Journal of Fish Health
Volume 5, Issue 2, May 2025
https://doi.org/10.29303/jfh.v5i2.6281

A Comprehensive Review of Bioinformatics Tools and Applications Revolutionising Aquatic Animal Health Management

Shijina Purushothaman Nambiar[1*], Surya Chaitanya Banuru[2]

[1]Department of Aquatic Animal Health Management, Kerala University of Fisheries and Ocean Studies, Kochi
[2]Department of Aquatic Animal Health Management, Central Institute of Fisheries Education, Mumbai
Panangad Road, Madavana, Junction, Kochi, Ernakulam, Kerala 682506, India

Correspondence:
shijinapnambiar@gmail.com

Received:
January 29th, 2025
Accepted:
March 6th, 2025

Keywords:
Aquatic Animal Health, Bioinformatics, Genetics, Molecular

ABSTRACT

Aquatic animal health management is a critical aspect of maintaining the sustainability and productivity of aquaculture systems. The increasing complexity of aquatic ecosystems and the challenges posed by the emerging diseases necessitate innovative approaches for effective health management. Molecular medicine necessitates the integration and analysis of genetic, molecular, cellular, and clinical data, posing a unique set of problems to bioinformatics. Omics technologies give researchers the tools they help to examine differences in DNA, RNA, proteins, and other biological components within and between different species. Bioinformatics is indispensable in the field of fish health as it enables a deeper understanding of the genetic, molecular, and environmental factors that affect fish well-being. This review explores the application of bioinformatics in understanding, monitoring, and managing the health of aquatic organisms.

INTRODUCTION

In aquaculture, bioinformatics is being used to scrutinize genetic data from fish and their microbial communities to better comprehend the fundamental mechanisms of fish health and disease. In this article, we will reconnoit how cutting-edge bioinformatics strategies are currently adopted for the advancement of fish health management system that can eventually aid in revolutionizing aquaculture farming practices. Bioinformatics tools are essential for managing the health of aquatic animals in aquaculture systems. They enable the analysis of genetic, genomic, and environmental data to monitor and improve the health and well-being of aquatic organisms.

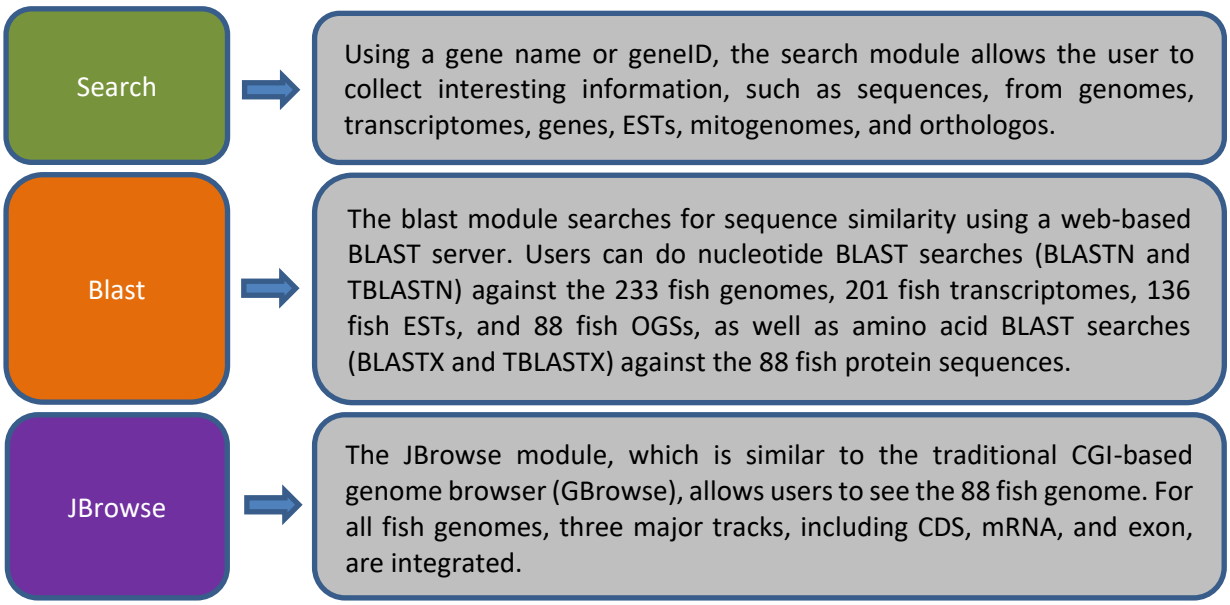
Bioinformatics plays a crucial role in disease diagnosis in fish by leveraging computational tools and techniques to analyze large volumes of biological data. Biological information, which includes sequences, structures, functions, and phylogeny, is analyzed, compared, integrated, visually displayed, modelled, stored, systematized, searched, and eventually distributed using computer based bioinformatic tools (Singh et al., 2011; Janicki et al., 2011). Bioinformatics make use of computational technology to comprehend and arrange

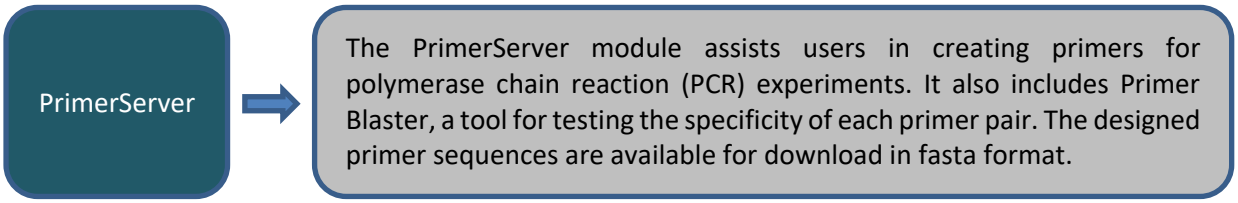
data relating to biological macromolecules (Luscombe, 2001). Hence, the primary purpose of bioinformatics tools is to organize, store, retrieve, and analyse enormous amounts of multiomics (genomics, proteomics, transcriptomics, metabolomics, etc.) data generated by high throughput technologies (Poetsch & Li, 2023). Bioinformatics tools are used to analyze the genomes of various aquaculture species, such as fish, shrimp, and oysters. This enables researchers to identify genes associated with desirable traits, understand genetic diversity, and develop breeding programs for improved strains. By comparing the genomes of different aquaculture species, researchers can gain insights into evolutionary relationships and identify genes responsible for specific adaptations or traits. This information can inform breeding strategies and conservation efforts.

Molecular medicine necessitates the integration and analysis of genetic, molecular, cellular, and clinical data, posing a unique set of problems to bioinformatics. Bioinformatics is now essential for comprehending genomic, transcriptomic, and proteomic data generated by high-throughput experimental technologies, as well as arranging information gathered through traditional biology and medicine (Molidor et al., 2003). Bioinformatics is a relatively new field that connects the biology sciences and computer sciences. The enormous increase of biological sequence information necessitates the integration of these two fields. Omics technologies give researchers the tools they help to examine differences in DNA, RNA, proteins, and other biological components within and between different species (Hasin et al., 2017). High throughput assays used in omics experiments provide enormous volumes of data on structural or functional changes to cells. Omics technologies include genomics, proteomics, transcriptomics, epigenomics, reactomics, metabolomics etc (Shea & Misra, 2020). Genomics involves analysis and comparison of entire genetic complement of a species by comparing more or less representative subsets of genes. Metabolomics is an important and powerful approach which reflects the underlying biochemical activity of cells/tissues, which is best representing the molecular phenotype of an organism.

Bioinformatics is indispensable in the field of fish health as it enables a deeper understanding of the genetic, molecular, and environmental factors that affect fish well-being.

METHODS





RESULTS AND DISCUSSION

1. Bioinformatics Tools and Databases

Bioinformatics tools and databases are invaluable resources for researchers, aquaculturists, and fisheries managers working to improve the health, productivity, and sustainability of fish populations. These tools and databases aid in data analysis, information retrieval, and the development of predictive models. Bioinformatics tools are software applications which are made to extract useful data from the vast amount of molecular biology and biological databases and perform structural or sequence analysis (Raza, 2012). The tools are broadly classified under four categories (Danish et al., 2017).

Homology and Similarity Tools

Homology and Similarity tools helps in comparing database sequences, making it possible to determine the similarity of an undiscovered structure and function (Herrero et al., 2016).

Protein Function Analysis

It compares protein sequences with secondary protein databases.

Structural Analysis

Structural analysis compares structures with known databases.

Sequence Analysis

Computational analysis of a DNA, RNA, or peptide sequence can reveal more about its properties, biological function, structure, and evolution (Rehm, 2001). This process is known as sequence analysis. Extensive analysis (such as evolutionary analysis, identification of mutations, hydropath regions, CpG islands, compositional biases, and so on) should be performed to discover the precise purpose of the query sequence (Rai et al., 2012).

According to Meena et al. (2020) the bioinformatics tools used in aquaculture are broadly classified to:

a. BLAST (The Basic Local Alignment Search Tool)

BLAST is a widely used tool for comparing DNA, RNA, or protein sequences against databases to identify homologous sequences. It helps in annotating and characterizing genetic elements in fish genomes (Madden., 2003). PSI-BLAST, PHI-BLAST, BLAST 2 sequences, and more variants are currently available (Korfet al., 2003). BLAST can be used in the field of fisheries in several purposes, these include:

Identifying Various Fish Species

Using BLAST, we may be able to identify a species correctly or to compare different species (Ross, 2008). When working with a DNA sequence from an unknown species, this can come in helpful.

Fish DNA Mapping

When working with known species and trying to sequence a gene from an unknown location, BLAST can compare the sequence of interest's chromosomal position to the relevant sequence in the database (BlastN, 2019).

Phylogeny Discovery

Using the BLAST results, we may generate a phylogenetic tree using the BLAST web page (Dereeper et al., 2010).

Comparison of Various Fish Species

When working with genes, Blast can find common genes in two related species and map annotations from one organism to another (BlastN, 2019).

b. FASTA (FAST-All)

FASTA compares nucleotide or peptide query sequences to a sequence database using the quick sequencing technique (Pearson, 2016). FASTA does a heuristic search using a protein query. A DNA query is translated by FASTX and FASTY. SSEARCH (local), GGSEARCH (global), and GLSEARCH (global query, local database) provide optimal results (Pearson, 2014).

c. EMBOSS (The European Molecular Biology Open Software Suite)

EMBOSS is a sophisticated, open-source software analysis package with over 100 programmes (applications) for sequence alignment and database searching using sequence patterns (Rice et al., 2011). The software effortlessly handles data in a variety of forms and even allows for the transparent retrieval of sequence data from the web. Furthermore, because the package contains extensive libraries, it can be used by other scientists to develop and deploy applications in the genuine open-source spirit. EMBOSS used in the field of fisheries for following purposes (Rana, 2020).

- Sequence alignment.
- Rapid data base searching with sequence patterns.
- Protein identification including domain analysis.
- Nucleotide sequence pattern analysis.
- Codon usage analysis for small genomes.

d. ClustalW

Multiple sequence alignment tools used to align DNA or protein sequences, aiding in the identification of conserved regions or genetic variations. Clustal omega is the latest version which is fast and scalable but whereas clustalW is the classical version. Through sequence weighting, position-specific gap penalties, and weight matrix selection, CLUSTALW improves the sensitivity of progressive multiple sequence alignment (Albayraktaroglu, 2005).

e. RasMol and Protein Explorer (derivative RasMol)

RasMol is commonly used for displaying the structural properties of DNA, proteins, and smaller molecules. Other tools utilised include Cn3D, Swiss PDB viewer, Hex, Vega, Bioeditor, Bioviewer, Chime, and others (Laskowski, 2018).

Other bioinformatic stools include:

Artemis

Artemis is a genome browser and annotation tool that allows researchers to visualize and annotate genes, regulatory elements, and other features in fish genomes.

MAFFT

Multiple Alignment using Fast Fourier Transform (MAFFT) is a tool for multiple sequence alignment. It is used to align DNA or protein sequences from different fish species to identify conserved regions and variations.

Bioconductor

Bioconductor is an open-source platform for the analysis and visualization of high-throughput genomics data. R packages in Bioconductor can be used for various statistical analyses on fish-related datasets.

Phylogenetic Software

Tools like RAxML, PhyML, and BEAST are used for phylogenetic analysis to study the

evolutionary relationships among different fish species and populations.

HMMER

Hidden Markov Model based on homology search is used to identify remote homologs and conserved protein domains in fish genomes.

Genome Annotation Tools

AUGUSTUS, GeneMark, and Prokka: Tools for predicting gene structures and annotating genomes. They help identify genes, coding regions, and functional elements in the genomes of aquatic organisms.

Variant Calling and Genotyping Tools

GATK (Genome Analysis Toolkit): Used for calling variants (SNPs and INDELs) from next-generation sequencing data to identify genetic variations relevant to health and disease resistance.

PLINK: A tool for genotyping and quality control of genetic data, useful in genome-wide association studies (GWAS) for identifying genetic markers associated with disease resistance.

Phylogenetic Analysis

PhyML, RAxML, and MrBayes: Phylogenetic tree construction tools used to infer evolutionary relationships among aquatic pathogens or host species, aiding in understanding disease transmission and relatedness.

Pathogen Identification and Characterization

NCBI BLAST and Virulence Factor Databases: Used to identify and characterize pathogens by comparing their genetic sequences to known pathogens or virulence factors.

Structural Biology and Drug Discovery

Rosetta, AutoDock, and Schrödinger Suite: Tools for protein structure prediction and molecular docking simulations, aiding in the design of drugs and vaccines against aquatic pathogens.

Functional Analysis

Gene Ontology (GO) Enrichment Analysis: Identifies overrepresented gene functions or pathways in high-throughput datasets, helping to understand the biological processes involved in disease resistance or health.

Data Visualization

R, Python (Matplotlib, Seaborn): Programming languages and libraries for creating plots and visualizing data, which is crucial for interpreting and presenting results from various bioinformatics analyses.

Machine Learning and Predictive Modeling

Scikit-learn, TensorFlow, and Keras: Machine learning libraries used for building predictive models to forecast disease outbreaks, identify risk factors, or predict treatment outcomes based on genetic and environmental data.

Databases and Repositories

NCBI GenBank, Ensembl, and UniProt: Publicly available databases that house genetic and genomic data for aquatic organisms and pathogens, facilitating data retrieval and analysis. These bioinformatics tools, along with high-throughput sequencing technologies, enable researchers and aquaculture practitioners to monitor the health of aquatic animals, identify disease risks, and develop strategies for disease prevention and management in aquaculture systems.

Sequence Annotation Tools

The identification and labelling of certain features within a DNA sequence, such as coding regions, regulatory elements, and non-coding regions, is referred to as sequence

annotation (Alexander, 2010). Bioinformatic tools are critical in automating the sequence annotation process, making it more efficient and accurate.

A tool that aids in sequence annotation is the Genome Annotation Transfer Utility (GATU). GATU is a programme that transfers functional annotations from well-annotated reference genomes to newly sequenced genomes or sequences of interest. GATU can predict gene architectures, identify regulatory elements, and mark functional areas in target sequences using information from reference genomes (Ejigu Jung, 2020).

The GeneMark programme, which predicts gene locations and structures inside a DNA sequence, is another extensively used tool. GeneMark identifies open reading frames (ORFs) and infers gene functions based on sequence patterns and codon using a combination of statistical models and algorithms (Haas et al., 2008).

Genome Assembly and Visualization Tools

The process of rebuilding whole genomes from broken DNA sequences is known as genome assembly (Jiao & Schneeberger, 2017). This approach is aided by bioinformatics tools that align and overlap sequence reads, locate overlaps, and generate consensus sequences.

- The Celera Assembler is a well-known tool for genome assembly. It accurately assembles complicated genomes by combining sequence alignment, graph theory, and statistical approaches. The Celera Assembler can handle huge datasets, delivering high-quality genome assemblies to researchers for further investigation (Rizzi et al., 2019).
- Genome browsers are essential for visualising DNA sequences and associated annotations. The Genome Browser at the University of California, Santa Cruz (UCSC) and the Genome Data Viewer at the National Centre for Biotechnology Information (NCBI) are two popular genome browsers (Haeussler et al., 2019). These browsers enable researchers to interactively visualise DNA sequences, annotations, genetic variants, and other pertinent genomic information.

Biological Databases

Biological databases are collections of information on life sciences gathered from scientific investigations, published literature, high throughput technology, and computational analysis (Kanehisa, 2019). Biological databases are useful in the following context

- It aids researchers in evaluating existing data and developing new ideas, anti-virus, beneficial bacteria, and cures, among other things.
- It aids scientists in their understanding of biological events.
- The database serves as a repository for information.

One can create database based on IMCA principle (Márquez et al., 2012) ie,

- Import - Collection of information from different resources
- Merging - Proper maintainance of database on a computer excel sheet or back end using computer tools
- Classification - Creation criteria in order to search databases
- Annotation - Attaching a biological function to micro molecules

Biological Data Formats

- Text, sequence data, protein structure and links are different formats in which biological data exists (Wilson et al., 2021). For example,
- Text formats - PubMed an OMIM
- Sequence data - GenBank, in terms of DNA, and Uniprot, in terms of protein
- Protein structures - PDB, SCOP, CATH

Major Classification of Biological Databases

1) Primary sequence databases- The primary database contains experimentally produced data such as genomic sequences, macromolecular structures, and so on. The data entered here is left uncurated (no changes are made to the data). It gathers unique data from the laboratory and makes these data available to typical users without modification (Sharma & Yadhav, 2022). Examples are DNA databank of Japan (DDBJ), European Molecular Biology Laboratory (EMBL), GenBank.
2) Meta Databases-Metadata is organised reference data that aids in the sorting and identification of attributes of the information it describes. Metadata organises a data object by utilising phrases that are associated with that thing. It also allows dissimilar things to be detected and linked with comparable ones to assist optimise the usage of data assets (Haynes, 2018).
3) Genome database - Collect organism genome sequence annotation, analyse them and provide public access.
4) Protein sequence databases - Protein amino acid sequences and related information are stored in the protein sequence database. A protein's amino acid sequence is crucial because it dictates the protein's three-dimensional shape, function, and identity (Wu et al., 2006). The following are some of the most popular protein sequence databases: UniProt, Protein Information Resource (PIR), Swiss-Prot, TrEMBL.
5) Protein structure databases -Protein structure databases are compilations of data about protein' three-dimensional structure and secondary structure. Protein structure databases come in a variety of forms (Paysan-Lafosse et al., 2023). Some examples are: Protein Data Bank (PDB), Class, Architecture, Topology, and Homologous Superfamily (CATH), Structural Classification of Protiens (SCOP).
6) Protein-protein interaction - Protein-protein interaction databases are collections of information on protein interactions. These databases contain useful information about the interactions of various proteins and their roles in biological systems (Kuhlman & Brandley, 2019). Protein-protein interaction databases include the following: STRING, Database of Interacting Protein (DIP), Biomolecular Interaction Network Database (BIND).
7) Protein pattern and profile databases: It contain information on sequence motifs. Sequence motifs are protein structural or functional properties. As a result, the utilisation of protein sequence patterns or profiles is a significant tool in determining protein function. Some of the examples are InterPro, PROSITE (Blum et al., 2021; Sigrist et al., 2010).

For protein structure prediction, bioinformatics provides a number of computational tools and algorithms. These methodologies can be broadly classified into two types: comparative modelling (homology modelling) and ab initio (de novo) modelling.

These bioinformatics tools and databases facilitate a wide range of research activities in fisheries, including genetic analysis, phylogenetics, functional genomics, and disease management. Researchers and practitioners can leverage these resources to make informed decisions and drive advancements in fish health and sustainability.

Computational Methods for Protein Structure Prediction

Comparative Modeling (Homology Modeling)

The idea behind comparative modelling is that proteins with comparable amino acid sequences have similar shapes and activities. This method employs already determined experimentally determined protein structures (templates) that have significant sequence similarity to the target protein (Xu et al., 2020). Comparative modelling entails aligning the target sequence with the template structure, transferring structural information from the template to the target, and optimising the anticipated model's accuracy. Comparative

modelling methods like as MODELLER, SWISSMODEL, and Phyre2 are commonly utilised (Webb & Sali, 2016).

Ab Initio (De Novo) Modeling

The goal of ab initio modelling is to predict protein structures based simply on the amino acid sequence, without the use of known template structures. This method predicts the most energetically favourable protein structure using physical and statistical concepts (Hardin et al., 2002). Ab initio approaches sometimes utilise fragment assembly, which involves reassembling tiny segments of known protein structures to construct the full-length protein model (Lee et al., 2017). To refine the predicted structures, physics-based force fields and optimisation methods are used. Rosetta and I-TASSER are two examples of ab initio modelling approaches (Yang & Zhang, 2015; Lee et al., 2017).

2. Gene Expression Profiling

Gene expression profiling is critical for assessing gene transcriptional activity under various situations, providing vital insights into cellular processes and disease mechanisms. Bioinformatics is important in analysing and interpreting gene expression data, allowing researchers to decipher complicated biological events and uncover prospective treatment targets (Cho &Cheng., 2007). Gene expression profiling dealts with data preprocessing, normalisation, and differential gene expression analysis.

Data Preprocessing

To assure data quality and reproducibility, gene expression data generated using techniques such as microarrays or RNA sequencing frequently require preprocessing processes. Data cleansing, background correction, and normalisation are all tasks performed by bioinformatic systems (Johnson & Krishnan, 2022).

Data cleaning is the process of removing noise, artefacts, and technological biases from raw gene expression data. The 'affy' package in Bioconductor for microarray data and the 'edgeR' or 'DESeq2' packages for RNA sequencing data are popular bioinformatic tools for data preparation (Li et al., 2022). These programmes include a variety of algorithms and statistical methods for preprocessing gene expression data and ensuring data quality for further studies.

Normalisation

Normalisation is an important step in gene expression profiling because it removes undesired technical variations that might mask important biological signals. To normalise gene expression data across samples, bioinformatic methods and approaches are used (Vallejos et al., 2017; Zhao et al., 2020). Normalisation approaches seek to compensate for discrepancies in sequencing depth, gene length, and other technical biases in order to allow for accurate comparison of gene expression levels across samples (Risso et al., 2011; Wolf, 2013).

Total counts normalisation, quantile normalisation, and normalisation utilising reference genes or spike-in controls are all common normalisation procedures (Pereira et al., 2018).

Bioinformatic tools in Bioconductor, such as 'RUVSeq' and 'DESeq2', provide functionality for implementing these normalisation approaches and resolving any batch effects or other confounding variables (Federico et al., 2020; Büttner et al., 2019; Cole et al., 2019).

Differential Gene Expression Analysis

Differential gene expression analysis seeks to find genes with significant differences in expression levels across biological circumstances or experimental groups. This analysis sheds light on the genes and pathways that are linked to various phenotypes or biological processes. To conduct differential gene expression analysis, bioinformatic tools and statistical methods

are used (Rapaport et al., 2013; Trapnell et al., 2012).

'limma' and 'edgeR' in Bioconductor are two popular bioinformatic tools for analysing differential gene expression. These tools use statistical models such as the moderated t-test and negative binomial models to discover statistically significant differentially expressed genes. The results are frequently corrected for multiple testing using methods such as the false discovery rate (FDR) or family-wise error rate (FWER) (Dong et al., 2021; Su et al., 2017).

Pathway Analysis

Pathway analysis is facilitated by bioinformatic tools and databases, which enable the identification of biological pathways, networks, and functional annotations associated with gene expression data (Chong et al., 2018).

Pathway analysis tools such as DAVID, Gene Set Enrichment Analysis (GSEA), and Enrichr are routinely employed. These tools examine the enrichment of differentially expressed genes in certain pathways or functional categories by using curated pathway databases such as Kyoto Encyclopaedia of Genes and Genomes (KEGG) and the Gene Ontology (GO). This study aids in understanding the underlying biological processes as well as identifying prospective areas for future research (Geistlinger et al., 2021; Jia et al., 2016; Evangelista et al., 2023).

Integration of Multi-omics Data

Researchers are progressively integrating gene expression data with other omics data, such as DNA methylation, histone modification, and protein-protein interaction data, to achieve a comprehensive knowledge of gene expression regulation (Aging Atlas, 2021; Hawe et al., 2019). Bioinformatics is critical in integrating and analysing multi-omics data, allowing for a comprehensive understanding of complicated molecular interactions and regulatory systems.

Integrative bioinformatic tools and techniques, like as network-based analyses, machine learning algorithms, and data integration platforms like Galaxy and Cytoscape, make multi-omics data integration and analysis easier (Tolani et al., 2021; Zeng et al., 2018). These methods allow for the investigation of relationships between molecular layers, the identification of regulatory networks, and the prediction of novel gene regulation mechanisms.

Gene Editing

Gene editing is a highly effective method in which DNA or nucleotide sequences are inserted, removed, or modified at a specific spot in the genome of living creatures or cells using a specific set of engineered nucleases that serve as molecular scissors (Khalil, 2020; Gupta & Shukla, 2017). Approaches such as gene editing and gene silencing have the potential to transform our understanding of fish biology and disease. Using such treatments to treat organ abnormalities and dysfunctions, it is feasible to change the phenotypic and remodel cells, tissues, and organs of animals (Gutási et al., 2023; Bayarsaikhan et al., 2021). Sequence-specific programmable nucleases are the most commonly used and widely successful approaches in gene editing. These molecular scissors cut the DNA precisely at a specific location.

Meganuclease

Endonuclease enzymes are distinguished by their ability to recognise and cleave long DNA sequences (Sternberg et al., 2014). The proteins in the LAGLIDADG family, which get their name from a conserved amino acid sequence, are the most common and well-known meganucleases. This is commonly found in microbial organisms (Grishin et al., 2010; Fajardo-Sanchez et al., 2008).

Zinc Finger Nuclease (ZFN)

Zinc finger nucleases are genetically modified hybrid proteins that are commonly used as a potential gene editing tool. The presence of zinc ions coordinated by cysteine and histidine amino acids forms a "finger-like" structure in these domains (Bora et al., 2023; Miglani, 2017). This hybrid protein is composed of specific DNA-binding domains that bind to the endonuclease Fok I, which was engineered to target specific genomic regions. According to the concept, various zinc fingers detect distinct sets of nucleotide triplets. ZFPs can recognise and bind to certain DNA sequences, and ZFN enzymes can cut the DNA in the targeted sequences (González Castro et al., 2021).

Fok-1: Fok -1 is a bacterial II restriction endonuclease found in *Flavobacterium okeanokoites* containing DNA binding domain in N terminal and non sequence-specific DNA cleavage domain in C terminal.

The main advantage of this method is to preserve temporal and tissue specific gene expression over the traditional gene therapy methods.

Transcription Activator-Like Effector Nucleases (TALENs)

TALENs are made up of particular effector proteins that contain the DNA-binding domain as well as the Fok 1 nuclease domain. These domains work as dimers in pairs, binding to opposite strand DNA and creating DSB. These distinct nucleases are released by the pathogenic bacteria Xanthomonas, which invade plant cells' cytoplasm (Mahfouz et al., 2014; Sprink et al., 2015). Each of these nuclease platforms has its own design for localization and activation, as well as a core domain for specialised DNA binding. The DNA-binding domain contains 10 to 30 repetitions of monomers in which each of them binds to one specific nucleotide of the target DNA sequence. TALENs attach to target sites in the nucleus as dimers, with the Fok1 domains at the c-termini and cleavage happening in the "spacer" sequence (Iqbal et al., 2023). Following that, the error-prone NHEJ technique is used to repair DNA breaks, which occurs essentially in the same way as ZFNs.

CRISPER-Cas9

One of the most recent developments in the genome editing toolbox is the clustered regularly interspaced short palindromic repeats (CRISPR) gene engineering approach. This most recent method of gene editing was discovered in 2012. The CRISPR/Cas9 system consists of a Cas9 endonuclease and a modified single guide RNA (sgRNA/gRNA) that contains a targeting crRNA (crRNA) and a transactivating crRNA (tracrRNA) (Gutási et al., 2023). Therefore, the Cas9 protein and sgRNA are the two important components. A 20-base pair guide RNA leads the Cas9 nuclease to its target sequence. As a result, one of the major advantages of this method is that it requires only a simple change of 20 nucleotide sgRNA "spacer" sequences, which are easier to modify, rather than the massive repetitive complex construction of DNA-binding arrays for each new genomic target site, as in the ZFN and TALEN systems. The target sequence's protospacer-adjacent motif (PAM), which binds Cas9, is also important (Uusi-Mäkelä, 2022). PAM is a short, specific sequence (NGG trinucleotide sequence) that appears after the target DNA sequence downstream of the crRNA binding site. It is needed for a Cas nuclease-mediated break (Rees et al., 2018).

Different CRISPER systems

Class 1 (CRISPR-Cas3) and Class III (CRISPR-Cas10): Makes use of various Cas proteins as well as the crRNA.

CRISPR-Cas9 (class 2) and CRISPR-Cpf1 (class 5): Combine a big single-component Cas-9 protein with crRNA and tracerRNA (Wada et al., 2022).

The Cas9 endonuclease accurately cuts the target DNA sequence and creates a DSB when directed by sgRNA. To make modifications, researchers can add or delete genetic

material sequences or replace an existing segment with a changed DNA sequence (Desai et al., 2022; Pickar-Oliver & Gersbach, 2019). A DBS can be corrected using either NHEJ or HDR.

Table 1. CRISPR Technology Used in Aquaculture

Fish	Target Gene	Result	Reference
Zebrafish (*D. rerio*) Nile tilapia (*O. niloticus*)	Ovarian aromatase (*cyp19a1a)*	Phenotype - all male phenotype	Nakamura *et al*. (2003); Li *et al.* (2013)
Atlantic salmon (*S. salar*)	*yr and slc45a2*	Pigment - loss of pigment	Edvardsen *et al.* (2014)
Red sea bream (*P. major*)	Myostain or GDF 8	skeletal mass increment	Kato *et al.* (2016)
Zebrafish (*D. rerio*)	Mesp genes (*mespaa, mespab, mespba and mespbb)*	Immune and growth function	Xiong *et al.* (2017)

Environmental DNA (eDNA)

Environmental DNA (eDNA) is nuclear or mitochondrial DNA that has been discharged into the environment by an organism. eDNA can be found in released faeces, mucus, and gametes, shed skin and hair, and carcasses. eDNA can be found in both cellular and extracellular (dissolved DNA) forms. eDNA analysis has emerged as a potentially significant method for gaining access to the architecture of aquatic communities (Antony et al., 2022; Jo et al., 2022; Moushomi et al., 2019). Through two methodologies, eDNA barcoding and eDNA metabarcoding, we may learn about organisms, their abundance, and biomass. Specific species are targeted in samples using conventional or quantitative PCR and the traditional Sanger sequencing method in the eDNA barcoding (LeBlanc et al., 2020; Gargan et al., 2017). The eDNA metabarcoding involves screening the entire community with several conserved primers and doing Next Gen Sequencing (NGS) (Ruppert et al., 2019). According to studies, eDNA metabarcoding beats standard survey approaches in terms of non-invasive sampling, sensitivity, and affordability.

eDNA may enable the collection of data on species distribution and relative abundance in a timely, cost-effective, and standardised manner, therefore it is an appealing option for aquatic inventory and monitoring programmes (Ramirez-Amaro et al., 2022; Hobbs & Bright, 2016). Thus, employing eDNA to detect species may improve biodiversity assessments and offer information about the status, distribution, and habitat requirements of less-known species (Ficetola et al., 2019).

3. Next Generation Sequencing (NGS)

NGS is a sophisticated platform that has enabled simultaneous sequencing of thousands to millions of DNA molecules. It is also known as high-throughput sequencing and is a catch-all term used to describe a type of sequencing a variety of contemporary sequencing technologies (Costa et al., 2016).

The enormous need for low-cost sequencing has fueled the development of high-throughput sequencing, which may generate dozens or millions of sequences all at once (Jing & Zu, 2021). They are intended to lower the cost of DNA sequencing beyond what conventional dye-terminator technology can achieve. Thereby, modern technologies enable us to sequence DNA and RNA considerably more swiftly and cheaply than previously employed Sanger sequencing, and have thereby revolutionised genomics and molecular biology research

(Kumar & Kocour, 2017).

Table 2. NGS Technology

Feature	Genome Sequencing	Exome Sequencing	Targeted Gene Panel
Coverage	All genes and non coding DNA	Entire exome (20-25k genes)	10-500 genes
Accuracy	Low	Good	High
Time	Longest turnaround time	Long turnaround time	Rapid turnaround time (few days)
Cost	Most expensive	Cost effective	Most cost effective
Depth	>30x	>50-100x	>500x

Table 3. Generation Classification of Sequencing

1st Generation	2nd Generation	3rd Generation	4th Generation
Sanger sequencing	• Pyrosequencing • Sequencing by Reversible terminator chemistry • Sequencing by ligation	• Single Molecule Fluorescent Sequencing • Single molecule real time sequencing • Semiconductor Sequencing • Nanopore Sequencing	Aims at conducting genomic analysis directly in the cell

Types of NGS

Lynx Therapeutics' Massively Parallel Signature Sequencing (MPSS)

MPSS is regarded as the first "next-generation" sequencing technology. MPSS is an ultra-high throughput sequencing technology (Zhou et al., 2006). When applied to an expression profile, it reveals nearly every transcript in the sample and provides an accurate expression level for each. MPSS is a bead-based approach that employs a sophisticated strategy of adapter ligation followed by adapter decoding, reading the sequence in four nucleotide increments; nonetheless, this method is susceptible to sequence-specific bias or loss of specific sequences (Neelapu & Surekha, 2016).

Polony Sequencing

Polony sequencing is a low-cost, high-accuracy multiplex sequencing technology that can read millions of immobilised DNA sequences in simultaneously. It uses an in vitro paired-tag library, emulsion PCR, an automated microscope, and ligation-based sequencing chemistry to sequence an *E. coli* genome with greater than 99.9999% accuracy and at a fraction of the cost of Sanger sequencing (Shendure et al., 2011).

Pyrosequencing

Pyrosequencing method uses water droplets in an oil solution to amplify DNA (emulsion PCR), with each droplet carrying a single DNA template coupled to a single primer-coated bead, which eventually forms a clonal colony (Pettersson et al., 2009). Pyrosequencing uses luciferase to generate light for detection of individual nucleotides added to the nascent DNA, and the combined data is used to generate sequence read-outs (Sharma et al., 2021; Yi et al., 2015).

Illumina (Solexa) Sequencing

In this process, DNA molecules are linked to primers on a slide before being amplified. This is referred to as bridge amplification. Solexa is a dye terminator-based sequencing

method. After photographing the fluorescently labelled nucleotides, the dye and the terminal 3' blocker are chemically removed from the DNA, allowing the next cycle to begin. In contrast to pyrosequencing, DNA can only be stretched one nucleotide at a time (Kumar et al., 2022; Wang et al., 2022).

SOLiD Sequencing

In this procedure, a pool of all possible oligonucleotides of fixed length is marked according to the sequenced position. This sequencing produces sequences of comparable lengths and quantities to illumination sequencing (Lindsay, 2016).

Helioscope Single Molecule Sequencing

Helioscope sequencing makes use of DNA fragments with polyA tail adapters that are attached to the flow cell surface. The following steps use extension-based sequencing with cyclic washes of fluorescently labelled nucleotides in the flow cell (Javed et al., 2021; MacCannell, 2019). The readings are brief, up to 55 bases per run, but recent advancements in the technology allow for more accurate homopolymer and RNA sequencing reads. The Helioscope sequencer performs the reads.

DNA Nanoball Sequencing

DNA nanoball sequencing technology is a high throughput sequencing technology used to ascertain an organism's full genetic sequence. Rolling circle replication is used to amplify fragments of genomic DNA molecules (Reuter et al., 2015). When compared to existing next generation sequencing platforms, this DNA sequencing platform can sequence a large number of DNA nanoballs per run at a low reagent cost (Kalendar et al., 2022).

Single Molecule SMRT Sequencing

The sequencing is carried out using unmodified polymerase and fluorescently labelled nucleotides that are freely flowing in the solution (Ardui et al., 2018). The DNA is synthesised in so-called zero-mode wave-guides (ZMWs), which are small well-like containers with capture instruments at the bottom. Nucleotide changes can be detected using SMTR technology (Zhang et al., 2022). This is accomplished by observing polymerase kinetics. This method allows for 1000 nucleotide readings (Zhang et al., 2022).

Microarray

DNA microarray is a molecular detection technology that consists of a collection of microscopic features (often DNA) bonded to a solid surface. DNA microarrays are solid substrates, typically composed of glass or silicon, on which DNA is affixed in a pre-arranged grid pattern (Taguchi et al., 2022; Carter., 2007). The term microarray contains DNA chips, gene chips, DNA arrays, gene arrays and biochips. DNA microarray technique is based on hybridisation technique and is originated form Southern blotting (Southern, 2001).

Complementary nucleic acid sequences are distinguished from one another by the creation of hydrogen bonds between complementary nucleotide base pairs (Dadkhah et al., 2015). Labelling the samples with fluorescent dyes is done, and at least two samples are hybridised to the chip. DNA microarrays use relative quantization, which compares the same character under two different environments and identifies that character based on its position. Using DNA microarrays, the presence of one genomic or cDNA sequence in 100,000 or more can be tested in a single hybridization (Rawat, 2017).

Types of Microarrays

cDNA Based Microarrays

In this technique cDNA is used for the chip preparation. The cDNAs are amplified using PCR. This is a parallel RNA expression analysis that permits quantitative study of RNAs transcribed from both known and unknown genes (Xiang & Chen, 2000).

Oligonucleotide Based Microarrays

Spotted probes of this sort are composed of short, chemically synthesised sequences of 20- 25 mers per gene. Shorter probe lengths result in fewer errors during probe synthesis and allow for the interrogation of small genomic areas as well as polymorphisms (Shi et al., 2003). Despite being easier to make than dsDNA probes, oligonucleotide probes must be carefully constructed so that all probes have similar melting temperatures (within 50°C) and no palindromic sequences exist (Von et al., 2015).

4. Biotechnological Tools in Fish Health Management

FishDB: An Integrated Functional Genomics for Fish

FishDB is a database designed to satisfy the demands of the fish scholarship community. It's ideal for taxonomy, phylogeny, evolution, development, and agriculture research. FishDB, as far as we know, obtains practically all of its fish genomes and the majority of its fish transcriptomes from public databases. FishDB not only includes commonly used web-services as a search tool, BLAST, JBrowse, and PrimerServer, but also a platform for comparative genomics analysis on orthologs. When new genome, transcriptome, and genetic datasets of fish become available, FishDB will be continuously updated, and more enhanced functionality will be possible in the future to create a more valuable resource for promoting comparative genomics, transcriptomes, and evolutionary biology studies (Pathak et al., 2019). The database contains 233 fish genomes, 201 fish transcriptomes, 5841 fish mitochondrial genomes, 88 fish gene sets, 16,239 miRNAs of 65 fishes, 1,330,692 piRNAs and 4852 lncRNAs of *Danio rerio*, 59,040 Mbuntranslated regions (UTR) of 230 fishes, and 31,918 Mb coding sequences (CDS) of 230 fishes. Among these a total of newely generated 11 fish genomes and 53 fish transcriptomes are also being included (Yang et al., 2020). FishDB is organised as follows:

FishDB provides web services such as a search engine, BLAST, JBrowse, and PrimerServer. Noncoding RNA (ncRNA), microRNA (miRNA), UTRs, and CDS gene information was collected and saved in the FishDB database. FishDB contains over 410,721.67 Mb sequences.

FisOmics

FisOmics is an online platform that features genomic databases and promotes information retrieval by querying these databases, analysing sequences, and sharing ASHAA for sequence/data analysis work(s). It contains:

HRGFish: A Database of Hypoxia-Responsive Genes in Fishes

HRGFish is a hypoxia-responsive gene database in fishes that includes a platform for assessing the functional annotation of gene, protein, mRNA, upstream sequences, and building primers for selected gene fragments by accumulating and curating information on the reported genes (Rashid et al., 2017).

Fish Karyome: A Chromosome Database of Fishes and Other Aquatic Organisms

'Fish Karyome' is a database that organises dispersed karyological knowledge about fishes. This database was created by gathering and organising existing chromosomal information from public repositories. It is a one-of-a-kind and useful online resource for cytogenetic characterisation, sex determination, chromosomal mapping, cytotaxonomy, karyoevolution, and fish systematics (Nagpure et al., 2016).

FMiR: Fish Mitogenome Resource

The FMiR database stores mitogenome sequences and characteristics for the purpose of examining mitogenome information of fishes reported worldwide. The system categorises and classifies the species based on general, biological, taxonomy, and conservation information, making it easier to browse mitogenome information according to the given

category (Nagpure et al., 2015).

FishMicrosat: Fish and Shellfish Microsatellite Database

The FishMicrosat database contains information on simple and compound microsatellites, as well as their clusters and locus orientation within sequences. This database is a valuable resource for fish and shellfish microsatellite sequence analysis and locus identification across species, with applications in population genetics, evolutionary studies, and species genetic relatedness (Nagpure et al., 2013).

FBIS: Fish Barcode Information System

FBIS was created to serve as a repository for fish species' DNA barcodes. Using the barcode sequence(s), FBIS provides a platform for identifying species and assessing genetic relatedness, taxonomic confirmation, and divergence within and between species (Nagpure et al., 2012).

5. Artificial Intelligence and Bioinformatics

Artificial Intelligence has been used to handle some of the most difficult bioinformatics tasks, including as protein structure prediction, homology search, multiple alignment and phylogeny construction, genomic sequence analysis, gene discovery, and more (Gasteiger, 2020).

Artificial Intelligence in Feeding Systems

The automated feeding system was the first application of artificial intelligence in fisheries (Barbedo, 2022). Even while freshwater aquaculture allows the farmer to stay close to the growing site, feeding the fish can be a time-consuming task. Farmers, too, can sometimes forget or become unavailable for various reasons. This is where AI can be immensely useful, stepping into the shoes of the cultivators via robotic feeding systems. AI can provide automatic feeds for the fish by utilising the technology of an automated feeding system (Wang et al., 2020). This is possible with AI-powered gadgets that release a predetermined amount of feed for fish to consume and sustain. This is also a great strategy to reduce food waste and optimise animal food intake (Vo et al., 2021).

a. An Indonesian aquaculture intelligence called 'eFishery' recently developed an AI feed dispenser that distributes the appropriate amount of feed at the appropriate moment. It detects the animal's appetite using a variety of sensors. The device can save roughly 21% on feed costs (Gladju et al., 2022).
b. 'Observe technologies' is a firm that develops artificial intelligence and data processing systems for measuring and tracking cattle feeding patterns (Neethirajan, 2020).
c. An aquaculture technology known as 'umitron cell' in Singapore and Japan manufactures a smart fish feeder that can be controlled by a remote (Rahimi-Midani., 2023; Das et al., 2022; Gladju et al., 2022).

Artificial Intelligence Drones Used in Aquaculture

Water quality parameters such as turbidity, temperature, dissolved oxygen, etc. and even heart rates of fishes can be monitored using drones (Choudhury et al., 2023; Wang et al., 2021). These data are easily accessible via a Smartphone linked to the drone. Shoal is a robotic fish drone which helps in analysing the water quality using low frequency sound waves (Setiyowati et al., 2022).

Prevention of Diseases

Diseases are the major thread to aquaculture.AI programmes can predict disease epidemics before they occur by comparing programmed data with site-collected data. They can even implement preventative measures. Norway's seafood innovation centre introduced 'Aquacloud' in April 2017, a cloud-based programme that assisted farmers in limiting the

development of sea lice in cages (Iden et al., 2021; Mustapha et al., 2021; Haukås, 2020; O'Donncha & Grant, 2019).

Screening of Fish Seeds

Identification of healthy fish seed is very important while starting aquaculture. Screening of pathogens, fish seed size etc has to be monitored properly before stocking (Rashid & Mithun, 2020). The Kindai University's Aquaculture Research Institute, Japan is using Microsoft Azure machine learning studio to identify and remove odd – shaped fish seed from the rearing cage (Lim, 2022).

Routine Check of Stocks

Vision based sensors are used for routine check of fish behaviour, swig patterns, size variation, any injuries etc. Xpertsea is an aquaculture innovation firm that provides an AI gadget called 'Xpercount' that uses machine learning and a camera to weigh, count, picture, and size prawns in seconds. These collected data are examined to determine the stock's periodic health (Das et al., 2022; Gladju et al., 2022; Mustapha et al., 2021; Antonucci & Costa, 2020).

Shrimp Culture Monitoring

Eruvaka, an Indian startup, offers shrimp farmers AI-based solutions such as real-time monitoring of water quality and voice call alert, appetite-based intelligent feeder, and autonomous control of aerators. Eruvaka's products have already been placed on around 1,000 hectares of shrimp farms scattered over Surat, Goa, Andhra Pradesh, and Pondicherry, and farmers are benefiting from AI-based shrimp farming solutions (Das et al., 2022).

Even though artificial intelligence provides many applications in the field of aquaculture but the investments are substantially costlier, and many people cannot afford them (Das et al., 2022; Mustapha et al., 2021; Duckett et al., 2018). AI system maintenance is also expensive. Another significant downside of AI is that it produces unemployment for workers. This may benefit farmers, but those who rely on fishing for a living will suffer.

CONCLUSION

In conclusion, the integration of bioinformatics into fisheries science offers unprecedented opportunities to advance our understanding and management of aquatic ecosystems. Through the use of genomic and computational tools, we can gain deeper insights into the genetic diversity, population dynamics, and evolutionary patterns of fish species. This knowledge is crucial for developing effective conservation strategies, sustainable fisheries management practices, and for addressing the challenges posed by climate change, habitat loss, and overfishing.

As bioinformatics continues to evolve, its application in fisheries is likely to expand, enabling more precise and data-driven decision-making. However, the success of these efforts will depend on continued interdisciplinary collaboration, investment in bioinformatics infrastructure, and the development of standardized methods for data analysis. By embracing these advancements, the fisheries sector can enhance its capacity to protect marine biodiversity and ensure the long-term viability of fishery resources for future generations.

ACKNOWLEDGEMENT

Author would like to thank all parties who have helped so that this research article can be completed properly. Our appreciation also extends to our colleagues and mentors for their

insightful guidance and constructive feedback throughout the research. Lastly, we thank our families and friends for their unwavering encouragement and support during this research process.

REFERENCES

Aging Atlas Consortium (2021). Aging Atlas: a multi-omics database for aging biology. *Nucleic acids research*, *49*(D1), D825–D830. https://doi.org/10.1093/nar/gkaa894

Albayraktaroglu, K., Jaleel, A., Wu, X., Franklin, M., Jacob, B., Tseng, C. W., & Yeung, D. (2005). Biobench: A benchmark suite of bioinformatics applications. In IEEE International Symposium on Performance Analysis of Systems and Software, 2005. ISPASS 2005. (pp. 2-9). IEEE.

Alexander, R. P., Fang, G., Rozowsky, J., Snyder, M., & Gerstein, M. B. (2010). Annotating non-coding regions of the genome. Nature Reviews Genetics, 11(8), 559-571.

Antonucci, F., & Costa, C. (2020). Precision aquaculture: a short review on engineering innovations. Aquaculture International, 28(1), 41-57.

Antony Dass, M., Sherman, C. D., Nai, Y. H., Ellis, M. R., van Oorschot, R. A., & Durdle, A. (2022). Assessing the use of environmental DNA (eDNA) as a tool in the detection of human DNA in water. Journal of Forensic Sciences, 67(6), 2299-2307.

Ardui, S., Ameur, A., Vermeesch, J. R., & Hestand, M. S. (2018). Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic acids research, 46(5), 2159-2168.

Barbedo, J. G. A. (2022). A review on the use of computer vision and artificial intelligence for fish recognition, monitoring, and management. Fishes, 7(6), 335.

Bayarsaikhan, D., Bayarsaikhan, G., & Lee, B. (2021). Recent advances in stem cells and gene editing: Drug discovery and therapeutics. Progress in Molecular Biology and Translational Science, 181, 231-269.

BlastN, G. (2019). BLAST: basic local alignment search tool. NUTRITIONAL AND PHYSIOLOGICAL DISORDERS IN HORTICULTURAL CROPS.

Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., ... & Finn, R. D. (2021). The InterPro protein families and domains database: 20 years on. Nucleic acids research, 49(D1), D344-D354.

Bora, J., Imam, S., Vaibhav, V., & Malik, S. (2023). Use of Genetic Engineering Approach in Bioremediation of Wastewater. In Modern Approaches in Waste Bioremediation: Environmental Microbiology (pp. 485-513). Cham: Springer International Publishing.

Büttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A., & Theis, F. J. (2019). A test metric for assessing single-cell RNA-seq batch correction. Nature methods, 16(1), 43-49.

Carter, N. P. (2007). Methods and strategies for analyzing copy number variation using DNA microarrays. Nature genetics, 39(Suppl 7), S16-S21.

Cho, W. C., & Cheng, C. H. (2007). Oncoproteomics: current trends and future perspectives. Expert review of proteomics, 4(3), 401-410.

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., ... & Xia, J. (2018). MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic acids research, 46(W1), W486-W494.

Choudhury, T. A., Kandra, H. S., Dinh, K., & Lim, S. (2023). Review of water quality monitoring using Internet of Things. Sustainable Civil Engineering: Principles and Applications, 309.

Cole, M. B., Risso, D., Wagner, A., DeTomaso, D., Ngai, J., Purdom, E., ... & Yosef, N. (2019). Performance assessment and selection of normalization procedures for single-cell RNA-seq. Cell systems, 8(4), 315-328.

Costa, V., Matarazzo, M. R., Gagliardi, M., Esposito, R., & Ciccodicola, A. (2016). High-Throughput Analysis of Noncoding RNAs: Implications in Clinical Epigenetics. In Epigenetic Biomarkers and Diagnostics (pp. 215-238). Academic Press.

Dadkhah, K., Anijdan, S. H. M., Karimi, M., Abolfazli, M. K., Rezaei, F., Adel, F., & Ataei, G. (2015). DNA microarray, types and its application in medicine. Scholars Academic Journal of Biosciences, 3(7), 598-599.

Das, B. K., Meena, D. K., Das, A., & Sahoo, A. K. (2022). Prospects of Smart Aquaculture in Indian Scenario: A New Horizon in the Management of Aquaculture Production Potential. In Smart and Sustainable Food Technologies (pp. 59-85). Singapore: Springer Nature Singapore.

De Las Rivas, J., & Fontanillo, C. (2010). Protein–protein interactions essentials: key concepts to building and analyzing interactome networks. PLoS computational biology, 6(6), e1000807.

Dereeper, A., Audic, S., Claverie, J. M., & Blanc, G. (2010). BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC evolutionary biology, 10(1), 1-6.

Desai, A., Thomas, N., & Bhat, V. (2021). Crispr Cas 9-a New Era in Genome Editing and its Application. Int. J. Livest. Res, 11(3), 17-24.

Dong, X., Tian, L., Gouil, Q., Kariyawasam, H., Su, S., De Paoli-Iseppi, R., ... & Ritchie, M. E. (2021). The long and the short of it: unlocking nanopore long-read RNA sequencing data with short-read differential expression analysis tools. NAR genomics and bioinformatics, 3(2), lqab028.

Duckett, T., Pearson, S., Blackmore, S., Grieve, B., Chen, W. H., Cielniak, G., ... & Yang, G. Z. (2018). Agricultural robotics: the future of robotic agriculture. arXiv preprint arXiv:1806.06762.

Ejigu, G. F., & Jung, J. (2020). Review on the computational genome annotation of sequences obtained by next-generation sequencing. Biology, 9(9), 295.

Evangelista, J. E., Xie, Z., Marino, G. B., Nguyen, N., Clarke, D. J., & Ma'ayan, A. (2023). Enrichr-KG: bridging enrichment analysis across multiple libraries. Nucleic Acids Research, gkad393.

Fajardo-Sanchez, E., Stricher, F., Paˆques, F., Isalan, M., & Serrano, L. (2008). Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences. Nucleic acids research, 36(7), 2163-2173.

Federico, A., Serra, A., Ha, M. K., Kohonen, P., Choi, J. S., Liampa, I., ... & Greco, D. (2020). Transcriptomics in toxicogenomics, part II: preprocessing and differential expression analysis for high quality data. Nanomaterials, 10(5), 903.

Ficetola, G. F., Manenti, R., & Taberlet, P. (2019). Environmental DNA and metabarcoding for the study of amphibians and reptiles: species distribution, the microbiome, and much more. Amphibia-Reptilia, 40(2), 129-148.

Gargan, L. M., Morato, T., Pham, C. K., Finarelli, J. A., Carlsson, J. E., & Carlsson, J. (2017). Development of a sensitive detection method to survey pelagic biodiversity using eDNA and quantitative PCR: a case study of devil ray at seamounts. Marine Biology, 164, 1-9.

Gasteiger, J. (2020). Chemistry in times of artificial intelligence. ChemPhysChem, 21(20), 2233-2242.

Geistlinger, L., Csaba, G., Santarelli, M., Ramos, M., Schiffer, L., Turaga, N., ... & Waldron, L. (2021). Toward a gold standard for benchmarking gene set enrichment analysis. Briefings in bioinformatics, 22(1), 545-556.

Gladju, J., Kamalam, B. S., & Kanagaraj, A. (2022). Applications of data mining and machine learning framework in aquaculture and fisheries: A review. Smart Agricultural Technology, 2, 100061.

González Castro, N., Bjelic, J., Malhotra, G., Huang, C., & Alsaffar, S. H. (2021). Comparison of the feasibility, efficiency, and safety of genome editing technologies. International Journal of Molecular Sciences, 22(19), 10355.

Grishin, A., Fonfara, I., Alexeevski, A., Spirin, S., Zanegina, O., Karyagina, A., ... & Wende, W. (2010). Identification of conserved features of LAGLIDADG homing endonucleases. Journal of bioinformatics and computational biology, 8(03), 453-469.

Gupta, S. K., & Shukla, P. (2017). Gene editing for cell engineering: trends and applications. Critical reviews in biotechnology, 37(5), 672-684.

Gutási, A., Hammer, S. E., El-Matbouli, M., & Saleh, M. (2023). Recent Applications of Gene Editing in Fish Species and Aquatic Medicine. Animals, 13(7), 1250.

Gutási, A., Hammer, S. E., El-Matbouli, M., & Saleh, M. (2023). Recent Applications of Gene Editing in Fish Species and Aquatic Medicine. Animals, 13(7), 1250.

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., ... & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome biology, 9, 1-22.

Haeussler, M., Zweig, A. S., Tyner, C., Speir, M. L., Rosenbloom, K. R., Raney, B. J., ... & Kent, W. J. (2019). The UCSC genome browser database: 2019 update. Nucleic acids research, 47(D1), D853-D858.

Hardin, C., Pogorelov, T. V., & Luthey-Schulten, Z. (2002). Ab initio protein structure prediction. Current opinion in structural biology, 12(2), 176-181.

Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. Genome biology, 18(1), 1-15.

Haukås, C. H. (2020). Preconditions to start and scale digital ecosystems: a study of aquacloud in the Norwegian seafood industry (Master's thesis).

Hawe, J. S., Theis, F. J., & Heinig, M. (2019). Inferring interaction networks from multi-omics data. Frontiers in genetics, 10, 535.

Haynes, D. (2018). Metadata for Information Management and Retrieval: Understanding metadata and its use. Facet Publishing.

Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., ... & Flicek, P. (2016). Ensembl comparative genomics resources. Database, 2016, bav096.

Hobbs, J., & Bright, D. (2016). Environmental DNA: implementation for resource development projects in BC and beyond.

Iden, J., Bygstad, B., Osmundsen, K. S., Costabile, C., & Øvrelid, E. (2021). Digital platform ecosystem governance: Preliminary findings and research agenda. In Norsk IKT-konferanse for forskning og utdanning (No. 2).

Iqbal, G., Quyoom, N., Singh, L. S., Ganpatbhai, A. V. K., Bhat, N. M., Gul, S., ... & Dar, S. A. (2023). Genome Editing Technology in Fishes.

Janicki, M., Rooke, R., & Yang, G. (2011). Bioinformatics and genomic analysis of transposable elements in eukaryotic genomes. Chromosome research, 19, 787-808.

Javed, M., Nadeem, A., & Hassan, F. U. (2021). Introduction to Molecular Genomics. Bentham Science Publishers.

Jia, Z., Liu, Y., Guan, N., Bo, X., Luo, Z., & Barnes, M. R. (2016). Cogena, a novel tool for co-expressed gene-set enrichment analysis, applied to drug repositioning and drug mode of action discovery. BMC genomics, 17, 1-11.

Jiao, W. B., & Schneeberger, K. (2017). The impact of third generation genomic technologies on plant genome assembly. Current opinion in plant biology, 36, 64-70.

Jing, W., & Yu, Z. (2021). Application of High-Throughput Sequencing Technology in New Coronary Pneumonia. MEDS Clinical Medicine, 2(1), 42-46.

Jo, T., Takao, K., & Minamoto, T. (2022). Linking the state of environmental DNA to its application for biomonitoring and stock assessment: Targeting mitochondrial/nuclear genes, and different DNA fragment lengths and particle sizes. Environmental DNA, 4(2), 271-283.

Johnson, K. A., & Krishnan, A. (2022). Robust normalization and transformation techniques for constructing gene coexpression networks from RNA-seq data. Genome biology, 23, 1-26.

Kalendar, R., Hunter, C., & Orbovic, V. (2022). Innovative applications of sequencing technologies in plant science. Frontiers in plant science, 13, 1058347.

Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. Protein Science, 28(11), 1947-1951.

Khalil, A. M. (2020). The genome editing revolution. Journal of genetic engineering and biotechnology, 18(1), 1-16.

Korf, I., Yandell, M., & Bedell, J. (2003). Blast. " O'Reilly Media, Inc."

Kuhlman, B., & Bradley, P. (2019). Advances in protein structure prediction and design. Nature Reviews Molecular Cell Biology, 20(11), 681-697.

Kumar, G., & Kocour, M. (2017). Applications of next-generation sequencing in fisheries research: A review. Fisheries Research, 186, 11-22.

Kumar, G., Reaume, A. M., Farrell, E., & Gaither, M. R. (2022). Comparing eDNA metabarcoding primers for assessing fish communities in a biodiverse estuary. PloS one, 17(6), e0266720.

Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. Protein science, 27(1), 129-134.

Lataretu, M., & Hölzer, M. (2020). RNAflow: An effective and simple RNA-seq differential gene expression pipeline using nextflow. Genes, 11(12), 1487.

LeBlanc, F., Belliveau, V., Watson, E., Coomber, C., Simard, N., DiBacco, C., ... & Gagné, N. (2020). Environmental DNA (eDNA) detection of marine aquatic invasive species (AIS) in Eastern Canada using a targeted species-specific qPCR approach. Management of Biological Invasions, 11(2).

Lee, J., Freddolino, P. L., & Zhang, Y. (2017). Ab initio protein structure prediction. From protein structure to function with bioinformatics, 3-35.

Li, H., Brouwer, C. R., & Luo, W. (2022). A universal deep neural network for in-depth cleaning of single-cell RNA-Seq data. Nature Communications, 13(1), 1901.

Lim, L. W. K. (2022). Implementation of Artificial Intelligence in Aquaculture and Fisheries: Deep Learning, Machine Vision, Big Data, Internet of Things, Robots and Beyond. Journal of Computational and Cognitive Engineering.

Lindsay, S. (2016). The promises and challenges of solid-state sequencing. Nature nanotechnology, 11(2), 109-111.

Liu, Z. J. (Ed.). (2017). Bioinformatics in aquaculture: principles and methods. John Wiley & Sons.

Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. Methods of information in medicine, 40(04), 346-358.

MacCannell, D. (2019). Platforms and analytical tools used in nucleic acid sequence-based microbial genotyping procedures. Microbiology spectrum, 7(1), 10-1128.

Madden, T. (2003). The BLAST sequence analysis tool. The NCBI handbook.

Mahfouz, M. M., Piatek, A., & Stewart Jr, C. N. (2014). Genome engineering via TALENs and CRISPR/Cas9 systems: challenges and perspectives. Plant biotechnology journal, 12(8), 1006-1014.

Márquez, J. R. G., Dormann, C. F., Sommer, J. H., Schmidt, M., Thiombiano, A., Da, S. S., ... & Barthlott, W. (2012). A methodological framework to quantify the spatial quality of biological databases. Biodiversity and Ecology, 4, 25-39.

Miglani, G. S. (2017). Genome editing in crop improvement: Present scenario and future prospects. Journal of Crop Improvement, 31(4), 453-559.

Molidor, R., Sturn, A., Maurer, M., & Trajanoski, Z. (2003). New trends in bioinformatics: from genome sequence to personalized medicine. Experimental gerontology, 38(10), 1031-1036.

Moushomi, R., Wilgar, G., Carvalho, G., Creer, S., & Seymour, M. (2019). Environmental DNA size sorting and degradation experiment indicates the state of Daphnia magna mitochondrial and nuclear eDNA is subcellular. Scientific Reports, 9(1), 12500.

Mustapha, U. F., Alhassan, A. W., Jiang, D. N., & Li, G. L. (2021). Sustainable aquaculture development: a review on the roles of cloud computing, internet of things and artificial intelligence (CIA). Reviews in Aquaculture, 13(4), 2076-2091.

Mustapha, U. F., Alhassan, A. W., Jiang, D. N., & Li, G. L. (2021). Sustainable aquaculture development: a review on the roles of cloud computing, internet of things and artificial intelligence (CIA). Reviews in Aquaculture, 13(4), 2076-2091.

Nagpure, N. S., Pathak, A. K., Pati, R., Rashid, I., Sharma, J., Singh, S. P., ... & Murali, S. (2016). Fish Karyome version 2.1: a chromosome database of fishes and other aquatic organisms. Database, 2016, baw012.

Nagpure, N. S., Rashid, I., Pathak, A. K., Singh, M., Pati, R., Singh, S. P., & Sarkar, U. K. (2015). FMiR: a curated resource of mitochondrial DNA information for fish. PLoS One, 10(8), e0136711.

Nagpure, N. S., Rashid, I., Pati, R., Pathak, A. K., Singh, M., Singh, S. P., & Sarkar, U. K. (2013). FishMicrosat: a microsatellite database of commercially important fishes and shellfishes of the Indian subcontinent. BMC genomics, 14, 1-10.

Neelapu, N. R. R., & Surekha, C. (2016). Next-generation sequencing and metagenomics. Computational biology and bioinformatics: gene regulation. CRC Press, Boca Raton, 331-351.

Neethirajan, S. (2020). The role of sensors, big data and machine learning in modern animal farming. Sensing and Bio-Sensing Research, 29, 100367.

O'Shea, K., & Misra, B. B. (2020). Software tools, databases and resources in metabolomics: updates from 2018 to 2019. Metabolomics, 16(3), 36.

O'Donncha, F., & Grant, J. (2019). Precision aquaculture. IEEE Internet of Things Magazine, 2(4), 26-30.

Pathak, A. K., Rashid, I., Nagpure, N. S., Kumar, R., Pati, R., Singh, M., ... & Rai, A. (2019). FisOmics: A portal of fish genomic resources. Genomics, 111(6), 1923-1928.

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A. Bateman, A. (2023). InterPro in 2022. Nucleic acids research, 51(D1), D418-D427.

Pearson, W. R. (2014). BLAST and FASTA similarity searching for multiple sequence alignment. Multiple sequence alignment methods, 75-101.

Pearson, W. R. (2016). Finding protein and nucleotide similarities with FASTA. Current protocols in bioinformatics, 53(1), 3-9.

Pereira, M. B., Wallroth, M., Jonsson, V., & Kristiansson, E. (2018). Comparison of normalization methods for the analysis of metagenomic gene abundance data. BMC genomics, 19, 1-17.

Pettersson, E., Lundeberg, J., & Ahmadian, A. (2009). Generations of sequencing technologies. Genomics, 93(2), 105-111.

Pickar-Oliver, A., & Gersbach, C. A. (2019). The next generation of CRISPR–Cas technologies and applications. Nature reviews Molecular cell biology, 20(8), 490-507.

Poetsch, A., & Li, Y. (2023). 3.1 Multi-Omics Data Types and Underlying Technology. Methodologies of Multi-Omics Data Integration and Data Mining: Techniques and Applications, 19, 33.

Rahimi-Midani, A. (2023). Deep Tech Practices in Aquaculture. In Deep Technology for Sustainable Fisheries and Aquaculture (pp. 17-60). Singapore: Springer Nature Singapore.

Rai, A., Bhati, J., & Lal, S. B. (2012). Software tools and resources for bioinformatics research. Appl. Comput. Biol. Stat. Biotechnol. Bioinforma, 1.

Ramirez-Amaro, S., Bassitta, M., Picornell, A., Ramon, C., & Terrasa, B. (2022). Environmental DNA: State-of-the-art of its application for fisheries assessment in marine environments. Frontiers in Marine Science, 9, 1004674.

Rana, K. M. S., Ahammad, K., & Salam, M. A. (2020). Bioinformatics: scope and challenges in aquaculture research of Bangladesh-a review. International Journal of Agricultural Research, Innovation and Technology (IJARIT), 10(2), 137-145.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., ... & Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome biology, 14(9), 1-13.

Rashid, I., & Mithun, M. H. (2020). Guidelines for freshwater pond management in Bangladesh. Bangladesh Journal of Multidisciplinary Scientific Research, 2(2), 1-9.

Rashid, I., Nagpure, N. S., Srivastava, P., Kumar, R., Pathak, A. K., Singh, M., & Kushwaha, B. (2017). HRGFish: A database of hypoxia responsive genes in fishes. Scientific Reports, 7(1), 42346.

Raut, S. A., Sathe, S. R., & Raut, A. (2010, April). Bioinformatics: Trends in gene expression analysis. In 2010 International Conference on Bioinformatics and Biomedical Technology (pp. 97-100). IEEE.

Rawat, P. (2017). Next-Generation Sequencing Technologies for Real-Time Genotyping and Targeted Sequencing Tailored for Precision Medicine (Doctoral dissertation, The University of New Mexico).

Raza, K. (2012). Application of data mining in bioinformatics. arXiv preprint arXiv:1205.1125.

Rees, H. A., & Liu, D. R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. Nature reviews genetics, 19(12), 770-788.

Rehm, B. (2001). Bioinformatic tools for DNA/protein sequence analysis, functional assignment of genes and protein classification. Applied microbiology and biotechnology, 57, 579-592.

Reuter, J. A., Spacek, D. V., & Snyder, M. P. (2015). High-throughput sequencing technologies. Molecular cell, 58(4), 586-597.

Rice, P. M., Rice, P. M., Bleasby, A. J., & Ison, J. C. (2011). EMBOSS User's Guide: Practical Bioinformatics with EMBOSS. Cambridge University Press.

Risso, D., Schwartz, K., Sherlock, G., & Dudoit, S. (2011). GC-content normalization for RNA-Seq data. BMC bioinformatics, 12, 1-17.

Rizzi, R., Beretta, S., Patterson, M., Pirola, Y., Previtali, M., Della Vedova, G., & Bonizzoni, P. (2019). Overlap graphs and de Bruijn graphs: data structures for de novo genome assembly in the big data era. Quantitative Biology, 7, 278-292.

Ross, H. A., Murugan, S., & Sibon Li, W. L. (2008). Testing the reliability of genetic methods of species identification via simulation. Systematic biology, 57(2), 216-230.

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. Global Ecology and Conservation, 17, e00547.

Setiyowati, H., Thalib, S., Setiawati, R., Nurjannah, N., & Akbariani, N. V. (2022). AN AQUACULTURE DISRUPTED BY DIGITAL TECHNOLOGY. AUSTENIT, 14(1), 12-16.

Sharma, D., Singh, V. P., Singh, R. K., Joshi, C. S., & Sharma, V. (2021). Isolation and characterization of bioactive compounds from natural resources: Metabolomics and molecular approaches. In Evolutionary Diversity as a Source for Anticancer Molecules (pp. 77-101). Academic Press.

Sharma, P. K., & Yadav, I. S. (2022). Biological databases and their application. In Bioinformatics (pp. 17-31). Academic Press.

Shendure, J. A., Porreca, G. J., Church, G. M., Gardner, A. F., Hendrickson, C. L., Kieleczawa, J., & Slatko, B. E. (2011). Overview of DNA sequencing strategies. Current Protocols in Molecular Biology, 96(1), 7-1.

Shi, H., Maier, S., Nimmrich, I., Yan, P. S., Caldwell, C. W., Olek, A., & Huang, T. H. M. (2003). Oligonucleotide-based microarray for DNA methylation analysis: principles and applications. Journal of cellular biochemistry, 88(1), 138-143.

Sigrist, C. J., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Bulliard, V., Bairoch, A., & Hulo, N. (2010). PROSITE, a protein domain database for functional characterization and annotation. Nucleic acids research, 38(suppl_1), D161-D166.

Singh, V. K., Singh, A. K., Chand, R., & Kushwaha, C. (2011). Role of bioinformatics in agriculture and sustainable development. Int J Bioinforma Res, 3(2), 221-226.

Southern, E. M. (2001). DNA microarrays: history and overview. DNA arrays: Methods and Protocols, 1-15.

Sprink, T., Metje, J., & Hartung, F. (2015). Plant genome editing by novel tools: TALEN and other sequence specific nucleases. Current opinion in biotechnology, 32, 47-53.

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Biophysical Journal, 106(2), 695a.

Su, S., Law, C. W., Ah-Cann, C., Asselin-Labat, M. L., Blewitt, M. E., & Ritchie, M. E. (2017). Glimma: interactive graphics for gene expression analysis. Bioinformatics, 33(13), 2050-2052.

Sun, J., Nishiyama, T., Shimizu, K., & Kadota, K. (2013). TCC: an R package for comparing tag count data with robust normalization strategies. BMC bioinformatics, 14(1), 1-14.

Taguchi, T., Ishikawa, M., Ichikawa, M., Tadenuma, T., Hirakawa, Y., Yoshino, T., ... & Tanaka, T. (2021). Amplification-free detection of bacterial genes using a signaling probe-based DNA microarray. Biosensors and Bioelectronics, 194, 113659.

Tolani, P., Gupta, S., Yadav, K., Aggarwal, S., & Yadav, A. K. (2021). Big data, integrative omics and network biology. Advances in Protein Chemistry and Structural Biology, 127, 127-160.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols, 7(3), 562-578.

Uusi-Mäkelä, M. (2022). Zebrafish as a Model for Human Genetic and Infectious Diseases.

Vallejos, C. A., Risso, D., Scialdone, A., Dudoit, S., & Marioni, J. C. (2017). Normalizing single-cell RNA sequencing data: challenges and opportunities. Nature methods, 14(6), 565-571.

Vo, T. T. E., Ko, H., Huh, J. H., & Kim, Y. (2021). Overview of smart aquaculture system: Focusing on applications of machine learning and computer vision. Electronics, 10(22), 2882.

Von Hippel, P. H., Johnson, N. P., & Marcus, A. H. (2013). Fifty years of DNA "breathing": Reflections on old and new approaches. Biopolymers, 99(12), 923-954.

Wada, N., Osakabe, K., & Osakabe, Y. (2022). Expanding the plant genome editing toolbox with recently developed CRISPR–Cas systems. Plant Physiology, 188(4), 1825-1837.

Wang, C., Li, Z., Wang, T., Xu, X., Zhang, X., & Li, D. (2021). Intelligent fish farm—the future of aquaculture. Aquaculture International, 1-31.

Wang, W., Gao, Z., Wang, H., Li, T., He, W., Lv, W., & Zhang, J. (2016). Transcriptome analysis reveals distinct gene expression profiles in eosinophilic and noneosinophilic chronic rhinosinusitis with nasal polyps. Scientific Reports, 6(1), 26604.

Wang, Y., Li, X. Y., Xu, W. J., Wang, K., Wu, B., Xu, M., ... & Gui, J. F. (2022). Comparative genome anatomy reveals evolutionary insights into a unique amphitriploid fish. Nature Ecology & Evolution, 6(9), 1354-1366.

Webb, B., & Sali, A. (2016). Comparative protein structure modeling using MODELLER. Current protocols in bioinformatics, 54(1), 5-6.

Weissenbach, J. (2016). The rise of genomics. Comptes rendus biologies, 339(7-8), 231-239.

Whisstock, J. C., & Lesk, A. M. (2003). Prediction of protein function from protein sequence and structure. Quarterly reviews of biophysics, 36(3), 307-340.

Wilson, S. L., Way, G. P., Bittremieux, W., Armache, J. P., Haendel, M. A., & Hoffman, M. M. (2021). Sharing biological data: why, when, and how. FEBS letters, 595(7), 847.

Wolf, J. B. (2013). Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. Molecular ecology resources, 13(4), 559-572.

Wu, C. H., Apweiler, R., Bairoch, A., Natale, D. A., Barker, W. C., Boeckmann, Suzek, B. (2006). The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic acids research, 34(suppl_1), D187-D191.

Xiang, C. C., & Chen, Y. (2000). cDNA microarray technology and its applications. Biotechnology Advances, 18(1), 35-46.

Xu, Y., Verma, D., Sheridan, R. P., Liaw, A., Ma, J., Marshall, N. M., ... & Johnston, J. M. (2020). Deep dive into machine learning models for protein engineering. Journal of chemical information and modeling, 60(6), 2773-2790.

Yang, J., & Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. Nucleic acids research, 43(W1), W174-W181.

Yang, L., Xu, Z., Zeng, H., Sun, N., Wu, B., Wang, C., ... & He, S. (2020). FishDB: an integrated functional genomics database for fishes. BMC genomics, 21(1), 1-5.

Yi, H., & Chun, J. (2015). Next generation sequencing-based metagenomics for monitoring soil microbiota. Biosafety and The Environmental uses of Microorganisms; OECD: Paris, France.

Zeng, I. S. L., & Lumley, T. (2018). Review of statistical learning methods in integrated omics studies (an integrated information science). Bioinformatics and biology insights, 12, 1177932218759292.

Zhang, J., Gao, S., Shi, Y., Yan, Y., & Liu, Q. (2022). Full-length transcriptome of anadromous Coilia nasus using single molecule real-time (SMRT) sequencing. Aquaculture and Fisheries, 7(4), 420-426.

Zhang, Y., Lou, F., Chen, J., Han, Z., Yang, T., Gao, T., & Song, N. (2022). Single-molecule Real-time (SMRT) Sequencing Facilitates Transcriptome Research and Genome Annotation of the Fish Sillago sinica. Marine Biotechnology, 24(5), 1002-1013.

Zhao, Y., Wong, L., & Goh, W. W. B. (2020). How to do quantile normalization correctly for gene expression data analyses. Scientific reports, 10(1), 15534.

Zhou, D., Rao, M. S., Walker, R., Khrebtukova, I., Haudenschild, C. D., Miura, T., ... & Vasicek, T. J. (2006). Massively parallel signature sequencing. Human Embryonic Stem Cell Protocols, 285-311.