A Novel Method for the Capture-based Purification of Whole Viral Native RNA Genomes

Chih Shen Cedric Tan[1,2,3], Sebastian Maurer-Stroh[4], Yue Wan[1], October Michael Sessions[5], Paola Florez de Sessions[1†]

[1] Genome Institute of Singapore (GIS), Agency for Science, Technology and Research (A*STAR), Singapore

[2]A*STAR Graduate Academy, Singapore

[3]University College London, United Kingdom

[4]Bioinfomatics Institute (BII), A*STAR, Singapore

[5]National University of Singapore, Saw Swee Hock School of Public Health, Singapore

[†]Correspondence to: P.F. de Sessions (desessions@gis.a-star.edu.sg)

ABSTRACT

Current technologies for targeted characterization and manipulation of viral RNA either involve amplification or ultracentrifugation with isopycnic gradients of viral particles to decrease host RNA background. The former strategy is non-compatible for characterizing properties innate to RNA strands such as secondary structure, RNA-RNA interactions, and also for nanopore direct RNA sequencing involving the sequencing of native RNA strands. The latter strategy, ultracentrifugation, causes loss in genomic information due to its inability to retrieve unassembled viral RNA. We developed a novel nucleic acid manipulation technique involving the capture of whole viral native RNA genomes for downstream RNA assays using hybridization baits in solution to circumvent these problems. This technique involves hybridization of biotinylated baits at 500 nucleotides (nt) intervals, stringent washes and release of free native RNA strands using DNase I treatment, with a turnaround time of about 6 h 15 min. Proof of concept was primarily done using RT-qPCR with dengue virus infected Huh-7 cells. We report that this protocol was able to purify viral RNA (561-791 fold). We also describe a successful application of our capture-based purification method to direct RNA sequencing, with a 77.47% of reads mapping to the target viral genome. We observed a reduction in human host RNA background by 1580 fold, a 99.91% recovery of viral genome with at least 15x coverage, and a mean coverage across the genome of 120x. This report is, to the best of our knowledge, the first description of a capture-based purification method for whole viral RNA genomes. The fundamental advantages of using our capture-based purification method makes it a superior alternative to conventional viral purification methods and would potentially pave a new path for the direct characterization and sequencing of native RNA molecules.

INTRODUCTION

Viruses with RNA genomes are the cause of many infectious diseases with serious consequences to human health and mortality such as flaviviruses, HIV-1, SARS-coronavirus, HTLV-1 and influenza virus (Cantera et al., 2014). It has been shown that RNA-RNA interactions (Romero-López et al., 2018) and intramolecular structure (Witteveldt et al., 2013) of viral RNA genomes play an important role in viral replication. However, extraction of viral RNA directly from culture often yields viral RNA with high host RNA background (Marston et al., 2013). Extracted viral RNA must therefore be amplified via polymerase chain reaction (PCR) to suit the sensitivities of the downstream assays (Ozsolak et al., 2011). Following conversion to cDNA and amplification, *in vitro* transcription is then performed to retrieve genetic information in its RNA form, since both RNA-RNA interactions and RNA intramolecular structure can only be observed whilst in this form (Peattie, 1979). Indeed, both Romero-López et al. (2018) and Witteveldt et al. (2013) amplified viral RNA obtained from cell culture via PCR amplification of viral cDNA clones followed by *in vitro* transcription. It has been shown, however, that such indirect methods for probing inherent properties of RNA increases experimental throughput at the expense of introducing PCR-induced artifacts and biases (Acinas et al., 2005), potentially clouding our understanding of these properties (Ozsolak et al.,2009).

With the advent of 3rd generation nanopore sequencing, direct RNA sequencing of native RNA strands is heralded as a PCR and cDNA-bias free sequencing technology (Garalde et al., 2018). However, the applicability of direct RNA sequencing to study viral genomes is similarly affected by high amounts of host RNA background in viral cultures. Current methods are therefore restricted to viral purification by the use of isopycnic gradients (Keller et al., unpublished) to obtain sufficient target read coverage for viable bioinformatics analysis. While viral purification by isopycnic gradients allow a high purity of viral RNA, only RNA from encapsulated viruses are obtained (Liu et al., 2011), making this technique unfeasible for the study of unassembled viral genomes, which have also been shown to be crucial to our understanding of viral replication machinery (Moon et al., 2015) (Manokaran et al., 2015).

From the abovementioned problems, we identified a need for a technique that isolates assembled and unassembled viral RNA molecules from host RNA background without PCR amplification or cDNA synthesis. In this study, we describe a capture-based purification method that leverages on current nucleic acid hybridization and enrichment techniques. This novel method is able to isolate native viral RNA molecules from a high host

RNA background without the use of PCR amplification or cDNA synthesis whilst retaining genetic information of both assembled and unassembled virus. Our capture-based purification method is distinguished from conventional hybridization-based or PCR-based enrichment techniques explored in the review paper by Mamanova et al. (2010) by virtue of not having any PCR amplification or cDNA conversion steps. Such a purification method, when merged with current technologies for probing and characterizing inherent RNA properties, could potentially catalyze the study of viral RNAs.

MATERIALS AND METHODS

Viral culture

Huh-7 hepatocellular carcinoma cells were cultured in DMEM media (10% Fetal Bovine Serum, 1% Penicillin Streptomycin) up to 70-90% confluency in Corning® 150 cm^2 cell culture flasks (Sigma Aldrich). Prior to infection, spent media was discarded and Huh-7 cell monolayer was washed with 10 mL 1x PBS twice. A multiplicity of infection of 1 was used, 2 mL dengue virus strain D1/SG/05K2402DK1/2005-1-I (DENV1) was diluted with 2 mL pure DMEM media and mixed via pipetting up and down 5 times. Four milliliter of diluted virus was pipetted into each flask and incubated at 37°C for 1 h. Flasks were rocked every 15 min. Subsequently, inoculum was removed via pipette and 30 mL of pure DMEM media was pipetted to each flask was incubated at 37°C for 30 h.

Trizol extraction of total RNA

Extractions were separately performed on viral supernatant and cell lysates. Media of infected flasks (viral supernatant) was transferred to a 50 mL Falcon tube. Subsequently the cell monolayer of each flask was treated with 8 mL of Trizol LS Reagent (Invitrogen) and 2.25 mL of 1x PBS (Thermo Fisher Scientific) followed by incubation at 37°C for 10 min. One milliliter aliquots of this cell lysate mixture were then pipetted to 1.5 mL Eppendorf tubes. 250 µL aliquots of viral supernatant were also pipetted into individual tubes and treated with 750 µL of Trizol LS Regeant. 200 µL of chloroform was then added to individual aliquots of viral supernatant and cell lysate before incubation at room temperature for 10 min. Tubes were shaken vigorously by hand for 30 s, and subsequently inverted by hand every 3 min. Each tube was centrifuged at 12 000 g, 4°C for 15 min. Upper aqueous phase in each tube was carefully removed and pipetted to a new 1.5 mL Eppendorf tube (approximately 500 µL). Two microliters of Glycoblue Coprecipitant (Invitrogen) and 500 µL of isopropanol (Sigma Aldrich) was added to each new tube and stored in -80°C overnight (or longer). Prior to the capture step, tubes were centrifuged at 12 000 g, 4°C for 15 min and supernatant was discarded. Pellet was re-suspended in 500 µL 75% ethanol diluted with nuclease-free water. Tubes were centrifuged again at 7 500 g, 4°C for 3 min and supernatant was discarded. Pellet was air dried. For the cell lysate RNA, total RNA concentration was determined via Nanodrop (Thermo Fisher Scientific) and re-precipitated in ethanol to obtain 30 µg total RNA pellets. Viral supernatant RNA either underwent capture or direct RNA sequencing library preparation without re-precipitation.

Bait design

Baits were designed using a method described in detail elsewhere (manuscript in submission). The bait stock in our experiments was previously used for hybridization-based target enrichment in a published study by The Singapore Zika Study Group (Ho et al., 2017). Briefly, 120mer biotinylated DNA baits (Integrated DNA Technologies) specific to conserved regions across all serotypes of dengue virus were used. Each bait was tiled along the viral genome with intervals of 500nt. A total of 6 pmol of baits were added to each capture.

Viral RNA capture

This protocol was modified from that for the use of xGEN Lockdown baits and reagents (Integrated DNA Technologies), optimized for our purposes of RNA capture. SeqCap Hybridization and Wash Kit (Roche) was used in place of Integrated DNA Technologies reagents. *Figure 1* demonstrates the workflow for the capture protocol.

Hybridization. RNA pellets extracted from either cell lysate or viral supernatant were re-eluted with reagents shown in *Table 1a* and incubated at 65°C for 4 h in a thermocycler. Total RNA input may be varied depending on the extraction and a single RNA pellet used is henceforth defined as a single capture.

Streptavidin-Biotin Binding. 100 µL of Dynabeads M-270 Streptavidin beads (Thermo Fisher Scientific) was pipetted to a 1.5 mL Eppendorf tube and equilibrated at room temperature for 30 min. The tube was put on a magnetic rack (Invitrogen) until eluate was clear and supernatant was discarded. 200 µL 1x Bead Wash Buffer was added and mixture was vortexed for 10 s before placing it on a magnetic rack and discarding the supernatant for a total of two washes. Hybridization mixture was immediately added to the beads and incubated at 65°C for 45 min.

Stringent Washes. All wash buffers were diluted to 1x prior to washing. Wash Buffer I and Stringent Wash Buffer was pre-heated for 1 hr to 65°C. 100 µL of Wash buffer I was added to the hybridization-bead mixture, agitated by hand and transferred to a new 1.5 mL Eppendorf tube. The new tube is placed on a magnetic rack. Once eluate is clear, supernatant was quickly discarded, taking caution not to remove any beads. 200 µL of Stringent Wash Buffer was added to the pellet and incubated at 65°C for 5 min, before discarding the supernatant as per the method above, for a total of two washes. Subsequent washes were performed as per the method above without the incubation step in the following sequence: Wash Buffer I, Wash Buffer II, and Wash Buffer III.

Dehybridization. Viral RNA was released from the beads via DNase I treatment. Beads were re-suspended in 20 µL of nuclease-free water, incubated at 95°C for 3 min then cooled to 37°C. Subsequently, 2.7 µL of 10x reaction buffer and 4 µL of TURBO DNase (Thermo Fisher Scientific) was added to the reaction and mixed by pipetting up and down 5 times, followed by further incubation at 37°C for 40 min. Subsequently, the tube was placed on a magnetic rack to allow the beads to aggregate. The supernatant was then transferred to a clean 1.5 mL Eppendorf tube for a 2.2x Agencourt RNAClean XP beads clean-up (Beckman Coulter). Briefly, 58.7 µL of beads were added to the supernatant and incubated for 8 min at room temperature, and then 2 min on a magnetic rack. The beads were washed twice with 75% Ethanol and residual ethanol was air-dried. The beads were then re-eluted in 30 µL of nuclease-free water.

RT-qPCR

The proof of concept for our capture-based purification method was to demonstrate that host RNA background could be removed from target viral RNA. Absolute quantification of host and viral RNA was done using reverse-transcription quantitative polymerase chain reaction (RT-qPCR). Two 1.5 ml Eppendorf tubes containing 30 µg of total RNA from cell lysate were prepared to represent pre and post-capture groups. 15 µL and 10 µL of nuclease-free water was added to the pre and post-capture group respectively. QuantiTect Probe RT-PCR Kit was used according to the manufacturer's instructions. Primer targets were selected in the nonstructural 1 protein region of DENV1 and human GAPDH (glyceraldehyde 3-phosphate dehydrogenase) as a reference gene.

The volumes of reagents, shown in *Table 1b*, were added to a 0.5 mL thin-walled PCR tube, and incubated in a ViiA 7 Real Time-PCR System (Applied Biosystems) with the program shown in *Table 1c* and analyzed using QuantStudio Real-Time PCR Software v1.1 (Applied Biosystems). Each plate included a non-template control for each pair of primers to check for PCR contamination. Samples were quantified in duplicate and non-template controls in triplicate.

A standard curve was constructed using a serial dilution of an oligo complementary to the primer target in accordance with MIQE Guidelines (Taylor et al., 2010). Custom oligoes of primer targets were ordered from Integrated DNA Technologies. Serial dilution was performed with nuclease-free water to obtain concentrations of 2.41×10^{n} copies/µL (where n=9,8,7,…,2). Each dilution was quantified in duplicate. The sequences of the oligoes, primers and probes used are shown in *Table 1d*. Average C_q values for the serial dilutions were linearly

regressed on $\log_2$(copies/µL) and Pearson's product-moment correlation coefficient was calculated. Equation (1) was used to determine primer efficiency:

$$Primer\ efficiency = 2^{\frac{-1}{m}} - 1 \qquad (1)$$

Where m is the gradient of the regression line of $<C_q>$ on $\log_2$(copies/µl)

Absolute quantity (copies/µL) was calculated using the respective equations of the regression lines.

The absolute quantity of DENV1 RNA, normalized to human GAPDH reference gene, was then compared between pre and post-capture groups. The ratio of the absolute quantity of DENV1 and human RNA pre and post-capture was compared, and purification factor calculated using Equation (2).

$$Purification\ factor = \frac{Post{-}capture\ \ DENV1\ amount}{Post{-}capture\ \ GAPDH\ amount} \div \frac{Pre{-}capture\ DENV1\ amount}{Pre{-}capture\ GAPDH\ amount} \quad (2)$$

In this study we defined purification factor as a measure of how well host RNA background is removed from viral RNA.

MinION Sequencing

Pre and post-capture RNA extracted from supernatant were sequenced on the MinION. Separately, we pooled 6 captures that increased the amount viral RNA used to generate the concentrated post-capture group prior to MinION sequencing using RNeasy Mini Kit (Qiagen). Prior to library preparation, RNA samples were polyadenylated using E. Coli Poly(A) Polymerase (New England Biolabs). The volumes of reagents shown in *Table 1e* were added to a thin-walled 0.5 mL PCR tube.

The reaction was then incubated in a thermocycler with the following program: 37°C for 30 min then cooled to 4°C before proceeding to a 2.2x Agencourt RNAClean XP cleanup (Beckman Coulter). The RNA was eluted in 11.5 µL nuclease-free water. Subsequently, library preparation and sequencing was performed according to the protocol for Direct RNA Sequencing Kit (SQK-RNA001) using 9.5 µL of RNA. All ligation reactions were extended to 20 min. The single pre and post-capture RNA libraries were loaded on a R9.4 MinION flow cell and sequencing was ran for 48 h using MinKNOW v1.14.1 (GUI 2.1.14) interface. The concentrated post-capture RNA library was sequenced using MinKNOW v1.15.0 (GUI 2.2.10). Live basecalling was turned off.

Bioinformatic Analysis

Raw fast5 files were basecalled and end adapters were trimmed using Poreplex v0.2 (https://github.com/hyeshik/poreplex) which internally calls Albacore v2.3.1 (Oxford Nanopore Technologies). Basecalled read metrics and plots were then generated using NanoPlot v1.0.0 (https://github.com/wdecoster/NanoPlot). Subsequently, mapping and alignment was performed using Graphmap v0.5.2 (https://github.com/isovic/graphmap) with default parameters. Mean coverage and mapping quality was calculated using Qualimap v0.2.0 (Garcia-Alcalde et al., 2012) respectively. Viral and human genome references used were EU081230.1 and GRCh38 respectively. All commands used are appended in the supplementary data.

RESULTS

RT-qPCR

Based on the ratios of the absolute quantity of viral to host RNA shown in *Table 2*, we report a 561-791 purification factor, suggesting successful retrieval of viral RNA and substantial removal of host RNA background. Capture yield, which is defined as the percentage of viral RNA amount captured, was calculated to be 6.98%-40.4%.

Supplementary Figure 1 shows the standard curves for DENV1 and GAPDH primers. The primer efficiency of DENV1 and GAPDH was 0.66 and 0.92 respectively. The DENV1 primer set used was a pre-optimised primer set that is able to detect all 4 dengue virus serotypes previously used and therefore contains ambiguous bases, explaining the lower primer efficiency. It was noted that relative quantification using the $2^{-\Delta\Delta CT}$ method (Schmittgen et al., 2008) was not appropriate as the primer efficiencies of DENV1 and GAPDH primer sets were not comparable. Absolute quantification and subsequent ratio comparison of viral and host RNA amounts is therefore more accurate in the determination of purification factor. All obtained values were within the dynamic linear range and limit of detection of the reaction conditions and primer sets used, suggesting reliability and accuracy of the reported RT-qPCR results.

MinION sequencing

Based on the mapping rates to human and DENV1 reference genomes for pre and post-capture groups, shown in *Table 3*, purification factor was calculated to be 272 fold. We also report a 3.89 fold increase in percentage recovery of DENV1 genome over 15x coverage from a single pre-capture to a single post-capture RNA sample. On the other hand, comparing the concentrated post-capture group (containing a pool of 6 captures) to the pre-capture group yielded a purification factor of 1580 fold and a 14 fold increase in the percentage recovery of DENV1 genome above with at least 15x coverage from 7.05% to 99.91%. *Supplementary figures 2, 3, and 4* show representative coverage plots for the DENV1 reference genome.

In Equation (3) and (4), a binomial distribution was used to determine the minimum read coverage, n, required to have a 0.95 confidence that k reads mapping to a particular base, X, are correctly called. We used the average error read rate of our runs as the probability, p, of correctly calling and mapping a single base.

$$X \sim B(n, p) \qquad (3)$$

$$P(X \geq k) \geq 0.95 \qquad (4)$$

Since the common heuristic for variant calling with Illumina reads requires 10x coverage (Illumina, https://www.illumina.com/Documents/products/technotes/technote_snp_caller_sequencing.pdf), we assumed the criterion that at least 10 reads were correctly called and mapped to a single base (*k*=10).

$$\frac{n!}{9!(n-9)!}(0.829)^9(0.171)^{n-9} \geq 0.95 \qquad (5)$$

Based on our sequencing runs, the average error read rate of pre and post-capture sequencing runs was 17.1±0.8% (calculated from *Table 3*). With this error read rate, we substitute 1-*p*=0.171 into Equation (5) and determined that the minimum coverage, *n*, required for variant calling to be at least 15x.

Discussion

We demonstrated with our RT-qPCR results that our novel RNA capture technique can be used to retrieve and purify viral RNA in its native form for downstream assays. This was corroborated by our direct RNA sequencing results. Indeed, unpurified viral RNA represented by the pre-capture group did not produce sufficient throughput for any meaningful downstream analysis due to the indiscriminate nature of the sequencing of both host and viral RNA. However, after treatment with our capture method, the amount of host RNA background has decreased greatly as seen from the 272 fold increase in mapping rates to target viral RNA. More significantly, the recovery of the whole DENV1 genome and the much higher read coverage across DENV1 obtained from sequencing concentrated post-capture RNA further demonstrates the superiority of our capture-based purification method even to PCR enrichment methods. In comparison, even with the incorporation of PCR amplification for shotgun sequencing of enterovirus, Jensen et al. (2015) reported only a 2-13 fold increase in mapping rates to enterovirus. This suggests that our capture-based purification method is highly successful in the removal of host RNA background. The minimum coverage required for variant calling was, as described above, benchmarked to that of Illumina reads so that the effectiveness of our capture-based purification method could be more accurately evaluated based on the higher error read rates of direct RNA sequencing technology. Even with the higher adjusted heuristic for variant calling, the 3.89 fold increase in percentage of the genome above 15x coverage suggests that application of our capture-based purification method increases the amount of target genetic information retrieved.

In light of the studies performed to date, only one study involving the use of direct RNA sequencing technology to study viral RNA genomes was found. Keller et al. (unpublished) utilized a custom sequencing adapter complimentary to the 3' end of influenza A viral genome to obtain sufficient sequencing throughput. However, this will be highly unviable for samples where viral strains or genotypes, as with viruses cultured from clinical samples. An appropriate example is the wide diversity in hepatitis C virus genotypes, which has caused issues in PCR primer design for target enrichment strategies due to sequence ambiguity (Thomson et al., 2016). On the same tangent, Cowan et al. (2005) noted that PCR-based enrichment techniques where *a priori* knowledge of target sequences is required for PCR primer design render the enrichment strategies ineffective in the characterization of novel viruses. It follows that the same sequence ambiguity would pose a problem for using customized sequencing adapters during direct RNA sequencing. Our method fundamentally allows for the

capture of unknown viral strains and genotypes due to the flexibility of using multiple baits complementary to different regions of the viral genome, or even multiple baits specific to different viruses in the hybridization step of our capture method. Indeed, Deviatkin et al. (2017) has reported the successful use of genus-specific degenerate baits for hybridization-based enrichment, which can be seamlessly incorporated into our bait design. By extension, the baits used can be appropriately modified for one's research focus. The robustness of this capture-based purification method therefore not only broadens the applicability of direct RNA sequencing on studying unknown viruses but also the applicability of a diverse range of direct RNA probing technologies.

Another major advantage to our capture-based purification method is that large machinery like an ultracentrifuge is not required. To perform our capture method, only a thermocycler and a heat block is required. This is especially applicable to research environments where an ultracentrifuge is not readily available. More fundamentally, our capture-based purification method allows for the capture of both unassembled and assembled viruses, which any method leveraging on the differential physiochemical properties of encapsulated viral RNAs cannot.

One point of concern for our protocol is the low capture yield. However, it must be noted that the capture yield is still significantly higher than that of conventional viral purification methods. Hall et al. (2014), in a study evaluating conventional viral purification methods, reported a 100 fold decrease in viral copy number post-treatment, whereas our method results in, at most, a 14 fold decrease in viral copy number. In evaluating our capture-based purification method, a wide range in capture yield was observed, which was due to the high sensitivity of RNA to potential contamination during handling. This is problem common to We recommend that the capture should be performed in a sterile workspace, strictly adhering to basic RNA handling procedures. We also recommend visual inspection to ensure that no beads are pipetted during stringent washing and clean-up steps are done to prevent accidental removal of bead-bound RNA during the capture. It is of utmost importance during the stringent washing that heated washes be done quickly to increase stringency, and that tubes are inverted and gently shaken to wash off any contaminants in the 1.5mL tubes. Aggressive vortexing is likely to cause some mechanical shearing of RNA and so we recommend gentle agitation of the tubes to re-elute during washing and clean-up steps.

Our capture-based purification method has a turnaround time of approximately 6 h 15 min (*Figure 1*), comparable to the time required for isopycnic gradient-based methods (Buclez et al., 2016). While there is a

significant loss of viral RNA post capture, we also note that our capture method is scalable. We found that the output of purified viral RNA from multiple captures could be pooled together to increase the amount of purified viral RNA for input into downstream assays.

CONCLUSION

In this study, we demonstrated a novel capture-based method for the purification of viral RNA from host RNA. We described a successful attempt of applying our capture method to the direct RNA sequencing of viral culture, using a DENV1 model, exemplifying how our capture method could improve viral-host RNA ratio in downstream assays involving native viral RNA strands. The fundamental advantages of using our capture-based purification method makes it a superior alternative to conventional viral purification methods and can potentially pave a new path for the direct characterization and sequencing of native RNA molecules.

FUNDING

We would like to thank the Biomedical Research Council for the transition fund entitled, Human Emerging Infectious Disease Initiative (HEIDI) [Ref No: H16/99/b0/013].

ACKNOWLEDGEMENT

We would like to acknowledge Ng Hui Qi Amanda and Niranjan Nagarajan from Computational Systems Biology 5, GIS; Aw Poh Kim Pauline (DENV1 primer design), Kwan Kam Weng Philip, Balamurugan Periaswamy, Png Eileen from GERMS Platform for their technical advice throughout the course of this study; Vithiagaran Gunalan and Dimitar Kenanov from Maurer-Stroh's group for helpful conversations. We would also like to acknowledge members of Wan Yue's team, in particular, Lim Wei Sheng Shaun, Aw Jong Ghut Ashley and other GIS members, like Yao Fei from the Genome Innovation Lab, for support during the sequencing process. Lastly, we would like to thank Andreas Wilm and Shih Chih Chuan from the GIS Bioinformatics Core.

SUPPLEMENTARY DATA

References

Acinas, Silvia G., et al. "PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample." *Applied and environmental microbiology* 71.12 (2005): 8966-8969.

Buclez, Pierre-Olivier, et al. "Rapid, scalable, and low-cost purification of recombinant adeno-associated virus produced by baculovirus expression vector system." Molecular Therapy-Methods & Clinical Development 3 (2016).

Cantara, William A., Erik D. Olson, and Karin Musier-Forsyth. "Progress and outlook in structural biology of large viral RNAs." Virus research 193 (2014): 24-38.

Cowan, Don, et al. "Metagenomic gene discovery: past, present and future." *TRENDS in Biotechnology* 23.6 (2005): 321-329.

García-Alcalde, Fernando, et al. "Qualimap: evaluating next-generation sequencing alignment data." *Bioinformatics* 28.20 (2012): 2678-2679.

Hall, Richard J., et al. "Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery." *Journal of virological methods* 195 (2014): 194-204.

Ho, Zheng Jie Marc, et al. "Outbreak of Zika virus infection in Singapore: an epidemiological, entomological, virological, and clinical analysis." *The Lancet Infectious Diseases* 17.8 (2017): 813-821.

Kafetzopoulou, Liana E., et al. "Assessment of Metagenomic MinION and Illumina sequencing as an approach for the recovery of whole genome sequences of chikungunya and dengue viruses directly from clinical samples." *bioRxiv* (2018): 355560.

Liu, Chia-Chyi, et al. "Purification and characterization of enterovirus 71 viral particles produced from vero cells grown in a serum-free microcarrier bioreactor system." PloS one 6.5 (2011): e20005.

Mamanova, Lira, et al. "Target-capture strategies for next-generation sequencing." *Nature methods* 7.2 (2010): 111.

Manokaran, Gayathri, et al. "Dengue subgenomic RNA binds TRIM25 to inhibit interferon expression for epidemiological fitness." *Science* 350.6257 (2015): 217-221.

Marston, Denise A., et al. "Next generation sequencing of viral RNA genomes." BMC genomics 14.1 (2013): 444.

Moon, Stephanie L., et al. "Flavivirus sfRNA suppresses antiviral RNA interference in cultured cells and mosquitoes and directly interacts with the RNAi machinery." *Virology* 485 (2015): 322-329.

Ozsolak, Fatih, et al. "Direct RNA sequencing." Nature 461.7265 (2009): 814.

Ozsolak, Fatih, and Patrice M. Milos. "RNA sequencing: advances, challenges and opportunities." Nature reviews genetics 12.2 (2011): 87.

Peattie, Debra A. "Direct chemical method for sequencing RNA." *Proceedings of the National Academy of Sciences* 76.4 (1979): 1760-1764.

Romero-López, Cristina, and Alfredo Berzal-Herranz. "A long-range RNA–RNA interaction between the 5’ and 3’ ends of the HCV genome." *Rna* (2009).

Quick, Joshua, et al. "Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples." *nature protocols* 12.6 (2017): 1261.

Schmittgen, Thomas D., and Kenneth J. Livak. "Analyzing real-time PCR data by the comparative C T method." *Nature protocols* 3.6 (2008): 1101.

Taylor, Sean, et al. "A practical approach to RT-qPCR—publishing data that conform to the MIQE guidelines." *Methods*50.4 (2010): S1-S5.

Thomson, Emma, et al. "Comparison of next generation sequencing technologies for the comprehensive assessment of full-length hepatitis C viral genomes." Journal of clinical microbiology (2016): JCM-00330.

Witteveldt, Jeroen, et al. "The influence of viral RNA secondary structure on interactions with innate host cell defences." *Nucleic acids research* 42.5 (2013): 3314-3329.

TABLES AND FIGURES

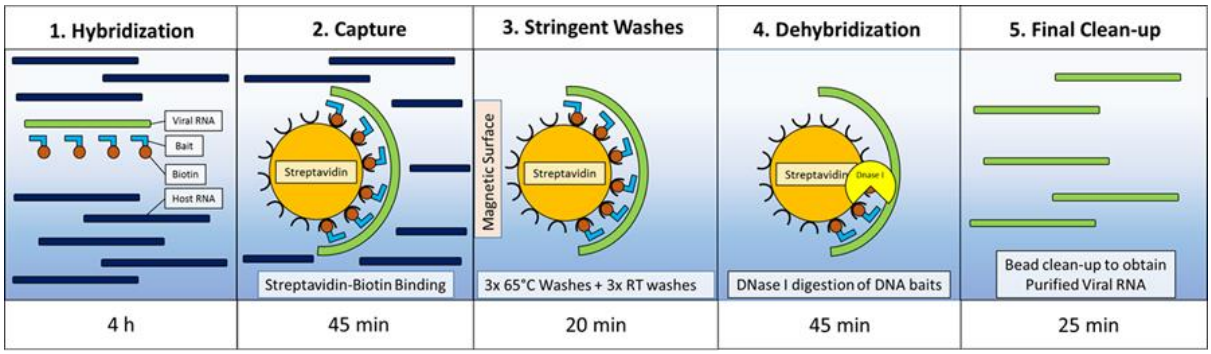


Figure 1: Capture-based purification workflow.

Components	Volume (uL)
2x Hybridization Buffer	30
Hybridization Component A	12
Diluted Baits	4
Nuclease-free water	10

Table 1a: Reagents and volumes for hybridization step.

Components	Volume (uL)
RNA in Nuclease-free Water	1.0
2x QuantiTect Probe RT-QPCR Master Mix Buffer	5.0
QuantiTect RT Mix	0.1
Nuclease-free Water	1.65
10 µM Forward Primer	0.9
10 µM Reverse Primer	0.9
10 µM Probe	0.45
Total Volume	10

Table 1b: Reagents and volumes for RT-qPCR.

Program name	Temperature/°C	Time/min	Number of Cycles
Reverse Transcription	48	30	1
PCR Initial Activation Step	95	15	
Denaturation	95	0.25	40
Combined Annealing/Extension	55	1	

Table 1c: RT-qPCR program set on Applied Biosystems ViiA 7 Real Time-PCR System.

Sequence Name	Sequence (5'-3')
DENV1 F	ACTAGYGGTTAGAGGAGACC
DENV1 R	GGTCTCCWCTAACCTCTAGT
DENV1 Probe	FAM-ACCAGGGRAAGCTGTAYCYT-BHQ
DENV1 Standard	ACTAGTGGTTAGAGGAGACCCCTCCCGAAACACAACGCAGCAGCGGGGCCCAACACCAGGGG AAGCTGTACCCTGGTGGTAAGGACTAGAGGTTAGAGGAGACC
GAPDH F	GATTCCACCCATGGCAAATTC
GAPDH R	ATTTCCATTGATGACAAGC
GAPDH Probe	FAM-CGTTCTCAGCCTTGACGGTGCCA-BHQ
GAPDH Standard	TATGATTCCACCCATGGCAAATTCCATGGCACCGTCAAGGCTGAGAACGGGAAGCTTGTCATC AATGGAAATCCCATCA

Table 1d: Primers, probes and standards used with the respective modifications for RT-qPCR.

Components	Volume (uL)
RNA	24
10x E. Coli Poly(A) Polymerase Reaction Buffer	3.2
ATP (10mM)	3.2
E. Coli Poly(A) Polymerase	1.6

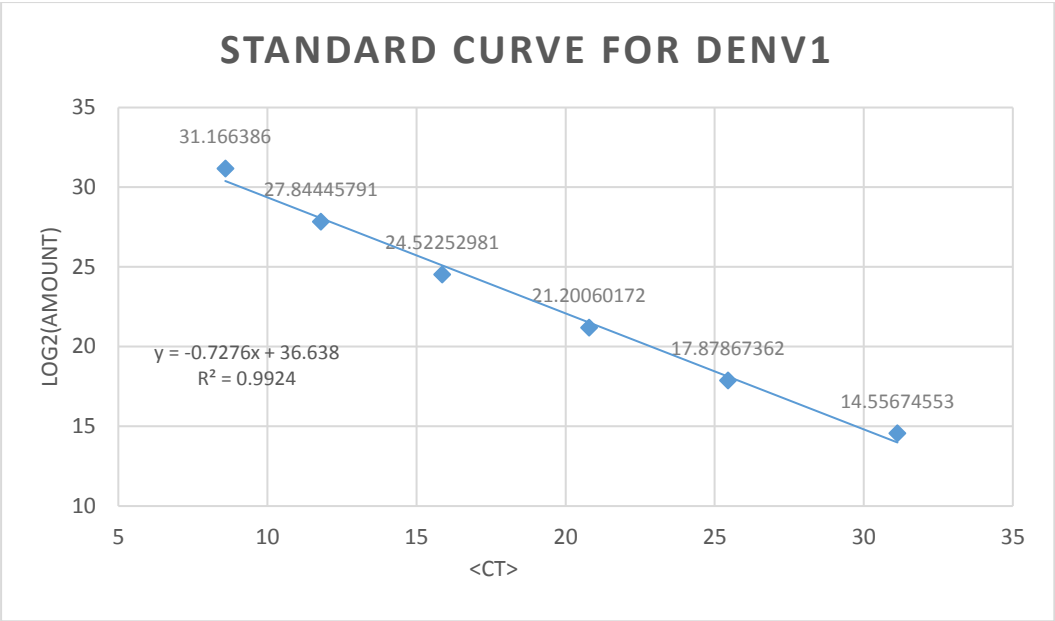
Table 1e: Reagents and volumes used for polyadenylation of RNA samples.

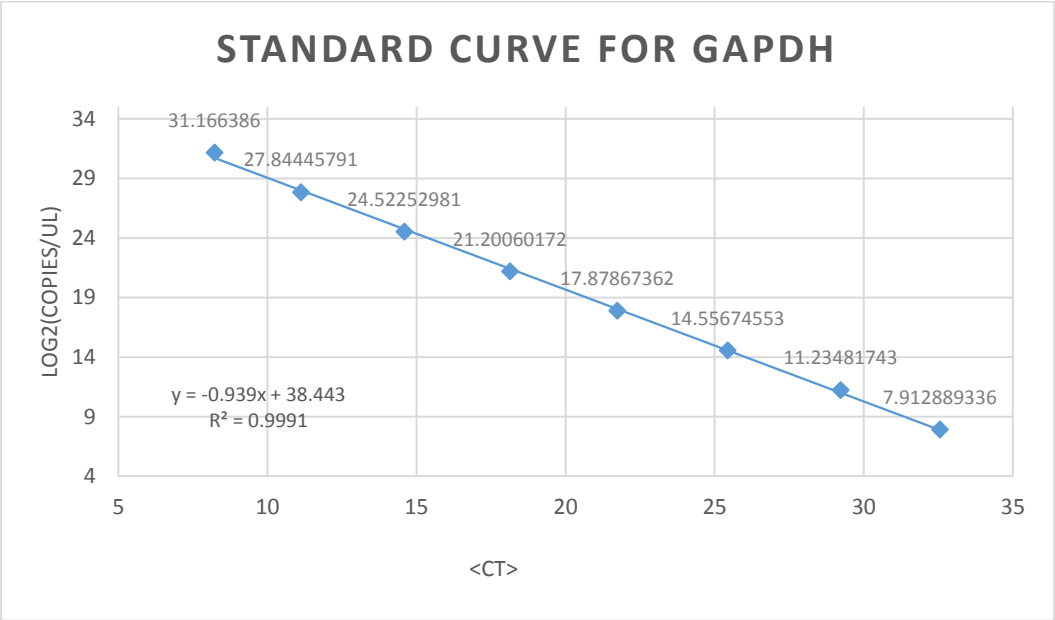
Treatment Group	Primer	<C_q>	Copies/µL	Total No. of Copies	Purification Factor
Pre-capture	DENV1	10.47*±0.08*	5.50×10^8	8.25×10^9	
	GAPDH	14.75*±0.1*	2.52×10^7	3.78×10^8	5.61×10^2
Post-capture	DENV1	17.14*±0.09*	1.92×10^7	5.77×10^8	
	GAPDH	29.63*±0.1*	1.57×10^3	4.72×10^4	
Pre-capture	DENV1	12.13*±0.1*	2.39×10^8	3.58×10^9	
	GAPDH	17.04*±0.4*	5.70×10^6	8.55×10^7	
Post-capture	DENV1	15.31*±0.1*	4.83×10^7	1.45×10^9	7.91×10^2
	GAPDH	29.75*±0.07*	1.46×10^3	4.37×10^4	

Table 2: Comparison of absolute quantities of DENV1 and GAPDH used for calculation of purification factor.

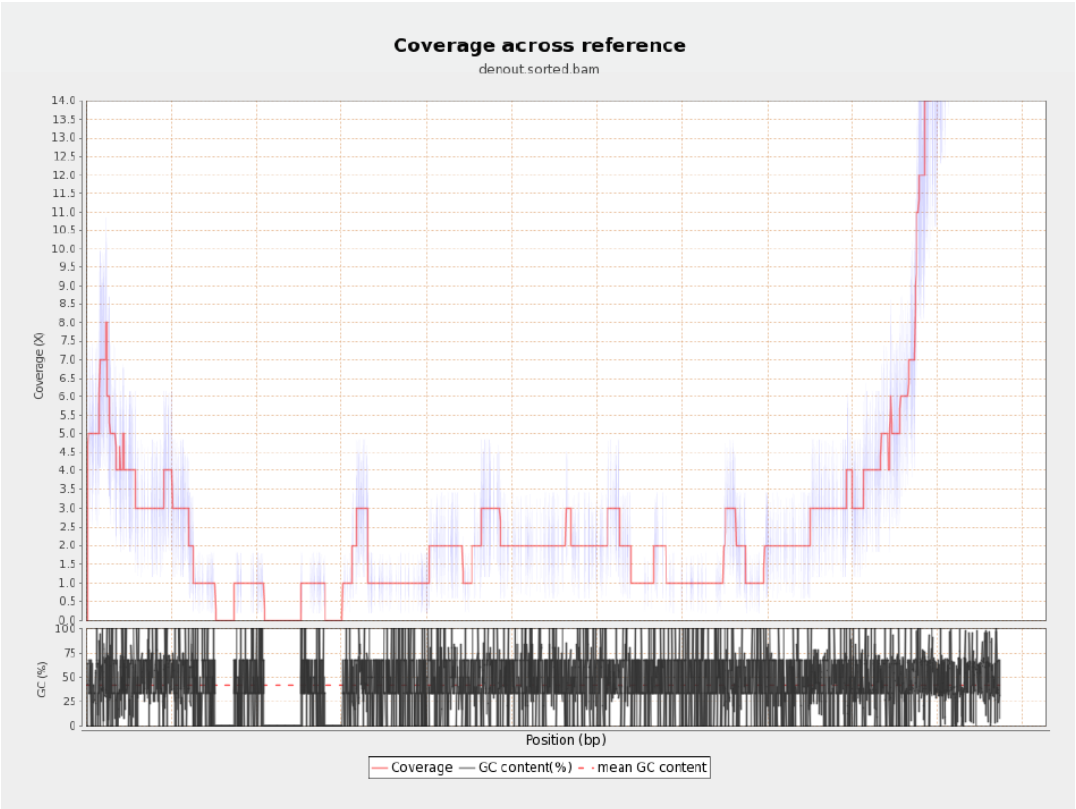
	Pre-capture	***Post-capture***	***Concentrated Post-capture***
<DENV1 Coverage>	*12.11*	*10.02*	*120.23*
<DENV1 Mapping Quality>	*34.31*	*31.60*	*36.99*
DENV1 Error Read Rate (%)	*16.49*	*17.61*	*15.27*
Mapped to DENV1 (%)	*0.63*	*31.35*	*77.47*
Mapped to Human (%)	*77.66*	*14.19*	*6.05*
Percentage of DENV1 Genome Recovered ≥15X coverage (%)	*7.05*	*27.45*	*99.91*
Purification factor	*272x*		*1580x*

Table 3: Summary of bioinformatic analysis for pre and post-capture MinION sequencing runs.

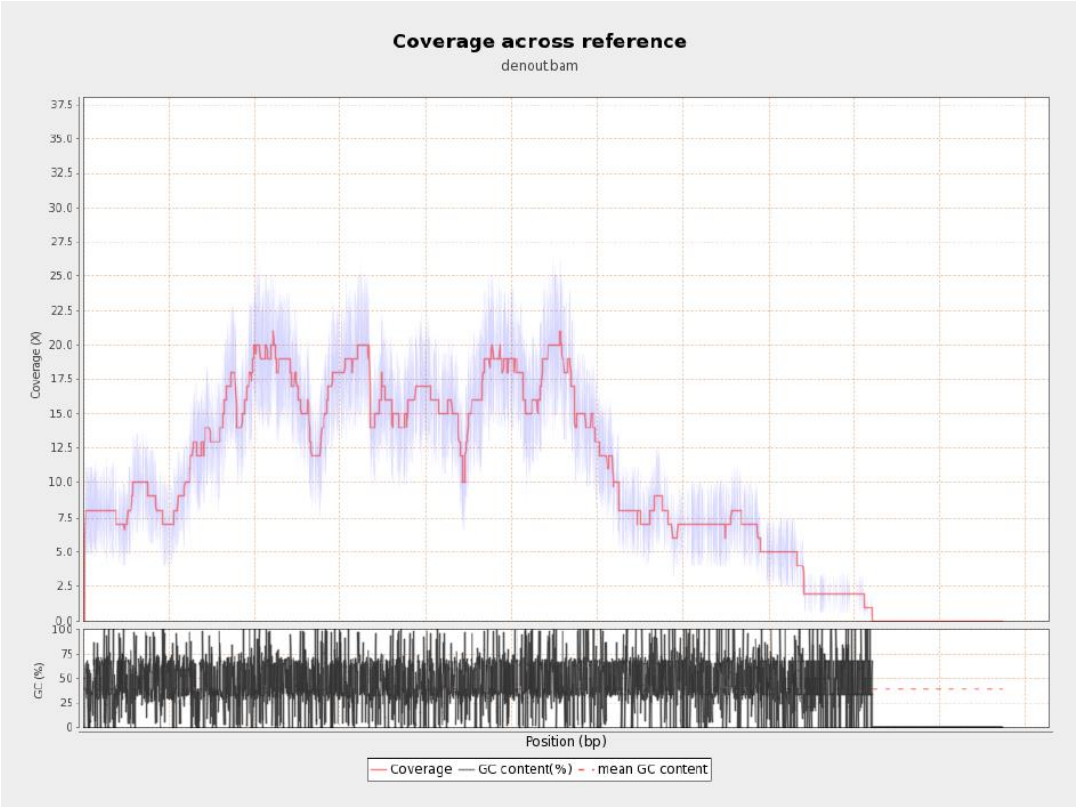




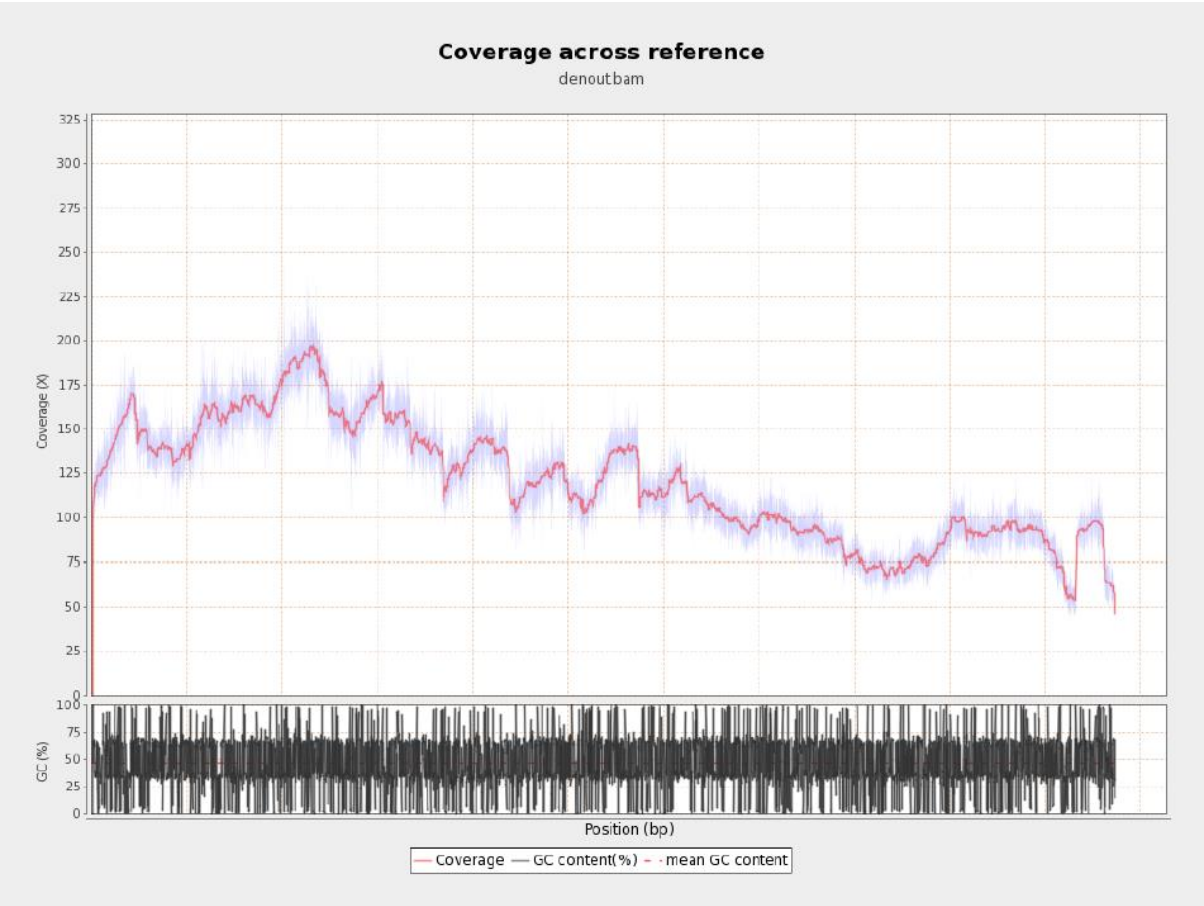
Supplementary Figure 1: DENV1 and GAPDH standard curves used for calculation of primer efficiency.



Supplementary Figure 2: Coverage plot against nucleotide position on DENV1 reference for pre-capture group MinION sequencing run.



Supplementary Figure 3: Coverage plot against nucleotide position on DENV1 reference for post-capture MinION group sequencing run.



Supplementary Figure 4: Coverage plot against nucleotide position on DENV1 reference for concentrated post-capture group MinION sequencing run.