GETTING GOOD-QUALITY NUCLEIC ACID FROM WASTEWATER SAMPLES AND ITS SIGNIFICANCE

Abstract

Next-generation sequencing (NGS) has brought about a revolutionary shift in the realm of genetic research, offering the ability to conduct high-throughput DNA sequencing with diverse applications across multiple fields. NGS encompasses two primary methods: short-read and long-read sequencing, each presenting distinct advantages. Short-read sequencing provides exceptional precision, while long-read sequencing allows for longer read lengths. Following the acquisition of raw DNA sequences through NGS, downstream processing becomes imperative for comprehensive analysis and interpretation. This crucial step involves various tasks such as aligning sequences, identifying genetic variations, assembling genomes, and annotating functionality. When comparing Illumina and Nanopore platforms, Illumina stands out for its exceptional data quality, high-throughput capabilities, and short read lengths. It proves to be exceptionally well-suited for applications that demand precise sequencing data. On the other hand, Nanopore platforms offer the advantages of longer read lengths, portability, and real-time sequencing capabilities, although they may exhibit higher error rates when compared to Illumina platforms. One of the major challenges encountered in NGS revolves around the quality of nucleic acid (NA) samples, which is a big concern when it comes to the extraction of the genomic material from environmental sources like water. Low-quality NA can introduce errors during sequencing, result in biased coverage, and diminish overall accuracy. To surmount this obstacle, it becomes paramount to ensure the integrity and purity of NA samples through the implementation of quality control measures such as optimized DNA

Authors

Aditi Nag
Dr. B. Lal Institute of Biotechnology
aditinag.bibt@gmail.com

Khushboo Sharma
Banasthali Vidyapith

Sudipti Arora
Dr. B. Lal Institute of Biotechnology

extraction and meticulous sample preprocessing. The quality of NA samples thus serves as a critical factor in obtaining reliable sequencing results. By addressing these challenges and capitalizing on continuous advancements in NGS technologies, the field of genomics research and its applications can foster further innovations.

Keywords: NA extraction protocols, Next-generation sequencing, high-throughput DNA sequencing, short-read sequencing, long-read sequencing, downstream processing, wastewater-based genomics.

I. INTRODUCTION TO NEXT GENERATION SEQUENCING

The phrase "next-generation sequencing" (NGS), commonly referred to as "high-throughput sequencing," is used to refer to a variety of contemporary sequencing technologies. The study of genomics and molecular biology has been completely transformed by these technologies, which make it possible to sequence DNA and RNA considerably more swiftly and affordably than was previously possible using Sanger sequencing.

NGS is a common technology in functional genomics and can be used to examine DNA and RNA materials. NGS-based approaches have a number of benefits over microarray techniques, including the following:

- The genome or its features do not need to be known in advance.
- Due to its single-nucleotide resolution, it is feasible to identify genes that are connected to one another, alternatively spliced transcripts, allelic gene variations, and single-nucleotide polymorphisms.
- Greater signal dynamic range.
- Requires less DNA/RNA input (materials in the form of nanoparticles are adequate).
- Increased reproducibility.

The general workflow followed typically for early generation NGS involves three steps as follows: sample preparation, library preparation and analyzing the data As shown in Figure 1.

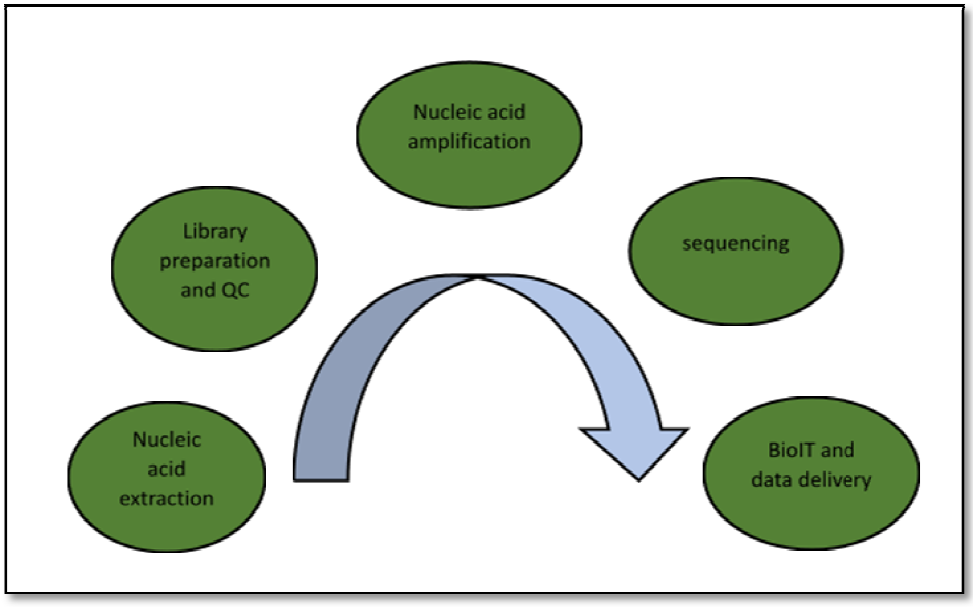


Figure 1: Basic steps of NGS

II. SHORT READ SEQUENCING: PRECISION AND APPLICATIONS

A potent tool for producing genomic data is short-read sequencing. Short-read sequencing allows DNA or RNA to be sequenced more quickly and inexpensively than with conventional techniques. Since short-read sequencing offers the most depth and quality of data at the lowest cost per base, it has long been recognized as the workhorse approach in NGS labs. Large genome sizes of organisms like humans, mice, bats, etc. are not sequenced by the Short Read Sequencing method for high throughput. As the number of nucleotides increases, the efficiency of accurate results will decrease.

Data coming from advanced molecular techniques such as targeted sequencing by next-generation sequencing (NGS) and third-generation sequencing (TGS) are more appropriate and valuable for DNA analysis(Syahzuwan Hassan,Rosnah Bahar et.al). The majority of the bacterial genome can be re-sequenced and de novo sequenced with read lengths of 20 to 30 nucleotides, and it is demonstrated that read lengths of 50 nucleotides can provide reconstructed contigs of 1000 nucleotides and greater, covering 80% of human chromosome 1. Reanalysis of short-read genome sequencing data was performed to improve the interpretation(Suzanne E. de Bruijn,Kim Rodenburg et.al). Sequence Bloom Trees (SBTs) are a newly developed method for short-read sequencing that is used for querying thousands of short reads 162 times faster than the general approach.

Additional Features of Short-Read Sequencing

- Increases time outcome and hinders the capacity to detect workflow mistakes prior to sequencing completion.
- Read length is typically 30 to 500 base pairs.
- There are additional practical difficulties related to handling and storing a large volume of data.
- Complete structural variants or significant classes of genomic aberrations are not covered by short reads.
- Gene duplication, transposons, and prophage sequences are examples of complicated genomic areas that may not be covered by short sequencing runs.
- Amplification can eliminate base modification and increase bias, which reduces coverage uniformity.
- Traditional sequencing techniques often cost a lot of money and require a lot of site infrastructure.
- Genomic assemblies that are in pieces and have unclear isoform identification
- It delays transmitting the result.

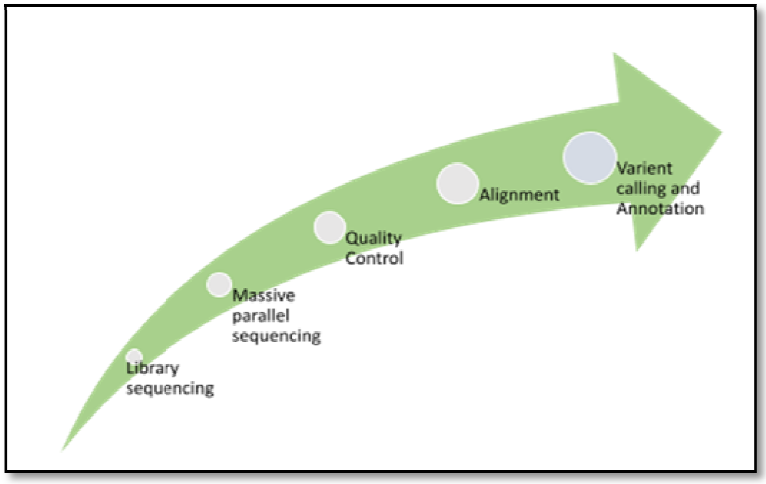


Figure 2: Steps Involved In Short Read Sequencing

III. LONG READ SEQUENCING: READ LENGTH AND ADVANTAGES

Third-generation sequencing, also known as long-read sequencing, is a DNA sequencing technique that is currently being explored. It can analyze long DNA sequences of between 10,000 and 100,000 base pairs at a time to discover their nucleotide sequence.

Long-read sequencing enables the detection of complex structural variants that may be challenging to detect with short reads, despite the fact that short reads can capture the majority of genetic variation. Large inversions, deletions, or translocations are some of these; some of these have been connected tothings like genetic illness.

This platform has proven their ability to resolve some of the most challenging regions of the human genome, detect previously inaccessible structural variants and generate some of the first telomere-to-telomere assemblies of whole chromosomes (Logsdon, G.A., College et. El).

SMART and Nanopore are some of the technologies which are used for long read sequencing.

The polymerase's persistence limits the read length in SMRT sequencing. A faster polymerase for the Sequel sequencer introduced with chemistry v3 in 2018 increased the read lengths to an average 30-kb polymerase read length(Shanika L Amarasinghe, Shian Su et.al)

Nanopore sequencing provides the longest read lengths, from 500 bp to the current record of 2.3 Mb (Payne A, Holmes N, Rakyan V,et.al, 2019). Read length in nanopore sequencing is mostly limited by theability to deliver very high-molecular weight DNA to the pore and the negative impact this has on run yield (Jain M, Koren S,et al).

1. Advantages and Benefits of Long Read Sequencing

Using longer reads to analyze genomic data has a number of inherent advantages, some of which may be beneficial for clinical genome study.

- Genome assembly: The length of the human genome is around three billion DNA base pairs, and it has numerous repeated genetic code segments. Reassembling the genome from short reads is analogous to putting together a difficult jigsaw puzzle since many of the pieces, when taken out of context, are quite identical. This work can be made easier by long-read data since the reads are more likely to appear different, reducing ambiguity and mistakes in how they are put together.

- Only long read sequencing can detect massive and complicated rearrangements, substantial insertions or deletions of DNA, repetitive sections, highly polymorphic regions, or regions with minimal DNA nucleotide diversity, for example.

- Haplotype phasing: Without further statistical inference, maternal/paternal sequencing, or sample preparation—all of which are necessary for an approximation of phasing using SRS—long reads can give the long-range information for resolving haplotypes.

- Portability: Nanopore's devices rely on detecting electronic signals rather than optical signals, in contrast to other sequencing technologies. This enables them to create gadgets that are as portable as a memory (USB) stick.

- Speed and real-time sequencing: PacBio and Oxford Nanopore both offer quicker sequencing runs than the fixed run times of SRS methods.

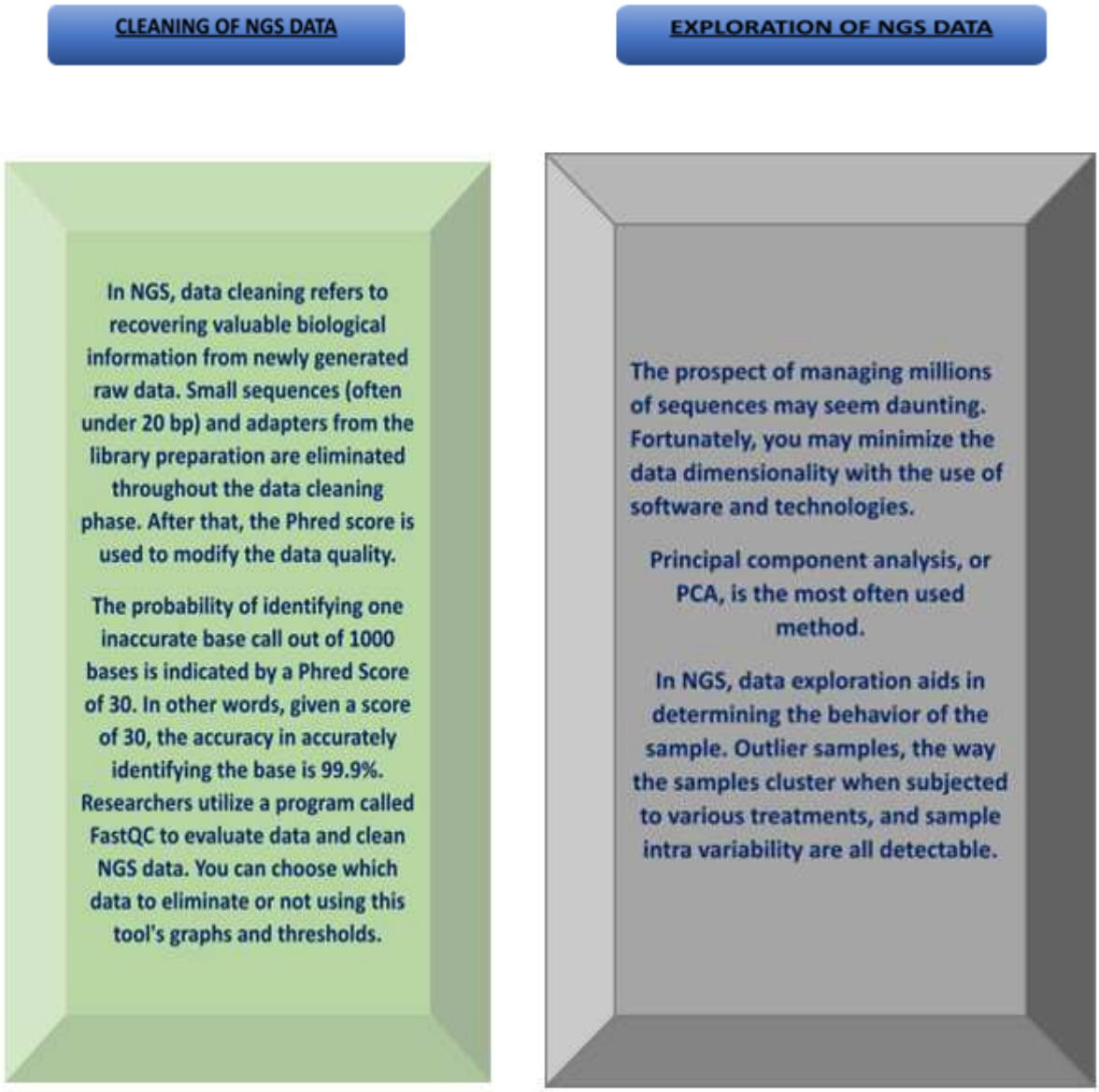
IV. DOWNSTREAM PROCESSING: ANALYZING AND INTERPRETING NGS DATA

Numerous biological and medical research projects have benefited from the analysis of gene expression. Microarrays have been widely utilized to profile the expression of genes during various developmental processes, medical conditions, and treatments.We now have a far better grasp of how gene regulation and signaling networks work thanks to new massively parallel sequencing techniques, also known as RNA-sequencing (RNA-seq) (Kukurba et al., 2015).

Any processes performed on the pre-processed data are categorized as downstream analysis, including simple descriptive analysis, hypothesis testing, grouping, and prediction. While downstream analysis is more focused, many preprocessing strategies are applicable to both clinical and tumor genetics.

1. **Analyzing and Interpretation of Ngs Data:** There are four major steps involved for analyzing the NGS data. These are:
   - cleaning of NGS data
   - Exploration of NGS data
   - Visualization of NGS data

- Deeper analysis of NGS data



V. COMPARATIVE ANALYSIS: ILLUMINE VS NANOPORE

S. NO		ILLUMINA	NANOPORE
1	Purpose	Short read	long read
2.	Read length	30 to 300 bp	More than 1kb
3.	Accuracy	99.9%	85%
4.	Typically used for	• GBS/RAD sequencing • Whole genome sequencing	• Studying structuralvariants • Genome assembly

		• Genome assembly scaffolding	
5.	Limit of detection	<500 copies/ml	10 copies/reaction
6.	Relative cost	intermediate	Low to high

The detailed differences between Illumina and Nanopore techniques are given in the following sections:

VI. ILLUMINA PLATFORM: DATA QUALITY AND HIGH THROUGHPUT CAPABILITIES

For quick and precise large-scale sequencing, Illumina's sequencing method makes use of clonal array construction and its own reversible terminator technology. A wide range of purposes in genomics, transcriptomics, and epigenomics are made possible by the novel and adaptable sequencing system.

1. Workflow of Illumina (Illumina resource page 2023)

- The DNA must first be divided into smaller, more manageable fragments, ranging in size from 200 to 600 base pairs, as the initial stage in this sequencing procedure.
- The DNA pieces are connected to one another by brief DNA segments referred to as adaptors.
- The adaptor-attached DNA fragments are then converted to single-stranded DNA. The fragments are incubated with sodium hydroxide to accomplish this.
- Fragments of DNA are produced and then washed through the flow cell. The primers on the flow cell's surface bond to the complementary DNA, and any non-attaching DNA is washed away.
- The flowcell's connected DNA is then replicated to create tiny DNA clusters with the same pattern. Each group of DNA molecules will release a signal after being sequenced that can be seen by a camera.
- The DNA strands connected to the flow cell are then lengthened and joined using unorganized bases of nucleotides and DNA polymerase.
- Heat is then used to convert the double-stranded DNA into single-stranded DNA, leaving behind a few million clumps of identical DNA strands.
- The flow cell is then supplemented with primers and fluorescently labeled terminators, which are nucleotide bases (A, C, G, or T) that terminate the synthesis of DNA.
- The DNA that is being sequenced is attached to the primer.
- The first fluorescently labeled terminator is then added to the new DNA strand by the DNA polymerase after it has bound to the primer. No extra bases can be introduced to the DNA strand after a base has been inserted until the final base has been removed.
- The fluorescent label on the nucleotide base is activated by passing lasers across the flowcell.
- Until countless clusters have been sequenced, the process keeps going.

- Illumina sequencing is a very accurate technique since the DNA sequence is examined base-by-base. The produced sequence can then be compared with an existing reference sequence to check for similarities or variations in the DNA sequence.

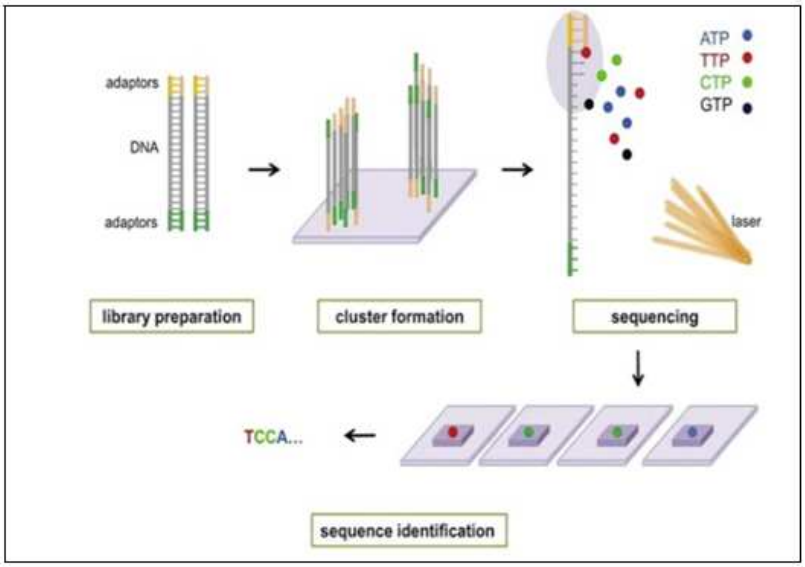


Figure 3: The picture shows the diagrammatic representation of illuminated sequencing (Zhou, X., Li, Y., et al.)

VII. NANOPORE PLATFORMS: LONG READS, PORTABILITY AND REAL TIME SEQUENCING

The most potent technique for quickly producing long-read sequences is nanopore sequencing technology, which was created by Oxford Nanopore Technologies Ltd. With Nanopore sequencing, a single molecule of DNA or RNA can be sequenced without the need for PCR amplification or chemical labeling of the sample (Purchase et al., 2019)
The fourth-generation DNA sequencing method is known as "nanopore sequencing," and its key benefits include label-free, ultralong reads (104–106 bases), high throughput, and little material required. Longer DNA or RNA segments can now be directly and immediately analyzed thanks to the innovative, scalable method known as nanopore sequencing. It functions by watching how an electrical current changes as nucleic acids travel through a protein nanopore. Decoding the resulting signal allows one to determine the exact DNA or RNA sequence.

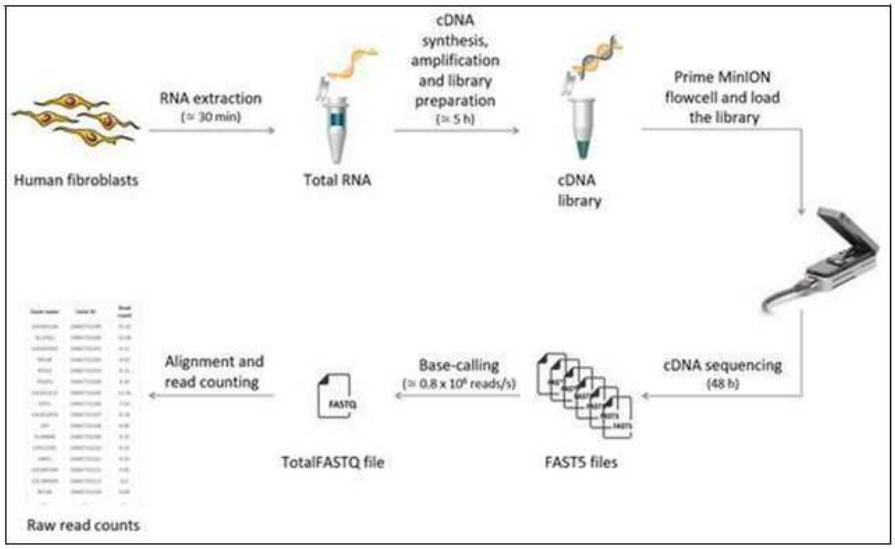


Figure : 4 A diagrammatic representation of Nanopore technique (Ilaria Massaiu Paola Songia et.al.)

1. **Real Time Sequencing:** The fact that there is no set run time for the MinION device, the PromethION, or the GridION systems—and that data is delivered in real time instead—is a crucial element of all three.Real-time insights are made possible by real-time data. The user can choose an experimental endpoint in advance and run the system for however long is necessary to gather enough information to answer that question.

   Each tiny pore in an array analyzes chemicals in the sample separately from the other nanopores during an experiment. The time it takes for one analyte molecule to successfully engage with one nanopore in the array is the quickest time to begin collecting experimental data. Data becomes available as soon as a DNA molecule begins to travel through the pore, even though it may take a single DNA molecule minutes to seconds to do so. Another DNA molecule will load once the previous one has exited the nanopore.

2. **Read Length:** Nanopore sequencing may span entire repetitive sections, resolve structural variations, and distinguish between different isoforms because it is only constrained by the size of the DNA/RNA fragment delivered to the pore. Native DNA and RNA can be sequenced without the need for amplification, which removes PCR bias and makes it possible to identify base changes like methylation alongside nucleotide sequence. The readings are processed by Oxford Nanopore devices; the lengths produced depend on sample preparation, and the longest read reported by a MinION user so far is >4 Mb in length.

   Long and ultra-long reads have a number of benefits, such as easier assembly and more thorough investigation of repetitive regions, phasing, or CNVs.

3. **Portability:** The MinION, which has the size of a stapler and is USB-powered, is utilized outside of the typical lab setting and enables users to bring analysis to the sample. The

MinION Mk1C is a full, mobile, networked device that performs both basecalling and analysis onboard as well as sequencing. In addition to studying the Antarctic microbiome, tracking viral outbreaks internationally, putting an end to marine poaching, and even boarding the International Space Station, MinION devices have been employed in a variety of other applications. VolTRAX further enables sequencing in various settings andby those with less lab expertise.

VIII. METHOD FOR PREPPING AND EXTRACTING HIGH-QUALITY RNA FROM SARS-COV-2 IN WASTEWATER FOR NEXT-GENERATION SEQUENCING

Wastewater, often underestimated in its significance, serves as a reservoir brimming with potential pathogens, including viruses, bacteria, and other microorganisms. This murky mixture, which combines domestic sewage, industrial effluents, and runoff, becomes a veritable melting pot for disease-causing agents shed by human and animal populations. In this complex aquatic environment, various pathogens can persist and even thrive, posing a latent threat to public health. Viruses like hepatitis, norovirus, andeven emerging threats like SARS-CoV-2 can persist in wastewater, while bacteria such as *E. coli* and *Salmonella* also find refuge. This underscores the importance of rigorous treatment and monitoring of wastewater, as inadequate management can transform it into a source of infection and environmental contamination, emphasizing the need for vigilant wastewater management strategies to protect both human health and the environment (Daughton 2018).

Wastewater-based epidemiology, often abbreviated as WBE, represents a groundbreaking approach to monitoring public health on a community or even a regional scale. It capitalizes on the fact that human and environmental health leave behind valuable traces in wastewater. By analyzing the various substances, pathogens, and biomarkers present in sewage and wastewater, scientists and public health officials can gain crucial insights into the health trends and well-being of a population. This innovative field has garnered increasing attention for its ability to detect and track diseases, including viral outbreaks like COVID-19, illicit drug use, environmental contaminants, and overall community health dynamics. Wastewater-based epidemiology serves as a powerful tool for early detection, rapid response, and evidence-based decision-making, bridging the gap between public health and environmental science in an effort to safeguard communities worldwide (Keshaviah et al., 2021).

Monitoring the presence and genomic variability of pathogens in wastewater offers a multitude of advantages in the realm of public health and epidemiology. Firstly, it provides an early warning system for disease outbreaks by detecting pathogens shed by infected individuals in the population, often before clinical cases become evident. This enables swift and targeted public health responses. Secondly, it offers a broader understanding of the epidemiological landscape by capturing data from entire communities, including asymptomatic carriers, helping to assess the true prevalence of diseases. Thirdly, analyzing genomic variability can track the evolution of pathogens, aiding in the identification of new variants that might be more transmissible or resistant to treatments. Moreover, wastewater-based surveillance is non-invasive and cost-effective, making it a valuable complement to traditional clinical surveillance systems. Finally, it can be applied to a range of pathogens beyond just viruses, including bacteria, parasites, and antimicrobial resistance genes, offering

a comprehensive perspective on community health. Thus, monitoring pathogens and their genomic diversity in wastewater provides an invaluable tool for proactive public health management and response (Daughton 2018).

Obtaining high-quality nucleic acids from wastewater samples is of paramount importance in various scientific fields, particularly in environmental microbiology, epidemiology, and wastewater-based epidemiology. The quality of nucleic acids directly impacts the accuracy and reliability of downstream genetic analyses, including PCR, sequencing, and genotyping. Here are key reasons why high-quality nucleic acids are essential (Carlson et al., 2028)l:

- In wastewater-based epidemiology, the presence and genetic characteristics of pathogens like viruses and bacteria are monitored. High-quality nucleic acids ensure that the detected genetic material indeed represents the target pathogens, reducing the risk of false positives or negatives.

- Researchers may want to analyze the genomic diversity and characteristics of microorganisms in wastewater, which can be crucial for understanding the spread of diseases, tracking antibiotic resistance genes, or studying environmental microbiota. Low-quality nucleic acids can lead to biased or incomplete genetic data.

- Quantitative PCR (qPCR) is commonly used to measure the abundance of specific genetic markers in wastewater. Reliable quantification depends on the quality of the starting nucleic acids.

- Next generation sequencing is widely used to find out newly emerged variants of viruses and identify the pathogen etc. As mentioned above a good quality of nucleic acid ensures reliable reads and data analysis.

Challenges in obtaining high-quality nucleic acids from wastewater samples include:

1. **Inhibitors:** Wastewater contains various inhibitors such as humic acids, heavy metals, and complex organic compounds, which can interfere with nucleic acid extraction and downstream analyses. These inhibitors must be effectively removed or neutralized.

2. **Low Concentrations:** Pathogens in wastewater are often present at low concentrations, making it challenging to extract sufficient nucleic acids for analysis. This requires sensitive and efficient extraction methods.

3. **Sample Variability:** Wastewater samples can vary widely in composition, depending on the source and time of collection. This variability can affect the quality and quantity of extracted nucleic acids.

4. **Contaminants:** Cross-contamination is a significant concern during nucleic acid extraction from wastewater samples. Strict laboratory practices and controls are necessary to prevent contamination from other samples or environmental sources.

5. **Degradation:** Nucleic acids in wastewater can degrade rapidly due to the presence of

nucleases and environmental conditions. Quick processing and preservation of samples are crucial to prevent degradation.

High-quality nucleic acids are essential for accurate and reliable genetic analyses in wastewater-based epidemiology and environmental microbiology. The challenges in obtaining such nucleic acids from wastewater samples necessitate rigorous sample preparation techniques and quality control measures to ensure the integrity and purity of the genetic material extracted for analysis.

Since a good quality of nucleic acid is important as well as tricky to extract. Therefore the protocol usedfor this extraction needs to be carefully modified and executed with precautions (Ahmed et al., 2020).

Following protocol has given consistently good yield for the authors:

1. **Sample Collection and Preparation:** Use wide-mouth, autoclavable polypropylene sample bottles (500 ml/1000 ml capacity) for wastewater(WW) sample collection. Ensure proper labeling of bottles, indicating sampling location, date, and time.Collect samples in the morning (between 8 am to 11 am) for efficient recovery.

   Wear appropriate personal protective equipment (PPE) including gloves, mask, full-sleeve apron, and optionally, a PPE suite and face shield to minimize contamination. Remove the bottle top just before sampling to prevent contamination.

   After collection, securely close the bottle and place it in a cooler with ice packs to maintain a cold chain. Record the sample temperature during collection and transportation using a thermometer.

   Sanitize gloves with 70% ethanol after closing the box for transportation.
   Process samples immediately upon arrival at the laboratory or store them at 4°C for no more than 24 hours to prevent significant degradation.

2. **Sample Processing:**

   - Conduct all steps in a Biosafety Level 2 (BSL-2) cabinet except for centrifugation and sample incubation.
   - Ensure the UV in the biosafety cabinet is on early to reach optimal conditions.
   - Use a refrigerated centrifuge capable of 10,000 rpm equivalent RCF (rotational centrifugal force), with the specific rpm value depending on the rotor's g-force.
   - Wear appropriate protective gear, including gloves, mask, full-sleeve apron, and optionally, a PPE suite and face shield.
   - Invert sample bottles to homogenize water and debris, then pasteurize the wastewater by placing sample bottles at 60°C in a water bath for 1 hour.
   - Following pasteurization, centrifuge at 5000 rpm (5000 x g) for 10 minutes at 4°C toremove larger debris particles.
   - Filter the supernatant twice using vacuum filtration assembly with a Whatman Grade 1 filter paper followed by a 0.22 μm PES membrane filter.

- Use disposable forceps for removing filters, changing them frequently to avoid clogging.

3. Sample Concentration by PEG-NaCl Method:

- Mix the filtrate with PEG 8000 (8%) and NaCl (1.7%) in a sterile container.
- Incubate the mixture at 4-8°C with agitation (150-160 rpm) for 2-4 hours until PEG completely dissolves.
- Centrifuge the mixture at 10,000 rpm (15,000 x g) for 30 minutes at 4°C.
- Decant the supernatant gently without disturbing the pellet to avoid loss.
- Resuspend the pellet in an elution buffer (140 µl for 40 ml protocol, 560 µl for 80 ml protocol).
- Store one set of resuspended pellets at -80°C for future use and immediately proceed with RNAextraction from the remaining half.

4. RNA Extraction and qRT-PCR:

- We Utilize the QIAamp Viral RNA Kit, following the manufacturer's protocol (refer to the QIAGEN QIAamp Viral RNA Mini Handbook).
- For the 40 ml protocol, elute RNA in 40 µl Buffer AVE. For the 80 ml protocol, elute RNA in 80 µl Buffer AVE.
- Store viral RNA at -80°C, avoiding frequent freeze-thaw cycles.
- Adhere to safety guidelines throughout the procedure, including wearing appropriate lab attire, gloves, and protective goggles.

5. Next-Generation Sequencing of Wastewater Samples:

For Nanopore sequencing, use the MinION/Mk1C sequencer with the midnight protocol (1200 base pair PCR amplicons) and the RAPID barcoding kit (SQK-RBK110.96).

For Illumina sequencing on the NextSeq 550 Sequencing Platform, prepare wet lab libraries and can be loaded using the Illumina Covid seq Ruo kits as per manufacturer instructions.

To conclude, the above steps have given us the optimized results and quality for RNA from an environmental sample along with a very small incidence of 0 coverage in Nanopore sequencing. Thus, this protocol is recommended for the WBE based studies as it can be a useful protocol for qRT-PCR as well as NGS and can prevent the aforementioned quality compromises mentioned.

REFERENCES

[1] Ahmed, W., Bivins, A., Bertsch, P. M., Bibby, K., Choi, P. M., Farkas, K., Gyawali, P., Hamilton, K. A., Haramoto, E., Kitajima, M., Simpson, S. L., Tandukar, S., Thomas, K., & Mueller, J. F. (2020). Surveillance of SARS-CoV-2 RNA in wastewater: Methods optimisation and quality control are crucial for generating reliable public health information. Current opinion in environmental science & health, 10.1016/j.coesh.2020.09.003. Advance onlin

[2] Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. Genome biology, 21(1), 1-16. https://doi.org/10.1186/s13059-020-1935-5

[3] Bharagava, R. N., Purchase, D., Saxena, G., & Mulla, S. I. (2019). Applications of metagenomics in microbial bioremediation of pollutants: from genomics to environmental cleanup. In Microbial diversity in the genomic era (pp. 459-477). Academic Press. https://doi.org/10.1016/B978-0-12-814849-5.00026-5

[4] Carlsson, J., Davidsson, S., Fridfeldt, J., Giunchi, F., Fiano, V., Grasso, C., Zelic, R., Richiardi, L., Andrén, O., Pettersson, A., Fiorentino, M., & Akre, O. (2018). Quantity and quality of nucleic acids extracted from archival formalin fixed paraffin embedded prostate biopsies. BMC medical research methodology, 18(1), 161. https://doi.org/10.1186/s12874-018-0628-1

[5] Daughton C. G. (2018). Monitoring wastewater for assessing community health: Sewage Chemical-Information Mining (SCIM). The Science of the total environment, 619-620, 748–764. https://doi.org/10.1016/j.scitotenv.2017.11.102

[6] de Bruijn, S. E., Rodenburg, K., Corominas, J., Ben-Yosef, T., Reurink, J., Kremer, H., ... & Roosing, S. (2023). Optical genome mapping and revisiting short-read genome sequencing data reveal previously overlooked structural variants disrupting retinal disease− associated genes. Genetics in Medicine, 25(3), 100345. https://doi.org/10.1016/j.gim.2022.11.013

[7] Hassan, S., Bahar, R., Johan, M. F., Mohamed Hashim, E. K., Abdullah, W. Z., Esa, E., Abdul Hamid, F. S., et al. (2023). Next-Generation Sequencing (NGS) and Third-Generation Sequencing (TGS) for the Diagnosis of Thalassemia. Diagnostics, 13(3),

[8] 373. MDPI AG. Retrieved from http://dx.doi.org/10.3390/diagnostics13030373

[9] Illumina resource page Pagehttps:// sapac.illumina.com/science/technology/ next-generation-sequencing/beginners/ngs-workflow.htm. (Last Accessed on 23 September 2023)

[10] Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., ... & Loose, M. (2018).Nanopore sequencing and assembly of a human genome with ultra-long reads. Nature biotechnology, 36(4), 338-345. https://doi.org/10.1038/nbt.4060

[11] Keshaviah, A., Hu, X. C., & Henry, M. (2021). Developing a Flexible National WastewaterSurveillance System for COVID-19 and Beyond. Environmental health perspectives, 129(4), 45002. https://doi.org/10.1289/EHP8572

[12] Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. Cold Spring Harbor protocols, 2015(11), 951–969. https://doi.org/10.1101/pdb.top084970

[13] Logsdon, G. A., Vollger, M. R., & Eichler, E. E. (2020). Long-read human genome sequencing and its applications. Nature Reviews Genetics, 21(10), 597-614https://doi.org/10.1038/s41576-020-0236-x

[14] Massaiu, I., Songia, P., Chiesa, M., Valerio, V., Moschetta, D., Alfieri, V., ... & Poggio, P. (2021). Evaluation of Oxford Nanopore MinION RNA-Seq performance for human primary cells. International Journal of Molecular Sciences, 22(12), 6317. https://doi.org/10.3390/ijms22126317

[15] Payne, A., Holmes, N., Rakyan, V., & Loose, M. (2019). BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. Bioinformatics, 35(13), 2193-2198. https://doi.org/10.1101/312256.

[16] Zhou, X., & Li, Y. (Eds.). (2015). Techniques for Oral Microbiology. In Atlas of Oral Microbiology (pp. 15-40). Academic Press. ISBN 9780128022344. doi:10.1016/B978-0-12-802234-4.00002-1.

