Metagenomic Next-generation Sequencing: Application in Infectious Diseases

Yan Li*, Qiao Yang* and Jianqiang Ding

Abstract

Infectious diseases have always been a difficult clinical problem, especially severe infections and those infections of unknown origin. The invention of a new detection method is particularly important and urgent. Metagenomic next-generation sequencing (mNGS) is a high-throughput sequencing method that sequences microbial DNA and RNAs from fluid or solid tissue samples in hours. All important sequence data that is present in a specimen can be used for pathogen identification after a series of bioinformatics analysis. Recently, mNGS has been used in preclinical trials to find and identify pathogens from the respiratory system, central nervous system, bloodstream, and other infections. mNGS technology has advantages of being faster, accurate, and detection of unknown pathogens over conventional laboratory methods for microbial identification and detection of antimicrobial resistance and virulence markers. However, mNGS has limitations that include human source DNA and RNA removal, intra-cellular bacteria extraction difficulties, and background pollution. mNGS technologies are innovative methods, especially when a bacterial culture was negative; however, a comprehensive collection of clinical evidence is required before they move from research into clinical laboratories.

Introduction

Infectious diseases involve all systems in the body and have a high incidence in all clinical departments, which lead to sepsis, infectious shock, and even death in serious cases. The ability to quickly identify the focus of the infection and pathogenic bacteria in patients and use targeted antibiotics is essential in treatment. Bacterial culture, serological immunology, and nucleic acid examination are widely used in clinical pathogenic microbiological identification; however, they all have different limitations, such as sensitivity, specificity, and detection cycle. In some hospitals, rare and unknown microorganisms are often not identified or cannot be identified quickly and accurately. Recently, metagenomic next-generation sequencing (mNGS) has been developing rapidly and has gradually been applied in medical practice. Compared with the first generation of sequencing technology, mNGS, which is a high-throughput technology, directly measures the millions of nucleic acid DNA’s sequences in a variety of specimens (Fig. 1).

According to the sequences obtained, which are compared with database information, the pathogenic bacteria can be identified. Therefore, mNGS is a faster and more accurate method than traditional laboratory microbial detection methods that provide greater assistance in the clinical diagnosis and treatment of patients with infectious diseases. The use of genetics and genomics to promote the study of infectious diseases is an important part of precision medicine, which could be the direction of development in the future; therefore, diseases could be studied at the genetic level.

Keywords: Next-generation sequencing; Metagenomic next-generation sequencing; Infection.

Abbreviations: mNGS, metagenomic next-generation sequencing; BALF, bronchoalveolar lavage fluid; CPF, cystic pulmonary fibrosis; FTLS, fever, thrombocytopenia and hyperleukocytosis syndrome; HNF virus, Henan fever virus; PCR, polymerase chain reaction.

*Correspondence to: Jianqiang Ding, Department of General Medicine, Central Hospital of Chongqing University, Chongqing Emergency Medical Center, Chongqing, China. ORCID: https://orcid.org/0000-0003-1964-1513. Tel: +86-023-63692190, Fax: +86-023-63854632, E-mail: jding18@foxmail.com

Yan Li and Qiao Yang are co-first authors.

How to cite this article: Li Y, Yang Q, Ding J. Metagenomic Next-generation Sequencing: Application in Infectious Diseases. Explor Res Hypothesis Med 2021;00(00):00–00. doi: 10.14218/ERHM.2021.00014.

© 2021 The Author(s). This article has been published under the terms of Creative Commons Attribution-Noncommercial 4.0 International License (CC BY-NC 4.0), which permits noncommercial unrestricted use, distribution, and reproduction in any medium, provided that the following statement is provided: "This article has been published in Exploratory Research and Hypothesis in Medicine at https://doi.org/10.14218/ERHM.2021.00014 and can also be viewed on the Journal’s website at https://www.xiaephpublishing.com/journal/erhm."
The pathogens were not detected by conventional pathogenic, immunological, or molecular biological means. Partial tissue samples from the donor and recipient were analyzed by next-generation sequencing (NGS) technology, both suggested the presence of an arenavirus infection, which was later confirmed by other means. In 2010, a whole-genome sequence analysis based on NGS data from the cholera outbreak in the Haiti earthquake showed that the Vibrio cholera that triggered the outbreak was not of local origin, but was more closely related to a strain isolated from Asia and later confirmed by single nucleotide polymorphism analysis to be of Nepalese origin. In addition, NGS provides strong evidence to control the spread of pathogenic bacteria, guiding the use of antimicrobials, and analyzing the drug resistance and virulence characteristics of bacteria. Recently, due to the continuous technological development of NGS (Table 1), it has shown unique advantages and broad application prospects in infectious agent detection, pathogen biological characterization, and molecular epidemiological analysis.

**Application of metagenomic NGS in infectious diseases**

**Infection of the respiratory system**

Respiratory system infections are extremely prevalent, and through chest imaging, bacterial culture, and smears, and many patients remain with unclear infection pathogens and insensitivity to the selected antibiotics. Some patients with the rapidly progressive disease develop life-threatening respiratory failure and infectious shock. mNGS detects pathogens from specimens such as bronchoalveolar lavage fluid (BALF), lung puncture/mediastinal lymph node puncture/tracheal biopsy tissue, pleural effusion, and sputum.

**BALF specimens**

BALF is obtained by sampling during bronchoscopy. Previously, BALF from healthy humans was considered an aseptic specimen; however, with the continuous development and improvement in the level of testing, there is a bacterial community in the lung, which becomes the microbiome, and the microbiome of diseased lungs is distinctly different from that of healthy lungs. In 2018, Miao et al. found causative pathogens in 34% of BALF that was obtained from infected or noninfected patients. Takeuchi et al. performed mNGS on BALF from 10 children with respiratory failure and detected significant bacterial or viral sequences in 8 children, and eventually identified the pathogens in 3 children, which were difficult to identify by conventional methods. In addition, several cases and small sample studies of second-generation sequencing have successfully detected the causative pathogens, which led to

![Flow chart for mNGS](image-url)
Infections of the central nervous system

Neurological infectious diseases, such as encephalitis or meningitis, often have unclear etiology in cerebrospinal fluid (CSF) and might be caused by pathogenic infections; however, there are many potential pathogens (bacteria, viruses, fungi, or parasites). Traditional testing, such as routine examination, biochemical examination, and bacterial culture often failed to detect them or missed them. Guo et al.28 found that the diagnostic sensitivity of bacterial meningitis increased from 55.6% to 68.7% when mNGS was used. In a case report published by Hu et al.26 a 31-year-old HIV-infected patient admitted to hospital in a critical condition with a Glasgow score of 3 was tested for Toxoplasma gondii by mNGS on CSF. Several cases have been reported in which the causative organism was found by mNGS of CSF.27

**Infection of the bloodstream**

Infectious diseases of the blood have a high mortality rate, such as infective endocarditis, with an annual incidence of 3:100,000 to 1:10,000, and one-third of these patients died 1 year after diagnosis. The key to the treatment of infective endocarditis is the ability to rapidly identify the pathogen. Approximately 31% of patients have negative blood cultures due to early application of antibiotics or other factors, and these blood culture-negative infective endocarditis often pose multiple diagnostic and therapeutic difficulties that cause significant harm to patients.28 In 2018, Cheng et al.29 performed mNGS testing on blood specimens that were collected from seven patients with suspected infective endocarditis and confirmed that mNGS was helpful in the diagnosis and treatment of patients with infective endocarditis despite the negative blood bacterial cultures. In addition to improving the diagnostic efficacy of the 3 days routine culture, mNGS helps to explore novel and unknown pathogens. From 2007 to 2010, a large number of cases of fever, thrombocytopenia, and hyperleukocytosis syndrome (FTLS) occurred in Henan Province. Xu et al.30 performed mNGS on the sera of patients and found that this was a new bunyavirus, eventually named Henan fever virus (HNF virus). However, bacterial nucleic acids have been reported in healthy volunteers tested by mNGS, which could be potential contamination of the sampling process, and therefore, requires careful screening in clinical applications.31

**Infections of the digestive system**

Acute cholecystitis, acute gastroenteritis, and liver abscess are

---

**Table 1. Sequencing platforms for the detection of pathogens**

| Sequencing Platforms | Principle | Advantage | Disadvantage |
|----------------------|-----------|-----------|--------------|
| Solid 5500 x 15 | Detect fluorescent signals based on sequential ligation of fluorescent probes | Realization of double base correction, proofreading while reducing initial data errors | Long sequencing time, partial read length and reference sequence cannot be matched |
| Illumina | In PCR, a DNA fragment is attached to a primer for the reaction | Short sequencing time, low cost and better uniformity | Short sequencing read length, sequencable sample size is limited |
| Ion Torrent | The technology is based on the use of semiconductor materials, semiconductor materials detect the release of H+ protons when DNA synthesis | Short sequencing time (2.5–4 h), low cost | Single-end sequencing |
| BGI | Sequencing while connecting, combinatorial probe anchor synthesis | Low cost, large amount of data, large-scale whole-genome sequencing | Read length is short (maximum read length 50–100 bp), poor operability of gene splicing |

PCR: polymerase chain reaction; BGI: Beijing Genomics Institution.
common infections of the digestive system. In 2018, Wu et al. performed mNGS on ascites and blood samples from 97 patients with acute abdominal disease; however, the bacteria analyzed by mNGS only partially matched those isolated using traditional culture methods and concluded that mNGS provided limited detection abilities of pathogens in ascites and blood specimens from patients with the acute abdominal disease. In a study on acute cholecytitis, metagenomic analysis was effective for rapid the diagnosis of the causative agent of acute cholecytitis, which included the assessment of potential antibiotic susceptibility. In addition, mNGS is useful for the detection of infectious diarrhea pathogens. mNGS was used to study the association between gut microbial alterations with type 2 diabetes and the association between atherosclerosis with the intestinal metagenome.

Infections of the ophthalmology

Diagnosis of ocular infections relies on intraocular fluid culture or polymerase chain reaction (PCR). In 2012, Eleinen et al. subjected corneal scraping samples that were collected from 88 patients with infectious corneal ulcers to bacterial culture, fungal culture, Gram stain, potassium hydroxide wet mount, and broad-range PCR with primer pairs targeted to the 16S (bacterial) and 18S (fungal) RNA genes. The results showed that for bacterial and fungal keratitis, the sensitivity of the culture was 57.58% and 59.09%, respectively and the PCR sensitivity was 87.88% and 90.91%, respectively. In 2017, Doan et al. detected PCR-positive and PCR-negative specimens from intraocular infections by mNGS and concluded that the mNGS results were in good agreement with the PCR results and mNGS could be used to detect bacteria and fungi and viruses and provided information on drug resistance.

Infections of the urinary system

Urinary tract infections have a high incidence and recurrence rate, and diagnosis mainly rely on urine culture. The most common pathogenic bacteria is Escherichia coli, followed by other Enterobacteriaceae, staphylococcus, streptococcus, fungi, and some pathogens that are difficult to cultures, such as viruses and tuberculosis. Moustafa et al. found that it was difficult to identify a large number of microorganisms in urine using traditional culture methods and that mNGS provided a more comprehensive and quantitative analysis of the microbial community in urine. Barraud et al. demonstrated that mNGS provided drug resistance information in addition to the analysis of the microbiota in urine. In addition, mNGS simultaneously detects viruses, fungi, Candida albicans, cryptococcus, Mycobacterium tuberculosis, and mycoplasma in urine specimens and the virulence and drug resistance of pathogens.

mNGS versus traditional detection methods

The cost of traditional detection methods is lower and these methods are mature and relatively simple to operate. The sensitivity and specificity of traditional detection methods are poor, and they require a long detection time. mNGS has the following advantages over traditional detection methods: (1) there is no bias, for example, if a specimen has C. albicans and Pneumocystis sp., faster growth of Pneumocystis sp. will affect the growth of C. albicans. Therefore, pathogenic cultures are biased and mNGS is unbiased for pathogenesis due to its different detection mechanism; (2) for infectious diseases, clinical symptoms, signs, history, epidemiology, and risk factors are used to infer which pathogens are probably present and to select the appropriate tests, but incomplete information in any one of the tests could lead to missed pathogens. Theoretically, mNGS detect the gene sequences of all pathogens in a specimen, and therefore, has broader coverage than other testing solutions; and (3) mNGS detects multiple pathogens simultaneously and in a shorter time.

Limitations of mNGS application

Despite the increasing clinical use and continuous enhancement of mNGS technology, the limitations or drawbacks of mNGS remain a major issue that cannot be ignored. The main disadvantage of mNGS is that most nucleic acids in patient specimens predominantly originate from human host backgrounds, with the majority of detected sequences (typically >99%) coming from human hosts, which limits the overall analytical sensitivity of pathogen detection due to the relatively low number of sequenced microbial detected sequences (reads). This drawback is dealt with by targeted sequencing. In addition, other disadvantages include nucleic acid contamination in reaction kits, or from colonized bacteria and laboratory procedures. Further research and optimization are required for mNGS sequencing depth, bioinformatics analysis, and clinical interpretation.

The mNGS test cannot distinguish background, colonized microbes from pathogenic organisms, and requires a clinician to decide based on the patient’s clinical manifestation, combined with other laboratory examination results.

Other limitation of mNGS use includes relatively low sensitivity for the detection of intracellular infected bacteria (e.g., M. tuberculosis, Legionella sp., and Brucella sp.), thick-walled bacteria and fungi, due to their rare release into body fluids. In addition, for pathogens with thick cell walls, such as Gram-positive bacteria, M. tuberculosis, fungi and parasites, the low efficiency of nucleic acid extraction results in lower positivity. RNA is susceptible to degradation and requires very low-temperature storage conditions that should be used in RNA sample transit and storage. Recently, several mNGSs have been reported for drug resistance detection, they are mainly for the study of individual pathogens, and difficulties remain when using mNGS to detect drug resistance genes in pathogenic bacteria simultaneously.

Future directions

Infectious diseases have always been a major threat to human life and health, and the continuous renewal of antibiotics has significantly reduced mortality. However, due to the development of human society, increase in human life expectancy, increase in malignant tumor prevalence, and the widespread use of glucocorticoids and immunosuppressant, infectious diseases have displayed the following characteristics: (1) an increase in the rate of severe infections; (2) an increase in rare or unknown pathogens; and (3) a significant increase in the virulence and drug resistance of pathogens. However, the widespread and early application of broad-spectrum antibiotics and because 20%–60% of human-associated microorganisms are not culturable, leads to a low positive rate in traditional cultures. The failure to identify pathogens has increased
the difficulty in the diagnosis and treatment of infectious diseases, significantly prolonging the patient’s treatment cycle and increasing morbidity and mortality. However, mNGS has disadvantages, such as it is expensive and might not provide a comprehensive or accurate report of antibiotic resistance, because of technical issues; however, due to the continuous improvement in the technology, mNGS remains a promising and practical means of clinical detection to help the clinical treatment of infectious diseases. The development of databases, the genetic profiling of rare pathogens, studies into the mechanisms of pathogens, and the establishment of standardized processes for mNGS require further research.

**Conclusions**

mNGS obtains nucleic acid sequence information on pathogens directly from the specimen without the need for bacterial culture to identify them. It is not influenced by the application of antibiotics and has higher sensitivity and specificity than traditional detection methods. In addition, because mNGS sequences millions of nucleic acids at one time it can detect bacteria, fungi, viruses, mycoplasma, chlamydia, and parasites in specimens simultaneously, which is not possible using other clinical detection methods.

**Acknowledgments**

None.

**Funding**

None.

**Conflict of interest**

The authors have no conflicts of interest to declare.

**Author contributions**

Study design (YL, QY), manuscript writing (YL), critical revision (JD). All authors have made a significant contribution to this study and have approved the final manuscript.

**References**

[1] de St Maurice A, Ervin E, Chu A. Ebola, Dengue, Chikungunya, and Zika Infections in Neonates and Infants. Clin Perinatol 2021;48(2):311–329. doi:10.1016/j.clp.2021.03.006.

[2] Patel RH, Acharya A, Chand HS, Mohan M, Byrareddy SN. Human Immunodeficiency Virus and Severe Acute Respiratory Syndrome Coronavirus 2 Coinfection: A Systematic Review of the Literature and Challenges. AIDS Res Hum Retroviruses 2021;37(4):266–282. doi:10.1089/aid.2020.0284.

[3] Duan H, Li X, Mei A, Li P, Liu Y, Li X, et al. The diagnostic value of metagenomic next-generation sequencing in infectious diseases. BMC Infect Dis 2021;21(1):62. doi:10.1186/s12879-020-06746-5.

[4] Shean RC, Greninger AL. One future of clinical metagenomic sequencing for infectious diseases. Expert Rev Mol Diagn 2019;19(10):849–851. doi:10.1080/14737159.2019.1658524.

[5] Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 1995;269(5223):496–512. doi:10.1126/science.7542800.

[6] Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med 2008;358(10):991–998. doi:10.1056/NEJMoa0737885.

[7] Hendriksen RS, Price LB, Schupp JM, Gilleece JD, Kaas RS, Englethaler DM, et al. Population genetics of Vibrio cholerae from Nepal in 2010: evidence on the origin of the Haitian outbreak. mBio 2011;2(4):e00157–11. doi:10.1128/mBio.00157–11.

[8] Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, et al. The origin of the Haitian cholera outbreak strain. N Engl J Med 2011;364(1):33–42. doi:10.1056/NEJMoa1012928.

[9] Gilmour MW, Graham M, Van Domselaar G, Tyler S, Kent H, Trut Yakel KM, et al. High-throughput genome sequencing of two Listeria monocytogenes clinical isolates during a large foodborne outbreak. BMC Genomics 2010;11:120. doi:10.1186/1471-2164-11-120.

[10] Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantarattin N, et al. Evolution of MRSA during transmission and intercontinental spread. Science 2010;327(5964):469–474. doi:10.1126/science.1182395.

[11] Lewis T, Loman NJ, Bingle L, Jumaa P, Weissstock GM, Mortiboys D, et al. High-throughput whole-genome sequencing to dissect the epidemiology of Acinetobacter baumannii isolates from a hospital outbreak. J Hosp Infect 2010;75(1):37–41. doi:10.1016/j.jhin.2010.01.012.

[12] Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Schweitz F, et al. Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. N Engl J Med 2011;365(8):709–717. doi:10.1056/NEJMoa1106920.

[13] Gardy JL, Johnston JC, Ho Sui SJ, Cook VK, Shah L, Brodkin E, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med 2011;364(8):730–739. doi:10.1056/NEJMoa1003176.

[14] Snitkin ES, Zelazny AM, Thomas PJ, Stock F, NISC Comparative Sequencing Program Group, Henderson DK, et al. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. Sci Transl Med 2012;4(148):148ra116. doi:10.1126/scitranslmed.3004129.

[15] Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, et al. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res 2008;18(7):1051–1063. doi:10.1101/gr.076463.108.

[16] Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456(7218):53–59. doi:10.1038/nature07517.

[17] Borewicz K, Pragman AA, Kim HB, Hertz M, Wendt C, Isaacson RE. Longitudinal analysis of the lung microbiome in lung transplantation. FEMS Microbiol Lett 2013;339(1):57–65. doi:10.1111/1574-6968.12053.

[18] Miao Q, Ma Y, Wang Q, Pan J, Zhang Y, Jin W, et al. Microbiological Diagnostic Performance of Metagenomic Next-generation Sequencing When Applied to Clinical Practice. Clin Infect Dis 2018;67(suppl_2):S231–S240. doi:10.1093/cid/ciy693.

[19] Takeuchi S, Kawada J, Horiba K, Okuno Y, Okumura T, Suzuki T, et al. Metagenomic analysis using next-generation sequencing of pathogens in bronchoalveolar lavage fluid from pediatric patients with respiratory failure. Sci Rep 2019;9(1):12909. doi:10.1038/s41598-019-49372-x.

[20] Fang X, Xu M, Fang Q, Tan H, Zhou J, Li Z, et al. Real-time Utilization of Metagenomic Sequencing in the Diagnosis and Treatment Monitoring of an Invasive Adenovirus B55 Infection and Subsequent Herpes Simplex Virus Encephalitis in an Immunocompetent Young Adult. Open Forum Infect Dis 2018;5(6):ofy114. doi:10.1093/ofid/ofy114.

[21] Fischer N, Rohde H, Indenbirken D, Günther T, Reumann K, Lütgethiem K, et al. The origin of the Haitian outbreak strain. N Engl J Med 2011;364(1):33–42. doi:10.1056/NEJMoa1012928.

[22] Li Y, Deng X, Hu F, Wang J, Liu Y, Huang H, et al. Metagenomic analysis identified co-infection with human rhinovirus C and bocavirus 1 in an adult suffering from severe pneumonia. J Infect 2018;76(3):311–313. doi:10.1016/j.jinf.2017.10.012.
