

# Evaluation of direct metagenomics and target enriched approaches for high-throughput sequencing of field rabies viruses

Anna Orłowska, Ewelina Iwan,  
Marcin Smreczak, Jerzy Rola

Department of Virology,  
National Veterinary Research Institute, 24-100 Puławy, Poland  
anna.orłowska@piwet.pulawy.pl

Received: March 6, 2019

Accepted: November 4, 2019

## Abstract

**Introduction:** High-throughput sequencing (HTS) identifies random viral fragments in environmental samples metagenomically. High reliability gains it broad application in virus evolution, host-virus interaction, and pathogenicity studies. Deep sequencing of field samples with content of host genetic material and bacteria often produces insufficient data for metagenomics and must be preceded by target enrichment. The main goal of the study was the evaluation of HTS for complete genome sequencing of field-case rabies viruses (RABVs). **Material and Methods:** The material was 23 RABVs isolated mainly from red foxes and one European bat lyssavirus-1 isolate propagated in neuroblastoma cells. Three methods of RNA isolation were tested for the direct metagenomics and RABV-enriched approaches. Deep sequencing was performed with a MiSeq sequencer (Illumina) and reagent v3 kit. Bioinformatics data were evaluated by Kraken and Centrifuge software and *de novo* assembly was done with metaSPAdes. **Results:** Testing RNA extraction procedures revealed the deep sequencing scope superiority of the combined TRIzol/column method. This HTS methodology made it possible to obtain complete genomes of all the RABV isolates collected in the field. Significantly greater rates of RABV genome coverages (over 5,900) were obtained with RABV enrichment. Direct metagenomic studies sequenced the full length of 6 out of 16 RABV isolates with a medium coverage between 1 and 71. **Conclusion:** Direct metagenomics gives the most realistic illustration of the field sample microbiome, but with low coverage. For deep characterisation of viruses, *e.g.* for spatial and temporal phylogeography during outbreaks, target enrichment is recommended as it covers sequences much more completely.

**Keywords:** rabies virus, HTS, complete genome, field samples.

## Introduction

Rabies virus (RABV), belonging to the *Lyssavirus* genus, has a negative-sense, single-stranded, non-segmented RNA genome approximately 12 kb long. Due to the low fidelity of viral polymerase used for replication, RNA viruses are characterised by a high level of diversity (25). RABV RNA codes five proteins and the N gene-coding nucleoprotein is the most conservative fragment within the RABV genome (13). Traditionally, the N gene was utilised as the favourite not only for RABV detection but also for viral speciation and phylogenetic analysis (4, 20, 21). Partial or full nucleoprotein gene sequencing using the Sanger method

is mostly valuable for preliminary phylogenetic studies and identification of RABV species. However, it provides limited information on viral genomes; therefore, the International Committee for Taxonomy of Viruses (ICTV) requires full-length genome sequencing when proposing new lyssavirus species.

Initially, whole-genome sequences (WGS) of RABV isolates were obtained with the genome-walking procedure. This procedure was based on amplification of multiple RABV fragments covering 12 kb of the RV genome and their sequencing using the Sanger method (27). From the first WGS of the RABV prototype in 1988 (26) to date, many advances in molecular technology have been made and alternative platforms for

high-throughput sequencing (HTS) have been developed (6, 18).

The application of HTS allows metagenomic-based identification of random viral fragments in environmental samples (15). HTS generates a multiple raw data set to obtain a consensus sequence, and therefore it increases the reliability of sequencing by avoiding mismatches generated during Sanger sequencing. Owing to the high reliability of HTS, nowadays it is broadly applied in studies concerning virus evolution, host-virus interaction, and pathogenicity (28). Whole genome population studies also offer great potential to provide deep investigation of phylogenetic relationships among isolates for better understanding of what determines virus spread and persistence in the field, as the disease spread is reflected in a genetic signature in pathogen genomes. Viral pathogens, particularly fast-evolving RNA viruses, are model systems for exploration of pathogen populations, as they rapidly accumulate genetic diversity on a timescale similar to epidemiological processes (1, 5, 9).

Rabies is an acute encephalitis, a fatal zoonosis affecting all warm-blooded animals. In Europe, it is most widespread in wild animals, particularly in red foxes. Brain samples delivered for rabies diagnosis are very often putrefied or autolysed, making the genetic material unsuitable for HTS studies through disintegration. Before being diagnosed in the laboratory, in many cases the carcasses of dead animals have stayed in unfavourable environmental conditions for a long period, resulting in decomposition of the brain tissue and bacterial contamination. Such decomposition, the presence of many pathogens both viral and bacterial, and host organism genetic material adversely affect both the quality of RABV RNA and the effectiveness of full-length sequencing of the viral genome. Appropriate homogenisation of the brain tissue and effective purification of RNA isolated from field samples are also significant for the success of whole genome sequencing or metagenomic studies. A limitation of HTS encountered subsequently is a low concentration of viral RNA in total RNA isolated from the brain tissue.

Taking into account the difficulties and limitations in deep sequencing, namely the high background of genetic material from host species and bacteria in field samples, the main objective of the work was to evaluate and validate the different protocols (the direct metagenomic and RABV-enriched approaches) used for deep sequencing of field RABV isolates and compare different protocols used for RNA extraction in terms of their application to HTS.

## Material and Methods

**Samples.** To develop a method for WGS and evaluate alternatives for maximal efficiency, a total of 23 animal brains diagnosed as positive for RABV infection in an immunofluorescence test (FAT) were

used in the study. The samples (21 fox brains, 1 cat brain, and 1 dog brain) were collected in the southern part of Poland (Lesser Poland and Subcarpathian provinces) between 1997 and 2017. To compare the utility of RABV propagation in cell culture as applicable to the metagenomics approach, European bat lyssavirus-1 isolate cultivated in a neuroblastoma cell line (passage 6) was also included in the study.

Based on the RNA extraction procedure as well as on the scheme of RNA preparation for HTS, two HTS approaches were applied: the direct metagenomics approach (groups I, II, and III) and the RABV-enriched approach (group IV) as shown in Table 1.

**RNA extraction.** Total RNA was extracted from 20% homogenates (w/v) of the brain tissue. Three different methods of RNA isolation were tested:

**A** – RNA isolation from 140  $\mu$ L of brain homogenates using a QIAmp Viral RNA Mini Kit (Qiagen, Germany) according to the manufacturer's instructions;

**B** – RNA extraction using a Direct-zol RNA MiniPrep Kit (Zymo Research, USA) preceded by a brain tissue digestion stage with DNase (TURBO DNase, Ambion, Thermo Fisher Scientific, USA). RNA extraction was carried out on tissue supernatant lysed with the TRIzol contained in the kit. All steps of RNA isolation were performed according to the manufacturer's instructions;

**C** – RNA isolation with a method combining TRIzol/chloroform/ethanol and a QIAmp Viral RNA Mini Kit. The initial step consisted of lysis of brain homogenates with TRIzol. Subsequently, RNA was extracted with chloroform and precipitated with 75% ethanol. RNA was washed on the columns and eluted with water for molecular biology. For each sample, three subsequent elutions of RNA were performed. The third RNA eluate was subjected to long RT-PCR. All of the extracted RNA was used immediately for further studies or stored frozen at  $-20^{\circ}\text{C}$ .

**Reverse transcription (RT) and double-stranded DNA (dsDNA) synthesis.** For complementary DNA (cDNA) synthesis, two protocols were evaluated, one of which used SuperScript III reverse transcriptase (Invitrogen, Thermo Fisher Scientific) and the other a PrimeScript RT-PCR Kit (TaKaRa Bio, Japan) with random hexamers. Synthesis of cDNA was performed as per the manufacturers' protocols. To digest the remaining RNA, cDNA was treated with RNase H (EurX, Poland). The second strand of DNA was synthesised with Klenow exonuclease (New England Biolabs, USA).

**Genome amplification RT-PCR.** To generate amplicons covering the complete viral genome of approximately 12 kb, three primer pairs suitable for generation of overlapping PCRs were designed. PCR products ranged between 3.7 kb and 4.5 kb and covered fragments of the RABV genome at position A, spanning nucleotides 1–4499; B, through 4418–8276; and C, 8172–11801. In addition, two sets of primers suitable for

overlapping hemi-nested PCR products on each of these three amplicons were also designed (amplicon length 1.6 kb to 2.3 kb). The primers targeted conserved regions of the rabies virus genome with particular consideration given to the sequences of Polish RABV reference genomes (GenBank accession nos. MF197743.1, MF197741.1 and MF197742.1). All primer details are shown in Table 2. For the purpose of this study, two distinct amplification protocols were tested:

**I – Two-step RT-PCR.** RNA in a 5 µL volume was mixed with 30 pmol of each of the two amplification primers from primer pairs A, B, or C and incubated at 70°C for 5 min and 37°C for 10 min. The hybridisation mixture was brought to 20 µL with the addition of 4 µL of 5X first strand buffer, 0.1 µL of 0.1 M DTT, 1 µL of RNase Out (Invitrogen, Thermo Fisher Scientific), 1 µL of SuperScript III enzyme, and 1 µL of 10 mM dNTPs.

**Table 1.** Details of samples in the comparative study of extraction methods. RNA concentration, virus detection using real-time RT-PCR, and the number of reads obtained during whole-genome sequencing

Group	Isolate	Sample origin	Collection date	Extraction method /RT-PCR procedure	Concentration of dsDNA after clean-up (Qubit HS) (ng/µL)	Verification of HTS library	Total number of reads	Number of viral reads	% of viral reads	Number of RABV reads (centrifuge)	Number of contigs	Average coverage
I	767121097L	red fox	1997	A – QIAmp Viral RNA Mini Kit/ RT + amplification of dsDNA with Klenow fragment	1.45	+	22,250	117	0.0525	3	-	-
	965180404L	red fox	2004		1.06	+	15,282	67	0.438	0	-	-
	1045120899L	red fox	1999		1.08	+	158,723	496	0.31	8	-	-
II	767121097L	red fox	1997	B – Direct-zol RNA MiniPrep Zymo Research/ RT + amplification of dsDNA with Klenow fragment	5.06	-	-	-	-	-	-	-
	965180404L	red fox	2004		1.21	+	160,354	1,1170	6.965	132	19	3.5
	1045120899L	red fox	1999		1.15	+	200,756	4,218	2.101	133	6	2.5
	1321180108L	red fox	2008		1.2	+	517,696	70,436	13.605	1,359	1	32
	1379120910L	red fox	2010		2.21	+	948,295	68,491	7.222	4,765	3	152
	1996181013L*	red fox	2013		1.18	+	839,440	47,118	5.613	28,116	1	495
III	1992121113L	red fox	2013	C – TRIzol/ chloroform/ethanol/ RT+ amplification of dsDNA with Klenow fragment	13.4	+	423,115	1,417	0.334	62	9	2
	1739120912L	red fox	2012		2.93	+	4,216,387	14,103	0.334	657	1	15
	1679180512L*	red fox	2012		2.04	+	4,543,264	21,683	0.4772	2,935	1	71
	1577121111L	red fox	2011		4.28	+	1,058,492	5,313	0.05019	564	1	13
	1525180711L	red fox	2011		4.38	+	849,062	2,023	0.238	23	3	1.5
	1391180910L	red fox	2010		3.86	+	2,401,009	8,487	0.3534	10	3	1
	EBLV-1	<i>Eptesicus serotinus</i>	2018			+	460,751	178,828	38.812	32,232	1	571
	IV	1045120899L	as above		as above		16.8	+	2,697,616	2,560,850	94.930	1,342,569
RABV-enriched approach	965180404L			C – TRIzol/chloroform/ ethanol + virus enrichment	56.5	+	2,951,939	2,765,165	93.67	1,503,424	1	41,961
	767121097L				76.6	+	3,106,953	2,871,569	92.42	1,686,411	1	41,697
	1379120910L				55.2	+	754,635	711,112	94.23	372,141	1	11,346
	1321180108L				25.1	+	1,115,667	951,915	85.322	557,949	3	5,962
	1525180711L				75.5	+	869,224	824,305	94.832	463,772	1	12,144
	1739120912L				48.5	+	682,577	645,311	94.54	362,514	1	9,500
	1996181013L*				65.4	+	760,390	718,626	94.507	423,322	1	10,144
	1577121111L				81.4	+	904,532	846,126	93.54	507,207	1	12,122
	1992121113L				75.9	+	978,751	922,826	94.286	525,370	1	13,536
	1391180910L				47.3	+	985,144	943,962	95.819	539,481	1	13,075
	2191180915L	red fox	2015		93	+	967,059	908,214	93.915	528,055	1	13,089
	1990121113L	red fox	2013		70.2	+	893,272	834,573	93.428	464,788	1	12,217
	2176120515L	red fox	2015		87.6	+	1,294,593	1,196,143	92.395	737,332	1	17,535
	2068180814L	red fox	2014		69.6	+	1,192,314	1,112,910	93.340	65,4514	1	16,190
	2214181115L	red fox	2015		87.6	+	1,230,082	1,140,045	92.680	709,820	1	16,316
	2067120814L	red fox	2014		62	+	1,265,627	1,150,880	90.933	703,662	1	17,091
	2066120814L	red fox	2014		68.4	+	1,934,161	1,741,947	90.062	1,062,064	1	25,144
2226120916L	red fox	2016	84	+	1,133,386	1,011,316	89.229	633,117	1	14,546		
2235181116L	red fox	2016	103	+	1,947,768	1,828,286	93.865	1,119,736	1	26,791		
2236181216P	dog	2016	81.2	+	612,073	564,607	92.245	346,010	1	7,980		
2237120117L	red fox	2017	96.4	+	1,336,571	1,225,265	91.672	758,770	1	16,869		
2238181117K	cat	2017	62	+	1,034,833	918,601	88.768	562,828	1	12,576		

\* – brain samples at a heavy decomposition stage; + – positive; – – negative

After 2 h incubation at 50°C, the RT was terminated by heating at 70°C for 15 min and chilled on ice. PCR was performed for 10 µL of RT product added to 40 µL of reaction mixture containing 5 µL of buffer, 1 µL of dNTPs 10 mM, 1 µL of TaKaRa PrimeSTAR GLX DNA polymerase (TaKaRa Bio) and 33 µL of water for molecular biology. The reactions were carried out in a ProFlex thermocycler (Thermo Fisher Scientific) with the following programme: 1 cycle at 98°C for 4 min, followed by 40 cycles at 98°C for 20 s, 55°C for 30 s, and 72°C for 5 min. Products of amplification were detected by separation in 1% agarose gel.

In cases of weak or no signal from expected amplicons, hemi-nested PCRs were performed. Using the first round residues as a template, the reaction was carried out with adequate primers and slightly modified cycling: 35 cycles at 98°C for 20 s, 55°C for 30 s, and 72°C for 2 min.

**II – One-step RT-PCR.** The protocol was based on One-Step SuperScript III RT-PCR Kit usage. The reaction was performed in 25 µL of mixture containing 2.5 µL of RNA, 12.5 µL of 2X reaction buffer, 1 pmol of each primer, 7 µL of water for molecular biology and 1 µL of SuperScript III/Platinum Taq enzyme mix. Again, the reactions were carried out in a ProFlex thermocycler. For the amplification of A, B or C fragments the following programme was applied: 1 cycle of reverse transcription at 50°C for 30 min, 1 cycle at 95°C for 15 min, 40 cycles at 95°C for 30 s,

55°C for 30 s, and 72°C for 5 min, and final elongation at 72°C for 10 min. The smaller products A1, A2, B1, B2, C1, and C2 were amplified under the following conditions: 1 cycle of reverse transcription at 50°C for 30 min, 1 cycle at 95°C for 15 min, 35 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 3 min, and final elongation at 72°C for 10 min. The amplicons were visualised under UV after separation in 1% agarose gel.

**Real-time RT-PCR (rtRT-PCR).** rtRT-PCR was performed to assess relative viral load based on the detection of over 100 bp of the N gene (Ct value). The reaction was performed as described previously (11, 29).  
**High-throughput sequencing**

**DNA pre-treatment and evaluation.** After the reverse transcription, dsDNA clean-up was performed with AMPure XP magnetic beads (Beckman Coulter, USA). For the purpose of discarding DNA fragments shorter than 1,000 bp, a 0.5:1 bead-to-sample ratio was applied.

The quantity and quality (A260/280 and A230/280) of DNA was measured with the use of a Qubit 3.0 fluorimeter and dsDNA HS Assay Kit, (Thermo Fisher Scientific) and NanoDrop One spectrophotometer (Thermo Fisher Scientific), respectively. In addition, the integrity of the RT product was assayed by capillary electrophoresis using a 5200 Fragment Analyser with a DNF-488 High Sensitivity Genomic DNA Analysis Kit (Agilent, USA). The samples which passed quality control were then normalised to equal concentrations.

**Table 2.** Primers employed for RT-PCR of field RABV samples

Amplicon	Primer name	Primer sequence 5'-3'	Location in RABV genome	Amplicon size
A	RVA_forward	ATGGATGCCGACAAGATTGTATT	1–23	4499
	RVA_reverse	CAGGGGGTGCATCAGGGGAAT	4478–4499	
B	RVB_forward	ATCCCAGAGATGCAATCATCC	4418–4439	3860
	RVB_reverse	TGAGTAGAATGGTAGGACTGGCACC	8251–8276	
C	RVC_forward	GAACCCAGATCTTGGAGAGAGAA	8172–8195	3631
	RVC_reverse	TTCGGATTCAAGATCTTGTTTT	11779–11801	
A1	RVA_forward	as above	2241–2267	2267
	RVA1_reverse	TGGAATTTCTTGGAAATTGGCCAAAGC		
A2	RVA2_forward	GCTCATGACGGATCCAAACTCCC	2193–2216	2300
	RVA_reverse	as above		
B1	RVB_forward	as above	6724–6750	2330
	RVB1_reverse	GATTCAGGAATCTCAAAGATTTGCGT		
B2	RVB2_forward	TTGACTCCTTATATCAAAAACCCAGA	6640–6665	1636
	RVB_reverse	as -above		
C1	RVC_forward	as above	10155–10188	2016
	RVC1_reverse	GTCATGGTTCTAGCTGCATGGCG		
C2	RVC2_forward	ATGAGGCAGGTGCTGGGTG	10054–10073	1750
	RVC_reverse	as above		

**Library preparation.** HTS libraries were prepared from 1 ng of dsDNA, according to the Nextera XT (Illumina, USA) protocol. The dual indexing system (Illumina) was used to label the samples uniquely. The libraries were then cleaned up with the use of AMPure XP magnetic beads (Beckman Coulter) at 0.8:1 ratio, removing fragments smaller than 300 bp. The quality and quantity of libraries were checked with the Qubit 3.0 fluorimeter and dsDNA BR Assay Kit, and an NGS DNF-473 Fragment Kit (Agilent), respectively. Each library was normalised with the use of library normalisation (LN) beads in the Nextera XT DNA Library Prep Kit (Illumina), then pooled and diluted to 20 pM concentration. PhiX Control v3 (Illumina) at 1% was used as an internal control for sequencing. Pair-end sequencing (2 × 300 bp) was performed on a MiSeq sequencer (Illumina) with a v3 kit (Illumina). A 10% portion of run capacity was dedicated to environmental samples of RABV and a 3% portion to amplified RABV genomes.

**Bioinformatics.** The quality check was done by FastQC. Data was trimmed by Trimmomatic; the operation consisted of removal of low quality reads (PHRED score below 33) and reads shorter than 36 bp (2). Non-viral data was filtered by BBDuk with three different approaches: positive filtration of virus reads based on the Kraken database, positive filtration according to the RABV reference sequence, and negative filtration of host reads. Evaluation of RABV data was performed in both Kraken and Centrifuge software (30, 14). Cleaned RABV data was then assembled *de novo* by metaSPAdes software (19).

## Results

In order to evaluate complete genome sequencing of field RABV isolates from brain samples, three RNA extraction methods were compared to select the most efficient for the metagenomic approach. The efficiency of RNA extraction was evaluated by the comparison of dsDNA properties: post-clean-up concentration (estimated by fluorimeter) and RABV genome integration (estimated by capillary electrophoresis). Measurements of dsDNA concentration on a spectrophotometer before clean-up is strongly discouraged due to misrepresentative dsDNA concentrations caused by the residue of reverse transcription. Approximate determination of dsDNA concentration on the spectrophotometer is possible after the clean-up procedure. During the study, HTS libraries were created from samples of dsDNA when the results of the quality check were positive. Additionally, proper size distribution, adapter remains, and quantity of libraries were evaluated. If these parameters were satisfactory, Illumina sequencing was carried out.

Testing of different RNA extraction procedures of the field samples began with the QIAmp Viral RNA Mini Kit and moved on to the Direct-zol RNA MiniPrep

Kit. Procedure testing revealed a higher concentration of dsDNA in samples processed with the Direct-zol RNA MiniPrep Kit – group II – than in samples of which the RNA was purified with the QIAmp Viral RNA Mini Kit – group I (Table 1). In the next step, HTS libraries were prepared, and if the quantity and quality of the library was sufficient, deep sequencing was performed. Unfortunately, the number of viral reads was insufficient to obtain full-length sequences with reliable coverage of all RABV isolates described in groups I and II. The isolate 1379120910L was successfully full-length sequenced with average coverage of 152, however, it was the sample kept the shortest in storage (over 8 years) of all samples in groups I and II, which have collection dates between 1997 and 2008. Due to the long storage period, RNA was considerably fragmented (capillary electrophoresis data not shown) and this negatively affected library preparation and ultimately the results of the metagenomics studies.

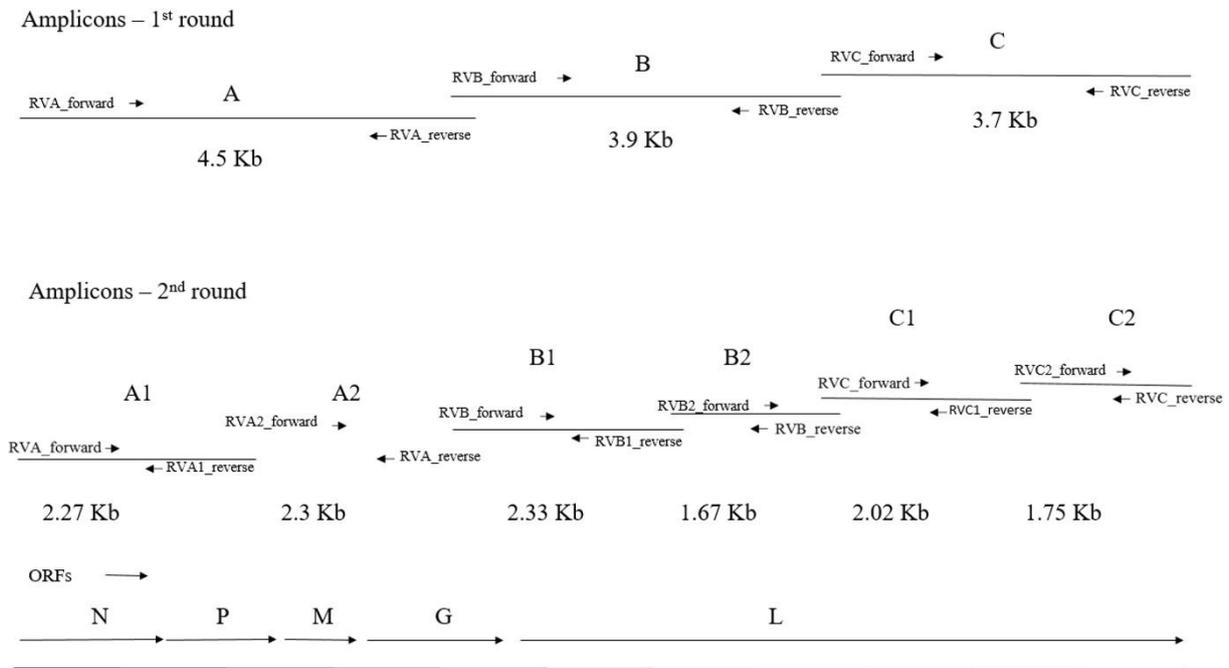
In the next step, the combined method of RNA isolation utilising TRIzol/chloroform/ethanol extraction and RNA purification on a column – group III – was subjected to testing. To assess the concentration of RABV in entire samples, a real-time RT-PCR was performed simultaneously. The dsDNA concentration of reverse-transcribed RNA from the combined TRIzol and column method was significantly higher than that of dsDNA obtained from Direct-zol RNA MiniPrep extracted equivalents. The relative Ct values ranged between 13.77 and 18.58 (data not shown), suggesting a high concentration of viral RNA. The number of total reads was significantly higher compared to the sequencing results obtained from group II. Nevertheless, the percentage of viral reads was much lower and complete RABV genomes were obtained for four out of seven brain samples with average coverage between 13 and 17 for three RABV isolates and coverage of 495 for the isolate 1996181013L (Table 1, group III). The isolate EBLV-1 propagated in the cell culture was deeply sequenced with viral reads numbering 178,828 (38.812% of total reads), of which 32,232 reads were from European bat lyssavirus with the average coverage of 571.

The two commercial kits exploited different techniques, one being based on the digestion of the brain homogenate with lysis buffer and carrier RNA and the second applying TRIzol and ethanol (95–100%). A comparison of the two RNA extraction procedures revealed higher quality and quantity of extracted RNA when the second was used. Therefore, a modification of the TRIzol method was made with chloroform and ethanol at 75% and purification on the column was staged next. Quality and quantity gains were achieved in the RNA extracted.

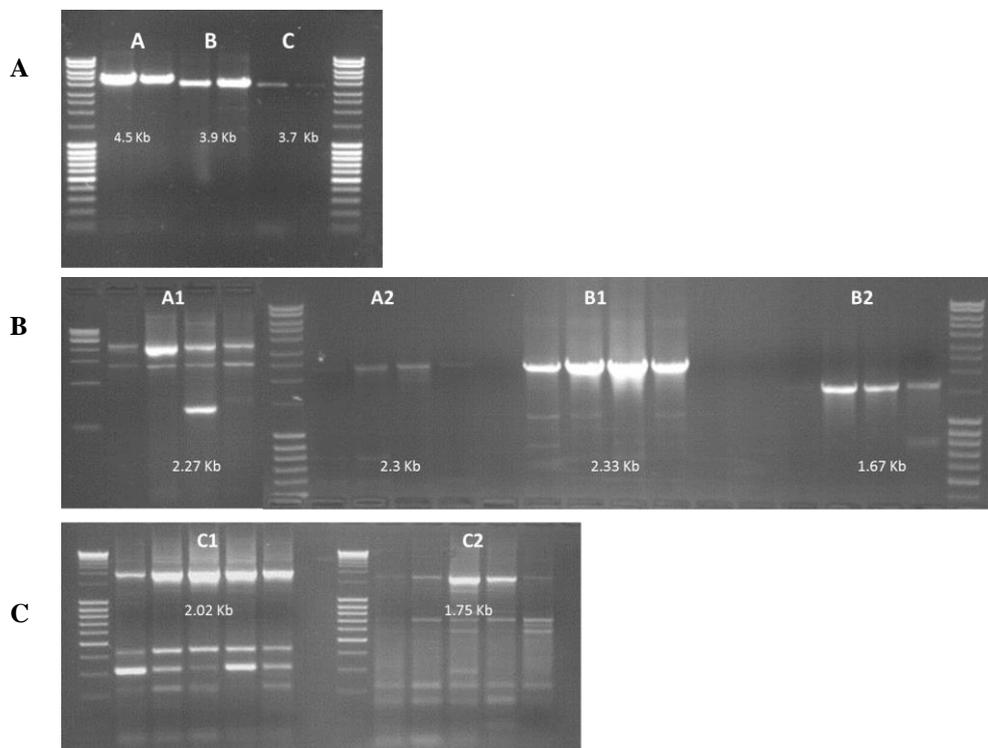
In the last of the investigated approaches, the combined extraction method followed by RABV enrichment was conducted – group IV. Specially designed primers (Table 1) were used in the RABV amplification. Initially, RT-PCR products up to 4.5 kb in

size were obtained, according to the scheme of overlapping amplicons (A, B, and C) as illustrated in Fig. 1. Due to the low yield of DNA polymerases, in the next step shorter products (A1, A2, B1, B2, C1, and C2) were amplified by RT-PCR, ranging in size from 1.67 kb to 2.33 kb. Typical RT-PCR products generated during the study are illustrated in Fig. 2.

A combination of amplification products covering complete genomes of RABV isolates was subjected to HTS. Full-length sequences were obtained for all 23 subjected samples. In this group the number of viral reads was significantly higher than that of group III. The average coverage of consensus sequences ranged between 5,968 and 41,961.



**Fig. 1.** A schematic of the RABV genome indicating position of five ORFs and primers



**Fig. 2.** Gel analysis of representative purified products of amplification. A – first round (amplicons A, B, and C). B – second round (amplicons A1, A2, B1, B2, C1, and C2)

## Discussion

High-throughput sequencing has been widely used for the characterisation of many pathogens, including viruses, bacteria, and parasites (10, 21–23). The direct metagenomic approach enables the study of the structure of a whole microbiome: bacteria, fungi, and all viruses present in an environmental sample where the host material is present as contamination. Whole metagenome sequencing provides a total approach for direct detection of specific viruses and makes possible an accurate survey of the virus structure. The main problem in this approach is the relatively small quantity of viral RNA compared to contamination by host and bacterial material.

Many HTS platforms and RNA preparation protocols were established for the WGS of RABV, however, the vast majority of samples were collected directly from fresh subjects without any decomposition signs or were tissue culture-propagated viruses like vaccine strains (7, 8). Relatively often, difficulties are faced when metagenomic studies are conducted on field samples collected several days after an animal's death. An additional obstacle which has to be taken into consideration in RABV research is the quality of material, which is often highly disintegrated due to prolonged contact with the environment and abundant presence of RNases.

Obtaining viral nucleic acids from the sample is a crucial prerequisite for successful pathogen detection. Therefore, three different methods of RNA extraction from the brain samples collected in the field were tested in this study in terms of their applicability for the metagenomic approach. Salient results evincing high quantity and quality of RNAs (data not shown) were obtained using the combined method of RNA extraction including initial sample lysis with TRIzol and extraction with chloroform and 75% ethanol, which bettered both column-based methods of RNA extraction (the QIAmp Viral RNA Mini and the Direct-zol RNA MiniPrep kits). It should consequently be assumed that treating brain samples with TRIzol significantly increases the efficacy and reliability of RNA extraction. The brain sample constitutes a difficult template for RNA isolation, mainly due to the high concentration of fats. Organic solvents sufficiently dissolve fats, improving of the efficacy of RNA isolation.

It is apparent that for indirect pathogen detection based on the presence of a gene fragment, it is most effective to extract the RNA using standard column-based kits, due to a faster and easier isolation procedure. But for high-throughput sequencing, the integrity of the genetic material is key for library preparation, and standard RNA extraction methods do not always provide sufficient quality of RNA for this approach. This is particularly true in the case of RNA viruses, which are much more sensitive to environmental conditions due to the fragile structure of RNA. Our results and suggestions correlate

with the findings of Wylezich *et al.* (31) that efficient RNA extraction is crucial in metagenomics studies.

Preliminary estimation of viral load in the samples was determined using RT-PCR for the comparison of Ct values between different RABV isolates. This comparison indicated high concentrations of RABV in the samples extracted with Direct-zol RNA MiniPrep and the combined TRIzol/column method (groups II and III). This prediction was not fully reflected in success in deep sequencing, probably because the estimation provided by RT amplification was too approximate. In addition, RT-PCR detection is based on the amplification of short fragments of a viral genome, making this method less sensitive to sample fragmentation. During deep sequencing, in our study, a large number of total reads and viral reads were obtained (mean 3.64%), and fragments of RABV nucleotide sequences were detected, but it was not possible to determine full-length consensus sequences. Even if a consensus was obtained, the average coverage of contigs was too low (below 3.5). Metagenomic studies allow direct detection of pathogens but are characterised by a defined detection limit in terms of sequencer throughput and percentage throughput per sample.

Due to the nature of a viral cycle based on a host cellular system, the main obstacle with all viral metagenomic studies is a low viral load. It is important to understand that viral genomic material constitutes only a small fraction of all extracted RNA, where the overwhelming majority of such material will be high background from host species and bacteria. To overcome this issue different strategies of target enrichment may be applied.

The first recommended solution is propagation of viruses in cell culture before HTS. A good HTS result was obtained for EBLV-1 isolate cultivated in a neuroblastoma cell line. However, it is not always possible to multiply viruses from decomposed field samples when RABV is not able to infect cells or be isolated in the cell culture (16). Indeed, only three out of six RABV isolates originating from brain samples collected in the field were able to propagate in cell culture, moreover, this was only at low titres (data not shown). Passaging viruses in a cell culture system adds new artificial diversity to a viral population. The extent of alteration to the original consensus sequence of a RABV population depends on the number of passages necessary to obtain virus at a high enough titre to harvest (3). The finding was previously published that the number of single-nucleotide polymorphisms (SNPs) observed in cell-cultured RNA preparations were greater than those in tissue-extracted samples (16). Therefore, virus generation in cell culture prior to WGS should be highly constricted when performing studies on phylogeography of the population or on genomic diversity or virus evolution.

The second approach to viral enrichment is amplification of the whole RABV genome of 12 kb. This tactic results in a significant increase in specific RABV reads and consequently much greater coverage, however,

it is limited to HTS of known pathogens only. Target enrichment, *i.e.* the amplification of viral RNA in a long-range PCR using specific primers designed for the detection of RABV, was a much more effective method of RNA preparation for HTS in our study. Primers were designed based on previously sequenced Polish RABV isolates. Two RT-PCR protocols were applied for the amplification of the RABV viral genome, of which three fragments of 4.5 kb, 3.8 kb, and 3.7 kb were amplified, covering almost 12 kb. We found that the TaKaRa PrimeSTAR GXL DNA polymerase could amplify longer amplicons than the SuperScript III One-Step RT-PCR Kit. However, there was no difference in RT-PCR results when smaller fragments of viral genome were amplified (A1, A2, B1, B2, C1, and C2) with molecular weight of DNA ranging between 1.6 kb and 2.3 kb. All tested field samples were successfully amplified using both protocols: either the two-step RT-PCR with TaKaRa PrimeSTAR GLX DNA polymerase or, for fragments shorter than 2.5 kb, the SuperScript III One-Step RT-PCR Kit. The amplification of shorter fragments of the genome is much easier, whereas for long-range PCR reactions high-fidelity and high-yield polymerases are required. Hence in our study, for the detection of fragments around 2–2.5 kb, a single-tube one-step RT-PCR with a mix of SuperScript III/Taq DNA polymerase enzymes was sufficient. It is paramount to take into consideration the occurrence of random errors generated during both reverse transcription and amplification, therefore, it is vital to use high-fidelity polymerases with high DNA replication accuracy to minimise amplification errors. High-fidelity amplification is essential for experiments of which the outcome depends upon the correct DNA sequence, *e.g.* cloning, single-nucleotide polymorphism (SNP) analysis, and HTS applications.

The HTS methodology described in this paper facilitated obtaining complete genomes of several RABV isolates originating from the brain tissue of animals collected in the field. Significantly greater rates of RABV genome coverage were obtained with the RABV-enriched approach. However, metagenomic studies enabled full-length sequencing of 6 out of 16 field viruses including EBLV-1 propagated in a neuroblastoma cell line. The direct metagenomic approach provides information on original genome sequences, but with lower coverage. Sequencing coverage describes the average number of reads that align to, or “cover,” known reference bases (12); if the coverage value for viral reads is over 20, it provides reliable nucleotide sequences. The enriched approach gives greater coverage, but with the risks of genome modification and artificial diversity caused by PCR amplification. Complete viral sequences with sufficient coverage provide the ability to discriminate between isolates that are very closely related both genetically and geographically. The application of such a powerful tool in rabies cases is crucial for better understanding of the outbreak as well as for implementing more effective rabies control strategies.

In conclusion, the study describes the comparison of two approaches to the HTS of field rabies viruses. The crucial issues are summarised here which should be considered before deep sequencing. Direct metagenomics offers the most realistic illustration of a microbiome and is a straightforward approach for surveying a viral community in environmental samples. Major issues that have to be overcome are high sequencing depth due to host contamination, insufficient viral load in original samples, and higher detection limits compared to amplification-based methods. Low quality of the samples results in a low number of total reads, decreases the sequencing efficiency, and increases total costs. To overcome those problems enrichment techniques may be applied: removal of host material, *e.g.* enzymatic digestion; amplification of target sequences; ultracentrifugation of viral particles; or accumulation of the viral load *via* cell culturing.

During the study a set of recommendations for sequencing RABV samples were derived. Careful sample processing is crucial for successful library preparation and sequencing. Appropriate storage and preservation of collected material and employment of a pretreatment method (digestion of host genetic material or ultracentrifugation) significantly increases the number of viral reads. An appropriate nucleic acid extraction method and control of RNA/DNA parameters, both of concentration (fluorimeter) and integrity (capillary electrophoresis) during each stage of sample processing are imperative for effective library preparation and sequencing.

If deep characterisation of viruses is intended, *e.g.* for spatial and temporal phylogeography of viral populations during outbreaks, target enrichment followed by deep sequencing is also recommended as it generates much greater coverage of obtained consensus sequences.

**Conflict of Interests Statements:** The authors declare that there is no conflict of interests regarding the publication of this article.

**Animal Rights Statement:** None required.

**Financial Disclosure Statement:** The authors declare that this study was funded by the KNOW (Leading National Research Centre) Scientific Consortium “Healthy Animal – Safe Food”, Ministry of Science and Higher Education Resolution No. 05-1/KNOW2/2015).

## References

1. Biek R., Real L.A.: The landscape genetics of infectious disease emergence and spread. *Mol Ecol* 2010, 19, 3515–3531.
2. Bolger A.M., Lohse M., Usadel B.: Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, 2014, 30, 2114–2120.
3. Boukadida C., Torres-Flores J.M., Yocupicio-Monroy M., Piten-Isidro E., Rivero-Arrieta A.Y., Luna-Villalobos Y.A., Martínez-Vargas L., Alcaraz-Estrada S.L., Torres K.J., Lira R., Reyes-Terán G., Sevilla-Reyes E.E.: Complete genome sequences, before and

- after mammalian cell culture, of Zika virus isolated from the serum of a symptomatic male patient from Oaxaca, Mexico. *Genome Announc* 2017, 5, e00072-17, doi: 10.1128/genomeA.00072-17.
4. Bourhy H., Kissi B., Audry L., Smreczak M., Sadkowska-Todys M., Kulonen K., Tordo N., Żmudziński J.F., Holmes E.: Ecology and evolution of rabies virus in Europe. *J Gen Virol* 1999, 80, 2545–2557.
  5. Bruncker K., Marston D., Horton D.L., Cleaveland S., Fooks A.R., Kazwala R., Ngeleja C., Lembo T., Sambo M., Mtema Z.J., Sikana L., Wilkie G., Biek R., Hampson K.: Elucidating the phylodynamics of endemic rabies virus in eastern Africa using whole-genome sequencing. *Virus Evol* 2015, 1, 1–11.
  6. Cheval J., Sauvage V., Frangeul L., Dacheux L., Guigon G., Dumey N., Pariente K., Rousseaux C., Dorange F., Berthet N., Brisse S., Moszer I., Bourhy H., Manuguerra C.J., Lecuit M., Burguiere A., Caro V., Eloit M.: Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. *J Clin Microbiol* 2011, 49, 3268–3275, doi: 10.1128/JCM.00850-11.
  7. Cliquet F., Picard-Meyer E., Mojzisz M., Dirbakova Z., Muizniece Z., Jaceviciene I., Mutinelli F., Matulova M., Frolichova J., Rychlik I., Celer V.: In-depth characterization of live vaccines used in Europe for oral rabies vaccination of wildlife. *PLoS One*, 2015, 10, e0141537, doi: 10.1371/journal.pone.0141537.
  8. Dedkov V.G., Lukashov A.N., Deviatkin A.A., Kuleshov K.V., Safonova M.V., Poleshchuk E.M., Drexler J.F., Shipulin G.A.: Retrospective diagnosis of two rabies cases in humans by high throughput sequencing. *J Clin Virol* 2016, 78, 74–81, doi: 10.1016/j.jcv.2016.03.012.
  9. Drummond A.J., Pybus O.G., Rambaut A., Forsberg A., Rodrigo A.G.: Measurable evolving populations. *Trends Ecol Evol* 2003, 18, 481–488.
  10. Elson R., Awofisayo-Okuyelu A., Greener T., Swift C., Painset A., Amar C.F.L., Newton A., Aird H., Swindlehurst M., Elviss N., Foster K., Dallman T.J., Ruggles R., Grant K.: Utility of whole genome sequencing to describe the persistence and evolution of listeria monocytogenes strains within crabmeat processing environments linked to two outbreaks of listeriosis. *J Food Prot* 2019, 82, 30–38, doi: 10.4315/0362-028X.JFP-18-206.
  11. Hoffmann B., Freuling C.M., Wakeley P.R., Rasmussen T.B., Leech S., Fooks A.R., Beer M., Müller T.: Improved safety for molecular diagnosis of classical rabies viruses by use of a TaqMan real-time reverse transcription-PCR “double check” strategy. *J Clin Microbiol* 2010, 48, 3970–3978, doi: 10.1128/JCM.00612-10.
  12. <https://www.illumina.com/science/education/sequencing-coverage.html>.
  13. Johnson N., McElhinney L.M., Smith J., Lowings P., Fooks A.R.: Phylogenetic comparison of the genus *Lyssavirus* using distal coding sequences of the glycoprotein and nucleoprotein genes. *Arch Virol* 2002, 147, 2111–2123.
  14. Kim D., Song L., Breitwieser F.P., Salzberg S.L.: Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res*. 2016, 26, 1721–1729.
  15. Lewandowska D.W., Zagordi O., Geissberger F., Kufner V., Schmutz S., Böni J., Metzner K.J., Trkola A., Huber M.: Optimization and validation of sample preparation for metagenomic sequencing of viruses in clinical samples. *Microbiome* 2017, 5, 1–13, <https://doi.org/10.1186/s40168-017-0317-z>.
  16. Marston D.A., McElhinney L.M., Ellis R.J., Horton D.L., Wise E.L., Leech S.L., David D., de Lambellerie X., Fooks A.R.: Next generation sequencing of viral RNA genomes. *BMC Genomics* 2013, 14, 444.
  17. McElhinney L.M., Marston D.A., Brookes S.M., Fooks A.R.: Effects of carcass decomposition on rabies virus infectivity and detection. *J Virol Methods* 2014, 207, 110–113.
  18. Nadin-Davis S.A., Colville A., Trewby H., Biek R., Real L.: Application of high-throughput sequencing to whole rabies viral genome characterization and its use for phylogenetic re-evaluation of a raccoon strain incursion into the province of Ontario. *Virus Res* 2017, 232, 123–133.
  19. Nurk S., Bankevich A., Antipov D., Gurevich A., Korobeynikov A., Lapidus A., Prjibelsky A., Pyshkin A., Sirotkin A., Sirotkin Y., Stepanauskas R., McLean J., Lasken R., Clingenpeel S.R., Woyke T., Tesler G., Alekseyev M.A., Pevzner P.A.: Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In: *Research in Computational Molecular Biology. RECOMB 2013. Lecture Notes in Computer Science*, vol 7821, edited by M. Deng, R. Jiang, F. Sun, X. Zhang, Springer, Heidelberg, 2013, pp 158–170.
  20. Orłowska A., Żmudziński J.F.: Molecular epidemiology of rabies virus in Poland. *Arch Virol* 2014, 159, 2043–2050.
  21. Picard-Meyer E., Robarded E., Moroz D., Trotsenko Z., Drozhzhe Z., Biarnais M., Solodchuk V., Smreczak M., Cliquet F.: Molecular epidemiology of rabies in Ukraine. *Arch Virol* 2012, 157, 1689–1698.
  22. Reis-Cunha J.L., Baptista R.P., Rodrigues-Luiz G.F., Coqueiro-Dos-Santos A., Valdivia H.O., de Almeida L.V., Cardoso M.S., D’Ávila D.A., Dias F.H.C., Fujiwara R.T., Galvão L.M.C., Chiari E., Cerqueira G.C., Bartholomeu D.C.: Whole genome sequencing of *Trypanosoma cruzi* field isolates reveals extensive genomic variability and complex aneuploidy patterns within TcII DTU. *BMC Genomics* 2018, 19, 816, doi: 10.1186/s12864-018-5198-4.
  23. Siira L., Naseer U., Alfsnes K., Hermansen N.O., Lange H., Brandal L.T.: Whole genome sequencing of *Salmonella chester* reveals geographically distinct clusters, Norway, 2000 to 2016. *Euro Surveill* 2019, 24, doi: 10.2807/1560-7917.ES.2019.24.4.1800186.
  24. Simon B., Pichon M., Valette M., Burfin G., Richard M., Lina B., Josset L.: Whole genome sequencing of A(H3N2) influenza viruses reveals variants associated with severity during the 2016–2017 season. *Viruses* 2019, 11, 108, doi: 10.3390/v11020108.
  25. Steinhauer D.A., Holland J.J.: Rapid evolution of RNA viruses. *Ann rev microbiol* 1987, 41, 409–433.
  26. Tordo N., Poch O., Ermine A., Keith G., Rougeon F.: Walking along the rabies genome: is the large G-L intergenic region a remnant gene? *Proc Natl Acad Sci USA* 1986, 83, 3914–3918.
  27. Tordo N., Poch O., Ermine A., Keith G., Rougeon F.: Completion of the rabies virus genome sequence determination: highly conserved domains among the L (polymerase) proteins of unsegmented negative-strand RNA viruses. *Virology* 1988, 165, 565–576.
  28. Tricou V., Berthet N., Nakoune E., Kazanji M.: Complete genome sequence of a rabies virus isolated from a human in Central African Republic. *Genome Announc* 2014, 2, e00598–e005914.
  29. Wakeley P.R., Johnson N., McElhinney L.M., Marston D., Sawyer J., Fooks A.R.: Development of a real-time, TaqMan reverse transcription-PCR assay for detection and differentiation of lyssavirus genotypes 1, 5, and 6. *J Clin Microbiol* 2005, 43, 2786–2792.
  30. Wood D.E., Salzberg S.L.: Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. 2014, 15, R46.
  31. Wylezich C., Papa A., Beer M., Höper D.: A versatile sample processing workflow for metagenomic pathogen detection. *Sci Rep* 2018, 30, 13108, doi: 10.1038/s41598-018-31496-1.