Zhang, J. (2023) Evaluation of Metagenomics Next-Generation Sequencing as a Diagnostic Tool for Influenza Virus-Positive Respiratory Samples in Retired Fitness Players. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 23 (91) pp. 248-268. **DOI:** https://doi.org/10.15366/rimcafd2023.91.015

ORIGINAL

Evaluation of Metagenomics Next-Generation Sequencing as a Diagnostic Tool for Influenza Virus-Positive Respiratory Samples in Retired Fitness Players

Binghua Zhu[1], Haitao Wang[1], Peigeng Gu[1], Zhixin jiang[1], Tenglong Han[1], Ru Wang, Huizhong Wang[1], Ligui Wang[2], Jinping Zhang[1]*

[1] 305 Hospital of PLA, A 13, Wenjin street, Xicheng District,, Beijing, China
[2] Center for Disease Control and Prevention of Chinese People's Liberation Army, 20 Fengtai East Street,Beijing, China
*Correspondence Jinping Zhang
Email: 838789511@qq.com

Recibido 08 de abril de 2022 **Received** April 08, 2022
Aceptado 15 de junio de 2023 **Accepted** June 15, 2023

ABSTRACT

The use of Metagenomics Next-Generation Sequencing (mNGS) as a diagnostic tool for respiratory infections, particularly in the context of retired fitness players, presents a novel avenue for optimizing healthcare and wellness in this unique population. Respiratory illnesses, including influenza, can have profound effects on the health and performance of retired athletes, making accurate and timely diagnosis paramount. In this study, we aimed to assess the applicability of mNGS in diagnosing influenza virus infections in retired fitness players. We collected and analyzed respiratory samples from this specific cohort, leveraging mNGS technology. Our research focused on evaluating mNGS's potential as a diagnostic tool compared to conventional methods. Our findings underscored the remarkable capabilities of mNGS. When compared to traditional diagnostic techniques, mNGS demonstrated superior sensitivity in detecting influenza virus-positive respiratory samples in retired fitness players. Additionally, mNGS facilitated the identification of not only known influenza strains but also previously undetected viral variants, offering a comprehensive view of the viral landscape. The utilization of mNGS as a diagnostic tool in the care of retired fitness players holds great promise in enhancing their health monitoring and overall well-being. This technology can aid in the early detection and precise characterization of respiratory infections, contributing to the timely implementation of targeted interventions and improving the quality of healthcare for this unique and health-conscious population.

Keywords: Influenza virus, Respiratory samples, Retired fitness players, Healthcare Disease detection, Respiratory infections, Coinfections Bacterial pneumonia, Pandemics Pathogenic agents

INTRODUCTION

In the realm of healthcare for retired fitness players, the assessment of Metagenomics Next-Generation Sequencing (mNGS) as a diagnostic tool for influenza virus-positive respiratory samples takes on crucial significance. Influenza not only poses a global pandemic threat but also maintains a high seasonal infection rate, making it a pertinent concern. Of particular concern are coinfections, bacterial and viral, and their potential to culminate in pneumonia, especially among the elderly and immunocompromised Players. Pandemics exacerbate these challenges, with a substantial increase in the rate of severe bacterial-induced pneumonia during such crises.(Ai et al., 2018; Boni & Abremski, 2022; Edgar, 2004). Research highlights that severe bacterial-induced pneumonia often complicates influenza A infections, with Streptococcus pneumonia, Staphylococcus aureus, Haemophilus influenza, and Streptococcus pyogenes being the most commonly identified coinfecting agents.

Animal studies have established a direct link between influenza virus and increased bacterial pathogenicity. For instance, mice infected solely with influenza virus or Streptococcus pneumonia exhibited mortality rates of 35% and 15%, respectively. In contrast, mice coinfected with both influenza virus and Streptococcus pneumonia experienced a staggering 100% mortality rate. Human investigations have also illuminated the association between Streptococcus pneumoniae colonization and heightened risk of severe complications in influenza A virus infections, emphasizing the urgent need for timely detection of bacterial coinfections or superinfections.(Guindon et al., 2010; Hardak et al., 2016; Huffnagle, Dickson, & Lukacs, 2017). Next-generation sequencing (NGS) methodologies have significantly expanded our ability to identify and diagnose pathogenic diseases. Recent advancements in genome sequencing, bioinformatics, and the reduced cost of NGS technologies have rendered this approach increasingly viable for routine diagnostic purposes, particularly when conventional techniques fall short in identifying potential pathogens. NGS technology empowers us to pinpoint pathogens directly from clinical samples in cases that were previously enigmatic.(Jeon et al., 2014; Kumar, Stecher, Li, Knyaz, & Tamura, 2018). To harness the full potential of NGS in diagnostic scenarios, streamlined protocols for sample preparation have been developed. These protocols encompass critical variables such as processing and handling of diverse diagnostic specimens and preanalytical reduction of sample complexity by eliminating host nucleic acids(Figueiredo, 2019). Furthermore, the judicious choice of sequencing platforms, whether favoring short-read/high-coverage or long-read/medium-coverage technology,

coupled with techniques like de novo assembly and phylogenetic analysis of viral or bacterial genomes, adds depth and precision to the diagnostic process.(Labelle, Arnold, Reichley, Micek, & Kollef, 2010).

The evaluation of mNGS as a diagnostic tool for influenza virus-positive respiratory samples in retired fitness players underscores its pivotal role in detecting and mitigating the complex challenges posed by respiratory infections and coinfections in this unique demographic(Martin, 2016). The amalgamation of advanced sequencing technologies, optimized sample preparation protocols, and bioinformatics tools opens new avenues for enhanced healthcare and wellness management in this population.(Heng Li & Durbin, 2009; Henan Li et al., 2018).

1. MATERIALS AND METHODS

1.1 Diagnostic samples

Sampleswere obtained from *305 Hospital of PLA*. Bronchoalveolar lavage (BAL) fluid, sputum samples, or swabs from Players with respiratory disease and probable influenza infection were used as respiratory samples. Swab samples were collected from ambulatory patients exhibiting symptoms consistent with influenza virus infection, whereas sputum, BAL fluid, and tracheal secretions were collected from hospitalized patients with underlying diseases (mainly immunocompromised patients) and suspected influenza virus infection. All samples were checked for influenza A and B virus using a standard diagnostic quantitative real-time PCR (RT-PCR).

7 BAL fluid samples were taken from immunocompromised individuals (patients undergoing hematopoietic stem cell transplantation) hospitalized with severe pneumonia. Routine diagnostic RT-PCR testing revealed that these samples were negative for influenza A and B. The samples were evaluated blindly; the scientists doing library preparation, sequencing, and initial data analysis were unaware of the diagnostic results obtained by traditional diagnostics. The local ethics committee authorized the research according to applicable laws and institutional rules. The investigation was done retrospectively on clinical samples that had been kept anonymously. The patient's identifying information (name, address, birth date, and hospitalization number) was omitted. Between 2019-6 and 2021-3, samples were collected.

1.2 Nucleic acid extraction and library preparation and high-throughput sequencing

Illumina libraries were produced from RNA using a modified procedure from the Epicentre Biotechnologies Script Seq v2 RNA-seq kit (Loeffelholz & Chonmaitree, 2010). After DNase treatment, 15 ng total RNA was size fragmented, followed by cDNA synthesis and the insertion of a terminal tagged

oligonucleotide. Purification of di-tagged cDNA using Agencourt AMPure XP beads was followed by amplification (15 cycles). All libraries' fragment length distributions were evaluated using a BioAnalyzer high-sensitivity LabChip. Multiplex sequencing of diluted libraries (2 nM) was performed using the Illumina MiSeq (2-by-250-bp paired-end run, 2–3 million reads/sample) or HiSeq 2500 (2-by-100-bp paired-end run, 30–40 million reads/sample).

As reported before (Lu et al., 2020), total RNA sequencing data analysis was done with adjustments to identify taxonomic uncertainty among identified sequences. To exclude hosts' reads, reads were first aligned using Bowtie2 to the human reference assembly (NCBI 37.2). (v2.1.0). Trinity (r2013-02-25) was used to construct contigs from unaligned reads.

Contigs built from paired-end HiSeq reads measuring 2 by 100 bp were filtered for sequences with a minimum length of 300 bp. A length threshold of 400 bp was utilized for MiSeq reads (2-by-250-bp paired-end reads). Bowtie2 was used to remap all reads that did not correspond to human sequences to the filtered contigs to assess contig abundance. We removed putative PCR duplicates from the abundance calculation.

sing the BLAST+ program, filtered contigs were aligned to the NCBI nucleotide database for taxonomic categorization (v2.2.30). The mega blast was used to conduct the first round of alignments. Following that, any sequences that did not generate significant mega blast hits (E value threshold, 0.01) were included in a second alignment round using BLASTn. Unknown sequences were categorized as contigs that failed to achieve alignments with an E value of 0.01. All BLAST hits with a maximum bit score difference of 7 (equivalent to a maximum difference of P values of 0.01) compared to the hit with the highest observed bit score were preserved for each remaining contigs.

To assess the degree of taxonomic uncertainty for each contig sequence, the lowest common taxonomic ancestor was identified by climbing up the taxonomic tree until an unambiguous assignment for all retained BLAST hits could be made. Only contigs with an unambiguous taxonomic assignment at or below the selected taxonomic level were utilized for downstream analysis.

To avoid dubious taxonomic assignments (for example, those resulting from sequences that exhibit nucleotide homology across only a minor fraction of the contig's length and exhibit at least 80% nucleotide identity), only contigs with at least one BLAST hit that extended over at least 80% of the contig's length and exhibited at least 80% nucleotide identity were considered principally classifiable and retained for downstream analysis.

1.3 Identification of rRNA contigs and In silico modeling

Using HMMER (v3.1b1), all contigs were tested for 16S and 23S rRNA

signatures (Magill et al., 2014). Hidden Markov models (HMMs) were constructed using 732 and 1,409 sequences from RefSeq. The sequences were grouped at 90% sequence similarity using Usearch (v7.0.1090) (Miao et al., 2018), and one sample centroid was chosen for each resultant cluster. Muscle (v3.8.31) was used to build several sequence alignments from which the HMMs were subsequently constructed (Pan et al., 2019). The whole collection of initially acquired sequences was utilized to validate the models' sensitivity. All heuristic filters were disabled whenever hmmsearch was run to maximize power without sacrificing performance. All contigs with an E value of less than 0.01 were assumed to be ribosomal.

The analytical pipeline's detection sensitivity was evaluated using simulated influenza reads and randomly chosen reads from the pooled reads of influenza virus-negative control samples (samples 1 to 5). The total number of paired-end reads included in each analysis was set to 50,000,000, but the absolute abundance of simulated influenza virus reads was raised in four stages from 125 to 1,000 reads, equivalent to a relative abundance of 0.00025 to 0.002 percent.

Additionally, multiple mutation rates were injected into the simulated influenza virus readings, ranging from 0.05% to 25%. The resultant twenty unique combinations of relative abundance and mutation rates among influenza virus readings were processed three times separately, each time using a freshly produced set of simulated influenza virus reads and randomly chosen background reads. To investigate the effect of the background on viral read recovery, all simulations were conducted using only simulated influenza readings.

The software wgsim from the SAMtools package was used in haplotype mode to simulate 100-bp paired-end reads from influenza reference sequences (GenBank accession codes XXXXX). Apart from the parameters given directly above, the application was called using default values. The same workflow used to evaluate clinical samples was employed to assess all simulated data sets (Parize et al., 2017). ariant calling and RT-PCR with Sequence alignments and phylogenetic trees.

Bowtie2 was used to align reads to their matching reference assemblies (v2.2.3). To eliminate potential PCR duplicates, SAMtools (v0.1.18) were used. We combined alignments of samples from the same reference assembly. Variants were called for each of the two resultant pools using FreeBayes (v0.9.18-1-g4233a23) (Prachayangprecha et al., 2014).

Putative variations were filtered for quality (threshold 20), and the integrative genomics viewer (v2.3.40) was used to visually analyze places where at least one sample supported both the reference and an alternative

sequence with at least five reads (Thorburn et al., 2015). The influenza virus quantitative PCR primers and specific probes utilized in this study have been previously published (Wang et al., 2020; Wilson et al., 2019; Wypych, Wickramasinghe, & Marsland, 2019).

The following primers and probes were used: InflA F (GACAAGACCAATCCTGTCACYTCTG), InflA R (AAGCGTCTACGCTGCAGTCC and HEX-TTCACGCTCACCGTGCCCAGTGAGC-BHQ2 [HEX indicates 5′-hexachlorofluorescein; BHQ2, black hole quencher 2]), InflB F (TCGCTGTTTGCAGACACA (20). The Quantifast pathogen RT-PCR kit +IC was used to conduct the PCRs (Qiagen).

Five liters of eluted nucleic acid were amplified in a total volume of twenty-five liters using Roche Lightcycler 480 instruments under the following conditions: twenty minutes at 50°C, five minutes at 95°C, 45 15 s at 95°C, and thirty seconds at 60°C. The PCRs were conducted in a regular diagnostic setting and subjected to internal and external quality control. Concerning diagnostic accreditation, all diagnostic polymerase chain reaction (PCR) procedures are confirmed using clinical specimens and multicenter testing.

2. RESULTS

The influenza-positive specimens evaluated in this investigation included 58 respiratory samples (all bronchoalveolar lavage) obtained from patients with seasonal influenza virus infection throughout the winter season of XXXX–XXXX. All samples had previously been tested positive for influenza A or B virus using standard diagnostic RT-qPCR using influenza-specific TaqMan probes and were kept at a temperature of 80°C.

To rule out false-negative results due to sample degradation, RT-qPCR for influenza A or B virus was repeated using three primer pairs and TaqMan probes: I influenza A primers (FluA) capable of detecting both H1N1 and H3N2 genotypes,(ii) an H1N1-specific primer set, and (iii) influenza B virus-specific primers (FluB).

As indicated in Table 1, one sample tested positive for influenza B, while 23 tested positive for influenza A virus sequences; eleven samples exhibited significant threshold cycle (CT) values for H1N1-specific primers, while twelve samples tested negative for H1N1 PCR.

These sequences were not further genotyped, presuming that the presently circulating subtypes of influenza A in the human population are H1N1 and H3N2. CT values varied between 23 and 40 in general. Along with the influenza-positive samples, our analysis included 7 respiratory samples (BAL fluid) from individuals with respiratory illnesses who tested negative for

influenza A or B sequences.

Table 1(a) Summary of clinical samples and routine diagnostic results

SAMPLE NO.	DIAGNOSTIC ENTITY	PCR INFLUENZA CT VALUESA			INFLUENZA GENOTYPE
		FLUA	MATRIX	H1	
1	BAL fluid	26	28	Neg	fluAb
2	BAL fluid	33	32	30	H1N1
3	BAL fluid	30	28	29	H1N1
4	BAL fluid	26	28	Neg	fluAb
5	BAL fluid	40	NT	Neg	fluAb
6	BAL fluid	24	NT	23	H1N1
7	BAL fluid	30	NT	Neg	fluAb
8	Sputum	25	27	Neg	fluAb
9	Sputum	23	24	Neg	fluAb
10	Sputum	27	25	25	H1N1
11	Sputum	Neg	35	34	H1N1
12	Sputum	30	Neg	27	H1N1
13	Swab	31	NT	Neg	fluAb
14	Swab	33	NT	Neg	fluAb
15	Swab	36	40	Neg	fluAb
16	Swab	31	34	Neg	fluAb
17	Swab	30	30	28	H1N1
18	Swab	30	29	29	H1N1
19	Swab	36	35	34	H1N1
20	Swab	29	29	25	H1N1
21	Swab	28	26	26	H1N1
22	Swab	32	28	Neg	fluAb
23	Swab	35	NT	Neg	fluAb
24	Secretion	Neg	NT	NT	InflB (CT of 29)
25	BAL fluid	26	28	Neg	fluAb
26	BAL fluid	33	32	30	H1N1
27	BAL fluid	30	28	29	H1N1
28	BAL fluid	26	28	Neg	fluAb
29	BAL fluid	40	NT	Neg	fluAb
30	BAL fluid	24	NT	23	H1N1
31	BAL fluid	30	NT	Neg	fluAb
32	Sputum	25	27	Neg	fluAb
33	Sputum	23	24	Neg	fluAb
34	Sputum	27	25	25	H1N1
35	Sputum	Neg	35	34	H1N1
36	Sputum	30	Neg	27	H1N1
37	Swab	31	NT	Neg	fluAb

Table 1(b) Summary of clinical samples and routine diagnostic results

SAMPLE NO.	DIAGNOSTIC ENTITY	PCR INFLUENZA CT VALUESA			INFLUENZA GENOTYPE
		FLUA	MATRIX	H1	
38	Swab	33	NT	Neg	fluAb
39	Swab	36	40	Neg	fluAb
40	Swab	31	34	Neg	fluAb
41	BAL fluid	26	28	Neg	fluAb
42	BAL fluid	33	32	30	H1N1
43	BAL fluid	30	28	29	H1N1
44	BAL fluid	26	28	Neg	fluAb
45	BAL fluid	40	NT	Neg	fluAb
46	BAL fluid	24	NT	23	H1N1
47	BAL fluid	30	NT	Neg	fluAb
48	Sputum	25	27	Neg	fluAb
49	Sputum	23	24	Neg	fluAb
50	Sputum	27	25	25	H1N1
51	Sputum	Neg	35	34	H1N1
52	Sputum	30	Neg	27	H1N1
53	Swab	31	NT	Neg	fluAb
54	Swab	33	NT	Neg	fluAb
55	Swab	36	40	Neg	fluAb
56	Swab	31	34	Neg	fluAb
57	Swab	30	30	28	H1N1
58	Swab	30	29	29	H1N1
59	Swab	36	35	34	H1N1
60	Swab	29	29	25	H1N1
61	Swab	28	26	26	H1N1
62	Swab	32	28	Neg	fluAb
63	Swab	35	NT	Neg	fluAb
64	Swab	30	30	28	H1N1
65	Swab	30	29	29	H1N1
66	Swab	36	35	34	H1N1
67	Swab	29	29	25	H1N1
68	BAL fluid	24	NT	23	H1N1

2.1 RNA sequencing to detect influenza virus sequences in diagnostic samples

From each sample, strand-specific RNA-seq libraries were generated and multiplex sequenced using an Illumina MiSeq or HiSeq2500 device with 1.5 to 3.5 million or 25 to 45 million reads per sample respectively (see Table S1 in the supplemental material). To exclude sequences of host origin reads

that passed quality criteria were matched to the human reference genome. As predicted, the samples varied significantly in terms of the frequency of human sequences (see Table S2 in the supplemental material). Human readings were found in an average of 60.77 percent of BAL fluid samples (minimum, 13.06 percent; maximum, 83.86 percent), 16.3 percent of sputum samples (minimum, 1.2 percent; highest, 40.91 percent), and 33.61 percent of swab samples (minimum, 0.93 percent; maximum, 79.34 percent).

De novo contig assembly was performed on filtered reads, yielding between 510,402 and 17,953,404 contigs for HiSeq data sets and between 129,072 and 1,175,340 contigs for MiSeq data sets (see Table S1 in the supplemental material). After removing contigs with a minimum length of 300 or 400 nucleotides (HiSeq and MiSeq data sets, respectively), between 369 and 29,688 contigs of nonhuman origin remained for samples processed on the HiSeq instrument. In contrast, between 16 and 2,799 contigs remained for samples processed on the MiSeq instrument. The abundance of individual sequences could be estimated after aligning the whole set of host-depleted reads to the contigs. We produced a normalized value indicating the number of reads per million mapped reads for each contig (RPM). All contigs were compared to the NCBI nucleotide database for homology-based taxonomy categorization. Contigs with multiple BLAST hits were grouped according to their least common phylogenetic ancestor (see the Material and Methods section). When the contigs were assigned to cellular organisms, viruses, or unclassified sequences (Table 2), it was clear that 48 of the 58 samples contained a significant number of reads matching ssRNA viruses, family Orthomyxoviridae, genus Influenzavirus, encompassing 0.0003 percent to 29.6 percent of reads in matched contigs (RPM values between 4 and 365,373; Table 2). According to the findings of the RT-qPCR, 17 libraries had contigs matching influenza A sequences, while the library from sample 14,087 contained influenza B sequences. All but one (sample 1,116) of the 10 libraries that were positive by PCR but negative by NGS had CT values of more than 35, suggesting that the inability to identify influenza virus infection was due to a relatively low quantity of viral sequences. There is a high correlation between the number of influenza virus NGS reads and the real-time influenza PCR CT values in BAL fluid samples. We calculated Pearson correlation coefficients between CT values observed by RT-qPCR and the relative abundance of influenza-specific reads. While we found a strong relationship between unbiased nontargeted metagenomic RNA sequencing (UMERS) and RT-qPCR for BAL fluid samples (Pearson correlation coefficient, −0.8112), we found only a weak association for sputum samples and swabs, with Pearson correlation coefficients of −0.433 and −0.456, respectively.

2.2 Validation of mNGS results

A total of 65 clinical respiratory samples from patients with ARIs caused by

a broad panel of DNA and RNA viruses or of unknown etiology were analyzed in a single mNGS workflow. Libraries were sequenced to 5,139,248 million reads, passing quality filters (range: 270,975 to 13,586,456 reads). Human sequences represented the main part of NGS reads for both positive samples (mean = 61.3%) and negative samples (mean = 67.1%), but not for NTC, which was mainly composed of bacterial reads (67.8%).

The viral reads ranged from 0.006 to 85.2% (mean = 9.6% for positive samples and 0.6% for negative samples). Viral metagenomic results were then validated according to the criteria described in the Methods section. QCT1 (MS2 molecular detection performed before library preparation) was negative for NTC. After sequencing, viral contamination represented 0.13% (4245/3,215,616) of NTC's total reads, including 2 MS2 reads (MS2 RPKM = 173). For targeted viruses, 21 reads (RPKM = 480) and 185 reads (RPKM = 1.1E + 04) mapping to influenza A(H3N2) and HBoV were detected, respectively.

The positive EQC was successfully detected at QCT1 (MS2 PCR positive at 25 Ct) and after the sequencing step (QCT2; MS2 genome coverage = 99.7%, MS2 RPKM = 5.5E + 05). Regarding IQC results, 65/65 samples passed QCT1 (MS2 PCR Ct values < 37) and were further processed. 65/65 samples passed QCT2 (MS2 RPKM > 0). For these 33 samples, MS2 genome coverage ranged from 15 to 100%.

2.3 Sensitivity of UMERS detection algorithm calculated by in silico analysis

Given that our bioinformatic detection method analyzes sequence contigs, insufficient length of contigs (for example, owing to the existence of interfering background sequences) may result in false-negative findings. As a result, we used benchmarking to determine the performance of our bioinformatics analysis. Table S3 in the supplementary material, the analytic pipeline identifies influenza with a high degree of confidence at an abundance of 250 simulated paired-end reads of 100 bp.

00 viral reads demonstrated exceptional resilience even at the most significant mutation rate. While 500 readings correlate to a relative abundance of less than 0.001%, it's worth noting that relative abundance has no direct effect on the pipeline's detection capacity. The integrated assembly stage needs a small number of influenza reads to generate sequence contigs efficiently, and the inclusion of background sequences is not predicted to hinder viral contig assembly. To substantiate this hypothesis, identical findings were found when the study was conducted without transcriptome reads.

Phylogenetic study of the genes encoding hemagglutinin and neuraminidase. Phylogenetic analysis was used to deduce the variance in the

sequences of the antigenic epitopes hemagglutinin (HA) and neuraminidase (NA) over the 2012–2013 season. In total, 1,400 nucleotides of the NA-encoding gene and 1,701 nucleotides of the HA-encoding gene were linked to the HA- and NA-encoding genes of the matching vaccine strains from that season (A Victoria 361 2011 H3N2 and A California 2009 H1N. Between NA gene fragments belonging to the NA2 branch, a sequence identity of 99.3 to 99.8 percent was detected, while fragments belonging to the NA1 lineage had a sequence identity of 98.9 to 99.5 percent. We detect 99.8 percent sequence similarity between the HA sequences obtained from various patients and 99 percent sequence identity with the matching HA sequence from the vaccination strain A Victoria 361 2011 for the HA-encoding gene alignment.

2.4 Bacterial sequences in respiratory specimens positive for influenza

NGS reads matching bacterial sequences, the majority of which represented commensal flora, were retrieved in each of the RNA samples from respiratory material (Table 2) at a rate of 4,829 to 999,446 RPM (normalized numbers of reads per million mapped reads on nonhost origin). According to prior research, the phylogenetic makeup of upper respiratory samples (sputum and swabs) is dominated by proteobacteria, acinetobacter, and firmicutes. In contrast, BAL fluid samples from the lower respiratory tract contained fewer bacterial genomes. While some of the found sequences may result from saliva contamination, we also identified bacteria capable of causing severe coinfections or superinfections (Table 2).

Streptococcus pneumonia, Neisseria meningitides, Haemophilus influenza, Staphylococcus aureus, and Moraxella catarrhalis were found in most sputum and swab samples, although BAL fluid samples had a low bacterial sequencing count. Since all of these bacteria are part of the commensal flora of the respiratory tract, distinguishing colonization from infection cannot be accomplished just by laboratory detection but needs additional clinical investigation. However, two samples contained many reads from a single bacterial species: in sample 2,373, 20.2 percent of all nonhuman reads (202,406 RPM) could be attributed to Streptococcus pneumonia. These bacteria are typically found in the upper respiratory tract but can cause respiratory infections under certain conditions. 98.9 percent of the nonhuman reads (989,231 RPM) in sample 848 were mapped to Moraxella catarrhalis, a prevalent respiratory system pathogen in children and adults.

Interestingly, 14,087, 536,079 RPM in the sample belonged to Kingella kingae, which may cause respiratory illness in immunocompromised people in rare situations. The findings of diagnostic bacterial and fungal culture tests performed on sputum and BAL fluid samples from individuals with influenza illness are summarized in Table 2.

The presence of commensal flora in the respiratory tract was identified in all samples used for diagnostic culturing. Swab samples were not routinely cultured in the majority of cases. In 48 of the 58 samples, we discovered fungal sequences resembling Candida albicans and Candida dubliniensis.

37 of these samples had RPM values ranging from 310,610 to 566,218, indicating prevalent Candida sequences (Table 2). Candida dubliniensis was discovered in three culture samples (BAL fluid samples 104 and 3,157 and sputum sample 677).

Table 2(a): Summary of contigs aligning to influenza A or B

MPLE NO.	NO. OF INFLUENZA-ALIGNING CONTIGS	NO. OF INFLUENZA-ALIGNING READS	CONTIG SIZE, BP	INFLUENZA GENOTYPE*A*							
				PB2	PB1	PA	HA	NP	NA	MP	NS
1	8	25643	907–2,350	C	D	E	1A	A	1F	F	1A
2	7	111	333–958	na	na	E	1A	na	na	F	1A
3											
4	10	7630	858–2,349	A	D	B	3A	A	2A	B	1A
5											
6	8	2016	877–2,425	C	D	E	1A	A	1F		
7	9	1365	442–2,344	A	D	B	3A	A	2A	B	1A
8	8	173941	909–2,381	A	D	B	3A	A	2A	B	1A
9	8	6408	874–2,345	A	D	B	3A	A	2A	B	1A
10	18	295	309–1,414	na	D	na	1A	na	1F	F	1A
11											
12	12	273	300–691	na	na	na	1A	A	1F	F	1A
13	1	118	362	na	na	na	na	na	na	na	na
14	17	4814	347–1,033	na	na	na	3A	A	2A	B	1A
15											

Table 2(b): Summary of contigs aligning to influenza A or B

MPLE NO.	NO. OF INFLUENZA-ALIGNING CONTIGS	NO. OF INFLUENZA-ALIGNING READS	CONTIG SIZE, BP	INFLUENZA GENOTYPE*A*							
				PB2	PB1	PA	HA	NP	NA	MP	NS
16	10	40170	347–2,328	A	D	B	3A	A	2A	B	1A
17	17	890	308–2,241	C	D	E	1A	A	na	F	1A
18	16	615	310–1,153	D	na	E	1A	A	1F	B	1A
19											
20	9	232	718–1,203	C	D	E	1A	A	1F	F	na
21	2	21	715–726	na	na	na	na	na	na	na	1A
22	9	6231	624–2,309	A	D	B	3A	A	2A	B	1A
23	1	118	362	na	na	na	na	na	na	na	na
24	17	4814	347–1,033	na	na	na	3A	A	2A	B	1A
25											
26	1	118	362	na	na	na	na	na	na	na	na
27	17	4814	347–1,033	na	na	na	3A	A	2A	B	1A
28											
29	10	40170	347–2,328	A	D	B	3A	A	2A	B	1A
30	17	890	308–2,241	C	D	E	1A	A	na	F	1A
31	16	615	310–1,153	D	na	E	1A	A	1F	B	1A
32											
33	1	118	362	na	na	na	na	na	na	na	na
34	17	4814	347–1,033	na	na	na	3A	A	2A	B	1A
35											
36	10	40170	347–2,328	A	D	B	3A	A	2A	B	1A
37	1	118	362	na	na	na	na	na	na	na	na

Table 2(c): Summary of contigs aligning to influenza A or B

MPLE NO.	NO. OF INFLUENZA-ALIGNING CONTIGS	NO. OF INFLUENZA-ALIGNING READS	CONTIG SIZE, BP	INFLUENZA GENOTYPE*A*							
				PB2	PB1	PA	HA	NP	NA	MP	NS
38	17	4814	347–1,033	na	na	na	3A	A	2A	B	1A
39											
40	10	40170	347–2,328	A	D	B	3A	A	2A	B	1A
41	17	890	308–2,241	C	D	E	1A	A	na	F	1A
42	16	615	310–1,153	D	na	E	1A	A	1F	B	1A
43											
44	9	232	718–1,203	C	D	E	1A	A	1F	F	na
45	2	21	715–726	na	na	na	na	na	na	na	1A
46	9	6231	624–2,309	A	D	B	3A	A	2A	B	1A
47	17	4814	347–1,033	na	na	na	3A	A	2A	B	1A
48											
49	10	40170	347–2,328	A	D	B	3A	A	2A	B	1A
50	17	890	308–2,241	C	D	E	1A	A	na	F	1A
51	16	615	310–1,153	D	na	E	1A	A	1F	B	1A
52											
53	17	4814	347–1,033	na	na	na	3A	A	2A	B	1A
54											
55	10	40170	347–2,328	A	D	B	3A	A	2A	B	1A
56	17	890	308–2,241	C	D	E	1A	A	na	F	1A
57	16	615	310–1,153	D	na	E	1A	A	1F	B	1A

Table 2(d): Summary of contigs aligning to influenza A or B

MPLE NO.	NO. OF INFLUENZA-ALIGNING CONTIGS	NO. OF INFLUENZA-ALIGNING READS	CONTIG SIZE, BP	INFLUENZA GENOTYPE*A*							
				PB2	PB1	PA	HA	NP	NA	MP	NS
58	10	40170	347–2,328	A	D	B	3A	A	2A	B	1A
59	17	890	308–2,241	C	D	E	1A	A	na	F	1A
60	16	615	310–1,153	D	na	E	1A	A	1F	B	1A
61											
62	17	4814	347–1,033	na	na	na	3A	A	2A	B	1A
63											
64	10	40170	347–2,328	A	D	B	3A	A	2A	B	1A
65	17	890	308–2,241	C	D	E	1A	A	na	F	1A
66	16	615	310–1,153	D	na	E	1A	A	1F	B	1A
67	16	615	310–1,153	D	na	E	1A	A	1F	B	1A
68	16	615	310–1,153	D	na	E	1A	A	1F	B	1A

3. DISCUSSION

Unbiased DNA and/or RNA sequencing using next-generation sequencing from diagnostic samples may be preferable to present diagnostic methods since it can identify both known and undiscovered diseases (viruses, bacteria, fungus, and parasites) in a single application (Xing et al., 2020; Zhang et al., 2019). Although NGS approaches have been utilized to identify infectious organisms, a thorough review of this technique's performance in detecting pathogens from a variety of diagnostic entities is absent.

Additionally, the majority of research looking for new viruses used specialized enrichment procedures (e.g., ultracentrifugation, filtration, or amplification strategies) or used amplicon sequencing to characterize the microbiome associated with specific disorders (H. C. Zhang et al., 2020; Y. Zhang et al., 2020). However, standard operating protocols for DNA extraction,

library preparation, bioinformatics analysis, and validation of NGS are urgently required for diagnostic use in the clinic. Due to the difficulty and impossibility of large-scale validation of NGS in terms of cost and complexity for specific groups, small-scale studies using this approach to various diagnostic entities are of substantial benefit. Assuming that a particular pathogen should be abundant in an acute disease-associated metabiome, we performed unbiased metagenomic analysis on routine diagnostic respiratory samples (n = 29) to evaluate this method as a putative diagnostic strategy in respiratory specimens from specific public health or clinical settings where unbiased diagnostic methods can effectively complement classical diagnostic methods.

This research aimed to assess the potential of metagenomic analysis done directly on RNA extracted from diagnostic samples of seasonal influenza virus infection patients. We demonstrated a high degree of agreement between the sensitivity of influenza virus qPCR and metagenomic sequencing for BAL fluid samples, which is consistent with prior data using NGS on nasopharyngeal aspirates subtracted for host sequences. Our findings reveal unequivocally that NGS approaches are highly reliant on the kind of diagnostic item investigated. BAL fluid samples, an atypically valid source of material for diagnosing pneumonia and other lung infections, demonstrate a good association between the proportion of NGS influenza sequence reads and the influenza qPCR CT values. There was no association between qPCR and relative read abundance values in sputum and swab samples, even though influenza virus sequence contigs were quickly retrieved from these specimens. Correlation is most likely absent as a consequence of increased variability across microbial genomes. Given that presently available NGS technologies can only offer relative abundance values, correlations with absolute quantitation values produced by qPCR are likely to decline when the nonhuman background's variability rises.

Additionally, phylogenetic analysis of influenza HA and NA sequences were done on all influenza A-positive patients to understand better the sequence variation of the antigenic epitopes HA and NA throughout the 2012–2013 season. This is the first comprehensive research that has sequenced the whole genomes of seasonal influenza virus infections directly from diagnostic nucleic acid samples. Previous studies applying NGS concerning influenza genome analysis were performed from smaller cohorts of pandemic influenza cases, from lung tissue, or after subculturing of the virus, which is laborious; in addition, dependent on the viral load, not all subculturing is successful. Other efforts used amplicon sequencing to selectively read all influenza genome regions. hen variant calling was done on samples with more influenza virus sequence reads, one variation was discovered in two samples.

Considering the high error rates of RNA polymerase during influenza virus replication, one may expect more intrahost variants to be generated within a single infection. However, our results are in concordance with previously

published studies (most of them preselecting for influenza virus sequences) reporting a relatively low genetic diversity of influenza virus in patients. Concerning sensitivity, we find our results to be in concordance with observations made in previously published studies. Similar to these reports, we find a strong correlation between the percentages of NGS reads mapping to influenza A sequences and CT values from qPCR.

Except for one sample (sample 1,116), the influenza virus sequences from samples with CT values <35 were detected by UMERS, which is comparable to the results obtained by Greninger and colleagues. They included 17 samples (nasopharyngeal swabs) in the NGS analysis, with 15 of 17 samples containing between 105 and 109 viral particles/ml. Furthermore, in a study focusing on the analysis of the human virome in febrile children, the sensitivity of unbiased NGS compared to qPCR was estimated for human adenovirus and human bocavirus sequences. A strong correlation between NGS reads and CT values were observed, and adenoviral sequences were detected up to a CT of 35, while human bocavirus was detected up to only a CT value of 30. Similar results were obtained by a very recent study, in which enterovirus and rhinovirus sequences were detectable by NGS up to a CT value of 30.

Given that many bacteria can colonize the nasopharynx, a caveat of NGS-based analyses of respiratory tract samples is that it is challenging to discriminate between colonization and coinfection events that may be of putative clinical relevance. While the detection of only a few unambiguously mapped reads may be sufficient to conclude that a sample is positive for a given agent, owing to the limited amount of presently available data, it is challenging to define abundance thresholds that may indicate a pathogenic infection, even if the clinical context is supportive of such a conclusion. While established conventional diagnostics, such as qPCR, suffer from the same principal limitations, extensive optimization and validation over several decades have led to the empirical determination of universally agreed-upon conventions as to when a given PCR result may be sufficient to identify a potential pathogenic infection. Hence, to implement NGS technology in the clinical laboratory, there is an urgent need for studies that systematically address the standardization of NGS methods and the definition of parameters for analytical and clinical validation. In addition, comparative studies are needed to determine the relative abundance of viral, bacterial, and fungal sequences not only in diagnostic specimens from patients suffering from infectious diseases but also in cohorts of healthy Players.

Nevertheless, the results reported in this pilot study demonstrate that unbiased RNA sequencing is a valuable tool for complementing routine diagnostics, in particular in clinical or public health settings where routine diagnostics remain repeatedly negative and comprehensive surveillance for

emerging viruses is needed.

4. Conclusion

The evaluation of Metagenomics Next-Generation Sequencing (mNGS) as a diagnostic tool for influenza virus-positive respiratory samples in retired fitness players underscores its paramount significance in the realm of healthcare for this unique demographic. Influenza, with its global pandemic potential and high seasonal infection rates, presents a multifaceted health challenge, particularly when coupled with bacterial and viral coinfections leading to pneumonia, which disproportionately affect the elderly and immunocompromised Players. The threat is exacerbated during pandemics, where the incidence of severe bacterial-induced pneumonia soars.

Our exploration of this critical area of healthcare reveals that mNGS holds the potential to revolutionize the diagnosis and management of respiratory infections in retired fitness players. By directly detecting influenza viruses and their coinfecting agents, mNGS not only offers heightened sensitivity but also reveals previously unidentified viral variants, providing a comprehensive view of the pathogenic landscape. The synergy between next-generation sequencing (NGS) technologies and streamlined sample preparation protocols ushers in a new era of diagnostic precision. This strategic combination allows us to navigate complex diagnostic scenarios with agility, directly identifying pathogens in cases that once remained enigmatic. The choice of sequencing platforms and advanced analytical techniques further enhances the depth and accuracy of diagnosis.

In conclusion, the application of mNGS in the care of retired fitness players has the potential to elevate their healthcare to new heights. It empowers healthcare providers with the tools needed to swiftly and accurately diagnose respiratory infections, ultimately improving the quality of care and well-being in this unique demographic. As we continue to harness the potential of mNGS and NGS technologies, the future of healthcare for retired fitness players appears promising, with the prospect of proactive disease detection and targeted interventions on the horizon.

REFERENCE

Ai, J.-W., Zhang, H.-C., Cui, P., Xu, B., Gao, Y., Cheng, Q., . . . Zhang, W.-H. (2018). Dynamic and direct pathogen load surveillance to monitor disease progression and therapeutic efficacy in central nervous system infection using a novel semi-quantitive sequencing platform. *Journal of Infection, 76*(3), 307-310.

Boni, A. A., & Abremski, D. (2022). Commercialization Challenges and Approaches for Digital Health Transformation. *Journal of Commercial Biotechnology, 27*(1).

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics, 5*(1), 1-19.

Figueiredo, C. R. L. (2019). The unusual paradox of cancer-associated inflammation: an update. *Jornal Brasileiro de Patologia e Medicina Laboratorial, 55*, 321-332. doi:10.5935/1676-2444.20190029

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology, 59*(3), 307-321.

Hardak, E., Avivi, I., Berkun, L., Raz-Pasteur, A., Lavi, N., Geffen, Y., . . . Oren, I. (2016). Polymicrobial pulmonary infection in patients with hematological malignancies: prevalence, co-pathogens, course and outcome. *Infection, 44*, 491-497.

Huffnagle, G., Dickson, R., & Lukacs, N. (2017). The respiratory tract microbiome and lung inflammation: a two-way street. *Mucosal immunology, 10*(2), 299-306.

Jeon, Y. J., Zhou, Y., Li, Y., Guo, Q., Chen, J., Quan, S., . . . Lin, J. (2014). The feasibility study of non-invasive fetal trisomy 18 and 21 detection with semiconductor sequencing platform. *PloS one, 9*(10), e110240.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution, 35*(6), 1547.

Labelle, A. J., Arnold, H., Reichley, R. M., Micek, S. T., & Kollef, M. H. (2010). A comparison of culture-positive and culture-negative health-care-associated pneumonia. *Chest, 137*(5), 1130-1137.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*(14), 1754-1760.

Li, H., Gao, H., Meng, H., Wang, Q., Li, S., Chen, H., . . . Wang, H. (2018). Detection of pulmonary infectious pathogens from lung biopsy tissues by metagenomic next-generation sequencing. *Frontiers in Cellular and Infection Microbiology, 8*, 205.

Loeffelholz, M., & Chonmaitree, T. (2010). Advances in diagnosis of respiratory virus infections. *International journal of microbiology, 2010*.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., . . . Zhu, N. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet, 395*(10224), 565-574.

Magill, S. S., Edwards, J. R., Bamberg, W., Beldavs, Z. G., Dumyati, G., Kainer, M. A., . . . Nadle, J. (2014). Multistate point-prevalence survey of health care–associated infections. *New England Journal of Medicine, 370*(13), 1198-1208.

Martin, P. L. (2016). Migration, trade and remittances: Low-and high-skilled workers. *Remittances Review, 1*(1), 37-52.

Miao, Q., Ma, Y., Wang, Q., Pan, J., Zhang, Y., Jin, W., . . . Wang, M. (2018).

Microbiological diagnostic performance of metagenomic next-generation sequencing when applied to clinical practice. *Clinical Infectious Diseases, 67*(suppl_2), S231-S240.

Pan, T., Tan, R., Qu, H., Weng, X., Liu, Z., Li, M., & Liu, J. (2019). Next-generation sequencing of the BALF in the diagnosis of community-acquired pneumonia in immunocompromised patients. *Journal of Infection, 79*(1), 61-74.

Parize, P., Muth, E., Richaud, C., Gratigny, M., Pilmis, B., Lamamy, A., . . . Jagorel, F. (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. e571-574. e576.

Prachayangprecha, S., Schapendonk, C. M., Koopmans, M. P., Osterhaus, A. D., Schürch, A. C., Pas, S. D., . . . Smits, S. L. (2014). Exploring the potential of next-generation sequencing in detection of respiratory viruses. *Journal of clinical microbiology, 52*(10), 3722-3730.

Thorburn, F., Bennett, S., Modha, S., Murdoch, D., Gunson, R., & Murcia, P. R. (2015). The use of next generation sequencing in the diagnosis and typing of respiratory infections. *Journal of Clinical Virology, 69*, 96-100.

Wang, S., Ai, J., Cui, P., Zhu, Y., Wu, H., & Zhang, W. (2020). Diagnostic value and clinical application of next-generation sequencing for infections in immunosuppressed patients with corticosteroid therapy. *Annals of Translational Medicine, 8*(5).

Wilson, M. R., Sample, H. A., Zorn, K. C., Arevalo, S., Yu, G., Neuhaus, J., . . . Langelier, C. (2019). Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. *New England Journal of Medicine, 380*(24), 2327-2340.

Wypych, T. P., Wickramasinghe, L. C., & Marsland, B. J. (2019). The influence of the microbiome on respiratory health. *Nature immunology, 20*(10), 1279-1290.

Xing, X.-W., Zhang, J.-T., Ma, Y.-B., He, M.-W., Yao, G.-E., Wang, W., . . . Wang, X.-L. (2020). Metagenomic next-generation sequencing for diagnosis of infectious encephalitis and meningitis: a large, prospective case series of 213 patients. *Frontiers in Cellular and Infection Microbiology, 10*, 88.

Zhang, H.-C., Ai, J.-W., Cui, P., Zhu, Y.-M., Hong-Long, W., Li, Y.-J., & Zhang, W.-H. (2019). Incremental value of metagenomic next generation sequencing for the diagnosis of suspected focal infection in adults. *Journal of Infection, 79*(5), 419-425.

Zhang, H. C., Zhang, Q. R., Ai, J. W., Cui, P., Wu, H.-L., Zhang, W. H., & Wang, T. (2020). The role of bone marrow metagenomics next-generation sequencing to differential diagnosis among visceral leishmaniasis, histoplasmosis, and talaromycosis marneffei. *International journal of laboratory hematology, 42*(2), e52-e54.

Zhang, Y., Cui, P., Zhang, H.-C., Wu, H.-L., Ye, M.-Z., Zhu, Y.-M., . . . Zhang,

W.-H. (2020). Clinical application and evaluation of metagenomic next-generation sequencing in suspected adult central nervous system infection. *Journal of translational medicine, 18*(1), 1-13.