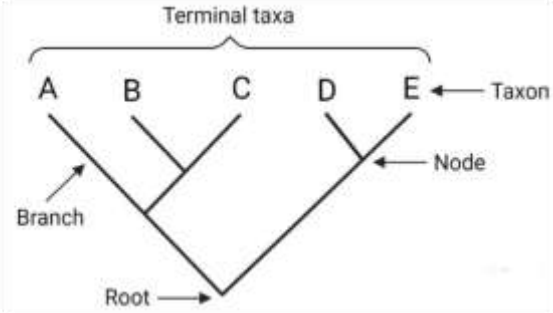



**Figure 1:** A typical phylogenetic tree

Phylogenetic trees are typically depicted as two-dimensional diagrams. The branches represent the evolutionary pathways and connections between organisms, while the endpoints of these branches, called taxa or operational taxonomic units (OTUs), represent present-day species or sequences (Figure 1). The points where branches intersect, known as nodes, indicate inferred common ancestors of the connected organisms or sequences. At the base of the tree, the root node represents the most recent common ancestor of all the organisms or sequences included in the tree.

**Types of phylogenetic trees**

Phylogenetic trees can be broadly classified into two types based on the presence or absence of a root (Figure 2 a&b). Rooted Trees: These trees have a designated root node, that signifies the common ancestor of all the organisms or sequences in the tree. Unrooted Trees: Unlike rooted trees, unrooted trees do not specify a root node. They illustrate only the branching patterns of evolutionary relationships among taxa or OTUs, without indicating a common ancestor.

Phylogenetic trees can also be classified based on their topology into two main types (Figure 2 c&d).

**Cladograms**: These trees focus solely on the branching patterns of evolutionary relationships. Cladograms are unscaled, meaning the lengths of the branches do not correspond to the degree of evolutionary divergence between taxa or OTUs. They are useful for highlighting the relative relationships among organisms.

**Phylograms**: In contrast to cladograms, phylograms provide more detailed information by depicting both the branching patterns and the extent of evolutionary divergence. The branch lengths in phylograms are scaled, meaning they are proportional to the amount of genetic or evolutionary change that has occurred.

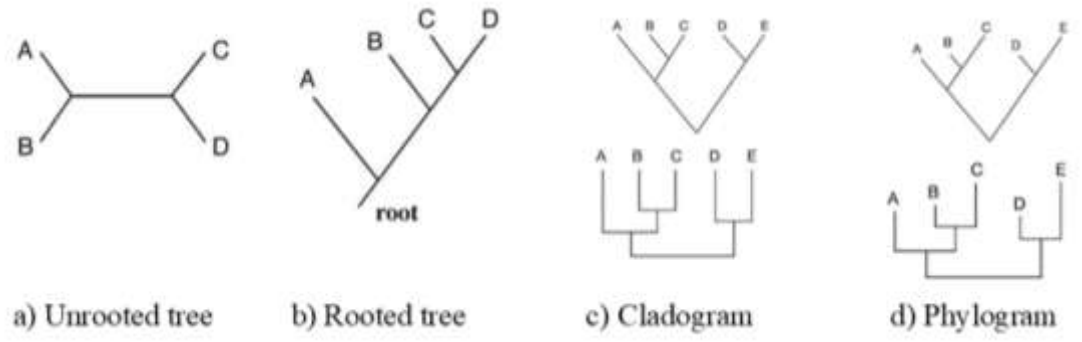



**Figure 2**: Different types of phylogenetic trees

**Phylogenetic Tree Construction and analysis**

Constructing a phylogenetic tree is a multi-step process that involves careful selection of molecular markers, sequence alignment, choice of evolutionary models, and the application of appropriate tree-building methods. Below is an overview of the key steps involved in this process

1. Selection of molecular marker

The first step in constructing a phylogenetic tree is selecting a suitable molecular marker. The choice depends on the nature of the sequences and the objectives of the study. Both nucleotide and protein sequences can be used. For closely related species, nucleotide sequences are often preferred due to their higher variability. In contrast, for distantly related groups, slowly evolving nucleotide sequences or protein sequences are more suitable, as they provide better resolution over longer evolutionary timescales.

2. Sequence alignment

The accuracy of a phylogenetic tree depends heavily on the quality of sequence alignment. This step involves comparing nucleotide or protein sequences by aligning them to identify similarities and differences, that can be scored for analysis.

The first issue to consider is Homology Assessment: Before alignment, it is essential to determine whether the sequences are homologous, meaning they share a common ancestor. Homologous sequences provide a valid basis for phylogenetic analysis.

The next step is Multiple Sequence Alignment: When comparing more than two sequences, a multiple sequence alignment is required. This is typically performed using computational tools. One of the most widely used programs for this purpose is Clustal (Jeanmougin et al., 1998). Additionally, tools like Gblocks (Castresana, 2000) can refine alignments by removing poorly aligned regions and divergent segments, improving the reliability of the analysis.

3. Selection of a model of evolution

The third stage in the construction of a phylogenetic tree involves the careful selection of a suitable evolutionary model. Evolutionary, or substitution, models are sophisticated statistical frameworks that describe the substitution and divergence patterns of sequences over time. There exist various substitution models for both nucleotide and amino acid sequences. Among the nucleotide substitution models, the Jukes-Cantor (JC) model and Kimura's two-parameter model stand out as commonly employed frameworks.

4. Construction of the phylogenetic tree

Phylogenetic trees can be constructed using two primary approaches: distance-based methods and character-based methods. Distance-based methods for constructing phylogenetic trees means calculating evolutionary distances between sequences using substitution models, which are then organized into a distance matrix. This matrix serves as the basis for the construction of phylogenetic tree. Two widely used techniques in this category are the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Neighbor-Joining (NJ). UPGMA function, assuming a constant evolutionary rate, across all taxa, positioning them equidistant from the root and implying a molecular clock. Whereas, the Neighbor-Joining method, the most commonly used distance-based approach, does not assume a molecular clock and generates an unrooted tree.

In distance-based method, degree of divergence between sequences were measured, whereas character-based methods analyze molecular sequences from individual taxa to infer ancestral character states. Although, character-based methods are generally considered as more precise, it is computationally demanding and requires advanced statistical modelling. The two most commonly used character-based techniques are Maximum Parsimony (MP) and Maximum Likelihood (ML).

The Maximum Parsimony (MP) method constructs a phylogenetic tree by selecting the one with fewest evolutionary changes which minimizes the total branch length. This approach is most effective when analyzing closely related sequences or a limited number of taxa. Maximum likelihood is a statistical method that uses probabilistic models to identify the most suitable phylogram that has the maximum probability of generating the observed data. Unlike MP, ML systematically assesses all possible trees, computing the probability for each tree and selecting the one that best explains the evolutionary relationships among the sequences.

**Assessing the accuracy of a reconstructed tree**

The final step in phylogenetic analysis involves assessing the reliability and confidence of the phylogenetic tree. Methods such as bootstrapping, jackknifing, and Bayesian simulation along with statistical methods such as Kishino-Hasegawa test are used to assess the confidence. Among these, bootstrapping is the commonly used statistical approach to measure the robustness of the phylogram topology.

Bootstrapping is a resampling technique used to assess the reliability of a phylogenetic tree. This involves generating multiple subsets of the original sequence data, known as bootstrap samples, which are then used to reconstruct phylogenetic trees using the same method as the initial analysis. Interior branches that consistently appear in the resampled trees are assigned a value of 1. This process is repeated numerous times, and the frequency with which each interior branch is supported is calculated as the bootstrap value or confidence score. Usually, a bootstrap value of 95 or higher is generally considered as reliable, and the values are expressed as percentages on the corresponding tree's branches

MEGA (molecular evolutionary genetics analysis) is a widely used software for phylogenetic analysis, developed by Tamura et al. (2011). It is freely available and user-friendly software. MEGA supports both distance-based and character-based tree construction methods and offers various analytical tools, including heuristic approaches and bootstrapping, to enhance phylogenetic studies.

**Molecular clocks**

The primary goal of phylogenetic analysis is to determine the evolutionary relationships among the DNA sequences being compared. The tree topology is a way of representing these relationships. A secondary objective is often to estimate the timing of divergence events that resulted in the modern sequences. Molecular clocks facilitate this process by assigning dates to branching points within the phylogenetic tree. To achieve accurate dating, the molecular clock must be calibrated by determining the rate of nucleotide substitutions per million years, typically using fossil records as a reference. The rate of molecular change varies across organisms and even within different genetic regions of the same organism. For instance, mitochondrial genes generally evolve at a faster rate than nuclear genes (Strauss, 1999).

For estimating divergence times in datasets that encompass both micro- and macro-evolutionary events, the multispecies coalescent (MSC) model is a suitable approach. However, its application is often limited due to the extensive computational time required for large-scale analyses. As an alternative, researchers commonly use standard Bayesian frameworks that do

not rely on the MSC model, such as those implemented in BEAST software (Bouckaert et al., 2014). Bayesian methods that incorporate coalescent tree priors have been shown to yield reliable molecular dating estimates, with high posterior densities and strong coverage probabilities.

**Applications and limitations of the phylogenetic tree**

Phylogenetic trees serve multiple purposes across various scientific fields, including a) Investigating evolutionary relationships among species and gaining evolutionary insights over time. b) assessing species diversity and distribution, which helps in the development of conservation strategies to protect endangered species and ecosystems. c) to determine the taxonomic classification of organisms based on genetic and morphological data. d) to assist in forensic investigations by determining the origin of biological samples found at crime scenes and linking suspects to criminal activities. e) Systematics. f) assessing the process of cryptic speciation in a species, understanding the history of life and resolving controversial history of life.

Despite its strengths, the molecular systematics has several limitations. Molecular systematics is a cladistic approach in which it is assumed that the classification must correspond to phylogenetic descent and that all valid taxa must be monophyletic. This is a restriction when attempt is made to determine the optimal tree(s), in which which often involves bisecting and reconnecting portions of the phylogenetic tree(s). The latest discovery of an extensive horizontal gene transfer between organisms provides a significant complication for molecular systematics, which indicates that different genes within the same organism can have different phylogenies. In addition, Molecular phylogenies are built on assumptions and models that are sensitive to variations in them. These may result in different results while applying different models to the same dataset.

**Suggested readings**

Fricke, R., Eschmeyer, W. N. & R. van der Laan (eds) 2025. Eschmeyer's catalog of fishes: genera, species, references.

NBFGR 2020. *Annual report 2020*, ICAR-National Bureau of Fish Genetic Resources, Lucknow, Uttar Pradesh, India, 160 pp.

Nelson G. Origin and diversification of teleostean fishes. Ann N Y Acad Sci. 1969;167:18–30.

Betancur-R, R., Wiley, E.O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G. and Orti, G., 2017. Phylogenetic classification of bony fishes. *BMC evolutionary biology*, *17*, pp.1-40.

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., & Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends in biochemical sciences*, *23*(10), 403-405.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, *17*(4), 540-552.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*, *28*(10), 2731-2739.

Brown, T. A. (2002). Molecular phylogenetics. In *Genomes. 2nd edition*. Wiley-Liss.
Strauss E. Can mitochondrial clocks keep time? Science. (1999);283: 1435–1438.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., ... & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology*, *10*(4), e1003537.

# Molecular Systematic Approaches for Phylogenetic Inference

Sandhya Sukumaran,
Marine Biotechnology Fish Nutrition and Health Division, ICAR- Central Marine Fisheries Research Institute, Ernakulam North P.O., Koch, Kerala, India- 682018
sandhya.sukumaran@icar.org.in

## Introduction

As the basis for the existence of species, an organism's evolutionary biology is crucial. Using scientifically proven techniques, systematics investigates the evolutionary relationships between organisms and divides them into distinct groups. They are classified by taking into account their phylogenetic or phenetic traits, giving them names, and grouping them into previously defined categories. In other words, systematics is the study of evolutionary patterns across time. Over the past ten years, enormous progress has been made in molecular biology and bioinformatics, which has produced powerful tools for the study of molecular systematics. Due to its many advantages over conventional morphology-based techniques, analysis of DNA or protein sequence variation has become the standard procedure for many molecular biology investigations. Molecular systematics has several advantages, including universal character types and states (character states are frequently similar due to homology, or inheritance from a common ancestor), increased character state availability, variable substitution rates across gene regions, well-studied molecular evolution, access to sophisticated statistical packages, and ease of data collection from a variety of taxa.

## Phylogenetic inference

Evolutionary models are used in molecular systematics to infer phylogenetic information. The majority of people are either directly or indirectly related to one another through shared ancestors, and the degree of genetic divergence reveals information about the kind and degree of relatedness within and between populations. Molecular phylogenetic inference begins with obtaining sequence information by sequencing a specific gene or by retrieving it from NCBI GenBank. After the sequences are aligned, mutations such as insertions or deletions can be found. To make the sequences more similar to one another, gaps could be added if necessary.
In addition to sequence data and alignment, a suitable model of sequence evolution must also be selected as there are several assumptions about substitution processes in molecular phylogenetics.

**Standard practices for phylogenetic inference**

Four approaches have been used extensively for phylogenetic inference: maximum parsimony, neighbor-joining, maximum likelihood, and Bayesian inference.

Compared to distance methods, maximum parsimony is one of the earliest inference techniques for phylogeny reconstruction that makes use of character states. It is an approach with the fewest presumptions. Since it does not estimate branch lengths and does not require any parameters, parsimony is a non-parametric approach. By choosing the fewest character state changes while minimizing homoplasy, the optimality criterion determines the optimal alternative tree. Fitch's algorithm, which traverses the tree and assigns one or more-character states to each internal node, is used to directly calculate the length of an unrooted tree.

**Distance trees**

Pairwise differences in species' sequences are the main focus of distance methods. Depending on the evolutionary model and the hierarchical relationships between species, a distance matrix is defined. Compared to distantly related species, closely related species exhibit greater similarity and higher similarity indices. However, the problem with these approaches is that a single distance matrix can mathematically define multiple trees, and we need algorithms to determine which tree is best for a given matrix. Distantly related species are divided, and the most closely related species are grouped.

One popular distance method is neighbor joining. The neighbor joining method is predicated on the idea that any sequence's dissimilarity is directly related to its phylogenetic relationship. The evolutionary distance between two species is the source of dissimilarity. After converting DNA or amino acid sequences to a distance matrix, a phylogenetic tree is investigated. The squared deviation of the pairwise observed distances between each pair of taxa estimated from the data matrix and the distance separating those taxa on the tree is used in optimality methods to determine a tree's score

The estimation of a likelihood function, which explains how a particular tree can be fitted to the sequence data, is the foundation of maximum likelihood. For character state evolution, an explicit model is selected, and a phylogenetic tree with branch lengths is suggested. The probability of observing the data given the tree and the evolutionary model is known as the likelihood of a phylogenetic tree. An optimality method is also used in this. The phylogenetic hypothesis with the highest chance of generating current sequences under the probabilistic

model of sequence evolution under consideration can be estimated by finding the tree with the highest likelihood.

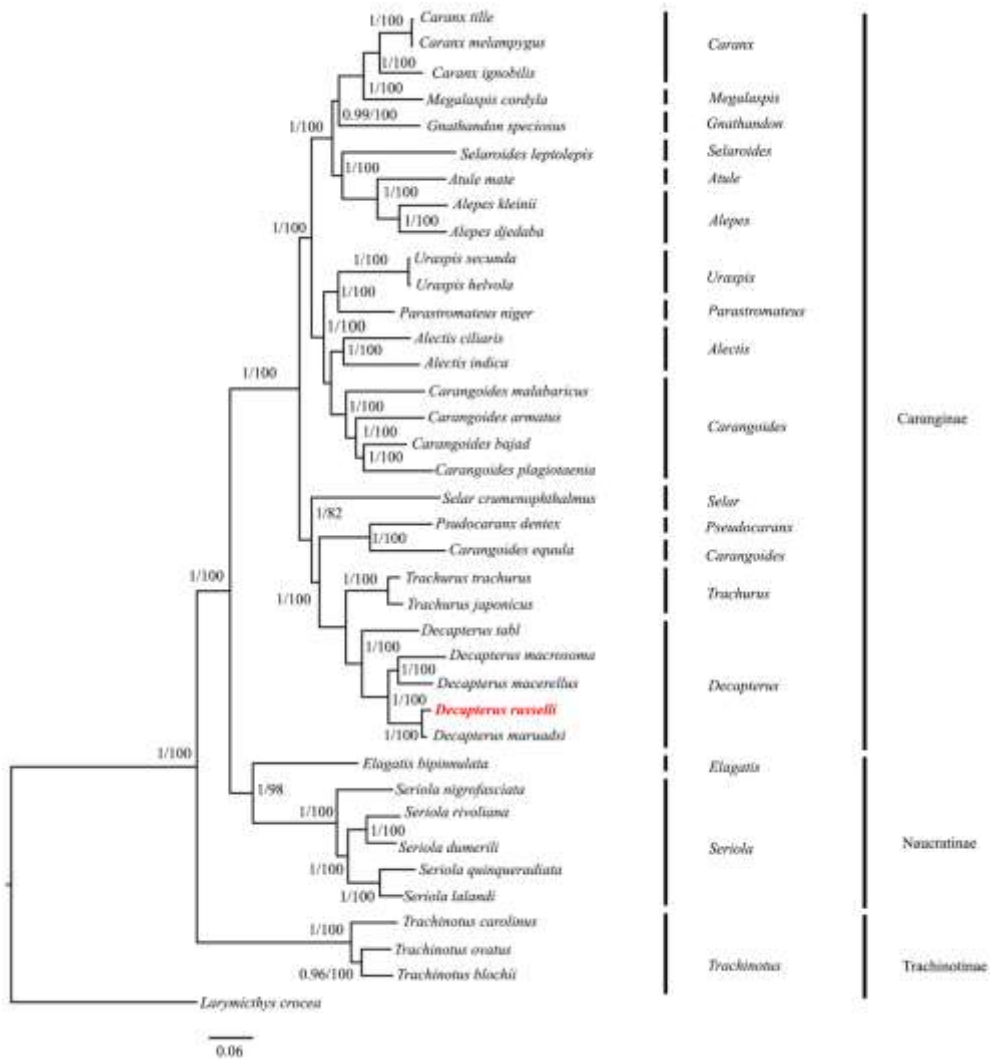



**Figure 1**: The phylogenetic tree showing the relationship between Carangids

The most recent phylogenetic inference technique is Bayesian inference, which is based on the Bayesian theorem. The best hypothesis is the one that can maximize the posterior probability, and it is most closely associated with the maximum likelihood approach. The likelihood multiplied by the prior probability of a specific hypothesis determines the posterior probability, according to Bayes' theorem. Complex models of sequence evolution can be used in Bayesian analysis for both the entire sequence data set and its various partitions.

A Monte Carlo simulation is used to generate the data in the parametric bootstrapping method. A simulation is performed using replicate data sets of the same size as the original based on the

null hypothesis being tested. For each replicate data set, likelihoods according to both the null and alternative hypotheses are estimated, and the LRT (Likelihood Ratio Test) statistic is derived and a significance test is implemented. Parametric bootstrap is computationally very demanding when the data sets are large. Non-parametric tests appear to be conservative, and parametric tests appear to be liberal in empirical comparisons. But generally, to assess branch support for phylogenetic trees, non-parametric bootstrapping is used.

**Conclusion**

Molecular systematics have made significant contributions to the fields of taxonomy and evolutionary biology. This field of study is also benefiting greatly from next-generation sequencing techniques, but handling massive amounts of data and integrating information at various genome levels will be difficult.

**Key references**

Askari, G.H., Shabani, H. and Miandare, H.K. Application of molecular markers in fisheries and aquaculture. Scientific Journal of Animal Science, 2(4): 82-88.

Chauhan, T and Rajiv. K. 2010. Molecular markers and their application is fisheries and aquaculture. Advances in Bioscience and Biotechnology 1: 281-291.

Davis, G.P. and Hetzel, D.J.S. 2000. Integrating molecular genetics technology with traditional approaches for genetic improvement in aquaculture species. Aquaculture Research, 31: 3-10.

Edwards, S.V. 2008. Is a new and general theory of molecular systematics emerging? Evolution 63(1): 1-19.

Hallerman, E.M. 2006. Use of molecular tools for research and improvement of aquaculture stocks. The Israeli Journal of Aquaculture – Bamidgeh, 58(4): 286-296.

Hillis, D.M. 1987. Molecular versus morphological approaches to systematic. Annual Reviews in Ecology and Systematics 18: 23-42.

Liu, Z.J. and Cordes, J.F. 2004. DNA marker technologies and their applications in aquaculture genetics. Aquaculture 1-37.

Mauro, D.S. and Agorreta, A. 2010. Molecular systematic: A synthesis of the common methods and the state of knowledge. Cellular and Molecular Biology Letters 311-341.

Okumus, I. and Ciftci, Y. 2003. Fish population genetics and molecular markers: II- Molecular markers and their applications in fisheries and aquaculture. Turkish Journal of Fisheries and Aquatic Sciences 3: 51-79.

Presti, R.L., Lisa,C. and Stasio, L.D. 2009. Molecular genetics in aquaculture. Italian Journal of Animal Science 8: 299-313.

Schwartz, J.H. and Maresca, B. 2006. Do molecular clocks run at all? A critique of molecular systematics. Biological Theory 1(4): 357-371.

Teletchea, F. 2009. Molecular identification methods of fish species: reassessment and possible applications. Reviews in Fish Biology and Fisheries 19: 265- 293.

# Demystifying Quantitative PCR: Working Principles, Technical Insights, and Common Sources of Error

Adnan H. Gora, Saima Rehman, Chandrasekar S., D. Linga Prabu, Sanal Ebeneezar and Qurat ul Ain, & Sajina K.A, *ICAR-Central Marine Fisheries Research Institute, Kochi, India*

**Abstract**

Quantitative PCR (qPCR) has emerged as a cornerstone technique in molecular biology, revolutionizing nucleic acid research through its ability to detect and quantify nucleic acids with exceptional precision and sensitivity. Building upon the foundational principles of conventional PCR, qPCR integrates real-time fluorescence detection, enabling simultaneous amplification and quantification of target DNA or RNA. This innovation not only mitigates contamination risks inherent in earlier PCR methodologies but also allows for dynamic, high-throughput analysis across diverse applications, from gene expression studies to clinical diagnostics.

The transformative potential of qPCR lies in its ability to detect differences as small as twofold between biological groups, with a quantification limit as low as 100 copies. However, its precision is contingent upon rigorous experimental design, including optimized reaction conditions, validated primers, and the inclusion of appropriate controls. Biological variability, timing of sample collection, and tissue-specific transcriptional profiles further underscore the complexity of qPCR studies, necessitating meticulous planning to ensure reproducibility and accuracy.

Despite its versatility, common sources of error, such as subsampling inaccuracies, measurement uncertainties, and improper primer validation, remain challenges. Addressing these issues through adherence to best practices is critical for generating reliable data that reflects the true biological context. Establishing standardized protocols and fostering transparency in qPCR experiments are essential for advancing scientific knowledge and ensuring the reproducibility of research findings.

**Keywords:** RT-qPCR, qPCR challenges, relative quantification, targeted gene expression

## Introduction

The advent of quantitative polymerase chain reaction (qPCR) represents a transformative milestone in molecular biology. The journey of this technology begins with the invention of

PCR by Kary Mullis in the early 1980s, an achievement that earned him the Nobel Prize in Chemistry (Morris, 2021). Despite the PCR platforms being a very common presence in most molecular biology laboratories around the world, the widespread adoption of PCR was delayed until the late 1990s due to intrinsic limitations (Zhu et al. 2020). Early PCR methods were labor-intensive, requiring post-amplification processes such as electrophoresis or high-performance liquid chromatography (HPLC) for the verification of amplification (Higuchi et al. 1992). Furthermore, the necessity to manipulate amplified DNA often resulted in significant contamination risks, particularly in high-throughput laboratory settings. In response to these challenges, efforts were undertaken to enhance the efficiency and practicality of PCR platforms that lead to the developments culminating in the emergence of qPCR, a method that integrates amplification with real-time detection, revolutionizing nucleic acid research.

The conceptual and technological advancements that underpin qPCR can be traced to pivotal work undertaken by researchers at F. Hoffmann-La Roche AG, Basel (commonly known as Roche) in the early 1990s. In 1992, they demonstrated that amplification could be confirmed without removing DNA from the reaction vessel by incorporating ethidium bromide, a fluorescent intercalating agent, into the reaction mixture (Higuchi et al. 1992). This innovation eliminated the need for physical manipulation of amplified DNA, thereby mitigating contamination risks. The researchers further refined the technique by automating fluorescence detection by using a camera with the thermocycler to capture fluorescence signals during each cycle, enabling real-time monitoring of the amplification process. Their observations revealed that the fluorescence intensity directly correlated both the number of amplification cycles and the initial concentration of DNA template (Higuchi et al. 1993), an observation that provided a novel means to quantify DNA (Figure 1). Results obtained with this study indicated that a kinetic approach to PCR analysis can quantify DNA sensitively and selectively and over a large dynamic range. This approach also provides a means of determining the effect of different reaction conditions like amount of target DNA and number of PCR cycles on the efficacy of the amplification process.

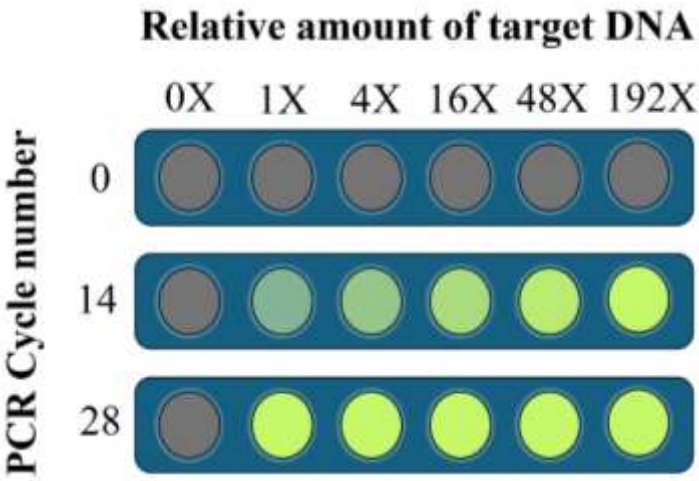



**Figure 1:** Variation in the Ethidium bromide-induced fluorescence in the PCR with reaction conditions like the number of cycles and amount of template DNA. The figure is created based on the original photograph in (Higuchi et al. 1993)

**Fundamentals of PCR and qPCR**

The basic principles of qPCR build upon those of conventional PCR, which operates through three cyclic phases: denaturation, annealing, and extension. In each cycle, the number of DNA molecules theoretically doubles, following an exponential amplification pattern under ideal conditions. However, practical limitations, such as reaction efficiency and template quality, often lead to deviations from this idealized model. qPCR distinguishes itself by incorporating fluorescent dyes, such as ethidium bromide or SYBR Green (Gudnason et al. 2007), which intercalate with double-stranded DNA (dsDNA). This fluorescence serves as a real-time indicator of amplification. Unlike traditional PCR, where the amplification product is detected only at the end of the reaction, qPCR monitors fluorescence at each cycle. The cycle at which fluorescence surpasses a predefined threshold (Figure 2) is referred to as the quantification cycle (Cq). This parameter forms the basis for quantitative analysis, offering insights into initial DNA concentration and reaction efficiency. One of the most prominent applications of qPCR lies in the comparative analysis of gene expression. When multiple samples are analyzed simultaneously, the fluorescence intensity reflects the initial concentration of the DNA template. Samples with higher template concentrations achieve detectable fluorescence earlier than those with lower initial concentrations. Importantly, while fluorescence levels ultimately converge by the end of the reaction, differences in Cq values serve as a basis for quantitative

measures of gene expression (Ruiz-Villalba et al. 2021). This capability has rendered qPCR an indispensable tool in fields ranging from genomics to clinical diagnostics.

**Critical Considerations in qPCR Experimental Design**

The successful implementation of qPCR necessitates meticulous planning and adherence to several critical parameters. Among these, the consistency of replicates, optimization of reaction

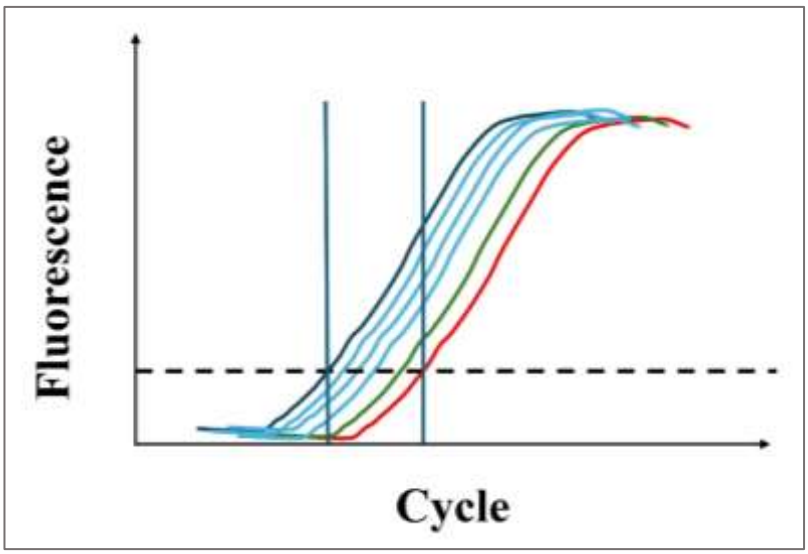



**Figure 2:** A typical amplification curve observed after running six samples with varying amounts of template in the qPCR. The blue curve represents the sample with more efficient PCR or higher amounts of template DNA, therefore causing an early rise of the amplification curve. Note that the fluorescent plateaus near the end of the PCR reaction in all samples where amplification has occurred.

conditions, and the inclusion of appropriate controls are paramount. Rigorous attention to these factors ensures the reliability and reproducibility of qPCR results, which are essential for robust scientific inquiry.

**Biological Variability**

The most challenging aspect of many gene expression experiments is the appropriate selection of biological replicates per biological group that ensure minimizing the effects of inherent variability in biological samples (Fraser, 2004). Gene expression profiles are highly sensitive to the inherent biological differences between individual samples so much so that gross differences in transcription profiles have been demonstrated between individual cells plated from the same passage in a single petri dish (Moignard and Gottgens, 2014). Therefore, careful thought and planning must go into the randomized selection, and number of samples per

biological group to ensure statistically significant and reproducible results that give precision and accuracy. In general, depending upon the complexity of samples and the strictness of experimental control exerted, the number of biological replicates can vary between >4 for *in vitro* cell based assays to >50 in case of human trials (Taylor et al. 2019).

**Time**

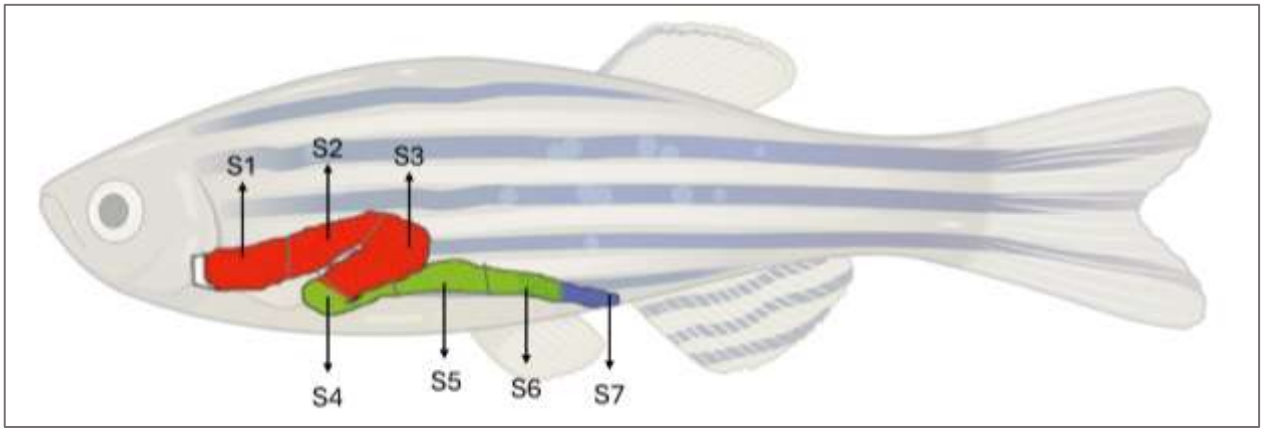
Time of sampling following treatment is a pivotal factor that profoundly influences the outcome of a qPCR study. Biological experiments frequently aim to unravel dynamic processes where gene expression profiles change over time. For example, the inflammatory response in the zebrafish brain exhibits a well-defined time-dependent pattern (Zheng et al., 2017). In such scenarios, it is crucial to meticulously plan sampling points to align with the transcriptional windows of the target genes. For genes that exhibit rapid or transient expression changes (e.g., hourly or daily fluctuations), conducting preliminary standardization experiments becomes an essential practice. These preliminary studies enable researchers to map the temporal expression patterns of target genes in response to specific treatments, ensuring that sampling captures the relevant transcriptional phases. Without such standardization, there is a heightened risk of obtaining inaccurate or irreproducible results, as the timing of sampling may fail to reflect the true biological response. In essence, careful consideration and calibration of sampling intervals, based on the kinetics of gene expression, are indispensable for generating reliable and meaningful qPCR data.

**Tissue sections**

A tissue is a complex of varied cell populations that are phenotypically oriented towards performing one or more specific physiological functions. Most tissues are also regionally specified, for instance the intestine and the brain in which the different sections of the tissue performed different functions and thus exhibit unique transcriptional profiles (Thompson et al. 2014, Wallace et al. 2005). In the zebrafish, it has been observed that even though the gut is morphologically divided into three regions (anterior, middle and posterior), the tissue can

divided into 7 different regions (Figure 3) based on the gene expression profile (Wang et al. 2010). This

**Figure 3: A graphical representation of functional organization of the zebrafish gut along**




**rostrocaudal axis.** The zebrafish gut is differentiated into seven unique regions based on the gene expression profile.

regional differentiation of tissues makes it critical to understand the precise tissue-specific location of the expression of the gene under study.

**Common Sources of error in qPCR studies**

Subsampling error

A subsample is a small portion of the original sample that is used as a template in PCR amplification to evaluate gene expression. The accuracy of PCR results heavily depends on the representativeness of the subsample. However, when the concentration of target DNA in the original sample is extremely low, the subsample may not adequately reflect the overall composition of the sample. This discrepancy introduces a significant source of error known as subsampling error. Subsampling error occurs when there is a considerable variation in PCR amplification among technical replicates derived from the same original sample. Such variability arises when the amount of target DNA template in the subsample is insufficient for consistent amplification. In cases where the DNA copy number is low, random differences in the distribution of target molecules within the subsample become more pronounced, leading to uneven amplification. Subsampling error is quantitatively significant and can severely affect the reliability of PCR data. For instance, when the DNA copy number in the subsample is below 100, subsampling error contributes to more than 10% of the total variance observed in the amplification results. As the copy number decreases further, the impact of subsampling error becomes even more pronounced. When the subsample contains fewer than ten copies of

the target DNA, the error contribution exceeds 30%, leading to highly variable and often unreliable amplification profiles (Taylor et al. 2019). This phenomenon underscores the importance of ensuring an adequate DNA concentration in subsamples to minimize variability and improve the accuracy of PCR-based gene expression analyses. Strategies such as increasing the initial amount of template DNA, pooling multiple subsamples, or employing more sensitive amplification techniques can help mitigate subsampling errors and enhance the reliability of experimental results.

Measurement Uncertainty

Measurement uncertainty in qPCR arises from the inherent randomness of the PCR amplification process. This variability is primarily a result of the stochastic nature of amplification, which is influenced by factors such as template DNA limitation, random primer binding, and amplification noise. These elements collectively introduce variability into the reaction, which can affect the precision and reliability of qPCR results. Among these factors, template DNA limitation plays a crucial role in determining the amplification efficiency in a qPCR reaction. The probability of successful amplification and doubling template molecules during each cycle depends significantly on the initial quantity of template DNA in the reaction. When the template DNA concentration is low, the likelihood that all template molecules will successfully bind to primers and initiate complementary strand synthesis is reduced (Figure 4). This limitation leads to significant variability in amplification efficiency across technical replicates and becomes particularly pronounced when the template DNA quantity is below 100 copies per reaction, corresponding to Cq values greater than 29 cycles. Under these conditions, the coefficient of variation (CV) among technical replicates can range widely, from 10% to as much as 200%, reflecting the challenges of consistent amplification at low template concentrations. To mitigate measurement uncertainty and subsampling error in qPCR, several strategies can be employed. Increasing the number of technical replicates is an effective approach, as it improves the statistical reliability of the results by averaging out random fluctuations. Another strategy is pre-amplification, which involves an initial amplification step to increase the quantity of template DNA before performing qPCR.

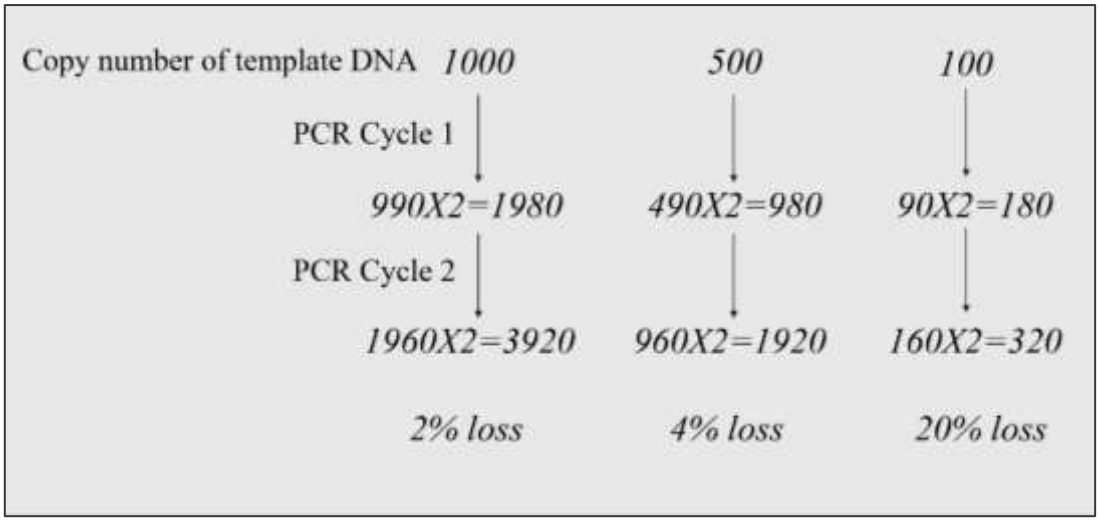



**Figure 4:** A visual representation of measurement uncertainty.

The lower the amount of template, the lower is the probability of binding the template with the primer. When the copy number of template is 1000, the loss after 2 amplification cycles in 2%, but it increases to 20% if the template has a copy number of 100.

This strategy ensures that the template DNA is present enough to reduce variability and improve the accuracy of amplification. Together, these measures enhance the robustness of qPCR data, particularly when working with low-template samples.

Inappropriate primer validation

Non-validated primers can result in variable and unreliable data that do not accurately reflect the true abundance of the target gene in each sample. This variability arises from factors such as amplification inefficiency, suboptimal annealing temperatures, and the amplification of non-specific products during a qPCR reaction. These issues highlight the critical importance of primer validation in ensuring accurate, reproducible, and biologically meaningful qPCR results. Primer validation begins with a thermal gradient PCR, an essential step to determine the optimal annealing temperature for efficient amplification. The annealing temperature is typically identified as the point where the lowest Cq (quantification cycle) values are observed, indicating the most efficient binding of primers to the target sequence. Efficient amplification not only minimizes background noise but also ensures the accuracy of quantification. The validation process also includes verifying the amplicon size by visualizing the PCR product on an agarose gel. This step confirms that the amplified product corresponds to the expected size, ruling out non-specific amplification or primer dimer formation. Additionally, a multi-point standard curve is generated using serial dilutions of the DNA template to assess the efficiency of the primer pair. According to Pfaffl (2001), primer efficiency should ideally fall within the range of 90–110%, with an $R^2$ value close to 1. This ensures that the amplification process is

both consistent and reliable across a wide range of template concentrations. To further confirm primer specificity, sequencing the amplified product through Sanger sequencing is highly recommended. This step verifies that the correct gene has been amplified, eliminating the possibility of non-target amplification. Melt curve analysis, a widely used and straightforward technique, complements these steps by confirming the specificity of the amplified product. It must indicate a single sharp melting peak corresponding to the target gene while detecting any additional peaks is indicative of primer dimers or non-specific amplification. Together, these rigorous validation steps are critical for ensuring that qPCR experiments produce accurate, reproducible, and biologically relevant results. Neglecting proper primer validation can compromise data integrity and lead to erroneous interpretations, underscoring its indispensable role in molecular biology workflows.

*Selection of internal reference genes*

Accurate normalization of qPCR data is essential for generating reliable and biologically meaningful results. At the core of this process lies the selection of stable and appropriate internal reference genes (often referred to as housekeeping genes). These genes serve as endogenous controls, compensating for variability in RNA quantity, quality, and reverse transcription efficiency across samples. Selecting suitable reference genes is a pivotal step, as improper internal controls can introduce bias and undermine the data interpretation. An ideal reference gene embodies several critical characteristics that ensure its suitability for normalization. It must exhibit consistent expression across the experimental conditions, treatments, and sample types, remaining unaffected by biological or environmental variations. Moreover, the expression levels of the reference genes should align closely with those of the target genes to ensure precise normalization. Crucially, the expression of the internal reference gene must be unregulated by experimental conditions, thereby providing a steadfast baseline for comparative analysis. Traditional housekeeping genes, such as *gapdh* (*glyceraldehyde-3-phosphate dehydrogenase*), *actb* (*beta-actin*), and 18S r*RNA* (*18S ribosomal rna*), have long been employed as internal controls. However, emerging evidence indicates that these genes do not always maintain stable expression under diverse experimental conditions, emphasizing the necessity of rigorous validation for each study (Sikand et al. 2012). The selection and validation of reference genes involve several meticulous steps. Initially, a panel of candidate reference genes is identified based on insights from existing literature or relevant databases. These candidates are then subjected to qPCR assays to assess their expression stability across all experimental conditions and sample types. Advanced statistical tools, such as geNorm,

NormFinder, or BestKeeper, must be used to evaluate and rank the stability of the candidate genes (Zhong et al. 2022). These algorithms provide metrics such as stability values, geometric means, or pairwise variations, which guide the identification of the most reliable reference genes for normalization. While a single reference gene may suffice in certain cases, exclusive reliance on one gene risks introducing bias if it exhibits even subtle variability under specific conditions. To enhance the robustness and reliability of normalization, it is advisable to use two or more validated reference genes. This approach ensures a more comprehensive and accurate normalization framework, reducing the likelihood of error. Several considerations are essential when selecting and validating reference genes. Validation should be conducted for every new type of tissue, species, or experimental condition to confirm their suitability. A gene that is stable in one tissue or species may display variability in another, necessitating independent validation. The application of robust statistical methods ensures that the chosen reference genes fulfill the necessary criteria for stability and consistency. Beyond qPCR assays and statistical evaluations, complementary techniques can further confirm the reliability of reference genes. For example, RNA-seq data offers a broader perspective on gene expression stability across experimental conditions, aiding in the identification of novel candidate genes for validation. The selection and validation of internal reference genes constitute an indispensable element of qPCR experimental design. Employing stable, unregulated, and appropriately expressed reference genes ensures accurate normalization, enabling precise quantification of target gene expression. By enhancing data reliability and contributing to the reproducibility and credibility of qPCR studies, validated reference genes serve as a cornerstone for uncovering meaningful biological insights.

## Conclusion

Quantitative PCR is a widely recognized technique for the detection and quantification of nucleic acids, capable of identifying differences as small as twofold between biological groups with exceptional precision. Additionally, it allows for a lower limit of quantification down to approximately 100 copies of the template, making it invaluable for both gene expression analysis and absolute quantification studies. However, despite its versatility and sensitivity, the method comes with certain limitations, primarily due to the lack of automation and the necessity for multiple standardizations throughout the process. Consequently, conducting qPCR experiments requires strict adherence to a meticulously planned workflow, including experimental design, sample verification, and primer validation. Failure to follow this rigorous

procedure can lead to significant issues, such as the generation of non-reproducible or artefactual data. These inaccuracies can misrepresent the biological reality of the samples under investigation, resulting in erroneous interpretations and highly variable results that compromise the validity of the research. These challenges emphasize the need for adherence to best practices to ensure that qPCR experiments yield reliable and reproducible outcomes. Moreover, the scientific community must address important questions regarding the implementation and enforcement of these best practices. Establishing standardized guidelines is critical for maintaining the integrity of data and ensuring consistency across studies. Researchers must also prioritize experimental transparency and accuracy, as foundational tenets to advancing scientific knowledge and supporting reproducibility in molecular biology and related fields.

**References**

Fraser, C. G. (2004). Inherent biological variation and reference values. Clinical Chemistry and Laboratory Medicine (CCLM), 42(7), 758-764.

Gudnason, H., Dufva, M., Bang, D. D., & Wolff, A. (2007). Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature. *Nucleic acids research*, *35*(19), e127.

Higuchi, R., Dollinger, G., Walsh, P. S., & Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences. Bio/technology, 10(4), 413-417.

Higuchi, R., Fockler, C., Dollinger, G., & Watson, R. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Bio/technology*, *11*(9), 1026-1030.

Moignard, V. and Gottgens, B. (2014) Transcriptional mechanisms of cell fate decisions revealed by single cell expression profiling. Bioessays 36, 419–426 15.

Morris, T. (2021). NOBEL LIFE: Conversations with 24 Nobel Laureates on their life stories, advice for future generations and what remains to be discovered. TLS. Times Literary Supplement, (6173), 24-25.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic acids research*, *29*(9), e45-e45.

Ruiz-Villalba, A., Ruijter, J. M., & van den Hoff, M. J. (2021). Use and misuse of Cq in qPCR data analysis and reporting. Life, 11(6), 496.

Sikand, K., Singh, J., Ebron, J. S., & Shukla, G. C. (2012). Housekeeping gene selection advisory: glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and β-actin are targets of miR-644a.

Taylor, S. C., Nadeau, K., Abbasi, M., Lachance, C., Nguyen, M., & Fenrich, J. (2019). The ultimate qPCR experiment: producing publication quality, reproducible data the first time. Trends in biotechnology, 37(7), 761-774.

Thompson, C. L., Ng, L., Menon, V., Martinez, S., Lee, C. K., Glattfelder, K., ... & Jones, A. R. (2014). A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. Neuron, 83(2), 309-323.

Wallace, K. N., Akhter, S., Smith, E. M., Lorent, K., & Pack, M. (2005). Intestinal growth and differentiation in zebrafish. Mechanisms of development, 122(2), 157-173.

Wang, Z., Du, J., Lam, S. H., Mathavan, S., Matsudaira, P., & Gong, Z. (2010). Morphological and molecular evidence for functional organization along the rostrocaudal axis of the adult zebrafish intestine. BMC genomics, 11, 1-13.

Zheng, J. L., Yuan, S. S., Wu, C. W., Lv, Z. M., & Zhu, A. Y. (2017). Circadian time-dependent antioxidant and inflammatory responses to acute cadmium exposure in the brain of zebrafish. Aquatic Toxicology, 182, 113-119.

Zhong, Z., Ao, L., Zhao, L., Zhang, Z., & Jiang, Y. (2022). Screening and validation of reference genes for qPCR analysis in gonads and embryos of Takifugu bimaculatus. Aquaculture and Fisheries, 7(3), 278-286.

Zhu, H., Zhang, H., Xu, Y., Laššáková, S., Korabečná, M., & Neužil, P. (2020). PCR past, present and future. *Biotechniques*, *69*(4), 317-325.

# Protein structure modelling and molecular docking: Principles and practical guidelines

Lailaja VP*, Sumithra TG, Krupesha Sharma

Marine biotechnology fish nutrition and health division., ICAR-CMFRI, Kochi-18, *lailaja@gmail.com

## Introduction

Proteins are vital macromolecules in biological systems and perform a variety of critical roles. The correct folding of the protein sequence into its tertiary and quaternary structures by achieving the lowest free energy determines its biological function. Understanding the structure and folding of a protein is necessary for determining its function. Computational methods such as protein modelling and molecular docking shed light on the structure and interactions of proteins. Drug development, biochemistry, and molecular biology studies depend on the knowledge of the structure and function of proteins. X-ray crystallography, nuclear magnetic resonance (NMR), and electron microscopy (EM) are used to identify the protein structure experimentally. However, due to the constraints of being time-consuming, expensive and fluctuating results, protein structure studies have shifted towards using computational models that can predict 3D protein structure with the available protein sequence (Nassar et al. 2021). Protein modelling and molecular docking are computational techniques that provide insights into protein structure and its interactions, facilitating the design of new therapeutic agents and the study of biological systems. Diverse algorithms, web-based platforms and software tools have been created to enhance the prediction of protein structures, structure analysis, and docking studies. The significance of these initiatives is highlighted by their potential impact on various applications, such as rational drug design, studies of mutations, structural comparisons, evolutionary analyses, and investigations into protein folding *etc*.

## Principles of protein modelling

The three-dimensional arrangement of a protein is vital for the detailed elucidation of its biochemical functions and interaction properties at the molecular level. However, it is noteworthy that the total number of protein sequences elucidated is substantially greater than the number of protein structures that have been resolved through experimental techniques. Generally, template-based and free modelling based on the usage or non-usage of available

protein templates are used to develop or predict a protein structure (Paiva et al. 2022). Further, there are intermediary prediction/hybrid methods that are based on both principles. Within template-based and free modelling groups, there are three algorithms, viz. *ab initio* (a free modelling procedure), threading/fold recognition (a template-based modelling procedure), and homology (template-based modelling procedure) (Agnihotry et al. 2022). Several computer tools and web servers automate modelling. CAMEO and biannual Critical Assessment of Techniques for Protein Structure Prediction (CASP) are used to validate the accuracy of model calculations by these servers (Haas et al. 2013).

**Homology modelling:** Any two amino acid sequences that share a homology in their sequences are likely to have a highly similar structural feature (Agnihotry et al. 2022). Comparative or homology modelling frequently yields a helpful three-dimensional model for a protein related to a minimum one-identified protein structure when an empirically determined structure is unavailable. A homology model for a query protein sequence is generated when the query sequence gets aligned to a protein sequence having an elucidated structure. The atomic model is created by developing loop regions and sidechains based on the available percentage of sequence identity to the selected template. In the last stage predicting model errors and submitting the model to energy minimization are done. The generated model is validated using various validation tools. The major computer programs and web servers used for the comparative modelling process are summarized below.

- **SWISS-MODEL** (swissmodel.expasy.org) is an automated tool that employs homology modelling techniques to forecast the three-dimensional structure of proteins. The Swiss model assimilates and justifies steps involved in the modelling process, establishing a completely automated workflow based on a PERL framework. A protein with an unidentified structure (target) sharing sequence homology with a protein of identified structure (template), helps to infer the structure of the target protein from the selected template (Kiefer et al. 2009).
- **ModeBase:** To create annotated comparative protein structure models, the ModBase database (http://salilab.org/modbase) uses ModPipe, a computerized pipeline that mainly uses Modeller (http://salilab.org/modeller/) for fold assignment, sequence-structure alignment, model creation, and evaluation. Through the ModWeb modelling server, users can demand the modelling for extra sequences and update models as needed. The ModBase gateway and the modelling Portal provide access to ModBase models. ModWeb, a key part of ModBase, accepts FASTA-format sequences for

modelling or can accept a user-provided protein structure to identify homologous sequences and model them based on the input structure (Eswar et al.2006).

- **ModPipe** (Smith et al. 1981): To find the ideal template structure, Modpipe employs various alignment techniques, including sequence-sequence, sequence-profile, and profile-profile. The Discrete Optimized Protein Energy (DOPE), statistical potential, and other quality metrics like sequence identity, GA341 score, Z-DOPE score, and ModPipe Quality Score (MPQS) are used to construct and analyse models. The MPQS integrates factors like sequence coverage, identity, gaps, model compactness, and statistical potential.

***Ab initio* protein modelling:** The *ab initio* modelling methodologies are grounded in the thermodynamic principle that the native conformation of a protein corresponds to the minimum free energy state (Agnihotry et al. 2022). To predict new protein folds, these techniques consider several physicochemical properties related to the folding process, such as hydrogen bonding, contact potential energy, and secondary structure traits gathered from Protein Data Bank (PDB), and the interactions that occur between bonded and non-bonded elements. The chief benefit of *ab initio* methods is the ability to find new protein folds. QUARK (Xu and Zhang 2012) follows *ab initio* protein modelling. To construct a comprehensive model, the application splits the protein sequence into 20 amino acid fragments, looks up the structure in the database, and subsequently employs replica-exchange Monte Carlo simulations to combine the pieces while taking the force field and energy terms into account.

**Threading and fold recognition:** Threading and fold detection techniques are developed based on the idea that a protein conformation is far more preserved compared to its fundamental amino acid sequences and that there is a limited number of structural folds are seen naturally (Agnihotry et al. 2022). These approaches involve selecting the most suitable three-dimensional template from known folds that aligns effectively with the query sequence, utilizing a grading scheme that incorporates pairwise potentials, secondary structure comparisons, and solvent characteristics. Consequently, the query protein sequence is correlated with the protein structure yielding the highest score, allowing for rearranging the target sequence's atoms according to the template. Compatibility of the query protein template with the deduced protein structural fold is subsequently assessed, followed by a manual review. Gen THREADER is an example of a computational tool that uses these algorithms.

**Hybrid modelling:** To refine existing modelling techniques with artificial intelligence, protein modelling platforms have been developed that combine template-based and free

modelling techniques. This strategy was made feasible by the advancement of fast computer processing. This method combines the joint efforts of already available computer tools for protein designing with that of artificial intelligence. The popular computer programs using hybrid modelling are summarized below.

- **I-TASSER** (Iterative Threading ASSEmbly Refinement): An advanced computational tool for predicting three-dimensional structure, function, and interactions of proteins. It employs a hybrid approach combining threading templates, iterative fragment assembly simulations, and structural refinement to construct accurate protein models. I-TASSER also predicts biological functions, such as ligand-binding sites and gene ontology terms, using structural alignments and function databases. It is primarily a comparative modelling tool that uses a threading-based algorithm to predict protein structures. While it incorporates elements of *ab initio* modelling during the assembly and refinement stages for regions without template information, it uses the PDB to identify structural templates from existing structures of proteins. Widely recognized for its high performance, I-TASSER has consistently ranked among the best methods in the Critical Assessment of Structure Prediction (CASP) trials (Yang et al. 2015).
- **AlphaFold2**: It is one of the most advanced algorithms for predicting protein structures. In 2018, AlphaFold gained prominence through Google startup DeepMind. This program was highlighted software for the 13th CASP edition. AlphaFold's initial iteration trained a neural network on protein sequences using deep learning to identify protein structures. To anticipate the spacing connecting residues, it employs a convolutional neural network developed on PDB structures. The algorithm uses multiple sequence alignment (MSA) data to create a datagram from the amino acid sequence and estimates backbone torsion distribution with a separate prediction network. Gradient descent optimizes both components, leading to accurate protein structure prediction (Senior et al. 2020). AlphaFold2 outperforms other methods, achieving low RMSD between predicted and experimental structures using advanced neural networks and incorporating evolutionary, physical, and geometrical limitations (Callaway 2022). It inputs the amino acid sequence and generates MSA from several protein sequence databases. This aids in identifying mutation-prone sections of the sequence and detecting relationships among them. It also searches for proteins with comparable structures to the input, which are then utilized to create a preliminary

template representing the input sequence. AlphaFold-Multimer is DeepMind's extension of this platform to many peptide chains. In 'AlphaFold-Multimer' the homomeric interfaces performed better than heteromeric ones (Yin et al. 2022). AlphaFold2 struggles with disorganized protein domains and loops, relying on MSA and known structure databases. It requires significant processing power, especially for longer proteins. ColabFold addresses this constraint using MMseq2 to reduce processing needs (Bertoline et al. 2023).

- **Protein structure prediction through the protein language model**: Protein language model techniques address AlphaFold2's restrictions, such as identifying new structures and long computation times. These models use natural language processing to analyze amino acid sequences and predict structure based on evolutionary, structural, and functional patterns. ESMfold (Lin et al. 2022) outperforms AlphaFold2 in speed and confidence by predicting protein structures from amino acid sequences using a masked transformer model with 15 billion parameters. While EMBER3D predicts 3D structures and outperforms ESMfold by correlating native and mutant structures, EMBER2 (Weissenow et al. 2022) employs embeddings and attention heads to predict 2D structures without MSA. By utilizing evolutionary, structural, and functional designs from sequence databases, these models improve prediction speed and accuracy while lowering processing requirements.

**Structure validation:** The computationally predicted structure must be validated by comparing various properties of the 3D protein structure with those of protein structures that are known through experimentation. Utilizing validation techniques, predicted models are evaluated for accuracy, guaranteeing that they are realistic and biologically significant. These procedures use several techniques to determine if the 3-D structure of a protein is experimentally feasible and physiologically significant. Accurate docking investigations to comprehend the interactions of medications or smaller molecules with proteins are made possible by dependable 3D protein structures. Inaccurate predictions regarding ligand binding and activity might result from poorly verified structures. Understanding a protein's functional mechanisms, such as conformational modifications, molecular interactions, and enzyme catalysis, requires a confirmed three-dimensional structure. Protein structural alterations can result from mutations in genetic disorders or illnesses like cancer. Using a validated 3D protein model, researchers can predict how these variations would affect the stability and functionality of proteins. The collection of frequently used protein 3D structure validation methods evaluates

several elements of protein structure quality, including fit, energy, and geometry. These tools can be used separately or in combination to evaluate the accuracy and structural quality of protein models. Selecting the appropriate tool depends on the particular characteristics, as different tools may concentrate on different aspects.

**Molecular docking:** These computer methods for the structural makeup of compounds made up of two or more different molecules. Docking studies aim to anticipate desirable three-dimensional configurations. Signal transduction in biological systems depends on the interactions of proteins, lipids, carbohydrates, and nucleic acids. By docking, predictions may be made about the kind and strength of signals produced by these interactions. Docking is frequently used in the pharmaceutical industry to predict how drug candidates will align with specific target molecules to regulate the function and selectivity of the tiny molecules. Docking, therefore, affects how drugs are structurally characterized. Docking research aims to minimize the system's free energy, achieving the most stable alignment by optimizing the ligand and protein shapes and their orientation. It involves multiple complex steps and is widely used to predict the binding position, affinity, and action of therapeutic molecules with their protein targets. These methods deposit tiny molecules over the active area of the protein. Scoring functions can evaluate a compound's biological activity by analyzing its interactions with potential targets. Molecular docking is used for lead discovery in both academic and medicinal settings. Docking predicts which molecule will be oriented more favourably than the others when they are linked to form a stable complex using mathematical models. There are two primary processes in the docking process: using a scoring function and sampling the ligand. Sampling algorithms assist in determining the ligand's most energetically advantageous conformations within the protein's active region while accounting for their binding manner. After that, a scoring algorithm is used to rank these confirmations. The scoring functions can be used to determine the binding affinity between two molecules (Shoichet et al. 2002).

**Types of docking and docking tools**

Models of molecular docking follow the standard concepts termed the "lock-and-key model" given by Fischer (1890) and "induced fit theory" by Koshland (1958). Similar to how a key is interleaved into a lock, a ligand molecule is placed into a receptor's active site. Additionally, both the ligand and the receptor undergo structural alterations to adjust to one another until a stable state is achieved. Proteins can experience substantial conformational changes in addition to slight induced-fit alterations. The conformation ensemble concept claims that proteins are

made up of an existing ensemble of conformal states and that they can change between these states because of their flexibility. Rigid docking and flexible docking are two different types of docking. Rigid docking predicts the optimal spatial match between the ligand and receptor by assuming that both are rigid. Flexible docking, on the other hand, assesses the adaptability of molecules by recognizing the receptor and ligand conformations as they arise inside the complex. Docking tools investigate orientations and conformations using search techniques, such as Tabu search, stochastic search, or systematic search. To determine the best structure and rotation for the receptor, virtual screening evaluates ligands according to their binding affinities. Four scoring functions are distinguished: empirical, force field, knowledge-based, and consensus. Among the most popular and efficient docking tools are AutoDock Vina, Glide, and AutoDock GOLD.

**Steps for docking**

**Recovery of protein and ligand structure**

**Check protein availability:** The three-dimensional structure of the target protein can be found in any of the familiar protein databases, such as the PDB (http://pdb.org) and the Swiss UniProt (http://expasy.org/sprot)

**Homology modelling:** If the target protein's structure is not present in PDB or Swiss UniProt, then the Swiss-Model Repository (http://swissmodel.expasy.org/repository/) will help build the protein structure by homology modelling, using sequence alignments with available templates. Alternatively, modelling programs such as I-TASSER or other similar software can be used to predict the 3D structure from comparable or available sequences.

**Find ligand(s):** Once the protein structure is found or modelled, the next step is to find ligands for docking. This can be found in PubChem Database: Search for available ligands on PubChem, (http://pubchem.org), Zinc Database (http://blaster.docking.org/zinc/) or ChEMBL Database. Chemical drawing tools like ChemDraw or ChemSketch can be used to design and create the ligand.

**Protein preparation:** Download the protein structure from a reliable database, such as the PDB. If the protein structure is unavailable, homology modelling tools like SWISS, and MODELLER or other modelling programs like I-TASSER can be used. Complete the protein structure by adding missing atoms/residues. This step ensures that the structure is complete for accurate docking simulations. Use modelling software to add or remove polar hydrogen atoms, missing atoms, or residues that were not included in the downloaded or modelled structure. Perform energy minimization to reduce any steric clashes or unfavourable geometries. This

step "loosens up" the protein structure and removes any high-energy conformations, ensuring that the protein is in a stable state. Assign protonation states to ionizable residues (such as histidines, lysines, and glutamates) based on the pH of the environment. This ensures correct electrostatic interactions between the protein and ligand during docking. Eliminate unnecessary water molecules and any extraneous ligands (co-factors, bound chemicals) that are not relevant to the docking simulation. This reduces the system complexity and makes the docking process more efficient. Apply the appropriate force field parameters to the protein structure. Force fields like AMBER, CHARMM, or GROMACS provide the necessary information to model molecular interactions and ensure the protein behaves accurately during docking. After completing the above steps, the protein structure becomes optimized and refined. This final structure is ready for docking simulations and provides a suitable starting point for identifying potential ligand binding sites.

**Preparation of ligands:** The selection of ligands for docking is dependent on several factors, including chemical variety, proven biological activity, and drug development potential. Following charge assignment, conformer generation, and shape optimization, the ligands are ready for docking.

**Active site prediction:** Molecular docking needs initial identification and post-docking validation of binding pockets. Three distinct methods can be used to confirm the binding pocket of interest. The first is site-directed docking, where the protein-ligand binding site is identified before docking the ligand. The second is blind docking, in which the ligand is loaded onto the receptor structure whose binding site is unknown, and a grid region is created to identify the docking area. Lastly, docking with a standard involves docking the protein with test ligands, allowing for predicting the relevant binding pocket.

**Protein-ligand docking:** After docking the prepared ligand against the prepared protein, the connections are examined. The binding interactions using grading functions that estimate the binding affinity are based on factors like hydrogen bonding, hydrophobic interactions, and electrostatic forces. The best grades access the protein-ligand complex and the corresponding ligand pose is selected and visualized using PYMOL or BIOVIA Discovery studio software.

**Post-docking analysis:** The most promising choices for additional research are determined by analyzing the findings of protein-ligand docking. The predicted interaction energy determines each ligand's binding affinity, and ligands are ranked following the results. Important connections between the ligands and the protein, including electrostatic forces, hydrophobic

interactions, and hydrogen bonds, are identified. These connections lead to the optimization of the ligands and deliver insightful information about their mode of action.

**References**

Nassar R, Dignon GL, Razban RM, Dill KA. The Protein Folding Problem: The Role of Theory. J Mol Biol. 2021 Oct;433(20):167126.

Paiva V de A, Gomes I de S, Monteiro CR, Mendonça MV, Martins PM, Santana CA, et al. Protein structural bioinformatics: An overview. Comput Biol Med. 2022 Aug; 147:105695.

Agnihotry S, Pathak RK, Singh DB, Tiwari A, Hussain I. Protein structure prediction. In: Bioinformatics. Elsevier; 2022. p. 177–88.

Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T. The SWISS-MODEL Repository and associated resources. Nucleic Acids Res. 2009 Jan 1;37(Database): D387–92.

Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M, et al. Comparative Protein Structure Modeling Using Modeller. Curr Protoc Bioinforma. 2006 Sep 2;15(1).

Smith TF, Waterman MS. Identification of common molecular subsequences. J Mol Biol. 1981; 147:195–197. DOI: 10.1016/0022-2836(81)90087-5

Xu, D., and Zhang, Y. (2012). Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins Struct. Funct. Bioinforma. 80, 1715–1735. doi:10.1002/prot.24065

Yang J, Zhang Y. Protein Structure and Function Prediction Using I-TASSER. Curr Protoc Bioinformatics. 2015 Dec 17; 52:5.8.1-5.8.15. doi: 10.1002/0471250953.

Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature 2020 Jan 30;577(7792):706–10.

Callaway E. What's next for AlphaFold and the AI protein-folding revolution. Nature 2022 Apr 14;604(7905):234–8.

Yin R, Feng BY, Varshney A, Pierce BG. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. Protein Sci. 2022 Aug;31(8): e4379. doi: 10.1002/pro.4379.

Bertoline LMF, Lima AN, Krieger JE and Teixeira SK (2023), Before and after AlphaFold2: An overview of protein structure prediction. Front. Bioinform. 3:1120370. doi: 10.3389/fbinf.2023.1120370.

Weissenow K, Heinzinger M, Steinegger M, Rost B. Ultra-fast protein structure prediction to capture effects of sequence variation in mutation movies. 2022.

Shoichet BK, McGovern SL, Wei B, Irwin JJ. Lead discovery using molecular docking. Curr Opin Chem Biol. 2002 Aug;6(4):439–46.

# Trained immunity approach for enhanced fish larval survival

Amritha Jagannivasan*, Krupesha Sharma S R, Sumithra T G., Ambarish P Gopidas
Marine biotechnology fish nutrition and health division., ICAR-CMFRI, Kochi-18,,
*amrithajagannivasan04@gmail.com

## Introduction

Aquaculture is one of the fastest-growing and most promising industries, providing a reliable source of high-quality animal protein and employment opportunities. Captive rearing of fishes is always associated with the risk of inviting exotic and opportunistic pathogens responsible for causing disease outbreaks. With the rising demand for aquatic animals, farmers face the challenge of enhancing growth rates while minimizing disease outbreaks. Due to the intensification of farming practices, diseases, mostly caused by infectious causes, have become a major threat to the sustainability of aquaculture practices. While opportunistic pathogens threaten intensive farming, the transboundary movement of aquatic animals also leads to the spread of infectious germs between nations. The use of antibiotics to treat infections has drawn criticism because it leads to the emergence of microbes resistant to antibiotics, environmental contamination, immune system dysfunction, and chemical buildup in aquatic animal tissues, all of which may pose a health risk to the general public. In fish culture, vaccination is a successful preventative measure against infectious diseases, although it can be highly costly and traumatic for the fish.

Larval rearing is one of the riskiest phases, with poor growth and survival rates (05-20%) affecting the supply of high-quality larvae in sufficient quantities (Kaliadas et al. 2012). The absence of a mature adaptive immune system makes the larval stage, the most vulnerable in the life cycle of fish. Accordingly, early-stage marine fish larvae are highly vulnerable to different stressors which can lead to larval mortalities, and poor larval viability and quality. As the fish larvae own only a functional innate immune system the approach targeting the innate immunity will be the appropriate viable means in the larviculture practices (Zhang et al. 2019).

## Fish immune system: An overview

The teleost immune system is a complex network of cells, tissues, and molecules designed to defend against pathogens and maintain homeostasis. Unlike mammals, fish inhabit an aquatic environment that constantly exposes them to diverse microorganisms, necessitating unique

adaptations to protect against infections. Fishes have two major immunological mechanisms namely: innate and adaptive immune systems.

**Innate immunity**

The initial line of defense is innate immunity, which provides instantaneous, general protection against infections. The innate immunity identify pathogens through the pathogen-associated molecular patterns including lipopolysaccharides, lipoteichoic acid, phospholipomannan, zymosan, beta-glucan, and hemagglutinin, that are not present in the host cells. Fish innate immune system components can be broadly divided into three categories: physical barriers, humoral, and cellular immunity. The mucous layer, scales, and the epithelial cells that line the gills, skin, and digestive tract are examples of physical characteristics in preventing infection. Humoral immune parameters constitute soluble compounds that have preventative effects by neutralizing the enzymes that the pathogen produces and preventing the growth of the bacteria. Teleost fish have a variety of non-specific defense compounds that can either kill or stop the growth of germs, such as transferrin, lectins, cytokines, antimicrobial peptides like lysozyme, and complements. Transferrin is an acute-phase protein that functions by binding iron and creating an environment that is inhospitable to bacteria. Transferrin is activated in the course of the inflammatory reaction for extracting iron from degraded tissue and acts as a fish macrophage activator. Lysozymes present in mucus, plasma, and lymphoid tissues, are mainly secreted by the liver and hepatic cells causing bacterial lysis by acting on the bacterial cell walls' peptidoglycan. The complement system performs several tasks, including phagocytosing foreign particles, lysis of bacteria and infected cells, apoptotic cell removal, and modulation of inflammatory vasodilation. It also acts as a link between innate and adaptive immune responses and enables an integrated host defense against infectious threats. The innate immune mechanism's cellular elements include a vast range of cells that are triggered when they identify a pathogen through the PAMP. Depending on their cell variant, they may engage in a variety of reactions that ultimately result in the pathogen's destruction.

**Adaptive immunity**

Fish have the most primitive adaptive immune systems among vertebrates. The genetic components necessary for the adaptive and combinatorial immune responses to function are present in all jawed vertebrates. The combinatorial immune system (CIS) comprises of Ag-recognizing lymphocytes, immunoglobulins (Abs and Ig-family TCR), MHC molecules, and recombination-activating (RAG) 1 and 2 genes, that have evolved in teleost fish around 450-500 million years ago. Cellular-mediated and humoral-mediated responses are the two major

constituents of the teleost fish adaptive immune system. Cell-mediated adaptive immunity employs T-cell lymphocytes and antigen-presenting cells to respond to infections rather than antibodies. T cells release factors to coordinate responses of other immune cells or cytotoxic factors to directly kill infected or abnormal cells. The main components of cellular-mediated adaptive immune responses are the T cells, which can develop into effector cytotoxic T lymphocytes or effector helper T cells. CD3, CD4, and CD8a and homologues CD28, CD3e, TCR-b, TCR-z, and TCR-g are the major T cell markers of fish. B cells are involved in humoral-mediated adaptive immunity as they generate antibodies that opsonize, neutralize, and bind complement through antibody-dependent cellular cytotoxicity (ADCC). In teleost fish, three antibody classes have been identified namely IgM, IgD, and IgT/Z. Since teleost fish do not have bone marrow and lymph nodes, their primary source of hematopoiesis is the anterior kidney, where B cells are formed.

**Concept of trained immunity**

According to the paradigm, innate immunity has a brief duration since it lacks immunological memory and specificity. Since innate immunity lacks specificity, it was previously believed that immunological memory was not functioning as in the adaptive immune system. However, Netea et al. (2011) demonstrated that repeated exposure to a homologous or heterologous microbe might cause innate immunity to establish more powerful responses and called this condition of innate hyper-responsiveness to the subsequent exposure as "trained immunity." In other words, trained immunity triggers the innate immune system that finally results in enhanced non-specific resistance to infection by similar or dissimilar infections. This is developed by providing immunomodulators which cause a vague memory, instead of specific memory developed by the adaptive immunological response. Several ligands of pattern recognition receptors, including flagellin, β-glucan, CpG containing oligodeoxynucleotides, and muramyl dipeptide are known to strengthen vertebrates' defence against invasion by heterologous infections. A demonstrative example is the prophylactic effect of β-glucan (laminarin) to protect fish against *Vibrio salmonicida* (Dalmo et al. 1998). So, the trained immunity concept evolved as a novel concept of the previous observations reported >15 years ago.

**Hallmarks of trained immunity**

Even though trained immunity is a novel concept that evolved from the previous observations, it has been enhanced with new features. It does not involve T- and B-cell during their activity, epigenetic changes and a change in metabolic profile act as major hallmarks. During the

process of immune training, triggered by the activation of particular pattern recognition receptors, mainly MAP kinase-dependent intracellular signalling, innate immune cells like macrophages, monocytes, and NK cells get (re) programmed causing epigenetic changes. It has been speculated that, in teleost fish, immune cells like macrophages may have a greater capacity to express traits of trained immunity because they belong to early vertebrates with both innate and adaptive immune components, and their innate immune arm is crucial in the fight against infectious agents. The typical hallmarks of a trained immune response include a) T and B lymphocyte-independent immune response to early infections and consequent defence against a secondary infection, b) an immune response that delivers more protection upon subsequent infections but is less unique than the adaptive immune response and, c) the involvement of cells involved in innate immune response, such as NK cells and macrophages, resulting in enhanced inflammatory reactions and better identification of pathogens. According to recent research, typical innate immune cells, exhibit long-lasting re-programmed response following infection or vaccination. These alterations increase the release of inflammatory signals and induce a stronger ability to eradicate the infection, which increases the capability to fight against further attacks by microbial agents. Furthermore, trained immunity probably plays a significant part in ontogeny, enabling the freshly hatched larvae's innate immune system to mature, a process in which the surrounding microbiota is crucial. This immunological memory, in contrast to adaptive immunity, is generic and depends on metabolic and epigenetic reprogramming.

**Trained immunity in finfish**

There are highly diverse innate receptors for innate training in teleost. Briefly, 21 distinct Toll-like receptors (TLRs) have been found in teleost, along with other splicing variations (subtypes/isotypes). The TLR variations in teleost are higher than in humans and mice. These diverse innate receptors consist of splice variants, including various C-type lectin receptors, NOD-like receptors, RIG-1-like receptors, and scavenger receptors may serve as innate immune training sites. Given that immunological training can strengthen the natural defense mechanism of fish, particularly against infections, it creates opportunities for several intriguing methods in fish larval rearing.

**Impact of epigenetic and metabolic modifications on trained immunity**

Numerous metabolic pathways are linked to trained immunity, which uses metabolic restructuring to enhance immune cells' defence against subsequent infections. Recent research indicates that epigenetic changes, such as chromatin remodelling, DNA methylation, histone

modifications, and non-coding RNA, function to preserve cell identity. By controlling gene expression, the mechanism of epigenetic changes can influence host-pathogen interactions suggesting that both metabolic and epigenetic reconstruction are crucial components of trained immunity. Durable epigenetic changes that remain even after the training cue is eliminated is one of the fundamental processes of trained immunity. The epigenetic changes in trained immunity depend on the various cellular metabolic reorganizations that are triggered by immunological signals. Also, several epigenetic enzymes are controlled by mitochondria. Additionally, lipid hydrolysis has been connected to the hydrolytic enzymes such as alkaline phosphatase and acid phosphatase, which regulate the dephosphorylation of proteins, alkaloids, and nucleotides, which fuels resistance against outside stimuli. Mitochondria supply metabolic intermediates and their byproducts, such as acetyl-CoA and S-adenosyl methionine, which promote epigenetic modifications including acetyl-CoA-induced histone acetylation. By altering the acetylation of the whole histone, quantity of acetyl-CoA influence histone acetyltransferases (HATs) activity, which control gene activity and are heavily reliant on the mitochondria's glucose supply and fatty acid metabolism. However, invertebrates' mitochondrial oxidative metabolism can be suppressed by mitochondrial malfunction brought on by environmental mutagen exposure or pathogen stimulation, which may be the cause of the shift in gene expression profiles following pathogen challenge.

Epigenetic modifications like methylation, acetylation, phosphorylation, and ubiquitination are post-translational moderations of histone proteins that may change the chromatin constitution at different positions. Histone acetylase/deacetylase (HDAC) catalyses the reversible process of histone acetylation, which involves the insertion or elimination of an acetyl group on lysine residues. While HDAC-catalysed histone deacetylation strengthens the chromatin structure to suppress gene expression, pathogen invasion may cause the structure to break down, which suppresses host gene expression. Methylation of DNA is a permanent covalent cytosine alteration in the cytosine-guanine dinucleotides (CpG) environment and is sparsely methylated in invertebrate genomes but strongly methylated in vertebrate genomes. DNA Methyltransferases (DNMTs) play a crucial role in regulating gene expression in several organisms by regulating genomic methylation. Long non-coding RNAs (LncRNAs) have a crucial physiological function in animals, including chromatin modification complexes, enhancers, promoters, and post-transcriptional levels of gene control including translation of mRNA or promoter activity, produces miRNA precursors to control the activity of target genes. These epigenetic changes lead to the activation of cells, increased generation of cytokines, and

a change in the metabolic pathway from oxidative phosphorylation to aerobic glycolysis. The final component supplies the energy required for optimal operation; immune system cells precisely regulated metabolic processes cause them to enter a dormant phase and change their metabolic processes to restore their energy (also known as Warburg effect) in response to inflammatory trigger. This change is associated with decreased basal rate of respiration, enhanced consumption of glucose and lactate production, and an increased ratio of nicotinamide-adenine dinucleotide (NAD) to its reduced form (NADH). The epigenetic rewiring mechanisms would lead to more rapid and heightened expression of pro-inflammatory genes after secondary exposure to the same or different stimuli, which is associated with prolonged innate immune memory.

Immunological responses following immunostimulation are evidenced by an increase in *tnf-α*, *il-1*, and *il-6* pro-inflammatory cytokines, which increase the acute phase response by producing complement components, C-reactive protein, ceruloplasmin, metallothionine, and other substances. Additionally, by promoting the production of adhesion molecules on vascular endothelium and strengthening the epithelial response to infections, these cytokines improve leucocyte diapedesis.

**Methods of immunostimulant administration to induce trained immunity**

The compounds that are likely to produce trained immunity are being evaluated by directly administering to freshly hatched alevins or larvae (before their first meal) and results have demonstrated improved larval survival(Zhang et al. 2019). However, to maintain a therapeutic dose in flow-through systems, where the tank water is continuously changed, the immunostimulant dose must also be changed making the use of immunostimulants too costly or unfeasible (Bricknell and Dalmo 2005). Accordingly, for fish species that consume particle feed, immunostimulants may be added to the prepared fish meal(Rojo-Cebreros et al. 2018). Major commercial feed manufacturers employ well-established feed delivery systems to integrate chemicals. Further, broodstock stimulation can also be used to stimulate trained immunity in fish embryos and larvae as it can pass the learnt innate immunity to their progeny (F1 generation). For example, parental gene expression patterns combined with the immunological gene and epigenetic regulation expression patterns of pipefish (*Syngnathus tyhple*) offspring showed a high correlation (Beemelmanns and Roth 2017)

**Commonly used immunostimulants in aquaculture**

Healthy fish with trained immunity can better resist bacterial and viral invasions. For many years, industrial fish aquaculture and lab settings have employed β-glucans as non-specific

immunostimulants (Vetvicka et al. 2013). Commonly used β-glucans for trained immunity studies include laminarin, curdlan, lentinan, schizophyllan, and scleroglucan. Since commercially available diets supplemented with nucleotides have been shown to decrease sea lice settlement and improve protection against *Aeromonas salmonicida* and *Vibrio anguillarum* infection, the use of in-diet immunomodulators has gained widespread acceptance in both salmonid and non-salmonid aquaculture (Burrells et al. 2001). When Atlantic cod (*Gadus morhua*) juvenile fish are fed cod milt cationic proteins, their resistance to this bacterial illness is enhanced (Pedersen et al. 2004). Compared to a control group, turbot larvae fed rotifers enriched alginate at first feeding exhibited protein synthesis fractional rates that are noticeably higher, which led to a three-fold increase in protein turnover in the immunostimulant-fed larvae. Boosting protein turnover probably results in increased larval survival and viability in the face of disease or stressful conditions (Conceicao et al. 2001). When *Dentex dentex* was fed with immunostimulants containing feeds, viz. “MakroGard” or “Vita-Stim”, mortalities caused by protozoan infections were significantly reduced to 10% and 15% compared to a control fish (30%).

**Conclusion**

Considering the presence of only a functioning innate immune system, fish larvae are the most vulnerable of all the life stages. According to the studies, aquaculture productivity may be significantly increased by raising the survival rate in these early life stages. Accordingly, it has been suggested that using a trained immune method is a key way to increase fish larval survival rates. The biological effects of several immunostimulants have been investigated since the discovery of pattern recognition receptors. Due to increased cellular and humoral activity, many immunostimulants employed in fish trials inevitably provide positive effects like disease prevention. However, care must be given concerning problems like tolerance, undesirable side effects like immunosuppression from using excessive amounts of immunostimulants, or undesirable effects brought on by long-term usage of such substances.

**References**

Beemelmanns A, Roth O (2017) Grandparental immune priming in the pipefish *Syngnathus typhle.* BMC Evol Biol 17:44. https://doi.org/10.1186/s12862-017-0885-

Bricknell I, Dalmo R (2005) The use of immunostimulants in fish larval aquaculture. Fish

Shellfish Immunol 19:457–472. https://doi.org/10.1016/j.fsi.2005.03.008

Burrells C, Williams P., Forno P. (2001) Dietary nucleotides: a novel supplement in fish feeds. Aquaculture 199:159–169. https://doi.org/10.1016/S0044-8486(01)00577-4

Conceicao LEC, Skjermo J, Skjåk-Bræk G, Verreth JA. (2001) Effect of an immunostimulating alginate on protein turnover of turbot *(Scophthalmus maximus L.*) larvae. Fish Physiol Biochem 24:207–212. https://doi.org/doi:10.1023/a:1014052611305

Dalmo RA, Martinsen B, Horsberg TE, et al (1998) Prophylactic effect of β(1,3)-D-glucan (laminaran) against experimental *Aeromonas salmonicida* and *Vibrio salmonicida* infections. J Fish Dis 21:459–462. https://doi.org/10.1046/j.1365-2761.1998.00120.x

Kaliadas C, Sakthivel M, Tamilmani G, et al (2012) Survival and growth of juvenile silver pompano *Trachinotus blochii* (Lacepède, 1801) at different salinities in tropical conditions. Indian J Fish 59:95–98

Netea MG, Quintin J, van der Meer JWM (2011) Trained Immunity: A Memory for Innate Host Defense. Cell Host Microbe 9:355–361. https://doi.org/10.1016/j.chom.2011.04.006

Pedersen GM, Gildberg A, Olsen RL (2004) Effects of including cationic proteins from cod milt in the feed to Atlantic cod (*Gadus morhua*) fry during a challenge trial with *Vibrio anguillarum.* Aquaculture 233:31–43. https://doi.org/10.1016/j.aquaculture.2003.09.026

Rojo-Cebreros AH, Ibarra-Castro L, Martínez-Brown JM (2018) Immunostimulation and trained immunity in marine fish larvae. Fish Shellfish Immunol 80:15–21. https://doi.org/10.1016/j.fsi.2018.05.044

Vetvicka V, Vannucci L, Sima P (2013) The effects of β - glucan on fish immunity. N Am J Med Sci 5:580. https://doi.org/10.4103/1947-2714.120792

Zhang Z, Chi H, Dalmo RA (2019) Trained Innate Immunity of Fish Is a Viable Approach in Larval Aquaculture. Front Immunol 10:. https://doi.org/10.3389/fimmu.2019.00042

# Importance of Restriction Site Associated DNA (RAD) sequencing in population genomics

Sandhya Sukumaran,
Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018
sandhya.sukumaran@icar.org.in

## Introduction

Restriction site associated DNA sequencing is a technique that utilizes restriction enzymes to cleave genomes at designated locations, thereby offering a comprehensive representation of each site cleaved by a specific restriction enzyme. This approach is also referred to as a genome 'complexity reduction protocol.' In recent years, this method has gained significant popularity for the genome-wide identification of polymorphic markers across multiple individuals.

## Method

RAD sequencing aims to integrate two methodologies in molecular biology alongside Illumina sequencing. DNA is fragmented using restriction enzymes, and sequence reads are associated with specific individuals through molecular identifiers (MID). The fragmentation of DNA from a single individual is performed with a specific restriction enzyme, resulting in a collection of sticky-ended fragments. Subsequently, these sticky-ended fragments are ligated to a P1 adapter, which possesses a complementary sticky end along with a molecular identifier. The tagged restriction fragments from various individuals are combined and randomly sheared to yield fragments with an average length of several hundred base pairs. The sheared fragments are then ligated to a second P2 adapter and amplified using P1 and P2 primers through PCR. The P2 adapter features a divergent “Y” structure that cannot bind to the P2 primer unless it has been fully amplified by the P1 adapter. This guarantees that all amplified fragments contain the P1 adapter and MID, a partial restriction site, the flanking sequence characterized by several hundred bases, and a P2 adapter. The amplified fragments, which are prepared for sequencing, are selected based on size (approximately 200-500 base fragments), and the RADseq library is sequenced on an Illumina platform. As a result, sequences are generated from the molecular identifier in the P1 adapter and across the restriction enzyme site, producing a dataset of RAD tags derived from a reduced segment of the original genome.

RAD sequencing offers numerous benefits over whole genome sequencing, as it enables the sequencing of specific regions of interest. This method, known as reduced representation, samples only a common set of sites across the genome from multiple individuals. Consequently, it facilitates population-level sequencing at a significantly lower cost compared to whole genome sequencing. Furthermore, NGS RAD sequencing can be utilized for fine-scale linkage mapping, phylogenetics, phylogeography, genome scaffolding, and studies in population genetics. Recently, this technique has been employed in species such as salmon, cutthroat trout, and rainbow trout to draw conclusions about their population genetic structure.

**Suggested Readings**

Amish, S.J., Hoehenlohe, P.A., Painter, S., Leary, R.F., Muhlfeld, C., Allendorf, F.W. and Luikart, G. 2012. RAD sequencing yields a high success rate for westslope cutthroat and rainbow trout species-diagnostic SNP arrays. Molecular Ecology Resources 12: 653-660.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. and Cresko, W.A. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22: 3124-3140.

Davey, J.W. and Blaxter, M.L. 2011. RADSeq: next generation population genetics. Briefings in Functional Genomics 9(5): 416-423.

Davey, J.W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K. and Blaxter, M.L. 2013. Special features of RAD sequencing data: implications for genotyping. Molecular Ecology 22: 3151-3164.

Pocwierz-Kotus, A., Bernas, R., Kent, M.P., Lien, S., Leliuna, E., Debowski, P. and Wenne, R. 2015. Restitution and genetic differentiation of salmon populations in the sourthern Baltic genotyped with the Atlantic salmon 7K SNP array. Genetics Selection Evolution 47: 1-9.

# Genomic DNA and the evolution of DNA extraction techniques

Krishna Priya S,
Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018
*krishnapriya1254@gmail.com*

**Genomic DNA**

An organism's genomic DNA represents its genetic information. Although almost all organisms' genomes are made of DNA, there are some viruses with their genomes made up of RNA. In most species, genomic DNA is extremely large and occurs in the form of DNA-protein complexes termed chromosomes. Different species have different genomic DNA properties, such as size, the number of chromosomes, and how they are arranged. Most eukaryote genomic DNA exists in the nucleus in the shape of different linear chromosomes of diverse sizes, while bacteria have a single circular chromosome.

**Genomic DNA extraction**

One key foundation of the molecular biology field is genomic DNA extraction. This ensures a consistent result in downstream applications. DNA extraction includes isolating and purifying DNA from biological samples while preserving the integrity of DNA.

Different methods for isolating DNA have been developed, from liquid-based DNA extraction (1955–1968) to nanotechnology-based DNA extraction in the 2020s. DNA-based methods are used in forensic science, disease diagnosis, and detecting pathogenic microorganisms in food, environmental, and clinical samples.

**Evolution of DNA Extraction Methods**

**1.Organic DNA extraction (1956)**

Organic DNA extraction technique isolates DNA from biological samples based on the use of organic solvents such as phenol, chloroform, and isoamyl alcohol. Cell lysis is the initial step involved and is usually done with a detergent and enzyme such as proteinase K to break down the proteins and cell membranes. An equal quantity of the phenol-chloroform mix is added to the sample. This is then centrifuged to separate the sample into two layers: an organic layer

that will contain proteins and lipids and an aqueous layer that will contain DNA after centrifugation. The DNA is left in the aqueous layer and the contaminants remain in the organic layer or interphase. The aqueous phase is slowly poured into a new tube, and cold ethanol or isopropanol is added to precipitate the DNA. The pellet of DNA is air-dried, washed with 70% ethanol, and suspended in water or buffer. The procedure is excellent for isolating good-quality DNA but uses toxic chemicals that should be used and eliminated properly to remain safe.

**2. Salting out extraction (1987)**

The salting-out procedure uses high salt concentrations to extract proteins and other impurities from DNA. It is an easy and low-cost DNA extraction process. In the process, the cells are opened up and DNA is released together with other cellular components using a lysis buffer with detergents and chelating agents. The proteins are then precipitated by adding a salt solution, e.g., potassium acetate or sodium chloride, that interferes with their solubility, leaving DNA in the supernatant. The protein pellet is spun out and cold isopropanol or ethanol is added to precipitate out the DNA. After removing any remaining contaminants with 70% ethanol, the DNA pellet is dried and resuspended in water or buffer. This technique is routinely used to purify high-quality DNA for procedures like PCR and sequencing and is a cheap substitute for commercial kits but may need to be carefully handled to prevent contamination and achieve good recovery.

**3. Column method (1990s)**

Column DNA extraction utilizes the silica-based affinity membrane to bind DNA at an appropriate pH, most often during or after cell lysis, in the presence of chaotropic salts (e.g., guanidine salts). These salts denature protein, break hydrogen bonds, and allow for the binding of DNA by breaking DNA's interaction with water. The DNA is immobilized on the membrane while impurities are removed by the washing operations, which purify the impurities with a detergent and ethanol solution. Then, an elution using a low-salt buffer (e.g., TE or Tris buffer) releases the DNA. Depending on the column structure, the elution can be performed either by centrifugation or gravity.

High purity, yield, high-quality inhibitor-free DNA for downstream applications, and reduced risk of contamination are obtained by this method. However, it is very expensive, time-consuming, and challenging, particularly when there are many pre-treatment steps.

Furthermore, some contaminants (e.g., phenolic and humic compounds) that bind to or resemble DNA will co-elute through silica-based membranes.

**4. Magnetic Beads (1990 s)**

The magnetic bead-based technique, which was first developed in 1998 by Hawkins T, involves using small beads, around 20-30 nanometers in size, made of iron oxide. These magnetic beads only show their magnetic properties when there is an external magnetic field. Many types of beads are available, and each type has different surface coatings, and these coatings affect how the beads bind with biomolecules. In the process, the DNA is attached to these beads when the magnetic field is applied, and then the DNA stays attached to the bead during the washing steps. Afterward, a buffer is added to this DNA, and the purified DNA can now be analyzed and quantified. The technique does not need any centrifugation, which makes it simpler. Also, it uses fewer chemicals and fewer steps compared to other methods. This method works well for automation purposes and is seen as effective and reliable in isolating nucleic acids. Automatic machines such as Maxwell 16 and MagNA Pure 96 make it easier but it is costly while manual extraction is less scalable It is also used in sample preparation for techniques like PCR and NGS and various other uses.

**5.FTA™ CARDS (Flinders Technology Associates) (1990s)**

FTA paper is a quick, easy, and cost-effective way of gathering, storing, and conserving biological samples at room temperature without refrigeration. It consists of cellulose paper treated with compounds that lyse cells, denature proteins, inactivate the microbe, and shield DNA from degradation. A small sample is deposited on the card to air dry. The DNA is separated and fixed for long-term storage.

This technique is usable for a broad range of samples including blood, saliva, tissues, plants, insects, microbes, and environmental components like soil and water. There can be pretreatment steps that have to be performed in order to facilitate DNA extraction from intricate samples. The DNA may be immobilized on the card or can be eluted for use in PCR, RFLP, or genome amplification.

There are two categories of FTA cards: elute cards, which facilitate DNA recovery into solution, and "classical" cards, which keep DNA on paper. Due to their environmental stress resistance, these cards can be transported and keep DNA stable for years in uncontrolled environments. The major disadvantage is the higher cost and a limited sample volume capacity.

**6. Automated DNA extraction (2000s)**

Automated DNA extraction equipment combines processes like cell lysis, washing, and elution into a single efficient process. They operate on several samples at a time using robotics and software. This reduces labour, time, and human error while providing reproducible results enhancing productivity, and reducing the risk of contamination making them best suited for high-throughput applications like forensic labs and clinical diagnostics. Their greater application is their value in achieving the need for fast, productive DNA isolation, ongoing genetic analysis, and forensic utility enhancement—but at a higher cost and reduced variability of the protocol.

**7. Microfluidic DNA extraction (2010s)**

Microfluidic DNA extraction is an approach that applies a lab-on-a-chip technology that includes cell lysis, purification, and elution together in the small microchannels of the chip. They are cost-effective, reduce the amount of reagent required, precisely control the amount of reagent, and aid with very precise sample manipulation resulting in a more efficient technique relative to the conventional ways. It is used in environmental monitoring, point-of-care diagnostics, forensics, etc where the sample size is small. It is however limited by the complexity of the chip design and the equipment and trained people required. However, microfluidic DNA extraction offers a promising technology since it is very efficient, portable, and cost-saving.

**8. Nanotechnology-Based Extraction (2020s)**

DNA binding and purification have improved with the use of nanotechnology. Purification is made easy and effective using nanoparticles, most of which are ligand-coated with some ligands that are capable of binding DNA. The magnetic or centrifugation processes allow effective separation thus rendering this process significant in clinical and forensic analysis. The process is currently under development but can be applied for ultra-sensitive DNA extraction where trace amounts are investigated from low-quality samples. The process can be applied to transform DNA extraction and analysis across different fields due to the flexible nanoparticle functionalities.

**References**

Barbosa, C., Nogueira, S., Gadanho, M., & Chaves, S. (2016). DNA extraction: finding the most suitable method. In *Molecular microbial diagnostic methods* (pp. 135-154). Academic Press.

Dairawan, M., & Shetty, P. J. (2020). The evolution of DNA extraction methods. *Am. J. Biomed. Sci. Res*, *8*(1), 39-45**.**

# DNA isolation by the phenol-chloroform method

Ramitha P A.

Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018 ramitha2394@gmail.com

## Introduction

Isolation of DNA through phenol-chloroform is one of the most common techniques in molecular biology. Researchers can obtain a good yield of high-molecular-weight DNA from any biological sample using a mixture of phenol, chloroform, and isoamyl alcohol. The basic principle involves the separation of DNA, RNA, and proteins in a sample, based on the differential solubilities of these molecules in different immiscible liquids. Different components of each step perform different functions in completing the lysis of the sample. Lysis buffer helps break open cells and regulates the acidity and osmolarity of the lysate by the salts used in the buffer. SDS can solubilise the protein and lipids that form the membrane and helps to release DNA from histones and other DNA-binding proteins by denaturing them . Further addition of Proteinase K helps in the destruction of proteins in cell lysate and in the release of DNA. The next phenol-chloroform mixture is added to separate the proteins from the DNA. Since DNA has net negative charge, it is more soluble in the aqueous portion of the organic aqueous mixture. When centrifuged, the unwanted proteins are separated away from the aqueous phase due to its hydrophobic and hydrophilic domains and will be present in the interphase. Finally, isolation of nucleic acid (DNA) was accomplished using alcohol precipitation and stored in TE buffer, which maintains the integrity of DNA. The isolated DNA can be used in gene amplification, fingerprinting, single-nucleotide polymorphism studies, RFLP, and gene quantification.

## Materials Required

- Forceps
- Surgical Blade
- Centrifuge
- Micro Centrifuge tube (1.5ml)
- Micropipette and Pipette tips
- Water bath
- Vortexer .

**Reagents**

1.Lysis buffer

- NaCl 23.36g (400mM)
- Tris-cl 1.2114g (10mM)
- EDTA 0.37224g (1mM)
- Distilled Water - 1000mL

2. Sodium Dodecyl Sulphate (10%)

3. Proteinase K (10mg/mL)

4. Phenol Chloroform Isoamyl Alcohol (25:24:1)

3. Chloroform Isoamyl Alcohol (24:1)

5. Isopropanol

6.70% alcohol

7. TE buffer

**Procedure**

1. 10-20mg of tissue of sample was taken, then stored in 95% ethanol at refrigerator for long term storage .
2. Tissue was minced to small pieces with sterilized surgical blade and was transferred to micro centrifuge tube.
3. 400µL lysis buffer and 100µL 10% SDS were added. Then 10µL proteinase k, was added to mixture, contents are thoroughly mixed by using vortex mixer.
4. The solution was kept in water bath at 60°C for complete lysis of the tissue till it becomes a clear homogenous solution.
5. When the tissue lysed completely, add equal amount of phenol: chloroform: isoamyl alcohol (25:24:1) was added, contents were then vortexed.
6. Tubes were then centrifuged at 10,000rpm for 10 mins at 4°C. After centrifugation, 3 phases were formed.
7. Transfer the upper layer carefully to a fresh tube with-out disturbing the interphase.
8. To this equal volume of chloroform: isoamyl alcohol (24:1) mixture was added and vortex thoroughly. Then contents again centrifuged at 10,000 rpm 10 minutes at 4°C. The supernatant collected and transferred to fresh tube.
9. Equal volume chilled isopropanol was added, mixed by inverting and incubate for 2 hours at $-20^0$c. Then tube is centrifuged 10,000 rpm for 10 mins at 4°C.

10. Supernatant was removed and pellets was washed with 70% ethanol, then centrifuged 10,000 rpm for 10 minutes at 4°C.
11. Remove the supernatant, allowed to air dry the pellet at room temperature.
12. Redissolved pellet in 50 µL 1X TE buffer, DNA was stored at deep freezer for further analysis.

NB: Wear gloves while handling the reagents and Make sure that the vials and tips are DNAase free

**Suggested Reading.**

Sambrook, J., & Russell, D. W. (2006). Purification of nucleic acids by extraction with phenol: chloroform. *Cold Spring Harbor Protocols*, *2006*(1), pdb-prot4455.

# Genomic DNA Isolation by the Salting-Out Method

Vysakh V G.
Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018
vishakvg99@gmail.com.

**Introduction**

Isolation of high-quality genomic DNA is the first and one of the most critical steps in molecular biology. The quality, purity, and integrity of extracted DNA directly influence the success of downstream applications such as polymerase chain reaction (PCR), DNA barcoding, DNA sequencing, genotyping, microsatellite analysis, population genetics, phylogenetic studies, and next-generation sequencing. Therefore, selecting an appropriate DNA extraction method is essential for obtaining reliable and reproducible molecular data.

For genomic DNA extraction several methods have been developed. For instance, phenol-chloroform extraction, cetyltrimethylammonium bromide (CTAB)-based protocols, silica column-based commercial kits, magnetic bead-based purification systems, and salting-out procedures. Even though phenol-chloroform extraction has long been regarded as an effective method for DNA purification, it involves hazardous organic solvents that are toxic, corrosive, and require careful handling and disposal. Commercial DNA extraction kits give rapid and reproducible results but are relatively expensive. This makes it less suitable for laboratories processing of large numbers of samples or operating under limited budgets. These limitations have encouraged the development of simpler, safer, and more economical alternatives for routine DNA isolation.

The salting-out method, first described by Miller and co-workers in 1988. Which is one of the most widely used non-organic methods for genomic DNA extraction. Unlike traditional organic solvent-based methods, this method removes cellular proteins by precipitation with a concentrated sodium chloride solution and eliminating the need for phenol or chloroform. This approach simplifies the extraction procedure and reduces health hazards and laboratory waste. This method consistently yields high-molecular-weight genomic DNA with good purity.

Making it suitable for PCR amplification, restriction enzyme digestion, sequencing, and numerous other molecular applications.

Several modifications of the original salting-out protocols are developed to improve DNA yield, reduce processing time, and simplify the extraction procedure for different sample types. Modified protocols have shown that the method produces DNA of sufficient quality for routine PCR-based analyses while remaining inexpensive, safe, and practical for laboratories handling large numbers of samples. These improvements have further established salting-out as a reliable alternative to more complex DNA extraction procedures.

In fisheries and marine biotechnology laboratories, the salting-out method has become one of the preferred techniques for isolating genomic DNA from fish, shellfish, crustaceans, molluscs, seaweeds, and other aquatic organisms. The protocol requires only basic laboratory equipment and readily available reagents, making it particularly suitable for routine research, biodiversity assessments, DNA barcoding, population genetic studies, conservation genetics, and aquaculture breeding programmes.

**Principle of the Salting-Out Method**

Salting-out method is based on the selective precipitation of cellular proteins and maintaining genomic DNA in solution. First, the cells disrupted using a lysis buffer containing Tris-HCl, EDTA, and sodium dodecyl sulfate (SDS). Tris-HCl maintains a stable pH during extraction, EDTA chelates divalent metal ions like $Mg^{2+}$ and $Ca^{2+}$ to inhibit DNase activity, and SDS disrupts cell membranes and denatures proteins, and releasing genomic DNA into solution. Proteinase K is then added to digest structural proteins, histones, nucleases, and other DNA-binding proteins. This step improves DNA purity and prevents enzymatic degradation.

After protein digestion, saturated sodium chloride (NaCl) solution is added to the lysate. The high salt concentration dehydrates and precipitates denatured proteins and other cellular debris and genomic DNA remains dissolved in the aqueous phase. Centrifugation separates the precipitated proteins as a compact pellet, enable the clear DNA-containing supernatant to be transferred into a fresh tube.

Genomic DNA is then recovered by adding two volumes of cold absolute ethanol, which reduces DNA solubility and promotes precipitation as visible strands or a compact pellet after

centrifugation. The DNA pellet is washed with 70% ethanol to remove residual salts and impurities, partially air-dried, and finally dissolved in Tris–EDTA (TE) buffer for long-term storage. The resulting DNA is typically of high molecular weight and suitable for a wide range of downstream molecular analyses, including PCR, DNA sequencing, DNA barcoding, microsatellite genotyping, and next-generation sequencing library preparation.

**Reagents Required**

All solutions should be prepared with double-distilled or nuclease-free water and stored under appropriate conditions to maintain reagent stability and prevent contamination.

**Table 1. Reagents used in the salting-out DNA extraction protocol.**

| Reagent | Composition | Function | Storage |
|---|---|---|---|
| **Solution 1 (Lysis Buffer)** | 50 mM Tris-HCl (pH 8.0), 20 mM EDTA (pH 8.0), 2% SDS | Cell lysis, membrane disruption, protein denaturation and protection of DNA from nuclease degradation | 4°C |
| **Solution 2** | Saturated NaCl (6 M) | Precipitation of proteins and cellular debris | 4°C |
| **Proteinase K** | 20 mg/mL | Digestion of proteins and nucleases | −20°C |
| **Absolute Ethanol** | Molecular biology grade | DNA precipitation | Room temperature |
| **70% Ethanol** | Prepared using nuclease-free water | Washing DNA pellet | Room temperature |
| **TE Buffer** | 10 mM Tris-HCl, 1 mM EDTA (pH 8.0) | Dissolution and storage of purified DNA | Room temperature or 4°C |

**Preparation of Reagents**

1. Solution 1 (Lysis Buffer)

Prepare the lysis buffer by dissolving 50 mM Tris-HCl (pH 8.0), 20 mM EDTA (pH 8.0), and 2% sodium dodecyl sulfate (SDS) in double-distilled water. Mix thoroughly until all components are dissolved completely. Sterilise the solution by autoclaving and store at 4°C.

2. Solution 2 (Protein Precipitation Solution)

Prepare a 6 M saturated sodium chloride (NaCl) solution in double-distilled water. Sterilise by autoclaving and store at 4°C.

3. Proteinase K

Prepare a 20 mg/mL Proteinase K solution according to the manufacturer's instructions. Aliquot into sterile microcentrifuge tubes to avoid repeated freeze–thaw cycles and store at −20°C.

4. Ethanol

Molecular biology grade 100% ethanol should be used for DNA precipitation. For washing the DNA pellet, freshly prepared 70% ethanol using nuclease free water should be used.

5. TE Buffer

Prepare TE buffer containing 10 mM Tris-HCl and 1 mM EDTA adjusted to pH 8.0. Store at room temperature or in 4°C until use.

**Protocol**

1. Sample Preparation

1. Remove nearly 20-30 mg of alcohol-preserved tissue.
2. Wash tissue once with Tris buffer (pH 8.0) in order to remove residual ethanol.
3. Centrifuge briefly and discard the wash buffer.

2. Tissue Homogenization

1. Transfer tissue into a sterile 1.5 mL microcentrifuge tube.
2. Add 500 µL of Solution 1 (lysis buffer).
3. Use sterile homogeniser for tissue homogenisation until no visible tissue fragments remain.

3. Protein Digestion

1. Add 5 µL Proteinase K (20 mg/mL).
2. Mix gently.
3. Incubate at 55°C for 2 hours in a water bath.
4. Occasionally mix during incubation in order to improve digestion.

4. Protein Precipitation

1. Cool the lysate on ice for 10 minutes.
2. Add 250 µL saturated NaCl solution (6 M).
3. Until thoroughly mixed, invert tube gently several times
4. Incubate on ice for 5 minutes.

5. Centrifugation

1. Centrifuge at 8,000 rpm for 15 minutes.
2. Transfer the clear supernatant carefully to a fresh labelled 1.5 mL microcentrifuge tube, avoid disturbance of the protein pellet.

6. DNA Precipitation

1. Add two volumes of ice-cold absolute ethanol to the recovered supernatant.
2. Until DNA begins to precipitate, mix gently by inversion u
3. Incubate at −20°C overnight to maximise DNA recovery.

7. DNA Recovery

1. Centrifuge at 11,000 rpm for 15 minutes.
2. Discard the supernatant carefully without disturbing the DNA pellet.

8. DNA Washing

1. Add 250 µL of ice-cold 70% ethanol.
2. Centrifuge at 11,000 rpm for 5 minutes.
3. Remove ethanol carefully

9. DNA Drying

Allow the DNA pellet to air-dry with the tube lid open for 5-10 minutes. Because excessively dried DNA is difficult to dissolve, avoid overdrying.

10. DNA Resuspension

1. Dissolve the DNA pellet in 50-200 µL TE buffer (pH 8.0) depending on pellet size.
2. Mix gently using a wide-bore pipette tip until it gets completely dissolved.
3. Store purified genomic DNA at −20°C for routine use or −80°C for long-term storage.

Laboratory Precautions

- Wear disposable gloves throughout the procedure.
- Use sterile, DNase and RNase free tubes and pipette tips.
- Avoid vigorous vortexing after DNA precipitation in order to prevent DNA shearing.
- Always use molecular biology-grade reagents.
- Label all tubes prior to extraction procedure.
- Minimise repeated freeze-thaw cycles of purified DNA to maintain its integrity.

**Reference**

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research, 16*(3), 1215. https://doi.org/10.1093/nar/16.3.1215

Nasiri, H., Forouzandeh, M., Rasaee, M. J., & Rahbarizadeh, F. (2005). Modified salting-out method: High-yield, high-quality genomic DNA extraction from whole blood using laundry detergent. *Journal of Clinical Laboratory Analysis, 19*(5), 229–232. https://doi.org/10.1002/jcla.20083

Shaik, M., Shivanna, D. K., Kamate, M., Vedamurthy, A. B., & Vinod, K. T. P. (2016). Single lysis-salting out method of genomic DNA extraction from dried blood spots. *Journal of Clinical Laboratory Analysis, 30*(6), 1009–1012. https://doi.org/10.1002/jcla.21972

# Effective RNA Extraction: Practical Insights

Vinayak S

Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018

vinayakmanu2@gmail.com

## Introduction

The central dogma of molecular biology emphasises the roles of DNA, RNA, and proteins. DNA serves as a repository of genetic information, while the functional products of translation appear in the form of proteins . RNA, often considered as a transcriptional product of DNA, plays a vital intermediary role in gene regulation, catalysis, and cellular processes, serving a vital component in bridging genetic information with protein synthesis and regulation.

Advancements in basic and clinical research, particularly in high-throughput technologies, depend on the availability of high-quality RNA.Among various RNA isolation methods Trizol method being a single-step approach has gained widespread popularity due to its ability to significantly reduce isolation time while maintaining RNA integrity . Being a Isotropic solution of phenol and guanidine isothiocyanate it preserves RNA integrity during tissue homogenization while simultaneously lysing cells and breaking down their components.Separation of nucleic acids between aqueous and organic phases is determined by the Phenol pH level. RNA retains in the aqueous layer, while DNA partitions into the organic phase due to the easier neutralization of its phosphate groups. An acidic environment also suppresses RNase activity, protecting RNA. Chloroform addition and centrifugation further aid phase separation, with RNA being retrained in the aqueous phase.TRIzol enables the simultaneous extraction of RNA, DNA, and protein from a common sample, including small RNAs like microRNAs, piRNAs, and siRNAs.

Isopropanol precipitation facilitates the recovery of RNA from the aqueous phase, while DNA and proteins are sequentially separated from the organic phase. The addition of isopropanol to the organic phase precipitates proteins, whereas DNA can be isolated from the interphase by treating with ethanol.

## Materials & methods

- DEPC-treated water (Ambion)

- TRIzol Reagent (Invitrogen)
- ce cold PBS
- Cell scraper
- 70% ethanol
  - Isopropyl alcohol

**Procedure**

1. Homogenize 50–100 mg of biological tissue in 1 mL of TRIzol reagent using a tissue homogenizer. Ensure that the sample volume does not exceed 10% of the total TRIzol volume used for homogenization.
2. Incubate the homogenate at room temperature (20–25°C) for 5 minutes to allow complete dissociation of nucleoprotein complexes.
3. Add 0.2 mL of chloroform for every 1 mL of TRIzol reagent used. Securely cap the tubes and vortex vigorously for 15 seconds.
4. Incubate the mixture at room temperature for 3 minutes, followed by centrifugation at 12,000 × g for 15 minutes at 4°C.
5. Carefully transfer the clear upper aqueous phase containing RNA into a new 1.5 mL RNase-free microcentrifuge tube, avoiding disturbance of the interphase.
6. Add 0.5 mL of isopropanol per 1 mL of TRIzol reagent initially used for homogenization. Mix gently by inverting the tube five times.
7. Incubate the mixture at room temperature (15–30°C) for 10 minutes, followed by centrifugation at 12,000 × g for 10 minutes at 4°C to pellet the RNA.
8. Carefully discard the supernatant and wash the RNA pellet with 1 mL of 75% ethanol for every 1 mL of TRIzol reagent used during homogenization.
9. Vortex briefly and centrifuge at 7,500 × g for 5 minutes at 4°C.
10. Discard the ethanol carefully and air-dry the RNA pellet for 5–10 minutes, ensuring that it is not over-dried.
11. Dissolve the RNA pellet in 20–50 µL of RNase-free (nuclease-free) water by gentle pipetting. If necessary, incubate at 55–60°C for 10 minutes to facilitate complete dissolution.
12. Quantify the RNA using a NanoDrop spectrophotometer by measuring the absorbance ratios at A260/A280 and A260/A230 to assess RNA concentration and purity.

13. Use the purified RNA immediately for downstream applications such as cDNA synthesis, RT-PCR, qPCR, RNA sequencing, or transcriptome analysis, or store it at −80°C for long-term preservation. (Storage at −20°C is suitable only for short-term use.)

**Suggested Readings**

Pagani, S., Maglio, M., Sicuro, L., Fini, M., Giavaresi, G., & Brogini, S. (2023). RNA extraction from cartilage: issues, methods, tips. International journal of molecular sciences, 24(3), 2120.

Gayral, P., Weinert, L., Chiari, Y., Tsagkogeorga, G., Ballenghien, M., & Galtier, N. (2011). Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. *Molecular ecology resources*, *11*(4), 650-661.

Botling, J.; Micke, P. Fresh Frozen Tissue: RNA Extraction and Quality Control. In *Methods in Biobanking*; Dillner, J., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, USA, 2011; pp. 405–413. ISBN 978-1-59745-423-0.

# cDNA synthesis

Abhinav Ashok
Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018
abhinavashok8@gmail.com.

## Introduction

cDNA synthesis is an important step in both gene expression studies and transcriptomic analyses. Therefore, understanding the principles underlying the synthesis of cDNA from a precursor RNA molecule is of considerable importance. The synthesis of cDNA is based on the process of reverse transcription, in which DNA is synthesized from an RNA template, typically messenger RNA (mRNA) or microRNA (miRNA). cDNA synthesis generally involves first-strand synthesis followed, in some applications, by second-strand synthesis (Cowell and Austin 2008).

cDNA has widespread applications in molecular biology. It is extensively used in gene expression studies because cDNA is considerably more stable than RNA, making it suitable for downstream molecular analyses (Zhumabayeva et al. 2001). Since cDNA lacks intronic sequences, it is also valuable in cloning experiments, allowing eukaryotic genes to be expressed in bacterial cloning systems (cDNA Cloning - an overview | ScienceDirect Topics n.d.). Another important application of cDNA is the construction of cDNA libraries, which serve as the foundation for transcriptomic studies (Cabanski et al. 2014). These libraries facilitate the identification of novel transcripts, differential gene expression analysis, and the characterisation of alternative splicing events. More recently, the importance of cDNA has been further highlighted in molecular diagnostics, particularly for RNA viruses, where reverse transcription-polymerase chain reaction (RT-PCR) forms the basis of many diagnostic assays (Resuehr and Spiess 2003).

Total RNA is generally used following its isolation from the tissue or organism of interest. To selectively synthesise cDNA from messenger RNA (mRNA), oligo(dT) primers are commonly employed. These primers specifically bind to the poly(A) tail at the 3′ end of eukaryotic mRNA. In certain cases, random hexamers or gene-specific primers are used instead of oligo(dT) primers. Random hexamers are useful when working with degraded RNA samples or RNA transcripts that lack a poly(A) tail. These primers initiate cDNA synthesis from the RNA template.

The enzyme reverse transcriptase catalyses the synthesis of complementary DNA by using RNA as the template. Following first-strand synthesis, RNase H is commonly used to degrade the RNA strand of

the RNA–DNA hybrid, thereby completing the first-strand cDNA synthesis. In some cases, a second complementary DNA strand is then synthesised using DNA polymerase to generate double-stranded cDNA.

If the yield of cDNA obtained after reverse transcription is low, necessitating troubleshooting of the reaction conditions. Poor cDNA yield is often associated with RNA degradation or inefficient primer annealing. In such cases, the use of RNase inhibitors and optimisation of primer concentration can improve cDNA yield (Kuang et al. 2018). Certain genes or transcripts may possess a high GC content, which can promote the formation of stable secondary structures that interfere with reverse transcription. The inclusion of additives such as dimethyl sulfoxide (DMSO) or betaine can help destabilise these secondary structures and improve the efficiency of cDNA synthesis (Spiess and Ivell 2002).

## Procedure

1. For first strand synthesis, prepare a mixture of the following reagents. Make sure to use a sterile and nuclease free PCR tube. Add template RNA (up to 1µg), an appropriate primer (either Oligo dT or Random Hexamer primers-2µL), 1µL of 10mM dNTP mix and make up the volume to 10µL using nuclease-free water.
2. The contents of the tube is to be gently mixed and then incubated at $65^0$C for 5 minutes. This will denature all the secondary structures formed in the template RNA.
3. Transfer the tube onto ice and keep it for 1 minute to slowly allow the primers to anneal.
4. Add 2µL of 10X Reverse Transcription Buffer, 2µL of reverse transcriptase enzyme, 0.2µL of RNase Inhibitor and make up the volume to 20µL using Nuclease-free water.
5. Incubate the reaction mixture for one hour at $42^0$C. In case the random hexamer primers are in use, it is advisable to incubate the samples at $25^0$C for 5 mins before the $42^0$C incubation.
6. After the incubation period is over, the samples need to be kept in $80^0$C for 5 minutes to ensure that the enzyme is inactivated.
7. Store the cDNA at $-20^0$C.
8. For second strand synthesis (optional), RNA removal must first be performed. For this, 1µL of RNase H to the cDNA.
9. For effective degradation, the mixture is to be incubated for 20 minutes at $37^0$C. If not for immediate use, store in $-20^0$C.
10. In a PCR tube containing first strand cDNA, add 5µL of 10X Second-strand synthesis Buffer, 1µL of dNTP mix, 2µL of DNA polymerase I, 1µL of RNase H, 0.5 µLs of T4 DNA ligase and T4 DNA polymerases each and make up the volume to 50 µL using Nuclease-free water.
11. Gently mix the contents of the tube and incubate the mixture for 2 hours at $16^0$C.

12. In order to terminate the ongoing reaction, either add EDTA or heat the reaction mixture up to $70^0$C for 10 minutes.
13. Store the DNA at -$20^0$C.

**Reference**

Cabanski, Christopher R., Vincent Magrini, Malachi Griffith, Obi L. Griffith, Sean McGrath, Jin Zhang, Jason Walker, et al. 2014. "cDNA Hybrid Capture Improves Transcriptome Analysis on Low-Input and Archived Samples." *The Journal of Molecular Diagnostics : JMD* 16(4): 440–51. doi:10.1016/j.jmoldx.2014.03.004.

Cowell, Ian G., and Caroline A. Austin. 2008. *cDNA Library Protocols*. Springer Science & Business Media.

Kuang, Jujiao, Xu Yan, Amanda J. Genders, Cesare Granata, and David J. Bishop. 2018. "An Overview of Technical Considerations When Using Quantitative Real-Time PCR Analysis of Gene Expression in Human Exercise Research." *PLoS ONE* 13(5): e0196438. doi:10.1371/journal.pone.0196438.

Resuehr, David, and Andrej-Nikolai Spiess. 2003. "A Real-Time Polymerase Chain Reaction-Based Evaluation of cDNA Synthesis Priming Methods." *Analytical Biochemistry* 322(2): 287–91. doi:10.1016/j.ab.2003.07.017.

Spiess, Andrej-Nikolai, and Richard Ivell. 2002. "A Highly Efficient Method for Long-Chain cDNA Synthesis Using Trehalose and Betaine." *Analytical Biochemistry* 301(2): 168–74. doi:10.1006/abio.2001.5474.

Zhumabayeva, B., L. Diatchenko, A. Chenchik, and P. D. Siebert. 2001. "Use of SMART-Generated cDNA for Gene Expression Studies in Multiple Human Tumors." *BioTechniques* 30(1): 158–63. doi:10.2144/01301pf01.

# Nucleic Acid Quantification by Agarose Gel Electrophoresis

Ramitha P A.
Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018
ramitha2394@gmail.com

## Introduction

Electrophoresis is a fundamental analytical technique used to separate charged biomolecules, including nucleic acids (DNA and RNA) and proteins, based on their electrophoretic mobility in an electric field. Biomolecules possess a net electrical charge that allows them to migrate toward the oppositely charged electrode when subjected to an electric field. Several factors, including the net charge, molecular size and shape, the strength of the applied electric field, and the properties of the supporting medium, influence the rate of migration.

A porous gel matrix, like agarose, acts as a molecular sieve, allowing smaller molecules to migrate more rapidly than larger ones. Consequently, separation is achieved through the combined effects of electrophoretic mobility and differential movement through the gel matrix.

It is more commonly used for DNA sequencing, restriction fragment length polymorphism (RFLP) analysis, molecular marker studies, DNA fingerprinting, polymerase chain reaction (PCR) product analysis, nucleic acid purification, and assessment of sample quality and integrity.

## Principle

Electrophoresis is a method used to separate charged molecules from one another based on difference in its charge and size in an applied electric field. The charged molecules under the influence of electric field migrate towards oppositely charged electrodes. Those molecules with a positive charge move towards the cathode and negatively charged molecules move towards Anode. The migration is due to charge on the molecules and potential applied across the electrodes. The fundamental driving force of electrophoresis is voltage applied to the system.

**Agarose gel electrophoresis (AGE)**

Agarose gel electrophoresis is an effective way to separate and analyse DNA. It is also used to separate other charged molecule such as RNA. Agarose is a linear polysaccharide consists of D-galactose and L-galactose purified from seaweed genera *Gelidium* and *Gracilaria*. Agarose gel is used to conduct heat evenly and provide an extra sieving effect. Agarose gels are cast by melting the agarose in the desired buffer until a clear, transparent solution is achieved. Chains of agarose forms helical fibers that aggregate into supercoiled structures which provide sieve like effect. This electrophoresis is used as a diagnostic tool to separate and visualize fragments based on its size. It is based on two key principles: 1) DNA has an overall negative charge (due to the phosphate backbone) and thus will migrate towards a positive charge. 2) A gel acts as a sort of microscopic sieve; smaller DNA fragments will travel through the gel more rapidly than larger fragments. Thus DNA can be separated based on the length (size) of the DNA segment.

**Gel concentration**

The concentration of agarose used depends upon the product size. The greater the percentage of agarose, the smaller the linear DNA that can be resolved. The greater the agarose concentration, the smaller the pores created in the gel matrix, and the more difficult it is for large linear DNA molecules to move through the matrix. Changing the agarose concentration changes the size of the sieve matrix of the gel. The concentration of agarose is inversely proportional to the rate of migration of the DNA fragments, i.e. lower the concentration, the faster is the DNA migration rate and vice versa. Generally, the agarose concentration used is 0.8% to separate DNA fragments of the range 2- 10 kb and 1.2% agarose for the separation of small fragments such as 0.1- 1 kb.

| Recommended % Agarose | Optimum Resolution for Linear DNA |
|---|---|
| 0.5 | 1,000–30,000bp |
| 0.7 | 800–12,000bp |
| 1.0 | 500–10,000bp |
| 1.2 | 400–7,000bp |
| 1.5 | 200–3,000bp |
| 2.0 | 50–2,000bp |

**Electrophoresis buffers**

For electrophoresis, several different buffers can be used. Most commonly used buffers are TAE (Tris-acetate EDTA) and TBE (Tris-borate EDTA). A buffer not only provides ions to support the conductivity, but also establishes a $p^H$. Electrophoresis buffers are usually made up as concentrated solutions and stored at room temperature. The working concentration was prepared as required, usually !X buffers were used for running the sample.

10X TBE

Tris - 108g

Boric acid - 58g

EDTA - 8.5g

Distilled water -1000ml

10X TAE

Tris - 48.4 g

Acetic acid -11.4 ml

EDTA -3.7 g

Distilled water-1000ml

**Staining of DNA**

Ethidium bromide (EtBr) is the most convenient and commonly used staining dye to visualize DNA in agarose gel. It is a fluorescent dye that intercalates between stacked bases of DNA. As it is florescent it emits light when subjected to UV light. Ethidium bromide is a potent mutagen and moderately toxic after an acute exposure. Therefore, it is highly recommended to handle it with considerable caution. Ethidium bromide is prepared as a stock solution of 10mg/ml in distilled water which is stored in amber coloured bottles or bottles wrapped in aluminium foil. Make 0.5µg/ml working solution according to the volume required.

**Gel loading buffer**

Gel loading buffer are mixed with the samples before loading into the well. These buffers are used to increase the density of the sample, ensuring that the DNA sinks evenly into the well and they add colour to the sample, which enable tracking the progress of the electrophoresis. It consists of Bromophenol blue, xylene cyanol and glycerol in water.

**Materials required**

- Agarose
- TAE Buffer
- DNA ladder standard
- Microwave oven
- Electrophoresis chamber
- Power supply
- Gel casting tray and combs
- DNA Loading dye
- Gloves
- Pipette and tips
- Samples

**Procedure**

1. To prepare 0.8% agarose gel, weigh about 0.32g agarose and dissolve in 50 ml of 1X TBE buffer by boiling it in a microwave oven. Melt agarose until it becomes clear solution.
2. Let the solution cool a little, and add ethidium bromide to a final volume of 0.5µg/ml. Mix the gel solution thoroughly by gentle swirling.
3. Place the combs in the gel casting tray in the most tight way to ensure there is no leakage of gel through the sides.
4. Pour the melted agarose solution into the casting tray and let cool until it is solid.
5. Carefully pull out the combs, and place the gel in the electrophoresis chamber. Add enough TAE Buffer so that there is about 2-3 mm of buffer over the gel.
6. Mix the sample of DNA with desired gel loading buffer.
7. Slowly load the sample mixture into the well using a disposable micropipette. Load DNA ladder into the well on either left or right side of the gel to calibrate.

8. Close the lid of the gel tank and attach the electrodes. Connect the electrode wires to the power supply, making sure the positive (red) and negative (black) are correctly connected.
9. Turn on the power supply to about 80-150 volts. The maximum allowed voltage will vary depending on the size of the electrophoresis chamber. Check to make sure the current is running through the buffer by looking for bubbles forming on each electrode.
10. Check to make sure that the current is running in the correct direction by observing the movement of the blue loading dye.
11. Perform electrophoresis until the tracking dye migrates close to the end of the gel, then switch off the power supply.
12. Disconnect the electrode leads from the power supply and carefully remove the gel tray from the electrophoresis chamber. Transfer the gel to a UV transilluminator or gel documentation system and visualise the DNA bands under UV illumination.

**Suggested Reading**

Green, M.R., & Sambrook, J. (2019). Agarose Gel Electrophoresis. *Cold Spring Harbor protocols, 2019 1*, pdb.prot100404 .

# Polymerase Chain Reaction: Practical Insights.

Ramitha P A.
Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018
ramitha2394@gmail.com

## Introduction

The Polymerase Chain Reaction (PCR) is a versatile molecular biology technique that allows the selective amplification of a specific DNA sequence, generating millions of copies from a small amount of template DNA. Developed by Kary Mullis in 1983, PCR has revolutionised molecular genetics by providing a fast, sensitive, and accurate method for DNA analysis.

PCR is based on repeated cycles of three fundamental steps: denaturation, annealing, and extension. During denaturation, double-stranded DNA is separated into single strands by heating. In the annealing step, sequence-specific primers bind to complementary regions of the target DNA. Finally, during extension, a thermostable DNA polymerase synthesises new DNA strands by incorporating deoxyribonucleotide triphosphates (dNTPs). These cycles are performed in a thermal cycler, resulting in exponential amplification of the target DNA fragment.

The success of PCR depends on several critical factors, including the quality and concentration of the template DNA, primer design, enzyme fidelity, reaction components, and optimised cycling conditions. Careful optimisation of these parameters is essential to achieve high specificity, efficiency, and reproducibility while minimising non-specific amplification and primer-dimer formation.

PCR has become an indispensable tool in biological and biomedical research, clinical diagnostics, forensic science, agriculture, veterinary medicine, environmental monitoring, and fisheries research. It is routinely employed for pathogen detection, genetic variation studies, DNA barcoding, gene expression analysis, genotyping, mutation detection, sequencing, and molecular marker-assisted studies.

**Principle**

The process involves amplifying a specific DNA fragment through a series of repeated cycles consisting of three steps, which is performed in a thermal cycler.

**Denaturation**

The initial denaturation step requires heating the reaction to 94–96°C for 3–4 minutes. This thermal elevation breaks the hydrogen bonds between the complementary base pairs, generating the single-stranded DNA templates essential for polymerisation.

**Annealing**

The annealing phase is considered the most critical step in the PCR cycle. During this stage, the reaction temperature is reduced to between 50°C and 65°C for less than one minute to facilitate the hybridisation of complementary oligonucleotide primers to their target sequences on the single-stranded DNA templates. For optimal binding and accuracy, this annealing temperature is generally set approximately 2°C below the calculated melting temperature of the primers.

**Extension**

Following the annealing stage, the PCR cycle proceeds to the extension phase at 72°C. At this optimal temperature, *Taq* polymerase synthesises new complementary DNA strands by sequentially adding nucleotides to the annealed primers in the 5' to 3' direction. To ensure accurate and complete amplification of the target template, the duration of this step is strictly calibrated to the length of the target sequence, generally requiring one minute per 1,000 base pairs.

**Materials Required**

- DNA template
  It contains the target sequence to be amplified. In general 10- 100 ng/ul of DNA template is ideal to get a proper reaction.

- PCR buffer
  The reaction utilises a 10X PCR buffer composed of 100 mM Tris-HCl (pH 8.3 at 25°C), 500 mM KCl, 15 mM $MgCl_2$, and 0.01% gelatin. This formulation provides the optimal chemical environment for DNA amplification by maintaining a stable pH, stabilising the enzyme, and supplying the essential ions necessary for the efficient function of *Taq* DNA polymerase.

- Primers

  Primers are short synthetic DNA strands, typically 16 to 30 nucleotides in length, that establish the precise start and end points of the target region to be amplified. By annealing to the template DNA, the forward and reverse primers restrict replication to a specific sequence and provide the essential starting point for polymerase extension. Optimal primer design requires careful consideration of sequence complementarity, specificity, and melting temperature ($T_m$), and length. Generally, primers should be designed to be as short as possible while maintaining target specificity, with the reaction's annealing temperature set at least 5°C below their calculated $T_m$.

- Taq Polymerase

  *Taq* polymerase is a thermostable enzyme isolated from the bacterium *Thermus aquaticus* that is integral to the Polymerase Chain Reaction (PCR). Its exceptional heat stability allows it to withstand the extreme temperatures required for DNA denaturation without degrading. During the extension phase, this enzyme drives the synthesis of new complementary DNA strands by adding free nucleotides to the annealed primers. Because it survives repeated thermal cycles of denaturation, annealing, and extension, *Taq* polymerase enables efficient and continuous DNA amplification without the need to replenish the enzyme.

- dNTPS

  Deoxynucleotide triphosphates (dNTPs) function as the fundamental substrates for DNA synthesis during the polymerase chain reaction. These monomeric nucleotides are sequentially incorporated into the nascent DNA strand by *Taq* DNA polymerase. The structural composition of dNTPs includes high-energy triphosphate groups; the hydrolysis of these phosphate bonds provides the thermodynamic energy necessary to drive the polymerisation process. To ensure optimal amplification fidelity and yield, the final concentration of dNTPs in the reaction is generally optimised within a range of 20 to 200 μM.

**Procedure**

1. For each fragment of target DNA to be amplified, combine the following reagents in a PCR tube according to the table given below.

| **Component** | **25µl reaction** | **Final concentration** |
|---|---|---|
| 10X standard Taq reaction buffer | **2.5µl** | 1X |
| 10mM dNTPs | **.5ul** | 200µM |
| 10µM forward primer | **.5µl** | .2µM(.05-1uM) |
| 10µM reverse primer | **.5µl** | .2µM(.05-1uM) |
| Template DNA | **variable** | <1000ng |
| Taq polymerase | **.125µl** | 1.25units per50µl PCR |
| Nuclease-free water | **To 25µl** | - |

Gently mix the components. Collect all liquid to the bottom of the tube by a quick spin if necessary.

2. Place the samples into a thermal cycler and perform the thermal cycling under appropriate conditions. Thermal cycling conditions for a routine PCR

| **Step** | **Temparature** | **Time** | |
|---|---|---|---|
| **Initial denaturation** | $94^0$C | 4-5 minutes | |
| **Denaturation** | $94^0$C | 30seconds | 32 CYCLES |
| **Annealing** | $42^0$C-$68^0$C | 30 seconds | |
| **Extension** | $72^0$C | 1 minute/kb | |

| | | |
|---|---|---|
| **Final extension** | $72^0$C | 7 minutes |
| **Hold** | 4-$10^0$C | |

The duration and temperature used for each phase of amplification and the total number of cycles required depend on a particular aspect of the reaction being performed.

3. If you need to visualise the amplified product before down stream process, you can perform 1- 1.5% agarose gel electrophoresis with a specific buffer (either TBE or TAE) and perform the analysis.

**Different Types of PCR and Their Applications.**

**Nested PCR**

To increase the target DNA sequences' PCR yield, nested sets of primers can be employed. As in Nestled PCR uses two primer sets, with one primer set being used for 15–30 cycles in the first round of PCR and a second set of primers being used for the second round. For example, internal region of the initially amplified DNA acts as a template for a further 15–30 cycles. Therefore, the nested PCR technique improves the sensitivity and specificity. Because this method nearly always removes any erroneous nonspecific amplification products, the intended target sequence is preferentially amplified because the products are unlikely to be complementary enough to the nested primers to act as a template for additional amplification. However, a disadvantage of this high sensitivity is the increased potential of contamination, caution must be used when doing such PCRs, especially in a diagnostic laboratory.

**Hot start PCR**

Hot-Start PCR is an advanced version of traditional PCR designed to enhance specificity and efficiency by addressing the problem of non-specific amplification caused by premature activation of the DNA polymerase at low temperatures. This issue often leads to the formation of undesired products during the reaction setup.

Hot-Start PCR prevents this by keeping the DNA polymerase inactive until the reaction is heated. Several methods are used to implement Hot-Start PCR. In physical separation, antibodies or other inhibitors block polymerase activity until high temperatures denature the

inhibitors, activating the enzyme. Chemical modification involves modifying the polymerase to remain inactive at low temperatures and activating it during the initial heating steps. Specialised "Hot-Start Taq Polymerases" are also commercially available, featuring mechanisms like antibody-mediated inhibition or reversible chemical modifications to ensure controlled activation.

For high accuracy in DNA amplification, high-fidelity polymerases like Pfu are used. These enzymes have proofreading activity, which allows them to detect and remove mismatched nucleotides during DNA synthesis, ensuring error-free amplification. Compared to traditional Taq polymerase, Pfu and other proofreading enzymes significantly reduce errors, making them ideal for applications like cloning or working with complex or low-concentration DNA templates. This combination of improved specificity and accuracy makes Hot-Start PCR highly valuable in research and diagnostics.

**Multiplex PCR**

Multiplex PCR uses two or more sets of primers in a single PCR to allow for the simultaneous amplification of several sequences or genes. Many issues, including the increased production of misprimed PCR products, "primer dimers," and the inability to distinguish longer DNA fragments during amplification, may arise from having too many PCR primers in one tube. The primers used for this kind of PCR amplification have comparable annealing temperatures. Differential yields of amplified products will come from significant differences in target DNA lengths, which will favour the amplification of the shorter target over the longer one**.** Furthermore, the Taq polymerase additive included in Multiplex PCR buffers reduces amplicon competition and the ability to distinguish longer DNA fragments during Multiplex PCR. It is possible to further hybridise multiplex PCR results using a probe specific to a gene for confirmation.

**Touch down PCR**

Touchdown PCR is a modified PCR technique that enhances specificity and yield by starting with a high annealing temperature and gradually lowering it during the initial cycles. The high starting temperature ensures that only highly specific primer-template pairs bind, reducing non-specific amplification. As the annealing temperature decreases incrementally, primer binding becomes more efficient, allowing the desired target sequence to be preferentially amplified. This method is highly effective for amplifying complex templates, low-abundance DNA, or

sequences prone to non-specific amplification. Its advantages include improved specificity, reduced background noise, and higher sensitivity, making it ideal for applications such as gene cloning, mutation detection, microsatellite analysis, and amplifying difficult or low-complexity DNA sequences.

**Quantitative PCR**

Quantitative PCR is used to measure the amount or quantity of target nucleic acid (DNA or RNA) in a sample. The amount of fluorescence generated during the phase of the true exponential stage directly measures the amount of nucleic acid or target. Special thermal cyclers fitted with a light source and suitable filters for wavelength selection are used for the real-time detection or monitoring of PCR products. Fluorescent dyes used are SYBR Green or fluorophore-anchored DNA probes, such as TaqMan, to measure the amount of amplified product as the amplification progresses. The creation of non-specific hybridisation is the primary drawback in this case. Using the DNA stocks, each reaction was carried out in quadruplicate.

**Multiplex PCR.**

The Multiplex PCR approach is used to identify exonic/intronic sequences in particular genes and detects multiple pathogens in a single sample. Because primers are designed to bind to specific DNA sequences, their designs differ. Since numerous DNA genes of different sizes are targeted in a single reaction, base pair lengths in multiplex PCR should differ to create discrete bands in order to save costs and time consumption while simultaneously identifying multiple pathogens. This method is employed to identify bacterial, viral, and other infectious agents. The likelihood of primer-dimer amplification and DNA fragment discrimination is increased when one or more primer pairs are present10. The Taq polymerase additive, which is included in multiplex PCR along with buffers, lessens amplicon competition**.**

**LAMP PCR**

LAMP (Loop-mediated Isothermal Amplification) is a DNA/RNA amplification technique performed at a constant temperature (60–65°C), eliminating the need for thermal cycling. It uses 4–6 primers targeting multiple regions of the target sequence and a strand-displacing polymerase, achieving rapid and highly sensitive amplification within 30–60 minutes. Resultscan be visualised using turbidity, fluorescence, or colourimetric changes. LAMP is

widely used in pathogen detection (e.g., COVID-19, malaria), agriculture, food safety, and environmental monitoring. Its simplicity and speed make it ideal for diagnostics in resource-limited settings. Key advantages include no thermal cycler requirement, high sensitivity, and rapid results. However, it requires careful primer design, and non-specific amplification may occur if conditions are not optimised. LAMP is a powerful tool for fast and reliable molecular diagnostics.

**Digital PCR**

Digital polymerase chain reaction (dPCR) is a cutting-edge technique used for the absolute quantification of nucleic acids by partitioning a sample into thousands or even millions of individual reactions. Each partition functions as an independent PCR reaction, allowing precise detection and amplification of single DNA or RNA molecules. This method is particularly valuable for applications requiring high sensitivity and specificity, such as detecting rare genetic mutations, monitoring minimal residual disease in cancer patients, and identifying low-abundance pathogens. Unlike quantitative PCR (qPCR), dPCR does not rely on standard curves for quantification, reducing variability and improving accuracy. Furthermore, dPCR has demonstrated superior resistance to inhibitors and increased reproducibility, making it an invaluable tool in clinical diagnostics, environmental monitoring, and research. For instance, demonstrated the potential of dPCR in detecting low-frequency mutations with remarkable precision.

**Applications Of PCR.**

- Medical Diagnostics: Detecting genetic mutations, identifying pathogens (e.g., viruses, bacteria), and diagnosing infectious diseases (e.g., COVID-19, HIV).
- Forensic Science: Amplifying DNA from crime scene samples for identification, paternity testing, and genetic fingerprinting.
- Genetic Research: Cloning genes, sequencing DNA, and studying gene expression, mutations, and genetic variation.
- Agriculture: Detecting GMOs, identifying plant pathogens, and aiding in crop breeding and improvement.
- Environmental Monitoring: Detecting microbial contamination, monitoring biodiversity, and studying environmental DNA (eDNA) in ecosystems.

**Suggested Readings**

Hindson, C. M., Chevillet, J. R., Briggs, H. A., Gallichotte, E. N., Ruf, I. K., Hindson, B. J., Vessella, R. L., & Tewari, M. (2013). Absolute quantification by droplet digital PCR versus analogue real-time PCR. *Nature Methods*, *10*(10), 1003–1005. https://doi.org/10.1038/nmeth.2633

Joshi, M., & Deshpande, J. D. (2010). Polymerase chain reaction: methods, principles and application. In the *International Journal of Biomedical Research*. www.ssjournals.com

L, L. S. (2023). A Review of PCR, Principle, and Its Applications. *International Journal of Pharmaceutical Research and Applications*, *8*(3), 3502–3513. https://doi.org/10.35629/7781-080335023513

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., & Hase, T. (2000). Loop-mediated isothermal amplification of DNA. In *Nucleic Acids Research* (Vol. 28, Issue 12).

# DNA Barcoding: A Molecular Approach for Species Identification

Vysakh V G.
Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018
vishakvg99@gmail.com.

## Introduction

Accurate species identification of biological species forms the foundation of taxonomy, biodiversity assessment, conservation biology, fisheries management, ecological monitoring and sustainable utilisation of biological resources. Conventional species identification is primarily based on morphological characteristics. Morphological identification is often constrained by sexual dimorphism, ontogenetic variation, phenotypic plasticity, cryptic species complexes, and damaged and processed biological materials. These challenges are pronounced in the case of marine organisms because many marine species exhibit highly similar morphological features and several species undergo complex life cycles involving morphologically distinct larval, juvenile and adult stages. In addition, several seafood products are often marketed as fillets, frozen meat, dried products, or processed foods. In these products, the morphological characters are absent, making species identification difficult or impossible using traditional taxonomic approaches.

Developments in molecular biology transformed species identification by providing DNA-based approaches. These approaches are independent of developmental stage, environmental influences, or specimen condition. Among these approaches, DNA barcoding is the most popularly adopted molecular tool for rapid, reliable, and standardised species identification. The concept of DNA barcoding was introduced by Hebert and colleagues in 2003. According to them, a short, standardised DNA sequence could serve as a universal molecular identifier for animal species, analogous to the Universal Product Code (UPC) barcode used for commercial products. This DNA barcoding concept revolutionised biological species identification and revealed that species can be accurately identified by using sequence variation within a small genomic region rather than relying solely on morphological features.

Genetic variation within the species is considerably lower than the variation observed between different species, which is the basic principle of DNA barcoding. Sequencing an appropriate DNA region from an unknown specimen and comparing it with an authenticated reference sequence from a database allows accurate identification of the species. The ultimate success of DNA barcoding depends on the selection of an appropriate genetic marker. The selected marker should be capable of reliably differentiating one species from another. An ideal barcode region should be present in all members of a taxonomic group, be easily amplified using universal primers, possess sufficient sequence variation to discriminate among species while remaining relatively conserved within species, and be short enough to allow amplification even from degraded DNA samples. In animals, mitochondrial DNA (mtDNA) satisfies these criteria more effectively than nuclear DNA and has therefore become the preferred source of barcode markers. Mitochondrial DNA is present in hundreds to thousands of copies per cell, greatly increasing the probability of successful DNA amplification, particularly from small, degraded, or processed biological samples. The mitochondrial genome is relatively small, lacks introns, undergoes limited recombination, and it is maternally inherited in most animals. These characteristics make sequence comparisons more straightforward. Mitochondrial genes evolve at a faster rate than many nuclear genes, generating sufficient nucleotide variation between closely related species and maintaining low variation within species. These characteristics make mitochondrial DNA an ideal target for the identification of species.

Among the thirteen protein-coding genes present in the animal mitochondrial genome, the cytochrome *c* oxidase subunit I (COI) gene has been established as the universal DNA barcode for animals. The COI gene encodes a subunit of cytochrome *c* oxidase, the terminal enzyme of the mitochondrial electron transport chain, and is present in virtually all animal taxa. Although the gene performs an essential biological function and is therefore sufficiently conserved to permit the design of universal primers, it also contains highly variable nucleotide regions that provide excellent species-level discrimination. A fragment of approximately 650 bp from the 5′ region of the COI gene was proposed by Hebert et al. (2003) as the standard DNA barcode for animals because it exhibits sufficient sequence variation to distinguish closely related species while remaining highly conserved within individuals of the same species. This distinct difference between low intraspecific variation and high interspecific variation, commonly known as the barcode gap, enables accurate and reliable species identification. The availability of standardised COI primers and extensive reference databases, such as the Barcode of Life

Data System (BOLD) and GenBank, has enabled the generation and comparison of barcode sequences across thousands of animal species.

Although the mitochondrial cytochrome *c* oxidase subunit I (COI) gene is regarded as the standard DNA barcode for animals, several other mitochondrial and nuclear genetic markers are also employed depending on the taxonomic group and research objective. Among mitochondrial markers, the 16S ribosomal RNA gene is commonly used for species identification and phylogenetic analysis in fishes, molluscs, crustaceans, amphibians, and other aquatic organisms, particularly when COI amplification is unsuccessful or when deeper evolutionary relationships are investigated. The 12S rRNA gene is frequently used in environmental DNA (eDNA) studies and metabarcoding because of its high amplification efficiency from degraded DNA samples. Other mitochondrial markers, such as cytochrome b (cytb), NADH dehydrogenase subunit genes (ND1, ND2, ND4, and ND5), and the highly variable mitochondrial control region (D-loop), are commonly utilised for population genetics, phylogeographic analyses, stock discrimination, and studies of genetic diversity, owing to their relatively high mutation rates and greater intraspecific variation.

In plants and algae, mitochondrial genes evolve too slowly to provide adequate species-level discrimination. Therefore, chloroplast and nuclear markers are generally preferred for DNA barcoding. The chloroplast genes ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*) and maturase K (*matK*) are recognised as the core plant barcode regions because they provide reliable amplification and complementary levels of sequence variation. Additional markers such as the internal transcribed spacer (ITS) region of nuclear ribosomal DNA, the chloroplast trnH–psbA intergenic spacer, and the chloroplast elongation factor gene *tufA* are frequently employed to improve species resolution in flowering plants, macroalgae, and other photosynthetic organisms. Selection of an appropriate barcode marker region depends on the organism interest, the desired taxonomic resolution, and the availability of validated reference sequences in public databases.

The increasing accessibility of polymerase chain reaction (PCR), automated DNA sequencing, and bioinformatics tools has greatly facilitated the widespread application of DNA barcoding. DNA barcoding requires only a small amount of biological material; it can be applied successfully to tissues, blood, fins, scales, eggs, larvae, environmental samples, museum specimens, and even highly processed biological products. The method has therefore become

particularly valuable when morphological identification is impractical or when rapid and accurate identification is required.

**DNA Barcoding Workflow**

Following the selection of appropriate barcode region, DNA barcoding involves a series of standardized laboratory and bioinformatic procedures which ultimately lead to accurate species identification. Individual protocols may vary depending on the organism, barcode gene, and laboratory facilities, the overall workflow remains largely similar across most studies. The major steps include sample collection and preservation, genomic DNA extraction, polymerase chain reaction (PCR) amplification of the barcode region, DNA sequencing, sequence quality assessment, database comparison, and species identification (Figure 1).

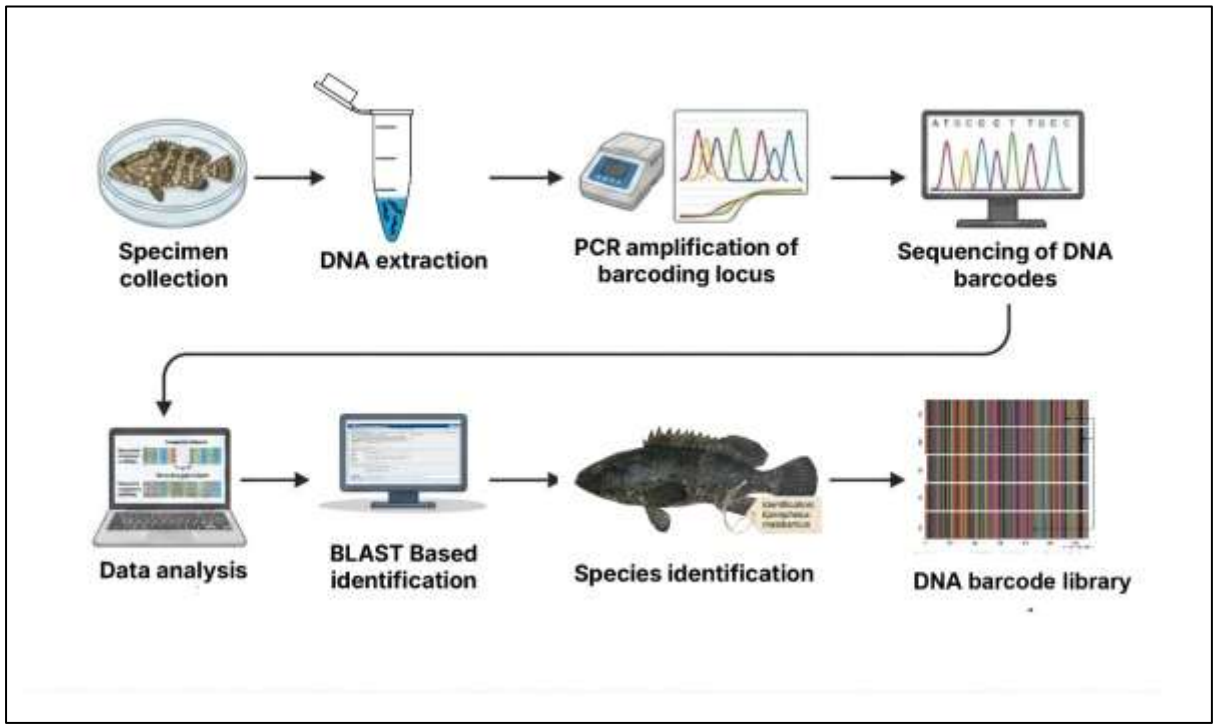



**Figure1**: DNA Barcoding work flow.

**Sample Collection and Preservation**

The quality of DNA obtained for barcoding depends largely on the quality of the biological sample. Fresh tissues generally provide the highest quality DNA and are therefore preferred whenever possible. In fishes, muscle tissue, fin clips, scales, blood, and eggs are generally used for DNA extraction. In molluscs and crustaceans tissues like the adductor muscle, mantle, foot,

pleopods, or walking legs are often used. Tissue samples should be handled aseptically to minimize contamination and should be preserved immediately after collection. Preservation in 95–100% molecular-grade ethanol is the most widely used method because it effectively inhibits nuclease activity and maintains DNA integrity. Samples also stored at −20°C or −80°C for long-term preservation. Formalin fixed specimens are generally not suitable for routine DNA barcoding because formaldehyde causes DNA fragmentation and cross-linking, which ultimately reduce PCR amplification efficiency.

**Genomic DNA Extraction**

After sample collection and preservation, genomic DNA is extracted from the selected tissue. The objective of DNA extraction is to obtain highly purified, intact DNA free from proteins, lipids, polysaccharides, and other contaminants that may inhibit downstream enzymatic reactions. DNA extraction is done using conventional methods such as the phenol–chloroform protocol, cetyltrimethylammonium bromide (CTAB) extraction, salting-out procedures, or commercially available silica membrane and magnetic bead-based extraction kits. Regardless of the extraction method employed, DNA quality should be evaluated before PCR amplification. NanoDrop or flurometric assays, like spectrophotometric instruments, were used for the determination of concentration and purity of DNA. Agarose gel electrophoresis is further used to evaluate the integrity of DNA. High quality DNA shows A260/280 ratio of around 1.8 indicating minimal protein contamination.

**PCR Amplification of the Barcode Gene**

The selected barcode region is amplified using the polymerase chain reaction (PCR), a technique that enables exponential amplification of specific DNA fragments using sequence-specific primers. In fishes and most other animals, the mitochondrial cytochrome *c* oxidase subunit I (COI) gene is amplified using universal or taxon-specific primer pairs. PCR reaction makes use of template DNA, primers (forward and reverse, deoxynucleotide triphosphates (dNTPs), magnesium ions ($Mg^{2+}$), reaction buffer, thermostable DNA polymerase, and nuclease-free water. PCR amplification involves repeated cycles of DNA denaturation, primer annealing, and strand extension, resulting in millions of copies of the target barcode fragment. The amplified product ranges between 600 and 700 bp, depend upon the primers used for amplification. Commonly used primer pairs for fish DNA barcoding are presented in Table 1.

**Table 1.** Commonly used DNA barcode markers and universal primer pairs for different taxonomic groups.

| Target Group | Barcode Marker | Primer Name | Primer Type | Primer Sequence (5′–3′) |
|---|---|---|---|---|
| **Vertebrates (Fish)** | COI | VF2_t1 | Forward | CAACCAACCACAAAGACATTGGCAC |
| | | FishR2_t1 | Reverse | ACTTCAGGGTGACCGAAGAATCAGAA |
| | | FishF1 | Forward | TCAACCAACCACAAAGACATTGGCAC |
| | | FishR1 | Reverse | TAGACTTCTGGGTGGCCAAAGAATCA |
| | | FishF2 | Forward | TCGACTAATCATAAAGATATCGGCAC |
| | | FishR2 | Reverse | ACTTCAGGGTGACCGAAGAATCAGAA |
| **Invertebrates** | COI | LCO1490_F | Forward | GGTCAACAAATCATAAAGATATTGG |
| | | HCO2198_R | Reverse | TAAACTTCAGGGTGACCAAAAAATCA |

**Abbreviations:** COI, cytochrome *c* oxidase subunit I; ITS, internal transcribed spacer; rbcL, ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit; matK, maturase K; tufA, elongation factor Tu. Degenerate nucleotide codes follow IUPAC nomenclature (e.g., R = A/G, Y = C/T, W = A/T, M = A/C, N = any nucleotide).

**Quantitative and qualitative assessment of PCR products**

The success of PCR amplification is confirmed by agarose gel electrophoresis. Amplified DNA fragments are separated in accordance with size under an electric field. These fragments are visualised after staining with DNA-binding dyes under ultraviolet or blue-light illumination. A single, well defined band of the expected size indicates successful amplification. Multiple bands, weak amplification, or smeared products may indicate non-specific amplification,

degraded DNA, or suboptimal PCR conditions. Verification of PCR products prior to sequencing ensures high-quality sequence data and minimise sequencing failures.

**PCR Product Purification and DNA Sequencing**

PCR products containing residual primers, nucleotides, enzymes, and reaction components are purified prior to sequencing using enzymatic cleanup methods or commercial purification kits. Purified amplicons are subsequently subjected to DNA sequencing. Most commonly, sequencing is done by the Sanger sequencing method, which remains the gold standard for routine DNA barcoding because of its high accuracy and reliability. Sequencing is generally performed in both forward and reverse directions to improve sequence quality and enable the generation of a consensus sequence by resolving ambiguous nucleotide positions.

**Sequence Editing and Quality Assessment**

Raw sequence chromatograms obtained from the sequencing platform require careful examination before species identification. Low-quality nucleotide regions at the beginning and end of the sequence are trimmed, ambiguous base calls are corrected, and forward and reverse reads are assembled to generate a high-quality consensus sequence. Sequence editing can be performed using several software such as Chromas, BioEdit, Geneious, MEGA, or similar bioinformatics tools. The final sequence should be examined for sequencing artefacts, insertions, deletions, unexpected stop codons, or frame-shift mutations that may indicate poor sequence quality or amplification of nuclear mitochondrial pseudogenes (NUMTs).

**Sequence Comparison and Species Identification**

The edited consensus sequence is compared with authenticated reference sequences available in public databases for example the Barcode of Life Data System (BOLD) and NCBI GenBank. Similarity searches are typically performed using the BOLD Identification Engine or the Basic Local Alignment Search Tool (BLAST). Species identification assigned on the basis of the highest sequence similarity together with sequence coverage, alignment quality, and taxonomic reliability of the reference sequence. Sequence identities of approximately 98–99% or higher are considered indicative of species-level matches. These threshold may vary among taxonomic groups. For closely related species or uncertain identifications, phylogenetic analyses using methods such as Neighbour-Joining, Maximum Likelihood, or Bayesian

Inference are further used to validate species assignments and confirm the taxonomic placement of the query sequence.

The adoption of standardized laboratory protocols, appropriate barcode markers, reliable reference databases, and rigorous quality control at each stage ensures accurate, reproducible, and globally comparable species identification, thereby making DNA barcoding one of the most powerful molecular tools in fisheries, marine biodiversity research, conservation biology, and aquatic resource management.

**Reference**

Hebert PDN , Cywinska A , Ball SL , deWaard JR ( 2003 ) Biological identifications through DNA barcodes . *Proceedings of the Royal Society of London. Series B, Biological Sciences* , 270 , 313 – 322.

Hebert PDN , Penton EH , Burns JM , Janzen DH , Hallwachs W ( 2004 ) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator* . *Proceedings of the National Academy of Sciences, USA* , 101 , 14812 – 14817 .

Inoue JG , Miya M , Tsukamoto K , Nishida M ( 2001 ) A mitogenomic perspective on the basal teleostean phylogeny: resolving higher-level relationships with longer DNA sequences . *Molecular Phylogenetics and Evolution* , 20 , 275 – 285.

Ivanova NV , deWaard JR , Hebert PDN ( 2006 ) An inexpensive, automation-friendly protocol for recovering high-quality DNA . *Molecular Ecology Notes* , 6 , 998 – 1002.

Lobo, J., Costa, P. M., Teixeira, M. A., Ferreira, M. S., Costa, M. H., & Costa, F. O. (2013). Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. *BMC ecology*, 13, 34. https://doi.org/10.1186/1472-6785-13-34

# Real Time PCR practical Aspects and Applications

Ramitha P A.
Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018
ramitha2394@gmail.com

**Introduction.**

Reverse Transcription Polymerase Chain Reaction (RT-PCR) is a modification of the traditional Polymerase Chain Reaction (PCR). Here, RNA is reverse transcribed to complementary Deoxy Nucleic Acid (cDNA) using specific RNA polymerases and expression of individual RNA molecules is visualised using RT-PCR. RT-PCR was created as a result of the discovery of reverse transcriptase in retroviruses by Howard Temin and Davis Baltimore replaces the old Northern blot technique for RNA detection. There is a significant difference observed in between Reverse Transcription PCR (RT-PCR) and qPCR (Real-time PCR). Here RT -PCR will convert the RNA to complementary DNA by specific RNA polymerase and quantified using specific probes, whereas the real-time quantification of amplified product is only done in the case of qPCR. It indirectly benefits the downstream analysis of the sample

**Principle**

In RT-PCR, RNA is first reverse-transcribed into complementary DNA (cDNA) using a reverse transcriptase enzyme, and the cDNA is then amplified through traditional PCR steps. The amplified cDNA is quantified using specifically designed probes for accurate detection using a specialised type of thermal cycler.

**Real-Time PCR Markers**

Real-Time PCR employs several markers, with the most common being the TaqMan probe and SYBR Green.

- **Taqman Probe**

The Taqman probe is a hydrolysis probe consisting of a reporter dye, typically fluorescein (FAM), attached to its 5’ end, and a quencher dye, tetramethylrhodamine (TAMRA), attached

to its 3’ end. Under normal conditions, the probe remains coiled, bringing the reporter dye close to the quencher, which suppresses its fluorescence.

The probe's oligonucleotide has a homologous region that binds to the target gene in the sample DNA. During the extension stage, Taq polymerase synthesises a new DNA strand and degrades the probe via 5’ nuclease activity. This releases the reporter dye from the quencher, generating a fluorescence signal. As the PCR cycles progress, the number of signal molecules increases, resulting in a fluorescence intensity that correlates with the amplification of the target DNA.

**SYBR Green**

SYBR Green is a dye that emits a strong fluorescent signal when it binds non-specifically to the minor groove of DNA. Up on binding with minor groove it emits flurosence 1000 times higher than in the solution.

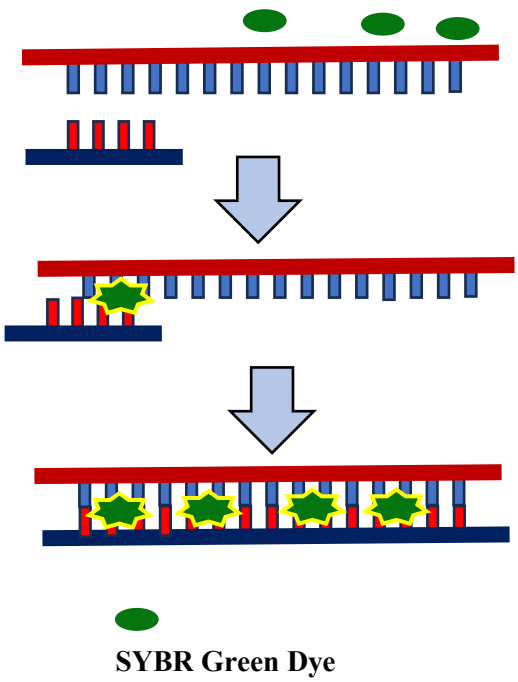



**Figure 1**: Mechanism of action: SYBR Green dye

SYBR Green is advantageous over the Taqman probe because it provides data on each amplification cycle and the melting temperature of the DNA. It also does not require a specific design of special reporters for signal identification because it has the ability for that by the dye

itself. which the Taqman probe does not offer. However, SYBR Green has lower specificity compared to the Taqman probe.

**Materials And Reagents Required**

- One – step RT-PCR kit
- RT-PCR primer
- Isolated RNA from the sample
- RNase inhibitor
- Nuclease-free water
- Thermal cycler
- PCR tube

**Procedure**

Real-time PCR steps have been divided into two steps: one-step and two-step, depending on the requirement of the enzyme and the reaction.

**One-Step RT-PCR Protocol**

The one-step RT-PCR protocol directly converts mRNA into cDNA via reverse transcription, followed by PCR amplification, all within a single reaction tube using sequence-specific primers.

1. Prepare a reaction mix containing dNTPs, sequence-specific primers, necessary enzymes (reverse transcriptase and DNA polymerase), and a buffer solution.
2. Dispense the reaction mix into PCR tubes for each reaction.
3. Add the RNA template to the reaction mix in each tube.
4. Place the PCR tubes in a thermal cycler and run the program. The thermal cycling includes reverse transcription to synthesize cDNA, followed by 40–50 cycles of amplification consisting of denaturation, annealing, and elongation. During the initial denaturation step, reverse transcriptase is inactivated, and Taq DNA polymerase amplifies the cDNA exponentially.
5. Visualize the RT-PCR products by agarose gel electrophoresis to confirm amplification and verify product size.

**Two-Step RT-PCR Protocol**

The two-step RT-PCR method separates reverse transcription and PCR amplification, offering increased sensitivity and flexibility by allowing the use of various primers, such as oligo(dT) primers (targeting the poly(A) tail of eukaryotic mRNA) or random hexamers (binding throughout RNA sequences, including mRNA, tRNA, or rRNA).

**Step 1: Reverse Transcription**

1. Prepare a reaction mix containing dNTPs, primers (oligo(dT) or random hexamers), reverse transcriptase, and a buffer solution.
2. Add RNase inhibitor to the mix to prevent RNA degradation.
3. Dispense the prepared reaction mix into PCR tubes for each reaction and add the RNA template.
4. Place the tubes in a thermal cycler. The program includes primer annealing, cDNA synthesis, and inactivation of the reverse transcriptase enzyme. The synthesized cDNA can be used immediately or stored at -20°C for later use.

**Step 2: PCR Amplification**

1. Prepare a reaction mix containing buffer, dNTPs, $MgCl_2$, Taq DNA polymerase, and nuclease-free water.
2. Dispense the mix into PCR tubes for each reaction and add sequence-specific primers.
   Note: Separate reactions for target and housekeeping need to be kept to assess the amplification as well as the expression of that particular gene in the sample.
3. Place the PCR tubes in the thermal cycler and run 30–35 cycles of amplification, each consisting of denaturation, annealing, and elongation.
   Note: The annealing temperature needs to be determined separately by running conventional PCR with your desired primer for the study.
4. Visualise the PCR products using agarose gel electrophoresis to confirm successful amplification and verify product size using the ladder.

**Quantification of the sample**

Relative quantification measures the fold change in target gene expression relative to a stable reference gene using qPCR for both. By calculating ΔCt (target - reference) and ΔΔCt (sample - control), fold changes are determined using the formula 2^(-ΔΔCt), which will provide the expression levels of each gene.

**Advantages of Real-Time PCR**

Real-Time PCR offers numerous advantages over traditional PCR techniques:

- It provides immediate insight into which reactions have worked and which have failed.
- The efficiency of the reaction can be accurately calculated.
- Melt curve analysis eliminates the need to run PCR products on a gel.
- It enables truly quantitative analysis of gene expression, whereas traditional PCR is only semi-quantitative.
- It is faster than traditional PCR.
- It simplifies the quantification of samples.

Unlike traditional PCR, Real-Time PCR allows for the automatic determination of multiple reaction successes after just a few cycles and avoids the issue of false negatives.

**Applications of Real-Time PCR**

Real-Time PCR has diverse applications across various fields:

- Gene expression analysis can be done in cancer and Drug research.
- Viral quantification in disease and diagnosis
- Gene copy number analysis.

**Suggested Readings**

Artika, I. M., Dewi, Y. P., Nainggolan, I. M., Siregar, J. E., & Antonjaya, U. (2022). Real-Time Polymerase Chain Reaction: Current Techniques, Applications, and Role in COVID-19 Diagnosis. In *Genes* (Vol. 13, Issue 12). MDPI. https://doi.org/10.3390/genes13122387

Arya, M., Shergill, I. S., Williamson, M., Gommersall, L., Arya, N., & Patel, H. R. H. (2005). Basic principles of real-time quantitative PCR. In *Expert Review of Molecular Diagnostics* (Vol. 5, Issue 2, pp. 209–219). https://doi.org/10.1586/14737159.5.2.20

Huggett, J. F., Foy, C. A., Benes, V., Emslie, K., Garson, J. A., Haynes, R., Hellemans, J., Kubista, M., Mueller, R. D., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., Wittwer, C. T., & Bustin, S. A. (2013). The digital MIQE guidelines: Minimum information for publication of quantitative digital PCR experiments. *Clinical Chemistry*, *59*(6), 89902https://doi.org/10.1373/clinchem.2013.20637

Tajadini, M., Panjehpour, M., & Javanmard, S. (2014). Comparison of SYBR Green and TaqMan methods in quantitative real-time polymerase chain reaction analysis of four adenosine receptor subtypes. *Advanced Biomedical Research*, *3*(1), 85. https://doi.org/10.4103/2277-9175.127998

# Overview of Next-Generation Sequencing Techniques

Bibin Xavier V F

Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018

bibinxavier28@gmail.com

## Introduction

Next-generation sequencing (NGS) is a contemporary technique for analyzing genetic material that allows large amounts of DNA or RNA to be sequenced quickly. Unlike traditional sequencing techniques, Next-generation sequencing can process millions of DNA fragments simultaneously, drastically increasing the scale and discovery power of genomic studies. Over the years, DNA sequencing has become a keystone of biology, supporting studies in fields such as population genetics, disease resistance, molecular breeding, healthcare, comparative evolution, etc. The advancement of sequencing technologies over the past three decades, from first-generation methods to advanced fourth-generation platforms, also marks the transformational journey of these fields..

The history of DNA sequencing began in the 1970s with Frederick Sanger's pioneering work at the MRC Centre, University of Cambridge. In 1977, Sanger introduced the chain termination method using dideoxynucleotide analogs, which led to the sequencing of the first complete DNA genome, bacteriophage φX174. Simultaneously, Maxam and Gilbert established a chemical degradation method for DNA sequencing. However, due to its technical complexity and reliance on hazardous chemicals, the Maxam–Gilbert method was overshadowed by Sanger's method, which became the standard due to its competence and practicality.

In 2003, the completion of the first draft of the human genome opened a new era for genomics. However, the low throughput sequencing nature and higher cost of Sanger sequencing emphasized the necessity for a more scalable and efficient approach. In 2005, Roche introduced the 454 sequencing platform which became the first commercially successful Next-generation sequencing platform that can do massively parallel sequencing. This has significantly reduced the time and cost of sequencing large genomes, opening up new opportunities for research and clinical applications.

The ability to process millions of DNA fragments simultaneously, which is not possible with traditional Sanger sequencing, is what led to next-generation sequencing later being named after that feature. This capability is what we call massively parallel sequencing. Unlike Sanger sequencing and microarrays, it enables direct sequencing of unknown genomic fragments. Similarly, the fine-grained nature of next-generation sequencing allows it to detect low-frequency variants and novel transcripts, and much more, not possible with traditional techniques. With this innovation, researchers can generate large datasets cost-effectively and quickly, a key factor that has led to advances in the fields of genomics, transcriptomics, metagenomics, and epigenomics.

Next-generation sequencing platforms are broadly categorized into short-read (second-generation) and long-read (third and fourth-generation) technologies. Short-read technologies such as Ion Torrent and Illumina offer exceptional accuracy and affordability, but the limitation of short read lengths can make resolving complex genomic regions challenging. However, long-read platforms such as Pacific Biosciences and Oxford Nanopore Technologies produce longer reads, which are indispensable for detecting structural variations and resolving repetitive sequences. Although these long-read platforms initially faced issues with accuracy and affordability, advancements in bioinformatics tools and hardware have significantly addressed these challenges.

**Basic Workflow of Next-Generation Sequencing**

Next-generation sequencing follows an organized process that includes four main steps: nucleic acid extraction, library preparation, sequencing, and data analysis. This workflow is designed to be compatible with a variety of sequencing technologies.

**1. Nucleic Acid Extraction**

The first step is to extract high-quality DNA or RNA from the sample. This involves lysing the cells to release the nucleic acids, followed by purification to remove impurities such as proteins, lipids, or other cellular debris. The integrity and quality of the extracted nucleic acids, as well as the extraction of high-molecular-weight DNA or RNA, are of great importance. This is crucial for ensuring reliable results for downstream applications using sequences generated from next-generation sequencing platforms.

**2. Library Preparation**

Library preparation converts purified nucleic acids into a form suitable for sequencing. Typically, this step involves cutting the DNA or RNA into shorter fragments and then attaching

special adapters to them after choosing an appropriate size. These adapters enable them to be ligated to the sequencing platform and facilitate downstream processes such as amplification or direct sequencing. Depending on the platform and application, amplification may be required to generate sufficient signal strength. Workflows will vary depending on the length of the strands chosen for library preparation, depending on whether the sequencing is for long-read sequencing or short-read sequencing.

**3. Sequencing**

The sequencing step involves reading the nucleotide sequence of a prepared library. Although the exact methods vary in technology, the general principle relies on detecting nucleotide sequences or monitoring molecular interactions in real-time. The nucleotide sequence can be read by detecting signals generated when a new complementary strand is made or by directly reading DNA molecules. These technologies can generate large amounts of data, and sequencing outputs range from short, highly accurate reads to ultra-long reads that can capture complex genomic regions.

**4. Data Analysis**

Once sequencing is complete, the raw data is subjected to computational processing to extract meaningful insights. It begins with the quality control of reads and the quality improvement interventions. This preprocessing step is essential for efficient downstream applications. This is followed by alignment to reference genomes, de novo assembly if the reference genome is not available, variant detection, transcriptome profiling, metagenomics, functional annotation, etc. The choice of pipeline depends on the goals of the study. It is also crucial to use modern bioinformatics software tools and pipelines that are geared towards our goals, so that researchers can translate their data into actionable conclusions.

**Major Platforms in Next-Generation Sequencing**

In the past few decades, many platforms have been developed since the advent of next-generation sequencing techniques, but only a few have had a lasting impact. The first next-generation sequencing (NGS) technology, the Roche 454, was released in 2005, followed by the Genome Analyzer in 2006, which later revolutionized the omics field after its acquisition by Illumina. This progressed further with the introduction of the PacBio sequencer in 2011, which introduced long-read sequencing, and the Nanopore sequencing device in 2014, both of which helped in the evolution of NGS platforms. Many of the platforms introduced during this

period have lost popularity over time, due to various practical reasons, such as scalability, accuracy, cost-effectiveness issues, and the development of new technologies.

In this section, we focus on four of the most widely used and innovative platforms today: IonTorrent, Illumina, PacBio, and Nanopore. Offering a combination of speed, accuracy, and versatility to meet the diverse needs of research and clinical applications, these platforms have become the cornerstone of modern omics research. Each platform brings unique capabilities and innovations to the field, making them indispensable tools for a wide range of genomic studies. With the constant competition between different platforms on the market, new technologies are emerging, making the future of next-generation sequencing platforms bright.

**1. Ion Torrent Sequencing**

Ion torrent sequencing uses semiconductor technology to detect nucleotide strand synthesis by measuring pH changes. When nucleotide molecules are incorporated into the growing nucleotide strand, H+ ions are released, which causes measurable changes in pH. This eliminates the need for optical detection systems and numerous chemicals, thus simplifying the sequencing process and reducing operational costs. The library preparation process involves DNA fragmentation to a specific length, adapter ligation, and fragment amplification by emulsion PCR to generate clusters. Sequencing is then performed in cycles, i.e., each nucleotide is added in steps, and then pH changes are measured in response to the addition of specific nucleotides.

This platform provides typical read lengths of 200 to 600 base pairs, making it particularly suitable for targeted sequencing and exome studies. Although Ion Torrent sequencing has high accuracy and is cost-effective, it has some challenges like high error rates in homopolymer regions, short read lengths, and low throughput. Despite these limitations, its ease of use and scalability make it a viable tool for genomic research where low-throughput sequencing is sufficient.

Ion Torrent technology uses a semiconductor chip to directly convert chemical information into digital data. This process of recording DNA molecules in real time is simple because it does not require light scanning or cameras to perform such a procedure. A few of the platforms under Ion Torrent include Ion Torrent Genexus Systems and Ion GeneStudio S5, which allow flexibility in sequencing for numerous applications, making the technology available worldwide. As it continues to develop, Ion Torrent is expected to play a major role in affordable and reliable diagnostics and genomic research.

**2. Illumina Sequencing**

Illumina sequencing is based on sequencing-by-synthesis technology. This process uses fluorescently labeled, reversible terminator nucleotides. As the nucleotides are incorporated into DNA strands during sequencing cycles, fluorescence signals are emitted and recorded by a camera, allowing for base identification. The library preparation process involves fragmenting the DNA, ligating adapters to DNA fragments of appropriate length, and performing bridge amplification to create clusters in the flow cell, ensuring sufficient signal for accurate base calling.

Illumina short-read sequencing produces results with read lengths ranging from approximately 50 to 300 base pairs with an error rate of 0.1%. This technology is ideal for high-throughput applications such as transcriptomics, metagenomics, and whole-genome sequencing. Although Illumina sequencing is known for its throughput and accuracy, shorter read length can pose challenges in resolving repetitive genomic regions while assembling complex genomes.

Illumina offers a diverse range of systems, from the MiniSeq and iSeq 100 for small-scale projects to the NextSeq 1000 and 2000 for large-scale complex sequencing applications. These platforms provide a variety of output levels to suit different research needs. High accuracy, cost-effectiveness, and flexibility make Illumina a leading player in the modern omics research world, especially in applications that require high accuracy.

**3. PacBio SMRT Sequencing**

PacBio's single-molecule real-time (SMRT) sequencing captures nucleotide incorporation in real time. During the library preparation process, DNA fragments are converted to SMRTbell templates using hairpin adapters, which form a circular template. Sequencing occurs within specially designed wells known as zero-mode waveguides (ZMWs), where fluorescence is recorded during nucleotide incorporation. The platform is known for generating long read lengths, currently averaging 18-20 kilobases, with some reads exceeding 20 kb. It is well suited for applications such as whole-genome sequencing, isoform sequencing, and structural variant analysis.

PacBio Long Read Sequencing mainly uses two approaches: continuous long read sequencing and circular consensus long read sequencing. Continuous long read sequencing maximizes read length but sacrifices accuracy, while circular consensus sequencing generates highly accurate long reads, known as HiFi reads, by sequencing the same DNA molecule multiple times. High-fidelity long reads achieve 99.9% accuracy, comparable to other short-read technologies, making them indispensable for assembling high quality reference genomes and the generation of longer transcript lengths.

Vega and Revio are some of the long-read sequencing systems under PacBio. The PacBio Vega system was recently designed to be used as a benchtop instrument in the lab, showing the future of how easily sequencing can be accessed. Both systems deliver high-quality HiFi sequencing outputs. By combining longer read lengths with exceptional accuracy, PacBio HiFi sequencing enables researchers to more confidently and easily study complex variants and genomic regions that are inaccessible to short-read technologies.

**4. Nanopore Sequencing**

Nanopore sequencing detects changes in ionic current as nucleotides pass through a nanopore embedded in an electro-resistant membrane. The passage of nucleotides creates a unique disruption in the current, which is translated into a signal that is decoded in real time by base calling algorithms. Library preparation for nanopore sequencing involves fragmentation of the nucleotide strand followed by attachment to adapters and a motor protein. Unlike other sequencing techniques, library preparation does not require amplification or further cluster generation. The motor protein guides the DNA into the pore, and hairpin adapters allow bidirectional sequencing. Nanopore sequencing performs direct sequencing of native DNA or RNA, which is a distinct advantage of the nanopore sequencing technique over other platforms. Nanopore sequencing generates reads ranging from 10 to 100 kb in the long-read approach and 100 to 300 kb in the ultra-long approach. This platform facilitates whole genome assembly, structural variant detection, and direct RNA sequencing. Read lengths up to 4 Mb have been recorded using nanopore technology. The nanopore sequencing platform offers a range of instruments, from the high-throughput PromethION to the pocket-sized MinION. These instruments enable researchers to sequence immediately, provide data in real time, and provide rapid insights, even in non-traditional lab environments.

Recently, nanopore sequencing has achieved an accuracy above 99% despite the initial error rate being high. Many factors have contributed to nanopore technology's high accuracy, including improved hardware, improved library preparation, and improved base calling algorithms. Their ability to sequence ultra-long fragments, combined with the scalability and portability of the technology, make them a versatile tool for a variety of applications, including complex genome assembly, pathogen detection, and epigenetic studies. Since nanopore sequencing has a simple underlying principle, they are progressing to the point where they can sequence proteins directly. Nanopore sequencing is expanding its potential in research and clinical settings due to its accessibility and real-time sequencing capabilities.

## Conclusion

With next-generation sequencing, bioscience and diagnostics have been revolutionized, offering unprecedented insights into genomes, transcriptomes, and epigenomes. Today, it is helping to advance personalized medicine and cancer genomics, as well as improving agricultural production with its wide range of applications. Its applications will assist researchers in addressing global challenges such as disease outbreaks, climate change, food security, and environmental sustainability.

Looking ahead, the future of next-generation sequencing holds great promise for continued innovation. Advances in ultra-precise data generation, real-time sequencing, low-cost sequencing, and portable platforms will further revolutionize sequencing. Growth in bioinformatics tools and software, combined with multi-omics approaches, will help advance our understanding of complex biological systems. These advances will enable faster, more accurate, and cheaper solutions, and will position next-generation sequencing as a cornerstone of progress for humanity and our sustainable future. Just as advances in telescope technology have enabled us to explore the vastness of the universe at a macroscopic level, next-generation sequencing technologies are revolutionizing our ability to probe the microscopic complexities of the biological world.

## Suggested Reading


Hu, T., Chitnis, N., Monos, D. and Dinh, A., 2021. Next-generation sequencing technologies: An overview. Human Immunology, 82(11), pp.801-811.

Mandlik, J.S., Patil, A.S. and Singh, S., 2024. Next-Generation Sequencing (NGS): Platforms and Applications. Journal of Pharmacy and Bioallied Sciences, 16(Suppl 1), pp.S41-S45.

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R.P., Banday, S., Mishra, A.K., Das, G. and Malonia, S.K., 2023. Next-generation sequencing technology: current trends and advancements. Biology, 12(7), p.997.

Tyagi, P. and Bhide, M., 2020. History of DNA sequencing. Folia Veterinaria, 64(2), pp.66-73.

# Foundations of Genomics: Core Concepts and Basic Principles

Vysakh V G.
Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018
vishakvg99@gmail.com.

## Introduction

Whole Genome Sequencing (WGS) is a powerful technology that provides a comprehensive view of the genetic makeup of an organism by sequencing its entire genome. This approach is crucial for understanding genetic variation, evolutionary patterns, and biological functions at the molecular level. In fisheries, WGS plays an essential role in studying genetic diversity, and identifying traits related to growth, disease resistance, and environmental adaptation. It aids in the development of improved breeding programs, conservation strategies, and the sustainable management of fish populations, contributing to both aquaculture and wild fisheries. Moreover, WGS allows for the identification of biomarkers, which can be valuable for monitoring environmental changes and ensuring the health of aquatic ecosystems.

## Materials and Methods

### *Sample Collection and DNA Extraction*

1. **Organism Selection**: Fish species are typically selected based on their ecological or evolutionary significance.
2. **Tissue Collection**: Commonly, fin clips, gill tissue, or muscle samples are obtained from adult specimens.
3. **DNA Extraction**: Extract high-molecular-weight genomic DNA with a kit (Qiagen Genomic DNA kit) or a custom protocol (CTAB). Ensure high-quality DNA suitable for both short and long-read sequencing.
4. **QC and Integrity Check**: Evaluate DNA purity and integrity with a NanoDrop spectrophotometer (A260/280 ratio should be ~1.8-2.0). Use a gel electrophoresis system to check for intact genomic DNA.

### *Library Preparation for Sequencing*

1. **Short-Read Sequencing (e.g., Illumina)**:
    - Fragment the DNA into the desired size (350-500 bp).

○ Perform library construction (end-repair, adapter ligation, and PCR amplification).

○ Example tool: *Illumina TruSeq library prep kit.*

2. **Long-Read Sequencing (e.g., PacBio or Oxford Nanopore)**:

○ Long-range DNA extraction methods must be employed.

○ For PacBio: Use the *SMRTbell* library prep kit.

○ For Nanopore: Utilize the appropriate Nanopore sequencing kit for library preparation.

***Sequencing Platforms***

- **Illumina**: Provides high-coverage, short reads. Best for resequencing and reference-based applications.
- **PacBio**: Provides long reads, ideal for de novo assembly and identifying large structural variations.
- **Oxford Nanopore**: Also generates long reads and can be used for on-site or field-based sequencing.

**Sequencing and Initial Data Preprocessing**

***1. Initial Data Assessment***

Run *FastQC* for raw sequence quality assessment:

*fastqc sample_R1.fastq sample_R2.fastq -o qc_reports/*

- Check for adapter contamination, GC content, and sequence duplication.

***2. Preprocessing: Trimming and Adapter Removal***

Trim the reads to remove adaptors and low-quality bases using *Trim Galore* or *cutadapt*:

*trim_galore --paired sample_R1.fastq sample_R2.fastq --output_dir cleaned_reads/*

**Genome Assembly**

***Chromosome-Level Assembly***

A chromosome-level assembly is important for reference genomes and extensive analysis of genomic features.

1. **Short-Read Assembly with Illumina (High Depth)**:

○ Use *SPAdes* for de novo genome assembly if no reference genome is available. For high-quality, large genomes, ensure a deep coverage (100X or more).

*spades.py -1 cleaned_reads/sample_R1.fastq -2 cleaned_reads/sample_R2.fastq -o spades_output/*

**Long-Read Assembly**:

- *Canu* and *Flye* are popular choices for long-read-based de novo assembly. Example with *Flye* for PacBio/ONT data:

*flye --nano-raw sample.fastq --out-dir flye_output/*

Once assembled, link scaffolds and order them using tools like *RA* or *NextFlow* with Hi-C data, if available, to generate a chromosome-level assembly.

**Hybrid Assembly (Illumina + PacBio or Nanopore)**:

- Use a hybrid assembly method combining both Illumina and long-read sequencing for greater assembly completeness. Tools like *Unicycler* can integrate long reads with short-read data for better resolution.

*unicycler -1 cleaned_reads/sample_R1.fastq -2 cleaned_reads/sample_R2.fastq --long sample_longreads.fastq -o unicycler_output/*

**Repeat Masking**: After genome assembly, repeat sequences often skew analysis and require masking. Use RepeatMasker:

*RepeatMasker -species fish_assemblies -pa 4 assembled_genome.fasta*

***Non-Chromosome-Level Assembly***

For genomes where chromosome-level assembly is not needed:

1. **De novo assembly** with tools like SPAdes or Trinity (for transcriptomes), particularly for draft genomes with fewer contigs.
2. **Mapping-based assembly**: Map the reads against a reference genome with tools like *BWA*, *Bowtie2*, or *Minimap2* (for PacBio/ONT long reads).

**Repeat Masking & Post-Assembly Processing**

Repeat Masking ensures that repetitive sequences are appropriately annotated and not considered in downstream analyses:

*RepeatMasker -species "Danio rerio" genome.fasta -xsmall*

After repeat masking, clean-up the assembly and evaluate assembly statistics, including N50, contig size, and other metrics that measure assembly completeness with tools like QUAST:

*quast.py assembly.fasta -o quast_output/*

**Genome Annotation**

1. **Gene Prediction (De Novo)**: Use *GeneMark* or *BRAKER* for genome-wide gene prediction.

*braker.pl --species=fish_species --genome=assembly.fasta --bam=aligned_reads.bam*

**Homology-Based Annotation**: BLAST the assembly against known gene databases like *NR* or *SwissProt* for homology-based functional annotation.

*blastn -query assembly.fasta -db nr -out blast_output.txt -outfmt 6*

**Functional Annotation**:

- Use tools like *InterProScan* to identify protein domains and annotations.

*interproscan.sh -i assembly.fasta -f tsv -o annotations.tsv*

**Gene Ontology (GO) Enrichment Analysis**: After obtaining a list of differentially expressed or annotated genes, you can perform GO term analysis.

*goseq()  # In R or Bioconductor*

**Visualization**

- Use *IGV* (Integrative Genomics Viewer) to explore alignment data and annotated genes visually.
- Create **circos plots** to represent genomic features, repeat distributions, and gene families across the assembled genome.

circos -conf circos_config.txt

**Suggested readings**

Masoudi-Nejad, A., Narimani, Z., & Hosseinkhan, N. (2013). *Next Generation Sequencing and Sequence Assembly: Methodologies and Algorithms* (Vol. 4). Springer. https://doi.org/10.1007/978-1-4614-7726-6

McGrath, A. (2007). Chapter 11 Genome Sequencing and Assembly. In K. R. Mitchelson (Ed.), *Perspectives in Bioanalysis* (Vol. 2, pp. 327–355). Elsevier. https://doi.org/10.1016/S1871-0069(06)02011-8

Ismail, H.D. (2023). Bioinformatics: A Practical Guide to Next Generation Sequencing Data Analysis (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003355205

# Principles of Transcriptome Analysis and Gene Expression analysis a practical approach

Vysakh V G.
Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018
vishakvg99@gmail.com.

## Introduction

Transcriptomics, the comprehensive study of RNA transcripts in a cell or organism, has become an indispensable tool for exploring gene regulation, cellular responses, and functional genomics. By capturing the dynamic landscape of gene expression, provides crucial insights into how organisms adapt to environmental changes, developmental processes, and stress conditions. Advances in RNA sequencing technologies have significantly expanded its accessibility and precision, enabling researchers to unravel molecular mechanisms across diverse scientific disciplines.

In fisheries science, transcriptomics plays a pivotal role in understanding stress resilience, reproduction, immunity, and metabolism in aquatic species. It aids in developing stress-tolerant aquaculture strains, optimising breeding programs, and mitigating disease outbreaks. Combining transcriptomics with other omics disciplines such as proteomics provides a systems-level understanding of biological processes, supporting sustainable fisheries management and aquaculture advancements. Its versatility ensures continued relevance in aquatic research, driving both fundamental discoveries and practical applications

## Methodology

### *1. Reference-Based Transcriptomics*

**Principle**: Reference-based transcriptomics aligns RNA-Seq reads to a reference genome, leveraging known gene annotations for transcript quantification and analysis. This approach assumes the completeness and accuracy of the reference genome.

**Steps**:

1. **RNA Extraction**: Extract total RNA using a commercial kit like TRIzol or RNeasy, ensuring high RNA integrity (RIN > 7).
2. **Library Preparation**: Use Illumina TruSeq or similar kits to create cDNA libraries.

3. **Sequencing**: Sequence libraries on Illumina or comparable platforms.

4. **Quality Control**: Removing adapter contaminations, low-quality base pairs etc using tools such as FastQC, fastp etc.

*fastp -i sample_R1.fastq -I sample_R2.fastq -o clean_R1.fastq -O clean_R2.fastq*

**5. Read Alignment**: align cleaned reads to a reference genome ort transcriptome

- Use HISAT2 or STAR for alignment:

*hisat2 -p 8 -x reference_genome -1 clean_R1.fastq -2 clean_R2.fastq -S output.sam*

**6. Transcript Assembly and Quantification**:

- Use StringTie for assembly:

*stringtie output.sam -G reference.gtf -o transcripts.gtf*

**7. Quantify transcripts with featureCounts:**

*featureCounts -a reference.gtf -o gene_counts.txt output.sam*

**8. Differential Expression Analysis**:

- Use DESeq2 in R:

*library(DESeq2)*

*countData <- read.table("gene_counts.txt", header=TRUE, row.names=1)*

*dds<-DESeqDataSetFromMatrix(countData=countData, colData=colData, design=~condition)*

*dds <- DESeq(dds)*

*results <- results(dds)*

***2. De Novo Transcriptomics***

**Principle**: De novo transcriptomics constructs transcriptomes without a reference genome, making it ideal for non-model organisms. This relies on de novo assemblers to generate transcripts from RNA-Seq reads.

**Steps**:

1. **RNA Extraction**: Same as reference-based.
2. **Library Preparation and Sequencing**: Same as reference-based.
3. **Quality Control**:

*fastp -i sample_R1.fastq -I sample_R2.fastq -o clean_R1.fastq -O clean_R2.fastq*

**4. De Novo Assembly**:

- Use Trinity for assembly:

*Trinity --seqType fq --max_memory 50G --left clean_R1.fastq --right clean_R2.fastq --CPU 16 --output trinity_out_dir*

**5. Transcriptome Evaluation**:

- Assess quality using BUSCO:

*busco -i trinity_out_dir/Trinity.fasta -l eukaryota_odb10 -o busco_output -m transcriptome*

**6. Filter lowly expressed transcripts:**

*seqtk seq -a Trinity.fasta > filtered_transcripts.fasta*

**7. Annotation**:

Perform suitable BLAST annotations:

- **BLASTn against NR database**:

*blastn -query Trinity.fasta -db nr -outfmt 6 -evalue 1e-5 -out blastn_output.txt*

- **BLASTx against NR database**:

*blastx -query Trinity.fasta -db nr -outfmt 6 -evalue 1e-5 -out blastx_output.txt*

- **BLASTp against UniProt/SwissProt database**:

*blastp -query filtered_transcripts.pep -db swissprot -outfmt 6 -evalue 1e-5 -out blastp_output.txt*

- **Use Trinotate for consolidated annotation:**

*Trinotate Trinotate.sqlite init --transcript_fasta Trinity.fasta --genpep filtered_transcripts.pep --gene_to_trans_map trinity_out_dir/Trinity.fasta.gene_trans_map*

- **Run InterProScan:**

*interproscan.sh -i Trinity.fasta -o interpro_output*

**8. Differential Expression**:

- Map reads back to transcripts with Salmon:

*salmon quant -i trinity_index -l A -1 clean_R1.fastq -2 clean_R2.fastq -p 8 -o quant_output*

Analyse with DESeq2: Same as reference-based.

***9. Integration with Pathway Analysis***

1. **Functional Annotation**:
    - Annotate transcripts using KEGG Mapper, EggNOG, or Reactome tools.
    - Example KEGG enrichment in R:

*library(clusterProfiler)*

*enrichKEGG(gene=gene_list, organism="hsa")*

**Visualisation**:

- Create heatmaps or PCA plots:

*library(pheatmap)*

*pheatmap(log2(countData + 1))*

**Suggested readings**

Azad, R. K. (Ed.). (2024). *Transcriptome Data Analysis* (Vol. 2812). Springer US. https://doi.org/10.1007/978-1-0716-3886-6

Wang, Y., & Sun, M. (Eds.). (2018). *Transcriptome Data Analysis* (Vol. 1751). Springer. https://doi.org/10.1007/978-1-4939-7710-9

# Integrating behavioral and physiological insights with omics in a zebrafish model

Saima Rehman, Adnan Hussain Gora, Jeena N. S., Qurat ul Ain, Sanal Ebeneezar, Chandrasekhar Selvam.

Marine biotechnology fish nutrition and health division., ICAR-CMFRI, Kochi-18,

## Introduction

The significant rise in global demand for seafood over recent decades has profoundly impacted both the fisheries capture industry and aquaculture production worldwide. This surge is driven by a growing human population, which has intensified the need for sustainable, high-quality protein sources in food supplies (Kidane et al., 2021). To meet this escalating demand, aquaculture systems have expanded, often leading to intensified farming practices. These practices can result in disease-related challenges due to factors such as inadequate biosecurity measures, high stocking densities, and compromised water quality, all of which can contribute to increased fish mortality (Action., 2020). Additionally, disease outbreaks may occur due to uncontrolled transmission from wild stocks, accidental transfers of infected stocks between farms, and management failures in identifying unhealthy fish stocks. If not properly managed, these issues can lead to mass mortality of cultured fish, potentially causing significant economic losses and threatening the stability of the aquaculture industry.

To address the biological challenges in aquaculture and meet the increasing global demand for seafood, advanced biotechnological tools, particularly omics technologies—such as genomics, transcriptomics, proteomics, and metabolomics—have been employed to enhance our understanding of aquaculture safety, production, quality, and health (Power et al., 2023). These omics technologies provide detailed information about organisms, aiding in the identification of molecular mechanisms and immune response pathways. In addition, integrating behavioural and physiological insights with omics approaches provides a comprehensive understanding of how external and internal factors influence an organism's overall health, development, and response to environmental stimuli (Sidira et al., 2024). Behavioural studies offer observable phenotypes such as locomotion, social interactions, stress responses, and learning, which can be quantitatively analysed using automated tracking systems. Physiological assessments, including stress biomarkers (e.g., cortisol), immune responses, organ function, and metabolic

rates, complement behavioural data by revealing underlying systemic changes (Egan et al.,2009).

Omics approaches like genomics, transcriptomics, proteomics, metabolomics, and microbiomics further enhance this integration by uncovering molecular mechanisms. Genomics and epigenomics help identify genetic and regulatory variations influencing behaviour and physiology, while transcriptomics reveals gene expression changes associated with stress or disease responses. Proteomics provides insights into protein pathways regulating neural or immune functions, and metabolomics identifies metabolic shifts linked to energy use or neurotransmitter activity (Sidira et al., 2024). Microbiomics adds another dimension by studying gut microbiota and its role in the gut-brain axis, impacting behaviour and immunity (Kim and Shim., 2022). A systems biology approach ties these datasets together, using computational tools and network analysis to correlate molecular changes with physiological and behavioural outcomes, allowing the identification of regulatory hubs that mediate responses to environmental stimuli. This multimodal integration is particularly valuable in fields like aquaculture, where it can improve stress tolerance, disease management, and probiotic efficacy by linking microbiota changes with host responses. It is also critical in neurobiology, environmental toxicology, and host-pathogen interaction studies, where understanding the interplay between molecular pathways and organismal behaviour is key. Despite challenges such as the complexity of data integration, ethical concerns, and standardization, advances in bioinformatics, machine learning, and non-invasive techniques are paving the way for more effective implementation of this approach. For example, studying stress in zebrafish could involve quantifying behaviour under heat stress, measuring cortisol levels, and analysing gene expression changes, neurotransmitter shifts, and microbiota composition, which are then integrated to map the full spectrum of responses. Such holistic insights offer profound implications for advancing basic biology, improving aquaculture practices, and addressing environmental and health challenges.

**Zebrafish for Gut-Brain Axis Studies**

Zebrafish (Danio rerio) has emerged as a powerful model for investigating the gut-brain axis, particularly in the context of intestinal inflammation and gut microbiota interactions. Its unique combination of a conserved immune system, well-characterized nervous system, and high genetic homology with humans makes it an ideal organism for studying how gut dysbiosis and inflammation influence brain function and behavior. Unlike traditional rodent models,

zebrafish offer advantages such as rapid development, external fertilization, and transparent larvae, allowing real-time visualization of gut-brain interactions at a cellular and molecular level. At around 4 days postfertilization (dpf), following the development of swim bladder, zebrafish larvae start to swim freely, and their swimming behavior is modulated by both internal and external stimuli (Kopp et al., 2018). With the use of high throughput tracking system, we can monitor the activity of an individual larva and split a particular movement into several measurable parameters like total distance travelled, total duration of movement, heading angle, turn angle, average velocity, angular velocity; all these parameters can be compared to understand the effect of different treatments on zebrafish behavior (Maeda et al., 2021). The light-dark locomotion test is adopted widely to study the behavioral response of zebrafish larvae to sudden changes in illumination. The alternating light and dark conditions prompt zebrafish to follow a specific pattern of movement: while a transition from a light to a dark regimen increases locomotion, a dark to light transition decreases its movement. This locomotor behavior depends on the coordination of eye and brain functions. Other tests like exploratory biting, T-maze test and diving response test are also used to understand zebrafish behavior (Fontana and Parker, 2022). Diet-induced changes can also affect the locomotor behavior in animals. For example, a high-fat, low-fiber diet reduced the locomotor activity of mice and a probiotic diet (with *Lactobacillus rhamnosus* IMC 501) increased the movement of zebrafish (). Furthermore, intestinal inflammation can also affect animal behavior as reported for a chemically-induced colitis mice model with stress associated behavior (Komoto et al., 2022).

**Key Advantages of Zebrafish in Gut-Brain Axis Research:**

**Conserved Gut-Immune System** – Zebrafish share key innate immune pathways with mammals, including Toll-like receptors and cytokine signaling, making them a relevant model for studying intestinal inflammation and immune responses.

**Microbiota Modulation** – The ability to generate germ-free (GF) zebrafish allows precise control over microbiome composition, enabling researchers to assess the impact of specific microbial communities on gut and brain function.

**Transparent Larvae & Live Imaging** – Zebrafish larvae remain transparent during early development, facilitating in vivo imaging of gut permeability, neuroinflammation, and microglial activation in response to microbial and inflammatory stimuli.

**Behavioral Assays** – Zebrafish exhibit well-characterized behaviors, including anxiety-like responses, social interaction, and cognitive performance, providing a measurable link between gut microbial composition and brain function.

**Multi-Omics Compatibility** – Zebrafish studies integrate advanced omics technologies, including transcriptomics, metabolomics, and microbiome sequencing, allowing for comprehensive mechanistic insights into gut-brain interactions at the molecular level.

By leveraging these advantages, zebrafish provide a versatile and cost-effective platform for investigating the mechanisms by which gut inflammation and microbiota alterations contribute to neurological disorders. This model is paving the way for novel therapeutic strategies targeting the gut-brain axis in both aquaculture and human health research.

## Key Research Applications

### 1. Intestinal Inflammation and Brain Function

Zebrafish models provide a powerful platform for studying the intricate relationship between intestinal inflammation and brain function. Experimental induction of gut inflammation using agents such as dextran sulfate sodium (DSS), 2,4,6-trinitrobenzenesulfonic acid (TNBS), or high-fat diets leads to significant alterations in intestinal homeostasis, triggering immune responses that extend beyond the gut. These models have demonstrated that inflammatory cytokines such as tumor necrosis factor-alpha (TNF-α) and interleukin-6 (IL-6) are upregulated during gut inflammation, which in turn can influence neuroinflammatory pathways. Transcriptomic profiling of both the gut and brain tissues in zebrafish reveals that the activation of pro-inflammatory markers in the intestine correlates with changes in brain gene expression, particularly in genes involved in neuroinflammation, neurotransmission, and blood-brain barrier integrity. This systemic inflammatory response highlights the gut-brain axis as a critical pathway through which gut disturbances can impact neural function, potentially leading to altered behavior, cognitive dysfunction, and increased susceptibility to neurological disorders. By leveraging zebrafish as a model, researchers can further explore the molecular and cellular mechanisms underlying this crosstalk, paving the way for novel therapeutic strategies targeting both gut and brain health (Collins et al., 2020).

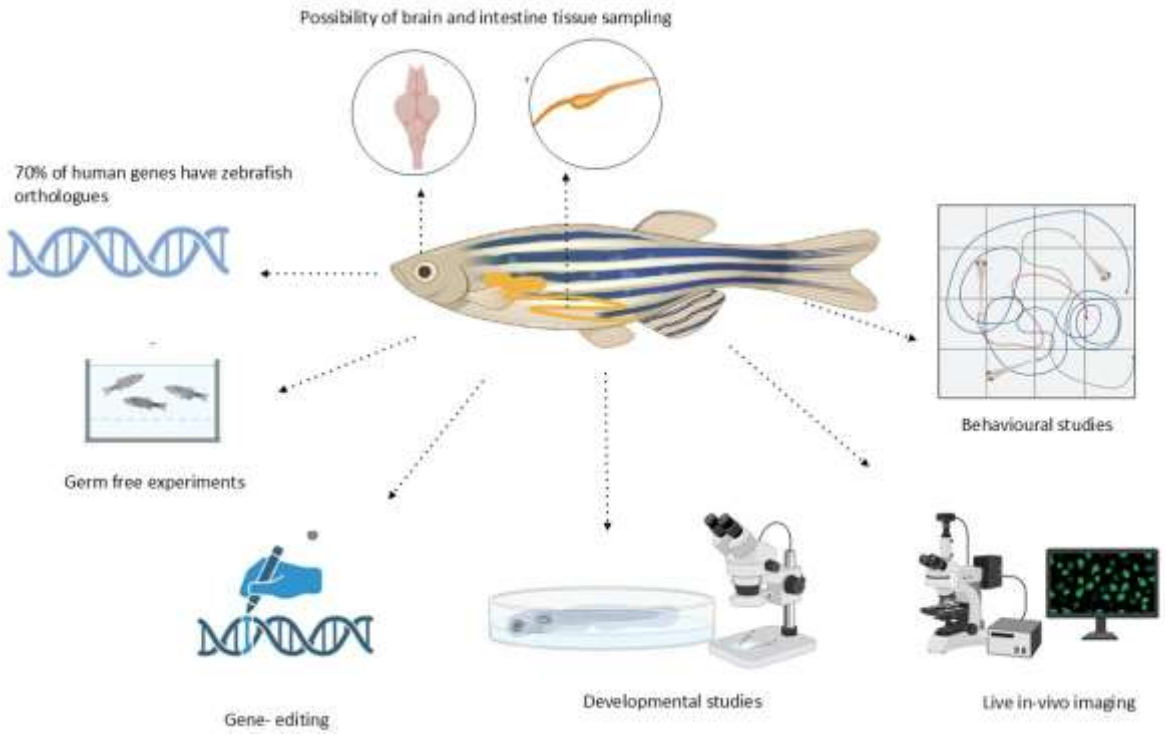



**Figure 1**: Zebrafish as model for Gut-Brain Axis Studies

## 2. Microbiota Dysbiosis and Behavioral Changes

Disruptions in gut microbiota composition, whether induced by antibiotic treatment or dietary modifications, have been shown to significantly alter zebrafish behavior, highlighting the crucial role of microbial communities in regulating brain function. Studies using germ-free (GF) zebrafish—raised in sterile conditions without microbial exposure—reveal profound behavioral abnormalities, including impaired locomotion, reduced social interaction, and cognitive deficits. These observations suggest that the presence of a healthy and diverse gut microbiome is essential for normal neural development and behavioral regulation. Metabolomic analyses further support this link by identifying key microbial-derived neuroactive compounds, such as short-chain fatty acids (SCFAs) and tryptophan metabolites, which are known to influence neurotransmitter synthesis, neuroinflammation, and brain plasticity. The absence or imbalance of these metabolites in dysbiotic conditions may lead to disruptions in neurochemical signalling pathways, ultimately manifesting as behavioral and cognitive impairments. By utilizing zebrafish as a model to investigate microbiota-brain

interactions, researchers can better understand how gut microbial dysbiosis contributes to neurodevelopmental disorders, mental health conditions, and neurological diseases, paving the way for potential microbiome-targeted therapeutic interventions (Xu et al., 2024) .

**3. Neuroinflammation and Neurodegeneration**

Zebrafish models of neuroinflammation provide compelling evidence that gut dysbiosis can serve as a precursor to neurodegenerative changes, establishing a strong link between gut health and brain pathology. Studies have shown that disruptions in the gut microbiota composition occur before the appearance of key neurodegenerative markers, such as α-synuclein aggregation and amyloid-beta accumulation, which are hallmarks of Parkinson's and Alzheimer's disease, respectively. This suggests that microbial imbalances may play a causative role in neurodegeneration by triggering chronic inflammation and altering immune signaling pathways. Additionally, early-life gut inflammation has been found to significantly impact neuronal plasticity and neurotransmitter levels, potentially leading to long-term cognitive deficits and behavioral abnormalities. Changes in synaptic function, reductions in key neurotransmitters like serotonin and dopamine, and increased oxidative stress further highlight the profound effects of gut inflammation on brain health. By leveraging zebrafish as a model system, researchers can dissect the molecular mechanisms underlying this gut-brain interaction, offering valuable insights into the development of neurodegenerative diseases and the potential for microbiome-based therapeutic interventions (Zhang et al., 2023).

**4. Role of Probiotics and Prebiotics in Modulating Gut-Brain Axis**

Probiotic and prebiotic interventions have emerged as promising strategies for restoring gut microbiota balance and mitigating the effects of gut dysbiosis on brain function. In zebrafish models, supplementation with beneficial probiotic strains such as Lactobacillus and Bifidobacterium has been shown to improve gut health by enhancing microbial diversity, strengthening gut barrier integrity, and modulating immune responses. These changes are accompanied by significant reductions in anxiety-like behaviors, suggesting a direct link between gut microbiota composition and neurobehavioral regulation. Similarly, prebiotic interventions, which involve the administration of non-digestible dietary fibers that promote the growth of beneficial microbes, have demonstrated positive effects on both intestinal and neurological health. Prebiotics help reinforce gut epithelial integrity, reducing intestinal permeability and preventing the leakage of pro-inflammatory cytokines into systemic circulation. As a result, neuroinflammatory markers such as TNF-α and IL-6 are

downregulated, leading to improved neurotransmitter balance and enhanced cognitive function. These findings highlight the potential of probiotic and prebiotic therapies in modulating the gut-brain axis, paving the way for microbiome-based interventions in managing neuroinflammation, anxiety disorders, and neurodegenerative diseases (Ansari et al., 2023).

**Conclusion:**

Integrating behavioral and physiological insights with omics approaches in zebrafish research offers a powerful framework for understanding the complex interplay between environmental factors, host responses, and disease mechanisms. The ability to combine behavioral assays, stress biomarkers, and multi-omics data provides a comprehensive perspective on how gut microbiota, neuroinflammation, and systemic immune responses influence fish health and neurobiology. Zebrafish, with their genetic similarity to humans, transparent larvae, and amenability to high-throughput analysis, serve as an ideal model for studying the gut-brain axis, microbiota dysbiosis, and neuroimmune interactions. This research has broad implications, not only for aquaculture disease management but also for advancing human health by identifying novel therapeutic targets for neurodegenerative diseases and mental health disorders. Despite challenges such as data integration complexity and ethical considerations, emerging bioinformatics tools and machine learning techniques continue to enhance our ability to interpret these multi-layered datasets. Moving forward, leveraging zebrafish models to study the effects of probiotics, prebiotics, and dietary interventions will be crucial in developing sustainable strategies to improve both aquatic and human health.

**References**

Action, S. I. (2020). World fisheries and aquaculture. *Food and Agriculture Organization*, *2020*, 1-244.

Ansari, F., Neshat, M., Pourjafar, H., Jafari, S. M., Samakkhah, S. A., & Mirzakhani, E. (2023). The role of probiotics and prebiotics in modulating of the gut-brain axis. *Frontiers in nutrition*, *10*, 1173660.

Borrelli L, Aceto S, Agnisola C, De Paolo S, Dipineto L, Stilling RM, et al. (2016) Probiotic modulation of the microbiota-gut-brain axis and behaviour in zebrafish. Sci Rep 6: 1-9.

Collins, Stephen M. "Interrogating the gut-brain axis in the context of inflammatory bowel disease: a translational approach." *Inflammatory bowel diseases* 26, no. 4 (2020): 493-501.

Egan, Rupert J., Carisa L. Bergner, Peter C. Hart, Jonathan M. Cachat, Peter R. Canavello, Marco F. Elegante, Salem I. Elkhayat et al. "Understanding behavioral and physiological phenotypes of stress and anxiety in zebrafish." *Behavioural brain research* 205, no. 1 (2009): 38-44.

Fontana BD, Parker MO. (2022) The larval diving response (LDR): validation of an automated, highthroughput, ecologically relevant measure of anxiety-related behavior in larval zebrafish (*Danio rerio*). J Neurosci Methods 381: 109706.

Kidane, D. G., & Brækkan, E. H. (2021). Global seafood demand growth differences across regions, income levels, and time. *Marine resource economics*, *36*(3), 289-305.

Kim, G. H., & Shim, J. O. (2022). Gut microbiota affects brain development and behavior. *Clinical and Experimental Pediatrics*, *66*(7), 274.

Komoto M, Asada A, Ohshima Y, Miyanaga K, Morimoto H, Yasukawa T, et al. (2022) Dextran sulfate sodium-induced colitis in C57BL/6J mice increases their susceptibility to chronic unpredictable mild stress that induces depressive-like behavior. Life Sci 289: 120217.

Kopp R, Legler J, Legradi J. (2018) Alterations in locomotor activity of feeding zebrafish larvae as a consequence of exposure to different environmental factors. Environ Sci Pollut Res 25: 4085-4093.

Maeda H, Fukushima N, Hasumi A. (2021) Standardized method for the assessment of behavioral responses of zebrafish larvae. Biomedicines 9: 884.

Power, D. M., Taoukis, P., Houhoula, D., Tsironi, T., & Flemetakis, E. (2023). Integrating omics technologies for improved quality and safety of seafood products. *Aquaculture and Fisheries*, *8*(4), 457-462.

Sidira, M., Agriopoulou, S., Smaoui, S., & Varzakas, T. (2024). Omics-Integrated Approach (Metabolomics, Proteomics and Lipidomics) to Assess the Quality Control of Aquatic and Seafood Products. *Applied Sciences*, *14*(22), 10755.

Xu, Q., Sun, L., Chen, Q., Jiao, C., Wang, Y., Li, H., ... & Chen, X. (2024). Gut microbiota dysbiosis contributes to depression-like behaviors via hippocampal NLRP3-mediated neuroinflammation in a postpartum depression mouse model. *Brain, Behavior, and Immunity*, *119*, 220-235.

Zhang, W., Xiao, D., Mao, Q., & Xia, H. (2023). Role of neuroinflammation in neurodegeneration development. *Signal transduction and targeted therapy*, *8*(1), 267.

# Genome Editing Using CRISPR Cas 9

Vysakh V G, Ramitha P.A, Sandhya Sukumaran.
Marine Biotechnology, Fish Nutrition and Health Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, India – 682018
vishakvg99@gmail.com.

## Introduction

CRISPR Cas9 is an emerging technique which revolutionize the molecular biology to a great extent. In earlier times, genome editing was done using Zinc Finger Nucleases( ZFNs) and Transcription Activator-Like Endonucleases (TALENs). It identifies specific nucleotide using nuclease dimerisation, with each nuclease usually fok1 spatially aligned at the target site monomer and allows for the DNA cleavage. The repair of DNA strand breaks is frequently associated with errors arising from nonhomologous end joining. The widespread application of ZFNs and TALENs has declined because of design complexity, cost, multiplexing difficulty, delivery, and off-target concerns. An adaptive system found in prokaryotes that can now be used as an advanced genome editing technique, Clustered Regularly Interspaced short palindromic Repeats - CRISPR associated system 9 (CRISPR - Cas9 ). It is mainly found in bacteria; whenever they encounter any alteration, they individually correct it by a particular mechanism.

## Structure

Three sequences make up the majority of the CRISPR cas structure: operons of CRISPR-associated genes that encode Cas proteins, a CRISPR array of repeat sequences and spacers, and the leader sequences, which are frequently AT-rich. When a phage infects bacteria, it keeps some DNA remnants, which act as memory for the bacteria to defend against a second attack. It will transcribe to CRISPR RNAs(crRNAs) that direct the binding and cleavage of nucleic acid targets in the host. The repeats help in the proper processing of the long precursor CRISPR RNA (pre-crRNA) into smaller crRNAs. They are palindromic, meaning they can form hairpin-loop structures in the RNA transcript, aiding in stability and processing. Cas proteins recognise repeats during CRISPR RNA maturation. These repeats were present in bacteria from a

previous infection; however, they are designed as guide RNA (gRNA) to carry out experiments, which help with cleavage and correction of sequences.

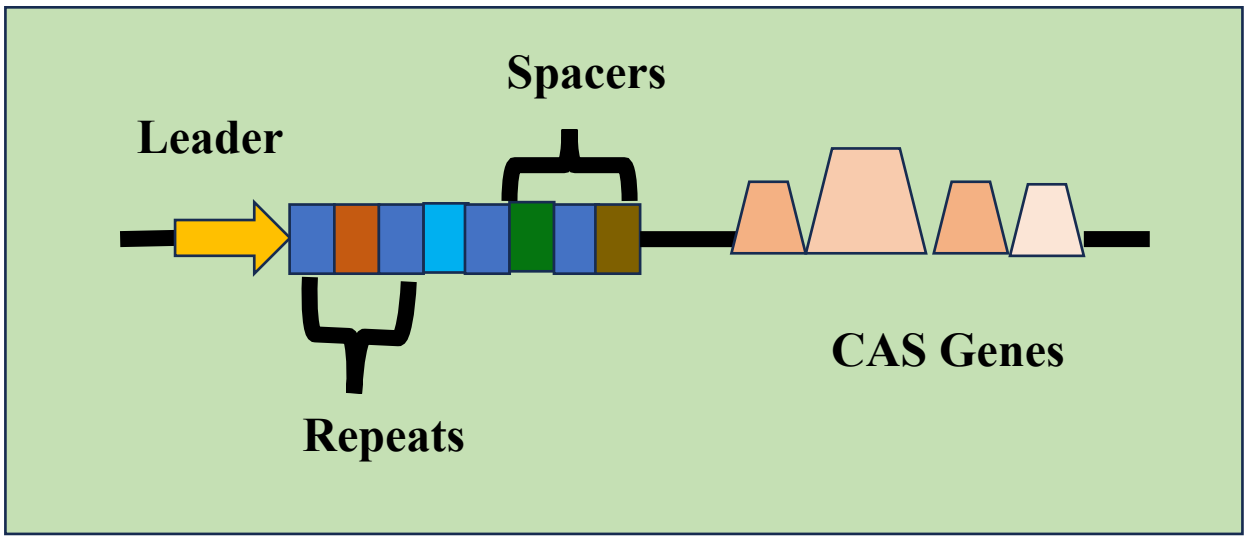



**Figure 1:** Graphical representation CRISPR Cas structure

**Classification**

CRISPR/Cas systems are broadly classified into two major classes according to the composition of their effector complexes, with Class 1 employing multi-subunit protein complexes and Class 2 relying on single, large effector proteins. Within Class 1, the most prominent is Type I, which comprises subtypes A through G and is characterised by the signature nuclease Cas3. These systems function without a tracrRNA and specifically target DNA substrates through the coordinated action of multiple protein subunits. Another important group within this class is Type III, represented by subtypes A to F, which also utilises multi-subunit effectors without a tracrRNA, but instead are defined by Cas10 and exhibit the unique ability to target both DNA and RNA molecules. Type IV, consisting of subtypes A to C, remains less well characterised; these systems use multi-protein complexes whose signature proteins and precise nucleic acid targets are still largely unknown. By contrast, Class 2 CRISPR systems are structurally simpler, as their effector activity is mediated by single proteins. Type II systems, comprising subtypes A to C, are represented by the signature protein Cas9, which requires a tracrRNA for its activity and specifically cleaves DNA. Cas9 has become the most widely utilised tool in genome editing because of its programmable DNA cleavage mechanism. Type V systems are more diverse, with subtypes ranging from A to U, and are defined by the effector Cas12. These nucleases also require tracrRNA in most subtypes, which are

predominantly DNA-targeting, although certain variants, such as subtype V-G, are capable of RNA targeting. Cas12 nucleases are also distinguished by their collateral cleavage activity, which has been harnessed for diagnostic applications. Type VI systems, consisting of subtypes A to D, are defined by Cas13 effectors that do not depend on tracrRNA and exclusively target RNA substrates. These properties have made Cas13 especially attractive for RNA editing and antiviral strategies. The details of the CRISPR protein are given in table 1

| **class** | Type | Number of subtypes | Effector molecules | TracrRNA | Signature protein | Target Substrate |
|---|---|---|---|---|---|---|
| 1 | I | 7 | Multiple Subunits | no | Cas3 | DNA |
| | III | 6 | Multiple Subunits | no | Cas10 | DNA/RNA |
| | IV | 3 | Multiple Subunits | no | unknown | Unknown |
| 2 | II | 3 | Single unit | Yes | Cas9 | DNA |
| | V | 11 | Single unit | Yes | Cas12 | DNA/RNA |
| | VI | 4 | Single unit | No | Cas13 | RNA |

**Table 1**: Different types of CRISPR-Cas Systems

**Mechanism of CRISPR-Cas9-Mediated Genome Editing**

Genome editing using the CRISPR-Cas9 system involves a series of coordinated molecular events which enable precise recognition and modification of target DNA sequences. The system has been adapted from the adaptive immune mechanism of bacteria and archaea. In bacteria, it functions as a defence against invading bacteriophages and plasmids. In genome engineering, the naturally occurring CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA) are fused into a single guide RNA (sgRNA), simplifying the editing process while retaining target specificity. The genome editing process begins with the formation of a ribonucleoprotein complex between the guide RNA and the Cas9 nuclease. The guide RNA contains a 20-nucleotide sequence that is complementary to the target DNA. Before binding, Cas9 first recognises a short protospacer adjacent motif (PAM), typically the sequence 5′-NGG-3′ for *Streptococcus pyogenes* Cas9. PAM recognition is an essential prerequisite because

Cas9 cannot bind or cleave DNA in its absence. Following PAM recognition, the guide RNA hybridises with the complementary DNA strand. This results in the formation of an RNA-DNA heteroduplex. This interaction activates the nuclease domains of Cas9. The HNH domain cleaves DNA strand complementary to the guide RNA, and the RuvC domain cleaves the opposite strand. subsequently generating a site-specific double-stranded DNA break nearly three nucleotides upstream of the PAM sequence. Once cleavage occurs, the endogenous DNA repair machinery of the cell repairs the break, ultimately determining the outcome of genome editing.

**DNA Repair Pathways Following CRISPR-Cas9 Cleavage**

The double-stranded DNA break generated by Cas9 is repaired through two endogenous cellular pathways, namely, non-homologous end joining (NHEJ) and homology-directed repair (HDR). The choice of repair mechanism determines the nature of the genomic modification. Non-homologous end joining is the predominant repair pathway in most cells. NHEJ ligates directly broken ends of DNA without the need for a homologous template. This mechanism is efficient, but inherently it is error-prone. It frequently introduces small indels (insertions or deletions) at the cleavage site. These mutations disrupt the reading frame of genes and lead to loss of gene function. NHEJ is popularly used for gene knockout experiments. Homology-directed repair is a template-dependent mechanism which uses a homologous DNA sequence as a repair template. When an exogenous donor DNA containing the desired genetic modification is supplied, HDR enables precise insertion, replacement, or correction of specific DNA sequences. HDR provides highly accurate genome editing; its efficiency is generally lower than that of NHEJ because it is primarily active during the S and G2 phases of the cell cycle. Consequently, improving HDR efficiency remains an active area of CRISPR research.

**Recent Advancements in CRISPR Technologies**

Beyond conventional genome editing, the continuous improvement and evolution in CRISPR technology have expanded its capabilities. To improve the precision of editing, broaden the target range, and minimise unintended genomic alterations, several engineered CRISPR systems have been developed.

Base editing is one of the major advancements. Base editing allows direct conversion of one nucleotide to another without generating double-stranded DNA breaks. Cytosine base editors (CBEs) convert C•G base pairs to T•A, and adenine base editors (ABEs) convert A•T to G•C. Because base editing avoids double-strand breaks, it reduces insertion-deletion mutations and thereby improves precision of editing.

Another important innovation is prime editing. It is a versatile genome editing approach that combines Cas9 nickase with a reverse transcriptase enzyme and a prime editing guide RNA (pegRNA). Prime editing enables targeted insertions, deletions, and all possible base substitutions without the need for donor DNA templates or inducing double-stranded DNA breaks. Due to its high precision, prime editing is considered one of the most promising next-generation genome editing technologies.

Catalytically inactive Cas9, known as dead Cas9 (dCas9), lacks DNA cleavage activity but retains its ability to recognise specific DNA sequences. Without altering the underlying DNA sequence, we can regulate gene expression by fusing dCas9 with transcriptional activators, repressors or epigenetic modifiers. These systems, known as CRISPR activation (CRISPRa) and CRISPR interference (CRISPRi), have become valuable tools in functional genomics.

Several alternative nucleases also gained importance. For example, Cas12 recognises different PAM sequences and creates a staggered DNA breaks, making it useful for genome editing and nucleic acid diagnostics. Cas13 targets RNA molecules, enabling programmable RNA editing and antiviral applications. These expanding CRISPR platforms have increased the versatility of genome engineering across biological research and biotechnology.

**CRISPR-Cas Genome Editing in Fisheries and Aquaculture**

The simplicity, efficiency, and programmability of CRISPR-Cas systems have transformed genome engineering in fisheries and aquaculture. To date, CRISPR-Cas9 genome editing has been successfully applied in more than 20 aquaculture species, targeting commercially significant traits such as growth, disease resistance, reproduction, sterility, and pigmentation. Although genome editing has been demonstrated in several commercially important species, including rohu (*Labeo rohita*) and grass carp (*Ctenopharyngodon idella*), most studies remain concentrated in model organisms and a limited number of high-value cultured fishes.

**Growth Improvement**

Growth enhancement is one of the primary objectives of genome editing in aquaculture. The most extensively targeted gene is myostatin (mstn). Which is a negative regulator of skeletal muscle development. CRISPR-Cas9-mediated knockout of the *mstn* gene successfully achieved in red sea bream, common carp, channel catfish, olive flounder, blunt snout bream, Nile tilapia, and mud loach (Roy et al. 2022; Zhu et al. 2024; Mokrani et al. 2023). In channel catfish, disruption of *mstn* resulted in around a 29.7% increase in mean body weight compared with unedited controls (Zhu et al. 2024), demonstrating the considerable potential of genome editing for improving aquaculture productivity.

**Reproductive Manipulation and Sterility**

Genome editing is also used to regulate reproduction and induce sterility in cultured fish. Genes associated with germ cell development and sex determination, dead end (dnd) and doublesex and mab-3 related transcription factor (dmrt), edited in species such as Nile tilapia and Atlantic salmon (Roy et al. 2022; Okoli et al. 2021; Mokrani et al. 2023). Sterility induction is a very important strategy for preventing genetic introgression in between farmed and wild populations and improving reproductive management in aquaculture.

**Disease Resistance**

Improving disease resistance represents another major application of CRISPR technology. Genome editing has been used to modify immune-related genes in grass carp, rohu (*Labeo rohita*), and channel catfish to enhance resistance against bacterial and viral pathogens (Okoli et al. 2021). Editing of immune regulatory genes, for instance TLR5, MHC II, and IFN-γ in Nile tilapia achieved editing efficiencies of up to 87.5% and increased relative percent survival to nearly 88% following pathogen challenge (Azizova et al. 2025). CRISPR based approaches have also been studied for combating bacterial diseases in channel catfish, white spot syndrome virus (WSSV) in shrimp, and nervous necrosis virus (NNV) in marine finfish (Ferdous et al. 2022).

**Pigmentation and Functional Genomics**

Genes involved in pigmentation, for example tyrosinase (tyr) and oca2, have been successfully disrupted in cichlids, goldfish, and several other fish species (Bayır et al. 2025; Mokrani et al. 2023; Li et al. 2021). Besides generating commercially attractive colour variants, these studies have contributed significantly to functional genomics by elucidating the molecular mechanisms driving pigmentation and developmental processes.

**Applications in Crustaceans**

Even though Most genome-editing studies have focused on fish, CRISPR-Cas systems have also been successfully applied in crustaceans. In the oriental river prawn (*Macrobrachium nipponense*), CRISPR/Cas9-mediated editing of the vitellogenin and eyeless genes produced mutation frequencies of approximately 66.7% in hatched individuals, showing the feasibility of genome editing for functional genetic studies in crustacean aquaculture (Qiao et al. 2023).

| Species | Target gene | Trait improved | Outcome |
|---|---|---|---|
| Channel catfish | *mstn* | Growth | 29.7% increase in body weight |
| Nile tilapia | *TLR5*, *MHC II*, *IFN-γ* | Disease resistance | Up to 88% survival |
| Atlantic salmon | *dnd* | Sterility | Germ cell ablation |
| Rohu | Immune-associated genes | Disease resistance | Enhanced immune traits |
| Red sea bream | *mstn* | Growth | Increased muscle development |

**References**

Azizova, F. (2025). Impact of ocean circulation changes on phytoplankton populations in the context of climate change. *International Journal of Aquatic Research and Environmental Studies*. https://doi.org/10.70102/ijares/v5i2/5-2-05

Bayır, A., Saoula, S., Almansour, A., Wang, D., Bayır, M., Uzun, B. N., Turhan, S., Arslan, H., & Arslan, G. (2025). CRISPR-Based Genetic Improvement of Aquaculture Species. *Fishes*. https://doi.org/10.3390/fishes10020084

Diwan, A., Ninawe, A., & Harke, S. (2017). Gene editing (CRISPR-Cas) technology and fisheries sector. 65-72. https://doi.org/10.24870/cjb.2017-000108

Ferdous, M. A., Islam, S. I., Habib, N., Almehmadi, M., Allahyani, M., Alsaiari, A., & Shafie, A. (2022). CRISPR-Cas Genome Editing Technique for Fish Disease Management: Current Study and Future Perspective. *Microorganisms, 10*. https://doi.org/10.3390/microorganisms10102012

Kho, K., Sukhan, Z., Cho, Y., Cho, D., & Choi, C. (2025). Genome-Edited Fish in the Field. *Current Issues in Molecular Biology, 47*. https://doi.org/10.3390/cimb47121013

Li, C.-Y., Steighner, J. R., Sweatt, G., Thiele, T., & Juntti, S. A. (2021). Manipulation of the Tyrosinase gene permits improved CRISPR/Cas editing and neural imaging in cichlid fish. *Scientific Reports, 11*. https://doi.org/10.1038/s41598-021-94577-8

Li, C., Chu, W., Gill, R., Sang, S., Shi, Y., Hu, X., Yang, Y., Zaman, Q., & Zhang, B. (2022). Computational Tools and Resources for CRISPR/Cas Genome Editing. *Genomics, Proteomics & Bioinformatics, 21*, 108 - 126. https://doi.org/10.1016/j.gpb.2022.02.006

Lin, N., Zhang, J., Wang, Y., & Zhang, Z. (2024). CRISPR/Cas Technology in Disease Prevention and Control of Aquatic Animals. *Reviews in Aquaculture*. https://doi.org/10.1111/raq.12995

Liu, K., Petree, C., Requena, T., Varshney, P., & Varshney, G. K. (2019). Expanding the CRISPR Toolbox in Zebrafish for Studying Development and Disease. *Frontiers in Cell and Developmental Biology, 7*. https://doi.org/10.3389/fcell.2019.00013

Marnis, H., & Syahputra, K. (2025). Advancing Fish Disease Research through CRISPR-Cas Genome Editing: Recent Developments and Future Perspectives.. *Fish & shellfish immunology*, 110220 . https://doi.org/10.1016/j.fsi.2025.110220

Mokrani, A., & Liu, S. (2023). Harnessing CRISPR/Cas9 system to improve economic traits in aquaculture species. *Aquaculture*. https://doi.org/10.1016/j.aquaculture.2023.740279

Moorthi, H. M., P, A. A. J., & Naduvathu, P. P. (2023). CRISPR Genome Editing Technology to Revolutionize Aquaculture and Fisheries. *Acta Scientific Veterinary Sciences*. https://doi.org/10.31080/asvs.2023.05.0604

Okoli, A. S., Blix, T., Myhr, A., Xu, W., & Xu, X. (2021). Sustainable use of CRISPR/Cas in fish aquaculture: the biosafety perspective. *Transgenic Research, 31*, 1 - 21. https://doi.org/10.1007/s11248-021-00274-7

Puthumana, J., Chandrababu, A., Sarasan, M., Joseph, V., & Singh, I. S. B. (2024). Genetic improvement in edible fish: status, constraints, and prospects on CRISPR-based genome engineering. *3 Biotech, 14*. https://doi.org/10.1007/s13205-023-03891-7

Qiao, H., Jiang, S., Fu, H., Xiong, Y., Zhang, W., Xu, L., Cheng, D., & Wang, J. (2023). CRISPR/Cas9 establishment-mediated targeted mutagenesis in Macrobrachium nipponense. *Frontiers in Physiology, 14*. https://doi.org/10.3389/fphys.2023.1141359

Roy, S., Kumar, V., Behera, B., Parhi, J., Mohapatra, S., Chakraborty, T., & Das, B. (2022). CRISPR/Cas Genome Editing—Can It Become a Game Changer in Future Fisheries Sector?. https://doi.org/10.3389/fmars.2022.924475

Sun, Y., & Zhu, Z. (2019). Designing future farmed fishes using genome editing. *Science China Life Sciences, 62*, 420 - 422. https://doi.org/10.1007/s11427-018-9467-x

Yang, Z., Yu, Y., Tay, Y. X., & Yue, G. (2021). Genome editing and its applications in genetic improvement in aquaculture. *Reviews in Aquaculture*. https://doi.org/10.1111/raq.12591

Zhang, Z., Wang, J., Li, J., Liu, X.-H., Liu, L., Zhao, C., Tao, W., Wang, D., & Wei, J. (2023). Establishment of an Integrated CRISPR/Cas9 Plasmid System for Simple and Efficient Genome Editing in Medaka In Vitro and In Vivo. *Biology, 12*. https://doi.org/10.3390/biology12020336

Zhu, M., Sumana, S., Abdullateef, M. M., Falayi, O. C., Shui, Y., Zhang, C., Zhu, J., & Su, S. (2024). CRISPR/Cas9 Technology for Enhancing Desirable Traits of Fish Species in Aquaculture. *International Journal of Molecular Sciences, 25*. https://doi.org/10.3390/ijms25179299

# Metagenomics from aquaculture context

Sumithra TG*, Krupesha Sharma SR, Gayathri Suresh

Marine biotechnology fish nutrition and health division., ICAR-CMFRI, Kochi-18,, *sumithravet@gmail.com

**Introduction:**

Microorganisms are the most metabolically adaptable creatures on the planet and are essential to all aspects of life and biosphere processes. Almost all of the existing knowledge in microbiology is based on the culture-based methods. The disparity between bacterial populations determined by microscopy and culture-based methods, however, suggested that cultured microbes made up a small percentage of the real microbial world. This knowledge led to the realization that microbiologists had only explored the very tip of the iceberg. Deciphering the non-culturable microorganisms will undoubtedly increase and promote scientists' capacity to identify and exploit microbial capabilities. The various methods to study non-cultivable microbes include metagenomics, metatranscriptomics, and whole metabolite analysis (metabolomics). Gene expressions and microbial community functionality are applied in the meta-transcriptomics. A complete list of the chemicals present in a particular habitat is provided by metabolomics, which allows the correlation of chemicals with microbial diversity.

**Metagenomics, a culture-independent microbiological methodology**:

Metagenomics is the direct examination of microbial genomes found in particular habitat. Metagenome, microbial environmental genome, microbial community genome or microbial population genome, is the total genome of the microbiota present in a specific habitat. As a result, metagenomics can investigate both unculturable and culturable microbiota. Metagenomics uses functional gene screening or sequencing analysis to investigate the microorganisms in a specific habitat. Microbial diversity, their cooperative interactions, relationship with the environment interactions, and functional activity are the primary topics of metagenomics research. Investigation on metagenomics is advancing very fast in several disciplines, including applied microbiology, environmental protection, veterinary and human medicine, aquaculture, and agriculture (Zhang et al. 2021).

DNA isolation from all the microbes present in a particular habitat is the first step in all metagenomics research. To create a library that contains the genomes of every microbe present in that habitat, the isolated DNA is either immediately analyzed by sequence analysis or cloned

and transformed into readily maintained host bacteria. The cloned DNA is then sequenced or the function of the cloned genes is ascertained in order to study the library. On the other hand, the metagenomics library contains millions of clones, each of which has a different DNA fragment. One of the biggest tasks for metagenomics-based research is organizing these clones. But the advent of affordable, high-throughput sequencing techniques and the sophisticated computational skills needed to interpret the millions of random sequences found in the libraries will aid future metagenomics.

**Fundamentals of metagenomics:**

There are two approaches for metagenomics namely, sequence-based and functional metagenomics.

**Sequence-based metagenomics**

Depending on the data type sequence-based metagenomics can be classified as amplicon or targeted gene and shotgun metagenomics. The amplified marker genes, *viz.* 16S rRNA, 18S rRNA, 26S rRNA, *recA,* and intergenic transcribed spacer (ITS) make the data of amplicon metagenomics. Whereas, shotgun metagenomics data includes random DNA sequences from the samples. Breitwieser et al. (2019) referred to "metataxonomics" for amplicon metagenomics and "metagenomics" for shotgun metagenomics. Sequence-based metagenomics has reconstructed the genomes of uncultured organisms and shown new connections between phylogeny and function (Venter et al., 2004). The *16S rRNA*-based metagenomics is the most common method for metagenomics. As it is too expensive to target the complete 1500+ bp some of its nine variable zones are commonly targeted. But, there has been no agreement on which of the variable zones should be sequenced. According to Yang et al. (2016), V2 and V8 are the least dependable zones, whereas V4 and V6 will be the most dependable. The limited numbers of reliable phylogenetic indicators restrict the application of phylogenetic indicators in amplicon metagenomics. This implies that the origin of DNA remains uncertain if the sequenced DNA segment lacks a phylogenetic indicator. Nonetheless, the strength of this strategy will be boosted by the anticipated advancements in phylogenetic markers throughout the next decades.

**Functional metagenomics**

In this approach, the clones are checked for particular feature expression. Success in functional metagenomics depends on the accurate transcription, translation, and secretion of the product. This approach has discovered several new antimicrobial compounds, antimicrobial resistance genes, and enzymes. This approach can discover completely novel gene classes for unknown

or well-known functions. The successful expression of a large diverse DNA has been recorded in *E. coli.* Nevertheless, several genes will not be sufficiently expressed in *E. coli*. The inadequacy of satisfactory clones highlights the necessity for efficient screening tests to enhance the identification. Novel antibiotic-resistant determinants were isolated through a selection method based on antibiotic resistance. Another tactic is the use of particular media that can be used to identify or select active clones. Pre-selecting environmental samples to enrich libraries with target genes (ecological enhancement or targeted metagenomics), is becoming more and more common in metagenomics (Maester *et al.* 2016). In this methodology, microbial communities are modified before DNA isolation to boost the *in-situ* occurrence of target functions. The approach was applied to detect novel chitinases (Stöveken *et al.* 2015), hemicellulases (Zhao *et al.* 2017), oxidoreductases (Nagayama *et al.* 2015) and lipases (Petrovskaya *et al.* 2016). Another emerging trend is the application of functional anchors which are the functions that can be assessed quickly. This is a hybrid technique, in which the clones sharing a common function are gathered together and sequenced to determine phylogenetic indicators in the flanking DNA.

**Road map for metagenomics study:**

**Sample preparation, processing and DNA extraction:** This is the initial and most important phase as the extracted DNA should represent all microbes of that sample and should contain adequate high-quality DNA. Since the samples are gathered from the field, we should take precautions to avoid cross-contamination. Due to the highly sensitive nature, the amplicon-based metagenomics is highly susceptible to errors through unsuitable sample preservation procedures. The importance of sample preservation methods in metagenomics was proved in human and fish larval microbiome (Choo et al. 2015; Sumithra et al. 2022). The sample size is based on the microbial concentration. While sea samples with low densities of microbes require vast volumes and filter-based concentrations, samples with high densities of microorganisms, like stool, may require only one sample. Decontamination of the samples is essential for certain samples. Humic acid must be eliminated from soil samples before DNA extraction. If the target microorganisms are associated with humans or animals, the host DNA has to be removed by fractionation or selective lysis. As a result, particular procedures were created for various samples, including fish tissues, human faeces, tropical soils, water samples, and plant tissues. Physical fractionation techniques such as centrifugation, flow cytometry, or selective filtration are employed to analyze viruses in seawater. Therefore, to guarantee the representative extraction of DNA, a comparison of several approaches must be carried out before using new

types of samples. Furthermore, numerous efforts are underway to use single DNA isolation techniques to identify the microbial diversity from different sample types. Direct extraction and indirect extraction are the two general techniques for extracting DNA. Direct extraction is a straightforward, effective technique that can extract 1 to 50 kb DNA, despite its low purity. It employs physical (example: freeze-thaw) and chemical (examples: protease, lysozyme, lysostaphin, mutanolysin, and SDS) lysing of microbes to release microbial DNA. Indirect extraction necessitates the initial isolation of the microbial cells followed by DNA extraction. Despite being laborious and ineffective, the technique is appropriate for extracting pure DNA of 20 to 500 kb in size. After DNA fragment extraction, enrichment techniques such as suppression subtractive hybridization, stable isotope probing, DNA chip technology, etc. are used for low-abundance DNA samples (Probst et al., 2015).

**Construction of metagenomic library**: Creating a metagenomic library involves first mechanically or digesting microbial DNA to a specific size, then randomly recombining it with a suitable clone vector and transforming it into the right host cells. Functional metagenomics can be carried out on clones with either small or big inserts. Shotgun sequencing is often performed on small insert clones. Plasmids, cosmids, fosmids, bacterial artificial chromosome (BAC), yeast artificial chromosome, and phages are the common vectors. For DNA pieces >10 kb, plasmids are appropriate and have been applied to identify several enzymes and antimicrobial resistance-encoding genes. Cloning big genes or several gene segments (20 to 40 kb) is possible with cosmid and fosmid. BAC and YAC are suitable for DNA segments of 200 to 450 kb. The host range can be extended by using the BAC shuttle vector. pCC1FOS and pWE15 are the two common vectors used for cloning big DNA segments. Moreover, phylogenetic interpretation of the fragment is made possible by the cloning DNA into fosmids or BAC. Plasmid-based small-insert libraries need 3 to 20 times more clones compared to fosmid or BAC-based clones to obtain similar microbial coverage. After selecting the required clones, they are transformed into appropriate hosts. *Escherichia coli* is the frequently used host. New host cells, including *Mycobacterium, Pseudomonas*, and *Streptomyces* (Rebets et al., 2017) are applied progressively. The main and most effective technique for delivering metagenomics DNA, particularly large-insert libraries is electroporation. The libraries are then preserved as glycerol stock broths containing vector-selective antibiotics and 10 to 15% glycerol.

In the amplicon-based metagenomics PCR amplification is done on the extracted metagenomic DNA. The DNA adapters are then annealed at the 3' and 5' ends of the amplicons. DNA adapters

are double-stranded with~ 20 to 40 bp size with known sequences. One adapter has the primer annealing spot, and the second adapter will attach the amplicon to a surface (beads with a complementary DNA sequence) for sequencing. The ligated DNA are selected by gel electrophoresis, columns, or magnetic beads. The final crucial step is to quantify the library which provides information on its concentration and various fragments. Real-time PCR is an accurate technique for library quantification but is unable to estimate the library size. Sequencing is the next step after preparing the metagenomic DNA library.

**Sequencing**: The sequencing platforms for metagenomics are determined by the research goal and application. The major considerations are

1. GC bias: In metagenomics research aimed at microbial diversity, the sequencing analysis platform must have little to no GC bias.
2. Error rate: As metagenomes are extremely diverse and contain several unknown sequences, we have to select a sequencing methodology with a low error rate.
3. Throughput: Since higher throughput (number of reads) helps in identifying rare microbes, a higher throughput sequencing methodology is apt for diversity research. Nevertheless, studies that focus on the differences between environmental trials before and after, like the gut microbiota pre- and post-antibiotic treatment do not require high throughput, where the study is to consider the impact on the primary species.
4. Read length: More read length improves the resolution (up to species level for long-reads and phyla for short-reads). The species-level discrimination in 16S rRNA-based metagenomics usually requires a read length of 500 to 700 bp. However, to find out novel genes or proteins, a minimum of 188 amino acids has to be covered.

**Sequence analysis:** The following procedures are used to analyze the metagenomic data after the sequencing: (1) read quality control; (2) assembly and binning; (3) annotation/taxonomic and functional profiling; and (4) data visualisation. The initial stage involves eliminating low-quality reads. The two primary components of quality control are the de-noising and de-host sequence. After that, short read segments are assembled to create longer genomic contigs either by de-novo assembly or reference-based assembly (using software programs like AMOS, MIRA, or Newbler). Larger computing resources, such as computer with RAM having hundreds of gigabytes are needed for de-novo assembly. Both assembly-based and assembly-free metagenomic profiling are used in data analysis. In metagenomics data representing a complex microbial community with less sequencing depth, several reads covering the same DNA segment will not be obtained, therefore assembly is not needed. However, it is impossible

to analyse longer and more complex genetic elements without assembly. Categorization of DNA sequences into groups representing a distinct genome or genomes from closely related creatures or phylogenetic study of taxonomic groups is known as binning. Binning techniques are of two types, viz. compositional binning (the sequence is binned based on the identity of the conserved nucleotide composition such as GC or specific abundance distribution of k-mers in the genome) and similarity-based binning (the sequence is binned based on the identity of the gene encoded by the unidentified DNA segment with identified genes in reference database). Phylopythia, PCAHIER, S-GSOM and TACAO are the commonly used compositional-based binning algorithms and IMG/M, MEGAN, SOrt-ITEMS, CARMA, MG-RAST, and MetaPhyler are the similarity-based binning software. PhymmBL and MetaCluster use both composition and similarity for binning. The composition-based binning is unreliable for short reads. However, if the data contains two or more genomes, there is possibility for generating "chimeric" bins by similarity-based binning. So, supplementary binning using compositional features can be used to separate the two genomes. There are two steps in annotation; initially, features of interest are identified (gene prediction) and, then, putative gene functions and taxonomic neighbours are allotted (functional annotation). A number of tools were created specially to deal with the metagenomic gene prediction of protein-coding and non-protein coding regions. For the majority of metagenomic experiments, functional annotation poses the greatest computational hurdle. The metagenomic datasets can be functionally annotated using a variety of reference databases, including KEGG, eggNOG, COG/KOG, PFAM, and TIGRFAM.

Microbial composition is the primary output of amplicon metagenomics, and VSEARCH, DADA2, and Deblur are some tools designed to clean, cluster, and quantify amplicon metagenomics. Compared to previous techniques, DADA2 can detect precise amplicon sequence variations and generate less false positive sequence variations. Databases for shotgun metagenomics are based on functional analysis include KEGG (Kyoto Encyclopedia of Genes and Genomes) and COG (Clusters of Orthologous Groups of Proteins). Furthermore, tools such as PICRUSt2, MEGAN, GeneMark-HM, and Prokka were created for functional profiling and prediction. Software for converting taxonomic profiles to putative functional profiles can be found in databases such as PICRUSt2 and Functional Annotation of Prokaryotic Taxa (FAPROTAX). MetaPhlAn 4 employs a database of >one million prokaryotic reference genomes. HUMAnN 3, StrainPhlAn 3, PanPhlAn 3, and PhyloPhlAn 3 are effective for the describing at the strain level. QIIME2 is another useful tool for evaluating and visualizing

metagenomic data. Megan is usually chosen to analyze and compare several sets of metagenomic data, while 'GeneMark' and 'metagenemark' tools are useful for gene prediction. It has been demonstrated that the introduction of MeganServer and the combination of DIAMOND + MEGAN offer an easily accessible and effective tool for investigating short and long metagenomic data. Antibiotic resistance features of microbes in the raw reads can be detected through 'ARG-ANNOT' (Antibiotic Resistance Gene-ANNOTation) and DeepARG. 'MEGAHIT, Minia, and Meraga' are good in assembling the genome across a broad range. 'PhyloPythiaS' and 'Kraken' perform well before the family level. Shotgun-FunctionalizeR package is good for statistical analysis of metagenomic data. Metaviz, Krona, and BURRITO are the major tools for illustrating the metagenomic results and giving interactive illustrations of metagenomic results. All metagenomic information must be kept with the National Center for Biotechnology Information (NCBI).

**Metagenomics in the aquaculture context**

The prospective uses of metagenomic tools in aquaculture include

1. **To investigate the microbial diversity within cultured fish and culture facilities**: The first step in comprehending the effects of microbial alterations in aquaculture systems is to document the bacteria found in healthy animals and their habitats (Sumithra et al., 2024). Since there will be a substantial loss of bacterial diversity and an increase in opportunistic pathogen abundance during stress and diseases, knowledge of bacterial diversity and abundance can shed light on the health status. To fully understand the complex host-microbe interactions, early disease diagnosis and to utilize the benefits of a healthy microbiome, these investigations are essential. The pathophysiology of different diseases can also be better understood by characterizing the relative effects of factors changing the composition of the symbiotic microbial communities. The metagenomics approach has been used to characterize the microbial communities of bio-filters in recirculating aquaculture systems to improve our understanding of applied microbial ecology (Huang et al. 2016). The management of various culture facilities to maximize the yield and sustainability of operation will be aided by knowledge of microbial community profiling in aquaculture environments, quantitative characterization of factors affecting these microbial communities' structure and function, changes over time, and whether and how community diversity changes when culturing the same species under different densities or organic loads etc.

2. **Understanding the core microbiome functions**: In addition to population analysis, it's critical to comprehend how these commensal microorganisms affect immunity, behaviour, digestion, and general fish health. These investigations will improve our understanding of fish health and aid in the development of improved probiotics (feed and water probiotics) as well as the use of microbial manipulation for disease detection, treatment, and prevention in culture systems. The majority of these studies have concentrated on culturable microbes (Martınez-Cordova et al. 2015); nevertheless, there is still little information available about their uncultivable microbes. In a similar vein, identifying the metabolic functions of microorganisms in aquaculture systems is crucial to improving comprehension and raising the likelihood of modifying the microcosms produced by aquaculture to achieve a better understanding and increase the possibility of manipulating the microcosms created by aquaculture to understand the biogeochemical cycles of nutrients within and outside of ponds, the modification of bacterial communities and the disruption of key processes that lead to disease. Numerous studies on the human microbiome have focused on identifying the core microorganisms, which is an intriguing goal in microbiome research. Apart from one study on the *Seriola lalandi* core microbiome, there are relatively few studies that identified the core microbiome of fish larvae (Walburn et al. 2019). We have recently thoroughly characterized the entire larval microbiome of *Trachinotus blochii* from the fertilized egg to the completion of metamorphosis, and gut from prejuveniles and the study identified the larval core microbes (Sumithra et al. 2024).

3. Metagenomics can be applied **to monitor water and sediment quality** to ensure optimal environmental conditions during aquaculture practices.

4. **Identification of novel pathogens and a comprehensive understanding of the pathophysiology of disease:** New diseases frequently affect aquaculture across almost all taxonomic groups. Through metagenomics, the microbial profile during the disease process helps to clarify the presence of rare pathogens, the relevant strains, the synergistic activities of numerous pathogens, and potential vectors that promote the pathogen's growth. For instance, early mortality syndrome and acute hepatopancreatic necrosis disease (AHPND) have cost the aquaculture industry billions of dollars. However, it is still unknown which strain of *V. parahaemolyticus* is causing the illness (Kondo et al. 2014). In these cases, a metagenomics profile could offer a clear picture.

5. **Finding new microbes**: Diverse microorganisms, such as bacteria, viruses, and fungi, are frequently present inn aquatic and marine habitats. The capacity of current detection techniques to detect new microorganisms, particularly viruses, is restricted. Thus, the use of metagenomics will aid in the identification of new microorganisms, particularly viruses and non-culturable bacteria.

6. **Biofloc formulation:** Since aquaculture has primarily relied on artificial feed, 70 to 80% of it currently ends up in the water column or sediments, wasting feed and money and producing harmful residues that restrict the growth rates of the cultured species and intensify cultivation methods. Biofloc technology has emerged as a viable approach to achieving aquaculture sustainability and intensification in this context. Metagenomics-driven diversity studies are required to characterize the bacteria that generate these bioflocs. Microbes of bioflocs are thought to contain genes involved in the metabolic pathways of nitrogenous and carbonous substances; however, shotgun metagenomics will be necessary to fully understand the mechanism of such compounds' metabolism. Metagenomics should be employed as a tool to determine whether the microbes that make up the bioflocs have gene families encoding for proteins linked to the antagonist activities or whether other mechanisms stop the pathogens' normal activity. Recent evidence indicates that bioflocs can prevent disease outbreaks. Finally, to identify the best bacterial species for flocculation, ascertain the nutritional makeup, and determine the relationships among the microbes contained in bioflocs metagenomics can play a key role (Martınez-Porchas and Vargas-Albores 2015).

7. **Probiotic designs**: Metagenomics will help with the identification of new probiotics and in the tracking the impact of probiotic use on commensal microbial populations.

8. **For identifying new molecules/genes**: Metagenomics can also be used to identify enzymes, therapeutics and other bioactive chemicals from different marine/aquatic habitats. Metagenomics can identify new compounds and new genes that code for proteins or enzymes that are more active than those currently known for a given biological reaction. Similar to this, new metabolically and ecologically important genes in the ocean, such as those about different biogeochemical cycles, photosynthesis, the response to phosphate and nitrogen stress, nucleotide scavenging *etc*. can be explored.

9. **Improved comprehension of natural scenarios:** We can better understand the scenarios that favour and prevent natural disasters and disease outbreaks by using metagenomics to investigate the imbalances in microbial communities during these

events. Similarly, no metagenomics-based research has been done on the microbial diversity found in the wastewater released into the environment by aquaculture and mariculture farms, nor on the alterations made to the native microbiota.

10. **For bioremediation and/or recirculation**: Millions of tons of trash are produced by aquaculture and mariculture operations, which are rich in carbon, phosphorous, sulfur, and nitrogenous substances. In order to understand the metabolic pathways of a particular microbial community to recycle nutrients and wastes in aquaculture, metagenomics can be used. Additionally, these genes may be investigated for recirculation and/or bioremediation.

11. **For understanding the mechanisms underlying antimicrobial resistance in marine ecosystem and aquaculture amenities:** Remarkably little is known about the diversity, origins, and distribution of resistance genes, especially those found in unculturable microorganisms. Studies on these topics are necessary for aquaculture, although they are uncommon, and they will be beneficial.

## Conclusion

Metagenomics and microbiome research can revolutionize our understanding of microbial ecosystems within aquaculture environments, offering transformative solutions to longstanding challenges in the industry. By unveiling the composition, functions, and dynamics of microbial communities, metagenomic approaches enable precise monitoring and management of aquaculture systems, fostering healthier and more sustainable production practices. Integrating microbiome-based strategies can enhance water quality, booster host immunity, mitigate the impacts of pathogens, and optimize nutrient cycling. These advancements not only improve productivity but also align with the global imperative for environmentally sustainable aquaculture. The development of tailored microbial interventions, such as probiotics and bioaugmentation techniques, further underscores the practical applications of metagenomic insights. Conversely, the tool has been underexplored in aquaculture field. As the aquaculture industry continues to expand, the importance of leveraging microbial science cannot be overstated. Future research should focus on refining analytical techniques, building comprehensive microbial databases, and translating findings into scalable, cost-effective solutions.

**References:**

Breitwieser FP, Pertea M, Zimin AV, Salzberg SL (2019). Human contamination in bacterial genomes has created thousands of spurious proteins. Genome Res 29:954-960.

Choo JM, Leong LE, Rogers GB (2015) Sample storage conditions significantly influence faecal microbiome profiles. Sci Rep 5:16350

Huang, Z., Wan, R., Song, X. et al. (2016). Metagenomic analysis shows diverse, distinct bacterial communities in biofilters among different marine recirculating aquaculture systems. Aquacult Int, 24, 1393–1408.

Kondo H, Tinwongger S, Proespraiwong P, Mavichak R, Unajak S, No-zaki R, Hirono I (2014). Draft genome sequences of six strains of *Vibrio parahaemolyticus* isolated from early mortality syndrome/acute hepatopancreatic necrosis disease shrimp in Thailand, Genome Announcement, 2, e00221-e00214.

Maester TC, Pereira MR, Machado Sierra EGet al., (2016). Characterization of EST3: a metagenome-derived esterase with suitable properties for biotechnological applications. Appl Microbiol Biot, 100, 5815-27.

Martínez-Córdova L. R., Emerenciano M., Miranda-Baeza A., and Martínez-Porchas M., (2015). Microbial-Based Systems for Aquaculture of Fish and Shrimp: An Updated Review, Reviews in Aquaculture, 7, no. 2, 131–148.

Nagayama H, Sugawara T, Endo R et al., (2015). Isolation of oxygenase genes for indigo-forming activity from an artificially polluted soil metagenome by functional screening using Pseudomonas putida strains as hosts. Appl Microbiol Biot, 99, 4453–70.

Petrovskaya, L.E., Novototskaya-Vlasova, K.A., Spirina, E.V., Durdenko, E.V., Lomakina, G.Y., Zavialova, M.G., Nikolaev, E.N., Rivkina, E.M., (2016). Expression and characterization of a new esterase with GCSAG motif from a permafrost metagenomic library. FEMS Microbiol. Ecol, 92.

Probst, A. J., Weinmaier, T., DeSantis, T. Z., Santo Domingo, J. W., & Ashbolt, N. (2015). New perspectives on microbial community distortion after whole-genome amplification. PloS one, 10(5), e0124158.

Rebets, Y., Kormanec, J., Luzhetskyy, A., Bernaerts, K., & Anné, J. (2017). Cloning and Expression of Metagenomic DNA in Streptomyces lividans and Subsequent Fermentation for Optimized Production. Methods in molecular biology (Clifton, N.J.), 1539, 99-144.

Stoveken J, Singh R, Kolkenbrock S, Zakrzewski M, Wibberg D, Eikmeyer FG, Puhler A, Schluter A, Moerschbacher BM. (2015). Successful heterologous expression of a novel chitinase identified by sequence analyses of the metagenome from a chitin-enriched soil sample. J Biotechnol, 201, 60-8.

Sumithra TG, Sharma SRK, Suresh G, Gop AP, Surya S, Gomathi P, Anil MK, Sajina KA, Reshma KJ, Ebeneezar S, Narasimapallavan I and Gopalakrishnan A. (2024) Mechanistic insights into the early life stage microbiota of silver pompano (Trachinotus blochii). Front. Microbiol, 15,1356828.

Sumithra, T.G., Sharma, S.R.K., Gayathri, S. et al. Comparative evaluation of fish larval preservation methods on microbiome profiles to aid in metagenomics research. Appl Microbiol Biotechnol 106, 4719–4735 (2022). https://doi.org/10.1007/s00253-022-12026-6

Venter JC, Remington K, Heidelberg JF et al (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science, 304, 66–74.

Walburn, J., Wemheuer, B., Thomas, T., Copeland, E., O'Connor, W., Booth, M., et al. (2019). Diet and diet-associated bacteria shape early microbiota development in yellowtail kingfish (Seriola lalandi), J. Microbial. Biotechnol,12, 275–288.

Yang, B., Wang, Y., & Qian, P. Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC bioinformatics, 17, 135.

Zhang L, Chen F, Zeng Z, Xu M, Sun F, Yang L, Bi X, Lin Y, Gao Y, Hao H, Yi W, Li M and Xie Y (2021). Advances in Metagenomics and Its Application in Environmental Microorganisms. Front. Microbiol, 12, 766364.

Zhao, X., Liu, W., Deng, Y., & Zhu, J. Y. (2017). Low-temperature microbial and direct conversion of lignocellulosic biomass to electricity: Advances and challenges. Renewable and Sustainable Energy Reviews, 71, 268-282

# The deep-sea microbiome: Invisible frontiers for biotechnological applications

Neethu N*, Vijayakumar SM, Gayathri Suresh, Sumithra TG, Krupesha Sharma SR, Ratheesh Kumar R,

Marine biotechnology fish nutrition and health division., ICAR-CMFRI, Kochi-18,, *neethukudallur@gmail.com

## Introduction

Microbes are tiny living organisms that are visible only under the microscope. They are metabolically versatile and essential in every part of life. They are beneficial in various areas, which include food production, clinical drug production, wastewater treatment, production of biofuels, chemicals and enzymes, maintaining soil fertility in agriculture, biofloc technology in aquaculture *etc*. They also influence climate change, nutrient cycling, and various ecosystems making it imperative to study. They are recognized as a rich niche for new enzymes, bioactive chemicals, and other useful compounds. Hence, studying microbes is essential for uncovering their roles in various ecosystems. The microbes are ubiquitous life forms and inhabitable in diverse habitats ranging from terrestrial soils and freshwater ecosystems to extreme environments like polar regions, deserts and the deep sea. Among these habitats, the deep sea stands out as one of the most challenging yet fascinating environments for microbial life. The deep sea is a harsh environment that lies below the depth of 200 m and is characterized by high pressure, cold temperatures, darkness, and restricted nutrient supply. (Danovaro *et al.*, 2017). Despite these harsh conditions, deep-sea microbes have evolved remarkable adaptations, and contribute to important biogeochemical processes. In addition to being vital to a number of biogeochemical processes, the diverse microbial population found in deep-sea habitats also produces a multitude of natural compounds and enzymes with significant scientific and commercial applications (Zhang *et al.*, 2024). Microbes that survive in extreme hydrostatic pressure are known as barophiles or piezophiles, which are characterized by evolved mechanisms to maintain cellular integrity under such pressures. Whereas, the microbes in deep-sea cold environments and those in hydrothermal vents (which reach >300°C temperature) are called psychrophiles and thermophiles, respectively. Recent advancements in

molecular biology and oceanographic technologies have begun to unveil the remarkable complexity and ecological and biotechnological significance of the deep-sea microbiome.

**Deep-sea microbial diversity**

The deep-sea microbes have various sub-habitats, such as deep pelagic waters, hydrothermal vents, cold seeps, deep sea sediments and abyssal plains. Anyhow, hydrothermal vents are hotspots of microbial diversity and activity, which are fueled by the geothermal energy released from Earth's crust. These microbes in the hydrothermal vents oxidize hydrogen sulfide and methane to generate energy, forming the foundation of ecosystems in that region (Orcutt *et al.*, 2011). Similarly, methanotrophic and sulfur-oxidizing bacteria dominate in the cold seep areas. Whereas, the microbes in the deep pelagic waters decompose the organic materials that sink from surface waters, which further helps maintain the deep-sea carbon cycle (DeLong and Pace, 2001). Deep-sea sediments contain diverse anaerobic microbial communities involved in processes such as methane oxidation, sulfate reduction, and decomposition of organic matter. Hence, the role of deep-sea microbes in carbon cycling and nutrient recycling highlights their significance in ecology and sustaining marine biodiversity (Orcutt *et al.*, 2011). The surface waters of the Sargasso Sea revealed nine bacterial phyla and two archaeal phyla, of the majority were from the α- and γ-*Proteobacteria (Venter et al., 2004)*.

The presence of life in hydrothermal vents in the deep sea raises crucial queries about the adaptive strategies that thermophilic microorganisms have developed, allowing them to flourish at temperatures far above the water's boiling point. The technique could be multifactorial involving all of the major cellular components (Poli *et al.*, 2017). Moreover, chemoautotrophic microbes are abundant in the areas of volcanically driven warm vents. For example, the microbes from the kingdom *Archaea* are the primary microbial species isolated from such extreme temperatures, namely *Euryarchaeota* and *Crenarchaeota*. Several microbes, which are hyperpiezophilic are also found in hydrothermal vents regions. *Thermococcus* species are the predominant deep-sea piezophiles. *Methanococcus jannaschii* was discovered at 2600 m depth in the deep sea having around 75 MPa pressure. *Marinitoga piezophila*, a sulfate-reducing thermospiezophile, was discovered at 2630 m depth in the East-Pacific Rise.

The microbes that grow at lower temperatures near zero are called psychrophiles. Taxonomically, the psychrophiles are listed in several genera of *Archaea* and *Bacteria* domains. Over 40% of psychrophiles have been discovered from marine settings. A variety of halophilic bacteria with great salinity tolerance have also been discovered from sea sands. The

deeper part of the Red Sea has shown the presence of a few interesting microorganisms, including *Salinisphaera shabanensis*, *Halorhabdus tiamatea etc* (Poli et al. 2017). Cold deep sea deposits were used to characterize two new species: the archaeon "*Candidatus Prometheoarchaeum syntrophicum*" in the *Asgard* superphylum and *Xianfuyuplasma coldseepsis* in the phylum *Bacillota* (Imachi *et al.* 2020).

There are conflicting outcomes of dominance between *Archaea* and *Bacteria* in the deep-sea sediments. The microbial community in the sediments varies according to the depth, latitude, oxygen and carbon content, hydrothermal influence etc. *The α*, *γ*, and δ-*Proteobacteria*, *Actinobacteria*, *Acidobacteria*, and *Planctomycetes* are the dominant microbes in oxygenated surface sediments, whereas Archaeal diversity is relatively low. The Archaeal community will dominate in the deep sediments with rich organic content. Surprisingly, a thermophilic bacterium (*Desulfotomaculum* sp.) in the cold surface sediments of the Arctic Ocean was discovered which is induced by the heating of sediments. Further, microbial communities from marine animals, including deep sea fishes, invertebrates, corals, and sponges have been widely investigated in several studies (Kennedy *et al.*, 2010). Nevertheless, studying the microbial diversity in the deep-sea environment is a massive work, so it is briefly described by scratching out the surface information.

**Characteristics of deep-sea microbes**

**Adaptations to extreme environment**

Deep sea microbes are extremophilic and can withstand environments like high temperature (thermophiles) or low temperature (psychrophiles), high salinity (halophiles), high pH (alkalophiles), low pH (acidophiles), anoxia, high pressure (piezophiles). The microorganisms living in these conditions are known as psychropiezophiles. On the other hand, microbes in areas with high temperatures (near hydrothermal vents), are called thermo-piezophiles. In these ecosystems, psychrotrophs are more adaptable than psychrophiles and can grow across a wider range of temperatures. Microbial adaptation mechanisms to high hydrostatic pressure (HHP) are diverse, including transcriptional regulation, membrane adjustments, and the production of osmolytes. The adaptations help the cells withstand stresses, like nutrient limitation and extreme pH, temperature, and salt conditions. HHP above 200 MPa can be lethal to most microorganisms and is even used to preserve food (Oger and Jebbar, 2010). Biological membranes play a critical role in microbial adaptation as they act as barriers, regulate ion gradients for energy processing, and are involved in metabolic and signalling pathways. Extremophiles of hot, cold, and salty habitats, exhibit specialized lipids, enzymes, and

biopolymers which help to adapt to these extreme conditions. The presence of fungi in deep-sea ecosystems has only recently begun to gain recognition, with much less research compared to bacteria, despite the first isolation of deep-sea fungi by Roth et al. happened in 1964.

**Unique metabolic pathways**

Sunlight is the primary source for photosynthesis in plants and terrestrial microbes to synthesise organic carbon that sustains life on Earth. Conversely, the diverse environmental constraints within the sea have made diverse microbes adopt varied survival and growth tactics. The pelagic photosynthetic microbes produce most of the organic matter within the sea. Whereas, deep sediment microbes degrade dead marine animals and marine snow. Additionally, extensive regions of the dark deep sea have the potential for primary production also through chemosynthesis (Das *et al.*, 2011). In these regions, reduced sulfur compounds react with free oxygen, providing a vital energy source and electron sink for sulfur-oxidizing bacteria. The chemosynthetic bacteria which reduce carbon dioxide to organic carbon and release energy by biological oxidation use chemical energy to synthesize biomass. Chemosynthesis is typically used to describe chemoautolithotrophic processes at hydrothermal vents and seeps, where they use inorganic compounds as electron donors. The deep-sea environment is abundant with elements which foster substantial primary productivity by chemosynthetic microorganisms as the result of anaerobic chemosynthesis. The major source of methane in the oceans is emitted from deep-sea hydrothermal vents. Hydrothermal vents, whale falls, wood falls, and cold seep communities thrive by utilizing the chemical energy in elements like methane and sulfide through symbiotic relationships with chemosynthetic microorganisms, allowing them to thrive in an environment where food is typically scarce. Another strategy is a minimalist approach, typically reported in a photosynthetic pelagic bacterium *Prochlorococcus marinus*, where the microbe has decreased its genome to 1.66 Mbp to have a competitive merit (Dufresne et al. 2005). Similarly, *Pelagibacter ubique* belonging to marine *α-Proteobacteria* is the free-living microbe with the smallest genome (Giovannoni *et al.*, 2005). Briefly, deep sea microbes have adapted novel metabolic and biochemical pathways to the extremes of temperature and pressure, which can be distinctively suited for several industrial applications.

**Specialized enzymes and proteins for survival**

Extremophiles can produce enzymes that remain functional under extreme environments. These include α-amylase, isocitrate dehydrogenase, lipase, β-lactamase, triose phosphate isomerase, and subtilisin. In addition to these, enzymes such as alkaline protease, alkaline cyclomaltodextrin-glucanotransferase, lipases, alkaline cellulases, alkaline phosphatase, endo-

1,3-β-D-glucanase, and alkaline metalloendopeptidase, as well as halotolerant enzymes like amylases, proteases, lipases, nuclease H, and cyclophilin-type peptidyl-prolyl cis-trans isomerase, have also been identified (Chandrasekaran and Kumar, 1997). Bioactive compounds present in deep-sea microorganisms help them to exhibit unique resistance to high pressure, low temperatures, and nutrient limitations. Cold-active enzymes, also known as psychrozymes, derived from these deep-sea microbes, can enable efficient food processing under refrigerated or cold-storage conditions in the food industry. Additionally, the survival of psychrophilic bacteria in cold environments might be supported by osmoprotectant and energy-generation strategies at low temperatures. The psychrophilic enzymes remain highly functional in the cold environment due to the reduced molecular motion. Specific structural modifications, including decreased core hydrophobicity, fewer prolyl residues in loops, enhanced surface hydrophobicity, and additional glycyl residues enhance conformational flexibility and support adaptations exhibited by the psychrophilic bacteria. Other factors contributing to this adaptation include weaker subunit and domain interactions, a lower arginyl: lysyl ratio, reduced secondary structure content, longer loops, increased prolyl residues in α-helices, decreased flexibility, weaker metal-binding sites, modifications in disulfide bridges and electrostatic interactions (Laye *et al.*, 2017).

A variety of membrane adaptations was observed in psychrophilic bacteria to cope with the low temperatures, which includes an increased ratio of polyunsaturated to saturated fatty acids in membrane phospholipids, alterations in the lipid classes present in the cell membrane, a reduction in the size and charge of lipid head groups, and the conversion of trans fatty acids to their cis-isomeric forms. These modifications affect the phospholipid packing within the membrane, helping to maintain membrane fluidity and functionality in cold environments. Moreover, many psychrophilic bacteria produce cholesterol-like molecules that can integrate into the membrane lipid bilayer. These cholesterol analogues play an important crucial role as membrane components, aiding the adaptation to cold temperatures. Further, cold shock proteins, especially cold-shock domain A (CsdA) are essential multifunctional proteins that support cellular function in the deep-sea microbes at low temperatures. These proteins work together with RNA-binding proteins to preserve cellular homeostasis and support the adaptation to cold environments (De Maayer *et al.*, 2014).

**Microbial interactions in the marine ecosystem**

The microbial communities coexist and compete in an environment with several networks of interactions, which are necessary to maintain the stability of the ecosystem. The interactions

would be mutual, commensal, competitive, predatory and/or syntrophic. For example, a mutual interaction was observed between sulfur-oxidizing bacteria and tube worms in hydrothermal vents, where the bacteria provide nutrients to the worms and they receive hydrogen sulfide and oxygen from the worms for metabolism. Some deep-sea mussels host the methane-oxidizing symbionts, a rare feature among bivalves. Few microbes benefit from the nutrients produced through the degradation of organic matter by bacteria, which shows the commensal interaction. In the syntrophic cooperative interaction, one microbe depends on another to produce or consume metabolic byproducts, which are usually exhibited by methanogenic and hydrogen-producing bacteria in an anaerobic environment. The deep sea microbes are also essential for carbon cycling and nutrient regeneration, making them vital for sustaining oceanic productivity. Viruses present an exciting opportunity to explore a largely new field of deep-sea ecology, which may be crucial for understanding ecosystem functioning. Viruses would infect all marine life, from the largest mammals to the smallest microbes. Since prokaryotes are the predominant members and dominate the oceans in terms of biomass, the interactions between viruses and members of the *Bacteria* and Archaea domains are key processes that influence ecosystem functioning on a global scale. In cold seeps, the free-living thiotrophic bacteria form visible mats on the seafloor, which has a greater ecological importance. They have a greater role in removing toxic hydrogen sulfide and contributing to chemosynthetic biomass production, which serves as a food source for other organisms. Microbial biofilms play an essential function in the primary establishment in the deep-sea hydrothermal vents, allowing microbes to exploit access to nutrients and energy, thus supporting various communities. Understanding the microbial interactions in the deep sea reveals several truths with applications in fields like bioremediation and biotechnology (Liu *et al.*, 2019).

**Molecular tools to study deep-sea microbes**

As only 0.0001% to 0.1% of the marine microbes are culturable, a culture-independent method is necessary to study the marine microbes. The advent of metagenomic tools has helped to explore the uncultured microbial diversity from different ecosystems including marine environments and permitted the entree to the metabolic pathways of uncultured microbes. Diversity and the richness of marine microbes have been studied using 16S rRNA-based metagenomics. However, whole metagenome shotgun sequencing is mainly employed for studying marine microbes, whereby phylogeny or putative functional genes of microbial DNA can be studied. If the putative functional genes are detected in proximity to phylogenetic marker genes on the same cloned segment, the taxonomy of the function can be known. The shotgun

sequencing usually gives read lengths of 600 to 900 bp, which can be increased using the fosmid. The sequences from fosmid clones produce a more complete aspect of the identified function, giving an increased chance for reconstructing the metabolic pathways within an environment (Kennedy et al. 2010). The taxonomic assignment to the detected functional genes from shotgun sequencing requires software, like 'Metagenomic and rDNA Taxonomy Assigner (MARTA)', or 'Genomic Signature-based Taxonomic Classifier (GSTax Classifier)'. MARTA gives the taxonomic status using percent-identity and thresholds; by comparing the MARTA results with 'CARMA' and 'RDP-II Classifier' (Horton, Bodenhausen and Bergelson, 2010). The 'GSTax Classifier' compares the genomic signatures of metagenomic query sequences to reference genome databases (Yu *et al.*, 2009). The function assignment needs profile-to-sequence searches, in which the predicted protein sequences are matched against the sequence alignments from the protein families using RPS-BLAST (Markowitz *et al.*, 2006). A dedicated data resource, namely the Community Cyber Infrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA) is present for analysing the marine-derived data. 'Integrated Microbial Genome/Microbiome (IMG/M)' is another useful tool for assessing the functional potential of microbial populations based on their metagenome through comparative analysis using the JGI-Integrated Microbial Genomes system; MetaBioME (a resource for novel enzyme discovery) and Orphelia (for identifying ORFs in short metagenomic sequences of unknown phylogeny). Further, the 'Metagenomics RAST server' gives an automated platform for annotating metagenome data. The 'JANE' resource enables the mapping of prokaryotic Expressed Sequence Tags from the DNA sequence to the RNA/transcriptomic data of the related template genomes. 'WebCARMA' resource enables the functional and taxonomic sorting of unassembled small reads (~35 bp). 'MEGAN' permits the comparison of several datasets, precisely permitting visual and statistical comparison.

In addition to sequence-based metagenomics, functional metagenomics, which focuses on phenotypic screening of the library clones, has been effectively utilized in marine materials. This technique does not rely on sequence analysis to identify target genes and can distinguish novel gene classes with known and unknown activities. Several enzymes were discovered in marine environments utilizing functional metagenomics. Future research on deep sea microbiomes is predicted to be revolutionized by novel methodologies such as single-cell genomics, metatranscriptomics, and metaproteomics (Kennedy et al. 2010).

**The potential of marine microbes in biotechnology**

Organisms inhabiting deep-sea hydrothermal vents with extreme physico-chemical conditions can be applied to develop new biotechnological applications. Microbes that live in deep-sea hydrothermal vents are capable of fixing $CO_2$ into biomass and producing the enzyme carbonic anhydrase and hold promise for advancements in biofuel production. Furthermore, barophiles and extremophiles are very useful in high-pressure bioreactors, crude oil degradation, toxic compound removal, and the production of steroid hormones from cholesterol (Kato, Inoue and Horikoshi, 1996). Cold-adapted microorganisms and their enzymes can be used in food processing, enzyme production, waste treatment, bioremediation, and molecular biology. For instance, the cold-active β-galactosidase has been effectively used in dairy products to hydrolyze lactose, alleviating lactose intolerance while maintaining nutritional content that would be lost during high-temperature treatments. The cold-adapted xylanase from *Sorangium cellulosum* has shown excellent performance in improving wheat bread quality without the excess water release commonly observed with thermophilic xylanases (Li *et al.*, 2022). Two marine DNA polymerases from thermophilic eubacteria (Tth polymerase) and archaea (VENT® polymerase) are effective alternatives to Taq DNA polymerase in PCR. A heat and acid stable α-amylase from deep-sea hydrothermal vent *Archaea* could facilitate the bioethanol production from corn (Mathur *et al.*, 2006). Similarly, marine cellulases have been applied to produce biofuels from renewable cellulosic substrates (Shanmughapriya *et al.*, 2010). The cold-active oxidoreductase enzyme from the marine environment has antioxidant properties and prevents oxidative damage to lipids and proteins and helps preserve the colour, texture, and flavour of food products. Marine lipases have potential applications in various food and industrial sectors, especially to enhance the flavour and texture of aged cheeses (de Pascale *et al.*, 2008). The cold-active transferases and proteases have potential applications in producing flavour compounds. Cold-active lyases are particularly applicable in industries like fruit and vegetable juice production. Similarly, alkane hydroxylase genes having the potential for hydrocarbon-bioremediation (oil spills) was discoovered from the marine metagenomics libraries (Wasmund *et al.*, 2009). Apart from the enzymes, novel bioactive compounds produced by deep-sea microbes show promise in drug development, particularly in antibiotics and anticancer therapies (Chen *et al.*, 2024).

**Challenges and prospects**

Studies on deep-sea microbes pose numerous challenges due to the extreme and inaccessible nature of their habitat. Even, the collection of samples from such an environment is a tedious process, which requires advanced expensive instruments and technologies like remotely

operated vehicles, autonomous underwater vehicles or deep-diving submersibles. Further, sampling under extreme conditions without contamination, replicating deep-sea environments in laboratory settings, and deciphering complex microbial interactions are significant hurdles. Culturing the deep-sea microbes in lab conditions is also difficult, as it is difficult to create the deep-sea environment in the lab. The researchers must use specialized equipment to mimic the extreme conditions of the deep ocean, including temperature and pressure. Yet, the equipment may fail to replicate the same conditions, which may lead to wrong interpretation as well. Moreover, studies related to the deep-sea environment also necessitate an understanding of the chemosynthesis process, as the deeper parts lack light. Further, conducting studies on the influence of temporal and spatial variations on the deep-sea microbial populations has become difficult, as it is difficult to monitor the deep-sea environment continuously. This limits the scope of long-term studies on deep-sea microbes. Some aspects such as analysing the complex interactions and metabolic pathways of deep-sea microbes are also challenging and complete understanding is difficult, even though several studies have been conducted. Maintenance of a research vessel, installation and usage of advanced equipment, and crew members are the major financial investments, which often limit the frequency of such studies.

Nevertheless, the deep-sea microbial communities hold enormous potential to encourage innovations in biotechnology and omic studies. The extraction technologies of enzymes from extremophiles, and their usage in industries ranging from pharmaceuticals to biofuels are in developmental stages. Furthermore, scientists are working on discovering novel antibiotics and anticancer compounds from these microbes, which could address several challenges faced in the medical field. Additionally, the study of the metabolic pathways of deep-sea microbes could help to identify strategies for mitigating greenhouse gas emissions. It is already known fact that they can clean up oil spills and other environmental contaminants in harsh marine environments. Hence, it is important to study and understand the deep-sea microbial diversity, their interactions, metabolic pathways, and their role in maintaining the deep-sea ecosystem.

## Conclusion

Deep-sea microbes have evolved remarkable adaptations to survive in challenging marine environments and contribute to important biogeochemical processes. This chapter gives insights into microbial diversity in the deep sea, their unique characters and potential for bioprospecting, and tools for studying the deep-sea microbiome. It is understood that the deep sea is dominated by bacteria and archaea, including extremophiles such as psychrophiles,

piezophiles, and chemo-lithoautotrophs. They have adapted themselves through diverse mechanisms, which are explained in the respective section. High-throughput next-generation DNA sequencing methods have been applied to study deep-sea microbes and the results demonstrated that the oceans carry an extraordinary microbial diversity. Genome-resolved research has also shown the amazing biosynthetic potential of deep-sea microbes for bioprospecting. Further research on the unexplored part of the deep sea is necessary to find novel microbes, their unique capabilities, interactions and their potential in advancing scientific fields such as biotechnology. At the same time, it is crucial to protect vulnerable marine ecosystems to ensure sustainable utilization, particularly in the present climate change scenario and increasing pollution.

**References**

Chandrasekaran, M. and Kumar, S.R. (1997) 'Marine Microbial Enzymes', in *Biotechnology*, pp. 47–79.

Chen, Q. *et al.* (2024) 'AI-based dynamic avoidance in deep-sea mining', *Ocean Engineering*, 311, p. 118945. Available at: https://doi.org/10.1016/j.oceaneng.2024.118945.Danovaro, R. *et al.* (2017) 'The deep-sea under global change', *Current Biology*, 27(11), pp. R461–R465. Available at: https://doi.org/10.1016/j.cub.2017.02.046.

Das, A. *et al.* (2011) 'Chemosynthetic activity prevails in deep-sea sediments of the Central Indian Basin', *Extremophiles*, 15(2), pp. 177–189. Available at: https://doi.org/10.1007/s00792-010-0346-z.

DeLong, E.F. and Pace, N.R. (2001) 'Environmental Diversity of Bacteria and Archaea', *Systematic Biology*. Edited by M. Kane, 50(4), pp. 470–478. Available at: https://doi.org/10.1080/10635150118513.

Dufresne, A., Garczarek, L. and Partensky, F. (2005) 'Accelerated evolution associated with genome reduction in a free-living prokaryote', *Genome Biology*, 6(2), p. R14. Available at: https://doi.org/10.1186/gb-2005-6-2-r14.

Giovannoni, S.J. *et al.* (2005) 'Genome Streamlining in a Cosmopolitan Oceanic Bacterium', *Science*, 309(5738), pp. 1242–1245. Available at: https://doi.org/10.1126/science.1114057.

Horton, M., Bodenhausen, N. and Bergelson, J. (2010) 'MARTA: a suite of Java-based tools for assigning taxonomic status to DNA sequences', *Bioinformatics*, 26(4), pp. 568–569. Available at: https://doi.org/10.1093/bioinformatics/btp682.

Imachi, H. *et al.* (2020) 'Isolation of an archaeon at the prokaryote–eukaryote interface', *Nature*, 577(7791), pp. 519–525. Available at: https://doi.org/10.1038/s41586-019-1916-6.

Kato, C., Inoue, A. and Horikoshi, K. (1996) 'Isolating and characterizing deep-sea marine microorganisms', *Trends in Biotechnology*, 14(1), pp. 6–12. Available at: https://doi.org/10.1016/0167-7799(96)80907-3.

Kennedy, H. *et al.* (2010) 'Seagrass sediments as a global carbon sink: Isotopic constraints', *Global Biogeochemical Cycles*, 24(4). Available at: https://doi.org/10.1029/2010GB003848.

Laye, V.J. *et al.* (2017) 'Key amino acid residues conferring enhanced enzyme activity at cold temperatures in an Antarctic polyextremophilic β-galactosidase', *Proceedings of the National Academy of Sciences*, 114(47), pp. 12530–12535. Available at: https://doi.org/10.1073/pnas.1711542114.

Li, X. *et al.* (2022) 'A novel cold-active GH8 xylanase from cellulolytic myxobacterium and its application in food industry', *Food Chemistry*, 393, p. 133463. Available at: https://doi.org/10.1016/j.foodchem.2022.133463.

Liu, L. *et al.* (2019) 'Mitigation of environmental pollution by genetically engineered bacteria — Current challenges and future perspectives', *Science of The Total Environment*, 667, pp. 444–454. Available at: https://doi.org/10.1016/j.scitotenv.2019.02.390.

De Maayer, P. *et al.* (2014) 'Some like it cold: understanding the survival strategies of psychrophiles', *EMBO reports*, 15(5), pp. 508–517. Available at: https://doi.org/10.1002/embr.201338170.

Markowitz, V.M. *et al.* (2006) 'An experimental metagenome data management and analysis system', *Bioinformatics*, 22(14), pp. e359–e367. Available at: https://doi.org/10.1093/bioinformatics/btl217.

Mathur, E.J. *et al.* (2006) 'A biodiversity-based approach to development of performance enzymes', *Industrial Biotechnology*, 1(4), pp. 283–287.

Oger, P.M. and Jebbar, M. (2010) ‘The many ways of coping with pressure’, *Research in Microbiology*, 161(10), pp. 799–809. Available at: https://doi.org/10.1016/j.resmic.2010.09.017.

Orcutt, B.N. *et al.* (2011) ‘Microbial Ecology of the Dark Ocean above, at, and below the Seafloor’, *Microbiology and Molecular Biology Reviews*, 75(2), pp. 361–422. Available at: https://doi.org/10.1128/MMBR.00039-10.

de Pascale, D. *et al.* (2008) ‘The cold-active Lip1 lipase from the Antarctic bacterium Pseudoalteromonas haloplanktis TAC125 is a member of a new bacterial lipolytic enzyme family’, *Extremophiles*, 12(3), pp. 311–323. Available at: https://doi.org/10.1007/s00792-008-0163-9.

Poli, A. *et al.* (2017) ‘Microbial Diversity in Extreme Marine Habitats and Their Biomolecules.’, *Microorganisms*, 5(2). Available at: https://doi.org/10.3390/microorganisms5020025.

Roth Jr., F.J., Orpurt, P.A. and Ahearn, D.G. (1964) ‘OCCURRENCE AND DISTRIBUTION OF FUNGI IN A SUBTROPICAL MARINE ENVIRONMENT’, *Canadian Journal of Botany*, 42(4), pp. 375–383. Available at: https://doi.org/10.1139/b64-037.

Shanmughapriya, S. *et al.* (2010) ‘Optimization, Purification, and Characterization of Extracellular Mesophilic Alkaline Cellulase from Sponge-Associated Marinobacter sp. MSI032’, *Applied Biochemistry and Biotechnology*, 162(3), pp. 625–640. Available at: https://doi.org/10.1007/s12010-009-8747-0.

Venter, J.C. *et al.* (2004) ‘Environmental Genome Shotgun Sequencing of the Sargasso Sea’, *Science*, 304(5667), pp. 66–74. Available at: https://doi.org/10.1126/science.1093857.

Wasmund, K. *et al.* (2009) ‘Novel Alkane Hydroxylase Gene ( alkB ) Diversity in Sediments Associated with Hydrocarbon Seeps in the Timor Sea, Australia’, *Applied and Environmental Microbiology*, 75(23), pp. 7391–7398. Available at: https://doi.org/10.1128/AEM.01370-09.

Yu, F. *et al.* (2009) ‘GSTaxClassifier: a genomic signature based taxonomic classifier for metagenomic data analysis’, *Bioinformation*, 4(1), pp. 46–49. Available at: https://doi.org/10.6026/97320630004046.

Zhang, C. *et al.* (2024) ‘Deep-sea microbial genetic resources: new frontiers for bioprospecting’, *Trends in Microbiology*, 32(4), pp. 321–324. Available at: https://doi.org/10.1016/j.tim.2024.01.002.

# Polar Microbiome: Life at the Extremes

Dr. Vipin Das P V

Arctic Ecology & Biogeochemistry, National Centre for Polar and Ocean Research, Goa

vipin@ncpor.res.in

## Introduction to Polar Microbiomes

- Polar regions cover **20% of Earth's surface**.
- Life exists despite extreme conditions:
    - Temperatures as low as -40°C
    - Prolonged darkness in winters
    - High UV radiation in summers

Polar microbiomes represent unique and resilient microbial communities that thrive in the Earth's extreme polar regions, which cover approximately 20% of the planet's surface. Despite the harsh environmental conditions—such as temperatures plunging to as low as -40°C, prolonged periods of darkness during the polar winters, and intense ultraviolet (UV) radiation during the summer months—life flourishes in these icy habitats. These regions are dominated by psychrophilic (cold-loving) and psychrotolerant microorganisms, including bacteria, archaea, fungi, viruses, and algae, which play a crucial role in maintaining ecosystem stability. Polar environments are nutrient-limited, yet microbial communities have adapted to efficiently recycle carbon, nitrogen, and other essential elements, sustaining local food webs. These microbes inhabit diverse niches, such as sea ice, glaciers, permafrost, and subglacial lakes, each presenting unique challenges and opportunities for survival. Their ability to endure such extremes not only broadens our understanding of life's adaptability but also provides insights into astrobiology, as polar microbiomes serve as analogs for potential life on icy planets and moons. Furthermore, studying these ecosystems is vital to monitoring the impacts of climate change, as the Polar Regions are warming at an accelerated rate, causing shifts in microbial diversity and their associated ecological roles.

The Polar Regions, encompassing the Arctic in the Northern Hemisphere and Antarctica in the Southern Hemisphere, collectively account for approximately 20% of Earth's surface area. These regions include vast expanses of sea ice, glaciers, permafrost, subglacial lakes, and polar oceans. The Arctic is predominantly an ice-covered ocean surrounded by continents, while Antarctica is a landmass covered by thick ice sheets, surrounded by the Southern Ocean.

Despite their seemingly inhospitable nature, these areas host diverse microbial life that has adapted to thrive in extreme conditions. The Polar Regions play a significant role in regulating Earth's climate, acting as global heat sinks and influencing oceanic and atmospheric circulation patterns. The microbial ecosystems in these regions are not only vital for local food webs but also contribute to global biogeochemical cycles, including carbon and nitrogen cycling. As such, studying these environments is crucial for understanding both Earth's biodiversity and the broader impacts of climate change.

**Polar Regions**

**The Arctic**

The Arctic, often referred to as the "Earth's refrigerator," plays a crucial role in regulating global climate. This region, encompassing the Arctic Ocean and parts of surrounding countries, is experiencing rapid and unprecedented changes due to climate change. The Arctic is warming nearly four times faster than the global average, making it a key indicator of environmental shifts caused by global warming.

**Arctic Amplification**

Arctic amplification is the term used to describe the accelerated warming of the Arctic compared to the rest of the planet. This phenomenon results in the Arctic warming at rates 2-4 times higher than the global average, driven by complex feedback mechanisms and environmental changes unique to the polar region.

The primary driver of Arctic amplification is the ice-albedo feedback. As sea ice melts due to rising temperatures, it exposes darker ocean waters that absorb more solar energy instead of reflecting it back into the atmosphere. This leads to further warming and ice loss in a self-reinforcing cycle. Additionally, warm air and moisture transported from lower latitudes amplify Arctic temperatures. Changes in cloud cover also contribute by trapping more heat in the atmosphere.

The effects of Arctic amplification are profound. The rapid melting of Arctic sea ice and glaciers contributes to global sea-level rise, threatening coastal regions worldwide. Furthermore, this warming disrupts global weather patterns, including the weakening of the jet stream, leading to more frequent extreme weather events in mid-latitudes. Arctic ecosystems and Indigenous communities face unprecedented challenges as the region's physical and biological systems are altered.

Understanding and addressing Arctic amplification is critical not just for protecting the Arctic but for mitigating its global impacts on climate and weather systems.

**Impact of Rising Temperatures**

- **Accelerated Warming**: The Arctic has warmed significantly over the past few decades. This phenomenon, known as Arctic amplification, is driven by feedback mechanisms such as the loss of reflective ice surfaces (albedo effect) and increased heat absorption by darker ocean waters.
- **Melting Glaciers and Ice Caps**: The Greenland Ice Sheet and other Arctic glaciers are melting at alarming rates, contributing to global sea level rise. If the Greenland Ice Sheet were to melt completely, it could raise sea levels by approximately 7 meters.
- **Ecosystem Disruption**: Rising temperatures threaten Arctic ecosystems, affecting species like polar bears, seals, and walruses that rely on sea ice for hunting and breeding.

**Changes in Arctic Sea Ice**

- **Declining Sea Ice Extent**: Arctic sea ice has been shrinking by about 13% per decade since satellite monitoring began in 1979. This reduction affects both the extent and thickness of the ice.
- **Transition to Younger Ice**: Multi-year ice, which is thicker and more resilient, is being replaced by thinner, seasonal ice that melts more easily.
- **Implications for Human Activities**:
    - New shipping routes are opening, such as the Northern Sea Route and the Northwest Passage, reducing travel time between major markets.
    - Increased opportunities for resource extraction, including oil and gas, which pose environmental risks.

**Permafrost Thaw**

- **Release of Greenhouse Gases**: Arctic permafrost stores vast amounts of organic carbon. As it thaws, methane and carbon dioxide are released into the atmosphere, further amplifying global warming.
- **Infrastructure Risks**: Thawing permafrost undermines roads, buildings, and pipelines, posing challenges to Arctic communities and industries.
- **Positive Feedback Loop**: The release of greenhouse gases from permafrost creates a feedback mechanism that accelerates warming, leading to more permafrost thaw.

**Impact on Indigenous Communities**

- **Threats to Traditional Lifestyles**: Indigenous Arctic communities rely on hunting, fishing, and herding for their livelihoods. Climate change is disrupting these activities by altering animal migration patterns and thinning ice.
- **Coastal Erosion and Relocation**: Rising sea levels and coastal erosion are forcing some communities to relocate, often at significant financial and cultural costs.
- **Need for Adaptation**: Indigenous communities are developing resilience strategies, but they require support through funding, infrastructure, and respect for traditional knowledge.

**Global Implications**

- **Weather Pattern Disruptions**: Changes in the Arctic affect global weather patterns. For example, the weakening of the polar jet stream can lead to prolonged extreme weather events such as heatwaves and cold spells.
- **Impact on Ocean Currents**: Melting Arctic ice affects thermohaline circulation, including the Atlantic Meridional Overturning Circulation (AMOC), which regulates temperatures across the Northern Hemisphere.
- **Loss of Albedo Effect**: The reduction in Arctic ice diminishes the Earth's ability to reflect sunlight, leading to increased heat absorption and accelerated global warming.

**The Antarctic**

The Antarctic, Earth's southernmost continent, is a frozen wilderness surrounded by the Southern Ocean. It is home to vast ice sheets, unique ecosystems, and a critical component of the global climate system. While less inhabited and studied than the Arctic, the Antarctic is also experiencing the impacts of climate change, with potential consequences for global sea levels, ocean currents, and biodiversity.

**Key Features of the Antarctic**

- **Geography**: The Antarctic is centered around the South Pole and is covered by the Antarctic Ice Sheet, which contains about 60% of the world's freshwater.
- **Climate**: The coldest, driest, and windiest continent on Earth. Temperatures can drop below −80°C (−112°F).
- **Unique Ecosystem**: Home to species such as penguins, seals, and krill, which form the foundation of marine food webs.
- **International Governance**: Protected under the Antarctic Treaty System, which prohibits military activities and mineral exploitation.

**Impact of Climate Change in the Antarctic**

1. **Melting Ice Sheets**
    - The West Antarctic Ice Sheet is thinning at an alarming rate due to warm ocean currents.
    - Antarctic ice loss has tripled in the past 30 years, contributing significantly to global sea level rise.
    - The potential collapse of major glaciers like Thwaites ("Doomsday Glacier") could raise sea levels by several meters.
2. **Warming Temperatures**
    - Surface temperatures in parts of Antarctica are increasing, particularly on the Antarctic Peninsula, one of the fastest-warming regions on Earth.
    - Record-breaking temperatures have been observed, such as 20.75°C (69.35°F) in 2020.
3. **Ocean Acidification and Ecosystems**
    - Warming oceans and increased CO2 levels are affecting Antarctic marine life, particularly krill, which are a vital food source for penguins, whales, and seals.
    - Melting sea ice disrupts habitats for species like emperor penguins and Adélie penguins.
4. **Changes in Sea Ice**
    - Unlike the Arctic, Antarctic sea ice has shown mixed trends, with some areas experiencing expansion and others significant loss.
    - The variability in sea ice affects global ocean circulation and ecosystems.
5. **Impact on Ocean Currents**
    - The melting of Antarctic ice sheets affects thermohaline circulation, including the Antarctic Bottom Water, a key driver of global ocean currents.
    - Disruption of these currents could impact weather patterns and marine ecosystems worldwide.

**Human and Scientific Significance**

- **Research Stations**: Over 40 permanent research stations operate in Antarctica, contributing to climate science, geology, and biology.
- **Global Implications**: Antarctic changes have far-reaching effects on global sea levels, weather patterns, and biodiversity.

- **Tourism**: Rising interest in Antarctic tourism poses risks to fragile ecosystems, requiring strict regulation.

**Conclusion**

The polar regions are a bellwether for climate change, offering stark evidence of the effects of global warming. Changes in this region have far-reaching implications for ecosystems, communities, and the global climate system. Addressing these challenges requires urgent action, international collaboration, and a commitment to sustainability to ensure the Arctic's future and the health of our planet.

**Polar Microbiomes**

**Habitats of Polar Microbiomes**

- **Sea Ice**:
  - Colonized by psychrophilic bacteria and algae.
  - Provides nutrients for polar marine ecosystems.
- **Permafrost**:
  - Contains microbes over 10,000 years old.
  - Potential reservoir of ancient pathogens.
- **Subglacial Lakes**:
  - E.g., Lake Vostok in Antarctica.
  - Supports chemoautotrophic microbes.
- **Cryoconite Holes**:
  - Microbial hotspots on glaciers.

**Polar Microbiomes: Key Features**

- **Low metabolic rates** allow survival in resource-scarce environments.
- Microbes have short but highly productive growth phases during polar summers.
- Dominant taxa:
  - **Bacteria**: Proteobacteria, Actinobacteria
  - **Archaea**: Methanogens and halophiles
  - **Fungi**: Cold-tolerant yeasts and molds

Adaptations of Polar Microbes

- **Production of Antifreeze Proteins and Cryoprotectants**

  Polar microbes synthesize antifreeze proteins and cryoprotectants like glycerol and trehalose, which prevent ice crystal formation in their cellular fluids, ensuring survival in subzero temperatures.

- **Enzyme Activity Optimized for Low Temperatures**

Their enzymes exhibit structural flexibility, allowing efficient catalytic activity even at freezing conditions. These enzymes, known as psychrophilic enzymes, are crucial for metabolic processes.

- **Slow Metabolic Rates to Conserve Energy**

Polar microbes adopt slow metabolic rates to minimize energy expenditure. This is a survival strategy in nutrient-scarce environments where energy resources are limited.

- **Biofilm Formation for Protection**

Many microbes form biofilms, which provide a protective matrix against extreme environmental conditions such as freezing, dehydration, and UV radiation.

- **Membrane Fluidity Adaptations**

Cell membranes are rich in unsaturated fatty acids to maintain fluidity at low temperatures, ensuring effective transport of nutrients and waste.

- **Efficient DNA Repair Mechanisms**

Enhanced DNA repair pathways protect these microbes from damage caused by high levels of UV radiation and oxidative stress in polar regions.

- **Pigment Production**

Polar microbes produce pigments like carotenoids, which act as antioxidants and protect against harmful UV rays while stabilizing cellular membranes.

- **Symbiotic Relationships**

Some microbes form symbiotic relationships with other organisms, such as algae or lichen, to access additional nutrients and protection.

- **Adaptation to Freeze-Thaw Cycles**

These organisms develop resilience to repeated freezing and thawing, allowing them to survive dynamic temperature fluctuations.

**Ecological Roles of Polar Microbes**

- Form the base of food webs (e.g., algae and cyanobacteria).
- Influence oceanic carbon sequestration.
- Degradation of organic matter in polar soils and water.
- Methane cycling in permafrost and subglacial lakes.

**Applications of Polar Microbiomes**

- **Industrial Enzymes**:
    - Cold-active enzymes for detergents and food processing.

- **Pharmaceuticals**:
    - Antifreeze proteins for organ transplantation.
- **Astrobiology**:
    - Analog for life on icy worlds (e.g., Europa, Enceladus).
- Insights into ancient Earth environments.

**Impact of Climate Change on Polar Microbiomes**

- Melting microbial primary production.
- Thawing permafrost releases greenhouse gases ($CO_2$, $CH_4$).
- Changes in microbial community composition.
- Potential for novel pathogens to emerge from thawed permafrost.

**Conclusion**

- Polar microbiomes are not just survivors but key contributors to Earth's systems.
- Research offers opportunities for biotechnology, climate science, and beyond.
- Urgent need to study and preserve these ecosystems amidst rapid climate change.
- Polar microbiomes provide a glimpse into the adaptability of life in extreme environments.

**Polar Microbiomes: How Do We Study Microorganisms?**

How Do We Study Microorganisms?

Microorganisms are vital to life on Earth, playing critical roles in ecosystems, human health, and biotechnology. To understand their diversity, functions, and interactions, scientists employ various methods. These approaches can be broadly categorized into **culture-based techniques** and **molecular-based techniques**. This lecture explores the principles, applications, and limitations of these methods and discusses how integrative approaches enhance our understanding of microbial life.

Culture-Based Approaches

Culture-based methods involve isolating and growing microorganisms in controlled laboratory conditions. This traditional approach remains essential for studying microbial physiology, metabolism, and interactions.

1. **Isolation and Culturing**
    - Microorganisms are grown on selective or differential media to isolate specific species.
    - Examples: Nutrient agar for general microbes, Sabouraud agar for fungi.
2. **Colony Morphology Observations**

- Characteristics such as colony size, shape, color, and texture provide initial clues about microbial identity.
- Example: Streptococcus species form small, round colonies, while fungi often produce fuzzy colonies.

3. **Biochemical Tests**
   - Analyze metabolic capabilities, such as carbohydrate fermentation, enzymatic activity, or gas production.
   - Examples: Catalase test, citrate utilization, or sugar fermentation tests.
4. **Antibiotic Susceptibility Testing**
   - Determines resistance or sensitivity to various antibiotics.
   - Used for clinical diagnostics and understanding resistance mechanisms.

*Applications:*

- Identification and classification of microorganisms.
- Studying microbial growth and metabolism.
- Testing the efficacy of antibiotics.
- Biotechnological applications, such as fermentation.

*Limitations:*

- **Low Cultivability:** Only a small fraction (estimated <1%) of microbes are culturable using standard laboratory techniques.
- **Environmental Bias:** Laboratory conditions often fail to replicate natural environments, leading to selective growth.
- **Time-Consuming:** Requires days to weeks for growth and analysis.

Molecular-Based Approaches

Molecular techniques have revolutionized microbiology by allowing scientists to study microorganisms without the need for cultivation. These methods rely on analyzing the genetic, transcriptomic, and proteomic information of microbial communities.

*Key Techniques:*

1. **DNA Sequencing**
   - Identifies microbial species based on genetic markers such as the 16S rRNA gene.
   - Example: Sanger sequencing or Next-Generation Sequencing (NGS).
2. **Metagenomics**

- Analyzes the collective genetic material of entire microbial communities directly from environmental samples.
- Applications: Exploring microbial diversity, identifying novel genes, and understanding community functions.

3. **Polymerase Chain Reaction (PCR)**
   - Amplifies specific DNA sequences for the identification and detection of microorganisms.
   - Variants include quantitative PCR (qPCR) for measuring gene abundance and reverse transcription PCR (RT-PCR) for studying RNA.
4. **Fluorescence In Situ Hybridization (FISH)**
   - Uses fluorescent probes to target specific DNA or RNA sequences, enabling visualization of microbial species within their natural environments.
5. **Transcriptomics and Proteomics**
   - Transcriptomics examines RNA to study gene expression and microbial activity.
   - Proteomics analyzes the protein profile of microbes, offering insights into functional roles and adaptations.
6. **Next-Generation Sequencing (NGS)**
   - High-throughput sequencing that provides comprehensive insights into microbial diversity, phylogeny, and functional potential.

*Applications:*

- Understanding microbial diversity, including unculturable species.
- Exploring functional pathways and metabolic potential.
- Identifying pathogenic microorganisms in clinical diagnostics.
- Studying microbial responses to environmental changes.

*Limitations:*

- **Cost and Equipment:** Advanced molecular techniques require expensive equipment and reagents.
- **Complex Data Analysis:** Interpreting large datasets demands bioinformatics expertise.
- **Activity Assessment:** DNA-based methods may not differentiate between live and dormant/dead microbes.

Integrative Approaches

Modern microbiology increasingly combines culture-based and molecular techniques to gain a more holistic understanding of microorganisms.

*Key Strategies:*

1. **Single-Cell Genomics**
    - Combines isolation of individual cells with molecular techniques to study unculturable microbes.
2. **Meta-Omics Approaches**
    - Integrates metagenomics, metatranscriptomics, and metaproteomics to link microbial diversity with function.
3. **Stable Isotope Probing (SIP)**
    - Tracks the incorporation of isotopically labeled substrates into microbial DNA, RNA, or proteins to study active populations.
4. **Functional Culturing**
    - Optimizes culture conditions based on molecular insights, improving the cultivation of previously unculturable microbes.
5. **Environmental Modeling**
    - Simulates natural environmental conditions in the laboratory to study microbial behavior and interactions.

*Benefits of Integration:*

- Captures a more complete picture of microbial communities.
- Links microbial identity with ecological functions and activity.
- Enhances the ability to study rare and unculturable organisms.

**Polar Microbiomes: Case Studies from the Arctic**

**Vertical Stratification of microbial communities in the Arctic Ocean:**

- Microbial diversity increases with depth in Arctic Ocean waters.
- Surface layers: Dominated by oligotrophic microbes.
- Deeper layers: Chemoautotrophs and mixotrophs play key roles.

**Environmental Context**: The study examined environmental interactions in the Arctic Ocean during the summer sea ice melting period, highlighting the effects of climate change.

**Microbial Community Stratification**: Clear vertical stratification of bacterial and archaeal communities observed across different water depths, influenced by distinct water masses (surface mixed layer, Atlantic water, and deep Arctic water).

**Bacterial Diversity**: Detection of 34 bacterial phyla in seawater and 10 in melt-ponds; Proteobacteria dominate seawater, while Bacteroidota are prevalent in melt-ponds.

**Microbial Dispersion**: Source tracking analysis reveals limited microbial dispersion between water layers; surface water contributes 21% of microbes to the deep chlorophyll maximum (DCM), while Atlantic water mass contributes 37% to deep Arctic water.

**Environmental Drivers**: Chlorophyll and ammonium shape surface microbial communities, while inorganic nutrient concentrations influence deep-water communities.

**Distribution of Surface-layer Prokaryotes in the Western Arctic Ocean: Responses to Pacific Water Inflow and Sea Ice Melting**

**Study Focus**: Investigated microbial community composition and functions along the Pacific water inflow path from the Bering Sea, through the Chukchi Sea, to the central Arctic Ocean.

**Key Environmental Drivers**: Pacific water inflow and sea ice melting significantly shape environmental conditions in the western Arctic Ocean, influencing prokaryotic community distribution patterns.

**Microbial Community Patterns**:

Heterotrophic populations dominated in nutrient-rich regions like the Bering Sea and Southern Chukchi Sea, relying on phytoplankton.

Chemoautotrophic bacteria and archaea thrived alongside heterotrophs in the oligotrophic central Arctic Ocean.

**Functional Insights**: While no universal functional genes were enriched across metagenome libraries, the relative abundance of functional genes varied across oceanic sectors, reflecting adaptation to local conditions.

**Ecological Implications**: Deterministic processes primarily govern prokaryotic community assembly, and increasing Pacific inflow and sea ice melting may favor fast-growing heterotrophic populations over native chemoautotrophic and oligotrophic taxa.

**Archaea in the Arctic Ocean**

**Significance of Archaea in Marine Ecosystems:** Archaea represent a significant fraction of marine microbial biomass and play critical roles in ocean biogeochemistry, yet their ecology in the Arctic Ocean remains understudied.

**Study Overview:** Investigated the distribution of archaea and the archaeal ammonia monooxygenase (amoA) gene in the western Arctic Ocean across a depth gradient from surface waters to 3040 m using amplicon sequencing.

**Key Findings:** Identified five archaeal phyla, with Ca. Thermoplasmatota (66.8%) dominant in surface waters and Nitrososphaerota (55.9%) dominating deeper waters, showing depth-dependent vertical segregation of communities.

**Ammonia Monooxygenase (amoA) Gene Insights:** 99% of amoA gene OTUs belonged to Nitrosopumilales, clustered into five subclades, with "NP-Epsilon" dominant throughout and "NP-Alpha" more abundant in deeper waters.

**Ecological Implications:** Vertical segregation of archaea is driven by salinity and inorganic nutrients, potentially influencing the partitioning of dark carbon fixation and nitrification in deeper waters versus organic matter degradation in surface waters.

**Microbiomes in the Arctic Sea Ice Melt Ponds:**

- **Role of Melt Ponds in the Arctic**
    - Formed from thawing snow and sea ice, covering a significant surface during summer.
    - Impact heat fluxes, ice-albedo feedback, and sea-ice energy balance.
- **Melt Ponds as Unique Microbial Ecosystems**
    - Microbial production begins immediately upon formation.
    - Serve as habitats for diverse prokaryotic communities.
- **Methodology**
    - Combined culture-dependent and -independent approaches.
    - 16S rRNA gene amplicon sequencing identified 14 bacterial phyla (146 genera) and 2 archaeal phyla.
    - Pure cultures isolated 24 bacterial genera.
- **Key Findings**
    - Dominant genera (e.g., *Flavobacterium*, *SAR11 clade*, *Polaribacter*).
    - Salinity, chlorophyll a, and dissolved organic carbon drive community distribution.
    - Community composition shifts with salinity changes during pond evolution.
- **Significance**
    - Melt pond microbial communities differ significantly from surface ocean communities.

- Highlight the role of dissolved organic carbon from primary production in oligotrophic waters.

**Distinct bacterial community structures sea ice types in the Central Arctic Ocean**

**Study Overview**: Analyzed bacterial community composition and metabolic functions from three distinct sea ice floes: the North Pole, western Amundsen Basin, and Nansen Basin, using long-read amplicon sequencing and metagenomics.

**Bacterial Diversity**: Identified 10 unique bacterial phyla, 114 genera, and 62 species, with *Bacteroidota*, *Proteobacteria*, and *Cyanobacteria* prevalent across all sites.

**Environmental Drivers**: Nitrate, nitrite, silicate concentrations, ice age, and origin were key factors shaping bacterial community structures.

**Metabolic Functions**:

- Prevalence of sulfur cycling (assimilatory and dissimilatory sulfur metabolism) and complex carbon degradation processes in sea ice.
- Heterotrophic bacteria dominated, contributing to nutrient recycling and organic matter degradation during the summer melt season.

**Novel Findings**: Discovered novel ecological roles of *Aquiluna* genomes through phylogenomic and pangenomic analysis, providing insights into their contributions to sea ice ecosystems.

**Key Takeaway**: Distinct bacterial community structures and ecological functions are strongly correlated with sea ice type, origin, and habitat characteristics in the Central Arctic Ocean.

**Microbiomes in Glacial Forelands:**

- Early colonization dominated by chemolithotrophic bacteria.
- Older sites show a shift towards nitrogen and carbon cycling.
- Discovery of *Candidatus ARS69*, a novel microbial phylum.

**Glacier Forelands as Natural Laboratories**: Expanding ice-free glacier forelands (GFs) provide a unique opportunity to study microbial succession, soil formation, and ecosystem development due to global glacier retreat.

**Gene-Centric and Genome-Resolved Metagenomics**: Comprehensive analysis of microbial diversity, community structure, and biogeochemical processes related to carbon, nitrogen, and sulfur cycling across three GF ecosystems.

**Metagenome Assembled Genomes (MAGs)**: Reconstruction of 903 MAGs (899 bacterial, 4 archaeal) from 27 shotgun metagenomic datasets collected from GFs in Svalbard, Greenland, and Sweden.

**Novel Microbial Lineages**: Taxonomic classification revealed 98% of MAGs were unclassified at the species level, highlighting the presence of previously unknown microbial lineages.

**Functional and Phylogenetic Insights**: MAGs showed close phylogenetic relationships across GFs but significant variations in abundance, distribution, and metabolic functions, underscoring their role in biogeochemical cycles.

**Challenges in Polar Microbiome Research and Future Directions**

**Research Challenges**

- Limited long-term datasets for microbial dynamics.
- Need for international collaboration in polar research.
- Leveraging **omics technologies** (e.g., transcriptomics, proteomics).
- Improved satellite and remote sensing for ecosystem monitoring.

**Logistical Challenges:**

1. Extreme environments and harsh weather.
2. High costs of polar expeditions.

**Analytical Challenges:**

1. Limited culturable diversity (<5%).
2. Data complexity from molecular approaches.

**Key Priorities:**

1. Link microbial ecology with climate models to predict ecosystem responses.
2. Develop sustainable research methodologies.
3. Explore novel microbial species and their potential applications.
4. Promote interdisciplinary collaborations to advance understanding and conservation

# Microalgae Biotechnology and Sustainable Aquaculture

Dr Preetha K Shenoy

AlgaGen Biosolutions LLP ,www.algagenbio.in,

## Introduction

Microalgae, a diverse group of microscopic photosynthetic organisms, play a pivotal role in aquatic ecosystems as primary producers and contributors to biogeochemical cycles. They are indispensable in aquaculture due to their ability to serve as live feed, enhance water quality, and sequester carbon. Microalgae biotechnology is increasingly recognized as a vital component in promoting sustainability in aquaculture. With the rising global demand for seafood and the environmental challenges posed by traditional aquaculture practices, microalgae offer a promising alternative. Their rich nutritional profile, ability to enhance water quality, and potential for reducing reliance on fishmeal, make them an attractive solution. It is promising to use microalgae biotechnology to support an eco-friendly, carbon neutral, and sustainable aquaculture.

### 1. Microalgae: general overview

Microbial organisms of size 2-200 $\mu$ m, which can photosynthesize and produce oxygen are generally termed microalgae. Microalgae are commonly found in aquatic habitats - planktonic or benthic, however there are terrestrial forms - epilithic, epiphytic, aerial or symbiotic (e.g. Lichens). This diverse group can live in extreme climatic conditions like salt pans, acidic lakes, hot springs and on ice, too.

Taxonomically, the prokaryotic forms are called Cyanobacteria or Blue green algae (Class Cyanophyceae) specifically for their blue-green colour due to phycocyanin pigment. Eukaryotic forms are green, brown, red coloured and classified as Chlorophyceae (green), Trebouxiophyceae (green), Bacillariophyceae (diatoms, brown), Prymnesiophyceae (golden), Rhodophyceae (red), Dinophyceae (dinoflagellates, brown), Prasinophyceae (green flagellates), Eustigmatophyceae (yellow green), etc.

### 2. Role of microalgae

Ecological: Microalgae are the unsung heroes of our planet's ecosystem, producing over 60% of the world's oxygen and sequestering significant amounts of $CO_2$. As primary producers in larger water bodies, they form the base of the food chain, supporting aquatic life and purifying water bodies by assimilating inorganic wastes. However, excessive growth of toxic microalgae

can lead to harmful algal blooms (HAB), eutrophication, toxicity, and mass mortality of aquatic life, highlighting the need to balance microalgae growth and mitigate HABs to maintain healthy aquatic ecosystems.

Commercial: Microalgae are treasure troves of bioactive compounds, offering a rich source of primary and secondary metabolites. These versatile microorganisms have vast commercial applications:

- Food and feed supplements
- Bio-fuels and energy solutions
- Organic fertilizers
- Nutraceuticals and functional foods
- Cosmeceuticals and skincare products
- Pharmaceuticals and therapeutic agents

Embracing microalgal biotechnology enables a seamless transition to a circular economy, where waste is transformed into wealth. This innovative approach promotes sustainable, eco-friendly, and carbon-neutral solutions, paving the way for a greener future.

### 3. Microalgae in Aquaculture

a) Nutritional source:

Microalgae are nutrient-dense organisms, rich in proteins, polyunsaturated fatty acids (PUFAs), vitamins, and pigments. With protein, oil, and carbohydrate contents reaching up to 60%, 70%, and 60% respectively, they offer a superior alternative to traditional fishmeal and fish oil in aquafeeds. The amino acid profile of microalgae often surpasses that of conventional feeds, promoting enhanced growth and health in aquatic organisms. Furthermore, microalgae's rich nutrient profile supports the development of robust immune systems in fish and shellfish, reducing the need for antibiotics and promoting eco-friendly aquaculture practices. Microalgae are versatile feed sources, used in live, processed, or formulated forms for zooplankton, larvae, and juvenile stages of fish and shellfish.

b) Water Quality Improvement

Microalgae play a vital role in maintaining optimal water quality in aquaculture systems. Through photosynthesis, they absorb excess nutrients like nitrogen and phosphorus, stabilize pH levels, and mitigate issues like eutrophication. Additionally, microalgae modify the bacterial community profile, working synergistic with their phycosphere bacteria to prevent water deterioration and recover nutrients from aquaculture effluents. Their ability to sequester $CO_2$ and produce $O_2$ also helps alleviate challenges posed by low dissolved oxygen levels in

extensive aquaculture systems, promoting a healthier and more sustainable aquatic environment. Furthermore, microalgae's capacity to absorb heavy metals and other pollutants makes them valuable in bioremediation efforts, enhancing water quality and the overall sustainability of aquaculture practices.

c) Sustainability

Microalgae are a beacon of sustainability in aquaculture, offering numerous eco-friendly benefits. They have an exceptional ability to capture $CO_2$ more efficiently than other plants, significantly reducing the carbon footprint of aquaculture operations. Microalgae-based feeds provide a sustainable solution by reducing production costs and feed waste due to their high digestibility and superior nutritional value. They also minimize the use of wild-caught fish and fishmeal, help conserve marine ecosystems, and provide a consistent and reliable source of nutrients. Furthermore, microalgae support circular economy principles by utilizing waste $CO_2$ and converting it into valuable biomass, promoting eco-friendly aquaculture practices that prioritize environmental stewardship and long-term sustainability.

d) Harmful algal blooms (HAB)

Harmful algal blooms (HAB) pose a significant threat to aquaculture systems. Excessive growth of toxic microalgae, such as cyanobacteria and dinoflagellates, can directly harm the health and quality of farmed fish and shellfish. Even non-toxic blooms can cause damage by depleting oxygen levels and damaging fish gills. Effective management of HAB is crucial, and control methods include physical, biological, and chemical interventions to mitigate the negative impacts on aquaculture operations.

4. **Microalgae - Biotechnological interventions:**

a) Microalgae Diversity and DNA barcoding;

Microalgae encompass prokaryotic cyanobacteria e.g., *Arthrospira* (Spirulina) and eukaryotic species like *Chlorella*, *Dunaliella* and *Nannochloropsis*. The complex microscopic cellular make-up and lack of expertise make morphological taxonomic delineation tedious and challenging. DNA barcoding involves sequencing specific regions of the genome, such as the 18S/16S rDNA and internal transcribed spacer (ITS) regions, which are commonly used for differentiating microalgal species. Other genes used in barcoding of microalgae are the 28S/23S rDNA, *rbc*L region (Rubisco gene), Cytochrome oxidase I (*CO* I) gene, etc. These genetic markers offer high sensitivity and specificity, allowing for accurate identification even among closely related species[1].

b) Next-Generation Sequencing (NGS):

The advent of NGS technologies has revolutionized the genomic analysis of microalgae[2]. This technology can be used to identify microalgae species and their abundance in a sample. NGS can also help researchers understand how environmental factors affect microalgae diversity. For instance, studies on *Chlorococcum* sp. FFG039 utilized NGS to uncover unique gene families associated with specific traits like biofilm formation and floating behavior. This approach enables comprehensive genome characterization, facilitating comparative genomics across different microalgal species. Genome sequencing can help in uncovering metabolic pathways for example, pathways for biofuel and nutraceutical development of *Chlorella vulgaris* and *Nannochloropsis gaditana* has uncovered by using the technology and further helped in metabolic engineering.

c) Comparative Genomics and multiple approaches:

Comparative genomics allows researchers to analyze the genetic similarities and differences between various microalgal species. For example, the genomic study of *Picochlorum* isolates revealed how specific genetic adaptations contribute to environmental resilience, highlighting the role of horizontal gene transfer (HGT) in microalgal evolution. In multiple polyphasic approaches more than one gene or genomic study of a particular species will be compared with their morpho-physiological features.

d) Genetic engineering:

- CRISPR/Cas9 and Other Gene Editing Tools[3]: The application of CRISPR/Cas9 and other gene editing tools has revolutionized microalgal genetic engineering, enabling precise and efficient modifications CRISPR-Cas9 / genome editing and mutagenesis for improving lipid production, stress tolerance, and biomass yield. CRISPR-Cas9 technology has enormous promises, and it has been employed to alter the metabolism of many algae and cyanobacteria.
- Design of Microalgal Expression Vectors: Researchers have made notable progress in designing expression vectors specifically tailored for microalgae, enabling more efficient genetic engineering.
- Discovery of Genetic Regulatory Elements: The identification of genetic regulatory elements, such as promoters and transcription factors, has improved our understanding of microalgal gene regulation and facilitated targeted genetic modifications.
- Optimization of Transformation Methods: Scientists have developed more efficient transformation methods, allowing for the introduction of desirable traits into microalgae with greater precision and success rates.

- New Strain Improvement Techniques: Novel strain improvement techniques have been developed, enabling researchers to tailor microalgae for specific applications, such as biofuel production, nutritional supplements, and bioremediation.

e) Metabolic Engineering:

Recent advancements in metabolic engineering of microalgae have focused on over-expressing or knocking out key genes that encode enzymes in pathways of interest, such as fatty acid and isoprenoid biosynthesis. This has led to significant improvements in the production of various bioproducts, including lipids, carotenoids, and other high-value compounds.

5. **History of microalgae biotechnology**

- Early Days (1960s-1980s)

a) Isolation and characterization: Researchers began isolating and characterizing microalgae strains for their potential in food, feed, and bioactive compound production.

b) Photobioreactors: The first photobioreactors were developed to cultivate microalgae for research and industrial applications.

- Evolution of Biotechnology in microalgae

First Wave (1990s-2000s)

a) Genetic engineering: Scientists began exploring genetic engineering techniques to improve microalgae strains for biofuel, nutritional supplements, and other applications.

b) Omics technologies: The advent of genomics, transcriptomics, and proteomics enabled researchers to better understand microalgae biology and identify targets for improvement.

Second Wave (2010s-present)

c) Next-generation sequencing: Advances in sequencing technologies have enabled rapid and cost-effective analysis of microalgae genomes, transcriptomes, and metabolomes.

d) CRISPR-Cas9: The development of CRISPR-Cas9 gene editing has revolutionized microalgae biotechnology, enabling precise and efficient genome editing.

e) Synthetic biology: Researchers are now designing and constructing new biological pathways and circuits in microalgae to produce novel compounds and improve existing processes.

f) Systems biology: Integrative approaches combining experimental and computational methods are being used to understand and optimize microalgae biology.

**6. Current Trends and Future Directions**

a) Precision biology: The use of CRISPR-Cas9 and other precision editing tools to develop microalgae strains with specific traits.

b) Synthetic genomics: The design and construction of new microalgae genomes with improved characteristics.

c) Microalgae-based biomanufacturing: The use of microalgae as cell factories for the production of biofuels, nutritional supplements, and other valuable compounds.

d) Wound healing: Live hydrogels embedded with microalgae offer a novel and promising approach to diabetic wound healing, where along with therapeutic compounds, microalgae can deliver soluble oxygen promoting accelerated wound healing and tissue repair in diabetic patients[4].

e) Integration with emerging technologies: Combining microalgae biotechnology with emerging technologies like artificial intelligence (AI), machine learning (ML), and the Internet of Things (IoT) to create more efficient and sustainable processes.

f) Quantum Dots[5] and Nanotechnology: Quantum dots enable real-time bioimaging of microalgal growth, while nanoparticles enhance nutrient uptake and photosynthetic efficiency.

g) Nitrogen Doping: Nitrogen-doped carbon materials improve light absorption and CO2 fixation, boosting growth under suboptimal conditions.

h) Omics Approaches: Genomics, proteomics, and metabolomics identify high-performing strains and metabolic pathways for targeted applications.

i) Artificial Intelligence (AI): AI-driven models optimize cultivation conditions, improving yield and reducing costs.

j) Liquid Trees[6]: Indoor or outdoor microalgae production systems or photobioreactors with an aim for better air quality, especially in urban cities.

k) Photobioreactor Innovations: Nanomaterial coatings and IoT-enabled sensors enhance light distribution, monitor culture conditions, and optimize productivity.

## 7. **Conclusion**

Microalgae biotechnology represents a promising avenue for achieving sustainability in aquaculture through innovative genetic approaches, chassis development, encapsulation techniques, and advanced technologies like CRISPR. By optimizing microalgal strains for improved nutritional profiles while utilizing advanced methods to protect and deliver bioactive compounds effectively, these innovations contribute significantly to enhanced feed quality, aquatic health, and overall sustainability in aquaculture practices. Continued research will be

essential to unlock the full potential of microalgae in supporting global food security while minimizing environmental impacts.

**References**

Takei-Idiaquez, D. H., Yupanqui-Morales, F. M., Chavez-Alberto, A. D., Ulloa-Osorio, A., Díaz-Pillasca, H. B., Ramírez-Viena, L., & Nerida, A. (2024). A panoramic review of DNA barcoding in microalgae: applications and challenge in the urgency of its use in Peru.

O'Neill, E. A., Fehrenbach, G., Murphy, E., Alencar, S. A., Pogue, R., & Rowan, N. J. (2022). Use of next generation sequencing and bioinformatics for profiling freshwater eukaryotic microalgae in a novel peatland integrated multi-trophic aquaculture (IMTA) system: Case study from the Republic of Ireland. Science of the Total Environment, 851, 158392.

Jeon, S., Lim, J. M., Lee, H. G., Shin, S. E., Kang, N. K., Park, Y. I., ... & Chang, Y. K. (2017). Current status and perspectives of genome editing technology for microalgae. Biotechnology for biofuels, 10, 1-18.

Kang, Y., Xu, L., Dong, J. et al. Programmed microalgae-gel promotes chronic wound healing in diabetes. Nat Commun 15, 1042 (2024). https://doi.org/10.1038/s41467-024-45101-9

Agnol, L. D., Neves, R. M., Maraschin, M., Moura, S., Ornaghi Jr, H. L., Dias, F. T. G., & Bianchi, O. (2021). Green synthesis of Spirulina-based carbon dots for stimulating agricultural plant growth. Sustainable Materials and Technologies, 30, e00347.

Shaji George. (2023). Liquifying Urban Lungs: Assessing the Air Purification Potential of Photobioreactor "Liquid Trees" in Highly Polluted Cities. Partners Universal Innovative Research Publication, 1(2), 1–14. https://doi.org/10.5281/zenodo.10111964

# Bacterial Transformation and its applications

Anusree V Nair*, Sumithra TG, Krupesha Sharma SR.

Marine biotechnology fish nutrition and health division., ICAR-CMFRI, Kochi-18, *anuadithya2004@gmail.com

## Introduction

Bacterial transformation is a cornerstone of genetic engineering, biotechnology, and genetic research, with significant implications for various fields. The method through which bacteria take foreign single-stranded DNA from their surroundings and incorporate it into their genome is known as bacterial transformation. The ability of bacteria to perform transformation is known as competence. This process is important in molecular biology because it makes the bacterium acquire new genetic traits, which can be advantageous for various applications. Bacterial transformation is the inherent capacity of bacteria to release DNA which is subsequently absorbed by another competent bacterium. The success of the transformation relies on the host cell competence.

## Natural transformation

Bacterial transformation idea was initially illustrated by Frederick Griffith in 1928. His investigations with *Streptococcus pneumoniae* showed that non-virulent bacteria might become pathogenic by absorbing the DNA of dead virulent strains. This landmark discovery paved the way for identifying DNA as the hereditary material by Avery, MacLeod, and McCarty in 1944. Natural transformation has been demonstrated in approximately 80 different bacterial species. (Johnston et al. 2014). Research has revealed that while phylogenetically distant species has conserved proteins involved in absorption and processing, however, they vary in the signals that induce transformation and the governing mechanisms that govern it. DNA internalization and chromosomal integration are the major steps of natural bacterial transformation. The adopted DNA into a cell is mostly involved in generating genetic diversity or repairing chromosomes. Naturally transformable organisms spontaneously release their DNA *via* autolysis. Johnston et al. (2014) made a comprehensive review to identify the diverse and similar principles which regulate the natural transformation procedure in various microbes. It is now widely recognized that transformation is an effective mechanism of horizontal gene transfer in bacterial populations, leading to the dissemination of several ecologically significant

activities in bacterial populations, such as antimicrobial resistance, heavy metal resistance, *etc*.

**Artificial transformation**

Artificial transformation involves laboratory techniques to introduce foreign DNA into bacterial cells. Competence can be artificially induced by using chemicals like calcium and applying various methods such as electroporation or heat shock. These techniques increase the permeability of the bacterial cell wall, making it more competent to absorb donor DNA. Transformation is considered physicochemical rather than simply chemical or electrical (physical) as it involves manipulating the cells with cations and by altering cell wall through temperature or electrical conditions to enable DNA uptake (Tantray et al., 2023). Once transformed, bacteria can be selected based on the presence of a marker in the inserted DNA, such as antibiotic resistance or the ability to utilize a specific amino acid. In naturally competent bacteria, the free DNA usually attaches to the host cell wall and incorporates into the chromosomal DNA. However, in some cases, the DNA is incorporated into a plasmid or fosmid (a hybrid structure that combines features of plasmids and bacteriophages) which can replicate independently from the chromosome, allowing the inserted DNA to remain outside the chromosome. The plasmid or fosmid used to carry foreign DNA into a host cell is referred to as a vector, which is subsequently introduced into the host cell through transformation. The factors, including the size of the insert, copy number, compatibility, selective markers, cloning sites, and any specialized functions the vector determine the vector to be used for a particular purpose (Preston, 2003). The recipient cell that successfully proliferates the cloned plasmid or fosmid with the foreign DNA is called the transformant. The plasmid or fosmid typically encodes enzymes and antibiotic resistance markers, which are expressed in the transformant. Once the transformation is complete, transformants can be identified by growing the cells on media supplemented with a specific antibiotic.

**Major vectors for transformation**

**Plasmids:** Plasmids are small, circular, double-stranded extrachromosomal DNA that are commonly found in many bacteria. They are the most common vectors used for cloning. They can carry genes that provide the host bacteria with a survival advantage, such as antibiotic resistance genes, but they can also be engineered to carry foreign DNA. Plasmids are widely used in bacterial transformation because they replicate independently of the bacterial chromosome, allowing for the stable propagation of introduced genes in large quantities (high copy number). Further, they can carry diverse genes, making them useful for versatile

applications, from protein production to gene editing. Cloning with plasmids is now a well-established and standardized technique in molecular biology, with many commercial plasmid vectors available for different purposes. Further, the transformation with plasmids is relatively cost-effective, especially when using *E. coli* as host for cloning and amplification. However, there is a size limit to the foreign DNA which may be cloned into a plasmid, typically a few 15 kb. Further, in some cases, the inserted gene may not be efficiently expressed in the host due to codon usage differences between species, so codon optimization may be necessary for efficient protein production. Furthermore, over time, plasmids may become unstable in certain host cells, especially if the plasmid carries large or toxic inserts. This can affect the consistency of gene expression.

**Features of vector plasmids**

- **Origin of replication (Ori)**: Vector plasmids contain an **origin of replication**, which allows them to replicate independently within a host cell, ensuring the DNA inserted into the plasmid is also copied and propagated.
- **Selectable markers**: Vector plasmids contain genes that confer resistance to certain antibiotics (mostly ampicillin or kanamycin). These markers allow for the selection of bacterial cells that have successfully taken up the plasmid, as only those cells will survive in the presence of the corresponding antibiotic.
- **Multiple Cloning Site (MCS)**: The **MCS** is a short region containing multiple restriction enzyme sites that allow easy insertion of the gene of interest. It is a key feature that simplifies the cloning process.
- **Reporter Genes**: Some plasmids may contain reporter genes, such as '**Green Fluorescent Protein'** or **'lacZ',** which can be used to track successful cloning events.

**Fosmids:** Fosmids combine the properties of plasmids and bacteriophages. Fosmids can carry larger DNA inserts (up to 50 kb) compared to standard plasmids, making them useful for cloning larger pieces of DNA. Fosmids are lower in copy numbers compared to plasmids. This makes them stable and well-suited to checking the quality of sequences generated from genome projects. Fosmids are particularly useful for cloning large genomic fragments, making them a valuable tool in genomic research, sequencing projects, and functional studies minimizing the risk of losing important genetic information

**Features of fosmids**

- Fosmids can accommodate DNA fragments typically between **30 to 45 kb.** This allows for the cloning of larger genes or more complex genomic regions, which are often needed for functional genomics and genome sequencing projects.
- A fosmid has **the F-factor origin of replication.** The F-factor is a natural plasmid in *E. coli* that facilitates the **high-efficiency replication and stable maintenance** of large DNA inserts.
- Fosmids typically contain **antibiotic resistance markers**, which allow for the selection of transformed bacteria.
- The vector includes an **MCS** for inserting DNA fragments and other elements such as **cosmid-like packaging sites** for DNA packaging into phage particles.
- Fosmids replicate independently of the bacterial chromosome.
- Fosmids are usually maintained at a **low copy number** in bacterial cells (about 1-2 copies per cell). This reduces the potential for recombination/genetic rearrangements and inserts loss/ instability that can occur with large inserts in high-copy-number plasmids on long-term culture.
- Fosmids can be packaged into **lambda phage particles** for efficient and easy delivery into bacterial cells, which is advantageous when working with large DNA fragments.
- The low copy number and large insert size reduce the likelihood of insert rearrangements and cloning bias, allowing for a more accurate representation of the original DNA fragment.
- Fosmid cloning is typically done in *E. coli* **strains** engineered for high transformation efficiency, such as *E. coli* DH10B or *E. coli* Stbl2.
- Since fosmids are usually present in low copy numbers, the transformation process can be less efficient compared to high-copy plasmids, and fewer colonies may be obtained after transformation.
- **Cloning Complexity**: The process of cloning with fosmids is slightly more complex than using traditional plasmids due to the larger insert size and the necessity of careful handling to avoid recombination or insert loss.

**Methodological perspectives**

**Materials required**

- Competent bacterial cells (*E. coli* strains vary based on the vector used. DH5α and EPI300 are examples of plasmid and fosmid-compatible strains, respectively)
- The vector containing the gene of interest or genomic DNA fragment

- Calcium chloride ($CaCl_2$) solution or electroporation buffer for electroporation method
- LB (Luria-Bertani) broth or SOC medium (for recovery)
- Selective LB agar plates with suitable antibiotics based on the vector
- Ice, water bath, and incubator
- Microcentrifuge tubes (1.5 mL)
- Sterile spreader or loop for plating

**Preparation of competent cells**

Competent cells are bacterial cells that have been treated to become permeable to foreign DNA, allowing them to take up plasmids, fosmids, or other DNA molecules during transformation. Selection of the appropriate bacterial strain **is t**he first step for transformation. The chosen strain should ideally be in a log phase of growth or a mid-exponential phase. Accordingly, the selected bacterial cells from the glycerol stock are subcultured in the LB agar followed by incubation overnight at 37°C (Sambrook and Russell, 2003). A single colony from the plate is then subcultured into LB broth (5 ml) and incubated at 37°C in the shaking incubator at 200 rpm for 14 h. Subsequently, around 200 µl of the culture is used to inoculate 20 ml of LB broth in a 100 ml flask, which is incubated in an orbital shaker at 37°C at 200 rpm for 3 h, until the culture reached an OD of 0.3 to 0.4 at 600 nm (mid-exponential phase), which is ideal for competence. The culture is then shifted to a pre-cooled 50 ml centrifuge tube and kept on ice for 1 h, followed by centrifugation at 3000 rpm for 10 min at 4°C. The culture is kept on ice to slow down bacterial metabolism. The culture pellet is then resuspended in cold, sterile $CaCl_2$ buffer to make the cells competent. This buffer helps in making the bacterial cell wall more permeable to DNA. After washing multiple times in the buffer the cells are resuspended in a small volume of ice-cold 0.1 M $CaCl_2$ with glycerol by mild tapping and retained in ice for 1 h. The competent cells are then aliquoted into 100 µl portions and placed in pre-cooled 1.5 ml microcentrifuge tubes, then stored at –80°C till use.

**Ligation of insert into the vector**

The gene or DNA fragment to be cloned is obtained by **PCR amplification or restriction enzyme digestion** of genomic DNA. The vector and the DNA fragment to be cloned are mixed in a **ligation reaction**. The enzyme **DNA ligase** catalyzes the establishment of phosphodiester bonds between the vector and the inserted gene, creating a **recombinant plasmid/fosmid.** The ligation reaction typically contains the vector (digested and dephosphorylated, if necessary), insert DNA (digested with compatible restriction enzymes), DNA ligase enzyme and buffer

containing ATP (required for ligase activity). Ligation is usually performed at 4°C for 18h or 16°C for 1-2 h to allow efficient ligation. While using the plasmid it's important to ensure the DNA is high quality and not degraded. Around 1 to 5 µg of fosmid DNA is required for ligation and to ensure the fosmid DNA is free from contaminants like salts, phenol, or RNA, as these can affect bacterial growth.

**Transformation of host cells**

**Heat shock or electroporation is** used to introduce the recombinant vector into bacterial cells via **transformation.** E**lectroporation** involves exposing the bacterial cells and vector DNA to an electric field, which creates transient pores in the bacterial membrane to facilitate DNA uptake. It requires an electroporator and specific cuvettes. Electroporation is particularly used for bacterial transformation with fosmids when higher transformation efficiency is required or when dealing with difficult-to-transform bacterial strains. Here, after mixing the competent cells with the fosmid DNA in an electroporation cuvette (usually 0.1 cm gap cuvette) a high-voltage pulse using an electroporator (typically 1.8-2.0 kV for *E. coli*) is applied. Immediately 900 µL of SOC medium or LB broth is added to the cuvette to recover the cells. SOC medium is preferred for better recovery and higher transformation efficiency. In the heat shock method, the cells and recombinant vector are exposed to a heat shock at 42°C for 90 sec and immediate cooling on ice for 10 min. After transformation, the bacterial cells are added with 900 µl of fresh LB or SOC medium and incubated at 37°C for 1 h in a shaker. Followed by the plating of transformed cells onto LB agar plates containing the specific antibiotic depending on the vector used. Chloramphenicol is commonly used for fosmids while ampicillin or kanamycin is commonly used for plasmids. Only the bacteria that have successfully taken up the vector (which contains the antibiotic resistance gene) will survive. If the vector contains a **reporter gene** or a **colourimetric marker (**such as **lacZ),** blue/white screening can be used to identify colonies with successful insertions. The plates are incubated for 12-14 h at 37°C for the recombinant colonies to develop. Since fosmids are large plasmids, they may take longer to propagate than smaller plasmids, so extended incubation is sometimes required.

**Screening**

After incubation, bacterial colonies are formed on the agar plate. The colonies that appear are picked and transferred to fresh LB broth containing the appropriate antibiotic for further growth and analysis. **Colony PCR using** primers that are specific to the gene or vector sequences is carried out to confirm the presence of desired gene of the insert. Further, the recombinant vector

is **isolated** from the bacteria using a **kit** (miniprep) and **the presence of insert** is confirmed by sequencing or performing additional analytical tests like, restriction digestion or PCR.

**Applications of bacterial transformation**

- Cloning and recombinant protein production: Scientists use plasmids or fosmids to clone genes of interest and express them in bacteria. This technique is fundamental for producing proteins such as insulin, growth hormones, enzymes, and vaccines.
- Genetic Studies: Transformation enables researchers to introduce specific mutations or reporter genes into bacterial cells, facilitating the study of gene function and regulation.
- Bioremediation: Engineered bacteria can be designed to degrade environmental pollutants, such as oil spills or heavy metals, through transformed metabolic pathways.
- Synthetic Biology: Transformation allows for the assembly of complex genetic circuits, enabling the design of bacteria with novel properties for industrial, medical, or environmental applications.

**Conclusion**

Bacterial transformation using plasmids and fosmids are versatile and powerful tools for cloning foreign DNA fragments, offering key advantages in genomic research, functional genomics, and gene discovery. It allows researchers to introduce foreign genes into bacteria for protein production, gene expression studies, or cloning. The procedure requires careful handling of competent cells, plasmid DNA, and the proper use of antibiotics for selecting successful transformants. Although there are limitations related to cloning efficiency and insert size, plasmids and fosmids remain one of the best cloning vectors for large-scale DNA cloning and sequencing applications. While bacterial transformation offers immense scientific and technological benefits, it raises certain ethical and ecological concerns. The release of genetically modified organisms into the environs may lead to unintended consequences, such as horizontal gene transfers to native microbial populations or the spread of antibiotic resistance genes. Thus, stringent regulatory frameworks and containment strategies are essential to mitigate potential risks. Briefly, bacterial transformation is a fundamental biological process with extensive applications in research, medicine, and industry. Its discovery marked a turning point in molecular biology, allowing scientists to work on genetic material with precision. As technology advances, bacterial transformation may continue to play a critical role in addressing

global challenges, from disease treatment to environmental sustainability, provided its use is guided by ethical considerations and responsible stewardship.

**References**

Johnston C, Martin B, Fichant G, Polard P, Claverys JP. Bacterial transformation: distribution, shared mechanisms and divergent control. Nature Reviews Microbiology. 2014 Mar;12(3):181-96.

Preston A. Choosing a cloning vector. E. coli Plasmid Vectors: Methods and Applications. 2003:19-26.

Tantray JA, Nissa NU, Wani RF, Shafi SM. Basic Life Science Methods: A Laboratory Manual for Students and Researchers. Elsevier; 2022 Sep 11.

Sambrook, J., Russell, D.W., 2003. Molecular cloning-A laboratory manual. 3rd ed. Cold spring harbor laboratory press, New York. 999 p.

# Microbiome manipulation: A promising strategy for improving the larval survival

Gayathri Suresh*, Sumithra TG, Krupesha Sharma SR, Ambarish P Gop
Marine biotechnology fish nutrition and health division., ICAR-CMFRI, Kochi-18,
*gayathrisureshmenon@gmail.com

**Introduction**

The larval stage is a crucial phase in the life cycle of aquatic organisms, particularly in aquaculture species. During this stage, larvae are highly vulnerable to environmental factors, malnutrition, and microbial imbalances, making low larval survival rates a significant challenge for productivity in aquaculture. In some species, mortality during the larval stage can reach as high as 70 to 90% due to poor water quality, environmental stress, disease outbreaks, and inadequate immune system development (Vadstein et al., 2018). Therefore, improving larval survival is critical for addressing the increasing worldwide demand for seeds to meet the increasing aquaculture goods. Addressing this critical challenge in aquaculture requires innovative strategies, particularly focusing on nutrition and health at the larval stages, where microbiome research holds promise.

**Microbiome concept**

The 'microbiome' is the community of microbes, including bacteria, fungi, and viruses occupying a specific environment. In animals, the term is used to designate the microbes occurring in specific body parts like skin, gill, or gut. The microbiome is highly dynamic and changes based on environmental factors, feed, and husbandry practices. The recognition that symbiotic microbes are integral factors in their host health and physiology leads to a new era of microbiome research in biology. Advances made in human microbiome research have revealed that microbiome manipulations and microbiome-based therapies can make innovative and eco-friendly ways to improve host and ecosystem health. Apart from human health, microbiome is an emerging paradigm for improving agriculture, aquaculture, livestock sectors, and ecosystem management sectors (coral protection and bioremediation of polluted areas).

**Fish larval microbiome**

The development and well-being of larvae are significantly influenced by the microbiome. These microbial communities colonize the larval gut, gills, skin, and surrounding environment, influencing digestion, nutrient absorption, and immunity. When the microbiome is balanced, it

performs a range of critical services, including the suppression of harmful pathogens, host development, immunity, support of metabolic processes, and the modulation of various physiological responses in the host (Ringø, 2020). However, when dysbiosis, a disruption in the balanced state of these microbial communities, occurs, it can lead to significant negative consequences. This imbalance can increase the larvae's susceptibility to infections and diminish their growth performance, underscoring the importance of maintaining healthy microbial communities in larval culture systems. Thus, careful management of these microorganisms is essential for promoting optimal larval health and development. Microbiome manipulation is becoming a promising approach to enhance larval survival and improve overall efficiency in aquaculture. Strategies such as using probiotics, prebiotics, and synbiotics are being investigated to boost beneficial microbial populations, while phage therapy and environmental modifications are employed to control harmful microbes. Advances in molecular tools and sequencing technologies have further enhanced our understanding of microbial dynamics, allowing for targeted interventions to optimize larval health. Therefore, prioritizing microbiome research is imperative for advancing sustainable fish farming practices. This chapter explores the critical role of microbiomes in larval survival and discusses innovative strategies for manipulating these communities to promote sustainable aquaculture.

**Composition and dynamics of larval microbiomes**

The larval microbiome is pivotal for the health and development of aquatic species, predominantly comprising bacteria from the phyla *Proteobacteria*, *Bacteroidota*, *Firmicutes*, and *Actinomycetes* (Bakke et al., 2015; Paralika et al., 2023; Sumithra et al., 2022, 2024; Wilkes Walburn et al., 2019). These bacteria are vital for nutrient digestion, the enhancement of immune function, and the provision of disease resistance. During the initial stages of development, microbial communities exhibit dynamic characteristics and are significantly influenced by environmental factors and dietary selections. Gaining a clear understanding of their composition, dynamics, and the factors that influence them is essential for improving larval health and survival in aquaculture.

Studies revealed that when fish developed from fertilized eggs to early larval stages, bacterial richness and diversity declined. However, the bacterial diversity in fish larvae tends to rise progressively with age until it peaks at the juvenile stage, regardless of the host species, rearing strategy, or diet regime (Borges et al., 2021). Research indicated that the early colonization of the fish gastrointestinal tract occurs primarily through the uptake of bacteria associated with feed and/or from waterborne bacteria. Li et al. (2017) established that the microbiome of gibel

carp (*Carassius auratus gibelio*) remained consistent among larvae fed the same diet. They observed the most pronounced changes in the microbiome during the transition from Artemia to a formulated pelleted feed. Furthermore, Sumithra et al. (2024) showed that dietary factors contribute more significantly to the overall gut microbiota of larvae of *Trachinotus blochii* compared to water.

The composition of microorganisms colonizing the larvae is also influenced by environmental conditions, including temperature and salinity, as well as the trophic level and dietary preferences of the fish species. For instance, turbot (*Scophthalmus maximus*) reared at 15°C exhibited reduced gut microbial richness and diversity compared to those maintained at 20°C (Guerreiro et al., 2016). In Atlantic salmon, temperature influenced both the α- and β-diversity of gut microbiota (Nguyen et al., 2020) and the prevalence of *Vibrio* and *Pseudomonas* species (Hatje et al., 2014). Additionally, elevated temperatures resulted in a reduction of lactic acid bacteria in the faeces of Atlantic salmon (Neuman et al., 2018), while yellowtail kingfish (*Seriola lalandi*) displayed microbial shifts that were temperature-dependent (Horlick et al., 2020). Salinity significantly affects the gut microbiota of fish, especially in euryhaline species, by shaping microbial community composition and triggering salinity stress responses. For example, *Trachinotus ovatus* exhibited distinct microbiota at varying salinities of 5, 15, 25, and 35‰ (Liu et al., 2019), while *Salvelinus alpinus* displayed microbial turnover associated with changes in salinity (Hamilton et al., 2019). In Atlantic salmon, fish in seawater exhibited lower α-diversity, characterized by 55.6% *Firmicutes* dominance (Dehler et al., 2017), while another study reported higher α-diversity with over 80% dominance of *Firmicutes* (Rudi et al., 2018). As larvae grow, their microbiome stabilizes, with specific bacterial taxa adapting to the distinctive conditions characteristic of the host's digestive system. Furthermore, host-specific factors, including genetic and physiological traits, significantly influence microbiome composition, resulting in distinct microbial profiles appearing at various developmental stages. Acknowledging the importance of these microbial communities is essential; understanding their dynamics may facilitate targeted interventions. Such interventions can enhance growth, bolster immunity, and improve survival rates within aquaculture systems.

**Host-microbiome interactions: insights into resilience and larval health**

The microbiome of larval organisms is crucial for their health and development, acting as a cornerstone of their physiological well-being. Beneficial microbes in the larvae form vital symbiotic relationships and perform several key functions, such as nutrient cycling and immune modulation. These microorganisms excel at breaking down complex organic matter

into simpler, more absorbable compounds, significantly enhancing larvae's nutritional intake. Researchers used Tax4Fun to show that early larval stages of yellowtail kingfish had more prominent secondary metabolic pathways than late stages, indicating stronger antibiotic synthesising capacity for early stages. On the other hand, the pathways for the metabolism of carbohydrates and thiamine were more prevalent at later stages (maybe as a result of the formulated meals' higher carbohydrate content compared to live feeds), and they need more thiamine as a cofactor in the metabolism of carbohydrates (Wilkes Walburn et al., 2019). Moreover, beneficial microbes play a pivotal role in shaping the larval immune response by interacting with immune cells, strengthening their defences against pathogens in constantly changing and often challenging environments.

Aquaculture productivity is increasingly threatened by infectious diseases, opportunistic pathogens, and microbial imbalances. The "pathobiome" concept underscores the interplay between host-associated commensals, environmental microbes, and the host in determining health outcomes. These interactions redefine diseases as complex ecological phenomena influenced by microbial communities rather than isolated pathogens. Stress-induced dysbiosis, where beneficial microbes are lost and pathogenic ones proliferate, is a significant challenge in aquaculture. For instance, in brook charr (*Salvelinus fontinalis*), stress decreases beneficial commensals like *Methylobacterium* and *Propionibacterium*, which produce antimicrobial compounds, while encouraging opportunistic pathogens like *Aeromonas* (Boutin et al., 2013). Similarly, environmental shifts such as salinity changes in Atlantic salmon farming disrupt microbial homeostasis, creating a transitional dysbiotic phase before a new balance is achieved. Defining a universally "healthy" microbiome is challenging due to variations across species and farming systems. Instead, a desirable microbiome adapts to environmental changes while maintaining functional stability. Such plasticity can enhance resilience to stressors like dietary shifts, temperature fluctuations, or salinity changes, reducing disease susceptibility. Addressing the challenges due to diseases and stress involves promoting microbial diversity, using probiotics, and implementing microbiome conditioning strategies. Non-invasive monitoring of microbial markers in water, mucus, or faeces enables early detection of dysbiosis, allowing timely interventions to restore microbial balance. Selective breeding programs enhancing host resilience and microbial diversity can further contribute to sustainable aquaculture practices. Future research should focus on understanding microbiome functionality and resilience, leveraging microbial diversity to combat diseases, and developing tools for real-

time microbiome monitoring. By harnessing the power of microbial dynamics, aquaculture systems can achieve improved health, productivity, and sustainability.

**Microbiome manipulation strategies to improve fish larvae survival**

Microbiome manipulation strategies have demonstrated significant potential in enhancing the health, survival, and growth of fish larvae. These approaches aim to promote beneficial microbial communities while controlling harmful pathogens, ultimately improving the development and resilience of larvae in aquaculture systems. Key strategies for microbiome manipulation include probiotics, prebiotics, paraprobiotics, synbiotics, phage therapy, and environmental modulation.

**Probiotics:** Probiotics are live microorganisms that provide health benefits to the host, including fish larvae, when administered in adequate amounts. Choosing the right probiotic strains is crucial for maximizing benefits. Key selection criteria include the ability to effectively colonize the gut, resistance to environmental stressors like salinity and temperature, and overall safety in aquatic habitats. There is an existing agreement that giving probiotics has several advantageous effects on fish, such as higher growth performance, feed utilization, antioxidant capacity, immune responses, and pathogen resistance through competitive exclusion or by producing antimicrobial compounds. The intestinal microbiome of Nile tilapia larvae varied considerably after the administration of probiotic-supplemented feeds (Xia et al., 2020). The research implied that probiotics can modify the composition and activity of the larval microbiome; however, the effects will vary depending on the host developmental stage, dosage, length of probiotic supplementation, and the compatibility of the probiotic–host system. Accordingly, continued research into host-specific applications promises more efficient and sustainable practices in fish farming. By integrating probiotics into aquaculture, we can foster healthier fish and enhance farming outcomes while promoting disease prevention through balanced microbial communities and improved fish immunity.

**Prebiotics:** Prebiotics are powerful non-digestible compounds that significantly enhance the growth of beneficial microorganisms in the fish gut, leading to improved microbial diversity and gut health. Well-known prebiotics, including fructooligosaccharides and mannan-oligosaccharides (MOS), not only improve feed efficiency and growth rates but also effectively reduce pathogen colonization (Yu et al., 2022). MOS, in particular, offers added protection by binding to harmful bacteria, thereby preventing their attachment to the gut epithelium (Mata-Sotres et al., 2016). Furthermore, prebiotics play a crucial role in stabilizing gut microbiota,

alleviating stress-induced disruptions, and enhancing overall disease resistance. Accordingly, future research focusing on the prebiotics for larval microbiome is warranted.

**Synbiotics:** Synbiotics refer to the combined use of probiotics and prebiotics to create a synergistic effect on the host's microbiome. Research shows that synbiotics, which combine probiotics like *Saccharomyces cerevisiae* with prebiotics such as mushroom extract, offer superior health benefits for zebrafish, including enhanced growth and immune responses (Hosseini et al., 2024). Tailoring such formulations to specific fish larvae is essential for maximizing effectiveness, given the influence of environmental factors.

**Bioencapsulation:** Bioencapsulation involves exposing live feed organisms to a suspension of probiotics or prebiotics. This allows the organisms to ingest and retain beneficial compounds, which are transferred to larvae when they consume the live feed. This method effectively delivers health-promoting agents through the larvae's natural feeding process. Research has demonstrated that larvae fed bioencapsulated live feed show improved survival rates, enhanced growth, and greater disease resistance compared to those fed with non-enriched diets. Loh and Ting (2015) investigated the bioencapsulation of *Lactococcus lactis* subspecies lactis in *Artemia franciscana nauplii* using sodium alginate. Bioencapsulation allowed the larvae to ingest beneficial compounds through live feed, improving survival, growth, and disease resistance.

**Phage therapy**: Bacteriophages are a powerful alternative to antibiotics for effectively managing bacterial infections in aquaculture. Phage therapy targets harmful bacteria, especially in fish larvae, while safeguarding beneficial microbes. Studies showed that VP-2 phage effectively inactivated *Vibrio anguillarum* and improved fish larvae survival without adverse effects (Silva et al., 2014). By specifically targeting harmful bacteria while protecting beneficial microbes, maintaining the natural microbial balance and decisively combating the threat of antibiotic-resistant bacteria, phages can contribute to healthier larvae and improved survival rates, ensuring a sustainable future for aquaculture.

**Microbial Consortia:** The application of microbial consortia (carefully selected combinations of various beneficial microbial species) offers another promising strategy to enhance the survival and health of fish larvae. Microbial consortia provide distinct advantages over single-species probiotics by promoting diverse and stable microbial communities in larval guts. This diversity enhances immunity, digestion, and pathogen resistance. Fish larvae often have underdeveloped immune systems, making them susceptible to infections. Introducing microbial consortia can significantly enhance their innate immune responses. These beneficial

microbes stimulate gut-associated lymphoid tissue (GALT), leading to increased immune cell production and improved gut barrier function. This results in greater infection resistance and better overall growth. Additionally, microbial consortia strengthen the stability of the larval microbiome, helping it withstand disturbances like pathogens and environmental stressors. Even though probiotic consortia administration appears to be beneficial in certain situations, it is evident that further research is still needed. Future trials of this strategy should evaluate the stability, compatibility, and environmental competency of the microbes of the probiotic consortia to improve the success rates (Borges et al. 2021).

**Environmental microbiome manipulation**: Manipulating the environmental microbiome to enhance fish larvae survival involves modifying the microbial composition of water to promote beneficial microorganisms and reduce harmful pathogens, thereby improving larval health and viability. The advantages encompass reduced pathogen burden, enhanced immune responses, and improved water quality, supporting larval development and survival. Nevertheless, challenges such as maintaining microbial equilibrium, ensuring long-term stability, and preventing pathogen resistance must be addressed. In summary, environmental microbiome manipulation is a promising approach for enhancing aquaculture outcomes by cultivating a healthy microbial environment that minimizes pathogens, strengthens immune function, and promotes overall larval well-being.

**Paraprobiotics:** Paraprobiotics are inactivated or non-viable microbial cells that confer a health benefit to the consumer (Siciliano et al., 2021). Notably, common concerns for probiotic usages, like cell viability in feed products, gut colonization efficiency, acquisition of virulence or antibiotic resistance genes due to horizontal gene transfer, and shelf-life do not apply to paraprobiotics. Simultaneously, paraprobiotics have the same positive benefits as their live probiotic counterparts, are safer for immunocompromised hosts, and are subject to less stringent regulatory regulations (Borges et al. 2021). Thanh Tung et al. (2010) illustrated the ability of a paraprobiotics-incorporated diet to enhance the survival rates, growth performance, and ability to withstand stress (only for postlarvae). The impact of paraprobiotics on host health most likely stems from a complex web of interactions between the host immune system, metabolome, and microbiome, eventually improving disease resistance. Borges et al. (2021) hypothesised that this network can be partially modulated by higher heterotrophic feeding on these newly available resources, which needs further verification.

**Challenges and limitations**

Larval microbiome manipulation presents significant potential for enhancing larval health and survival through the promotion of growth and increased disease resistance. However, a key concern is the introduction of non-native microorganisms may alter the natural microbiome and cause ecological imbalances or weakened host immunity. Another issue is the possibility of changing the microbial environment due to probiotics may make it possible for infections to adapt and outcompete helpful microbes. This problem calls for a well-rounded strategy incorporating probiotics with other tactics like immunization and environmental control. Furthermore, microbiome-based therapies in commercial aquaculture continue to face major cost and scalability challenges. The cost of producing, maintaining, and stabilizing probiotics and prebiotics in aquatic environments can be high, and the cost is further increased by controlled delivery methods such as encapsulation. Infrastructure investment is required for broad adoption. Finally, addressing the risks associated with non-native microorganisms, disease resistance, and cost-effectiveness is critical to realize the full potential of larval microbiome modification in aquaculture.

**Future Perspectives**

Modification of microbiomes in larval aquaculture has a bright future ahead, since new technologies and continuous study will create more effective and sustainable methods. It will be possible to develop a complete strategy that enhances water quality management, reduces the usage of antibiotics, and improves larvae health while minimizing environmental effects by combining microbiome modification with other sustainable aquaculture practices. For example, a balanced microbial habitat will be promoted while promoting sustainable growth if microbiome modification is combined with plant-based or locally derived food, as well as techniques like biofiltration and recirculating aquaculture systems. Using artificial intelligence (AI) is also revolutionary because it can analyse large datasets to predict microbial community dynamics and optimize interventions based on factors like water quality and host health. AI might make it possible to track microbial populations in real-time and modify interventions as necessary to preserve the best parameters for larval survival. However, regulatory and ethical issues need to be carefully considered as microbiome modification in aquaculture grows. To ensure food safety and the environment, precise recommendations for adding probiotics, prebiotics, and other microbiome-modulating agents need to be developed. As these technologies advance, aquaculture could undergo a revolution in sustainability, efficiency, and

environmental responsibility through the use of microbiome modification, artificial intelligence, and responsible regulation.

## Conclusion

The health, growth, and survival of aquatic species are greatly influenced by the balance of the larval microbiome, which has a direct impact on immune responses, nutrient absorption, and disease resistance. Our understanding of microbial community dynamics, composition, and functional roles in larval aquaculture systems has improved, allowing for more targeted manipulations. In addition to environmental microbiome alterations, strategies including probiotics, prebiotics, synbiotics, paraprobiotics, and phage therapy have shown great promise in promoting beneficial microbial communities, reducing dysbiosis, and enhancing overall larval performance. Nonetheless, there are still issues, such as species-specific variability, environmental effects, and the requirement for economical, sustainable solutions that are suited to aquaculture systems. Future studies should concentrate on microbial consortia optimization, host-microbiome interactions, and real-time monitoring methods for microbiome analysis to guarantee the health and resilience of fish larvae. By combining microbiome-based strategies with conventional aquaculture methods, aquaculture industries can become more sustainable, productive, and long-term viable.

## References

Bakke, I., Coward, E., Andersen, T., & Vadstein, O. (2015). Selection in the host structures the microbiota associated with developing cod larvae ( G adus morhua ). *Environmental Microbiology*, *17*(10), 3914–3924. https://doi.org/10.1111/1462-2920.12888

Borges, N., Keller-Costa, T., Sanches-Fernandes, G. M. M., Louvado, A., Gomes, N. C. M., & Costa, R. (2021). Bacteriome Structure, Function, and Probiotics in Fish Larviculture: The Good, the Bad, and the Gaps. *Annual Review of Animal Biosciences*, *9*(1), 423–452. https://doi.org/10.1146/annurev-animal-062920-113114

Boutin, S., Bernatchez, L., Audet, C., & Derôme, N. (2013). Network Analysis Highlights Complex Interactions between Pathogen, Host and Commensal Microbiota. *PLoS ONE*, *8*(12), e84772. https://doi.org/10.1371/journal.pone.0084772

Dehler, C. E., Secombes, C. J., & Martin, S. A. M. (2017). Seawater transfer alters the intestinal

microbiota profiles of Atlantic salmon (Salmo salar L.). *Scientific Reports*, *7*(1), 13877. https://doi.org/10.1038/s41598-017-13249-8

Guerreiro, I., Enes, P., Rodiles, A., Merrifield, D., & Oliva-Teles, A. (2016). Effects of rearing temperature and dietary short-chain fructooligosaccharides supplementation on allochthonous gut microbiota, digestive enzymes activities and intestine health of turbot (Scophthalmus maximus L.) juveniles. *Aquaculture Nutrition*, *22*(3), 631–642. https://doi.org/10.1111/anu.12277

Hamilton, E. F., Element, G., van Coeverden de Groot, P., Engel, K., Neufeld, J. D., Shah, V., & Walker, V. K. (2019). Anadromous Arctic Char Microbiomes: Bioprospecting in the High Arctic. *Frontiers in Bioengineering and Biotechnology*, *7*. https://doi.org/10.3389/fbioe.2019.00032

Hatje, E., Neuman, C., Stevenson, H., Bowman, J. P., & Katouli, M. (2014). Population Dynamics of Vibrio and Pseudomonas Species Isolated from Farmed Tasmanian Atlantic Salmon (Salmo salar L.): A Seasonal Study. *Microbial Ecology*, *68*(4), 679–687. https://doi.org/10.1007/s00248-014-0462-x

Horlick, J., Booth, M. A., & Tetu, S. G. (2020). Alternative dietary protein and water temperature influence the skin and gut microbial communities of yellowtail kingfish ( Seriola lalandi ). *PeerJ*, *8*, e8705. https://doi.org/10.7717/peerj.8705

Hosseini, S. S., Sudaagar, M., Zakariaee, H., Paknejad, H., Baruah, K., & Norouzitalab, P. (2024). Evaluation of the synbiotic effects of Saccharomyces cerevisiae and mushroom extract on the growth performance, digestive enzyme activity, and immune status of zebrafish danio rerio. *BMC Microbiology*, *24*(1), 331. https://doi.org/10.1186/s12866-024-03459-2

Li, X., Zhou, L., Yu, Y., Ni, J., Xu, W., & Yan, Q. (2017). Composition of Gut Microbiota in the Gibel Carp (Carassius auratus gibelio) Varies with Host Development. *Microbial Ecology*, *74*(1), 239–249. https://doi.org/10.1007/s00248-016-0924-4

Liu, B., Guo, H., Zhu, K., Guo, L., Liu, B., Zhang, N., Jiang, S., & Zhang, D. (2019). Salinity effect on intestinal microbiota in golden pompano trachinotus ovatus (Linnaeus, 1758). *Israeli Journal of Aquaculture - Bamidgeh*, *71*(January). https://doi.org/10.46989/001c.20995

Loh, J. Y., & Ting, A. S. Y. (2015). Bioencapsulation of probiotic Lactococcus lactis subsp. lactis on Artemia franciscana nauplii: Effects of encapsulation media on Nauplii survival and probiotic recovery. *Malaysian Journal of Microbiology*. https://doi.org/10.21161/mjm.12314

Mata-Sotres, J. A., Moyano, F. J., Martínez-Rodríguez, G., & Yúfera, M. (2016). Daily rhythms of digestive enzyme activity and gene expression in gilthead seabream (Sparus aurata) during ontogeny. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *197*, 43–51. https://doi.org/10.1016/j.cbpa.2016.03.010

Neuman, C., Hatje, E., Smullen, R., Bowman, J. P., & Katouli, M. (2018). The effect of fishmeal inclusion and prebiotic supplementation on the hindgut faecal microbiota of farmed Tasmanian Atlantic salmon ( Salmo salar L.). *Journal of Applied Microbiology*, *125*(4), 952–963. https://doi.org/10.1111/jam.13930

Nguyen, C. D. H., Amoroso, G., Ventura, T., Minich, J. J., & Elizur, A. (2020). Atlantic Salmon (Salmo salar L., 1758) Gut Microbiota Profile Correlates with Flesh Pigmentation: Cause or Effect? *Marine Biotechnology*, *22*(6), 786–804. https://doi.org/10.1007/s10126-019-09939-1

Paralika, V., Kokou, F., Karapanagiotis, S., & Makridis, P. (2023). Characterization of Host-Associated Microbiota and Isolation of Antagonistic Bacteria from Greater Amberjack (Seriola dumerili, Risso, 1810) Larvae. *Microorganisms*, *11*(8), 1889. https://doi.org/10.3390/microorganisms11081889

Ringø, E. (2020). Probiotics in shellfish aquaculture. *Aquaculture and Fisheries*, *5*(1), 1–27. https://doi.org/10.1016/j.aaf.2019.12.001

Rudi, K., Angell, I. L., Pope, P. B., Vik, J. O., Sandve, S. R., & Snipen, L.-G. (2018). Stable Core Gut Microbiota across the Freshwater-to-Saltwater Transition for Farmed Atlantic Salmon. *Applied and Environmental Microbiology*, *84*(2). https://doi.org/10.1128/AEM.01974-17

Siciliano, R. A., Reale, A., Mazzeo, M. F., Morandi, S., Silvetti, T., & Brasca, M. (2021). Paraprobiotics: A New Perspective for Functional Foods and Nutraceuticals. *Nutrients*, *13*(4), 1225. https://doi.org/10.3390/nu13041225

Silva, Y. J., Costa, L., Pereira, C., Mateus, C., Cunha, Â., Calado, R., Gomes, N. C. M., Pardo,

M. A., Hernandez, I., & Almeida, A. (2014). Phage therapy as an approach to prevent Vibrio anguillarum infections in fish larvae production. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0114197

Sumithra, T. G., Sharma, K. S. R., Gangadharan, S., Suresh, G., Prasad, V., Amala, P. V., Sayooj, P., Gop, A. P., Anil, M. K., Patil, P. K., & Achamveetil, G. (2022). Dysbiosis and Restoration Dynamics of the Gut Microbiome Following Therapeutic Exposure to Florfenicol in Snubnose Pompano (Trachinotus blochii) to Aid in Sustainable Aquaculture Production Strategies. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.881275

Sumithra, T. G., Sharma, S. R. K., Suresh, G., Gop, A. P., Surya, S., Gomathi, P., Anil, M. K., Sajina, K. A., Reshma, K. J., Ebeneezar, S., Narasimapallavan, I., & Gopalakrishnan, A. (2024). Mechanistic insights into the early life stage microbiota of silver pompano (Trachinotus blochii). *Frontiers in Microbiology*, *15*. https://doi.org/10.3389/fmicb.2024.1356828

Thanh Tung, H., Koshio, S., Ferdinand Traifalgar, R., Ishikawa, M., & Yokoyama, S. (2010). Effects of Dietary Heat-killed Lactobacillus plantarum on Larval and Post-larval Kuruma Shrimp, Marsupenaeus japonicus Bate. *Journal of the World Aquaculture Society*, *41*(s1), 16–27. https://doi.org/10.1111/j.1749-7345.2009.00329.x

Vadstein, O., Attramadal, K. J. K., Bakke, I., Forberg, T., Olsen, Y., Verdegem, M., Giatsis, C., Skjermo, J., Aasen, I. M., Gatesoupe, F.-J., Dierckens, K., Sorgeloos, P., & Bossier, P. (2018). Managing the Microbial Community of Marine Fish Larvae: A Holistic Perspective for Larviculture. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.01820

Wilkes Walburn, J., Wemheuer, B., Thomas, T., Copeland, E., O'Connor, W., Booth, M., Fielder, S., & Egan, S. (2019). Diet and diet-associated bacteria shape early microbiome development in Yellowtail Kingfish ( Seriola lalandi ). *Microbial Biotechnology*, *12*(2), 275–288. https://doi.org/10.1111/1751-7915.13323

Xia, Y., Wang, M., Gao, F., Lu, M., & Chen, G. (2020). Effects of dietary probiotic supplementation on the growth, gut health and disease resistance of juvenile Nile tilapia (Oreochromis niloticus). *Animal Nutrition*, *6*(1), 69–79. https://doi.org/10.1016/j.aninu.2019.07.002

Yu, W., Yang, Y., Zhou, Q., Huang, X., Huang, Z., Li, T., Wu, Q., Zhou, C., Ma, Z., & Lin,

H. (2022). Effects of dietary Astragalus polysaccharides on growth, health and resistance to Vibrio harveyi of Lates calcarifer. *International Journal of Biological Macromolecules*, *207*, 850–858. https://doi.org/10.1016/j.ijbiomac.2022.03.176

# Decoding lipid metabolism in Teleost: Omics insights into the Impact of Novel Microbe Derived Ingredients

Adnan H. Gora, Saima Rehman, Chandrasekar S., D. Linga Prabu, Sanal Ebeneezar and Qurat ul Ain, & Sajina K.A.

Marine biotechnology fish nutrition and health division., ICAR-CMFRI, Kochi-18,

**Abstract**

Animal models, particularly teleosts, have emerged as indispensable tools for studying lipid metabolism and its dysregulation. While mammals such as mice and rats have long been the gold standard in metabolic research, zebrafish (*Danio rerio*) have gained increasing recognition as a powerful model for lipidomics and metabolic studies. Their genetic tractability, optical transparency, high fecundity, and physiological similarities to mammals in lipid processing make them an invaluable system for investigating lipid-associated disorders. Recent studies leveraging diet-induced hyperlipidemia models in zebrafish have successfully recapitulated key features of human dyslipidemia, further validating their relevance in metabolic research. Despite species-specific differences in lipoprotein biology and lipid metabolism pathways, zebrafish continue to offer novel insights into lipid-related disease mechanisms and potential therapeutic strategies. This chapter explores the critical role of lipids in cellular structure, energy metabolism, and physiological regulation, emphasizing their impact on metabolic health. It highlights the use of teleost models, particularly zebrafish, in lipid metabolism research, showcasing its relevance in studying diet-induced dyslipidemia and screening potential therapeutic interventions.

***Keywords*****:** Lipid metabolism, Dyslipidemia, Zebrafish, Lipidomics, Transcriptomics

## Background

Lipids constitute a diverse class of hydrophobic molecules, including triglycerides, phospholipids, waxes, and steroids, distinguished by their solubility in organic solvents and insolubility in polar solvents like water (Ahmad et al., 2019). These biomolecules serve essential biological functions, with one of their most fundamental roles being the structural integrity of cell membranes. Comprising a glycerol backbone, two hydrophobic fatty acid chains, and a hydrophilic phosphate group, lipids exhibit amphipathic properties that facilitate the formation of bilayer structures, ensuring cellular stability and selective permeability. Beyond their

architectural function, lipids are indispensable for various physiological processes, including energy storage, hormone synthesis, nerve impulse transmission, organ protection, and nutrient transport. As highly efficient energy reservoirs, triglycerides store more than twice the energy per gram compared to carbohydrates and proteins, playing a critical role in metabolic homeostasis. Lipids also underpin endocrine signaling, with steroid hormones such as testosterone, estrogen, and cortisol regulating growth, metabolism, and immune responses. Additionally, phospholipids and sphingolipids are vital for neurotransmission and brain function, while cholesterol serves as a precursor for vitamin D and bile acids, facilitating lipid digestion and nutrient absorption. Given their central role in metabolism, disruptions in lipid homeostasis are major contributors to metabolic disorders. Dysregulation in lipid synthesis, storage, or degradation can lead to obesity, insulin resistance, and dyslipidemia, significantly elevating the risk of cardiovascular diseases (CVDs), diabetes, and neurodegenerative conditions. CVDs, in particular, remain the leading cause of mortality worldwide, accounting for over 20 million deaths annually, with poor dietary habits further exacerbating the burden—over six million deaths are attributed to diet-induced CVDs alone (Vaduganathan et al., 2022). Modern dietary patterns, characterized by excessive consumption of processed red meat, butter, high-fat dairy products (Hu, 2002), and refined sugars, alongside insufficient intake of soluble and insoluble fibers (Wilson et al., 2016), are key drivers of noncommunicable diseases (Kopp, 2019). Prolonged adherence to such diets increases the likelihood of dyslipidemia (Asadi et al., 2020; Na et al., 2019), a condition marked by imbalances in lipid profiles—including elevated total cholesterol, low-density lipoprotein (LDL), and triacylglycerol levels, coupled with reduced high-density lipoprotein (HDL) cholesterol. These abnormalities accelerate the development of atherosclerosis, a pathological process that significantly heightens the risk of CVDs (Gómez-Delgado et al., 2021; Hedayatnia et al., 2020).

A deep understanding of lipid metabolism and its regulatory pathways is essential for developing therapeutic strategies to mitigate metabolic diseases. Animal models play a crucial role in this research, offering controlled *in vivo* systems where the intricate interplay of genetic, dietary, and environmental factors in lipid regulation can be meticulously examined. These models provide invaluable insights into the mechanisms underlying lipid-associated disorders, paving the way for targeted interventions and novel therapeutic approaches.

**Teleost models of lipid research**

Animal models such as mice and rats share remarkable physiological and genetic similarities with humans, making them indispensable tools for investigating complex metabolic pathways

and evaluating potential therapeutic interventions. These models allow precise experimental manipulations—including dietary modifications, genetic engineering, and targeted drug therapies—facilitating the dissection of fundamental mechanisms underlying lipid-related diseases. Although zebrafish have long been employed in biomedical research, their application in lipid metabolism studies is relatively recent compared to well-established mammalian models. Despite lacking a functional stomach, zebrafish possess a segmented gut—comprising the foregut, midgut, and hindgut—resembling that of mammals (Wallace et al., 2005). The regulation of epithelial cell transcripts across these gut segments is highly conserved with that of mammals (Lickwar et al., 2017), further underscoring their suitability for metabolic studies. Additionally, zebrafish offer several advantages as a research model, including small size, optical transparency, a fully sequenced genome, high fecundity, and a short generation time, making them particularly well-suited for studying metabolic disorders.

One of the earliest studies validating zebrafish as a model for dyslipidemia was conducted by Stoletov et al. (2009). In this pioneering work, adult zebrafish were fed a 4% cholesterol diet for 8 to 12 weeks, leading to elevated total cholesterol and triacylglycerol levels in the blood. The study demonstrated hallmark features of mammalian atherosclerosis, such as lipid accumulation in the vasculature and cholesterol uptake by myeloid cells, confirming the model's relevance to dyslipidemia research. This seminal finding paved the way for the development of diet-induced hyperlipidemia models in zebrafish. Following this, researchers increasingly leveraged the zebrafish model to elucidate the molecular mechanisms of dyslipidemia and to screen for potential therapeutic agents (Baek et al., 2012; Jin & Cho, 2011). The identification of zebrafish orthologs of key lipoprotein metabolism genes, including MTTP, LDLR, SRB1, and LCAT (O'Hare et al., 2014; Marza et al., 2005), has further reinforced the model's significance in lipid research. A key advantage of zebrafish in dyslipidemia studies is the ease with which hyperlipidemic conditions can be induced through dietary interventions, particularly high-fat diets.

Despite its growing utility, the zebrafish model exhibits inherent differences from human lipoprotein biology. For instance, in humans, lipoproteins constitute approximately 10% of total plasma proteins, whereas in teleosts, they account for about 36% (Figure 1). In mammals, the APOB gene encodes two major protein isoforms: APOB-100 and APOB-48. APOB-100 is a 512 kDa protein translated from full-length mRNA, while APOB-48, a truncated 246 kDa version, arises from RNA editing in the intestine. In contrast, zebrafish possess three orthologs of APOB—apoba, apobb.1, and apobb.2—with sequence similarities to human APOB of

54.2%, 43.0%, and 29%, respectively (Otis et al., 2015). Notably, there is no evidence of an APOB-48 protein in zebrafish, likely due to the absence of an RNA-editing enzyme or other necessary factors required to introduce a premature stop codon in the apob pre-mRNA (Conticello et al., 2005; Chester et al., 2003). The absence of APOB-48 may contribute to the rapid onset of dyslipidemia in zebrafish, as APOB-100-like chylomicrons are likely cleared more slowly from circulation. Furthermore, in zebrafish, LDL particles can be derived directly from chylomicrons (Schlegel, 2016).

Biochemically, zebrafish LDL differs considerably from that of humans. In human LDL, cholesterol esters comprise approximately 45–50%, whereas in teleosts, they constitute only about 15%. Conversely, triacylglycerols make up 4–8% of human LDL but approximately 23% in teleosts (Babin & Vernier, 1989). These biochemical differences highlight fundamental distinctions between mammalian and teleost lipid metabolism. Additionally, whole-genome duplication in teleosts has resulted in multiple orthologs of human lipid-related genes. In

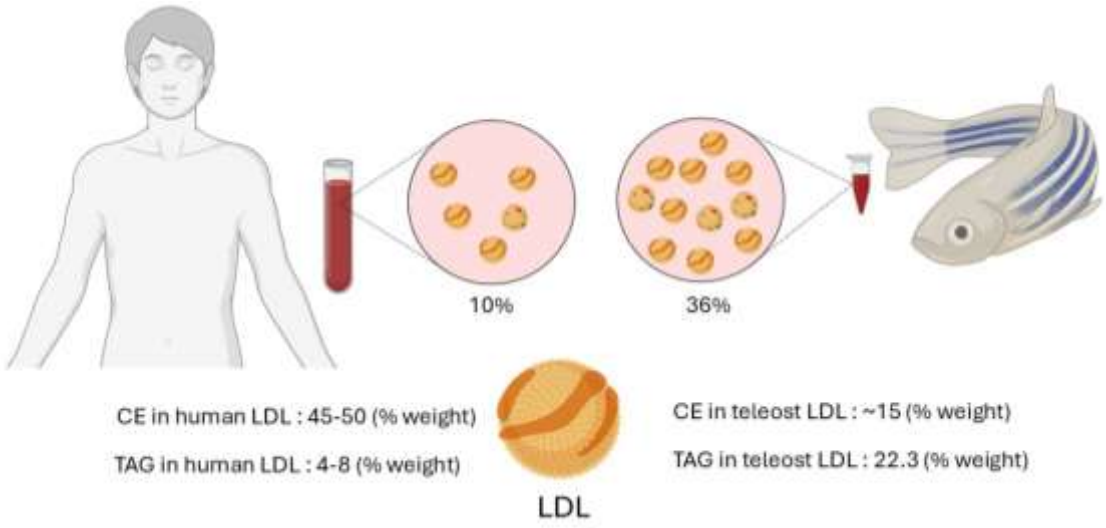



**Figure 1**: Differences in the lipoprotein profile of humans with fishes. The fish lipoprotein are more than three-fold more abundant than humans. The figure is generated based on raw data presented in Babin & Vernier, 1989.

zebrafish, for instance, apoa1a and apoa1b correspond to the human APOA-I gene, while apoea and apoeb are orthologous to APOE. Likewise, apoa4a, apoa4b.1, apoa4b.2, and apoa4b.3 correspond to the human APOA4 gene (Otis et al., 2015).

Despite these differences, zebrafish remain a powerful and versatile model for lipid metabolism and dyslipidemia research. Their unique genetic attributes, ease of experimental manipulation,

and conservation of key metabolic pathways make them an indispensable tool for unraveling lipid-related diseases and identifying novel therapeutic targets. As research continues to refine the zebrafish model, it holds immense promise for advancing our understanding of lipid biology and metabolic disorders.

**Advancements in Omics Technologies for Understanding Lipid Metabolism in Teleosts**

Recent advancements in high-throughput omics technologies—encompassing genomics, transcriptomics, proteomics, metabolomics, and lipidomics—have revolutionized our understanding of lipid metabolism in fish species. These integrative approaches have not only enhanced the utility of model organisms but have also provided profound insights into gene regulation by dietary lipids and the molecular mechanisms underlying lipid utilization and lipid-associated disorders.

Lipidomics, a specialized branch of metabolomics, enables high-throughput screening and quantification of lipid species, revealing both novel and well-characterized lipid molecules within biological systems. This approach has proven instrumental in identifying biomarkers of cardiovascular diseases (CVDs) that extend beyond traditional risk indicators. For instance, from a comprehensive analysis of 184 lipids in human plasma, eight lipid species—including four phosphatidylcholines (PC), one sphingomyelin (SM), one diacylglycerol, one phosphatidylinositol (PI), and one cholesterol ester (CE)—were recognized as predictive biomarkers of CVD (Ottosson et al., 2021).

Similarly, lipidomic profiling of rats subjected to a high-fat diet (HFD) revealed significant alterations in seven fatty acid species, including palmitic acid, hexadecenoic acid, hexanoylcarnitine, tetracosahexaenoic acid, cervonoyl ethanolamide, 3-hydroxytetradecanoic acid, and 5,6-dihydroxy eicosatrienoic acid. Furthermore, five sterols—cholesterol ester (CE 18:2), cholesterol, hydroxytestosterone, 19-hydroxydeoxycorticosterone, and cholic acid—also exhibited marked changes. However, natural bioactive compounds have demonstrated potential in mitigating the adverse effects of HFD. For example, an ethanolic extract (95%) of the medicinal mushroom *Poria cocos* was found to restore lipid levels disrupted by HFD (Miao et al., 2016). Additionally, *Chrysanthemum morifolium* and its active flavonoid, luteolin, were shown to counteract HFD-induced imbalances in specific lipid classes, reinstating the concentrations of lysophosphatidylcholines (LPC 18:1, LPC 22:6), phosphatidylcholine (PC 18:1/22:6), cholesterol ester (CE 22:6), ceramide (CER 34:1 (d18:1/16:0)), and sphingomyelin (SM 34:1 (d18:1/16:0)) (Shon et al., 2020). These findings highlight the potential of lipidomic

analysis in elucidating the mechanisms of action of natural bioactive compounds in lipid homeostasis.

The zebrafish model has emerged as a powerful tool for lipidomic studies, particularly in the context of metabolic disorders. A recent plasma lipidomic analysis conducted by Gora et al. (2023) identified 331 distinct lipid species spanning 11 lipid classes, shedding light on the impact of microbial additives in managing dyslipidemia (Gora et al., 2023a, 2023b). In zebrafish liver, glycerophospholipids constitute the predominant lipid class, followed by fatty acyls and glycerolipids (Zhang et al., 2024). Additionally, researchers have demonstrated the protective effects of β-sitosterol against HFD-induced disruptions in triglyceride homeostasis, highlighting its potential therapeutic role in metabolic disorders. To investigate the direct effects of obesity on kidney metabolism, an overfeeding model in zebrafish was employed, followed by untargeted metabolomic analysis of kidney tissues. This study identified 235 significantly altered metabolites between overfed and control groups. Pathway analysis of these metabolites revealed a strong link between diet-induced obesity and disruptions in glycolysis and tryptophan metabolism—both of which are implicated in the pathogenesis of obesity-related kidney disease (Zeitler et al., 2023). Further metabolomic investigations have provided new insights into sex-specific metabolic responses in diabetic zebrafish. Benchoula et al. (2024) explored the differential metabolic profiles of male and female diabetic zebrafish following treatment with limonene, a natural compound with promising anti-diabetic and anti-obesity effects. Using HFD-fed zebrafish as a model for type 2 diabetes (T2D), the study identified several endogenous metabolites that could serve as potential diagnostic biomarkers for diabetes. Interestingly, despite both sexes exhibiting hyperglycemia and elevated BMI, their endogenous metabolic profiles differed significantly. Limonene treatment effectively normalized blood glucose levels and mitigated diabesity in zebrafish by reversing HFD-induced metabolic disruptions in both genders. Notably, limonene also downregulated the aberrant expression of AKT, a key player in insulin signaling, further reinforcing its therapeutic potential in metabolic disorders. The integration of omics technologies, particularly lipidomics, has significantly enhanced our understanding of lipid metabolism in teleosts. By enabling comprehensive lipid profiling, these approaches have facilitated the identification of novel biomarkers, elucidated metabolic pathways associated with dyslipidemia and obesity, and advanced the development of targeted nutritional and therapeutic interventions. The zebrafish model has proven invaluable in metabolic research, offering new perspectives on lipid-related disorders and potential treatment strategies.

While lipidome analysis provides a comprehensive snapshot of all lipid species within a tissue, transcriptomics offers a dynamic perspective by capturing the complete set of RNAs transcribed by cells at a given time. Since cells modulate their transcriptomic profiles in response to external stimuli, transcriptomic analyses can unveil critical insights into different physiological states. By quantifying transcripts, researchers can assess an organism's adaptive capacity or susceptibility to challenges such as high-fat intake or the efficacy of novel bioactive compounds. The advent of next-generation sequencing technologies has significantly enhanced our ability to conduct in-depth transcriptomic studies, facilitating a deeper understanding of disease mechanisms (Wang et al., 2009).

Transcriptomic approaches have been particularly instrumental in identifying biomarkers associated with dyslipidemia. A recent blood transcriptomic study identified 40 differentially expressed genes (DEGs) linked to dyslipidemia, primarily involved in metabolism, mitochondrial function, and immune system regulation (Plaza-Florido et al., 2021). Among these DEGs, several were identified as hub genes playing pivotal roles in conditions such as type 2 diabetes, macrophage phospholipid metabolism, and inflammation. Similarly, transcriptomic analyses of blood cells have provided valuable insights into the relationship between dyslipidemia and other diseases, including type 2 diabetes and periodontitis (Corbi et al., 2020). These human studies highlight the potential of transcriptomics in elucidating novel aspects of dyslipidemia, though most findings remain observational. Notably, there are relatively few randomized controlled trials exploring the impact of dyslipidemia on the transcriptomic landscape of specific tissues.

In human studies, transcriptomic analyses are often constrained by the difficulty of obtaining tissue samples, leading researchers to primarily rely on blood cells. However, the use of animal models has significantly expanded our understanding of the tissue-specific molecular effects of dyslipidemia and its modulation through dietary interventions. Transcriptome profiling has revealed that a cholesterol-rich diet suppresses genes involved in endogenous cholesterol biosynthesis while inducing endoplasmic reticulum stress in the zebrafish intestine (Gora et al., 2022). In another study, Ka et al. (2020) examined transcriptomic changes in the liver of zebrafish subjected to a high-fat diet and conducted a comparative analysis with mice. Their findings indicated that hepatic transcriptomic responses in zebrafish differed substantially from those observed in mice. Despite these species-specific differences, KEGG pathway analysis revealed the upregulation of similar signaling pathways in both species in response to high-fat diets. This suggests an evolutionary convergence between poikilotherms and homeotherms in

lipid metabolism regulation, reinforcing the relevance of zebrafish as a model for human hyperlipidemia and related metabolic disorders.

Further supporting the utility of zebrafish as a model, Zhong et al. (2022) investigated the effects of serotonin on energy homeostasis using a diet-induced obesity model in zebrafish. The fish were fed a high-fat diet for eight weeks and administered intraperitoneal injections of PBS, 0.1 mM, or 10 mM 5-hydroxytryptamine (serotonin). Transcriptomic analysis of muscle tissues revealed 1,134, 3,713, and 2,535 differentially expressed genes (DEGs) across the treatment groups. Enrichment analysis identified these DEGs as being significantly associated with multiple metabolic pathways, including ribosome biogenesis, oxidative phosphorylation, proteasome activity, PPAR signaling, and ferroptosis. These findings provide valuable insights into the transcriptional responses of muscle tissue in obese zebrafish exposed to serotonin, offering novel perspectives on the peripheral regulatory effects of serotonin in teleosts. Such studies also open new avenues for the prevention and treatment of obesity-related metabolic dysfunction.

**Conclusion**

The integration of omics technologies—particularly lipidomics and transcriptomics—has significantly advanced our understanding of lipid metabolism in teleosts. By enabling comprehensive lipid profiling, these approaches have facilitated the identification of novel biomarkers, elucidated metabolic pathways associated with dyslipidemia and obesity, and supported the development of targeted nutritional and therapeutic interventions. The zebrafish model, in particular, continues to provide valuable insights into lipid-related disorders, reinforcing its role as a powerful platform for metabolic research.

**References**

Ahmed S, Shah P, Ahmed O. Biochemistry, Lipids. [Updated 2023 May 1]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK525952/

Asadi Z, Moghbeli M, Khayyatzadeh SS, Mohammadi Bajgiran M, Ghaffarian Zirak R, Zare Feyzabadi R, et al. (2020) A positive association between a western dietary pattern and high LDL-C among Iranian population. J Res Health Sci 20: e00485.

Babin, P. J., & Vernier, J. M. (1989). Plasma lipoproteins in fish. J Lipid Res. 30(4):467-89.

Baek JS, Fang L, Li AC, Miller YI. (2012) Ezetimibe and simvastatin reduce cholesterol levels in zebrafish larvae fed a high-cholesterol diet. Cholesterol 2012: 564705.

Benchoula, K., Serpell, C. J., Mediani, A., Albogami, A., Misnan, N. M., Ismail, N. H., ... & Hwa, W. E. (2024). 1H NMR metabolomics insights into comparative diabesity in male and female zebrafish and the antidiabetic activity of DL-limonene. *Scientific Reports*, *14*(1), 3823.

Chester A, Somasekaram A, Tzimina M, Jarmuz A, Gisbourne J, O'keefe R, et al. (2003) The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. EMBO J 22: 3971-3982.

Conticello SG, Thomas CJF, Petersen-Mahrt SK, Neuberger MS. (2005) Evolution of the AID/APOBEC family of polynucleotide (Deoxy)cytidine deaminases. Mol Biol Evol 22: 367-377.

Corbi SCT, De Vasconcellos JF, Bastos AS, Bussaneli DG, Da Silva BR, Santos RA, et al. (2020) Circulating lymphocytes and monocytes transcriptomic analysis of patients with type 2 diabetes mellitus, dyslipidemia and periodontitis. Sci Rep 10: 8145.

diseases. Diabetes Metab Syndr Obes 12: 2221-2236.

Gomez-Delgado F, Katsiki N, Lopez-Miranda J, Perez-Martinez P. (2021) Dietary habits, lipoprotein metabolism and cardiovascular disease: From individual foods to dietary patterns. Crit Rev Food Sci Nutr 61: 1651-1669.

Gora, A. H., Rehman, S., Dias, J., Fernandes, J. M., Olsvik, P. A., Sørensen, M., & Kiron, V. (2023a). Protective mechanisms of a microbial oil against hypercholesterolemia: evidence from a zebrafish model. *Frontiers in Nutrition*, *10*, 1161119.

Gora, A. H., Rehman, S., Dias, J., Fernandes, J. M., Olsvik, P. A., Sørensen, M., & Kiron, V. (2023b). Microbial oil, alone or paired with β-glucans, can control hypercholesterolemia in a zebrafish model. Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids, 1868(11), 159383.

Gora, A. H., Rehman, S., Kiron, V., Dias, J., Fernandes, J. M., Olsvik, P. A., ... & Cardoso, M. (2022). Management of hypercholesterolemia through dietary ß-glucans–insights from a zebrafish model. *Frontiers in Nutrition*, *8*, 797452.

Hedayatnia M, Asadi Z, Zare-Feyzabadi R, Yaghooti-Khorasani M, Ghazizadeh H, Ghaffarian Zirak R, et al. (2020) Dyslipidemia and cardiovascular disease risk among the MASHAD study population. Lipids Health Dis 19: 42.

Hu FB. (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol 13: 3-9.

Jin S, Cho KH. (2011) Water extracts of cinnamon and clove exhibits potent inhibition of protein glycation and anti-atherosclerotic activity in vitro and in vivo hypolipidemic activity in zebrafish. Food Chem Toxicol 49: 1521-1529.

Ka, J., Pak, B., Han, O., Lee, S., & Jin, S. W. (2020). Comparison of transcriptomic changes between zebrafish and mice upon high fat diet reveals evolutionary convergence in lipid metabolism. *Biochemical and Biophysical Research Communications*, *530*(4), 638-643.

Kopp W. (2019) How western diet and lifestyle drive the pandemic of obesity and civilization

Lickwar CR, Camp JG, Weiser M, Cocchiaro JL, Kingsley DM, Furey TS, et al. (2017) Genomic dissection of conserved transcriptional regulation in intestinal epithelial cells. PLoS Biol 15: e2002054.

Marza E, Barthe C, André M, Villeneuve L, Hélou C, Babin PJ. (2005) Developmental expression and nutritional regulation of a zebrafish gene homologous to mammalian microsomal triglyceride transfer protein large subunit. Dev Dyn 232: 506-518.

Miao H, Zhao Y-H, Vaziri ND, Tang D-D, Chen H, Chen H, et al. (2016) Lipidomics biomarkers of diet-induced hyperlipidemia and its treatment with Poria cocos. J Agric Food Chem 64: 969-979.

Na W, Chung B, Sohn C. (2019) A relationship between dietary patterns and dyslipidemia in urban-dwelling middle-aged korean men: using korean genome and epidemiology study (KOGES). Clin Nutr Res 8: 219-228.

O'hare EA, Wang X, Montasser ME, Chang YP, Mitchell BD, Zaghloul NA. (2014) Disruption of ldlr causes increased LDL-C and vascular lipid accumulation in a zebrafish model of hypercholesterolemia. J Lipid Res 55: 2242-53.

Otis JP, Zeituni EM, Thierer JH, Anderson JL, Brown AC, Boehm ED, et al. (2015) Zebrafish as a model for apolipoprotein biology: comprehensive expression analysis and a role for APOA-IV in regulating food intake. Dis Model Mech 8: 295-309.

Ottosson F, Emami Khoonsari P, Gerl MJ, Simons K, Melander O, Fernandez C. (2021) A plasma lipid signature predicts incident coronary artery disease. Int J Cardiol 331: 249-254.

Plaza-Florido A, Altmäe S, Esteban FJ, Cadenas-Sanchez C, Aguilera CM, Einarsdottir E, et al. (2021) Distinct whole-blood transcriptome profile of children with metabolic healthy overweight/obesity compared to metabolic unhealthy overweight/obesity. Pediatr Res 89: 1687-1694.

Schlegel A. (2016) Zebrafish models for dyslipidemia and atherosclerosis research. Front Endocrinol 7: 159.

Shon JC, Kim WC, Ryu R, Wu Z, Seo J-S, Choi M-S, et al. (2020) Plasma lipidomics reveals insights into anti-obesity effect of Chrysanthemum morifolium ramat leaves and its constituent luteolin in high-fat diet-induced dyslipidemic mice. Nutrients 12: 2973.

Stoletov, K., Fang, L., Choi, S. H., Hartvigsen, K., Hansen, L. F., Hall, C., ... & Miller, Y. I. (2009). Vascular lipid accumulation, lipoprotein oxidation, and macrophage lipid uptake in hypercholesterolemic zebrafish. Circulation research, 104(8), 952-960.

Vaduganathan, M., Mensah, G. A., Turco, J. V., Fuster, V., & Roth, G. A. (2022). The global burden of cardiovascular diseases and risk: a compass for future health. Journal of the American College of Cardiology, 80(25), 2361-2371.

Wallace KN, Akhter S, Smith EM, Lorent K, Pack M. (2005) Intestinal growth and differentiation in zebrafish. Mech Dev 122: 157-173.

Wang Z, Gerstein M, Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10: 57-63.

Wilson MM, Reedy J, Krebs-Smith SM. (2016) American diet quality: where it is, where it is heading, and what it could be. J Acad Nutr Diet 116: 302-310.e1.

Zeitler, E. M., Li, Y., Schroder, M., Falk, R. J., & Sumner, S. (2023). Characterizing the metabolic response of the zebrafish kidney to overfeeding. American Journal of Physiology-Renal Physiology, 325(4), F491-F502.

Zeitler, E. M., Li, Y., Schroder, M., Falk, R. J., & Sumner, S. (2023). Characterizing the metabolic response of the zebrafish kidney to overfeeding. *American Journal of Physiology-Renal Physiology*, *325*(4), F491-F502.

Zhang, P., Liu, N., Xue, M., Zhang, M., Xiao, Z., Xu, C., ... & Zhou, Y. (2024). β-Sitosterol Reduces the Content of Triglyceride and Cholesterol in a High-Fat Diet-Induced Non-Alcoholic Fatty Liver Disease Zebrafish (Danio rerio) Model. *Animals*, *14*(9), 1289.


Indian Council of Agricultural Research
**Central Marine Fisheries Research Institute**

(Ministry of Agriculture and family welfare)
PB No: 1603, Ernakulam North P.O, Kochi-682 018, Kerala, India
Phone: +91 484 2394357, 2394867 Fax +91 4842394909
E-mail: contact@cmfri.org.in www.cmfri.org.in

ICAR

सी एम एफ आर आइ
CMFRI