NEXT GENERATION SEQUENCING: A REVOLUTION TECHNOLOGY IN FISHERIES AND ENVIRONMENTAL DNA (EDNA) STUDY

Abstract

Next-generation sequencing (NGS) technologies have revolutionized the field of genomics and have significantly impacted various disciplines, including fisheries and aquaculture. The technology is used to determine the order of nucleotides in entire genomes or targeted regions of DNA or RNA. NGS platforms provide researchers with unprecedented capabilities to unravel the genetic makeup of species, investigate genetic diversity, identify molecular markers, and understand the genetic basis of important traits. In this paper, we explore the diverse applications of NGS in fisheries and aquaculture, including genome sequencing, transcriptomics, metagenomics, and population genetics. We also discuss the potential of NGS in addressing challenges faced by the industry, such as disease management, selective breeding, and conservation efforts. Finally, we provide a comprehensive conclusion summarizing the key findings and future prospects of NGS in fisheries and aquaculture.

Keywords: Sequence Of DNA, RNA to study genetic variation, scalability, and speed, Environmental DNA

Authors

Ashish Sahu
Faculty of fisheries
Kerala University of Fisheries & Ocean Studies
Panangad, Cochin, Kerala, India.
Sahu81862@gmail;.com

Durgesh Verma
ICAR-Central Inland Fisheries Research Institute, Regional Center
Prayagraj, Uttar Pradesh, India.

Satendra Kumar
College of Fisheries Science
Chaudhary Charan Singh Haryana Agricultural University
Hisar, Haryana, India.

Shiva Ji
College of Fisheries
Acharya Narendra Deva University of Agriculture and Technology
Kumarganj, Ayodhya, Uttar Pradesh, India.

Sagar Satkar
Faculty of fisheries
Kerala University of Fisheries & Ocean Studies
Panangad, Cochin, Kerala, India.

Saiprasad Bhusare
Fish Nutrition
Biochemistry and Physiology Division
ICAR-Central Institute of Fisheries Education, Mumbai, India.

Neelesh Kumar
College of Fisheries, G. B. Pant University of Agriculture and Technology
Pantnagar, Uttarakhand, India.

I. INTRODUCTION

Next-generation sequencing, also known as high-throughput sequencing, is a powerful set of technologies that have revolutionized the genomics sector. It enables the rapid and cost-effective sequencing of DNA or RNA molecules on a massive scale, providing researchers with unprecedented capabilities to explore genetic information in numerous flora and fauna. The development of NGS platforms, such as Illumina's sequencing-by-synthesis, Roche's pyrosequencing, and Ion Torrent's semiconductor sequencing, has significantly increased the speed, accuracy, and throughput of DNA sequencing compared to traditional Sanger sequencing methods. These advancements have opened up new avenues for scientific discovery and have found extensive applications in numerous fields, including biomedical research, agriculture, ecology, and forensics.

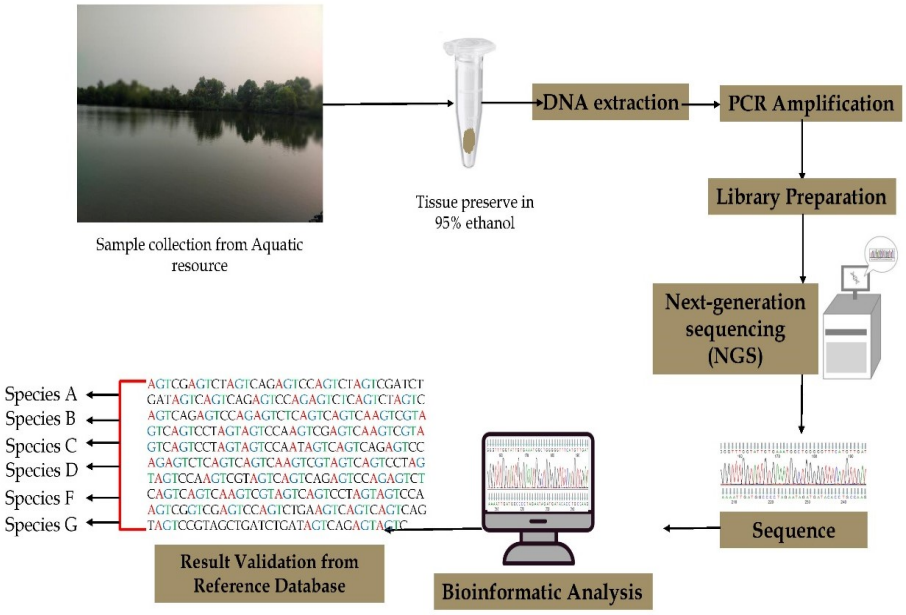

Figure 1. Concept of Next-generation sequencing

II. GENERATION OF DNA SEQUENCING

Sequencing technologies have undergone remarkable advancements over the years, leading to increased speed, accuracy, and cost-effectiveness. Here's an overview of the evolution of DNA sequencing technologies:

1. **Sanger Sequencing (1977):** Sanger sequencing, also known as dideoxy or chain termination sequencing, was the first widely used sequencing method. It relies on the incorporation of chain-terminating nucleotides during DNA replication. Sanger sequencing played a pivotal role in the Human Genome Project and other early sequencing efforts.

2. **Next-Generation Sequencing (2005 onwards):** NGS technologies, also referred to as second-generation sequencing, emerged in the mid-2000s. They revolutionized the field of genomics by enabling high-throughput, massively parallel sequencing of DNA fragments. The key NGS platforms include Roche 454, Illumina Genome Analyzer, Ion Torrent, and SOLiD. These technologies dramatically reduced the cost and time required for sequencing and allowed for large-scale genomics research.

3. **Third-Generation Sequencing (2010 onwards):** Third-generation sequencing technologies offer long-read sequencing capabilities, which provide a significant advantage in resolving complex genomic regions and structural variations. These technologies include Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT). PacBio utilizes single-molecule, real-time (SMRT) sequencing, while ONT employs nanopore-based sequencing. Third-generation sequencing has been instrumental in completing de novo genome assemblies, characterizing repetitive regions, and studying epigenetic modifications.

4. **Fourth-Generation Sequencing (emerging):** Fourth-generation sequencing technologies are still in the early stages of development and are expected to further enhance sequencing capabilities. These technologies aim to overcome the limitations of previous generations by providing higher throughput, longer reads, and improved accuracy. Several innovative approaches are being explored, such as synthetic nanopores, single-molecule imaging, and novel chemistry formulations.

5. **Single-Cell Sequencing:** Single-cell sequencing has gained prominence as a specialized application within sequencing technologies. It allows for the analysis of individual cells, providing insights into cellular heterogeneity, rare cell populations, and cellular dynamics. Different sequencing platforms, including NGS and emerging technologies, have been adapted for single-cell sequencing, enabling researchers to unravel the complexities of cell types and cell states.

III. ADVANTAGES OF NGS

NGS Technologies offer several key advantages over traditional sequencing approaches making them a preferred choice for genomics research and various applications. The key advantages of NGS include

1. **High Throughput:** NGS able to read millions to billions of sequences of DNA in a single run. This high throughput enables the analysis of large genomes, comprehensive profiling of genetic variants, and the study of complex biological systems. This capacity has facilitated the generation of comprehensive catalogue of genetic variants, including single nucleotide polymorphisms (SNPs), insertions, deletions, and structural variations, which are essential for understanding the genetic basis of diseases, traits, and evolutionary processes.

2. **Speed and Efficiency:** NGS technologies dramatically accelerate the sequencing process compared to traditional methods. Sequencing runs that previously took weeks or months can now be completed within hours or days. This speed allows researchers to obtain results quickly and facilitates the execution of large-scale projects.

3. **Cost-Effectiveness:** NGS has significantly reduced the cost per base of DNA sequencing, enabling broader accessibility and affordability. The decreased cost has democratized genomics research and facilitated the expansion of sequencing projects with large sample sizes or extensive coverage.

4. **Scalability:** NGS platforms are highly scalable, accommodating a wide range of project sizes and sequencing needs. Researchers can scale up or down their sequencing experiments based on the specific requirements, from targeted gene panels to whole-genome sequencing. This scalability allows for flexibility in experimental design and cost optimization.

5. **Data Quantity and Quality:** NGS generates vast amounts of sequencing data, providing comprehensive coverage and increasing the statistical power of genomic analysis. The high read depth obtained from NGS platforms enhances the accuracy of variant calling, enabling the detection of rare genetic variants and low-frequency mutations.

6. **Versatility:** NGS technologies are versatile and can be applied to various genomic applications, including whole-genome sequencing, targeted sequencing, transcriptomics, epigenomics, metagenomics, etc. This versatility allows researchers to explore diverse aspects of genomics and address different research questions using a single technology.

7. **Discoverability:** NGS enables unbiased sequencing, which means that it can detect both known and unknown genetic variants. This unbiased approach increases the likelihood of discovering novel genomic features, rare mutations, and unexpected genetic phenomena. It facilitates the identification of genetic factors associated with diseases, traits, and evolutionary processes.

8. **Flexibility in Experimental Design:** NGS platforms offer flexibility in experimental design, allowing researchers to tailor sequencing approaches to specific research goals. Researchers can select different sequencing depths, read lengths, and library preparation methods based on the requirements of their study, optimizing the sequencing strategy for cost-effectiveness and data quality.

9. **Integration with Bioinformatics:** NGS generates massive amounts of raw sequencing data, and its integration with bioinformatics tools and computational analysis pipelines has become an integral part of the sequencing process. The availability of bioinformatics tools and databases specifically designed for NGS data analysis enables efficient data processing, variant calling, annotation, and interpretation.

10. **Technological Advancements:** NGS technologies continue to evolve rapidly, with advancements in sequencing chemistry, platform performance, read length, and error rates. Continuous improvements in NGS platforms and protocols enhance sequencing accuracy, reduce bias, and expand the range of applications, further increasing the value and utility of NGS in genomics research.

IV. DISADVANTAGES OF NGS

NGS also have certain disadvantages that researchers should consider. The disadvantages of NGS include:

1. **Complexity of Data Analysis:** NGS generates large volumes of sequencing data, often requiring sophisticated bioinformatics analysis pipelines and computational resources. The analysis and interpretation of NGS data can be computationally intensive and require specialized bioinformatics expertise, which may pose challenges for researchers who are not well-versed in data analysis or lack access to appropriate computational infrastructure.

2. **Error Rates and Accuracy:** Despite advancements in sequencing technologies, NGS platforms are not completely error-free. Sequencing errors can occur during library preparation, amplification, or base calling, resulting in incorrect base identification or low-level background noise. Error rates vary among NGS platforms and sequencing chemistries, and certain genomic regions, such as repetitive or GC-rich regions, can be more prone to errors or sequencing biases.

3. **Short Read Lengths:** NGS platforms typically produce short read lengths compared to traditional Sanger sequencing. Although read lengths have improved over time, they may still be insufficient for certain applications, such as de novo genome assembly or the accurate identification of structural variations. Short read lengths can limit the ability to accurately resolve repetitive regions, complex genomic rearrangements, or haplotype phasing.

4. **Reference Genome Bias:** Most NGS-based analyses rely on a reference genome for read alignment and variant calling. However, reference genomes may not represent the full diversity of a given species, particularly for non-model organisms or populations with high genetic variation. This reference bias can lead to misalignments, inaccurate variant calling, or the inability to detect novel or rare variants not present in the reference.

5. **Coverage and Depth Variability:** NGS platforms generate read coverage that may not be uniform across the genome, resulting in regions with lower coverage or uneven distribution of reads. Uneven coverage can affect the detection of variants, particularly in low-covered regions, and may require additional sequencing or alternative strategies to improve coverage depth.

6. **Library Preparation Artifacts and Biases:** The library preparation step in NGS workflows can introduce biases and artifacts. Variability in library preparation efficiency can lead to uneven representation of genomic regions, affecting the accuracy of variant calling and gene expression quantification. Additionally, certain library preparation methods may introduce biases, such as PCR amplification bias or GC-content bias, which can impact downstream analysis and interpretation.

7. **Repetitive and Structural Variations:** NGS technologies have limitations in accurately resolving repetitive regions and complex structural variations in the genome. Highly repetitive sequences can hinder read alignment and lead to ambiguous mapping or

misinterpretation of genomic features. Additionally, the detection of large-scale structural variations, such as inversions or translocations, may require complementary techniques, such as long-read sequencing or optical mapping.

8. **Sample Contamination and DNA Quality:** NGS results can be influenced by sample contamination or degraded DNA quality. Contamination by exogenous DNA during sample collection, extraction, or library preparation can introduce errors and compromise the accuracy of downstream analysis. Similarly, degraded or low-quality DNA samples can result in increased sequencing errors, reduced coverage, and lower data quality.

9. **Cost Considerations:** While NGS has become more affordable compared to traditional sequencing methods, it still requires a significant investment in equipment, reagents, and computational resources. The cost of NGS experiments can vary depending on the desired coverage, read depth, and data analysis requirements. Researchers must carefully consider the cost implications of NGS and balance their experimental design and budget constraints.

10. **Ethical and Privacy Considerations:** NGS technologies can generate vast amounts of personal genomic data, raising ethical and privacy concerns. Proper data management, informed consent, and protection of sensitive information are crucial to ensure the responsible use and storage of genomic data generated by NGS.

V. VERSATILE APPLICATIONS OF NEXT-GENERATION SEQUENCING IN DIFFERENT FIELDS

- Rapidly sequence whole genomes
- Deeply sequence target regions
- Utilize RNA sequencing (RNA-Seq) to discover novel RNA variants and splice sites, or quantify mRNAs for gene expression analysis
- Analyse epigenetic factors such as genome-wide DNA methylation and DNA-protein interactions
- Sequence cancer samples to study rare somatic variants, tumor subclones, and more
- Study the human microbiome
- Identify novel pathogens.

VI. WORKFLOW OF NGS

Certainly! Here's an illustration of how NGS can be utilized in fisheries:

1. **Sample Collection:** we have to collect tissue samples (e.g., fin clips, muscle tissue) or environmental samples (e.g., water, sediment) from fish populations or aquatic environments of interest.

2. **DNA Extraction:** The collected samples undergo DNA extraction to isolate the genetic material (DNA) from the tissues or environmental samples. Various DNA extraction methods are available depending on the sample type and the downstream sequencing applications.

3. **Library Preparation:** NGS requires the preparation of DNA libraries, which involve fragmenting the extracted DNA and attaching specific adapters to the fragments. This step prepares the DNA for sequencing.

4. **DNA Sequencing:** The prepared DNA libraries are loaded onto an NGS platform, such as Illumina, Ion Torrent, or Oxford Nanopore Technologies. The sequencing platform then performs the sequencing reaction, generating millions to billions of short DNA reads (fragments) or long-read sequences, depending on the technology used.

5. **Data Analysis:** The generated sequencing data undergoes bioinformatics analysis. This step includes quality control, read alignment to a reference genome (if available), de novo assembly (if reference genome is not available), variant calling, and downstream analyses.

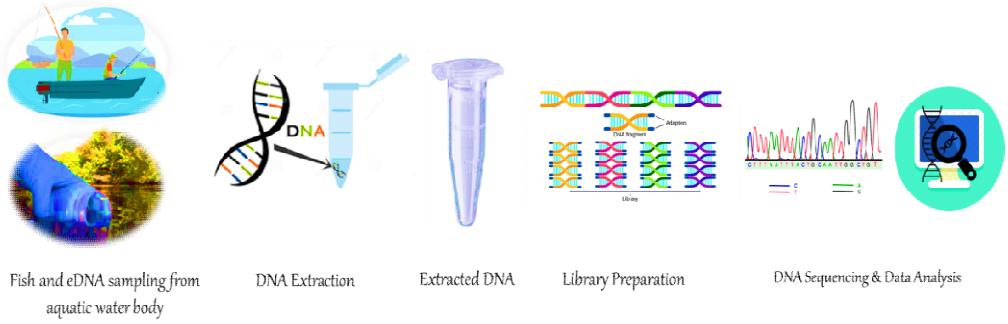

Figure 2: Schematic diagram representing Meta bar-coding steps involved in eDNA bar-coding

VII. APPLICATION NGS IN FISHERIES AND EDNA

Next-generation sequencing has revolutionized the field of genomics and has had a significant impact on fisheries research and management. Here's a brief overview of the history of NGS in fisheries:

1. **Early Adoption of NGS Technologies:** The first NGS platforms, such as Roche 454 and Illumina Genome Analyser, were introduced in the mid-2000s. Fisheries researchers quickly recognized the potential of these technologies to generate large amounts of genomic data in a cost-effective manner and started adopting them for various applications.

2. **Genome Sequencing of Fishes:** NGS technologies enabled the rapid and cost-effective sequencing of the genomes of several model fish species. Model organisms like zebrafish (*Danio rerio*) and medaka (*Oryziaslatipes*) have been extensively sequenced, providing a rich genomic resource for comparative genomics and understanding the genetics of fish development and physiology.

3. **Population Genomics and Phylogenetics:** NGS has played a crucial role in advancing our understanding of the population structure, genetic diversity, and evolutionary relationships of fish species. Researchers have used NGS to generate genome-wide datasets to investigate population dynamics, phylogeography, and speciation in various fish species. These studies have provided insights into the effects of habitat fragmentation, climate change, and human activities on fish populations.

4. **Transcriptomics and Gene Expression Analysis:** NGS technologies have facilitated transcriptome sequencing (RNA-seq) in fisheries research. RNA-seq allows for the analysis of gene expression patterns in different tissues or under different conditions, providing insights into the molecular mechanisms underlying important traits, such as growth, reproduction, and immune response in fish species. It has also been used to identify candidate genes associated with economically valuable traits in aquaculture.

5. **Genomic Selection in Aquaculture:** NGS has facilitated the implementation of genomic selection strategies in aquaculture. By genotyping large numbers of individuals using NGS-based genotyping platforms, researchers can identify genetic markers associated with economically important traits, such as disease resistance, growth rate, and fillet quality. This information can then be used to make selective breeding decisions, accelerating the genetic improvement of aquaculture species.

6. **Environmental DNA (eDNA) Metabarcoding:** NGS is panacea for eDNA metabarcoding because it can read short fragments in environmental samples. Researchers can detect and identify fish species, which are present in aquatic ecosystems without using any kind of fishing. NGS involves the DNA extraction and sequencing of DNA from environmental samples such as soil, water, or air. Environmental DNA analysis has been used for biodiversity monitoring, species detection, detecting invasive species, rare or elusive fish species and ecological studies also. We are elaborating the application of NGS use in eDNA research;

7. **Species Detection and Identification:** NGS allows for the simultaneous sequencing of DNA from multiple organisms present in an environmental sample. By targeting specific genetic markers such as the mitochondrial DNA (e.g., COI gene) or ribosomal DNA (e.g., 12S, 16S rRNA gene), it is possible to detect and identify various species within a community. This approach is particularly useful for detecting rare, elusive, or endangered species that may be difficult to observe directly.

8. **Biodiversity Assessment:** NGS-based eDNA analysis provides a comprehensive view of biodiversity within an ecosystem. By sequencing DNA from a sample, researchers can estimate species richness, abundance, and even assess the composition and structure of ecological communities. This information is valuable for understanding ecosystem health, assessing the impacts of environmental changes, and guiding conservation efforts.

9. **Invasive Species Monitoring:** NGS can be used to detect and monitor invasive species in different environments. By analysing eDNA samples from water bodies, soil, or other habitats, researchers can identify the presence of invasive species and track their spread. This information is crucial for implementing timely management strategies to control and mitigate the impacts of invasive species on native ecosystems.

10. **Environmental Monitoring and Assessment:** NGS-based eDNA analysis enables the monitoring of environmental conditions and changes over time. By analysing DNA from environmental samples, researchers can assess the presence of pathogens, pollutants, or indicators of environmental stress. This information aids in understanding ecosystem dynamics, identifying emerging threats, and guiding environmental management and remediation efforts.

11. **Metagenomics and Community Ecology:** NGS allow for the analysis of entire microbial communities present in environmental samples, known as metagenomics. By sequencing and analysing the DNA of bacteria, fungi, viruses, and other microorganisms, researchers can study community structure, functional diversity, and ecological interactions. This information contributes to our understanding of microbial ecology, nutrient cycling, and the role of microorganisms in ecosystem functioning.

12. **Genome Sequencing:** NGS platforms have enabled the sequencing of complete genomes of numerous fish and shellfish species, providing valuable insights into their genetic composition, structural variation, and evolutionary history. Whole-genome sequencing facilitates the identification of genes responsible for specific traits, including disease resistance, growth rate, and adaptation to environmental conditions. This information can be utilized to enhance breeding programs, develop robust management strategies, and promote sustainable aquaculture practices.

13. **Transcriptomics**: Transcriptomic studies using NGS techniques have allowed researchers to explore the gene expression profiles of aquatic organisms under different physiological and environmental conditions. By sequencing and analysing the transcriptome, researchers can identify key genes involved in important biological processes, such as development, immune response, and stress tolerance. These findings contribute to a better understanding of the molecular mechanisms underlying various traits and can be utilized to improve aquaculture productivity and resilience.

14. **Metagenomics:** Metagenomics, which involves the sequencing of microbial communities present in aquatic ecosystems, has gained prominence in fisheries and aquaculture research. NGS-based metagenomic analysis provides insights into the microbial diversity, functional potential, and ecological roles of microorganisms associated with fish health, nutrition, and overall ecosystem stability. This information aids in the development of probiotics, disease prevention strategies, and sustainable aquaculture practices.

15. **Population Genetics**: NGS technologies have revolutionized population genetics studies by enabling the analysis of genetic variation at unprecedented resolution and scale. Through techniques such as restriction-site-associated DNA sequencing (RAD-seq) and genotyping-by-sequencing (GBS), researchers can assess genetic diversity, population structure, and patterns of gene flow in wild and farmed fish populations. This information is crucial for effective fishery management, conservation efforts, and the identification of locally adapted populations for selective breeding.

16. **Disease Management**: Disease outbreaks pose significant challenges to the aquaculture industry. NGS plays a vital role in disease management by enabling the rapid and accurate identification of pathogens, understanding their virulence factors, and tracking their transmission dynamics. Metagenomic sequencing can detect multiple pathogens simultaneously, facilitating early detection, diagnosis, and the development of targeted treatments and vaccines.

17. **Selective Breeding**: Selective breeding is a powerful tool for improving desirable traits in aquaculture species. NGS technologies enable the identification of genetic markers associated with important traits, such as disease resistance, growth rate, and flesh quality. Genomic selection, based on marker-assisted breeding, enhances the accuracy and efficiency of selecting superior individuals for breeding programs, resulting in improved productivity and reduced environmental impact.

18. **Conservation and Environmental Monitoring**: NGS techniques have significant applications in conservation genetics, enabling the assessment of genetic diversity, population structure, and effective population size of endangered and commercially important species. This information aids in designing targeted conservation strategies, monitoring population health, and preventing the loss of genetic diversity. Additionally, NGS-based environmental DNA (eDNA) analysis provides a non-invasive method for monitoring the presence and abundance of aquatic species, including rare and invasive species.

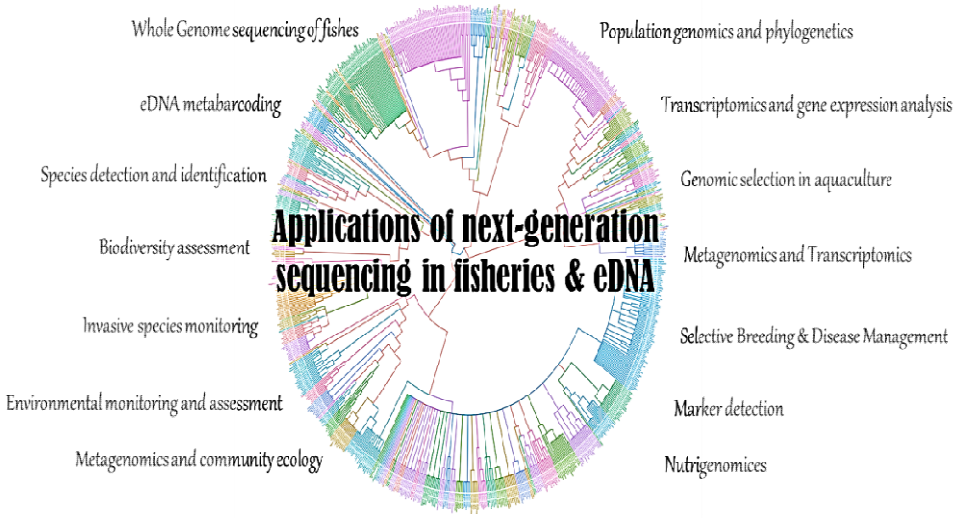

Figure 3: Application NGS in fisheries and eDNA study

VIII. USE OF NANOPORE SEQUENCING IN FISHERIES RESEARCH

Nanopore sequencing is a powerful DNA sequencing technology that has various applications in different fields, including fisheries. In the context of fisheries, nanopore sequencing can provide valuable insights into the genetic diversity, population structure, and migration patterns of fish species.

IX. FUTURE PERSPECTIVE OF NGS WITH THE CONCLUSION

Next-Generation Sequencing has revolutionized the field of genomics by enabling rapid and cost-effective sequencing of DNA and RNA. Over the years, NGS has had a profound impact on various areas of research, including medical genetics, personalized medicine, agriculture, and environmental studies. Future trends in NGS-based eDNA analysis include advancements in sequencing technologies, improved bioinformatics tools for data analysis, and the integration of multi-omics approaches.These developments will enhance the sensitivity, accuracy, and efficiency of eDNA analysis, further expanding its applications in biodiversity conservation, environmental monitoring, and ecological research.

In conclusion, the applications of next-generation sequencing in fisheries and aquacultureoffer the understanding of aquatic species, their genetic diversity, and the complex interactions within their ecosystems. From genome sequencing and transcriptomics to metagenomics and population genetics, NGS technologies have provided valuable insights that can be leveraged for disease management, selective breeding, and conservation efforts. These applications have the potential to address various challenges faced by the industry and shaping its future trajectory. As NGS technologies continue to advance, their impact on fisheries and aquaculture is expected to grow, further enhancing our understanding of aquatic ecosystems and supporting the sustainable development of the industry. NGS can generate large amounts of sequence data in one shot and opens new areas of research while on the other hand; it presents significant challenges for data analysis and interpretation into meaningful insights. So, NGS demands strong bioinformatics support.

REFERENCE

[1] Avarre, J. C. (2020). Next-generation sequencing: a revolution in the field of fish diseases. *Bulletin of the European Association of Fish Pathologists*, *40*(2), 62-69.
[2] Buermans, H. P. J., & Den Dunnen, J. T. (2014). Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1842*(10), 1932-1941.https://doi.org/10.1016/j.bbadis.2014.06.015.
[3] Ke, R., Mignardi, M., Hauling, T., & Nilsson, M. (2016). Fourth generation of next-generation sequencing technologies: Promise and consequences. *Human mutation*, *37*(12), 1363-1367.https://doi.org/10.1002/humu.23051.
[4] Kumar, G., &Kocour, M. (2017). Applications of next-generation sequencing in fisheries research: A review. *Fisheries Research*, *186*, 11-22.https://doi.org/10.1016/j.fishres.2016.07.021.
[5] Sahu, A., Kumar, N., Singh, C. P., & Singh, M. (2022). Environmental DNA (eDNA): Powerful technique for biodiversity conservation. *Journal for Nature Conservation*, 126325.https://doi.org/10.1016/j.envadv.2023.100370.
[6] Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature biotechnology*, *26*(10), 1135-1145.
[7] Thomsen, P. F., &Willerslev, E. (2015). Environmental DNA–An emerging tool in conservation for monitoring past and present biodiversity. *Biological conservation*, *183*, 4-18.https://doi.org/10.1016/j.biocon.2014.11.019.
[8] Thomsen, P. F., Kielgast, J. O. S., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., &Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular ecology*, *21*(11), 2565-2573.https://doi.org/10.1111/j.1365-294x.2011.05418.x.
[9] Yao, M., Zhang, S., Lu, Q., Chen, X., Zhang, S. Y., Kong, Y., & Zhao, J. (2022). Fishing for fish environmental DNA: Ecological applications, methodological considerations, surveying designs, and ways forward. *Molecular Ecology*, *31*(20), 5132-5164.https://doi.org/10.1111/mec.16659.

