Latest RNA and DNA nanopore sequencing allows for rapid avian influenza profiling

Albert Perlas[1,2], Tim Reska[1,2,3], Guillaume Croville[4], Ferran Tarrés-Freixas[5,6,8], Jean-Luc Guérin[4], Natàlia Majó[6,7], Lara Urban[1,2,3*]

[1]Helmholtz AI, Helmholtz Zentrum Muenchen, Neuherberg, Germany

[2]Helmholtz Pioneer Campus, Helmholtz Zentrum Muenchen, Neuherberg, Germany

[3]Technical University of Munich, School of Life Sciences, Freising, Germany

[4]IHAP, Université de Toulouse, INRAE, ENVT, Toulouse, France

[5]IRTA. Programa de Sanitat Animal. Centre de Recerca en Sanitat Animal (CReSA), Campus de la Universitat Autònoma de Barcelona (UAB), Bellaterra, Spain

[6]Unitat mixta d'investigació IRTA-UAB en Sanitat Animal, Centre de Recerca en Sanitat Animal (CReSA), Campus de la Universitat Autònoma de Barcelona (UAB), 08193, Bellaterra, Catalonia, Spain

[7]Departament de Sanitat i Anatomia Animals, Facultat de Veterinària, Universitat Autònoma de Barcelona (UAB), Campus de la UAB, 08193, Bellaterra, Catalonia, Spain

[8]University of Vic-Central University of Catalonia (UVic-UCC), Vic, Spain.

*Corresponding author

E-mail: lara.urban@helmholtz-muenchen.de

Abstract

Avian influenza virus (AIV) currently causes a panzootic with extensive mortality in wild birds, poultry, and wild mammals, thus posing a major threat to global health and underscoring the need for efficient monitoring of its distribution and evolution. Here, we utilized a well-defined AIV strain to systematically investigate AIV characterization through rapid, portable nanopore sequencing by (i) benchmarking the performance of fully portable RNA extraction and viral detection; (ii) comparing the latest DNA and RNA nanopore sequencing approaches for in-depth AIV profiling; and (iii) evaluating the performance of various computational pipelines for viral consensus sequence creation and phylogenetic analysis. Our results show that the latest RNA-specific nanopores can accurately genomically profile AIV from native RNA while additionally detecting RNA epigenetic modifications. We further identified an optimal laboratory and bioinformatic pipeline for reconstructing viral consensus genomes from nanopore sequencing data at various rarefaction thresholds, which we validated by application to real-world environmental samples for AIV monitoring in livestock.

Author Summary

We tested portable, rapid, and easy-to-use technology to obtain more information about the potentially zoonotic RNA virus avian influenza virus, or AIV. AIV has spread globally via the migratory paths of wild birds, and endangers domestic birds, mammals, and human populations given past evidence of infections of different animal species. We here used novel genomic technology that is based on nanopores to explore the genomes of the virus; we established optimized ways of creating the viral genome by comparing different laboratory and computational approaches and the performance of nanopores that either sequence the viral RNA directly or the converted

DNA. We then applied the optimized protocol to dust samples which were collected from a duck farm in France during an AIV outbreak. We showed that we were able to use the resulting data to reconstruct the relationship between the virus responsible for the outbreak and previously detected AIV. Altogether, we showed how novel easy-to-use genomic technology can support the surveillance of potentially zoonotic pathogens by accurately recreating the viral genomes to better understand evolution and transmission of these pathogens.

Introduction

Avian influenza virus (AIV) currently causes the largest and deadliest panzootic on the European and American continents [1]; it is known to have spilled over from wild bird populations to poultry and humans, posing a risk for causing a future pandemic [2]. Wild birds are the main reservoir of low-pathogenicity AIV (LPAIV), in particular the Anseriformes and Charadriiformes orders [3]. These birds are asymptomatic to LPAIV and can spread the virus to poultry around the globe [4]. Once in gallinaceous species, LPAIV can evolve into high pathogenicity AIV (HPAIV), resulting in animal welfare, financial and social issues due to high poultry mortality, economic loss, and food insecurity [1]. LPAIV and HPAIV further have the potential to adapt and spread to mammalian species. Since the emergence of H5N1 HPAIV in a domestic goose in Guangdong China in 1996 (“Gs/GD lineage”), it has become clear that HPAIV can also be transmitted back to and subsequently maintained in wild bird populations [5]. As many Anseriformes and Charadriiformes populations perform long-distance migrations, they can rapidly spread AIV variants across countries and continents [6].

AIV is a segmented, negative-strand RNA virus from the *Orthomyxoviridae* family. Its error-prone polymerase, and therefore high mutation rate, as well as its segmented genome in combination with mixed infections allow this virus to be in continuous evolution due to antigenic drift and antigenic shift [4]. One such example is the frequent mutation of LPAIV into HPAIV after recurrent replication in poultry, which provides the perfect environment for the virus to mutate due to the high density of susceptible, genetically similar hosts [7]. This evolutionary plasticity of AIV means that the application of fast genomic technologies to determine their nucleotide composition can help to quickly characterize AIV genomic variation including low-frequency variants, predict virulence, reconstruct transmission dynamics, and determine an outbreak's origin [8].

The application of *in situ* real-time nanopore sequencing technology by Oxford Nanopore Technologies provides a currently unique genomics-based approach to characterize AIV in a fast, straightforward, and cost-efficient manner all around the world [9], which makes viral surveillance accessible in low- and middle-income countries as well as in remote field setting for wild bird monitoring. This technology has been established for AIV profiling through sequencing of complementary DNA (cDNA) after retro-transcription (RT) and multi-segment PCR amplification (M-RTPCR) [8,10,11]. While ligation-based and transposase-based rapid sequencing library preparations ("DNA-nanopore" chemistry R9) have been applied to nanopore-sequence cDNA from AIV [12], a variety of computational pipelines have subsequently been used for data analysis and consensus sequence generation, which have however not yet been systematically assessed and compared [8,10,13–17]. Keller et al. [14] have further applied direct RNA nanopore sequencing to the viral RNA (vRNA)

of AIV, which could circumvent biases introduced through cDNA synthesis [18]. This protocol is faster due to the omission of M-RTPCR, and further allows for the detection of RNA modifications [19]. Direct RNA sequencing through nanopore technology (DNA-nanopore chemistry R9) has, however, suffered from high sequencing error rates as well from high RNA input requirements and from a lack of multiplexing options for efficient sample processing [14,20].

Here, we used a well-defined viral culture to conduct a systematic study for AIV characterization through nanopore sequencing by (i) comparing cDNA and vRNA sequencing of AIV in terms of sequencing data throughput, data quality and consensus sequence accuracy, and by (ii) systematically assessing the performance of different computational analysis pipelines. Besides the current gold standard nanopores for DNA (DNA-nanopore chemistry R10) and RNA (DNA-nanopore chemistry R9) sequencing, we have for the first time applied RNA-specific nanopores ("RNA-nanopore" chemistry) for RNA virus profiling; this chemistry is based on completely new nanopores that have been optimized for direct RNA sequencing – in contrast to the previous DNA-nanopore chemistry R9 which relies on nanopores optimized for DNA sequencing and therefore suffers from a high sequencing error rate of >10% [14]. We further included several protocols to compare portable approaches for *in situ* applications with standard laboratory-based approaches. Finally, in order to test the application of nanopore sequencing to samples with lower AIV loads such as environmental samples, we optimized our approaches through data rarefaction simulations, and finally used our results to characterize AIV from non-invasively collected dust samples.

Results

Similar performance of laboratory-based and portable RNA extraction and quantification protocols

Using LPAIV H1N1 viral cultures, we found that the NucleoSpin RNA Virus kit was the most efficient RNA extraction approach, yielding the lowest Ct (cycle threshold) values. While we therefore continued our analyses with this kit, the portable Biomeme M1 Sample Prep Cartridge Kit, which allows for RNA extraction in just 5 minutes, yielded only slightly higher Ct values (S1). The standard RT-PCR and portable Mic qPCR systems further showed comparable performance (S1); we therefore continued our analyses with the Mic qPCR machine.

Nanopore sequencing genome coverage and rarefaction analysis

While the sequencing read length distributions were largely consistent across the cDNA, RNA002, and RNA004 nanopore sequencing approaches (Methods), the RNA002 dataset comprised a lower number of sequencing reads (**Fig 1A**). The alignment of the reads to the AIV reference segments further showed an uneven coverage distribution across the genome, with all sequencing approaches leading to similar coverage of the polymerase segments; every sequencing approach further led to increased coverage at the ends of each segment (**Fig 1B**).

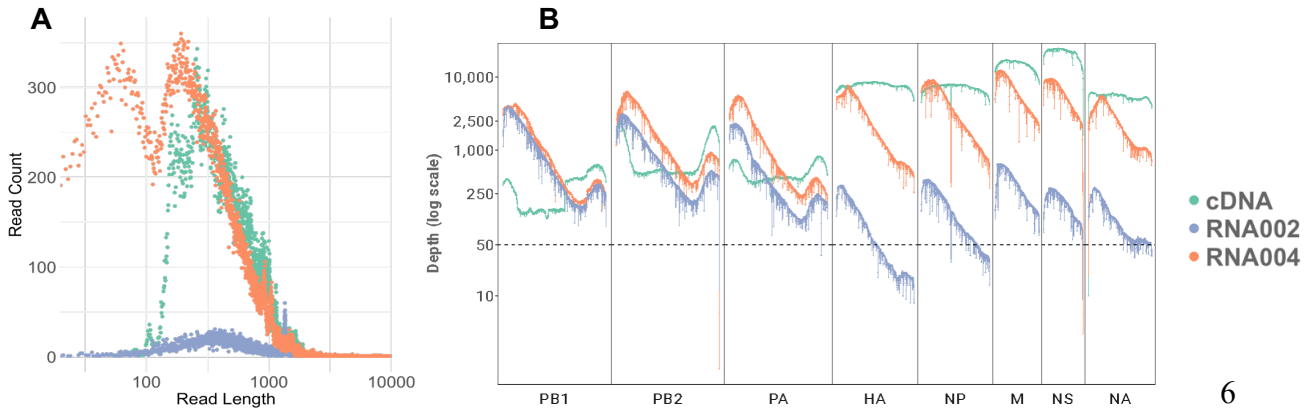


Fig 1. Nanopore sequencing results of an AIV viral culture using DNA-nanopores ("cDNA" sequencing through R10 chemistry; direct RNA sequencing through "RNA002" R9 chemistry) and RNA-nanopores (direct RNA sequencing through "RNA004" RNA chemistry). **A.** Sequencing read length distribution across the cDNA, RNA002, and RNA004 datasets. **B.** Reference genome coverage of the three sequencing datasets across all AIV segments (PB1: Polymerase basic 1, PB2: Polymerase basic 2, PA: Polymerase acidic, HA: Hemagglutinin, NP: Nucleoprotein, NA: Neuraminidase, M: Matrix, NS: Nonstructural). The horizontal line indicates a coverage of 50x.

Given the uneven throughput and coverage across the three sequencing datasets, we performed rarefaction for all downstream analyses and re-assessed the AIV genome coverage (Methods; **Table 1**; **S2**).

Table 1. Mean AIV reference genome coverage and (in brackets) total number of reads across each sequencing dataset (rows: cDNA, RNA002, RNA004) and respective rarefaction (columns: *raw* for total dataset; *max* for same mean coverage; *med* for 10% of the *max* data; *min* for 1% of the *max* data).

	Rarefaction Mean coverage (total reads)			
Dataset	*raw*	*max*	*med*	*min*
cDNA	5281 (326956)	530 (32660)	53 (3249)	8 (378)
RNA002	NA	537 (33255)	57 (3344)	6 (382)
RNA004	2809 (269621)	535 (51135)	54 (5140)	7 (511)

AIV consensus sequence creation

We first used the BIT score to evaluate the quality of the viral consensus sequence created by different computational pipelines in comparison to the known AIV reference (Methods). Even at maximum coverage (*max* rarefaction), only the reference-based

171 approaches BCFtools or iVar, and the iterative reference-based assembly tool IRMA
172 were able to create the full consensus sequence of all eight viral segments (**Fig 2A**).
173 Reference-based EPI2ME did not assemble the NS and HA segments. The hybrid
174 approach CZID as well as the *de novo* assembler metaFlye only assembled the largest
175 segments PA, PB1, and PB2, and did not create consensus sequences of the
176 unassembled reads mapping to the other segments. Flye (without the metagenomics
177 configuration) only assembled the PA and PB2 segments.
178
179 For our rarefied datasets (*med* and *min*), we found further performance differences
180 between BCFtools, iVar, and IRMA (**Fig 2A**). For all three sequencing approaches
181 (cDNA, RNA002, RNA004), BCFtools performed best across all viral segments, with
182 IRMA being unable to create certain segments at all at *min* rarefaction, namely PB1
183 for cDNA, and HA and NS for RNA002. Across the sequencing approaches, we only
184 found differences at *min* rarefaction where RNA004 outperformed RNA002, and cDNA
185 surpassed both RNA-based methods with the exception of the polymerase segments.
186
187 We next calculated the evolutionary distance between the whole-genome consensus
188 sequences and the known AIV reference (Methods). We found that BCFtools again
189 performed best in that it achieved the smallest evolutionary distance from the true
190 reference across all nanopore sequencing approaches and rarefaction thresholds (**Fig**
191 **2B**). The good performance of BCFtools was especially pronounced at *min* rarefaction,
192 where it clearly outperformed the other consensus sequence creation tools. Across
193 the nanopore sequencing approaches, the RNA-nanopore-based RNA004 and cDNA-
194 based approaches worked equally and were quite robust to the viral coverage; the

worse performance of the DNA-nanopore-based RNA002 approach was most noticeable for the minimum-coverage data (*min* rarefaction) (**Fig 2B**).

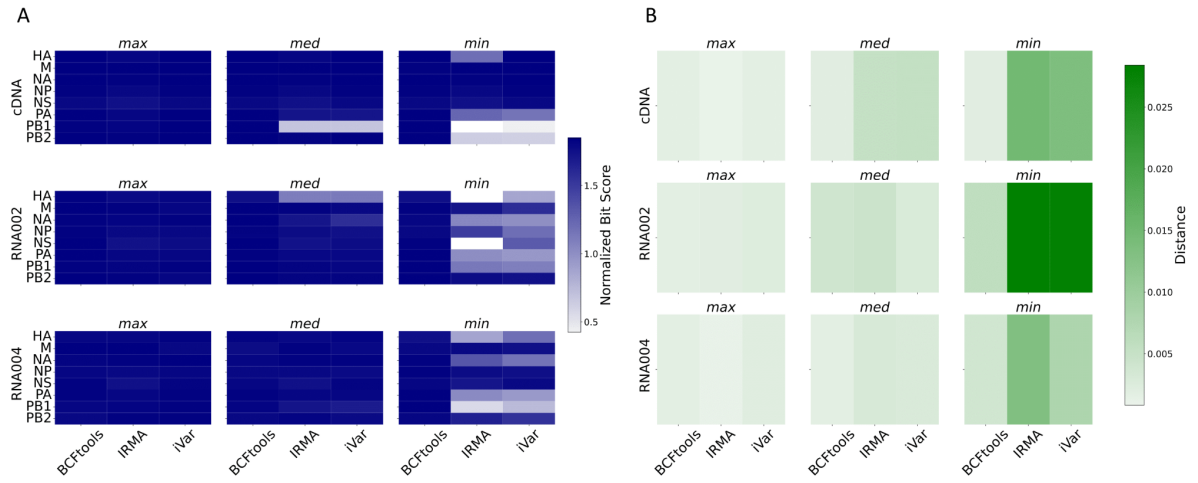


Fig 2. Evaluation of viral consensus sequence creation from nanopore sequencing datasets (cDNA, RNA002, RNA004) across all data rarefactions (*min*, *med*, *max*). The performance of the computational tools BCFtools, iVar, and IRMA, which were the best-performing approaches for the *max* datasets, is visualized. **A.** Consensus sequence evaluation across the eight viral AIV segments through normalized BIT scores calculated based on the known AIV reference. **B.** Consensus sequence evaluation through whole-genome evolutionary distance comparisons with the known AIV reference.

Nanopore sequencing-based AIV profiling in environmental samples

Given the good performance of cDNA and RN004 for viral consensus sequence creation also from smaller genomic datasets (*min* rarefaction), we next simulated sequencing data from environmental samples by rarefaction to 0.01% (*env_max*) and 0,001% (*env_min*). We focused this analysis on HA as the viral segment that is normally used for AIV lineage identification via phylogenetic analysis, and which plays an important role in host cell penetration. We found that the cDNA sequencing approach was very robust to extremely low viral coverage, yielding lower evolutionary

distances of the HA segment to its known reference. Specifically, with *env_max* at a distance of 0.0000025 and a mean coverage of 76, and *env_min* at a distance of 0.00000025 and a mean coverage of 10, these outcomes were significantly better than those achieved through direct RNA sequencing using RNA004. For RNA004, *env_max* resulted in a distance of 0.0012 with a mean coverage of 24, and *env_min* resulted in a distance of 0.0058 with a mean coverage of 5.

Given the good performance of AIV cDNA sequencing and BCFtools analysis for simulated environmental samples, we finally used this approach to process real environmental samples, namely four dust samples from a turkey farm in France. The samples ranged from Ct values of 24 to 26, and all resulted in similar read length distributions and coverage distributions across viral segments (S3). BCFtools was able to create consensus sequences of all eight viral segments, and an evolutionary based on known AIV strains from the NCBI influenza database (Methods) elucidated the phylogenetic relationship between the farm's viral strain and known AIVs (**Fig 3**).

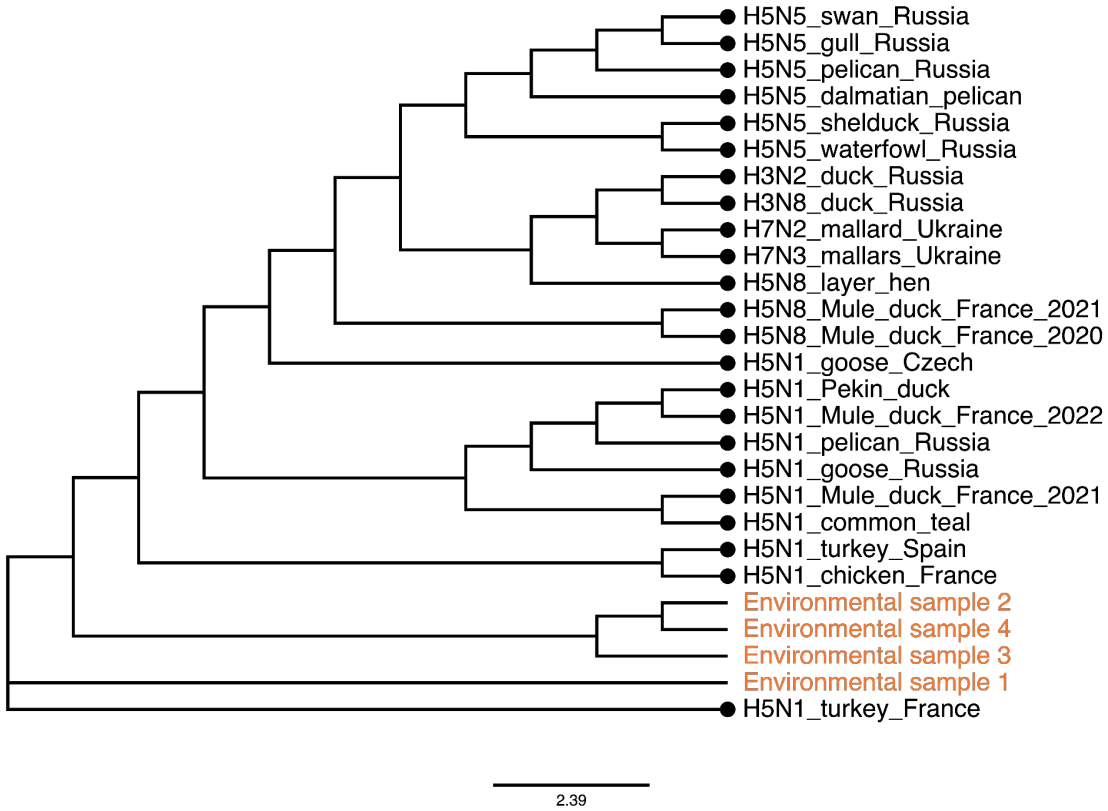


Fig 3. Phylogenetic tree of AIV consensus HA segments from four environmental samples (dust samples from turkey farm in France) and known European AIV strains. AIV of the environmental samples was assessed by cDNA nanopore sequencing, and consensus sequence was created through BCFtools.

RNA modification profiling

Finally, we identified m6A modifications in the raw RNA004 nanopore data as the best-performing direct RNA sequencing approach in our study. In total, we identified 2145 modifications, with the distribution of total modifications and modifications per base across segments as follows: HA (311 modifications, 0.179 modifications per base), M (227 modifications, 0.230 modifications per base), NA (307 modifications, 0.216 modifications per base), NP (253 modifications, 0.165 modifications per base), NS (185 modifications, 0.208 modifications per base), PA (237 modifications, 0.108 modifications per base), PB1 (282 modifications, 0.122 modifications per base), and

PB2 (343 modifications, 0.149 modifications per base). The M segment exhibited the highest modification density, while the PA segment displayed the lowest.

Discussion

Here, we present an optimized nanopore sequencing pipeline suitable for rapid field studies from non-invasively collected environmental samples. Our protocols aim to identify AIV strains as well as their evolution processes and potential transmission patterns. The implementation of such strategy for AIV monitoring – including in remote areas along long-distance migration routes of potential avian hosts – is very promising for rapidly and appropriately informing control measures as part of a "One Health" strategy.

Our study shows that direct viral RNA as well as cDNA nanopore sequencing provide robust genomic approaches to rapidly create viral consensus sequences *in situ*, to assess the virus' evolutionary trajectory. While previous studies have explored the application of nanopore sequencing to AIV, they have often been limited to one or a few processing pipelines [8,10,13–17]. Here we identify the optimal computational analysis pipeline for robust analysis of viral data across a range of simulated viral loads.

While direct RNA sequencing has previously been applied to AIV analysis [14], the high sequencing error rate of these protocols, that are based on standard DNA-nanopores, impeded meaningful analyses of the data. Here we find that the latest direct RNA nanopore sequencing technology (which is based on a unique RNA-nanopore specifically designed for transcriptomic rather than genomic research),

provides similar results to cDNA sequencing using Oxford Nanopore Technologies' established high-accuracy DNA-nanopores (R10 chemistry). Our study is therefore, to the best of our knowledge, the first to show that viral genomes can be profiled with high accuracy directly from their RNA without any cDNA reverse-transcription which is laborious, time-consuming and might introduce biases. We also show that direct RNA sequencing simultaneously allows for RNA modification calling [19]. To the best of our knowledge, this study marks the first in-stance of identifying m6A modifications in AIV using direct RNA nanopore sequencing technology. Such modifications play a critical role in viral RNA viruses, allowing them to mimic host RNA and thereby evade the host's immune system, which underscores the significance of direct RNA sequencing for epidemiology and immunology [21,22].

We compared the performance of several reference-based, *de novo* assembly, and hybrid computational approaches to reconstruct the viral consensus sequence from nanopore data at various rarefaction thresholds. While web-based tools such as EPI2ME and CZID are more user-friendly than the remaining tools which rely on the usage of the command line, they did not perform well in creating the consensus sequence of all AIV segments – even in the datasets with a high genome-wide coverage of >500x. In the case of the reference-based EPI2ME tool, the poor performance could be due to the analysis' restriction to US AIV references. In the case of the hybrid-assembler CZID, only assemblies from the longer viral segments could be obtained while many un-assembled reads aligned to the reference sequences of the smaller segments. We faced the same problem when using the *de novo* assembly command line tool Flye, which can be explained by Flye's incompatibility with reads shorther than 1 kb, which exceeds the entire length of some viral segments. We found

298 that the Flye version for metagenome assembly (using the –meta flag) worked better
299 for reference reconstruction; this might be related to the fact that metaFlye does not
300 assume even coverage across the genome, which is a suitable configuration for highly
301 diverse RNA viruses where amplification or targeting biases might result in uneven
302 coverage across segments [23–25].
303
304 The reference-based command line tools BCFtools and iVar as well as IRMA, which
305 relies on iterative refinement, were able to create high-quality viral consensus
306 sequences for all nanopore data if available at high-coverage. Although we could have
307 evaluated other reference-based tools, we chose to utilize the most commonly
308 employed ones for generating consensus sequences from AIV in this study. While
309 some of these computational pipelines have previously been applied to AIV nanopore
310 sequencing data, they have not yet been compared to each other, especially in
311 application to different nanopore sequencing modalities [13,14,16,17]. Rarefaction of
312 this high-coverage viral data identified BCFtools as the best tool in generating
313 consensus sequences across viral segments similar to the known reference
314 (measured by high BIT scores and small evolutionary distances). The good
315 performance of BCFtools throughout our analyses might be due to its – in comparison
316 to iVar and IRMA – relatively strong reliance on reference data [26,27] and our
317 incorporation of a comprehensive reference database. This means that the
318 performance of BCFtools might worsen in the case of highly divergent and previously
319 unseen RNA viruses.
320
321 Given the good performance of BCFtools in combination with cDNA and RNA004
322 nanopore sequencing data, we further simulated AIV sequencing from environmental

samples. We show that cDNA sequencing leads to relatively higher BIT scores and smaller evolutionary distances than RNA004 sequencing for such low-concentration samples, which might be due to the still relatively decreased sequencing accuracy of direct RNA sequencing in comparison to direct DNA sequencing (96% for RNA004 using the RNA chemistry, vs. 99% for cDNA using the R10 chemistry;). When testing our AIV profiling pipeline on real environmental samples, namely dust samples from duck farms, we therefore employed the combination of cDNA sequencing and BCFtools-based analysis. We were able to reconstruct complete viral consensus sequences from this data, which we leveraged to reconstruct a phylogenetic tree and to show evolutionary similarity between our AIV strains and contemporary H5 AIV strains. These H5 strains are responsible for the ongoing severe HPAIV panzootic [1], and one of the most closely related strains has actually been responsible for a H5 HPAIV outbreak in another french farm.

An additional challenge for the analysis of low-concentration viral samples can be the uneven segment coverage that we observed across and within viral segments. Our cDNA data showed decreased coverage of the polymerase segments, while the RNA002 data showed decreased coverage of the respective other segments. We hypothesize that these coverage disparities stem from biases introduced through the use of universal primers for cDNA amplification and through the oligo-nucleotide adapters targeting AIV for direct RNA sequencing, respectively. The newest direct RNA sequencing protocol RNA004, on the other hand, relies on an alternative ligase enzyme, which might explain its more even coverage across segments. Within segments, all nanopore sequencing approaches result in uneven coverage. This especially applies to the direct RNA sequencing approaches, where the systematic

348 decrease in coverage towards the end of the segment might be explained by
349 sequencing adapters targeting the segments' conserved 3'-end and by rapid RNA
350 fragmentation [14].
351
352 Our study additionally showcases the field applicability of our nanopore sequencing
353 protocols by benchmarking fully portable equipment. While we found that standard
354 column-based viral RNA extraction outperformed more portable alternatives, the
355 Biomeme M1 Sample Prep Cartridge approach only led to slightly increased Ct values,
356 suggesting its potential for future field studies. We further show that the MIC qPCR [8],
357 MinION MK1c nanopore sequencing device, and rapid library preparation protocols
358 provide a fully portable framework to conduct AIV profiling at the point of interest all
359 around the world, even without internet access.
360

361 Methods

362 RNA extraction and quantification

363 H1N1 LPAIV was isolated from a duck sample in 2006 (strain
364 A/duck/Italy/281904/2006) and isolated in specific pathogen-free (SPF) eggs as
365 previously described [30]. The high-quality reference genome was obtained from
366 Sanger sequencing data ([31], GenBank accession number: FJ432771). We extracted
367 RNA from egg allantoic fluids using Macherey-Nagel's NucleoSpin RNA Virus
368 extraction kit, and quantified the extracted RNA using the Qubit RNA BR assay. We
369 additionally used Biomeme's M1 Sample Prep Cartridge Kit For RNA 2.0 [8] and
370 Lucigen's Quick Extract DNA Extraction Solution kit to assess the performances of
371 faster and portable RNA extraction approaches. For Quick Extract, we followed the

manufacturer's instructions and, additionally, an alternative method adapted for SARS-CoV-2 RNA extraction [32]. We then compared the performance of the different kits in terms of detection and quantification rates using standard RT-PCR (Applied Biosystems 7500 Fast Instrument, Thermo Fisher) and portable RT-PCR (Magnetic Induction Cycler quantitative PCR (Mic qPCR), Bio Molecular Systems). We targeted a highly conserved region of 99 bases of the AIV MP gene using previously established approaches to detect and quantify AIV using RT-PCR [33,34].

Nanopore sequencing

We then performed nanopore sequencing of the NucleoSpin RNA extracts. First, we performed direct vRNA sequencing using the DNA-nanopore (R9 chemistry; “RNA002” kit) and the RNA-nanopore chemistry (RNA chemistry; “RNA004” kit). We specifically targeted AIV RNA following the protocol described by Keller et al. [14]. Briefly, direct RNA nanopore sequencing requires a reverse transcriptase adapter (RTA) which usually captures poly(A) tails of the messenger RNA (mRNA); a sequencing adapter then ligates to the RTA and directs the mRNA to the nanopore. To target AIV RNA, we used a modified RTA, i.e. a custom oligo-nucleotide that is complementary to the 3’-region that is conserved across all AIV segments. As these conserved regions differ slightly across segments, we used two custom oligo-nucleotides, RTA-U12 and RTA-U12.4, which were mixed at a molar ratio of 2:3 to a total concentration of 1.4µM [14]. We subsequently used the portable MinION Mk1c device for nanopore sequencing; for the R9 chemistry sequencing, we used a FLO-MIN106 R9.4.1 flow cell, and for the RNA chemistry, we used a FLO-MIN004RA flow cell.

Second, we performed cDNA sequencing using the latest DNA-nanopore chemistry (R10 chemistry) and rapid barcoding library preparation (SQK-RBK114.24) after cDNA conversion of the extracted RNA and multi-segment amplification through M-RTPCR. M-RTPCR was performed as described previously, targeting the conserved regions across all AIV segments [35,36]. Briefly, the extracted RNA was mixed with Superscript III One-Step PCR reaction buffer and the previously defined primers, the PCR reactions were run on a portable Mic qPCR device. For sequencing, we used three barcodes with the same sample to increase the total quantity of cDNA added to the final sequencing library. We subsequently used the portable MinION Mk1c device and a FLO-MIN114 R10.4.1 flow cell for nanopore sequencing.

Data processing

We obtained raw nanopore sequencing data in fast5 format for the DNA-nanopore, and in pod5 format for the RNA-nanopore sequencing runs. For the DNA-nanopore runs, we used the Guppy (v6.4.8+31becc9) high-accuracy basecalling model (HAC; rna_r9.4.1_70bps_hac model for vRNA, dna_r10.4.1_e8.2_400bps_hac model for cDNA); for the RNA-nanopore run, we used the Dorado (v0.4.3+656766b) HAC model for RNA (rna004_130bps_hac). After removing short reads (<50 bases) using SeqKit (v2.4.0) [37], we used Minimap2 (v2.26) [38] with the *-ax map-ont* configuration for cDNA and the *-ax splice -uf -k7* configuration for vRNA reads to align the resulting fastq files to our ground-truth reference genome (GenBank accession number: FJ43277). We converted the resulting sam files to bam files, indexed, and sorted them using SAMtools (v1.17) [39] to obtain the genome coverage distribution.

Data rarefaction

To compare all nanopore sequencing results, we rarefied the three genomic datasets (cDNA, vRNA by DNA-nanopore: RNA002, vRNA by RNA-nanopore: RNA004) from the "*raw*" data to the same mean coverage ("*max*" data). After rarefying the cDNA fastq file to 10% and the RNA004 fastq file to 20% of its original number of reads, a similar mean coverage to the RNA002 fastq file was achieved (mean genome coverage of 537). We further rarefied this *max* data to simulate results from samples with lower viral load, namely to 10% of the *max* data ("*med*") and to 1% of the *max* data ("*min*"). Finally, to simulate real environmental samples with potentially extremely low viral loads, we additionally rarefied the raw cDNA and RNA004 data to 0.01% ("*env_max*") and 0.001% ("*env_min*") for follow-up analyses.

Consensus sequence construction

For reference-based consensus sequence creation, we mapped each dataset to a reference database generated for each segment from the NCBI Influenza Virus Database, which contains all AIV nucleotide sequences from Europe (as of 04/03/2023). We excluded the true reference sequence of our H1N1 virus from all segment-specific reference databases in order to simulate a realistic situation where the true genomic sequence of our AIV strain would not yet be known. We indexed the reference databases and mapped our sequencing reads against the databases using Minimap2. We then indexed and sorted the sam files and converted them to bam files using samtools. Using samtools idxstats, we selected the reference to which most reads mapped across segments. All our reads were then mapped to the full best reference genome using Minimap2. We then tested two standard reference-based

445 computational pipelines to create the consensus sequence from this alignment,
446 BCFtools (v1.17) [26] and iVar (v1.4.2) [27].
447 We additionally used the Iterative Refinement Meta-Assembler (IRMA; v1.0.3) [40] that
448 iteratively refines the reference used in the analysis to increase the accuracy of the
449 consensus sequence obtained. Using this pipeline, the consensus of each segment
450 can be obtained directly from the fastq file without intermediate steps required by the
451 user. We used the "FLU-minion" configuration for nanopore sequencing data, which
452 drops the median read Q-score filter from 30 to 0, raises the minimum read length
453 from 125 to 150, raises the frequency threshold for insertion and deletion refinement
454 from 0.25 to 0.75 and 0.6 to 0.75, respectively, and lowers the Smith-Waterman
455 mismatch penalty from 5 to 3 as well as the gap open penalty from 10 to 6.
456 We further applied Oxford Nanopore Technologies' EPI2ME (v.5.1.9.) workflow for
457 influenza viruses ("wf-flu") to our data, which is also based on a reference-based
458 consensus sequence creation approach, but which uses a specific influenza reference
459 database which only focuses on the FluA and FluB segments
460 (https://labs.EPI2ME.io/influenza-workflow/).
461
462 For *de novo* consensus sequence creation, we used Flye (v2.9.2) with and without –
463 meta flag [41] followed by assembly polishing using racon (v1.4.3) [42]. We
464 additionally applied the Chan Zuckerberg ID (CZID) [43,44] pipeline to our data, which
465 performs a combination of *de novo* and reference-based approaches: It uses metaFlye
466 to assemble the data and generate contigs, followed by Minimap2-alignments of the
467 still unassembled reads against the NCBI database [45].
468

To evaluate the consensus sequence creation pipelines, we first used blastn (v2.15) [46] to align every consensus segment to our known reference and we normalized the resulting BIT score by segment size to facilitate comparisons across different segments. Then, we assembled the whole-genome consensus sequences by concatenating the segment-specific consensuses derived from each of the rarefied datasets. We next performed a phylogenetic analysis of all reconstructed AIV whole-genome consensus sequences together with the known reference using a maximum likelihood approach as implemented in IQ-TREE (Jukes-Cantor nucleotide substitution model) (v2.0.6) [47] to obtain pairwise likelihood distances between our consensus sequences and the reference.

Environmental samples

We finally obtained real environmental samples (surface dust collected with dry wipes on building's walls and feeders) from 4 HPAIV H5N1 Gs/GD lineage outbreaks in 2022 and 2023 in duck farms in South-west and West regions of France [48]. The environmental samples were processed and analyzed as described above for the LPAIV H1N1 viral cultures. For the phylogenetic tree reconstruction, we incorporated all recent AIV strains from Europe (from January 1st 2020 until May 1st 2023) from the NCBI Influenza Virus Database; visualization was done using IROKI [49]. Due to the relevance of the HA segment for host cell penetration and phylogenetic analysis, we exclusively focused this analysis on this segment.

Detection of RNA modifications

We additionally searched for N^6-methyladenosine (m6A) RNA modifications in the RNA-nanopore data using the respective Dorado basecalling model

(rna004_130bps_sup v3.0.1_m6A_DRACH@v1). Subsequent analysis was performed using Modkit (v0.2.4.)(https://github.com/nanoporetech/modkit).

Data access

Original fastq files from all the sequencing runs are available on the European Nucleotide Archive (ENA) under the accession number PRJEB72673.

All our computational scripts are available via the GitHub repository *real-time_surveillance_of_avian-influenza*: https://github.com/Albertperlas/Latest-RNA-and-DNA-nanopores-allow-for-rapid-avian-influenza-profiling

Acknowledgement

We would like to acknowledge Dr. Ana Moreno from the Istituto Zooprofilattico Sperimentale della Lombardia e dell' Emilia Romagna (Brescia, Italy) who kindly provided us with the A/duck/Italy/281904/2006 LPAIV H1N1 isolate.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

AP conducted the nanopore sequencing experiments and subsequent bioinformatic analysis. TR, GC, FT, and LU provided essential advice on experimental design and bioinformatics. NM and JG collected the samples for the study. AP and LU significantly contributed to the discussion and manuscript writing. All authors, AP, TR, GC, FT, NM, JG, and LU, contributed to writing, reviewing, editing, and approving the final manuscript.

References

1. Authority EFS, European Centre for Disease Prevention and Control, Influenza EURL for A, Adlhoch C, Fusaro A, Gonzales JL, et al. Avian influenza overview September–December 2023. EFSA J. 2023;21: e8539. doi:10.2903/j.efsa.2023.8539

2. Mostafa A, Abdelwhab EM, Mettenleiter TC, Pleschka S. Zoonotic potential of influenza A viruses: A comprehensive overview. Viruses. 2018. doi:10.3390/v10090497

3. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. Evolution and ecology of influenza A viruses. Microbiol Rev. 1992;56: 152. doi:10.1128/mmbr.56.1.152-179.1992

4. Swayne DE. Avian Influenza. Swayne DE, editor. Oxford, UK: Blackwell Publishing Ltd.; 2008. doi:10.1002/9780813818634

5. Verhagen JH, Fouchier RAM, Lewis N. Highly Pathogenic Avian Influenza Viruses at the Wild–Domestic Bird Interface in Europe: Future Directions for Research and Surveillance. Viruses. 2021;13: 212. doi:10.3390/v13020212

6. Fourment M, Darling AE, Holmes EC. The impact of migratory flyways on the spread of avian influenza virus in North America. BMC Evol Biol. 2017;17: 118. doi:10.1186/s12862-017-0965-4

7. Van Oosterhout C. Mitigating the threat of emerging infectious diseases; a coevolutionary perspective. Virulence. 2021;12: 1288–1295. doi:10.1080/21505594.2021.1920741

8. de Vries EM, Cogan NOI, Gubala AJ, Mee PT, O'Riley KJ, Rodoni BC, et al. Rapid, in-field deployable, avian influenza virus haemagglutinin characterisation tool using MinION technology. Sci Rep. 2022;12: 11886. doi:10.1038/S41598-022-16048-Y

9. Urban L, Perlas A, Francino O, Martí-Carreras J, Muga BA, Mwangi JW, et al. Real-time genomics for One Health. Mol Syst Biol. 2023;19: e11686. doi:10.15252/msb.202311686

10. Crossley BM, Rejmanek D, Baroch J, Stanton JB, Young KT, Killian ML, et al. Nanopore sequencing as a rapid tool for identification and pathotyping of avian influenza A viruses. J Vet Diagn Invest. 2021;33: 253–260. doi:10.1177/1040638720984114

11. King J, Harder T, Beer M, Pohlmann A. Rapid multiplex MinION nanopore sequencing workflow for Influenza A viruses. BMC Infect Dis. 2020;20: 1–8. doi:10.1186/S12879-020-05367-Y/FIGURES/1

12. Ip HS, Uhm S, Killian ML, Torchetti M. An evaluation of avian influenza virus whole genome sequencing approaches using nanopore technology. 2023. doi:10.20944/preprints202301.0480.v1

13. Croville G, Walch M, Lèbre L, Silva S, Filaire F, Guérin J-L. An amplicon-based nanopore sequencing workflow for rapid tracking of avian influenza outbreaks, France, 2020-2022. bioRxiv; 2023. p. 2023.05.15.538689. doi:10.1101/2023.05.15.538689

14. Keller MW, Rambo-Martin BL, Wilson MM, Ridenour CA, Shepard SS, Stark TJ, et al. Direct RNA Sequencing of the Coding Complete Influenza A Virus Genome. Sci Rep. 2018;8: 14408. doi:10.1038/s41598-018-32615-8

15. King J, Harder T, Globig A, Stacker L, Günther A, Grund C, et al. Highly pathogenic avian influenza virus incursions of subtype H5N8, H5N5, H5N1, H5N4, and H5N3 in Germany during 2020-21. Virus Evol. 2022;8: veac035. doi:10.1093/ve/veac035

16. Rambo-Martin BL, Keller MW, Wilson MM, Nolting JM, Anderson TK, Vincent AL, et al. Influenza A Virus Field Surveillance at a Swine-Human Interface. Lowen AC, editor. mSphere. 2020;5. doi:10.1128/mSphere.00822-19

17. Nabeshima K, Asakura S, Iwata R, Honjo H, Haga A, Goka K, et al. Sequencing methods for HA and NA genes of avian influenza viruses from wild bird feces using Oxford Nanopore sequencing. Comp Immunol Microbiol Infect Dis. 2023;102: 102076. doi:10.1016/j.cimid.2023.102076

18. Viehweger A, Krautwurst S, Lamkiewicz K, Madhugiri R, Ziebuhr J, Hölzer M, et al. Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. Genome Res. 2019;29: 1545–1554. doi:10.1101/gr.247064.118

19. Abebe JS, Verstraten R, Depledge DP. Nanopore-Based Detection of Viral RNA Modifications. mBio. 2022;13: e03702-21. doi:10.1128/mbio.03702-21

20. Jain M, Abu-Shumays R, Olsen HE, Akeson M. Advances in nanopore direct RNA sequencing. Nat Methods 2022 1910. 2022;19: 1160–1164. doi:10.1038/s41592-022-01633-w

21. Courtney DG, Kennedy EM, Dumm RE, Bogerd HP, Tsai K, Heaton NS, et al. Epitranscriptomic Enhancement of Influenza A Virus Gene Expression and Replication. Cell Host Microbe. 2017;22: 377-386.e5. doi:10.1016/j.chom.2017.08.004

22. Lu M, Zhang Z, Xue M, Zhao BS, Harder O, Li A, et al. N6-methyladenosine modification enables viral RNA to escape recognition by RNA sensor RIG-I. Nat Microbiol. 2020;5: 584–598. doi:10.1038/s41564-019-0653-9

23. Hunt M, Gall A, Ong SH, Brener J, Ferns B, Goulder P, et al. IVA: accurate de novo assembly of RNA virus genomes. Bioinformatics. 2015;31: 2374–2376. doi:10.1093/bioinformatics/btv120

24. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods. 2020;17: 1103–1110. doi:10.1038/s41592-020-00971-x

25. Meleshko D, Hajirasouliha I, Korobeynikov A. coronaSPAdes: from biosynthetic gene clusters to RNA viral assemblies. Bioinformatics. 2021;38: 1–8. doi:10.1093/bioinformatics/btab597

26. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10: 1–4. doi:10.1093/GIGASCIENCE/GIAB008

27. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 2019;20: 8. doi:10.1186/s13059-018-1618-7

28. Zhang T, Li H, Ma S, Cao J, Liao H, Huang Q, et al. The newest Oxford Nanopore R10.4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. Appl Environ Microbiol. 2023;89: e0060523. doi:10.1128/aem.00605-23

29. Diensthuber G, Pryszcz L, Llovera L, Lucas MC, Delgado-Tejedor A, Cruciani S, et al. Enhanced detection of RNA modifications and mappability with high-accuracy nanopore RNA basecalling models. bioRxiv; 2023. p. 2023.11.28.568965. doi:10.1101/2023.11.28.568965

30. Brauer R, Chen P. Influenza virus propagation in embryonated chicken eggs. J Vis Exp JoVE. 2015; 52421. doi:10.3791/52421

31. Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database. 2020;2020: baaa062. doi:10.1093/database/baaa062

32. Ladha A, Joung J, Abudayyeh OO, Gootenberg JS, Zhang F. A 5-min RNA preparation method for COVID-19 detection with RT-qPCR. medRxiv; 2020. p. 2020.05.07.20055947. doi:10.1101/2020.05.07.20055947

33. Sánchez-González R, Ramis A, Nofrarías M, Wali N, Valle R, Pérez M, et al. Pathobiology of the highly pathogenic avian influenza viruses H7N1 and H5N8 in different chicken breeds and role of Mx 2032 G/A polymorphism in infection outcome. Vet Res. 2020;51: 113. doi:10.1186/s13567-020-00835-4

34. Spackman E, Senne DA, Myers TJ, Bulaga LL, Garber LP, Perdue ML, et al. Development of a real-time reverse transcriptase PCR assay for type A influenza virus and the avian H5 and H7 hemagglutinin subtypes. J Clin Microbiol. 2002;40: 3256–3260. doi:10.1128/JCM.40.9.3256-3260.2002

35. Thielen P. Influenza Whole Genome Sequencing with Integrated Indexing on Oxford Nanopore Platforms. 2022 [cited 16 Jun 2023]. Available: https://www.protocols.io/view/influenza-whole-genome-sequencing-with-integrated-wykffuw

638 36. Kampmann M-L, Fordyce SL, Ávila-Arcos MC, Rasmussen M, Willerslev E,
639 Nielsen LP, et al. A simple method for the parallel deep sequencing of full
640 influenza A genomes. J Virol Methods. 2011;178: 243–248.
641 doi:10.1016/j.jviromet.2011.09.001

642 37. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for
643 FASTA/Q File Manipulation. PloS One. 2016;11: e0163962.
644 doi:10.1371/journal.pone.0163962

645 38. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinforma Oxf
646 Engl. 2018;34: 3094–3100. doi:10.1093/bioinformatics/bty191

647 39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The
648 Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009;25:
649 2078–2079. doi:10.1093/bioinformatics/btp352

650 40. Shepard SS, Meno S, Bahl J, Wilson MM, Barnes J, Neuhaus E. Viral deep
651 sequencing needs an adaptive approach: IRMA, the iterative refinement meta-
652 assembler. BMC Genomics. 2016;17: 708. doi:10.1186/s12864-016-3030-6

653 41. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al.
654 metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat
655 Methods. 2020;17: 1103–1110. doi:10.1038/s41592-020-00971-x

656 42. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome
657 assembly from long uncorrected reads. Genome Res. 2017;27: 737–746.
658 doi:10.1101/gr.214270.116

659 43. Kalantar KL, Carvalho T, de Bourcy CFA, Dimitrov B, Dingle G, Egger R, et al.
660 IDseq-An open source cloud-based pipeline and analysis service for
661 metagenomic pathogen detection and monitoring. GigaScience. 2020;9:
662 giaa111. doi:10.1093/gigascience/giaa111

663 44. Simmonds SE, Ly L, Beaulaurier J, Lim R, Morse T, Thakku SG, et al. CZ ID: a
664 cloud-based, no-code platform enabling advanced long read metagenomic
665 analysis. bioRxiv; 2024. p. 2024.02.29.579666. doi:10.1101/2024.02.29.579666

666 45. Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, et al.
667 Database resources of the National Center for Biotechnology Information.
668 Nucleic Acids Res. 2021;49: D10–D17. doi:10.1093/nar/gkaa892

669 46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment
670 search tool. J Mol Biol. 1990;215: 403–410. doi:10.1016/S0022-2836(05)80360-
671 2

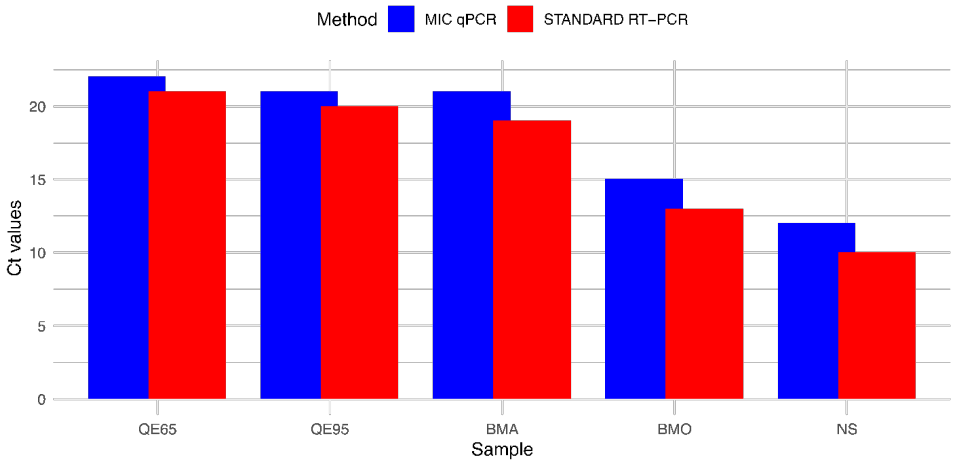
672 47. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and
673 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.
674 Mol Biol Evol. 2015;32: 268–274. doi:10.1093/molbev/msu300

675 48. Croville G, Walch M, Sécula A, Lèbre L, Silva S, Filaire F, et al. An amplicon-
676 based nanopore sequencing workflow for rapid tracking of avian influenza

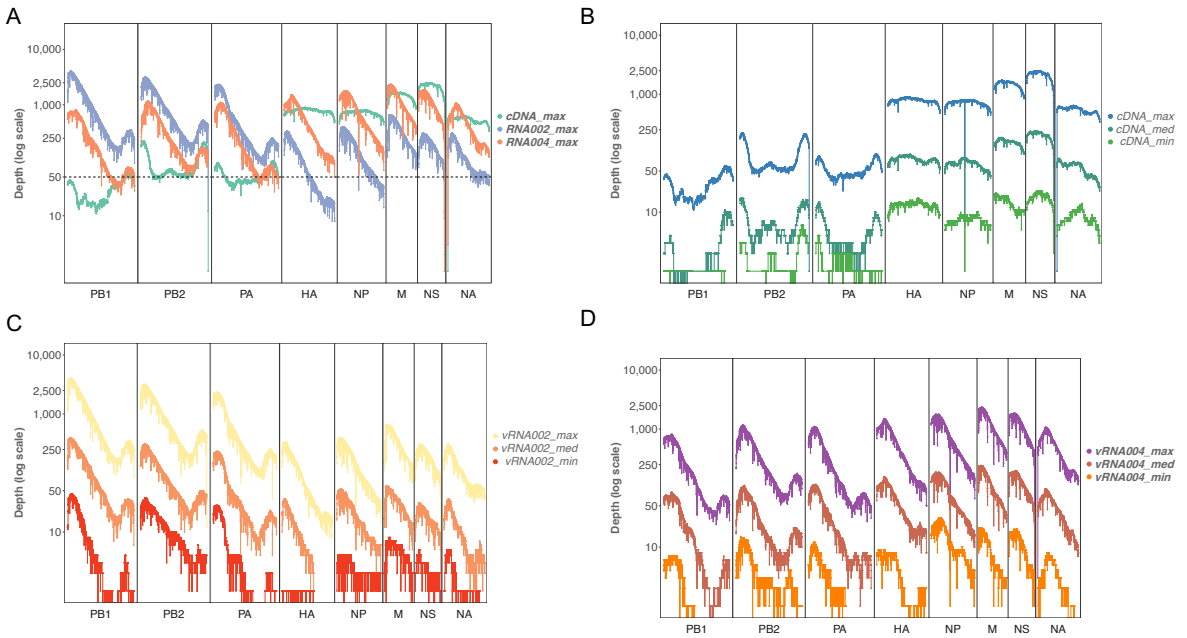
outbreaks, France, 2020-2022. Front Cell Infect Microbiol. 2024;14. Available: https://www.frontiersin.org/articles/10.3389/fcimb.2024.1257586

49. Moore RM, Harrison AO, McAllister SM, Polson SW, Wommack KE. Iroki: automatic customization and visualization of phylogenetic trees. PeerJ. 2020;8: e8584. doi:10.7717/peerj.8584

Supporting Information



S1 Figure. Comparison of Ct values from different RNA extraction kits and quantification methods. The Ct values were determined using the NucleoSpin RNA Virus extraction kit (NS), the Biomeme M1 Sample Prep Cartridge Kit for RNA 2.0 with the manufacturer's protocol (BMO) and a modified protocol by de Vries et al. (2022) (BMA), and the Quick Extract DNA Extraction Solution with the manufacturer's protocol (QE95) and an alternative method for SARS-CoV-2 RNA extraction by Ladha et al. (2020) (QE65). Quantification was performed using standard real-time PCR (red columns) and a portable real-time PCR (Mic qPCR, blue columns) targeting the M segment of the virus.



694

695 **S2 Figure.** Coverage of the different datasets after rarefaction. **A.** Maximum coverage with similar mean coverage

696 of all datasets (*cDNA-max, vRNA004-max*, and *vRNA002-max*). **B.** Coverage from cDNA datasets after rarefaction

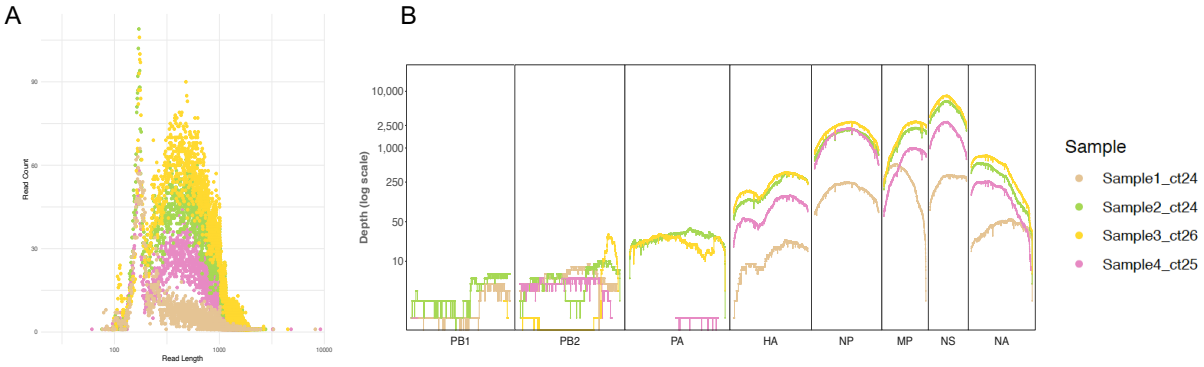
697 (*cDNA-max*, *cDNA-med*, and *cDNA-min*). **C.** Coverage from vRNA002 dataset (*vRNA002-max*) after rarefaction

698 (*vRNA002-med*, and *vRNA002-min*). **D.** Coverage from vRNA004 dataset after rarefaction (*vRNA004-max*,

699 v*RNA004-med*, and *vRNA004-min*).

700

701



702

703 **S3 Figure. A.** Read length distribution plots of the environmental samples. **B.** Coverage of each nucleotide

704 position in each segment from cDNA sequencing of our environmental samples.

705

706