bioRxiv preprint doi: https://doi.org/10.1101/644591; this version posted May 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Next-generation sequencing of double stranded RNA is greatly improved by treatment with the inexpensive denaturing reagent DMSO

Alexander H. Wilcox[1], Eric Delwart[2,3], and Samuel L. Díaz Muñoz[1,4#]

[1]Department of Microbiology and Molecular Genetics, University of California, Davis
One Shields Avenue
Davis, CA 95616, USA

[2]Vitalant Research Institute, San Francisco, USA

[3]Department of Laboratory Medicine, University of California San Francisco, San Francisco, USA

[4]Genome Center, University of California, Davis
One Shields Ave
Davis CA 95616 USA

#Address correspondence to: Samuel L. Díaz Muñoz, samdiazmunoz@ucdavis.edu

Abstract

Double stranded RNA (dsRNA) is the genetic material of important viruses and a key component of RNA interference-based immunity in eukaryotes. Previous studies have noted difficulties in determining the sequence of dsRNA molecules that have affected studies of immune function and estimates of viral diversity in nature. Dimethyl sulfoxide (DMSO) has been used to denature dsRNA prior to the reverse transcription stage to improve RT-PCR and Sanger sequencing. We systematically tested the utility of DMSO to improve sequencing yield of a dsRNA virus (Φ6) in a short-read next generation sequencing platform. DMSO treatment improved sequencing read recovery by over two orders of magnitude, even when RNA and cDNA concentrations were below the limit of detection. We also tested the effects of DMSO on a mock eukaryotic viral community and found that dsRNA virus reads increased with DMSO treatment. Furthermore, we provide evidence that DMSO treatment does not adversely affect recovery of reads from a single-stranded RNA viral genome (Influenza A/California/07/2009). We suggest that up to 50% DMSO treatment be used prior to cDNA synthesis when samples of interest are composed of or may contain dsRNA.

Data Summary

Sequence data was deposited in the NCBI Short Read Archive (accession numbers: PRJNA527100, PRJNA527101, PRJNA527098). Data and code for analysis is available on GitHub

bioRxiv preprint doi: https://doi.org/10.1101/644591; this version posted May 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(https://github.com/awilcox83/dsRNA-sequencing/, doi:10.5281/zenodo.1453423). Protocol for dsRNA sequencing is posted on protocols.io (doi:10.17504/protocols.io.ugnetve).

Introduction

Ribonucleic acid (RNA) is a ubiquitous biological molecule involved in transcription and translation, which also serves as the genetic material of a large number of important viruses. The double-stranded form of RNA (dsRNA) is believed to be less abundant in nature but is a crucial component of a number of biological systems. It has a central role in the RNA interference system [1], which modulates innate immunity in plants and animals. Moreover, it also serves as the genetic material of a number of virus lineages (including the families *Reoviridae*, *Cystoviridae*, *Picobirnaviridae*) that infect humans, animals, plants, fungi, and bacteria; which play important medical, ecological, and scientific roles. A number of dsRNA viruses are of clinical and agricultural significance, such as Bluetongue virus, which causes high morbidity and mortality in ruminants [2], and rotavirus, which causes acute gastroenteritis in humans [3]. There are indications that the overall diversity of RNA viruses may be underestimated [4, 5] and difficulties sequencing dsRNA in particular have been noted in the literature [6,7]. As a consequence, dsRNA virus lineages may be underrepresented and dsRNAs involved in immunity may be underestimated.

Theory and implementation

The extent of microbial diversity has been revealed by powerful whole genome sequencing tools that are quickly becoming standard tools in biology. However, next-generation sequencing has known biases according to the nucleic acid composition [8]. A major limitation is that most sequencing platforms cannot sequence RNA directly, requiring that it is first reverse-transcribed into its complementary DNA (cDNA). cDNA synthesis is typically achieved by hybridising oligonucleotide DNA primers to the RNA and using a reverse transcriptase to synthesise the remainder of the complementary DNA strand. This step poses a particular problem to dsRNA because the presence of a complementary strand blocks the ability of these primers to bind. The blocking has a direct effect on the amount of dsRNA converted to cDNA, resulting in many fewer sequencing reads being generated relative to the true amount of RNA present. Additionally, many dsRNA viruses have small genomes that limit the amount of RNA that can be extracted, further complicating the determination of dsRNA sequences.

bioRxiv preprint doi: https://doi.org/10.1101/644591; this version posted May 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

As early as 1968, dimethyl sulfoxide (DMSO) has been shown to have a denaturing effect on nucleic acids [9]. DMSO has been successfully used to improve the performance of RT-PCR [10] and Sanger Sequencing [11]. However, to our knowledge DMSO has not been used for next-generation sequencing approaches and many dsRNA sequencing studies omit DMSO treatment [6, 12, 13, 14]. Moreover, there is no standard protocol for DMSO treatment of samples and previous methods vary greatly in their conditions, particularly DMSO concentration, which has ranged from 15% [11] to 90% [15].

This paper investigated four questions regarding the effect of DMSO treatment on next-generation sequencing: 1) Does DMSO treatment improve recovery of dsRNA reads and at what concentration? 2) Does DMSO affect read coverage of a viral genome? 3) Is the effect of DMSO independent of RNA concentration? 4) Does DMSO treatment negatively affect the recovery of single-stranded RNA (ssRNA) genomes? Our results suggest that treatment with a high concentration of DMSO greatly increases the number of reads generated when sequencing dsRNA, without adversely affecting sequencing of ssRNA virus genomes on the Illumina short read sequencing platform.

We carried out the methods in this paper on two different viruses: *Pseudomonas* phage Φ6, which has a dsRNA genome made up of three segments, and human influenza virus A H1N1, which contains eight ssRNA segments. We also used a mock eukaryotic viral community, manufactured by the National Institute for Biological Standards and Control (NIBSC, UK) as a reference material for multiplex viral detection (NIBSC reagent 11/242-001). This reagent was expected to contain 25 human pathogenic viruses and has been used to investigate viral detection methods on mixed and metagenomic samples [16].

Sample Preparation

Φ6 lysate was prepared by plating phage and its *Pseudomonas syringae* host using double agar overlay. Phages were harvested by selecting a plate with semi-confluent lysis, transferring the soft agar layer to 3ml LB (Lennox) media, and centrifuging to remove host cells and agar. An influenza virus lysate was generated from egg-passaged stock of influenza A/California/07/2009(H1NI) (generously provided by Ted M. Ross, University of Georgia), which was expanded by passage at a low multiplicity of infection in Madin-Darby canine kidney (MDCK) cells in culture. Viral lysates and the mock viral community (NIBSC reagent 11/242-001) were passed through a 0.22µm filter to remove debris and contaminants. 1ml of each filtrate was treated with 25µl DNAse I (Thermo

bioRxiv preprint doi: https://doi.org/10.1101/644591; this version posted May 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Scientific) and 50µl RNAse A/T1 mix (Thermo Scientific) with 1X DNAse I Buffer (Thermo Scientific) at 37ºC for 1 hour 30 minutes to degrade extracapsular nucleic acids. Viral RNA was extracted using an RNeasy Mini Kit (Qiagen), passing a total of 900µl of nuclease-treated lysate through a column, and eluting into 100µl of elution buffer.

DMSO Treatment, reverse transcription and sequencing

Viral RNA samples were divided into 20µl aliquots. DMSO was added to concentrations (v/v) of 15%, 50% and 90% for each sample, followed by 1 hour and 30 minutes of incubation at 65ºC. DMSO was removed using a RNeasy MinElute Cleanup Kit (Qiagen, Valencia CA), following the manufacturer's instructions.

First strand cDNA synthesis was carried out by adding 5µl of each RNA sample (including a control that had not undergone DMSO treatment or column cleanup) to 1µl of random hexamer oligos, 1µl dNTPs and 3µl DEPC-treated water. Reactions were incubated at 65ºC for 5 minutes then placed on ice for 1 minute. 1X reverse transcriptase buffer, 5mM $McCl_2$, 0.01M DTT, 1µl RNAseOUT and 1µl Superscript III RT enzyme were added to each reaction for a total volume of 20µl. Reactions were incubated on a thermal cycler at 25ºC for 10 minutes, 50ºC for 50 minutes and 85ºC for 5 minutes. Second strand synthesis was carried out by adding 1µl dNTPs, 0.5µl DNA ligase, 2µl DNA polymerase I, 0.5µl RNAse H in 1X second strand synthesis buffer and made up to a total volume of 40µl with nuclease free water. Reactions were incubated at 16ºC for 5 hours and cDNA was purified with a Nucleospin Gel and PCR Clean-up kit (Macherey-Nagal, Düren Germany).

Libraries were prepared for Illumina sequencing using the Nextera XT DNA Sample Preparation Kit (Illumina, San Diego CA), with a 1/5 "scaled" library preparation protocol after Baym *et al*. [17].

Bioinformatics

Reads were trimmed for adapters and quality using Cutadapt [18] and Sickle [19]. Due to a short fragment size, reads overlapped and so paired-end libraries were merged into single-end libraries using PEAR [20]. For phi6 and influenza virus lysates, these libraries were mapped to the reference genomes using Bowtie2 [21], and bam-readcount [22] was used to determine the read depth at each position. Plots were generated using the ggplot2 package in R.

bioRxiv preprint doi: https://doi.org/10.1101/644591; this version posted May 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

For the mock viral community, a custom virus discovery pipeline was used to analyze sequencing reads [23]. Reads were translated and aligned to a viral proteome database (consisting of all annotated full or near full viral genomes) using BLASTx. The significant hits to the virus database were then aligned to a non-virus-non-redundant (NVNR) universal proteome database using BLASTx. Hits with more significant E-value to NVNR than to the virus database were removed.

Results and Discussion

Our findings show that DMSO treatment has a dramatic effect on dsRNA sequencing (Table 1). When we prepared viral lysates for sequencing without DMSO, we did not obtain sufficient reads to cover the entire Φ6 genome. Treatment with 15% DMSO increased the number of mapping reads over sevenfold, whilst the 50% and 90% treatments increased the number of reads by over two orders of magnitude, allowing the full genome to be sequenced at high coverage (average read depth of 1727 for 50% DMSO and 1493 for 90% DMSO).

Table 1. *The total number of reads generated during sequencing of influenza and Φ6 after treatment at varying DMSO concentrations, and the number of those reads that mapped to reference genomes.*

DMSO %	Total reads	Reads that map to phi6	Reads that map (%)
0	1551220	1373	0.1%
15	1314520	10320	0.8%
50	1077133	263641	24.5%
90	1393569	224191	16.1%

These increases in genome coverage occurred despite very low starting nucleic acid concentrations. We used a Qubit RNA High Sensitivity Assay kit to quantify RNA immediately after extraction; in all cases RNA was undetectable and so assumed to be under the kit's limit of detection of 5ng/µl. Additionally, we used a Qubit DNA High Sensitivity Assay kit to quantify the amount of cDNA synthesised. Despite this kit's lower limit of detection of 200pg/µl, DNA was still not detected. Thus, for dsRNA the raw quantity of starting material may not be as important as the efficiency of cDNA synthesis, a fact that should accounted for when preparing quality control thresholds before next generation sequencing.

Table 2. *The average read depth for each influenza segment under varying concentrations of DMSO.*

DMSO %	Total reads	Reads that map to flu	Reads that map (%)
0	1614418	931915	57.7%
15	1369886	353752	25.8%
50	1510801	295244	19.5%
90	1257059	446549	35.5%

bioRxiv preprint doi: https://doi.org/10.1101/644591; this version posted May 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We next sought to determine whether DMSO treatment affected other types of RNA present in the sample, which would occur in metagenomes, clinical samples, or transcriptomes. DMSO treatment did not appear to affect the recovery of ssRNA-derived reads from the influenza virus genome. There was no discernible effect when the DMSO concentration was varied (Table 1). The number of reads mapping to the reference genome did decrease from the 0% DMSO treatment to the 15% DMSO treatment. However, this loss in mapping reads was most likely caused by the extra column cleanup step required to remove the DMSO (note that the total reads also decreased). We note that this decrease in reads in the influenza virus sequencing was less than one order of magnitude and still resulted in extremely high read depth, with an average of over 1000x for every influenza segment after DMSO treatment (Table 2). This effect was obscured in Φ6 due to the large increase in mapping reads from DMSO treatment.

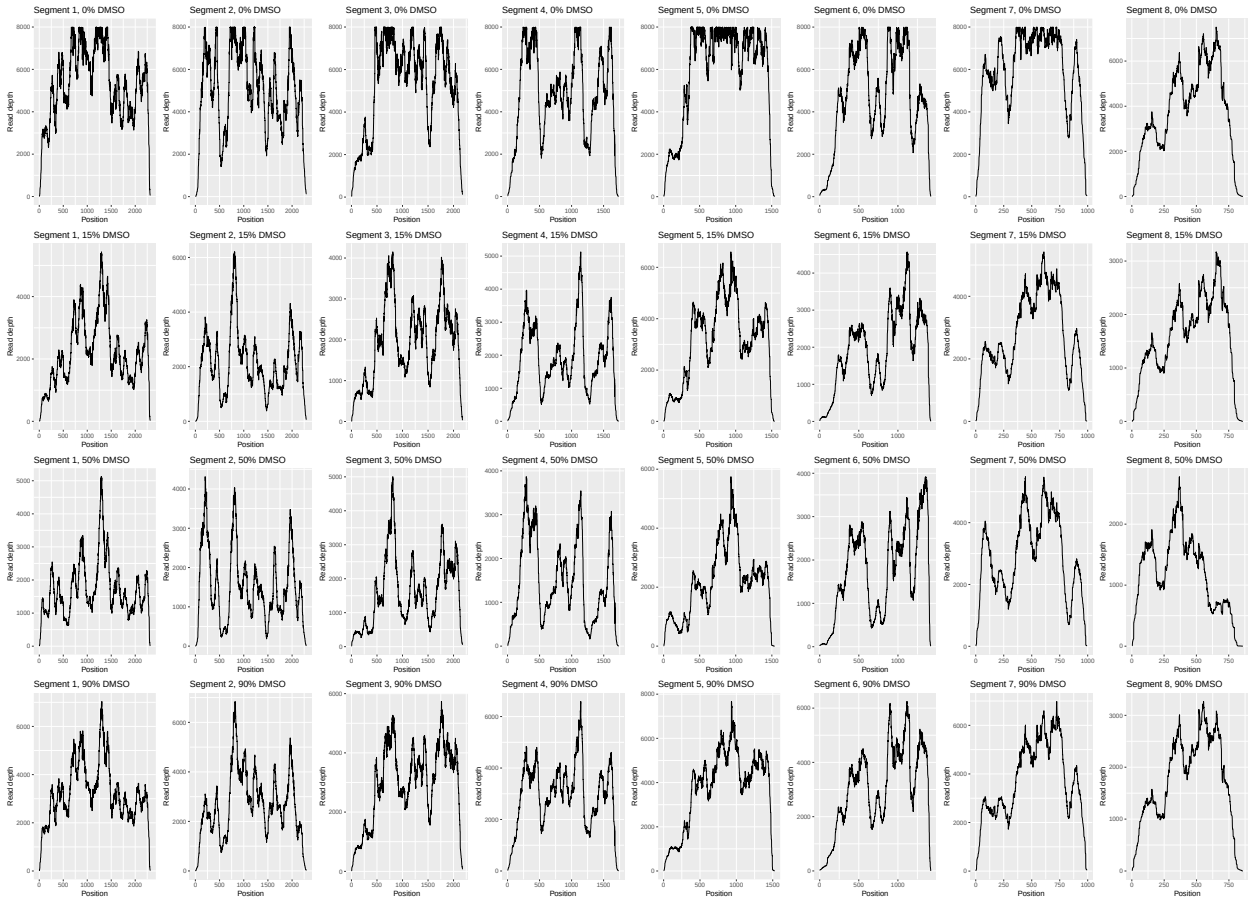

Figure 1. DMSO treatment does not affect sequencing read coverage across the ssRNA influenza genome. Read depth at each position in the influenza genome under varying concentrations of DMSO. Note that 8000 is the maximum read depth supported by the SAM/BAM file format, so some peaks in the 0% DMSO plots have been truncated.

bioRxiv preprint doi: https://doi.org/10.1101/644591; this version posted May 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In order to determine if the presence of DMSO affected any other properties of the RNA when used for downstream sequencing, we plotted the coverage at every nucleotide position for each DMSO concentration used. The plot for influenza (Figure 1) showed a distinct, repeating pattern for each segment, with DMSO concentration appearing to have no effect on relative coverage. This indicates that there is no bias in which reads are affected by the DMSO treatment, and that the reads generated are still representative. A similar pattern could be observed for the 50% and 90% DMSO treatments of Φ6 (Figure 2, the low number of reads made this pattern harder to discern in the 15% treatment). Therefore, DMSO treatment did not adversely affect genome-wide coverage patterns of dsRNA or ssRNA viruses.

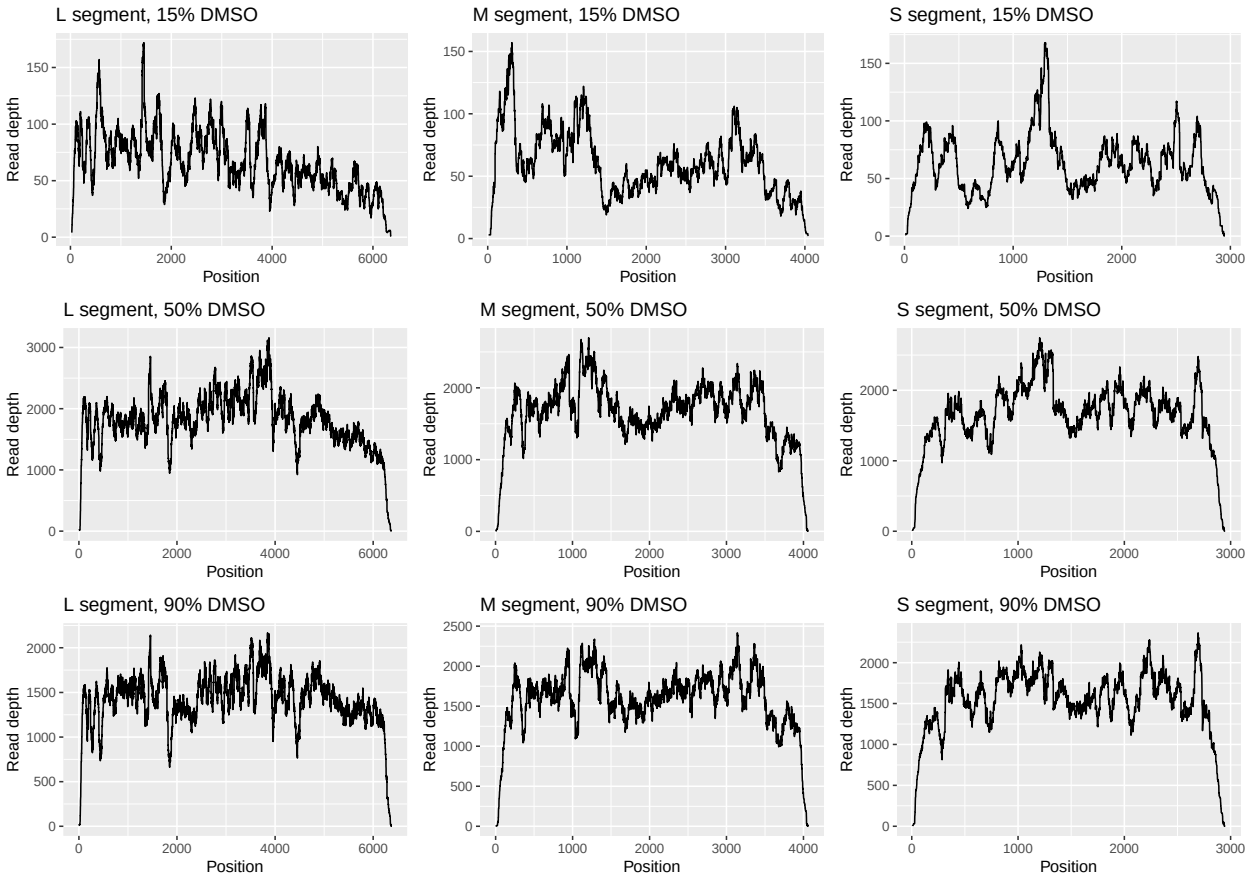

Figure 2. DMSO treatment greatly increases sequence coverage of the dsRNA Φ6 genome. Read depth at each position in the Φ6 genome under varying concentrations of DMSO. There were insufficient reads to generate a plot for 0% DMSO. Read depth at each position in the influenza genome under varying concentrations of DMSO. Note that 8000 is the maximum read depth supported by the SAM/BAM file format, so some peaks in the 0% DMSO plots have been truncated.

To determine if this method worked with higher starting concentrations of RNA, we used an Amicon centrifugal filter unit to concentrate approximately 10ml of Φ6 lysate into 50µl. The concentrated lysate contained 37.6ng/µl RNA (measured by Qubit) and was prepared for sequencing using 90%

bioRxiv preprint doi: https://doi.org/10.1101/644591; this version posted May 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

DMSO (as above), as well as a control without DMSO (heat-treated for 90 minutes at 65ºC). While only 0.08% of reads mapped to the reference genome in the non-DMSO treated control, there was an increase to 72.45% in the DMSO-treated sample, again demonstrating the importance of the DMSO treatment over raw RNA concentration (Figure 3). The concentrated sample had a much higher proportion of reads than the non-concentrated sample (16.1%, Table 1), most likely due to the concentration step increasing the ratio of viral RNA to extracapsular RNA. Therefore, DMSO treatment works at varying RNA concentrations and is likely to improve any dsRNA sequencing regardless of starting concentration.

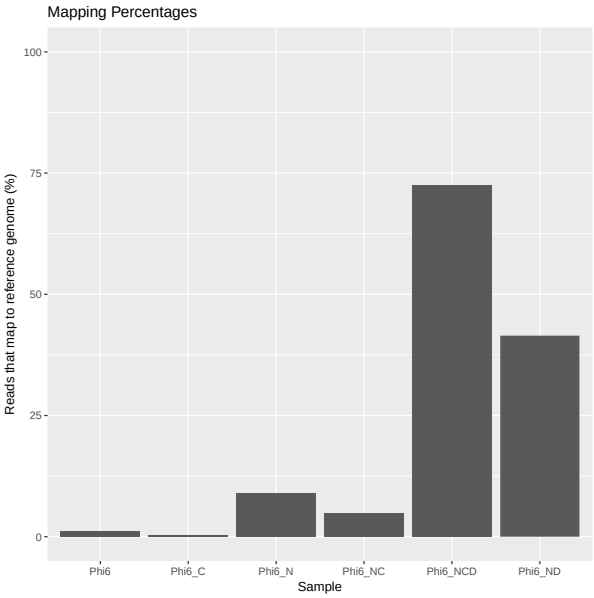

Figure 3. DMSO treatment has a greater effect than RNA concentration on generating dsRNA reads using next generation sequencing. Percent of sequencing reads mapping to phi6 reference genome with unconcentrated, unconcentrated nuclease-treated, concentrated nuclease-treated, and unconcentrated nuclease-treated DMSO-treated phi6 sequencing libraries.

The mock viral community (NIBSC reagent 11/242-001) was made by mixing 25 eukaryotic viruses [16]. Of those with RNA genomes, the most abundant in the sample is thought to be the dsRNA rotavirus, based on real time PCR results [23], which we selected for comparison with the ssRNA virus human parechovirus 3. Using a virus discovery pipeline, we only detected 1 read from a dsRNA virus (rotavirus A) in the sample untreated with DMSO (Table 3). However, under the 15% treatment, this increased to over 203 rotavirus reads, as well as reads from dsRNA viruses not thought to have been in the reagent, including human picobirnavirus and some totiviruses. This result is notable because the mock viral community was made such that rotavirus A was the most abundant virus [23]. If the most abundant virus can be so underrepresented in a known sample, this suggests many metagenomic studies will miss almost all dsRNA viruses. Furthermore, we indeed find dsRNA viruses that had previously evaded detection in the mock viral community [16]. The 50% and 90% treatments contained significantly fewer dsRNA reads than the 15% treatment, but reads from all viruses were reduced in these samples. Although the number of reads detected in the DMSO treated samples were low compared with the untreated sample, dsRNA was practically undetectable in the latter. Therefore, despite some caveats we suggest that the power of this method for detecting dsRNA viruses in metagenomic samples is clear.

bioRxiv preprint doi: https://doi.org/10.1101/644591; this version posted May 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 3. *Total number of reads detected in the mock viral community (reagent 11/242-001) under each DMSO treatment for human parechovirus, rotavirus A, and other dsRNA viruses.*

DMSO %	Parechovirus hits	Rotavirus A hits	Other dsRNA virus hits
0	3263	1	0
15	2089	203	27
50	293	10	9
90	785	11	4

Conclusion

Sequencing of dsRNA can frequently be problematic, with traditional cDNA synthesis being highly inefficient. We have demonstrated that a simple treatment with a cheap and common laboratory reagent can increase the number of sequencing reads from dsRNA organisms by over two orders of magnitude. Importantly, the positive effect of the DMSO treatment occurred independent of RNA concentration, even when RNA was undetectable. Furthermore, DMSO treatment was more important than RNA concentration in determining dsRNA read yield and it did not affect viral genome coverage. We suggest that samples to be sequenced that contain or are suspected to contain dsRNA are treated with at least 50% DMSO prior to cDNA synthesis. This treatment should also improve sequencing of dsRNA involved in innate immunity in plants and animals. We suspect this treatment can be successfully applied to other DNA sequencing technologies, because the DMSO treatment occurs at the cDNA synthesis step and has been shown to improve other procedures such as Sanger sequencing [11].

When preparing an environmental sample for sequencing, it is possible that there may be dsRNA viruses present that are undetectable when following standard protocols. Previous data have shown dsRNA viruses to be underrepresented in metagenomic samples [23]. Our data on the mock viral community (where the putatively most abundant virus was not detected without DMSO treatment) suggests dsRNA viruses will almost invariably go undetected in environmental samples. We have shown that not only will DMSO treatment increase representation of these organisms, the effect on ssRNA representation is minor. It may well be that dsRNA viruses are more numerous than thought but remain undetected using traditional methods.

Acknowledgements

Authors and contributions

Conceptualisation, A.H.W and S.L.D.M.; Methodology, A.H.W.; Software, A.H.W, E.D, and S.L.D.M.; Formal Analysis, A.H.W. and E.D.; Investigation, A.H.W.; Resources, E.D. and S.L.D.M.;

bioRxiv preprint doi: https://doi.org/10.1101/644591; this version posted May 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Data Curation, A.H.W.; Writing - Original Draft Preparation, A.H.W.; Writing - Review and Editing, A.H.W. and S.L.D.M.; Supervision, S.L.D.M.; Project Administration, S.L.D.M.; Funding, S.L.D.M.

Conflicts of interest

The authors declare no conflicts of interest.

Funding information

This study was funded by NIH R00 grant (4R00AI119401-02) to SLDM.

References

1. **Hannon GJ**. RNA interference. *Nature*. 2002 418: 244-251 Doi: 10.1038/418244a
2. **Coetzee P, Stokstad M, Venter EH, Myrmel M, Van Vuuren M.** Bluetongue: a historical and epidemiological perspective with the emphasis on South Africa. *Virology Journal* 2012 9(1): 198 Doi: 10.1186/1743-422X-9-198
3. **Crawford SE, Ramani S, Tate JE, Parashar UD, Svensson L, Hagbom M, Franco MA, Greenberg HB, O'Ryan M, Kang G, Desselberger U, Estes MK**. Rotavirus infection. *Nature Reviews* 2017 3: 17083 Doi:10.1038/nrdp.2017.83
4. **Steward GF, Culley AI, Mueller JA, Wood-Charlson EM, Belcaid M, Poisson G.** Are we missing half the viruses in the ocean? *The ISME Journal* 2013 7: 672-679 Doi: 10.1038/ismej.2012.121
5. **Woolhouse MEJ, Adair K.** The diversity of human RNA viruses. *Future Virol* 2013 8(2): 169-171 Doi: 10.2217/fvl.12.129
6. **Kleiner M, Hooper LV, Duerkop BA.** Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genomics* 2015 16:7 Doi: 10.1186/s12864-014-1207-4
7. **Peyambari M, Warner S, Stoler N, Rainer D, Roossinck MJ.** A 1000 year-old RNA virus. *Journal of Virology* 2018 doi:10.1128/JVI.01188-18
8. **Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB.** Benchmarking viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 2017: 5:e3817 Doi: 10.7717/peerj.3817
9. **Strauss JH, Kelly RB, Sinsheimer RL.** Denaturation of RNA with dimethyl sulfoxide. *Biopolymers* 1968 6: 793-807. Doi: 10.1002/bip.1968.360060604
10. **Varadaraj J, Skinner DM.** Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases. *Gene* 1994 140(1): 1-5. Doi: 10.1016/0378-1119(94)90723-4
11. **Mäntynen S, Laanto E, Kohvakka A, Poranen JKH, Ravantti JJ.** New enveloped dsRNA phage from freshwater habitat. *Journal of General Virology* 2015 96: 1180-1189 Doi: 10.1099/vir.0.000063

bioRxiv preprint doi: https://doi.org/10.1101/644591; this version posted May 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

12. **Díaz-Muñoz SL, Tenaillon O, Goldhill D, Brao K, Turner PE, Chao L**. Electrophoretic mobility confirms reassortment bias among geographic isolates of segmented RNA phages. *BMC Evolutionary Biology* 2013 13: 206 Doi: 10.1186/1471-2148-13-206
13. **Yang Y, Lu S, Shen W, Zhao X, Shen M, Tan Y, Li G, Li M, Wang J, Hu F, Le S**. Characterization of the first double-stranded RNA bacteriophage infecting *Psuedomonas aeruginosa*. *Scientific Reports* 2016 6:38795 Doi: 10.1038/srep38795
14. **Tang Y, Lu H, Sebastian A, Yeh YT, Praul CA, Albert IU, Zheng SY** Genomic characterization of a turkey reovirus field strain by Next-Generation Sequencing. *Infection, Genetics and Evolution* 2015 32: 313-321 Doi: 10.1016/j.meegid.2015.03.029
15. **Chiba S, Salaipeth L, Lin YH, Sasaki A, Kanematsu S, Suzuki N.** A Novel Bipartite Double-Stranded RNA Mycovirus from the White Root Rot Fungus *Rosellinia necatrix*: Molecular and Biological Characterization, Taxonomic Considerations, and Potential for Biological Control. *Journal of Virology* 2009 83(24): 12801-12812 Doi: 10.1128/JVI.01830-09
16. **Mee ET, Preston MD, CS533 Study Participants, Minor PD, Schepelmann S.** Development of a candidate reference material for adventitious virus detection in vaccine and biologicals manufacturing by deep sequencing. *Vaccine* 2016 34: 2035-2043 Doi: 10.1016/j.vaccine.2015.12.020
17. **Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R**. Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. *PLOS ONE* 2015 10(5): e0128036 Doi: 10.1371/journal.pone.0128036
18. **Martin M.** Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 2011 17(1): 10-12 Doi: 10.14806/ej.17.1.200
19. **Joshi NA, Fass JN**. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. 2011 (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.
20. **Zhang J, Kobert K, Flouri T, Stamatakis A.** PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 2014 30(5): 614-620 doi:10.1093/bioinformatics/btt593
21. **Langmead B, Salzberg SL**. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012 9(4): 357-359 Doi:10.1038/nmeth.1923
22. **The McDonnell Genome Institute**. bam-readcount. 2018 GitHub repository: https://github.com/genome/bam-readcount
23. **Li L, Deng X, Mee ET, Collot-Teixeira S, Anderson R, Schepelmann S, Minor PD, Delwart E.** Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. *Journal of Virological Methods* 2015 213: 139-146 Doi: 10.1016/j.jviromet.2014.12.002