

SCIENTIFIC REPORTS



OPEN

Enrichment of Viral Nucleic Acids by Solution Hybrid Selection with Genus Specific Oligonucleotides

Andrei A. Deviatkin^{1,2}, Alexander N. Lukashov^{1,3}, Mikhail M. Markelov², Larisa V. Gmyl¹ & German A. Shipulin⁴

Despite recent advances, our knowledge of potential and rare human pathogens is far from exhaustive. Current molecular diagnostic tools mainly rely on the specific amplification of marker sequences and may overlook infections caused by unknown and rare pathogens. Using high-throughput sequencing (HTS) can solve this problem; but, due to the extremely low fraction of pathogen genetic material in clinical samples, its application is only cost-effective in special, rather than routine, cases. In this study, we present a method for the semi-specific enrichment of viral conservative sequences in a HTS library by hybridization in solution with genus-specific degenerate biotinylated oligonucleotides. Nucleic acids of the test viruses (yellow fever virus and Japanese encephalitis virus) were enriched by solution hybrid selection using pan-flavivirus oligonucleotides. Moreover, enterovirus (family: *Picornaviridae*, genus: *Enterovirus*) sequences were successfully enriched using foot-and-mouth disease virus (family: *Picornaviridae*, genus: *Aphthovirus*) oligonucleotide. The enrichment factor relative to the background nucleic acid was about 1,000-fold. As hybridization has less stringent oligonucleotide match requirements than PCR, few oligonucleotides are sufficient to cover the potential sequence variation in the whole genus and may even enrich nucleic acids of viruses of other related genera. Efficient enrichment of viral sequences makes its use in diagnostics cost-efficient.

Increasing human-animal interactions, migration and rising population density have resulted in the growing threat from emerging infections. Novel pathogens are among the major challenges to public health¹.

The majority of emerging infectious diseases (EIDs) among humans are viral zoonoses. This is exemplified by recent outbreaks of zoonotic diseases with high lethality rates (SARS and MERS coronaviruses, ebolavirus etc.)². Fast and accurate identification of infectious agents is a critical step in any surveillance strategy.

Most of the current diagnostic systems, such as PCR, NASBA and LAMP, are based on the specific amplification of nucleic acids (NAs)^{3,4}. However, all these methods have one serious limitation: diagnostic systems efficiently detect specific NA markers only if the pathogen genome was previously identified, sequenced and annotated⁵.

According to some estimates, there are hundreds of thousands viruses awaiting discovery⁶. Some of these viruses can pose a real threat to human health. HTS techniques (also termed next-generation sequencing, NGS) can directly identify genetic markers of unknown infectious agents in a pool of NAs from biological samples without using classical time-consuming virological or microbiological methods. However, total NA sequencing is not used in routine infectious disease diagnostics because of certain limitations. The genetic markers of a pathogen usually account for only a few reads among a huge number of background host-related sequences (human/animal and non-pathogenic environmental bacteria)⁷. For example, when HTS was used to identify the Schmallenberg virus, only 192 out of 94,722 reads (0.2%) were mapped to its genome⁸.

The following strategies have been developed to increase the viral/host NA ratio before or after NA extraction: physical enrichment methods (ultrafiltration, centrifugation, virus particles precipitation); biological methods (exonuclease treatment); host NA depletion^{7,9–12}. Some tools for viral NA enrichment are based on PCR with non-specific or genus-specific degenerate oligonucleotide primers. The preferential amplification of

¹Chumakov Federal Scientific Center for Research and Development of Immune and Biological Products of Russian Academy of Sciences, Moscow, Russian Federation. ²Research Institute of Occupational Health, Moscow, Russian Federation. ³Institute of Molecular Medicine, Sechenov First Moscow State Medical University, Moscow, Russia. ⁴Federal Budget Institute of Science Central Research Institute for Epidemiology, Moscow, Russian Federation. Correspondence and requests for materials should be addressed to A.A.D. (email: andreideviatkin@gmail.com)

pathogenic sequence (PATHseq) is based on using a set of 88 specific 8-mer oligonucleotides (which do not match the sequences of the 2,000 most abundant human transcripts) as primers for the synthesis of secondary cDNA strands¹³. Alternatively, the VIDISCA method based on the presence of virus-specific endonuclease restriction sites was suggested¹⁴. Advanced VIDISCA uses non-rRNA binding hexamers in a reverse transcription (RT) reaction¹⁵. It is also possible to enrich viral sequences by a broad-range RT-PCR using a set of degenerate oligonucleotides^{16, 17}.

The recently developed virome capture sequencing platform for vertebrate viruses (VirCapSeq-VERT) technique increases the sensitivity of sequence-based virus detection by HTS¹⁸. The authors described target selection of genome sequences of known viruses by hybridization in a solution using 1,993,200 biotinylated oligonucleotides representing 342,438 viral coding sequences. However, as this approach used specific (non-degenerate) oligonucleotides, theoretically, it can only detect viruses that have regions with > 80% sequence identity to known viruses.

In this study, we describe a cost-effective method for the significant enrichment of conservative fragments of viral genomes by hybridization with degenerate genus-specific biotinylated oligonucleotides. We analyze the utility of the technique for NA library enrichment in order to increase HTS efficiency and decrease the cost per sample.

Materials and Methods

Selection of probe sequences. All full-genome flavivirus sequences available in Genbank in October 2014 were extracted. After alignment with MAFFT¹⁹, sequences sharing more than 90% identity were discarded. The data set was refined manually to exclude pegiviruses and incomplete sequences. Foot-and-mouth disease virus alignment was prepared similarly. Hybridization oligonucleotides matching the most conserved genome fragments were designed according to the following principles:

- length between 50–90 nt
- 90% match with all target sequences (degenerate nucleotides were used where necessary)
- whenever possible, degenerate positions were simplified to G or K (T/G), given that mismatches involving T and G result in a minimal penalty to hybridization efficiency²⁰.

The potential ability of oligonucleotides to hybridize with selected target sequences was checked using the Mfold web server²¹.

Finally, three pan-flavivirus biotinylated oligonucleotides were chosen: Flavi-8469 ATGACNGAYACYACNVCNTTGGRCARCARMGRGKTTYAARGARAARGTKGACAC (genome positions: 8,709–8,764 in KF297915); Flavi-8844 GCNAARGGNAGNMGNCGYATNTGGT WYATGTGGCTNGGNDN MGNTTYTNGARTTYGARGCYTNGGNTTYTNAAYGARGAYCAYTGG (9,084–9,173 in KF297915); and Flavi-9834 GGNGMNTGGATGACNACWGANGACATGYTNNNGTNTGGAAYVRRGTNTGGAT (10,092–10,144 in KF297915).

The hybridization probe aimed at all FMDV sequences was designed similarly: ggTggKWTgCCKTCTggKT-gTKCTggKACYTSTgTgTTKAATTCATKWTCAA (genome positions: 6,803–6,855 in X80059).

Samples, nucleic acid extraction and preparation of artificial samples containing human RNA and viral RNA.

Donor whole blood samples were collected with written informed consent from volunteers in the CRIE (Central Research Institute for Epidemiology, Moscow). All experiments were performed in accordance with relevant guidelines and regulations and were approved by the CRIE ethics committee. Total human RNA was extracted from the human mononuclear cell fraction of the samples using the RNeasy Mini Kit (Qiagen).

The Gagar Japanese encephalitis virus (JEV) strain was cultured in pig embryo kidney cells. The 17D yellow fever virus (YFV) strain was passaged in Vero cells. The Gregory Echovirus 11 (E11) strain, which was originally obtained from ATCC, was cultured in human rhabdomyosarcoma cells. Viral RNA was extracted from cell culture medium using the Viral RNA Mini Kit (Qiagen).

The relative concentrations of isolated viral RNA and host RNA were estimated using in-house qPCR assays. Then, total human RNA (100 ng) was mixed with either 20 pg of YFV, 30 pg of JEV or 100 pg of E11. The total volume of each sample was adjusted to 10 µL with water.

Preparation of cDNA libraries. First-strand cDNA was synthesized from RNA using Reverta-L RT kit (AmpliSens) according to the manufacturer's instructions. The kit uses MMLV reverse transcriptase and random hexamer primers.

Second-strand cDNA synthesis was performed by adding 10 µL of RT mixture, 8 µL of 5x buffer C (AmpliSens), 5 µL of dNTP (2.5 mM) (AmpliSens), 1 µL of random hexamer oligonucleotide (168 pmol/µL) (in-house) and 1 µL of RNase H (New England Biolabs), made up to 50 µL with milli-Q water. The mixture was incubated at 37 °C for 10 min. The enzymes were inactivated by heating at 95 °C for 2 min. After cooling on wet ice, 1 µL of Bst polymerase (New England Biolabs) and 0.5 µL of Klenow exo- (New England Biolabs) were added. Incubation conditions were as follows: 4 °C for 45 secs, 20 °C for 5 min, 37 °C for 5 min, 45 °C for 5 min, 50 °C for 5 min and 68 °C for 5 min. The reaction product was purified with Agencourt AMPure XP beads (Beckman Coulter) using the manufacturer's protocol, with the exception of an altered bead-to-DNA ratio, where we used a 1.5:1 beads-to-DNA solution ratio, instead of a 1.8:1 ratio. DNA was eluted to 17 µL of milli-Q water.

To incorporate sequencing adapters, purified cDNA libraries (15 µL) were mixed with 5 µL of 5x transposase buffer (in-house) and 0.5 µL of the transposase and transposon complex, termed transposome (in-house)²². Total reaction volume was adjusted with milli-Q water up to 25 µL. The mixture was incubated at 55 °C for 5 min. The reaction was stopped by adding 1 µL of 200 mM EDTA. The reaction product was purified with Agencourt AMPure XP beads (Beckman Coulter) using the manufacturer's protocol. DNA was eluted to 22 µL of milli-Q water.

The cDNA libraries were amplified by real-time PCR. PCR reactions (50 μ L) included 200 nM of NTS1 primer (5' TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG 3'), 200 nM of NTS2 primer (5' GTC TCG TGG GCT CGG A GAT GTG TAT AAG AGA CAG 3'), 0.25 mM of dNTP, 14 μ L of PCR-mix-2-FL (AmpliSens), 1.2X of EvaGreen (Biotium Inc.), and a 20 μ L purified cDNA sample.

The oligonucleotides (total volume: 5 μ L) were separated from other parts of the PCR mixture (total volume: 45 μ L) by a wax layer. Amplification conditions were as follows: 30 °C for 1 min; 37 °C for 1 min; 72 °C for 1 min; denaturation at 95 °C for 5 min; 18 cycles at 95 °C for 20 s, 65 °C for 30 s and 72 °C for 60 s. Real-time PCR amplification was performed using a Rotor Gene 6000 instrument (Qiagen). The amplification was monitored to stop the reaction during the exponential phase and before the reaction reached the plateau²³.

Aliquots of PCR product were stored at -20 °C for further analysis.

Solution hybrid selection of viral NAs. The cDNA libraries containing human and YFV or human and JEV genetic material were hybridized with three pan-flavivirus biotinylated oligonucleotides (Flavi-8469, Flavi-8844, Flavi-9834). The library containing human and E11 genetic material was hybridized with FMDV-specific biotinylated oligonucleotide.

Hybrid selection was performed similarly to a previously published method²⁴ with minor modifications.

Briefly, a 20 μ L mix containing 1X hybridization buffer (0.25 M sodium phosphate buffer: pH 7.0, 0.5% SDS, 5 mM EDTA, 1xSSC, 2xDenhardt's solution), biotinylated oligonucleotide(s) (final concentration: 20 nM each), 1 μ g Cot-1 DNA (Invitrogen), 50 μ M oligonucleotide Nts1_SpC3 (5' CTG TCT CTT ATA CAC ATC TGA CGC TGC CGA CGA spC3 3'), 50 μ M oligonucleotide Nts2_SpC3 (5' CTG TCT CTT ATA CAC ATC TCC GAG CCC ACG AGA C spC3 3') and 3 μ L of an amplified cDNA library was heated at 97 °C for 5 min. Next, the mixture was slowly cooled to 47 °C (3 °C every 10 min). After 15 h at 47 °C, the hybridization mixture was added to 100 μ g of M-270 streptavidin Dynabeads (Invitrogen), which had been initially processed as follows: the beads were washed twice in 150 μ L of a buffer containing 2 M NaCl, 10 mM Tris-HCl (pH 7.5) and 1 mM EDTA, then incubated in 14 μ L of a solution containing 1X hybridization buffer, 2.2% casein (Stereospecific Detection Technologies GmbH) and 100 mg/L of Cot-1 DNA (Invitrogen) for 60 min at 37 °C, and then resuspended in 80 μ L of a buffer containing 0.4 M NaCl and 6.25 mM Tris-HCl (pH 7.5).

After 15 min at 37 °C, the beads were pulled down by the magnetic rack and washed once at 47 °C for 5 min with 1 mL of prewarmed 1xSSC (saline and sodium citrate)/0.1% SDS (sodium dodecyl sulfate). Next, the beads were washed twice at 47 °C for 5 min with 1 mL of prewarmed 0.2xSSC/0.1%SDS, once at room temperature for 5 min with 1 mL 0.2xSSC/0.1%SDS, and once at room temperature for 5 min with 1 mL of 0.2xSSC/0.1% Tween 20. The beads were resuspended once after each washing step and collected by the magnetic rack. Hybrid-selected DNA, which was attached to the beads, was resuspended with 20 μ L of 0.1% Tween 20.

The cDNA libraries captured on the beads were reamplified by PCR. PCR reactions (50 μ L) consisted of 200 nM NTS1 primer (5' TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG 3'), 200 nM NTS2 primer (5' GTC TCG TGG GCT CGG A GAT GTG TAT AAG AGA CAG 3'), 0.25 mM dNTP, 14 μ L of PCR-mix-2-FL (AmpliSens), 3 μ L of 20X EvaGreen (Biotium Inc.) and 20 μ L of bead suspension.

The oligonucleotides and dNTP (total volume: 13 μ L) were separated from other components of the PCR mixture (total volume: 37 μ L) by a wax layer. Amplification conditions were as follows: denaturation at 95 °C for 5 min; 19 cycles at 95 °C for 20 s, 65 °C for 40 s and 72 °C for 60 s. Real-time PCR amplification was performed using the Rotor Gene 6000 instrument (Qiagen).

Enrichment evaluation by specific qPCR. The relative concentration of viral cDNA before and after hybridization was measured by a specific qPCR using GAPDH (Glyceraldehyde-3-Phosphate Dehydrogenase) as the housekeeping gene. The cycle threshold (Ct value) was chosen as being at the middle of the exponential phase of the amplification curve. The ratio between the host and viral NAs was assessed in terms of the Ct value difference. Oligonucleotide primers and probes for viral cDNA quantification (Table 1) were chosen in close proximity to the location of hybridization oligonucleotide probes. The efficiency of PCR amplification was assumed to be 100% (twofold amplification on each cycle) for further calculations.

For HTS, the initial and enriched cDNA libraries were indexed by PCR using Nextera-compatible oligonucleotide primers. The libraries were pooled at equal ratios. Paired-end 250-base sequencing by synthesis was performed on the MiSeq System (Illumina) using protocols provided by the manufacturer. Samples were de-multiplexed using the Illumina software, with FASTQ files generated.

Bioinformatic analysis. Reads in the FASTQ format were filtered by quality (Q30), and the adapter sequences were removed by Trimmomatic²⁵. The Bowtie 2²⁶ tool was used for aligning sequencing reads to references. Alignments were processed using SAMtools²⁷ and pysamstats software, and figures were generated in the R environment.

Results

First, enrichment efficiency was estimated by comparison of the host:viral cDNA ratio in libraries before and after hybridization (Table 2).

In unprocessed samples, delta Ct, indicating the host:pathogen ratio, was 6.2, 5.0 and 9.3 for human/YFV, human/JEV and human/E11 libraries, respectively. As a result of hybridization, this ratio changed to -5.1, -7.5 and 2.3 for human/YFV, human/JEV and human/E11 libraries, respectively. According to qPCR data, the ratio of target sequences to host sequences increased by approximately 2¹¹, or 2,000 times, in the case of the human/YFV library, approximately 2¹², or 4,000 times, in the case of the human/JEV library after hybridization with three pan-flavivirus oligonucleotides, and about 2⁷, or 100 times, for the human/E11 library after hybridization with the FMDV-specific probe. Due to the use of the semi-quantitative PCR, the enrichment efficiency

Primer name	Sequence 5'–3'	Position in the reference genome
YFV 8740F	AGCGGGAAGCTAGGAAGATCATGAAAGTT	8,740–8,767 in KF907504
YFV 8881R	TTCCAGGTAAGCTCCAATGGCTGCATGACTTC	8,850–8,881 in KF907504
YFV 8792Pr	Fam-CTGGCCAGAGAAAAGAACCCAGACTGTGCA-BHQ1	8,792–8,822 in KF907504
JEV 8890F	TCAATAGCAACGCGGCTCTTGGGA	8,890–8,912 in KF297915
JEV 9004R	TTTCCCTCTCTTCATCAACCATCTCCCAA	975–9,003 in KF297915
JEV 8935Pr	Fam-AATGGAGCACGGCGGTGAGGCT-BHQ1	8,935–8,957 in KF297915
Enterov_6003F	CCTGTGTAATTCACCACCTGTACAGA	6,754–6,781 in X80059
Enterov_6182R	CCGTATGCAATCATCCTAACTGGTCCAA	6,905–6,933 in X80059
Enterov_6037Pr	FAM-CACTACTTTGAGCGGGTGTATGCCCTCAGG-BHQ1	6,788–6,819 in X80059
GAPDH_F	CAACAGCGACACCCACTCCT	1,046–1,065 in NM_002046.5
GAPDH_R1	CACCCTGTGCTGTAGCCAAA	1,140–1,160 in NM_002046.5
GAPDH_prb	Fam-TGGGGCTGGCATTGCCCTCAAC-BHQ1	1,079–1,100 in NM_002046.5

Table 1. Primers and probes used for specific qPCR.

Sample	PCR assay	Ct before hybridization	dCt1 (Ct[virus]-Ct[human])	Ct after hybridization	dCt2 (Ct[virus]-Ct[human])	ddCt (dCt1-dCt2)
Human/YFV	Human	21.3	6.2	21.8	-5.1	11.3
	YFV	27.5		16.7		
Human/JEV	Human	22.0	5.0	22.3	-7.5	12.5
	JEV	27.0		14.8		
Human/E11	Human	21.4	9.3	21.2	2.3	7.0
	E11	30.7		23.5		

Table 2. Enrichment of viral targets assessed by quantitative real-time PCR detection.

Sample	Organism	Share of reads in HTS data, % and reads number (before hybridization)	Share of reads in HTS data, % and reads number (after hybridization)
Human/YFV	Human	92.47 (659,449/713,161)	84.43 (548,426/649,532)
	YFV	0.00 (1/713,161)	0.2 (1,315/649,532)
Human/JEV	Human	92.12 (473,999/514,540)	84.09 (631,867/751,393)
	JEV	0.00 (38/514,540)	3.37 (25,364/751,393)
Human/E11	Human	91.96 (377,820/410,873)	79.72 (197,183/247,357)
	E11	0.07 (303/410,873)	9.65 (23,880/247,357)

Table 3. Enrichment of viral targets assessed by HTS read counts.

estimates should be considered as approximate; however, they at least indicate the order of magnitude of enrichment efficiency.

HTS was also performed to validate the hybridization-induced change in the host:virus cDNA ratio. Significant enrichment of target viral sequences was observed in samples after hybridization (Table 3). The number of reads mapped to viral genomes increased from single accidental reads to a significant fraction of the total. The increase in the number of viral reads corresponded well to the enrichment ratio suggested by qPCR (Table 2).

The pattern of HTS reads coverage along the YFV, JEV and ECHOV genomes corresponded to an expected enrichment in the genome regions, which were complementary to the biotinylated oligonucleotides (Figs 1, S1). Interestingly, the pattern of reads coverage in the E11 library shows a bimodal distribution. After omitting reads shorter than 200 nucleotides from the analysis (Fig. S2), only one peak close to the probe binding site was left. Therefore, the second sequence density peak in Fig. 1 likely was due to short reads from longer cDNA library fragments that were hybridized to the oligonucleotide.

Discussion

Rapidly evolving methods in molecular diagnostics face limitations associated with extremely high genetic diversity of viruses. Furthermore, most of the global virome remains undiscovered, resulting in a lack of reference sequences for novel pathogen identification. One way to overcome this situation is by developing a system, which allows specific enrichment of unknown virus content in processed biological material based on known homologous sequences.

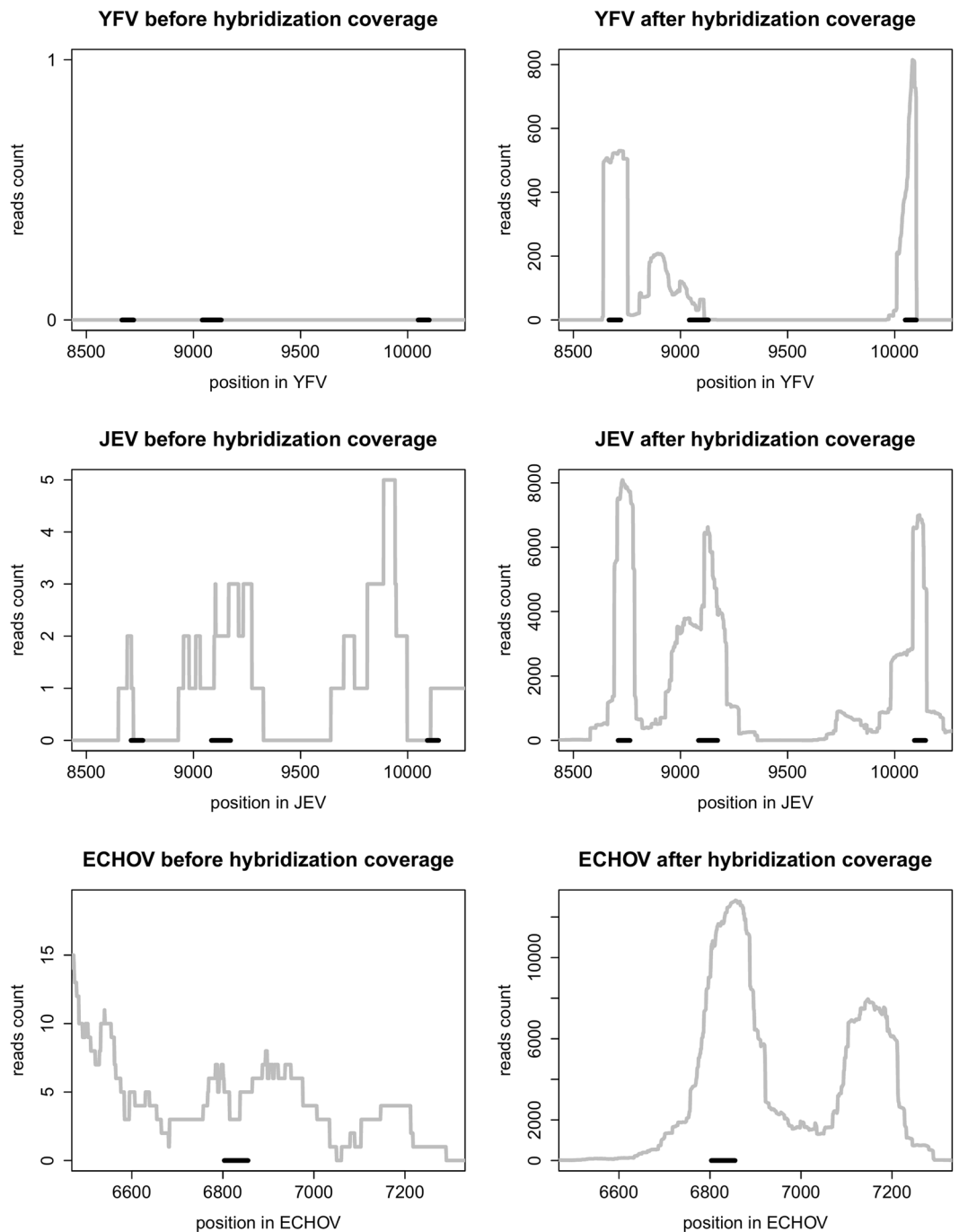


Figure 1. YFV, JEV, ECHOV genome fragment coverage (gray) before and after hybridization. Genome fragments, which are complementary to the biotinylated oligonucleotides, are indicated by black bars. Positions are indicated relatively to Genbank entries KF907504, KF297915, X80059, respectively.

We were able to increase viral cDNA content for two distinct flaviviruses (YFV and JEV) by hybridization of HTS-ready libraries with pan-flavivirus oligonucleotides. Comparison of the genomes of these viruses (KF907504 vs KF297915, respectively) using BLASTn shows a 66% identity level at 27% of the genome coverage. Oligonucleotide probes were designed based on the most conserved genome fragments and had 96.4–100.0% sequence identity (considering degenerate positions as matches) with the viruses (54/56 positions in Flavi-8469 probe vs. YFV; 88/90 in Flavi-8844 vs. YFV; 53/53 in Flavi-9834 vs. YFV; 56/56 in Flavi-8469 vs. JEV; 89/90 in Flavi-8844 vs. JEV; 52/53 in Flavi-9834 vs. JEV). The hybridization oligonucleotides showed no preference towards YFV and JEV in the enrichment experiment. The data set for the oligonucleotide design included over 100 highly diversified flavivirus sequences, therefore, comparable enrichment efficiency may be expected for any known flavivirus. To further test the specificity range of the method, we tested whether it was possible to enrich the cDNA of one virus genus with hybridization probes designed for another genus of the same family. Indeed,

it was possible to enrich the content of enterovirus cDNA using an oligonucleotide designed to to the FMDV virus (79.2% sequence identity between the FMDV oligonucleotide and the enterovirus). The *Enterovirus* and *Aphthovirus* genera are very distantly related within the *Picornaviridae* family. Therefore, the specificity breadth of the method goes beyond a genus and could cover distantly related genera within a highly diversified virus family.

Importantly, cDNA enrichment by hybridization lacks theoretical specificity limitations typical of PCR, such as the strict requirement of the precise match of the three 3'-end nucleotides of the PCR primer. Any small number of additional mismatches in any part of the hybridization oligonucleotide will only reduce hybridization efficiency, but not completely prevent it. Therefore, this approach is theoretically more robust in terms of sequence variation than genus-specific PCR with degenerate primers. Moreover, the degeneracy of the probes (the number of possible unique oligonucleotides within a degenerate oligonucleotide preparation) was between 2¹² and 2⁴⁶, thereby greatly exceeding the maximum degeneracy of 2⁸, which is acceptable in a PCR primer.

Enrichment of cDNA libraries by hybridization involves four steps of NA copying and amplification: reverse transcription, second-strand synthesis and two rounds of library amplification before and after hybridization. Even a standard library preparation procedure results in the uneven amplification of genome fragments²⁸. As expected, this bias was further deepened in our protocol. Therefore, it is advisable to use several hybridization oligonucleotides per taxon. On the other hand, using specifically designed degenerate probes, which are complementary to conserved genome regions, is much more affordable than synthesizing multiple specific oligonucleotides to the whole genome, as suggested earlier¹⁸.

At the present time, molecular diagnostic tools mostly rely on the specific amplification of marker sequences. In other words, while a standard molecular test answers the question about whether pathogen X is present in the sample, this does help investigators to answer the more important question, “which pathogen is present in the sample?”²⁹. Moreover, the correctness of assumption about a pathogen's presence in a sample completely depends on the qualification of a medical specialist.

To date, many tools for unbiased HTS-based pathogen detection exist. However, due to the extremely low fraction of a pathogen's genetic material in clinical samples, its usage is only cost-effective in very special cases. The proposed technique can enrich the share of pathogen genome fragments in a library, even if the NA sequence of a pathogen is currently unknown. It can also significantly enhance the utility of HTS for diagnostics.

References

1. Stacey, K., Adel, M., Stanley, L. & Leslie, P. *The Impact of Globalization on Infectious Disease Emergence and Control*. (National Academies Press). doi:10.17226/11588 (2006).
2. Team, G. P. T. F. Global Health Security: The Lessons from the West African Ebola Virus Disease Epidemic and MERS Outbreak in the Republic of Korea. *Osong Public Heal. Res. Perspect.* **6**, S25–S27 (2015).
3. Yan, L. *et al.* Isothermal amplified detection of DNA and RNA. *Mol. Biosyst.* **10**, 970–1003 (2014).
4. Chiendo, A. M. *et al.* Better Tests, Better Care: Improved Diagnostics for Infectious Diseases. *Clin. Infect. Dis.* **57**, S139–S170 (2013).
5. Kennedy, P., Quan, P.-L. & Lipkin, W. Viral Encephalitis of Unknown Cause: Current Perspective and Recent Advances. *Viruses* **9**, 138 (2017).
6. Anthony, S. J. *et al.* A Strategy To Estimate Unknown Viral Diversity in Mammals. *MBio* **4**, e00598-13–e00598-13 (2013).
7. Hall, R. J. *et al.* Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *J. Virol. Methods* **195**, 194–204 (2014).
8. Rosseel, T. *et al.* DNase SISPA-next generation sequencing confirms Schmallenberg virus in Belgian field samples and identifies genetic variation in Europe. *PLoS One* **7**, 1–6 (2012).
9. Grard, G. *et al.* A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. *PLoS Pathog.* **8**, e1002924 (2012).
10. Sauvage, V. & Eloit, M. Viral metagenomics and blood safety. *Transfus. Clin. Biol.* **23**, 28–38 (2016).
11. Hasan, M. R. *et al.* Depletion of human DNA in spiked clinical specimens for improvement of sensitivity of pathogen detection by next-generation sequencing. *J. Clin. Microbiol.* **54**, 919–927 (2016).
12. Matranga, C. B. *et al.* Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. *Genome Biol.* **15**, 519 (2014).
13. Ge, F. *et al.* Preferential Amplification of Pathogenic Sequences. *Sci. Rep.* **5**, 11047 (2015).
14. de Souza Luna, L. K. *et al.* Identification of a contemporary human parechovirus type 1 by VIDISCA and characterisation of its full genome. *Virol. J.* **5**, 26 (2008).
15. van der Hoek, L. *et al.* Performance of VIDISCA-454 in feces-suspensions and serum. *Viruses* **4**, 1328–1334 (2012).
16. Dedkov, V. G. *et al.* Retrospective diagnosis of two rabies cases in humans by high throughput sequencing. *J. Clin. Virol.* **78**, 74–81 (2016).
17. Maganga, G. D. *et al.* Molecular typing of PPRV strains detected during an outbreak in sheep and goats in south-eastern Gabon in 2011. *Virol. J.* **10**, 82 (2013).
18. Briese, T. *et al.* Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. *MBio* **6**, 1–11 (2015).
19. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
20. Hooyberghs, J., Van Hummelen, P. & Carlon, E. The effects of mismatches on hybridization in DNA microarrays: Determination of nearest neighbor parameters. *Nucleic Acids Res.* **37**, (2009).
21. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
22. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
23. Bowman, S. K. *et al.* Multiplexed Illumina sequencing libraries from picogram quantities of DNA. *BMC Genomics* **14**, 466 (2013).
24. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182–189 (2009).
25. Bolger Anthony, M., Marc, L. & Bjoern, U. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* **30**, 2114 (2014).
26. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
27. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
28. Van Dijk, E. L., Jaszczyszyn, Y. & Thermes, C. Library preparation methods for next-generation sequencing: Tone down the bias. *Exp. Cell Res.* **322**, 12–20 (2014).
29. Flygare, S. *et al.* Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. *Genome Biol.* **17**, 111 (2016).

Acknowledgements

Development of hybridization oligonucleotides and data analysis were supported by the Russian Scientific Foundation Grant 14-15-00619.

Author Contributions

G.A.S., A.N.L. and M.L.M. conceived and designed the study; A.A.D. and A.N.L. wrote the manuscript; A.A.D., M.L.M. and L.V.G. conducted the experiments and all authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-10342-w](https://doi.org/10.1038/s41598-017-10342-w)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017