American Journal of Medical and Natural Sciences
Content Available at www.ajmns.com ISSN (O): 2582-6182
(An International online peer reviewed Referred Journal)
Review Article Open Access

CLINICAL METAGENOMICS IN MICROBIOLOGY DIAGNOSTICS: ANALYTICAL CHALLENGES, QUALITY ASSURANCE, AND EMERGING FUTURE-READY APPROACHES

DEEPANSHU RAWAT

Assistant Professor, Faculty of Paramedical and Allied Health Sciences,
Motherhood University Roorkee, Haridwar, Uttarakhand 247661 (India)

Article History: Received: 19 Dec 2025, Revised: 14 Jan 2026, Accepted: 13 Feb 2026

***Corresponding author**
Deepanshu Rawat

DOI: https://doi.org/10.70604/ajmns.v6i1.22

Abstract

Metagenomic sequencing has quickly developed into a revolutionary method in clinical microbiology by making it possible to identify and characterise pathogens directly from clinical specimens without the need for culture. In contrast to focused PCR tests, which need previous information of suspected species, metagenomics facilitates the wide identification of bacteria, viruses, fungi, and parasites and can offer insights into strain-level epidemiology and antibiotic resistance determinants. However, obstacles such low microbial biomass, host DNA dominance, contamination, variable extraction efficiency, bioinformatics complexity, interpretative ambiguity, and expense continue to limit routine application in diagnostic laboratories.The modern metagenomic workflows-sample preparation, nucleic acid extraction, library building, sequencing techniques, and computational pipelines-are summarised in this study, which also emphasises the crucial quality assurance and control procedures needed for accurate reporting. Meningitis/encephalitis, respiratory infections, sepsis, bloodstream infections, and outbreak investigations are among the clinical applications we cover. We also assess translation obstacles including standardisation, regulatory frameworks, reference databases, and result interpretation.

Keywords: *Clinical metagenomics, Next-generation sequencing, Pathogen detection, Bioinformatics, Quality control,antimicrobial resistance, Diagnostic microbiology.*

This article is licensed under a Creative Commons Attribution-Non-commercial 4.0 International License. Copyright © 2026 Author(s) retains the copyright of this article.

1. INTRODUCTION

The specialised area of clinical microbiology links the enormous universe of microbes to patient treatment. It serves as the diagnostic link between an individual's disease and the microscopic, invisible biological source, such as bacteria, viruses, fungus, or parasites. Finding the infectious agent in a patient sample fast and precisely is the main goal [1]. This enables doctors to make well-informed treatment decisions, guaranteeing that medical actions are focused and supported by science. The majority of this discipline's work is done at public health facilities, commercial reference laboratories, and hospital labs [2]. In order to find disease-causing substances, practitioners continuously examine patient specimens such as blood, urine, tissue, and cerebrospinal fluid. This emphasis enables medical personnel to comprehend the microbe's behaviour and take appropriate action to enhance patient outcomes.A popular molecular diagnostic technique in clinical labs, polymerase chain reaction (PCR) may quickly determine if DNA and RNA are present in a clinicalsample without the requirement for microbial culture.Although targeted PCR is quick, it can only be used on species that are part of the test panel [4]. By sequencing every nucleic acid in a material, metagenomics fills in these gaps and enables the objective identification of both known and perhaps new species. Metagenomics is useful in applied microbiology for monitoring antibiotic resistance, tracking outbreaks, controlling infections, and environmental surveillance in addition to diagnosis [5].

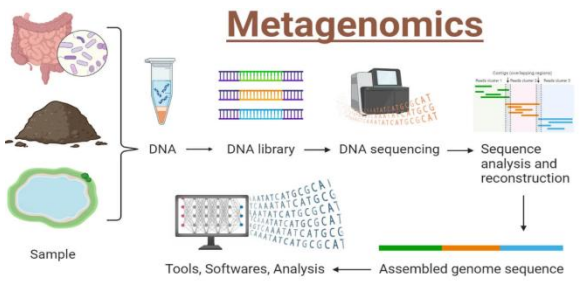

Fig. 01 Metageninics

2. TYPES OF CLINICAL METAGENOMIC TESTING

2.1 Shotgun DNA Metagenomics (mNGS)

a traditional approach focussing on a single marker gene, shotgun DNA metagenomics is a culture-independent method that extracts all DNA directly from a sample (such as stool, soil, water, sputum, or blood) and sequences it at random (a process known as "shotgun"). Bioinformatics is then used to process the sequencing reads: low-quality reads are eliminated, host (human) DNA is filtered out, and the remaining sequences are compared with reference databases to determine the presence of microbes (viruses, fungi, bacteria, and occasionally parasites) and to predict their functional potential [6].

- **Detects mixed infections and unexpected pathogens:**Shotgun metagenomics can identify many microorganisms simultaneously (co-infections) and uncommon or unexpected diseases that wouldn't be targeted in standard testing since it reads all of the DNA in the sample rather than just one suspected germ.
- **Enables species/strain identification (depends on depth + database quality):** The technique may match DNA fragments to species-level and occasionally even strain-level signals with sufficient sequencing reads (high depth/coverage) and a robust reference database. Identification may be ambiguous or limited to the genus level if sequencing depth is inadequate if the organism is poorly documented in databases.
- **Infers antimicrobial resistance genes (ARGs), but genotype ≠ phenotype:**It can identify potential resistance by identifying DNA sequences of known resistance genes (the genotype). Confirmatory phenotypic testing is frequently required because this does not necessarily predict real medication response (phenotype) because the gene may not be expressed, may be nonfunctional, may be in low quantity, or may not belong to the underlying disease-causing organism [7,8,9].

2.2 RNA Metatranscriptomics

In order to investigate the active gene expression of microbial communities at a certain moment, RNA metatranscriptomics is a culture-independent method that examines the total RNA isolated straight from a sample. Following RNA extraction, ribosomal RNA is often depleted. The residual RNA is then transformed into cDNA, sequenced, and subjected to bioinformatics analysis to determine which microorganisms are metabolically active and which genes and pathways are being expressed. Although this method is particularly helpful for identifying RNA viruses, comprehending host-microbe and microbe-microbe interactions, and connecting microbial presence to functional activity, it is technically difficult because of RNA instability, increased expense, complicated data analysis, and the challenge of differentiating between microbial and host RNA [10].

2.3 Targeted Amplicon Sequencing

Targeted amplicon sequencing is a metagenomics technique that profiles microbial communities by amplifying and sequencing a particular genetic marker from a sample. It is quicker and less expensive than shotgun sequencing, but it primarily yields community composition rather than complete functional/AMR information.

- 16S rRNA sequencing: Identifies and contrasts bacterial species in a sample using the bacterial 16S ribosomal RNA gene. It works well for genus-level profiling, but it frequently fails to accurately distinguish closely related species or strains, and it typically fails to identify viruses and most fungi (because they lack the bacterial 16S marker).
- ITS sequencing: Emphasises on the extremely variable Internal Transcribed Spacer (ITS) region of fungi, which makes it valuable for evaluating fungal diversity (mycobiome) and detecting fungus. It does not profile viruses or bacteria, and its accuracy is dependent on the quality of the database and primer selection [11,12].

3. ANALYTICAL CHALLENGES IN CLINICAL METAGENOMICS

3.1 Sample Complexity and Host DNA Contamination: Clinical specimens such as blood, cerebrospinal fluid, and respiratory samples contain a high proportion of host nucleic acids, often overwhelming microbial reads. This reduces analytical sensitivity and complicates pathogen detection, particularly for low-abundance organisms.

3.2 Nucleic Acid Extraction Bias: Extraction methods can selectively enrich or deplete certain microbial taxa, leading to skewed community profiles. Differences in cell wall composition (e.g., Gram-positive bacteria, fungi, mycobacteria) further exacerbate bias.

3.3 Sequencing Errors and Platform Variability: Short-read sequencing platforms may struggle with repetitive regions, plasmids, and complex resistance loci, while long-read technologies face higher raw error rates. Platform-specific artifacts can impact taxonomic classification and resistance gene identification.

3.4 Bioinformatics and Data Interpretation: Metagenomic analysis pipelines vary widely in

reference databases, classification algorithms, and reporting thresholds. Distinguishing true pathogens from contaminants or commensal flora remains a critical interpretive challenge.

3.5 Clinical Relevance and Actionability

Detection of microbial DNA does not necessarily indicate active infection. Differentiating colonization, contamination, and clinically significant infection requires integration with clinical metadata and laboratory findings [13,14,15].

Table 01:

Challenge	Diagnostic impact	Common mitigation approaches
Host DNA contamination	Low microbial read depth; missed low-biomass pathogens	Host depletion, microbial enrichment, deeper sequencing, optimized sample processing
Extraction bias	Skewed profiles; false negatives for hard-to-lyse organisms	Mechanical/enzymatic lysis optimization, validated extraction kits, internal spike-ins
Platform variability	Inconsistent calls; difficulty resolving plasmids/AMR context	Platform-specific validation, hybrid (short+long) strategies, standardized metrics
Bioinformatics variability	Pipeline-dependent results; contamination misclassification	Version-controlled pipelines, curated databases, negative controls, fixed thresholds
Clinical actionability	Over-calling colonizers/contaminants; uncertain significance	Clinical metadata integration, sterile-site vs non-sterile interpretation frameworks, confirmatory testing

4. QUALITY ASSURANCE AND QUALITY CONTROL (QA/QC)

4.1 Contamination Control

Contamination control is crucial in clinical metagenomics because the technology is so sensitive-it can identify DNA from a single microbial cell. DNA contamination can come from a variety of sources, such as operator handling, airborne particles, lab surfaces, and chemicals (the so-called "kitome"). Inadequate control over contamination can result in false-positive pathogen identification, misunderstanding of microbial abundance, or inaccurate projections of antibiotic resistance, all of which might have a direct effect on patient treatment. Strict contamination control procedures are used in labs to ensure the accuracy and dependability of results (Fig. no. 02).

- **Unidirectional workflow (pre-PCR to post-PCR separation):** Laboratory processes are organized in a strict linear sequence-from **sample preparation and nucleic acid extraction** to **library preparation** and finally **PCR/sequencing**. This prevents amplified DNA from contaminating upstream steps, which is a common source of false-positive reads in highly sensitive metagenomic assays.
- **Dedicated clean rooms for extraction and library preparation:** Physical separation of workspaces for different stages of the workflow minimizes the risk of **cross-contamination** between samples or from previously amplified DNA. Extraction, library preparation, and sequencing setup are performed in **independently ventilated areas** with controlled access.
- **UV and chemical decontamination:** Surfaces, equipment, pipettes, and consumables are routinely treated with **UV irradiation** or **DNA-degrading chemicals** (e.g., bleach, DNA-away solutions) to eliminate residual nucleic acids. This reduces background DNA that could otherwise be amplified and sequenced erroneously.
- **Frequent use of no-template controls (NTCs) and extraction blanks:** Including **blank samples** in each batch helps monitor **background contamination** from reagents or the laboratory environment. By comparing microbial reads in these controls with actual samples, bioinformatic pipelines can distinguish true pathogens from contaminants, ensuring that reported organisms are genuinely present in the patient sample [16,17,18,19].

Fig 02:

4.2 Controls and Validation

To provide precise, trustworthy results in clinical metagenomics, strict controls and validation are crucial. Negative controls (extraction blanks and library blanks) track background contamination, whereas positive controls (such as a mimic microbial

community) verify that the procedure can identify anticipated species. Sequencing yield and extraction efficiency may be monitored via an internal spike-in. The assay's validation should determine the reportable range and assess the limit of detection (LoD), repeatability, and accuracy across several operators and days. In order to verify clinical agreement and make sure the metagenomic test is reliable and appropriate for diagnostic usage, data should be compared with gold standard techniques (culture, PCR) [20].

4.3 Database and Software Governance

In clinical metagenomics, careful governance of **databases and software** is critical to ensure consistent and reliable results. Since reference databases are frequently updated, it is essential to **track versions** so that analyses are reproducible and results can be interpreted accurately. **Bioinformatics pipelines** must be **locked and validated**, just like wet-lab procedures, to prevent inadvertent changes from affecting outcomes. Additionally, laboratories should **maintain audit trails and documented updates** for both databases and software, providing a clear record of changes for regulatory compliance and quality assurance [21].

- **Databases change frequently; version tracking is essential:** Regular updates in microbial reference databases can alter classification results. Version tracking ensures that results are reproducible and interpretable over time.
- **Bioinformatics pipelines must be locked and validated:** Pipelines should be standardized, validated, and protected against unapproved modifications, similar to laboratory protocols, to maintain analytical reliability.
- **Maintain audit trails and documented updates:** All changes to software, pipelines, or databases should be logged with timestamps and responsible personnel, supporting traceability, quality assurance, and regulatory compliance [22].

5. EMERGING FUTURE-READY APPROACHES

5.1 Host DNA Depletion and Targeted Enrichment

The predominant presence of host nucleic acids in many clinical specimens, which can obscure microbial signals and lower diagnostic sensitivity, is one of the biggest technical obstacles in clinical metagenomics. Therefore, host DNA depletion and targeted enrichment techniques have become essential methodological developments to improve pathogen identification, especially in low-biomass materials like tissue biopsies, blood, and cerebrospinal fluid.The goal of host DNA depletion is to preserve microbial genetic material while selectively eliminating or reducing human nucleic acids [23]. Common methods include enzymatic destruction of methylated host DNA, selective lysis of host cells followed by nuclease digestion of liberated human DNA, and differential centrifugation or filtration to extract intact microbial cells. More recently, methods for adaptive sequencing and CRISPR-Cas-based depletion have been developed to increase the percentage of useful microbial genomes by dynamically rejecting host-derived reads during sequencing. Even though these techniques significantly increase microbial read recovery, unintentional loss of microbial DNA and partial host clearance are still issues, especially for intracellular infections [24].

5.2 Long-Read and Real-Time Sequencing

Complex genomic areas that are challenging to rebuild with short reads can be better resolved thanks to long-read sequencing methods, which provide reads that span several kilobases. Resolving repetitive sequences, mobile genetic elements, plasmids, and integ-ons-which typically include virulence and antimicrobial resistance (A-R) genes-requires this skill. Long-read sequencing provides more accurate antimicrobial stewardship and increases confidence in genotype–phenotype correlations by directly connecting resistance determinants to their host organism. Furthermore, it is possible to collect near-complete genome assemblies directly from clinical samples, which makes strain-level identification, outbreak investigation, and transmission monitoring easier. Despite these benefits, strong error-correction algorithms and meticulous validation are required due to greater raw error rates and poorer throughput when compared to short-read systems [25].

5.3 Artificial Intelligence and Machine Learning

By tackling the intricacy and noise present in sequencing data, artificial intelligence (AI) and machine learning (ML) are being employed more and more to enhance the precision, speed, and clinical interpretability of metagenomic diagnostics. By identifying patterns that go beyond straightforward rule-based matching, ML-assisted techniques can improve taxonomy classification by resolving unclear read assignments and enhancing discriminating between closely related taxa. By using negative controls, batch information, and laboratory-specific signatures, AI models are also useful for detecting and correcting contamination and background signals. This reduces false-positive pathogen calls, particularly in low-biomass sterile-site specimens [26].

5.4 Standardized Reference Databases and Global Surveillance

Global monitoring networks and standardised reference databases are critical to enhancing the clinical value and dependability of metagenomic diagnostics. Variability in database content, taxonomy, genome quality, and annotation standards might result in conflicting findings across laboratories since organism identification, strain typing, and antimicrobial resistance (AMR) annotation rely largely on reference genomes and curated gene catalogues. Therefore, harmonised, clinically curated reference resources with

quality-controlled genomes, stable nomenclature, transparent versioning, validated AMR/virulence gene annotations, and explicit guidelines for managing closely related species and emerging taxa are necessary for future-ready clinical metagenomics. Metagenomic data can provide near real-time epidemic detection, tracking of transmission chains, and monitoring of changing resistance mechanisms across hospitals and communities when such standardised databases are combined with regional and global surveillance systems [27].

5.5 Point-of-Care and Workflow Automation

Important developments for integrating clinical metagenomics from specialised reference laboratories into standard diagnostic procedures include workflow automation and point-of-care testing. Rapid nucleic acid extraction, library preparation, sequencing, and data analysis are examples of automated and integrated processes that improve turnaround time and reproducibility while lowering operator-dependent variability, hands-on time, and contamination risk. Near-patient testing is made possible by small, portable sequencing devices, which also facilitate prompt pathogen detection in acute clinical situations including emergency rooms, critical care units, and epidemic investigations. These systems can provide clinically actionable data in clinically relevant timeframes when paired with automated quality control, real-time analytic pipelines, and standardised reporting frameworks [28].

6. CLINICAL APPLICATIONS (APPLIED MICROBIOLOGY)

6.1 CNS Infections (Meningitis/Encephalitis)

Metagenomics is useful when culture/PCR is negative or when rare agents are suspected. It can detect viral, bacterial, fungal, and parasitic pathogens from CSF, particularly in complex cases.

6.2 Respiratory Tract Infections

Useful for atypical organisms, mixed infections, and immunocompromised patients. The major challenge is differentiating colonizers (normal flora) from true pathogens.

6.3 Sepsis and Bloodstream Infections

Plasma cell-free DNA sequencing is a promising approach, especially when blood cultures are negative or antibiotics were started early. Sensitivity can still be limited by low pathogen biomass.

6.4 Hospital Outbreak Investigation

Shotgun sequencing can support strain tracking, transmission inference, and identification of resistance determinants, complementing infection control programs.

6.5 Antimicrobial Resistance Surveillance

Metagenomics can detect resistance genes and monitor AMR trends, but reporting must clarify that gene detection does not always predict phenotypic resistance [29,30].

7. CONCLUSION

Clinical metagenomics is fundamentally reshaping applied microbiology diagnostics by delivering rapid, broad-spectrum pathogen identification and resistance profiling directly from patient samples, filling critical gaps left by culture-dependent and targeted molecular methods in complex, culture-negative, or polymicrobial infections. While remarkable progress in shotgun DNA/RNA sequencing, amplicon panels, host depletion techniques, long-read platforms, AI-driven analysis, automated workflows, and curated global databases has enhanced sensitivity, specificity, and speed, persistent challenges necessitate rigorous quality controls-including contamination mitigation via clean rooms and blanks, pipeline governance, clinical correlation for actionability, and multidisciplinary interpretation-to ensure reliable, reproducible results that directly inform therapy. Overcoming remaining translational barriers through harmonized standards, regulatory approvals, cost reductions, and real-time surveillance integration will solidify metagenomics as a routine adjunct to traditional microbiology, particularly for high-stakes scenarios like meningitis-encephalitis, ventilator-associated pneumonia, empiric antibiotic failures, and outbreak responses, ultimately improving patient outcomes, antimicrobial stewardship, and public health preparedness.

FUNDING

Nil

INFORM CONSENT AND ETHICAL CONSIDERATIONS

Not applicable

ACKNWLEDGEMENT

Not Applicable

REFERENCES

1. Chiu, C. Y., & Miller, S. A. (2019). Clinical metagenomics. *Nature Reviews Genetics, 20*(6), 341–355.
2. Xu, C., & others. (2025). Effect of interpretation of positive metagenomic next-generation sequencing results on clinical decision-making. *Open Forum Infectious Diseases, 12*(2), ofaf076.
3. Yin, C., & others. (2024). Pathogenic detection by metagenomic next-generation sequencing in spinal infections. *Orthopedic Surgery*.
4. Miller, S., & others. (2020). Retrospective review of clinical utility of shotgun metagenomic sequencing in cerebrospinal fluid. *Journal of Clinical Microbiology, 59*(2), e02593-20.
5. Li, Z., & others. (2024). Diagnostic performance of metagenomic sequencing in infectious diseases. *Journal of Infection*.

6. Cao, X., & others. (2025). Diagnostic performance and clinical utility of metagenomic next-generation sequencing. *Infection and Drug Resistance*.
7. Hong, D. K., Blauwkamp, T., Kertesz, M., Bercovich, A., & Chow, R. J. (2021). Multicenter assessment of shotgun metagenomics for pathogen detection. *BMC Biology, 19*(1), 267.
8. Gast, D. C., & others. (2023). Clinical metagenomics for infectious diseases. *Journal of Clinical Microbiology, 61*(3), e0177622.
9. Simmon, K. E., & others. (2020). Advances and challenges in clinical metagenomics. *Journal of Infectious Diseases, 221*(Supplement_3), S318–S323.
10. others. (2025). Metagenomic next-generation sequencing in infectious disease diagnostics. *Frontiers in Microbiology*.
11. Chen, S., & others. (2025). Application of metagenomic next-generation sequencing in pulmonary infections. *BMC Pulmonary Medicine*.
12. Rajendhran, J., & others. (2024). Clinical metagenomics-based diagnostics for infectious diseases. *Frontiers in Cellular and Infection Microbiology, 14*, 1459621.
13. others. (2025). Shotgun metagenomic sequencing analysis as a diagnostic tool. *Diagnostics, 15*(12), 1456.
14. Govender, K. N., & others. (2021). Metagenomic sequencing as a pathogen-agnostic clinical diagnostic tool. *PLoS Neglected Tropical Diseases, 15*(8), e0009666.
15. Blauwkamp, T. A., & others. (2019). Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nature Medicine, 25*(7), 1188–1195.
16. Wilson, M. R., & others. (2019). Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. *New England Journal of Medicine, 380*(18), 1737–1746.
17. Gu, W., & others. (2021). Optimized viral metagenomics for clinical diagnosis. *Genome Biology, 22*(1), 41
18. Charalampous, T., & others. (2019). Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory tract infection. *Nature Medicine, 25*(11), 1854–1860.
19. Street, T. L., & others. (2020). How to approach the diagnosis and management of lower respiratory tract infection. *Journal of Clinical Microbiology, 58*(7), e00421-20.
20. Leo, Y. S., & others. (2022). Clinical validation of metagenomic next-generation sequencing in routine diagnostic practice. *Clinical Infectious Diseases, 75*(1), 107–115
21. Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews, 68*(4), 669–685.
22. Chen, E. C., & others. (2017). Cross-talk between Akkermansiamuciniphila and intestinal epithelium controls diet-induced obesity. *PNAS*. (Early metagenomics influence)
23. Forbes, J. D., Knox, N. C., & Tang, P. (2019). The diagnostic accuracy of metagenomic next generation sequencing in routine clinical diagnosis. *Clinical Chemistry*.
24. Simner, P. J., & others. (2021). Development and clinical validation of a host gene expression test for distinguishing sepsis from non-infectious systemic inflammation. *Journal of Clinical Microbiology*.
25. Kalantar, K., & others. (2020). IDseq—an open source cloud-based pipeline and analysis service for metagenomic data. *bioRxiv*.
26. Farnaud, S., & others. (2023). Clinical metagenomics implementation challenges. *Journal of Molecular Diagnostics*.
27. Vissers, J. H., & others. (2024). Metagenomic next-generation sequencing outperforms cultures in identifying pathogens in pediatric osteoarticular infections. *Clinical Infectious Diseases*.
28. Zhang, H. C., & others. (2022). Metagenomic next-generation sequencing for the diagnosis of central nervous system infections: A real-world prospective observational study. *Clinical Microbiology and Infection*.
29. Wang, S., & others. (2021). Comparison of metagenomic next-generation sequencing and targeted detection for etiological diagnosis. *Journal of Infection*.
30. Parize, P., & others. (2017). Untargeted next-generation sequencing-based first-line diagnosis of infection in immunocompromised adults: A multicentre, blinded, prospective study. *Clinical Microbiology and Infection, 23*(8), 574.e1–574.e6.