

Metagenomics characterization of respiratory viral RNA pathogens in children under five years with severe acute respiratory infection in the Free State, South Africa

Ayodeji E. Ogunbayo¹ | Milton T. Mogotsi¹ | Hlengiwe Sondlane¹ |
Saheed Sabiu² | Martin M. Nyaga¹

¹Next Generation Sequencing Unit and Division of Virology, Faculty of Health Sciences, University of the Free State, Bloemfontein, South Africa

²Department of Biotechnology and Food Science, Durban University of Technology, Durban, South Africa

Correspondence

Martin M. Nyaga, Next Generation Sequencing Unit and Division of Virology, Faculty of Health Sciences, University of the Free State, Bloemfontein 9300, South Africa.
Email: nyagamm@ufs.ac.za

Funding information

Poliomyelitis Research Foundation; Bill and Melinda Gates Foundation; South African Medical Research Council; National Research Foundation

Abstract

Prompt detection of viral respiratory pathogens is crucial in managing respiratory infection including severe acute respiratory infection (SARI). Metagenomics next-generation sequencing (mNGS) and bioinformatics analyses remain reliable strategies for diagnostic and surveillance purposes. This study evaluated the diagnostic utility of mNGS using multiple analysis tools compared with multiplex real-time PCR for the detection of viral respiratory pathogens in children under 5 years with SARI. Nasopharyngeal swabs collected in viral transport media from 84 children admitted with SARI as per the World Health Organization definition between December 2020 and August 2021 in the Free State Province, South Africa, were used in this study. The obtained specimens were subjected to mNGS using the Illumina MiSeq system, and bioinformatics analysis was performed using three web-based analysis tools; Genome Detective, One Codex and Twist Respiratory Viral Research Panel. With average reads of 211323, mNGS detected viral pathogens in 82 (97.6%) of the 84 patients. Viral aetiologies were established in nine previously undetected/missed cases with an additional bacterial aetiology (*Neisseria meningitidis*) detected in one patient. Furthermore, mNGS enabled the much needed viral genotypic and subtype differentiation and provided significant information on bacterial co-infection despite enrichment for RNA viruses. Sequences of nonhuman viruses, bacteriophages, and endogenous retrovirus K113 (constituting the respiratory virome) were also uncovered. Notably, mNGS had lower detectability rate for severe acute respiratory syndrome coronavirus 2 (missing 18/32 cases). This study suggests that mNGS, combined with multiple/improved bioinformatics tools, is practically feasible for increased viral and bacterial pathogen detection in SARI, especially in cases where no aetiological agent could be identified by available traditional methods.

KEYWORDS

children, metagenomics, nasopharyngeal swab, next-generation sequencing, severe acute respiratory infection, viral RNA pathogens

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Journal of Medical Virology* published by Wiley Periodicals LLC.

1 | INTRODUCTION

Severe acute respiratory infections (SARI) are a leading cause of morbidity and mortality, especially in vulnerable new-borns and young children.^{1–3} The detection of viruses and bacteria implicated in respiratory infections could be challenging due to their nonspecific clinical presentation nature and the possibility of co-infection.^{4,5} The traditional testing methods, such as culture, have limitations, including low sensitivity and are time demanding.^{6,7} Also, molecular approaches such as real-time multiplex PCR (mRT-PCR) can only detect targeted pathogens with limited genotyping capacity.^{1,8} These limitations, combined with the possible emergence of novel or atypical pathogens, underscore the need for less restrictive/constrained methods with improved pathogen detection capacity and sufficient genomic information that could assist in proficient diagnosis and possible outbreak management.^{4,9,10}

On the other hand, however, metagenomic next-generation sequencing (mNGS) has helped to integrate and expand upon numerous benefits of molecular testing and culture-based techniques.^{7,11–14} For instance, host-and pathogen-derived nucleic acids can be sequenced without a prior knowledge, enabling simultaneous detection of a potentially limitless number of microorganisms with sequence homology to reference sequences.^{13,15,16} In addition, mNGS-based detection methods offer considerable diagnostic potential because alternative causes can be detected or ruled out with better certainty.^{17,18} Furthermore, mNGS enables pathogen subtyping/genotyping, identification of drug resistance markers, and molecular-based epidemiological studies.^{9,19,20}

Although recent studies have demonstrated the power of mNGS in diagnosing respiratory infection in different parts of the world,^{6,9,14,21} the current data from South Africa/Africa at large on the clinical diagnostic utility of mNGS in SARI are still limited,²² especially in comparison to the performance of mRT-PCR in paediatric patients. Additionally, the increasing need for prompt analysis of generated mNGS data has birthed several user-friendly, nontechnical, yet sensitive web-based mNGS bioinformatic analysis tools.^{23,24} Nevertheless, different bioinformatic tools may produce varying results for the same FASTQ file.²⁴ Therefore, evaluating the viral detectability of some of these mNGS tools is imperative.

In this study, we report the performance of mNGS with three rapid web-based bioinformatic tools for detecting respiratory viruses in a cohort of paediatric subject's ≤ 5 years admitted with SARI in the Free State Province, South Africa. While the pathogen profile of these patients has just been recently reported using a new 21 target mRT-PCR (QIAstat-Dx- Respiratory SARS-CoV-2 panel),²⁵ the current study evaluated and further expanded the clinical diagnostic utility of mNGS in children with SARI and established a benchmark for future research in the study settings.

2 | METHODOLOGY

2.1 | Study settings and demography

The patients, children ≤ 5 years of age, admitted with SARI were recruited from Botshabelo Hospital, Pelonomi Hospital, National Hospital, and Universitas Academic Hospital in the Free State Province, South Africa. The World Health Organization (WHO) case definitions for SARI²⁶ were implemented.

2.2 | Ethical considerations

Institutional (number UFS = HSD2019/1129/2910) and Provincial ethical approvals were processed and obtained before commencement of the study. Information document with a clear description of the study as well as permission and consent form in the language of choice was provided for the parent/guardian of the children recruited for the study.

2.3 | Patient enrollment and sample collection

Patient's enrollment and sample collection are as previously described,²⁵ Briefly, a total of 84 children admitted with SARI as per the WHO definition, whose nasopharyngeal swab samples in viral transport media (BD Diagnostics) were collected between December 2020 and August 2021 and had previous mRT-PCR result, were included in this study.

2.4 | Sample processing, enrichment, and extraction of RNA viruses

Briefly, 1000 μ L of the VTM (BD Diagnostics) was centrifuged at 10 000g for 10 min to remove cellular debris. The supernatant was filtered through a 0.22 μ m filter to remove the remaining possible host cellular debris and bacteria. The filtrate was treated with a nuclease mixture of 0.1U_L-1 Turbo DNase (Life Technologies), 0.1U_L-1 RNase One (Promega) and 1X DNase buffer (Life Technologies) and incubated at 37°C for 90 min, to remove nonincorporating nucleic acid. Nucleic acid extraction was performed using the PureLink® viral RNA/DNA Mini Kit (ThermoFisher Scientific) following the manufacturer's instructions but without using carrier RNA. Viral-extracted RNA was eluted in 50 μ L of DNase/RNase-free water.

2.5 | Pathogen analysis using mRT-PCR

Pathogen profiling in the samples collected was performed as previously described,²⁵ using QIAstat-Dx Respiratory-SARS-CoV-2 Panel (Qiagen). The panel detects, generates cycle threshold (Ct) value, and differentiates

nucleic acid from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and other 20 respiratory pathogens. The assay was performed as per the manufacturer's instructions.

2.6 | DNase treatment, ribosomal RNA (rRNA) depletion and purification

The extracted RNA was quantified using the Qubit RNA High Sensitivity (HS) Assay Kit (ThermoFisher Scientific) with the Qubit® 3.0 Fluorometer (ThermoFisher Scientific) as per the manufacturer's instructions. The quantified RNA was treated with TURBO DNA-free™ reagents (ThermoFisher Scientific) according to the manufacturer's instructions. The DNase-treated samples were purified using the RNeasy Mini Kit (Qiagen) as per the manufacturer's instructions. Depletion of rRNA was performed to decrease the human, mouse, and rat rRNA by using NEBNext rRNA depletion kit (New England Biolabs) according to the manufacturer's instructions.

2.7 | Reverse transcription and whole transcriptome amplification

For library preparation purposes, isolated viral RNA was reverse-transcribed to generate complementary DNA (cDNA). For cDNA synthesis and random amplification, a QIASeq FX Single Cell RNA Library Preparation Kit (Qiagen) was used as per the manufacturer's instructions. Amplified cDNA was subsequently quantified using the Qubit™ 1X dsDNA HS Assay Kits with the Qubit 3.0 fluorometer (ThermoFisher Scientific) as per the manufacturer's instructions and normalized to the required 100–1000 ng/μL input DNA in 10 μL volume.

2.8 | Library preparation and next-generation sequencing

Libraries were prepared using the QIASeq FX Single Cell RNA Library Kit (Qiagen) according to the manufacturer's instructions. The DNA libraries were generated and analysed for average fragment size distribution using the Agilent 2100 Bioanalyzer (Agilent Technologies). Metagenomic sequencing was performed on the Illumina MiSeq system with the reagent kit v3 (Illumina) for 600 cycles to generate 2 × 250 bp paired-end reads.

2.9 | Bioinformatic analysis for pathogen detection

The generated paired-end reads were analysed using three mNGS analysis tools, including the Genome detective (<https://www.genomedetective.com/>); an automated web system for virus identification from high-throughput sequencing data,²⁷ One codex analysis platform, where detected organisms were displayed as high abundance (HA) (>25% of the sample), medium abundance (5%–25% of the sample), and low

abundance (0.5%–5% of the sample),²⁸ and the Twist Respiratory Virus Research Panel (TRVRP). The TRVRP is a culture-free, hybrid capture enrichment workflow embedded as a sub analysis panel in One Codex analysis platform and is capable of detecting 29 common human respiratory viruses simultaneously. Briefly, for the TRVRP, raw sequencing reads (FASTQ) were aligned to each individual viral reference and viruses were reported as “detected” if ≥20% of the genome was observed with an overall average of ≥10x depth. Respiratory viruses were reported as “indeterminate” if there is coverage of ≥5% of the genome, but the data falls short of that required for a “detected” call. All other viruses were noted as “not detected” but the corresponding number of reads and genome coverage are still made available. A BLAST analysis of matching sequences was further performed for some of the detected pathogens (those with very few reads/genomes coverage) to exclude false positive results. Characteristic details of the mNGS analysis tools used in this study are described in Table S1.

2.10 | Inclusion of negative control

To evaluate for contamination, a no template control (nuclease free water) was subjected through the same mNGS workflow as the samples to assess the presence of cross-contamination or kitome contamination from reagents.

2.11 | Statistical analysis

Raw data were calculated using Microsoft Excel 365 version 2206, and results are presented as simple percentages, mean, and standard deviations of replicate determinations.

3 | RESULTS

3.1 | Cohort clinical and demographic characteristics

The detailed demographic and clinical characteristics of the participants enrolled in this study are as previously described.²⁵ Demographically, 34 (40.4%) were females, while 50 (59.6%) were males. The ages of the participants ranged from 0 to 60 months, with most participants, 57 (67.8%) aged 0–24 months. Also, 35 (41.6%) of the participants required oxygenation, and 17 (20%) were HIV infected, 24 (28.5%) had chest indrawing, 9 (10.7%) had feeding difficulty, 3 (3.5%) were asthmatic, and none of them required the need for admission into the intensive care unit.

3.2 | Viral detection by mNGS

Sequencing of the cDNA libraries from all samples generated an average of 211323 reads after quality trimming and filtering (as per

genome detective). Based on the results of the three analysis tools, mNGS detected viral pathogen/s in 82 (97.6%) of 84 participants included in the study. Further insight into the results revealed that respiratory syncytial virus (RSV) B (32.5%), human rhinovirus (HRV) C (14.3%), HRV A/89 (16%), and SARS-CoV-2 (10%) were the most frequent pathogens detected (Figure 1 and Table S2). Of the 82 positive samples, single, double, triple, quadruple, and quintuple viral co-infections were detected in 33% ($n = 27$), 32% ($n = 26$), 19.5% ($n = 16$), 13.4% ($n = 11$), and 2.4% ($n = 2$) patients, respectively (Table S2). The highest genome coverages were mostly for RSV. The detailed mNGS results based on the three analysis tools is presented in Table S2, and additional detection spectrum of mNGS is shown in Table 1.

3.3 | Viral pathogen detectability of multiple mNGS analysis tools

The use of multiple analysis tool increased the viral detection rate of mNGS. Using the detection of at least one viral respiratory pathogen as a positivity ratio, TRVRP detected more viral co-infections, detected more viral pathogen present at low abundance (Ct value ≥ 35) and had the highest positive detection rate (98%) compared to One Codex (48%) and Genome detective (71%) (Figure 2 and Table 2).

3.4 | Bacteria detection

Via One Codex Analysis, bacterial co-infection was detected in 24/84 (28.5%) patients, most of which were *Streptococcus pneumoniae* 8/84 (9.5%), *Klebsiella pneumoniae* 3/84 (3.5%), *Escherichia coli* 2/84 (2.3%), and one each of the remaining bacteria pathogens shown in Figure 3. Complete data on reads and abundance is shown in Table S2. The remaining two analysis tools (Genome Detective and TRVRP) only detects viral sequences.

3.5 | Detection of nonrespiratory/nonhuman viruses and bacteriophages

In addition to the detected common respiratory viruses. Reads for other viruses constituting the respiratory virome were detected. This includes human endogenous retrovirus k113, plant, fungal, animal, and bacteriophages that primarily infect bacteria known to colonize the upper respiratory tract (Table S3).

3.6 | Evaluation of the negative control

No respiratory or human-pathogenic viral reads were detected in the negative control (nuclease free water) across the three mNGS analysis tools. However, One Codex analysis tool detected some bacteria reads in the negative control (Table S4). These bacteria reads at low abundance were filtered out of the actual results during bacteria analysis.

4 | DISCUSSION

The respiratory tract is a target of many human viruses, especially RNA viruses. The current molecular techniques for detecting respiratory viruses are largely targeted-dependent tests, which detect a limited number of viruses. Therefore, it often leaves clinicians with negative results and the question of alternative aetiology for which testing was not done.¹¹ The untargeted nature of mNGS allows the query of theoretically unlimited number of pathogens in parallel, resulting in increased viral detection and a higher positivity rate. The detected percentage average reads (4.8%) in this study was in the same range (2%) for similar sample types previously reported,²⁹ but higher than the 0.32% and 0.05% reported by Romero-espinoza et al.,³⁰ and Yang et al.,³¹ respectively. Also, the observation regarding the performance of mNGS (97.6%) viral

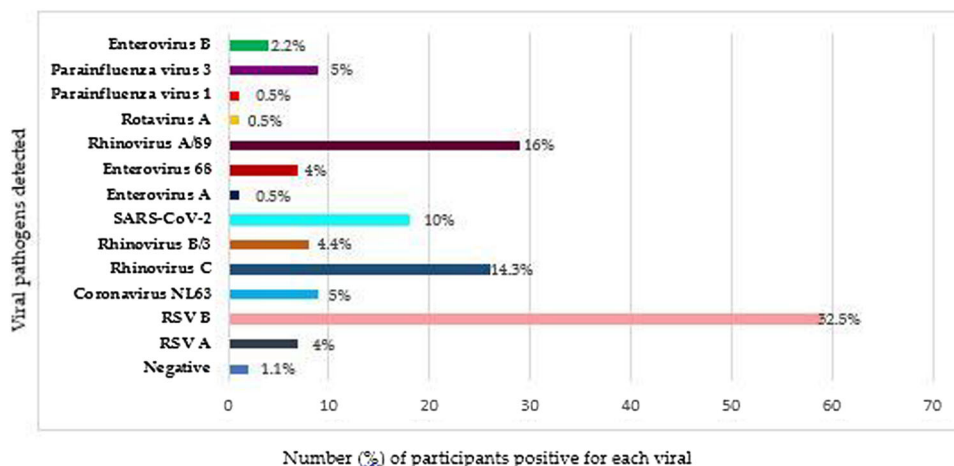


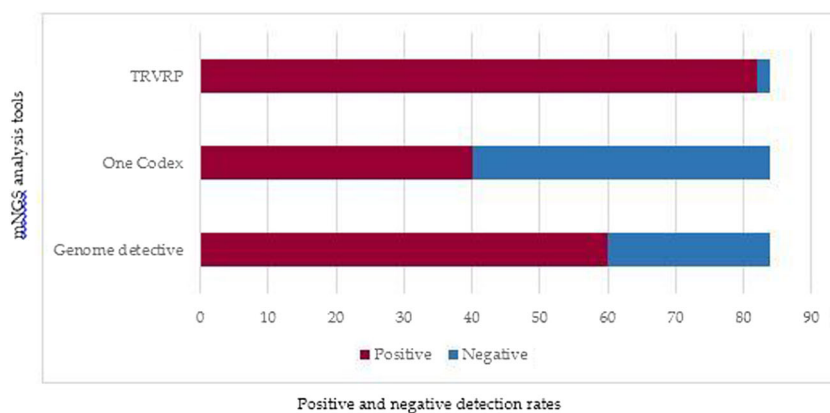
FIGURE 1 Clustered bar chart of the percentage of patients positive for the detected respiratory viruses via mNGS using the combined results of the three analysis tools. mNGS, metagenomics next-generation sequencing.

TABLE 1 Detection spectrum of the viral pathogens using mNGS versus mRT-PCR.

Detection spectrum	Number of patients (%)
Both methods detected exact number of viral pathogens	22 (26.1)
mNGS detected all viral pathogen detected by mRT-PCR with at least 1 additional.	15 (18)
mRT-PCR detected all viral pathogen detected by mNGS with at least 1 additional	19 (22.6)
No viral pathogen detected by mRT-PCR but detected by mNGS	9 (10.7)
No viral pathogen detected by mNGS but detected by mRT-PCR	1 (1.2)
mNGS detected more viral pathogen but missed at least 1 detected by mRT-PCR	7 (8.3)
mRT-PCR detected more viral pathogen but missed at least 1 detected by mNGS	3 (3.5)
Negative in both methods	1 (1.2)
Both detected one or more viruses but missed by the other	6 (7.1)
Both detected totally different viruses	1 (1.2)

Abbreviations: mNGS, metagenomic next-generation sequencing; mRT-PCR, multiplex real time polymerase chain reaction.

FIGURE 2 Viral pathogen detectability of the three mNGS analysis tools; Genome Detective, One Codex, and TRVRP (Twist Respiratory Virus Research Panel). mNGS, metagenomics next-generation sequencing.



pathogen detection rate in this study is similar to those previously reported by Graf and colleagues (91.8%)⁹ but higher than the 43%, 32.6%, and 63% reported by Thorburn et al.,¹⁹ Madi et al.,²⁹ and Graf et al.,⁹ respectively. The variability in the reported detection rate may be due to several factors that tend to influence the amount of virus in each sample. These factors include location, sampling technique, and preanalytical sample preparation. All of which may greatly impact the relative amount of the viral genome available to be sequenced.³² Enrichment of viral nucleic acids from clinical samples before sequencing as adopted in this study is thus crucial.

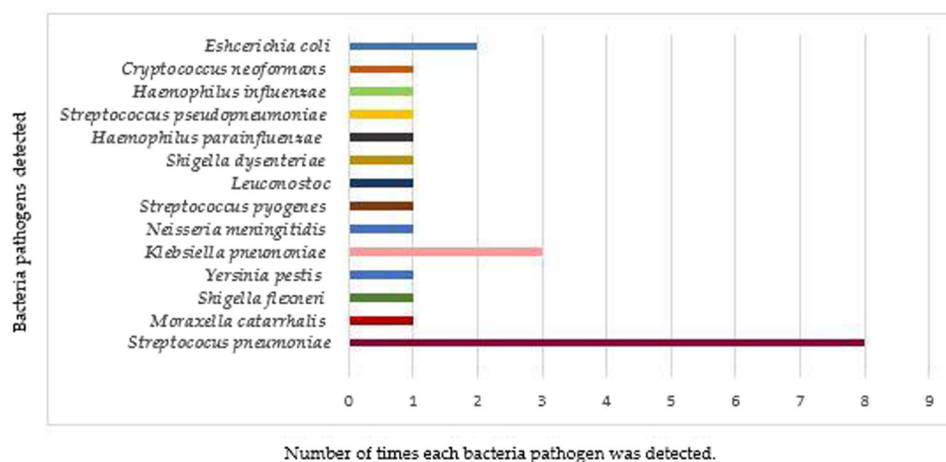
The need for immediate molecular subtyping of viral species continues to garner interest; this is due to the several advantages of subtyping, including identification of genotypic markers of drug resistance or pathogenicity. In this study, relevant information derived from mNGS typing included the following: (i) almost 80% of HRV in this study belonged to HRV A and the more pathogenic species C; (ii) RSV-B was more prevalent than RSV-A, which may be relevant, as strain-specific differences in pathogenicity have been

suggested³³ or may be due to infection prevalence of RSV-B in the community; (iii) detection of Enterovirus 68 subtype in 4% of the patients; a pathogen once linked to more severe and fatal respiratory illnesses than other enterovirus genotypes³⁴; (iv) detection of Enterovirus B subtype in 4 patients; a pathogen previously linked with viral meningitis in children and adults.³⁵ The specific identification of these viruses by mNGS (especially of HRV and Enterovirus with sequence similarities in the commonly used conserved targets for molecular assays) can have implications for the supportive management of patients and can become more significant when specific antiviral drugs become available.³⁶ Moreover, as genotype-phenotype correlations become better understood, genotypic strain characterization will gain importance and may further facilitate epidemiologic investigations or studies of vaccine effectiveness.^{9,37,38} Furthermore, the uncovered respiratory virome (constituting bacteriophages, plant virus, animal virus, human endogenous retrovirus k113 and algae viruses) in each participant in this study reflects the transient or permanent

TABLE 2 Viral pathogen detection spectrum of the three mNGS analysis tools (Genome Detective, One Codex, and TRVRP [Twist Respiratory Virus Research Panel]).

Detection spectrum	Number of subjects
Instances where similar number of pathogens were detected by the three methods per subjects	12
Instances where the three methods detected different viral pathogens	3
Instances where the three methods had no viral pathogen detected	2
Instances where similar number of pathogens were detected by Genome detective and TRVRP	9
Instances where similar number of pathogens were detected by One Codex and TRVRP per subject.	4
Instances where Genome Detective detected more viral pathogens per subjects	3
Instances where One Codex detected more viral pathogens per subjects	3
Instances where TRVRP detected more viral pathogens per subjects	30
Instances where TRVRP detected at least a viral pathogen where others had no viral detection	18
Instances where Genome Detective detected at least a viral pathogen where others had no viral detection	0
Instances where One Codex detected at least a viral pathogen where others had no viral detection	0
Instances where a pathogen present at very low abundance (as seen by mRT-PCR Ct value ≥ 35.0) was detected by Genome Detective	0
Instances where a pathogen present at very low abundance (as seen by mRT-PCR Ct value ≥ 35.0) was detected by One Codex	0
Instances where a pathogen present at very low abundance (as seen by mRT-PCR Ct value ≥ 35.0) was detected by TRVRP	6

Abbreviations: mNGS, metagenomic next-generation sequencing; mRT-PCR, multiplex real time polymerase chain reaction.

**FIGURE 3** A bar chart of the detected bacteria pathogens using One Codex mNGS analysis platform. mNGS, metagenomics next-generation sequencing.

colonization of nonhuman viruses in the respiratory milieu. While some of these nonhuman viruses were sparsely detected, reads for bacteriophages were consistently detected at varying amount in all participants. This support previous reports on HA of bacteriophages in the respiratory tract with the potential to prime and modulate host immune responses in children.^{39,40} Additionally, mNGS further enabled the simultaneous detection of bacteria co-infection in 24 (28.5%) patients despite enrichment for RNA viruses. This phenomenon is due to the significantly known high representation of bacterial DNA including pathogenic and commensals bacteria which

may be present in clinical samples, and often outnumbers viral nucleic acids. Despite strict sample enrichments for viral RNA, present bacterial nucleic material is still often carried over and may be amplified during the nonspecific whole transcriptome amplification step. Nonetheless, the ability of mNGS to concurrently detect pathogenic bacterial nucleic material in a single test despite sample enrichment for RNA viruses as shown in this study further points to the power of syndromic testing of mNGS⁴¹ which could be harnessed and may futuristically serve as an add-on test to several culture and PCR-based tests.⁹

Recently, mNGS reports have been benchmarked against traditional methods such as mRT-PCR or culture to further establish the utility of mNGS in clinical diagnostics. Compared with the 97.6% detectability rate in this study by mNGS, mRT-PCR results in the same participants yielded 88% positivity detection rate.²⁵ Besides, mNGS identified more viral pathogens (179 vs. 154) than the mRT-PCR. Further, mNGS identified at least a viral pathogen in 9 of 10 patients who were negative to the 21 targeted pathogens by mRT-PCR.²⁵ These results are in line with the previous findings, confirming that mNGS detection can improve the positive rate of pathogen detection. A study by Lewandowska³² and colleagues confirmed a viral aetiology of infection by mNGS approach in four patients despite initial negative results in specific multiplex PCR. Another study by Li et al.⁴² performed NGS detection on BALF samples from 32 patients, and the positive detection rate of NGS reached 100%. Zhang et al.⁴³ also reported that the total positive rate detected by mNGS method (91.1%) in the diagnosis of acute respiratory distress syndrome was significantly higher than that detected by the culture method (62.2%). Similarly, Huang and colleagues compared NGS with traditional pathogen detection methods in diagnosing peripheral pulmonary infectious lesions and reported a detection accuracy of 86.25% vs 43.75% for mNGS and traditional detection methods, respectively.⁴⁴ Despite the reported increased favorable detection rates reported for mNGS in this study and other studies, lower performances of mNGS compared to PCR has been documented. A study by Thorburn et al.¹⁹ reported 43% versus 54% positivity rates for NGS and RT-PCR, respectively. Similarly, Madi et al.²⁹ reported 75.6% versus 32.6% positivity rates for mRT-PCR and mNGS, respectively. However, both studies highlighted the several advantages of mNGS over the mRT-PCR and suggested that lower abundance of the viruses detected by RT-PCR could have resulted in a lower detection rate by mNGS.

Furthermore, corresponding with the previously reported data from the mRT-PCR assay in the same patients,²⁵ RSV and HRV were the most predominant viral pathogens detected in the respiratory samples by the mNGS-based approach. This observation agrees with others who showed that RSV and HRV are the most prevalent viruses in children with SARI.^{4,8,45,46} Notably, the mRT-PCR detected 32 cases of SARS-CoV-2 in the study participants.²⁵ On the contrary, mNGS only detected 18 of the 32 SARS-CoV-2 cases. This comparatively lower detectability rate is not surprising, considering that most of the SARS-CoV-2 cases detected by mRT-PCR had Ct values mostly ≥ 33 .²⁵ Such higher Ct values (correlative of low viral load) have been previously reported to impact viral detection by mNGS.^{19,29} Similar to the mRT-PCR result from the same participants, there was no detection of the seasonal influenza virus by mNGS. This could be attributed to the sampling/study period where zero influenza detection was reported in the Free State Province from week 50 (December) of 2020 to week 34 (August) of 2021 as per the National weekly respiratory pathogen surveillance report by the National Institute for Communicable Diseases, (NICD) South Africa (<https://www.nicd.ac.za/diseases-a-z-index/disease-index-covid-19/surveillance-reports/weekly-respiratory-pathogens-surveillance-report-week/>) (a summary of the weekly

influenza surveillance data per province for the study period is shown in Table S5). Contrary to the nondetection of influenza, nationally, an outbreak of RSV outside of its season followed by its sporadic detection⁴⁷ and waves of SARS-CoV-2 transmission (<https://www.nicd.ac.za/wp-content/uploads/2021/11/Proposed-definition-of-COVID-19-wave-in-South-Africa.pdf>) were documented during the study period. This observed phenomenon is, however, attributed to the nonpharmaceutical interventions implemented against COVID-19 and was also corroborated in a previous study.²⁵

Also, the nondetection of adenovirus by the mNGS approach may be due to the initial DNase step outlined in the methodology. Additionally, mNGS in this study identified several viral co-infection rates correlative with the mRT-PCR result. Viral co-infections detected by the mRT-PCR in the same participants ranged from double to quadruple,²⁵ while mNGS detected double to quintuple. However, the association of co-infection to disease severity or clinical outcome could not be established due to the unavailability of information on death history, if any, or exacerbated clinical course. Of interest, of the two patients with no viral disease-related pathogen detected by mNGS, one had a viral pathogen detected, and the other was equally missed by mRT-PCR. Hypothetically, the viral aetiology could be present at a low abundance below the sensitivity of the mNGS, could be a novel virus not detected in the bioinformatic analysis, a DNA virus not targeted in the workflow, or rather an aetiology other than of viral origin as seen by the detection of *Escherichia coli* in the patient (Z44) missed by both methods.

The incremental value of multiple rapid, user-friendly mNGS analysis tools has been previously reported.^{23,34} Currently, there are no gold standard mNGS analysis tools, and different methods can provide varying outcomes for the same FASTQ file due to differences in the databases used by each tool.^{24,48} In this study, a higher detection rate was noted for the TRVRP; it detected higher co-infection rate, detected more viral pathogen present at low abundance and reported reads for at least one viral pathogen in 98% of the patients compared to 71% and 48% by Genome Detective and One Codex analysis tools, respectively. The detection rate exhibited by the TRVRP may be due to the use of the GenBank database. However, unlike Genome detective and One Codex analysis tools, any untargeted viral pathogen will not be detected. Notably, the number of detected reads and genome coverage for each viral pathogen varied significantly between the three mNGS analysis tools, explaining in part the variations in detection limits of each tool.²³ Although study-specific cut-off points for defining positive detection were defined for the three mNGS analysis tools. Nonetheless, this study detected all pathogens well above the cut-off points. While the use of multiple mNGS analysis tools as described may bolster mNGS results, further improvements/benchmarking of these analysis tools is required and upgrading of the data library to overcome deficiencies of the individual methods or databases is equally important.

Finally, the case of a participant (patient A25) further highlights the complex microbial basis of SARI. This female patient, aged 4 years and 5 months old, was admitted with SARI (upper respiratory

infection of pneumonia and difficulty in breathing) and had a nasopharyngeal sample collected on 01/01/2021. The patient was HIV-negative and did not need intensive care or oxygen. The pathogen detection using mRT-PCR was negative for all the targeted pathogens, and only HRV was detected via mNGS using TRVRP at 0.44% genome coverage. However, reads for *Neisseria meningitidis* was detected via One Codex at a HA of 97.91%. The history of meningococcal vaccination is unknown in the child. However, meningococcal vaccination is not included in the South Africa vaccine schedule and is only available for infants and adolescents in the private sector. While latent colonization has been reported in children,⁴⁹ a detection at 97.91% abundance may suggest replication/active infection and could be the main factor for symptoms exhibited.

Despite the apparent advantages of mNGS in viral pathogen detection, there are still some gaps to bridge. For instance, the observed low sensitivity of mNGS for the detection of samples with low viral load, especially of SARS-CoV-2, as seen in this study impedes the unequivocal reliability of this method. As such, the requirement for improved sample preparation, for low biomass pathogens and optimized analysis tools, is still imperative.¹⁰ More so, the detection of viral reads may not translate to the presence of replicable viruses. Nonetheless, in some cases, viral infection may be limited to a specific area of a tissue or part of susceptible cells. As such, detection of a single or very few viral reads may be indicative of viral infection.⁵⁰ As seen in this study, there were many correlative viral detections by mRT-PCR and mNGS. However, in some cases, the viral reads from mNGS analysis were comparatively low, despite being a true positive. Consequently, it raises the question of what limit of detection is applicable/should be implemented to enable intra/inter-laboratory comparisons.^{7,23,50}

A notable limitation of this study is the strict enrichment for RNA viruses, which makes it impossible to detect adenovirus (as detected by the mRT-PCR) or account for other co-infecting DNA viruses using mNGS. Consequently, further studies may be required to evaluate the detection of DNA viruses using mNGS compared to mRT-PCR. Additional limitation is the absence of the inclusion of positive control which could have been used in the evaluation of false negatives.

5 | CONCLUSION

This study further reiterates that mNGS-based detection holds promise as a diagnostic tool in severe paediatric infectious viral respiratory diseases, enabling the unbiased detection and molecular typing of the viral respiratory pathogens with simultaneous elucidation of the respiratory virome. The study presented much-needed information on the clinical utility of mNGS in SARI in the study setting where limited data on usability currently exist. Although further improvement of mNGS workflow (wet and dry lab onuses) may be required, mNGS can be used to supplement PCR-based tests, especially in cases where traditional methods have yielded no result

or instances where symptoms are severe, and the detected pathogen is not truly reflective of the symptoms presented. Imperatively, future studies where mNGS fails to yield a positive result may benefit from increased sequencing depth that could even obviate the need for enrichment steps, target of both DNA and RNA virus and differential sampling to include lower respiratory samples. Furthermore, the power of mNGS in syndromic testing and the significant incremental value of multiple user-friendly mNGS analysis tools was strongly demonstrated. Moving on, further implementation of trials to benchmark mNGS analysis tools and ensure comparable results and upgrade of the different analysis tools are required. Excitingly, sequencing costs are beginning to decline, and as evidence of clinical utility accumulates, the use and experience with this approach will probably become more widespread over time, stimulating the continuous development and optimization of mNGS for clinical diagnostic use.

AUTHOR CONTRIBUTIONS

Project conceptualization, supervision, and administration: Martin M Nyaga, Sabiu Saheed, and Ayodeji E Ogunbayo. *Sample delivery:* Ayodeji E Ogunbayo and Milton T Mogotsi. *Data curation:* Ayodeji E Ogunbayo, and Hlengiwe Sondlane. *Laboratory investigation:* Ayodeji E Ogunbayo, Milton T Mogotsi, Hlengiwe Sondlane, and Martin M Nyaga. *Data analysis and presentation:* Ayodeji E Ogunbayo. *Original draft:* Ayodeji E Ogunbayo. *Writing—review and editing:* all co-authors. *Funding:* Martin M Nyaga.

ACKNOWLEDGMENTS

We would like to thank the following senior nurses: Palesa Kortman, Yolanda Porota, and Mahlapenyane Botsane for their assistance with the patient recruitment and sample collection. We would also like to thank the Qiagen company (Qiagen, Hilden Germany) for availing us with the QIAstat-Dx Respiratory SARS-CoV-2 panel.

This research was funded by the Bill and Melinda Gate Foundation, grant number BMGFOPP1180423_2017, the National Research Foundation (NRF), South Africa, grant number 120814, the Poliomyelitis Research Foundation (PRF), South Africa, grant number 19/16 and the South Africa Medical Research Council Self-Initiated Research (SAMRC-SIR) grants; all awarded to Martin M Nyaga.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data presented in this study are available in Tables S2–S4.

ETHICS STATEMENT

The study was conducted in accordance with the Declaration of Helsinki and approved by the Health Science Research Ethics Committee (HSREC) of the University of the Free State, South Africa (Approval no UFS-HSD2019/1129/2910 and approval date of 24 October 2019). The study was also approved by the Free State Department of Health.

REFERENCES

- Lamrani Hanchi A, Guennouni M, Rachidi M, et al. Epidemiology of respiratory pathogens in children with severe acute respiratory infection and impact of the multiplex PCR film array respiratory panel: a 2-year study. *Int J Microbiol.* 2021;2021(2276261):1-9. doi:10.1155/2021/2276261
- Abubakar II, Tillmann T, Banerjee A. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the global burden of disease study 2013. *The Lancet.* 2015;385(9963):117-171. doi:10.1016/S0140-6736(14)61682-2
- Zhao Y, Lu R, Shen J, Xie Z, Liu G, Tan W. Comparison of viral and epidemiological profiles of hospitalized children with severe acute respiratory infection in Beijing and Shanghai, China. *BMC Infect Dis.* 2019;19(1):729. doi:10.1186/s12879-019-4385-5
- Tsitsiklis A, Osborne CM, Kamm J, et al. Lower respiratory tract infections in children requiring mechanical ventilation: a multicentre prospective surveillance study incorporating airway metagenomics. *The Lancet Microbe.* 2022;3(4):e284-e293. doi:10.1016/S2666-5247(21)00304-9
- Tang J, Chen J, He T, et al. Diversity of upper respiratory tract infections and prevalence of *Streptococcus pneumoniae* colonization among patients with fever and flu-like symptoms. *BMC Infect Dis.* 2019;19(1):1-10. doi:10.1186/s12879-018-3662-z
- Shao J, Hassouna A, Wang Y, et al. Next-generation sequencing as an advanced supplementary tool for the diagnosis of pathogens in lower respiratory tract infections: an observational trial in Xi'an, China. *Biomed Reports.* 2022;16(2):14. doi:10.3892/br.2021.1497
- Watts GS, Hurwitz BL. Metagenomic next-generation sequencing in clinical microbiology. *Clin Microbiol Newsletter.* 2020;42(7):53-59. doi:10.1016/j.clinmicnews.2020.03.004
- Juliana AE, Tang MJ, Kemps L, et al. Viral causes of severe acute respiratory infection in hospitalized children and association with outcomes: a two-year prospective surveillance study in Suriname. *PLoS One.* 2021;16(2):e0247000. doi:10.1371/journal.pone.0247000
- Graf EH, Simmon KE, Tardif KD, et al. Unbiased detection of respiratory viruses by use of RNA sequencing-based metagenomics: a systematic comparison to a commercial PCR panel. *J Clin Microbiol.* 2016;54(4):1000-1007. doi:10.1128/JCM.03060-15
- Zheng Y, Qiu X, Wang T, Zhang J. The diagnostic value of metagenomic next-generation sequencing in lower respiratory tract infection. *Front Cell Infect Microbiol.* 2021;11:694756. doi:10.3389/fcimb.2021.694756
- Gu W, Miller S, Chiu CY. Clinical metagenomic next-generation sequencing for pathogen detection. *Annu Rev Pathol: Mech Dis.* 2019;14(14):319-338. doi:10.1146/annurev-pathmechdis-012418-012751.Clinical
- Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet.* 2019;20(6):341-355. doi:10.1038/s41576-019-0113-7
- Diao Z, Han D, Zhang R, Li J. Metagenomics next-generation sequencing tests take the stage in the diagnosis of lower respiratory tract infections. *J Adv Res.* 2022;38(1):201-212. doi:10.1016/j.jare.2021.09.012
- Qian YY, Wang HY, Zhou Y, et al. Improving pulmonary infection diagnosis with metagenomic next generation sequencing. *Front Cell Infect Microbiol.* 2021;10:567615. doi:10.3389/fcimb.2020.567615
- Wang J, Han Y, Feng J. Metagenomic next-generation sequencing for mixed pulmonary infection diagnosis. *BMC Pulm Med.* 2019;19:252. doi:10.1186/s12890-019-1022-4
- Deurenberg RH, Bathoorn E, Chlebowicz MA, et al. Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotech.* 2017;243(10):16-24. doi:10.1016/j.jbiotec.2016.12.022
- Li CX, Li W, Zhou J, et al. High resolution metagenomic characterization of complex infectomes in paediatric acute respiratory infection. *Sci Rep.* 2020;10(1):3963. doi:10.1038/s41598-020-60992-6
- Liang M, Fan Y, Zhang D, et al. Metagenomic next-generation sequencing for accurate diagnosis and management of lower respiratory tract infections. *Int J Infect Dis.* 2022;122:921-929. doi:10.1016/j.ijid.2022.07.060
- Thorburn F, Bennett S, Modha S, Murdoch D, Gunson R, Murcia PR. The use of next generation sequencing in the diagnosis and typing of respiratory infections. *J Clin Virol.* 2015;69:96-100. doi:10.1016/j.jcv.2015.06.082
- Gong YN, Yang SL, Chen GW, et al. A metagenomics study for the identification of respiratory viruses in mixed clinical specimens: an application of the iterative mapping approach. *Arch Virol.* 2017;162(7):2003-2012. doi:10.1007/s00705-017-3367-4
- Bartlow AW, Stromberg ZR, Gleasner CD, et al. Comparing variability in diagnosis of upper respiratory tract infections in patients using syndromic, next generation sequencing, and PCR-based methods. *PLOS Global Public Health.* 2022;2(7):e0000811. doi:10.1371/journal.pgph.0000811
- Wang X, Braid SS, Scotch M, Rawlinson WD. Detection of respiratory viruses directly from clinical samples using next-generation sequencing: A literature review of recent advances and potential for routine clinical use. *Rev Med Virol.* 2022;1(e2375):1-12. doi:10.1002/rmv.2375
- de Vries JJC, Brown JR, Fischer N, et al. Benchmark of thirteen bioinformatic pipelines for metagenomic virus diagnostics using datasets from clinical samples. *J Clin Virol.* 2021;141:104908. doi:10.1016/j.jcv.2021.104908
- de Vries JJC, Brown JR, Couto N, et al. Recommendations for the introduction of metagenomic next-generation sequencing in clinical virology, part II: bioinformatic analysis and reporting. *J Clin Virol.* 2021;138:104812. doi:10.1016/j.jcv.2021.104812
- Ogunbayo AE, Mogotsi MT, Sondlane H, Nkwadipo KR, Sabiu S, Nyaga MM. Pathogen profile of children hospitalised with severe acute respiratory infections during COVID-19 pandemic in the free state province, South Africa. *Int J Environ Res Public Health.* 2022;19:10418.
- Fitzner J, Qasmieh S, Mounts AW, et al. Revision of clinical case definitions: influenza-like illness and severe acute respiratory infection. *Bull World Health Organ.* 2018;96(2):122-128. doi:10.2471/BLT.17.194514
- Vilsker M, Moosa Y, Nooij S, et al. Genome detective: an automated system for virus identification from high-throughput sequencing data. *Bioinformatics.* 2019;35(5):871-873. doi:10.1093/bioinformatics/bty695
- Minot SS, Krumm N, Greenfield NB. One codex: a sensitive and accurate data platform for genomic microbial identification. *bioRxiv.* Published online 2015:027607. <http://biorxiv.org/content/early/2015/09/28/027607.abstract>
- Madi N, Al-nakib W, Mustafa AS, Habibi N. Metagenomic analysis of viral diversity in respiratory samples from patients with respiratory tract infections in Kuwait. *J Med Virol.* 2018;90(3):412-420. doi:10.1002/jmv.24984
- Romero-esposito JA, Moreno-valencia Y, Coronel-tellez RH, et al. Virome and bacteriome characterization of children with pneumonia and asthma in Mexico City during winter seasons 2014 and 2015. *PLoS One.* 2018;13(2):e0192878.
- Yang J, Yang F, Ren L, et al. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. *J Clin Microbiol.* 2011;49(10):3463-3469. doi:10.1128/JCM.00273-11
- Lewandowska DW, Schreiber PW, Schuurmans MM, et al. Metagenomic sequencing complements routine diagnostics in identifying

- viral pathogens in lung transplant recipients with unknown etiology of respiratory infection. *PLoS One*. 2017;12(5):e0177340.
33. Moore ML, Stokes KL, Hartert TV. The impact of viral genotype on pathogenesis and disease severity: respiratory syncytial virus and human rhinoviruses. *Curr Opin Immunol*. 2013;25(6):761-768. doi:10.1016/j.coi.2013.09.016
 34. Huang W, Wang G, Lin H, et al. Assessing next-generation sequencing and 4 bioinformatics tools for detection of Enterovirus D68 and other respiratory viruses in clinical samples. *Diagn Microbiol Infect Dis*. 2016;85(1):26-29. doi:10.1016/j.diagmicrobio.2016.01.013
 35. B'Krong NTT, Minh NNQ, Qui PT, et al. Enterovirus serotypes in patients with central nervous system and respiratory infections in Viet Nam 1997-2010. *Virol J*. 2018;15(1):69. doi:10.1186/s12985-018-0980-0
 36. Pabbaraju K, Wong S, Wong AA, Tellier R. Detection of enteroviruses and parechoviruses by a multiplex real-time RT-PCR assay. *Mol Cell Probes*. 2015;29(2):81-85.
 37. Zhou Y, Fernandez S, Yoon IK, et al. Metagenomics study of viral pathogens in undiagnosed respiratory specimens and identification of human enteroviruses at a Thailand hospital. *Am J Trop Med Hyg*. 2016;95(3):663-669. doi:10.4269/ajtmh.16-0062
 38. Reyes A, Carbo EC, Harinxma thoe Slooten JS, et al. Viral metagenomic sequencing in a cohort of international travellers returning with febrile illness. *J Clin Virol*. 2021;143:104940. doi:10.1016/j.jcv.2021.104940
 39. Tay CJX, Ta LDH, Ow Yeong YX, et al. Role of upper respiratory microbiota and virome in childhood rhinitis and wheeze: collegium internationale allergologicum update 2021. *Int Arch Allergy Immunol*. 2021;182(4):265-276. doi:10.1159/000513325
 40. Li Y, Fu X, Ma J, et al. Altered respiratory virome and serum cytokine profile associated with recurrent respiratory tract infections in children. *Nat Commun*. 2019;10(2288):2288. doi:10.1038/s41467-019-10294-x
 41. Kustin T, Ling G, Sharabi S, et al. A method to identify respiratory virus infections in clinical samples using next-generation sequencing. *Sci Rep*. 2019;9(1):2606. doi:10.1038/s41598-018-37483-w
 42. Li Y, Sun B, Tang X, et al. Application of metagenomic next-generation sequencing for bronchoalveolar lavage diagnostics in critically ill patients. *Eur J Clin Microbiol Infect Dis*. 2020;39(2):369-374. doi:10.1007/s10096-019-03734-5
 43. Zhang P, Chen Y, Li S, et al. Metagenomic next-generation sequencing for the clinical diagnosis and prognosis of acute respiratory distress syndrome caused by severe pneumonia: a retrospective study. *PeerJ*. 2020;8:e9623. doi:10.7717/peerj.9623
 44. Huang J, Jiang E, Yang D, et al. Metagenomic next-generation sequencing versus traditional pathogen detection in the diagnosis of peripheral pulmonary infectious lesions. *Infect Drug Resist*. 2020;13:567-576. doi:10.2147/IDR.S235182
 45. O'Brien KL, Baggett HC, Brooks WA, et al. Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. *The Lancet*. 2019;394(10200):757-779. doi:10.1016/S0140-6736(19)30721-4
 46. Hatem A, Mohamed S, Abu Elhassan UE, et al. Clinical characteristics and outcomes of patients with severe acute respiratory infections (SARI): results from the Egyptian surveillance study 2010-2014. *Multidiscip Respir Med*. 2019;14(11):11. doi:10.1186/s40248-019-0174-7
 47. Bents S, Viboud C, Grenfell B, et al. The impact of COVID-19 non-pharmaceutical interventions on future respiratory syncytial virus transmission in South Africa. *medRxiv Prepr Serv Heal Sci*. Published online 2022. doi:10.1101/2022.03.12.22271872
 48. Carbo EC, Sidorov IA, van Rijn-Klink AL, et al. Performance of five metagenomic classifiers for virus pathogen detection using respiratory samples from a clinical cohort. *Pathogens*. 2022;11(3):340. doi:10.3390/pathogens11030340
 49. Maleki A, Mirnaseri Z, Kouhsari E, et al. Asymptomatic carriers of *Neisseria meningitidis* and *Moraxella catarrhalis* in healthy children. *New Microbes New Infect*. 2020;36:100691. doi:10.1016/j.nmni.2020.100691
 50. Cantalupo PG, Pipas JM. Detecting viral sequences in NGS data. *Current Opinion in Virology*. 2019;39:41-48. doi:10.1016/j.coviro.2019.07.010

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ogunbayo AE, Mogotsi MT, Sondlane H, Sabiu S, Nyaga MM. Metagenomics characterization of respiratory viral RNA pathogens in children under five years with severe acute respiratory infection in the Free State, South Africa. *J Med Virol*. 2023;95:e28753. doi:10.1002/jmv.28753