Short communication

Metagenomic sequencing of zoonotic viruses: evaluation of a CRISPR-Cas–based rRNA depletion system

Gianpiero Zamperin[1*], Elisa Palumbo[1], Martina Castellan[1], Sabrina Marciano[1], Alice Fusaro[1], Isabella Monne[1]

[1]Istituto Zooprofilattico Sperimentale delle Venezie - SCS5 - IT

[*]Corresponding author at: Istituto Zooprofilattico Sperimentale delle Venezie - SCS5 - IT

E-mail: gzamperin@izsvenezie.it

Abstract

Pathogen-agnostic diagnostics are crucial for the early detection of emerging viruses. Shotgun metagenomic sequencing enables unbiased detection of viral genomes but is frequently constrained by the abundance of host and microbial ribosomal RNA (rRNA), which reduces sensitivity and increases sequencing costs. CRISPR-Cas9–based rRNA depletion has emerged as an alternative to enzymatic methods; however, its performance for the characterization of zoonotic viruses across diverse animal hosts and tissues remains underexplored. We compared CRISPR-Cas9 (Jumpcode CRISPRclean™ Plus) and RNase H–based enzymatic depletion (Ribo-Zero Plus, Illumina) using 12 samples positive for rabies lyssavirus, influenza A virus, West Nile virus or norovirus, from multiple host species and tissues, including both high-quality and degraded RNA. CRISPR-Cas9 efficiently reduced rRNA content (14.5%) but recovered fewer viral reads than Ribo-Zero, which achieved up to 60.7× enrichment. Both methods produced complete viral consensus genomes when RNA quality and viral load were sufficient. However, based on the data generated here, enzymatic depletion currently remains more efficient and cost-effective for viral metagenomics. Further optimization of CRISPR-Cas9 workflows could enhance its utility for viral surveillance and diagnostics.

Keywords

CRISPR-Cas, metagenomics, NGS, rRNA, virus

The recent COVID-19 pandemic underscored the critical need for readily accessible, pathogen-agnostic diagnostic tools capable of controlling a novel viral disease prior to its developing into a global pandemic event (The White House, 2021). Next generation sequencing (NGS)-based viral metagenomics -in particular shotgun metagenomics - has proven to meet the agnostic requirement, as it enables sequencing of all the genetic material, both DNA and RNA, without any prior knowledge of the sequence itself (Schlaberg et al., 2017). Despite its proven value for the detection and characterization of both known and novel viruses (Babiker et al., 2020; Fischer et al., 2015; Rajagopala et al., 2021; Schlaberg et al., 2017; Wilson et al., 2019), shotgun metagenomics still remains somehow inefficient because it usually requires large sample input and subsequent high sequencing depth. These demands result in substantially higher sequencing costs compared with other approaches, such as amplicon-based methods. To reduce costs while improving viral detection sensitivity, it is necessary to deplete uninformative genetic material - primarily host- and microbiota-derived sequences - before sequencing. Among these, ribosomal RNA (rRNA) is one of the most abundant nucleic acid species in the majority of clinical and environmental samples and therefore represents an ideal target for depletion strategies aimed at enriching viral nucleic acids.

For an adaptive immune response against phages, bacteria have evolved a CRISPR (clustered regularly interspaced short palindromic repeats) and Cas (CRISPR-associated) system. Cas nucleases such as Cas9 use this molecular machinery to eliminate phage DNA while sparing the bacteria's own genetic material. This system can be reprogrammed to target unwanted nucleic acid molecules, such as rRNA, through the use of specific pools of single-guide RNAs that direct Cas9 to defined sequences, enabling selective cleavage and consequent removal of the targeted molecules. Previous evaluations of this CRISPR–Cas9–based rRNA depletion system (e.g., Jumpcode CRISPRclean™ Plus) have mainly examined human respiratory specimens for SARS-CoV-2 (Cerón et al., 2023), with

limited evidence across diverse zoonotic RNA viruses, hosts, and tissue types. To address this gap, we performed a head-to-head evaluation between a CRISPR–Cas9 rRNA depletion workflow (Jumpcode CRISPRclean™ Plus Stranded Total RNA Prep) and an RNase H enzymatic depletion workflow (Illumina Stranded Total RNA Prep with Ribo-Zero Plus). Our comparison used 12 clinically or experimentally positive samples representing four epidemiologically relevant RNA viruses (i.e. rabies lyssavirus, influenza A virus, West Nile virus and norovirus) from multiple host species (mammals: mouse, pig, ferret; birds: turkey, pigeon, collared dove) and diverse tissue matrices. Notably, both kits are designed to deplete human, mouse, and rat rRNA, and their performance on non-target species has not been previously evaluated.

Both depletion strategies were applied using commercial kits (https://emea.illumina.com/products/by-type/sequencing-kits/library-prep-kits/stranded-total-rna-prep.html; https://www.jumpcodegenomics.com/products/crisprclean-plus-rna-prep) prior to sequencing on an Illumina MiSeq (Figure 1) and were evaluated against an un-depleted shotgun approach. In the first experimental round (round I), we analyzed samples with high viral loads and good RNA integrity (RIN > 4.6), testing two input amounts (10 ng and 100 ng) for each workflow. In addition, the turkey was selected for evaluating rRNA depletion on a non-target species. Sequencing depth was sufficient to generate the consensus genomes of the target virus, which were used to compare the two protocols. In the second round (round II), the depletion workflows were challenged with highly degraded samples (RIN 2.0–3.2) from non-target host species, collected from multiple tissues/matrices and containing low viral loads. These samples were processed using 100 ng of total RNA and sequenced at a depth sufficient only to assess viral enrichment. Reconstructing full viral consensus sequences was not the goal at this stage.
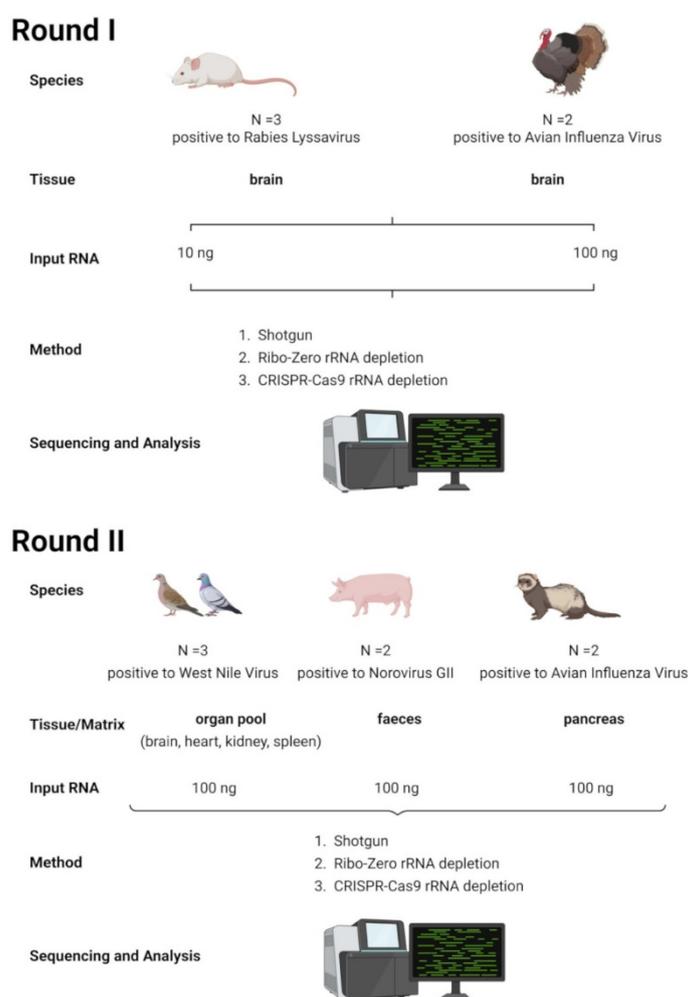

Figure 1. Experimental design of the study. In Round I, brain tissues from mouse and turkey were chosen as the target and non-target species for the rRNA depletion kits, respectively. Round II extended the evaluation to additional mammalian and avian species and to lower-quality tissue types and sample matrices compared with Round I.

For round I of analyses, total RNA was extracted from brains of mouse (*Mus musculus*, N = 3) and turkey (*Meleagris gallopavo*, N = 2) with known viral loads of rabies lyssavirus and avian influenza virus, respectively. For round II, total

RNA was extracted from organ pools (brain, heart, spleen, kidney) of wild avian species (*Columba palumbus*, N = 2, *Streptopelia decaocto* N = 1), pig (*Sus domesticus*, N = 2) faeces and ferret (*Mustela furo*, N = 2) pancreas. All samples had previously tested positive by quantitative real-time PCR for West Nile virus, norovirus, and avian influenza virus, respectively. RNA extraction was performed with NucleoSpin RNA Mini kit for RNA purification (MACHEREY-NAGEL GmbH & Co, Germany). RNA integrity number (RIN) of the samples was determined using the Agilent RNA 6000 Nano Kit on the 2100 Bioanalyzer Instrument (Agilent, CA, USA). In the shotgun approach, cDNA was synthesised using the NEBNext® Ultra™ II RNA First and Second Strand Synthesis Modules (New England Biolabs, MA, USA) and libraries were produced using the Illumina DNA Prep kit (Illumina, CA, USA) according to the manufacturer's instructions. Ribosomal RNA depletion was performed using the CRISPRclean™ Plus Stranded Total RNA Prep with rRNA Depletion-HMRPB (Jumcode Genomics Inc., CA, USA) and the Illumina Stranded Total RNA Prep with Ribo-Zero Plus (Illumina, CA, USA) according to the manufacturer's instructions. Sequencing was performed in 300 bp paired-end mode on a MiSeq instrument, assigning approximately 2 million reads per sample.

Raw data were filtered by removing: (a) reads with more than 10% of undetermined ("N") bases; (b) reads with more than 100 bases with a Q score below 7; (c) duplicated paired-end reads. The remaining reads were clipped from Illumina adaptors with Scythe v0.991 (https://github.com/vsbuffalo/scythe) and trimmed with Sickle v1.33 (https://github.com/najoshi/sickle). Reads shorter than 80 bases or unpaired after previous filters were discarded. For rRNA content estimation, we aligned high-quality reads using Bowtie2 v2.5.4 (Langmead et al., 2009, 2019; Langmead and Salzberg, 2012) with standard parameters against a database built from the rRNA sequences collected from the NCBI nucleotide database using the following query: "biomol_rrna[PROP]" (as of August 13, 2025). Paired-end reads in which at least one read aligned were classified as ribosomal, while the remaining reads were extracted for metagenomic analysis. Taxonomic assignment of high-quality rRNA-free reads was carried out using Kraken2 v2.1.5 (Wood et al., 2019) with standard parameters against "core_nt" database (as of June 09, 2025). For computing target viral read fraction and reference-based assembly, for each sample we selected and extracted all reads classified as belonging to the target virus (as reported in Table I) at the taxonomical level of species. High-quality rRNA-free viral reads were aligned against the corresponding reference genome (Table I) using BWA v0.7.12 (Li and Durbin, 2010) and standard parameters. Alignments were processed with SAMtools v1.6 (Li et al., 2009) to convert them in BAM format and sort them by position. SNPs were called using LoFreq v2.1.2 (Wilm et al., 2012).

Sample	Round	Source	Host	RIN	Virus	Accession Number (NCBI GenBank or GISAID)	Viral Titer/Ct
24RS414	I	brain	*Mus musculus*	7,0	Rabies lyssavirus	OQ787037.1	1.88E+06 gc/µL
24RS415	I	brain	*Mus musculus*	6,9	Rabies lyssavirus	OQ787037.1	1.07E+08 gc/µL
24RS416	I	brain	*Mus musculus*	4,7	Rabies lyssavirus	OQ787037.1	1.72E+08 gc/µL
23RS2515-9	I	brain	*Meleagris gallopavo*	6,4	Influenza A virus	EPI_ISL_18513244	3.51E+08 gc/µL
23RS2515-51	I	brain	*Meleagris gallopavo*	6,1	Influenza A virus	EPI_ISL_18513244	7.94E+07 gc/µL
23VIR8019-3	II	brain, heart, spleen, kidney (organ pool)	*Columba palumbus*	2,0	West Nile virus 1	PX315796	23.6
23VIR7607-3	II	brain, heart, spleen, kidney (organ pool)	*Streptopelia decaocto*	2,3	West Nile virus 1	PX315795	26.59
23VIR8704-3	II	brain, heart, spleen, kidney (organ pool)	*Columba palumbus*	2,3	West Nile virus 2	PX315797	26.96
23RS1211-8	II	faeces	*Sus scrofa domesticus*	2,4	Norovirus - GII	PV806862	26.8
23RS1878-1	II	faeces	*Sus scrofa domesticus*	3,2	Norovirus - GII	PV769083	22.88
24VIR10444-75	II	pancreas	*Mustela furo*	2,5	Influenza A virus	EPI_ISL_19767156	2.96E+05 gc/µL
24VIR10444-85	II	pancreas	*Mustela furo*	2,5	Influenza A virus	EPI_ISL_19767156	2.39E+05 gc/µL

Table I. Metadata information for the samples investigated in the present study.

According to LoFreq usage recommendations, the alignment was first processed with Picard-tools v2.1.0 (http://broadinstitute.github.io/picard/) and GATK v3.5 (McKenna et al., 2010) in order to correct potential errors, realign reads around indels and recalibrate base quality. LoFreq was then run on fixed alignment with option "--call-indels" to produce a vcf file containing both SNPs and indels. From the final set of variants, indels with a frequency lower than 50% and SNPs with a frequency lower than 25% were discarded. To produce the consensus sequence, we changed the reference genome in agreement with the following rules: (a) for a position j, if coverage was not enough (<10X) to reliably call variants, we added an "N" base; (b) for a position j, if coverage was enough (>=10X) to reliably call variants but no SNP were called, we added a reference genome base at position j; (c) for a position j, if coverage was enough (>=10X) to reliably call variants and at least one SNP were called, we added the nucleotide using the IUPAC nucleotide code (http://www.bioinformatics.org/sms/iupac.html) according to the bases present. Consensus sequences belonging to the same sample were compared using Mafft v7.526 (Katoh and Standley, 2013) and "--auto" parameter. Degenerated bases at corresponding positions were treated as identical if at least one of the nucleotides

composing the degenerated code matched between consensus sequences.

In Round I, samples processed without any enrichment strategy showed, on average, 84.0% of reads classified as rRNA (Figure 2, round I), ranging from 74.3% to 90.2%. In contrast, samples treated with the CRISPR-Cas9 depletion kit exhibited a marked reduction in rRNA content, with an average of 22.5% of reads (ranging from 7.3% to 37.7%), demonstrating the effectiveness of this approach in removing the majority of ribosomal RNA. The most efficient rRNA removal, however, was achieved with the Ribo-Zero depletion kit, where the average rRNA fraction dropped to 1.1%, with values ranging between 0.2% and 2.4%. Regarding target viral reads, both depletion strategies enhanced viral RNA recovery compared to sequencing without any depletion (Figure 3). On average, the viral read fraction reached 9.98% and 16.02% for CRISPR-Cas9– and Ribo-Zero–treated samples, respectively, while untreated samples showed only 2.21%. The Ribo-Zero method recovered, on average, 14.4 times more viral reads than untreated samples, whereas CRISPR-Cas9–based depletion achieved an 8.2-fold enrichment (Figure 4, round I). Furthermore, differences in input material during Round I had no significant effect on rRNA depletion efficiency or viral read recovery. Since the primary objective of Round I sequencing was to generate sufficient viral data for downstream analyses, we also compared the two depletion methods based on their capacity to assemble high-quality consensus sequences of the target virus, characterized by uniform and complete coverage. Both methods successfully achieved this outcome (Table II): all consensus sequences generated corresponded to complete coding genomes. Moreover, consensus sequences derived from both depletion-treated and untreated samples were 100% identical within coding regions at positions with a coverage greater than 10×.
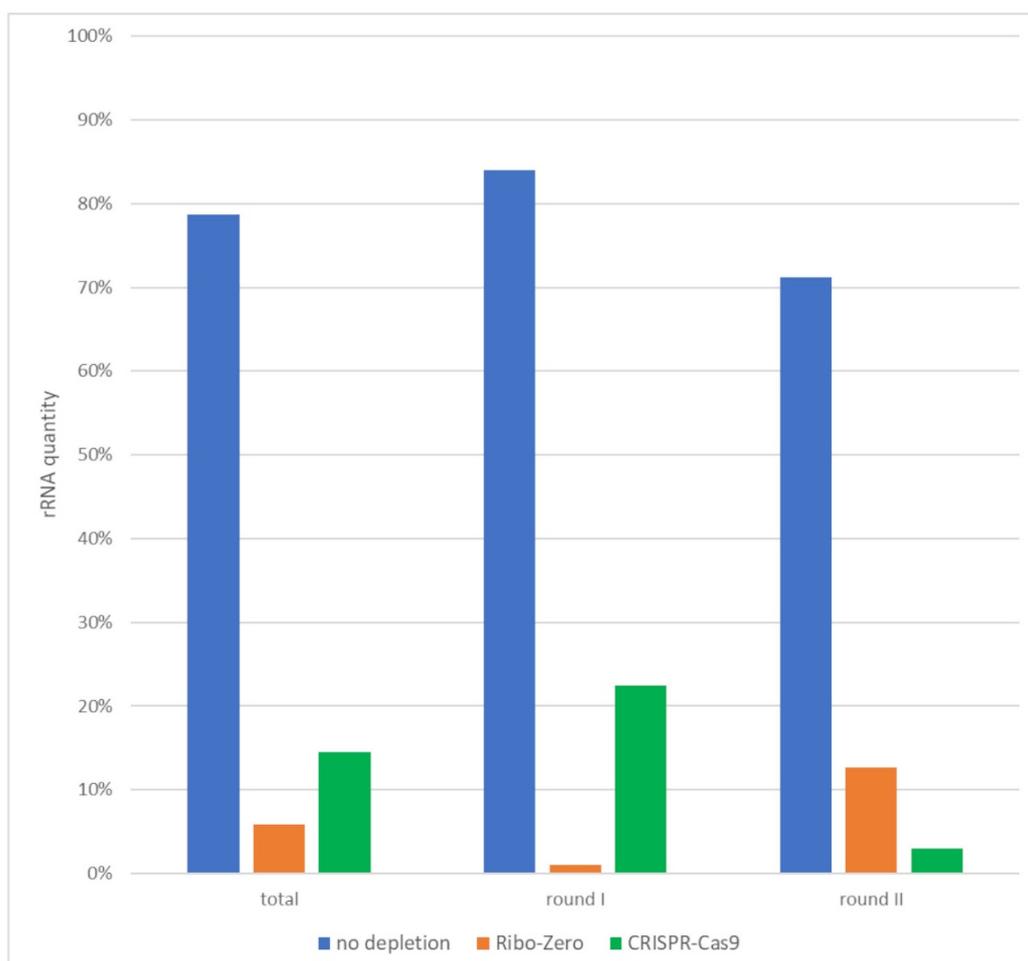

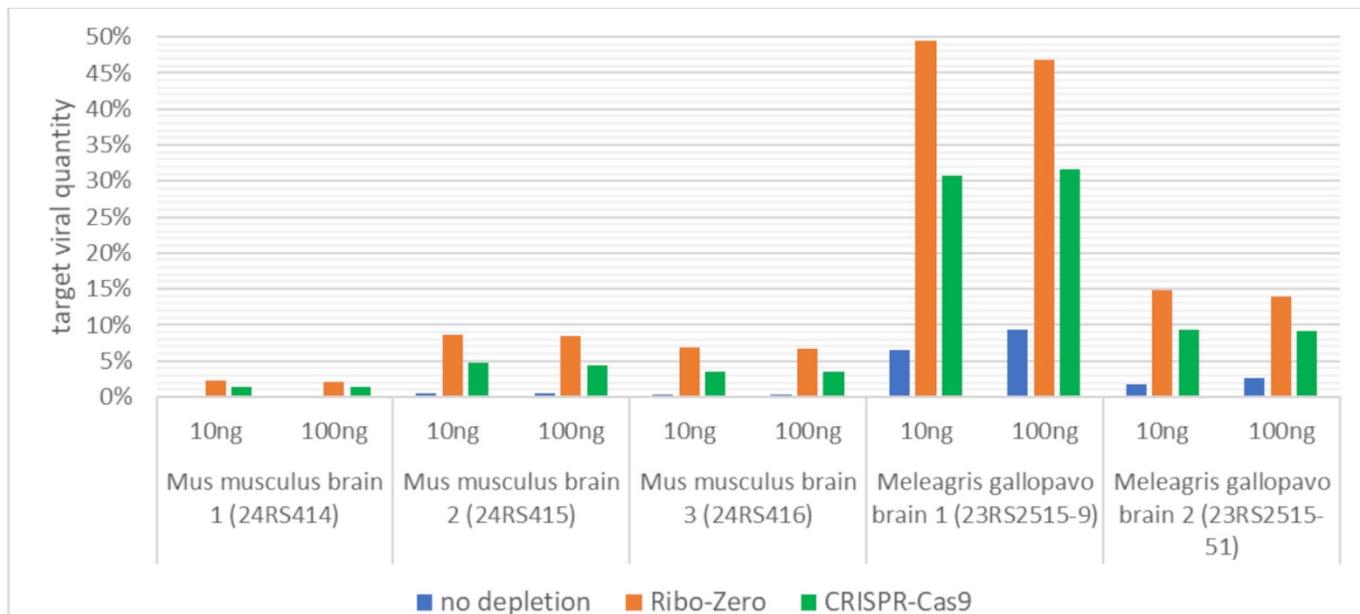**Figure** 2. Average rRNA content calculated per round and across all samples.

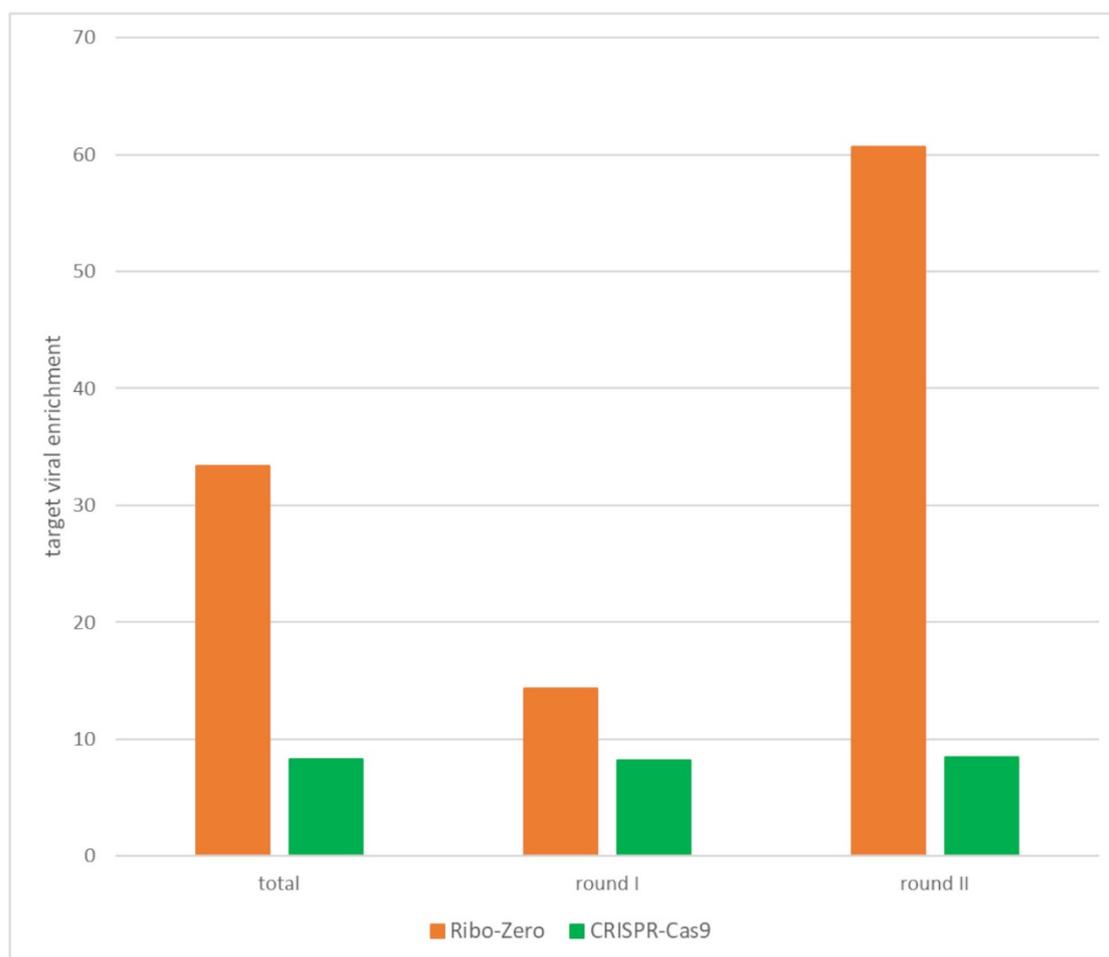**Figure** 3. Target viral content detected in samples analyzed in Round I.

Figure 4. Average enrichment of target viral content calculated per round and across all samples.

Sample	Virus	Quantity	Approach	Number of raw reads	Horizontal coverage
24RS414	Rabies lyssavirus	10 ng	no depletion	3,081,341	99.46% of the genome with at least 10X coverage
			Ribo-Zero	1,056,845	99.99% of the genome with at least 10X coverage
			Crispr-Cas9	2,230,715	100% of the genome with at least 10X coverage
		100 ng	no depletion	2,217,043	99.24% of the genome with at least 10X coverage
			Ribo-Zero	1,937,606	100% of the genome with at least 10X coverage
			Crispr-Cas9	2,368,959	99.93% of the genome with at least 10X coverage
24RS415	Rabies lyssavirus	10 ng	no depletion	2,122,222	100% of the genome with at least 10X coverage
			Ribo-Zero	1,836,139	100% of the genome with at least 10X coverage
			Crispr-Cas9	2,359,930	100% of the genome with at least 10X coverage
		100 ng	no depletion	1,505,722	99.88% of the genome with at least 10X coverage
			Ribo-Zero	2,338,387	100% of the genome with at least 10X coverage
			Crispr-Cas9	1,712,099	100% of the genome with at least 10X coverage
24RS416	Rabies lyssavirus	10 ng	no depletion	2,652,675	99.88% of the genome with at least 10X coverage
			Ribo-Zero	1,017,607	100% of the genome with at least 10X coverage
			Crispr-Cas9	2,105,776	99.93% of the genome with at least 10X coverage
		100 ng	no depletion	1,844,419	99.61% of the genome with at least 10X coverage
			Ribo-Zero	1,951,316	100% of the genome with at least 10X coverage
			Crispr-Cas9	1,760,423	99.93% of the genome with at least 10X coverage
23RS2515-9	Influenza A virus	10 ng	no depletion	1,791,950	100% of the genome* with at least 10X coverage
			Ribo-Zero	2,120,493	100% of the genome* with at least 10X coverage
			Crispr-Cas9	2,766,464	100% of the genome* with at least 10X coverage
		100 ng	no depletion	1,548,866	100% of the genome* with at least 10X coverage
			Ribo-Zero	2,152,895	100% of the genome* with at least 10X coverage
			Crispr-Cas9	267,112	100% of the genome* with at least 10X coverage
23RS2515-51	Influenza A virus	10 ng	no depletion	959,936	99.98% of the genome* with at least 10X coverage
			Ribo-Zero	2,316,845	100% of the genome* with at least 10X coverage
			Crispr-Cas9	3,034,076	100% of the genome* with at least 10X coverage
		100 ng	no depletion	2,128,897	99.97% of the genome* with at least 10X coverage
			Ribo-Zero	2,215,575	100% of the genome* with at least 10X coverage
			Crispr-Cas9	3,072,937	100% of the genome* with at least 10X coverage

Table II. Horizontal coverage of viral consensus sequences in Round I samples. *For segmented viruses, such as influenza A virus, the reported genome coverage represents the combined coverage across all genomic segments composing the viral genome.

Assessment of rRNA removal efficiency in samples selected for Round II revealed that the CRISPR-Cas9 depletion kit reduced rRNA content to an average of 3.1%, whereas the Ribo-Zero approach yielded an average of 12.7% ribosomal material, compared with 71.2% in untreated samples (Figure 2, Round II). However, the greater rRNA depletion achieved with the CRISPR-Cas9 method compared to the Ribo-Zero one was not reflected in viral content metrics, the most relevant measure of epidemiological utility, defined here as the fraction of reads taxonomically classified as belonging to the viruses of interest (West Nile virus, norovirus, and influenza A virus) at species level. Consistent with observations from Round I, both depletion strategies increased viral RNA recovery compared to sequencing without any depletion (Figure 5). On average, the target viral read fraction was 0.08% for CRISPR-Cas9–treated samples and 1.8% for Ribo-Zero–treated samples, while untreated samples showed only 0.01%. Across all rounds, the fraction of viral reads was consistently higher in Ribo-Zero–treated samples than in those processed with CRISPR-Cas9, indicating that Ribo-Zero ultimately provides more virologically relevant data at equivalent sequencing depth, despite its comparatively lower rRNA removal efficiency, particularly in Round II experiments. The Ribo-Zero depletion approach enriched viral reads by 60.7-fold in Round II compared to untreated samples, whereas CRISPR-Cas9 achieved an 8.5-fold increase (Figure 4, Round II). Together with Round I data, these results underscore the superior performance and cost-effectiveness of the RNase H–based Ribo-Zero method, which generated up to sevenfold more viral reads than the CRISPR-Cas9 approach at comparable sequencing depth.
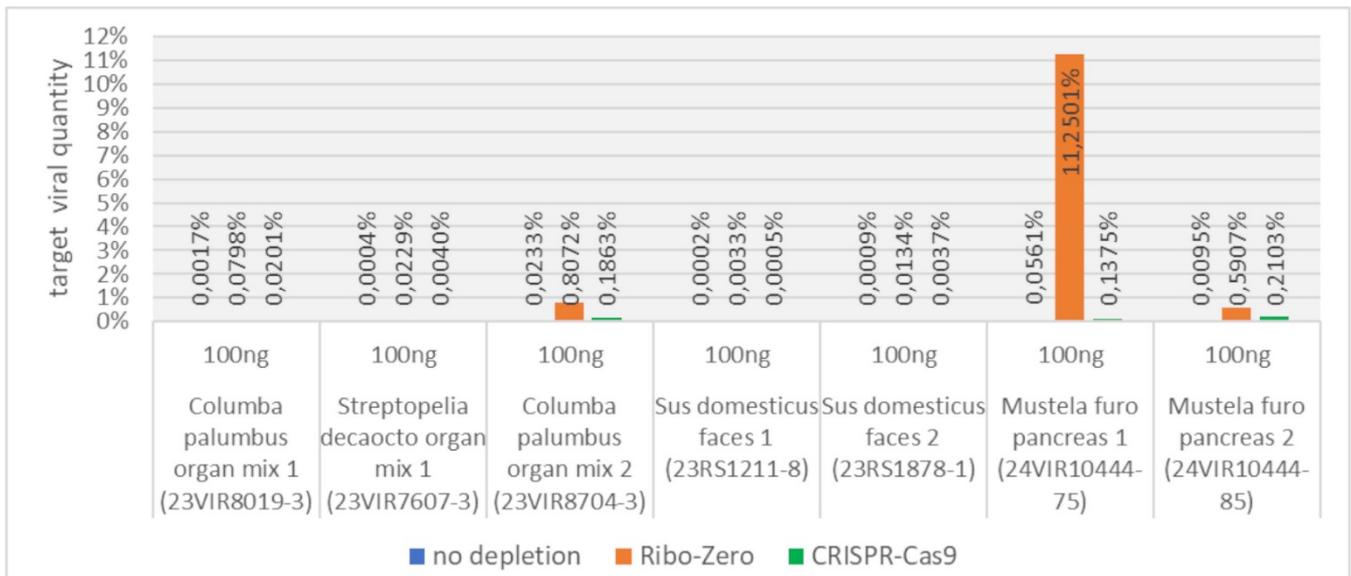
Figure 5. Target viral content detected in samples analyzed in Round II.

Finally, we compared hands-on time, overall turnaround time and per-sample costs for the three workflows (Table III). Ribo-Zero and the no-depletion workflow showed comparable overall turnaround times (8:30), whereas the CRISPR–Cas9 workflow required a longer turnaround time (10:30). Compared with CRISPR–Cas9, Ribo-Zero was slightly faster and less expensive (€400 vs €450 per sample).

Approach	Hands-on time (hrs:min)	Overall turnaround time(hrs:min)	Costs per sample(€) *
no depletion	2:30	8:30	370
Ribo-Zero	3:35	8:30	400
Crispr-Cas9	4:00	10:30	450

Table III. Time and cost assessment of the evaluated approaches tested in the study. *Cost estimates are based on 2 million reads per sample (N = 24) using two MiSeq Reagent Kits v3.

Owing to its inherent molecular flexibility, CRISPR-Cas9 is a highly adaptable technology for a range of molecular applications. In this study, we employed it to deplete host- and bacterial rRNA from viral samples in order to enhance metagenomic sequencing for the detection and characterization of RNA viruses in diverse samples of animal origin. Our results show that CRISPR-Cas9 efficiently removes targeted sequences, such as rRNA, in shotgun metagenomics applications, although its overall depletion efficiency remains lower than that of the Ribo-Zero kit. Importantly, genome reconstruction was unaffected and remained unbiased, producing results comparable to those obtained with both the enzymatic depletion method and unenriched shotgun sequencing. At present, cost is a limiting factor for the CRISPR-Cas9 depletion method. This approach requires greater sequencing depth to achieve results similar to those of Ribo-Zero, and the CRISPRclean kit itself is approximately 10% more expensive than the Ribo-Zero kit used in this study. A potential advantage of the CRISPR-Cas9 depletion method is the timing of the depletion step, which occurs after library preparation, in contrast to the Ribo-Zero protocol that acts directly on the input RNA. This distinction may be particularly relevant for low-input samples, as performing rRNA depletion post-library preparation may help to maximize library yield. However, CRISPR-Cas9 depletion is generally more sensitive to RNA quality than the Ribo-Zero kit, particularly at low viral loads. This aspect was deliberately evaluated in Round II, which included samples with degraded material- a common scenario in veterinary diagnostics, where carcasses of wildlife species at various stages of decomposition, collected for passive surveillance, are frequently submitted to the laboratory. It is therefore likely that low sample quality contributed to the reduced viral read recovery observed with the CRISPRclean kit in Round II, despite efficient rRNA removal. Viral read percentages were consistent with the viral loads. Among the Influenza A virus–positive samples, the viral copy numbers in Round I were 100—1000 fold higher than those in Round II (Table I), a difference that was reflected in the NGS results. The correlation ($R^2 = 0.997$) between viral copy number and viral read percentage confirmed the reliability of NGS data for quantitative assessment. Overall, enrichment of target viral reads—critical for veterinary applications—demonstrated that Ribo-Zero outperformed

CRISPR-Cas9 across all tested conditions, particularly under challenging scenarios such as those in Round II.

In conclusion, CRISPR-Cas9–based depletion shows promising potential for viral metagenomic sequencing, although it currently remains less efficient and more costly than enzymatic methods. To our knowledge, this is the first study to evaluate its performance in rRNA-rich tissues, extending previous work largely limited to low-rRNA respiratory human samples (Cerón et al., 2023; Chan et al., 2023; Gu et al., 2016). Improvements in efficiency and specificity for rRNA depletion in veterinary-relevant species, along with reduced cost and improved tolerance to degraded samples, could establish CRISPR-Cas9 as a powerful tool for untargeted surveillance of animal pathogens.

Acknowledgments

Ethical approval

Brain samples employed in Round I derived from mice treated following animal experimental procedures conducted in strict accordance with Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. The procedures were authorized by the Italian Ministry of Health (Decrees 505/2015-PR and 491/2020-PR) before experiments were initiated and carefully evaluated and endorsed by the IZSVe Ethics Committee.

Pancreas samples employed in Round II derived from ferrets treated following animal experimental procedures conducted in strict accordance with the Decree of the Italian Ministry of Health n. 26 of 4 March 2014 on the protection of animals used for scientific purposes, implementing Directive 2010/63/EU, and approved by the Institute's Ethic Committee (protocol n. 1/19 obtained on the 4/02/2019). Animal experiments were approved by the Ministry of Health (733/2020-PR, further amended by 16526-23/07/2020-DGSAF-MDS-P).

Author Contributions

Conceptualization: IM; Methodology: EP, GZ; Formal analysis: EP, GZ, MC, SM; Investigation: EP, GZ; Writing original draft preparation: IM, EP, GZ; Writing, review and editing: IM, EP, GZ, CM, SM, AF; Visualization: EP, GZ; Supervision: IM; Project administration: IM; Funding acquisition: IM. All authors have read and agreed to the published version of the manuscript.

Data availability

The datasets generated for this study can be found in the Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1362999).

Fundings

References

Babiker, A., Bradley, H., Stittleburg, V., Key, A., Kraft, C. S., Waggoner, J., & Piantadosi, A. (2020). Metagenomic sequencing to detect respiratory viruses in persons under investigation for COVID-19. medRxiv : the preprint server for health sciences, 2020.09.09.20178764. https://doi.org/10.1101/2020.09.09.20178764.

Cerón, S., Clemons, N. C., von Bredow, B., & Yang, S. (2023). Application of CRISPR-based human and bacterial ribosomal RNA depletion for SARS-CoV-2 shotgun metagenomic sequencing. American Journal of Clinical Pathology, 159, 111–115. https://doi.org/10.1093/ajcp/aqac137.

Chan, A. P., Siddique, A., Desplat, Y., Choi, Y., Ranganathan, S., & Choudhary, K. S. (2023). A CRISPR-enhanced metagenomic NGS test to improve pandemic preparedness. Cell Reports Methods, 3, 100463. https://doi.org/10.1016/j.crmeth.2023.100463.

Fischer, N., Indenbirken, D., Meyer, T., Lütgehetmann, M., Lellek, H., & Spohn, M. (2015). Evaluation of unbiased next-generation sequencing of RNA (RNA-seq) as a diagnostic method in influenza virus-positive respiratory samples. Journal of Clinical Microbiology, 53, 2238–2250. https://doi.org/10.1128/JCM.02495-14.

Gu, W., Crawford, E. D., O'Donovan, B. D., Wilson, M. R., Chow, E. D., & Retallack, H. (2016). Depletion of abundant sequences by hybridization (DASH): Using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biology, 17, 41. https://doi.org/10.1186/s13059-016-0904-5.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution, 30, 772–780. https://doi.org/10.1093/molbev/mst010.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9, 357–359. https://doi.org/10.1038/nmeth.1923.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology, 10, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Langmead, B., Wilks, C., Antonescu, V., & Charles, R. (2019). Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics, 35, 421–432. https://doi.org/10.1093/bioinformatics/bty648.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics, 26, 589–595. https://doi.org/10.1093/bioinformatics/btp698.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics, 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., & Daly, M. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20, 1297–1303. https://doi.org/10.1101/gr.107524.110.

Rajagopala, S. V., Bakhoum, N. G., Pakala, S. B., Shilts, M. H., Rosas-Salazar, C., & Mai, A. (2021). Metatranscriptomics to characterize respiratory virome, microbiome, and host response directly from clinical samples. Cell Reports Methods, 1, 100091. https://doi.org/10.1016/j.crmeth.2021.100091.

Schlaberg, R., Chiu, C. Y., Miller, S., Procop, G. W., & Weinstock, G. (2017). Validation of metagenomic next-generation sequencing tests for universal pathogen detection. Archives of Pathology & Laboratory Medicine, 141, 776–786. https://doi.org/10.5858/arpa.2016-0539-RA.

The White House. (2021). American pandemic preparedness: Transforming our capabilities (Report). https://www.whitehouse.gov/wp-content/uploads/2021/09/American-Pandemic-Preparedness-Transforming-Our-Capabilities-Final-For-Web.pdf.

Wilm, A., Aw, P. P. K., Bertrand, D., Yeo, G. H. T., Ong, S. H., & Wong, C. H. (2012). LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Research, 40, 11189–11201. https://doi.org/10.1093/nar/gks918.

Wilson, M. R., Sample, H. A., Zorn, K. C., Arevalo, S., Yu, G., & Neuhaus, J. (2019). Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. New England Journal of Medicine, 380, 2327–2340. https://doi.org/10.1056/NEJMoa1803396.

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biology, 20, 257. https://doi.org/10.1186/s13059-019-1891-0.