Characterization of an integrated, endogenous mouse mammary tumor virus-like (MMTV) betaretrovirus genome in a black Syrian hamster *(Mesocricetus auratus)*

Katalin B. Horváth[a,b], Ákos Boros[a,b], Endre Kálmán[c], Péter Pankovics[a,b], Eric Delwart[d,e], Gábor Reuter^*

[a] *Department of Medical Microbiology and Immunology, Medical School, University of Pécs, Pécs, Hungary*

[b] *Regional Laboratory of Virology, National Reference Laboratory of Gastroenteric Viruses, ÁNTSZ Regional Institute of State Public Health Service, Pécs, Hungary* [c] *Department of Pathology, Medical School, University of Pécs, Pécs, Hungary* [d] *Blood Systems Research Institute, San Francisco, CA, USA* [e] *University of California, San Francisco, CA, USA*

* Corresponding author at: Department of Medical Microbiology and Immunology, Medical School, University of Pécs, H-7624 Szigeti út 12, Pécs, Hungary. *E-mail address:* reuter.gabor@gmail.com (G. Reuter).

Retroviruses (family *Retroviridae)* are important agents of humans and animals. This study reports the detection and complete genome characterization of a novel endogenous retrovirus from the black Syrian hamster (*Mesocricetus auratus)* with a squamous cell skin tumor. The proviral genome, tentatively named black Syrian hamster retrovirus (BSHRV/2013/HUN, MK304634), was 8784 nucleotide in length with typical full-length betaretrovirus genome organization of 5'LTR-gag-pro-pol-env-3'LTR and with a characteristic mouse mammary tumor virus-like (MMTV) betaretrovirus dUTPase domain but without a sag gene. The BSHRV gag (534aa), pro/ pol (~1099aa) and env (672aa) proteins had 56%/63%/50% aa identity to the corresponding proteins of MMTV (AF228552). The proviral DNA is detectable in tumor as well as in tumor-free cells by conventional PCR and qPCR but only visible in the tumor cells by *in situ* hybridization. Low level retroviral RNA expression was found only in the DNase-treated RNA tumor samples using RT/nested PCR. BSHRV/2013/HUN-like betaretrovirus DNA was also identified from a faecal and tissue samples from 1 of the further 3 tested individuals by nested-PCR and qPCR. Further research is needed to investigate the distribution, activity and etiological role of this novel MMTV-like betaretrovirus species in hamster.

1. Introduction

The members of the family *Retroviridae* are important pathogens of humans and animals. The ability of some retroviruses to induce tumors has been known since the turn of the 20th century. In 1908, Ellerman and Bang described a chicken erythroleukemia that was caused by a retrovirus followed by isolation of Rous sarcoma virus from a chicken fibrosarcoma by Peyton Rous (Ellerman and Bang, 1908; Rous, 1911). These discoveries mark the beginning of experimentation that led to our current understanding of retroviruses as cancer-causing agents. Subsequent studies extended the general paradigm to mammalian hosts. The discoveries of Bittner and Gross revealed that mouse mammary tumor virus (MMTV) retroviruses were associated with mammary epithelial tumors and thymic lymphomas in mice (Bittner, 1936). The list of animals affected by oncogenic retroviruses expanded as the 20th century progressed to include cats, cows, rats, mice, sheep, goats, several primates, fish and humans (Dudley, 2011; Feller and Chopra, 1969; Wang et al., 1995).

Endogenous retroviral elements comprise 8% of the total human genome (Lander et al., 2001). Modern-day exogenous retroviruses, as well as the infectious predecessors of endogenous retroviruses (ERVs), are demonstrably oncogenic (Young et al., 2013; Gonzalez-Cao et al., 2016). Further, replication-competent ERVs contribute to cancer development (Kassiotis and Stoye, 2017). Moreover, human cancers are characterized by transcriptional activation of ERVs. These observations combine to incriminate ERVs as potential causative agents of some cancers, too (Kassiotis, 2014).

MMTV, formerly known as Bittner agent, has a causal role in the majority of mouse mammary tumors in feral and experimental mice (Ross, 2008). MMTV is a prototype species of the genus *Betaretrovirus* in

Table 1

List of oligonucleotide primers used in this study. ISH: *in situ* hybridization, AS: anti-sense probe, S: sense probe, Lowercase letters in 5'-3' sequence column: T = RNA polymerase recognition sites.

Application	Primer name	5'-3' sequence	PCR-product size (bp)
1st PCR round	HamREV-972-F-out	CTGCAGTCCCGCCAGTGAAGATAG	672
1st PCR round	HamREV-1643-R-out	CTCTGCCAATGCCCCTTGAA	
2nd PCR round (nested)	HamREV-1036-F-in	ACTCGGGCTCAAAGGGACAGAAAA	587
2nd PCR round (nested)	HamREV-1622-R-in	GCCTATGGGTGGCAGGTATTC	
ISH AS probe production	HamREV-2847-F-ISH	GCTCTTGGGAAAGAGCTCCA	191
ISH AS probe production	HamREV-3038-R-T7-ISH	taatacgactcactatagggGCCTGTAGCTCAGGATTGGG	
ISH S probe production	HamREV-6242-F-T7-ISH	taatacgactcactatagggTTACATCCGGTGGTTTGGGG	193
ISH S probe production	HamREV-6435-R-ISH	CTGACAGGCAAACAACCAGC	
qPCR standard	HamREV-2540-R	TTTACATCTTGGGCAAATGCCA	428
qPCR standard	HamREV-2113-F	CAGGGATTCARCAAGGCAATGA	
qPCR reaction	HamREV-2310-F	GTATGTGCAGATGCCTCACC	104
qPCR reaction	HamREV-2413-R	CCTCCCCCATATGTCTGTTTC	

the family *Retroviridae* (ICTV, 2011). Multiple double-stranded MMTV DNA copies are also found in the chromosomal DNA (called integrated or endogenous proviruses) of laboratory mice (Coffin, 1992). The virus is mostly transmitted vertically (endogenous virus) when embryonic germ-line cells are infected (Cohen and Varmus, 1979), but it is also usually transmitted horizontally through maternal milk, as

exogenous virus or milk-borne virus (Bittner, 1936). MMTV-induced mammary tumorigenesis is mediated by proviral integration, usually by enhancer-mediated activation of nearby cellular oncogenes, although coding region insertions that alter the gene product have also been described (Callahan and Smith, 2008; Ross, 2010). Additionally, the MMTV envelope (env) protein has been implicated in mammary tumorigenesis (Katz et al., 2005). The identification of a retrovirus that causes breast cancer in mice created great interest in determining whether similar viruses exist in other species, particularly in humans (Spiegelman et al., 1970; Axel et al., 1972; Mesa-Tejada et al., 1982; Keydar et al., 1984). MMTV-like sequences were identified in humans using nested PCR techniques (Wang et al., 1995; Liu et al., 2001; Lawson et al., 2010): a human endogenous retrovirus K (HERV-K) human mouse mammary tumor virus(HMTV)-like 2 (HML-2) (Hohn et al., 2013; Garcia-Montojo et al., 2018) 90-98% homologous to MMTV has been detected in approximately 40% of breast cancers in American women (Wang et al., 1995) but interpretation of these reports remains controversial.

Hamsters are popular pet animals with Syrian hamsters *(Mesocricetus auratus)* and Djungarian hamsters *(Phodopus sungaros)* being among the most frequently kept species. Reports of spontaneous tumors in domestic hamsters are scarce, and most are individual case reports (Kondo et al., 2008). The incidence of tumors in these two species is quite different. In the integumentary system of Syrian hamsters' melanomas and cutaneous lymphoma are the most frequently detected tumors, but both are low in incidence. However, the incidence of cutaneous neoplasias is five times greater in Djungarian than in Syrian hamsters and these are predominantly represented by mammary tumors, atypical fibromas and papillomas (Jelinek et al., 2013).

In the 1960-1970's there were several reports of the isolation of type-C-like retroviruses from hamster cells including the hamster leukaemia virus (HaLV), the hamster sarcoma virus (HaSV) and the D-9 virus was described (Bassin et al., 1968; Kelloff et al., 1970; Somers et al., 1973). These viruses have been reported to have little or no endogenous reverse transcriptase activity and to be unable to transform murine cells (Kelloff et al., 1970; Verma et al., 1974). In 1979, a hamster retrovirus (HaRV) with typical type-C characteristics was described in the continuous culture of natural hamster melanoma cell line (Y-22). The virus differs from all other known hamster viruses in its ability to transform murine as well as rat and hamster cells with apparent one-hit kinetics (Russell et al., 1979). Since then there was no further data in this research area.

This study reports the identification and characterization of the complete genome of a novel MMTV-like betaretrovirus in an integrated form (provirus) in black Syrian hamsters *(Mesocricetus auratus)by* molecular and *in situ* hybridization methods.

2. Materials and methods

2.1. Specimens

Post mortem tumor specimen and tumor-free tissue samples from the tail, liver, testes and lung were collected from a male black Syrian hamster *(Mesocricetus auratus)* pet animal, euthanized for clinical reasons, and used for molecular and histo/immunopathological investigation (Animal 1).

To acquire cellular nucleic acids in a non-invasive way, fresh faecal specimens which are known to contain intestinal epithelial cells were collected from another three, clinically healthy, 12-21-month-old black Syrian hamsters (n = 2 female and *n* = 1 male) kept as house pets by three different owners (medical students) in the town of Pécs (Hungary), in May and June 2018 to test for the study sequence. One of these animals (a 1.5-year-old female) died - because of old age - on November 22, 2018. Different *postmortem* tissue specimens (skin, lung, ovaries and liver) collected from this animal (Animal 2) were also tested by (RT/) nested PCR (Table 1) for the studied virus.

2.2. *Viral metagenomics*

A tumor sample from the face (Fig. 1) of Animal 1 was subjected to viral metagenomic analysis using random nucleic acid amplification of enriched tissue associated viral particles (Victoria et al., 2009). Briefly, phosphate-buffered saline-diluted (PBS) tissue homogenates were passed through a 0.45 urn sterile filter (Millipore) and centrifuged at 6000 xg for 5 min (Phan et al., 2013). Then the filtrate was treated with a mixture of DNases and RNases to digest unprotected nucleic acids (Phan et al., 2013). Nucleic acids were then extracted (RNA extraction without DNAase step) using the QIAamp spin-column technique (Qiagen, Hilden, Germany) and subjected to sequence independent random PCR/RT-PCR amplification. A cDNA library was constructed from the extracted RNA by ScriptSeqTM v2 RNA-Seq Library Preparation Kit (Epicentre). The random PCR/RT-PCR amplicons from the cDNA library and the extracted DNA were sequenced by the MiSeq Illumina platform according to the manufacturer's instruction, and as described previously (Phan et al., 2013). Paired-end reads of 250 bp generated by MiSeq are debarcoded using Illumina vendor software. Using an in-house analysis pipeline running on a 36-nodes Linux cluster human and bacterial reads were subtracted by mapping to human reference genome hg38 and bacterial nucleotide sequences from nt using Bowtie 2 (Langmead and Salzberg, 2012). Adaptor and primer sequences were trimmed using the default parameters of VecScreen (Ye et al., 2006). The cleaned reads were then de-novo assembled using Ensemble Assembler (Deng et al., 2015). The assembled contigs > 100- bp, along with singlets, were aligned to an in-house viral proteome and nucleotide database using GenBank BLASTx and BLASTn using E-value cut-off of 10^{-10}.

Fig. 1. Gross photography of the affected (euthanized) black Syrian hamster (Animal 1). (A) Primary skin tumor masses are visible on the face, neck and metastasis on the chest area (black arrows). The tumor was circumscribed and white-brown in color. On the cut surface, a large central necrosis circumscribed by homogeneous grayish tissue was apparent. (B) The metastasis on the chest was solid, well circumscribed, measured 0.7 cm in diameter and of white color on cut surface (white arrow). No other metastasis was observed in other parts or organs of the body. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The significant hits to virus were then aligned to an inhouse non-virus-non-redundant universal proteome database using DIAMOND (Buchfink et al., 2015). A web-based graphical user interface was used to show the viral matches, along with taxonomy information and processing meta-information. Metagenomic raw reads were submitted and deposited in the NCBI short read archive (SRA) under Bio-Project PRJNA550124 (Experiment SRX6101754 and Run SRR9335418).

2.3. Complete genome determination and sequencing

The freshly dissected tumor specimen and tumor-free tissue samples from the tail, skin, liver and lung were homogenized separately using a Potter-Elvehjem pestle (Sigma-Aldrich Co., Munich, Germany).

Total RNA and DNA were extracted in separate reactions from approximately 50 mg of freshly homogenized tissue samples and 150150 ulof 40 w/v% faecal suspensions (in 0.1 M PBS) using TRI Reagent (Molecular Research Centre, Cincinnati, OH, USA) for RNA isolation and innuPREP DNA Mini Kit (Analytik Jena, Germany) for DNA isolation according to the manufacturer's instructions. To remove the genomic DNA the extracted total RNA samples were digested with DNase I (Thermo Scientific USA Waltham, MA) in the presence of Mg^2 + for an hour at 37 °C. The efficiency of the DNase treatment was verified by RT-nested PCR without the use of reverse transcriptase enzyme which showed negative results. For the RT-nested PCR outer and inner primer pairs targeting the 5'LTR (long terminal repeat) and P10 junction was used (Table 1).

The complete genome of the black Syrian hamster retrovirus (BSHRV/2013/HUN, GenBank accession number MK304634) was determined by a primer-walking method and PCR techniques using DNA extract from the tumor tissue from Animal 1. Sequence-specific oligo-nucleotide primers were designed using the retroviral reads and contigs determined by viral metagenomics. Both ends of the genome and the flanking

chromosomal regions of the provirus were determined using Thermal Asymmetric Interlaced PCR (TAIL-PCR) method (Liu and Whittier, 1995). The sequences determined by TAIL-PCR and viral metagenomics were verified by conventional and Long Range PCR using reagents and methods described previously (Boros et al., 2012).

PCR products were sequenced directly in both directions using a BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Warrington, UK) with the specific primers and run on an automated sequencer (ABI PRISM 310 Genetic Analyzer; Applied Biosystems, Stafford, USA).

2.4. Sequence- andphylogenetic analysis

ClustalX (version 2.1) and GeneDoc (version 2.7) software were used to align, assemble and compare the study strain and the prototype betaretroviral sequences available in GenBank. The phylogenetic tree from the aligned amino acid sequences of the pol gene was constructed by the neighbor-joining method using MEGA 6 (Tamura et al., 2013). Bootstrap analysis of 1000 replicates was done to measure the significance of branching.

2.5. Integrated viral genome copy number determination using quantitative PCR (qPCR)

For the absolute quantification of integrated viral copy number of tumor, tail, liver and lung tissue samples (10 ng DNA/specimen), the SYBR Green-based qPCR method (Maxima SYBR Green qPCR Master Mix; Thermo Scientific, Waltham, MA, USA) was used according to the manufacturer's instructions. The qPCR reaction mix contained 12.5 ul 2 x Maxima SYBR Green/ROX qPCR mastermix and 0.3-0.3 ulof20 uM strain specific HamREV-2310-F and HamREV-2413-R oligonucleotide primers (gag gene, Table 1) and 10-10 ng of isolated total DNA as templates in the final volume of 25 ul. The qPCR thermal program contained a Pre-Incubation step (95 °C-10 min) followed by a 39 cycles of quantification steps [95°C-15s, 57°C-15s, 72°C-33s] and a dissociation assay and ran both in a plate format on an ABI 7500 qPCR thermal cycler. For the generation of the standard curve, a 10-fold dilution series (2.26E + 06-2.26E + 03) of purified (GeneJet Gel Extraction Kit, Thermo Scientific, Waltham, MA, USA) and spectro-photometrically (NanoDrop ND-1000) quantified 428 bp long PCR product of gag gene were used. The PCR product used for the standard curve includes the target sequences of HamREV-2310-F and HamREV-2413-R oligonucleotide primers (Table 1) of the qPCR reactions. The qPCR assay contained three technical repeats of all samples and standards. The slope of the standard curve was — 3.398550 and the calculated PCR efficiency was 99.92% (Supplementary Fig. S1).

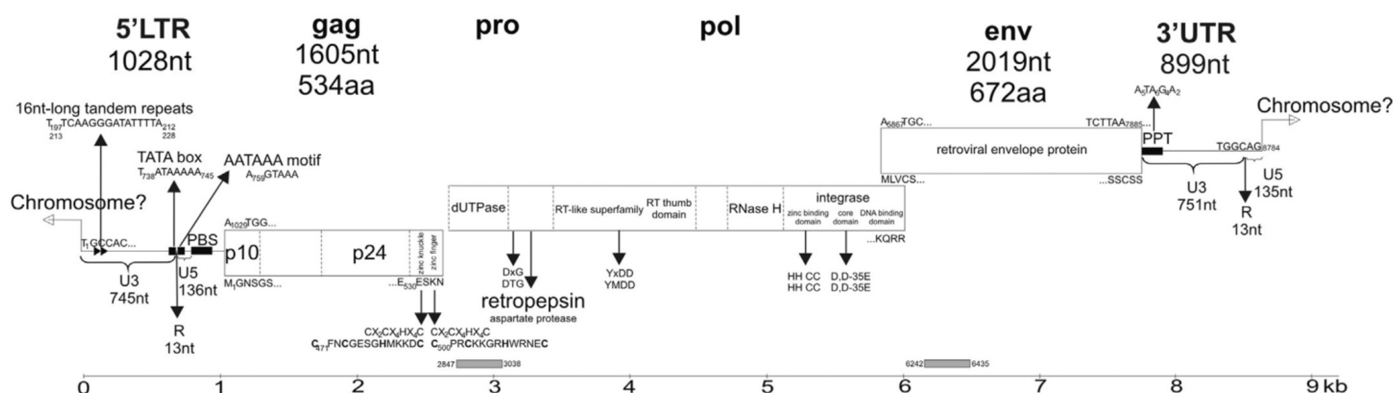

Fig. 2. Schematic organization of the betaretroviral genome of black Syrian hamster retrovirus (BSHRV/2013/HUN, GenBank accession number MK304634). The full length of the provirus is 8784 nt. The sense and antisense ISH probes were located in the presumed dUTPase (sense) and the retroviral envelope protein-coding (antisense) regions (gray lines above the scale). The figure is drawn to scale. Abbreviations: dUTPase=trimeric dUTP (deoxyuridine 5′-triphosphate nucleotide hydrolase) diphosphatase domain; LTR=long terminal repeat; PBS=primer-binding site; PPT=polypurine tract; U3=unique 3 region; R=terminal repeat; U5=unique 5 region.

2.6. Histopathology

Samples (Animal 1) for histopathological examination were fixed in 10% neutral buffered formalin, embedded in paraffin wax, sectioned at 5 urn thick slices and stained with haematoxylin and eosin.

2.7. In situ hybridization

2.7.1. Probe production

Two, 191-nt-long sense, and 193-nt-long antisense Digoxigenin-11-UTP (DIG, Sigma-Aldrich, Saint Louis, MI, USA) labeled RNA probes (Table 1, Fig. 2) were produced using TranscriptAid T7 High Yield Transcription Kit (Thermo Scientific USA Waltham, MA) and T7-pro-moter-tagged silica-column (Qiagen, Germany) purified PCR products as templates. The sense and antisense probes were located in the pre-sumed dUTPase (sense) and the retroviral envelope protein-coding (antisense) regions (Fig. 2).

The produced DIG-labeled RNA probes were purified using the RNeasy Mini Kit (Qiagen, Germany). For the evaluation of the integrity, length and yield of the transcripts the purified samples were run on a 1% agarose gel with ethidium bromide along with RiboRuler™ RNA Ladder (Thermo Scientific USA Waltham, MA). The produced DIG-labeled RNA probes were used for the detection of retroviral DNA and/or RNA using *in situ* hybridization of formalin fixed paraffin embedded slides.

2.7.2. Chromogenic in situ hybridization of FFPE sections

Slides of 10 urn thick formalin-fixed paraffin embedded (FFPE) sections were deparaffinized and rehydrated using xylene and descending concentrations of ethanol (100% to 50%). The rehydrated sections were digested with Proteinase K (Sigma-Aldrich, Saint Louis, MI, USA) at 37 °C for 10 min. After extensive washing with 1 x phosphate-buffered saline (1xPBS, pH 7.4) the samples were post-fixed with freshly prepared 4% paraformaldehyde for 5 min at room temperature (RT). After washing in 1xPBS the slides were acetylated using 0.1 M triethanolamine buffer containing 0.25% acetic anhydride

(Sigma-Aldrich, Saint Louis, MI, USA) (pH 8.0). The pre-hybridization was done at 42 °C for 2 h in a pre-hybridization buffer containing 50% deionized formamide, 4 x saline-sodium citrate (SSC), 2.5x Denhardt's solution, 0.25% yeast RNA. After pre-hybridization the slides were incubated at 80 °C for 5 min (DNA denaturation) then at 42 °C for overnight (probe hybridization) in a hybridization buffer containing 50% deionized formamide, 4 x saline-sodium citrate (SSC), 2.5x Denhardt's solution, 0.5% yeast RNA, 10% Dextran sulphate and 1,4 ug/ml of the sense or antisense DIG-labeled probes. The slides were washed in 2xSSC at 42 °C for 2 x 15 min followed by washing in 1xSSC at 42 °C for 2 x 30 min in a rocking shaker. After stringency washing the sections were washed in1xPBS-0.1% Triton X-100 (1xPBST) and blocked in 1xPBST containing 10% fetal bovine serum (Sigma-Aldrich, Saint Louis, MI, USA) for an hour at RT. After blocking the slides were incubated overnight in an antibody solution of 1xPBST, 1% FBS and 1:1000 alkaline phos-phatase conjugated anti-DIG Fab (Sigma-Aldrich, Saint Louis, MI, USA) at 4 °C. After four times washing with 1xPBST for 15 min, and 100 mM Tris(pH 9.5), 100 mM NaCl for 3x5min the sections were developed with NBT/BCIP diluted in 100 mM Tris (pH9.5), 100 mM NaCl and 1 mM levamisole for three hours. The chromogenic reaction was stopped by extensive washing in 1xPBS and dH_2O. The sections were covered in 87% glycerol. In case of negative control slides the hybridization procedure was the same except in the probe hybridization step where probe-free hybridization buffer was used.

3. Results

3.1. Case description, clinical and gross pathology

A juvenile male black Syrian hamster *(Mesocricetus auratus)* (Animal 1) was purchased in a pet shop in Slovakia, in January 2013 and kept as a pet in Hungary. Three months later a skin tumor developed on the face and neck area of the hamster. During the following three weeks the neoplasia reached 25 mm in diameter and a skin metastasis on the chest. The hamster survived for 6 months and was euthanized by a veterinarian because of unfavorable prognosis (Fig. 1). At *postmortem* pathology the black Syrian hamster body weight was 40 g. The tumor was circumscribed and white-brown in color. On the cut surface a large central necrosis circumscribed by homogeneous grayish tissue was apparent (Fig. 1). The metastasis on the chest was solid, well circumscribed, measured 0.7 cm in diameter and of white color on the cut surface. No other metastasis was observed in other parts of the body.

3.2. Detection of viral sequences by viral metagenomics

The primary tumor tissue from the face (Animal 1) was subjected to viral metagenomics analysis. After *in silico* analysis and *de novo* assembly, 5593 sequence reads were obtained showing similarity to viruses (BLASTx cut-off E score <10^{-10}). The detected sequences containing > 100 reads were only from virus family *Retroviridae (N* = 2284) and human endogenous retrovirus elements (N = 613). The other (N = 1622) and unclassified (N = 1074) reads were non-viral repeats and non-relevant sequences by *in silico* sequence analysis. The reads corresponding to virus family *Retroviridae* were further analyzed.

3.3. Determination of the retrovirus sequence

Using multiple sequence-specific primer pairs designed against the metagenomic retroviral sequences the viral DNA sequence was detectable from the DNA extracts of the tumor tissue as well as the additional analyzed tissue samples of liver, lung, testes and tail by PCR. These results suggested that retroviral sequence(s) are present in the DNA form likely integrated as a provirus in the hamster genome. Viral RNA was only detectable in the 2nd PCR round of RT nested-PCR in the DNase-treated RNA extract of the tumor but not in the other RNA samples of tumor-free organs.

Using a primer walking method, TAIL-PCR, Long Range PCR (up to 8800 base pairs in length), multiple control PCRs and sequencing the continuous, single and complete genome of the provirus, tentatively named black Syrian hamster retrovirus (BSHRV/2013/HUN, GenBank accession number MK304634), was determined and verified from the DNA sample of the tumor tissue. The BSHRV genome is 8784 nucleotide (nt) in length and the genome organization follows the typical retro-viral genome structure, 5'LTR-gag-pro-pol-env-3'LTR with intact coding regions (Fig. 2). BSHRV encodes the capsid/nucleocapsid (gag) proteins; the enzymes (reverse transcriptase and integrase) needed for genome replication (pro/pol) and the envelope proteins (env) that bind the cell surface molecules used for virus entry. Because the BSHRV genome encodes at least all known fundamental retrovirus genome elements and proteins (Fig. 2) it is considered a simple betaretrovirus
(MK304634).

3.4. Analysis of the long terminal repeats (LTRs)

Betaretrovirus LTRs are one of the longest known among retro-viruses, with a length exceeding 1000 nt. LTRs contain regulatory sequences such as promoters, polyadenylation signals/sites and enhancers, and can therefore dramatically influence the RNA expression of both retroviral and nonretroviral sequences (Guntaka, 1993; Seifarth et al., 2005). At both ends of the BSHRV genome there are 1028 nt-long and 899 nt-long 5' and 3' long terminal repeats (LTRs), respectively. LTR sequences of the virus predicted to start with the sequence T1GCCAC and ended with TGGCAG8784. The 5' and 3' LTRs (in U3-R-U5 genome part) of BSHRV has 96% sequence identity to each other, respectively.

At the beginning of the BSHRV 5'LTR, a 16 nt-long tandem repeat ($T_{197/213}$TCAAGGGATATTTT$A_{212/228}$) with a high T content (44%) were found (Fig. 2). Conserved elements like TATA box (TATAAAAA) at nts 738-745, an AATAAA motif (AGTAAA) at nts 759-764 and C/T rich stretch can be recognized in BSHRV 5'LTR, respectively (Fig. 2). The predicted lengths of the unique 5' region (U5'), the terminal repeat (R) and the U3' regions are 745 nts, 13nts and 270 nts within the 5'LTR, respectively (Fig. 2). An 18 nt-long conserved primer-binding site (PBS) that bound a cellular tRNA (tRNALys3) is located at the terminal part of the BSHRV 5'LTR (TGGCGCCCGAACGGGGAC at nts 895-912). This PBS (PBS(Lys3)) compare to the PBS of HIV, SIV and MMTV (Das et al., 1995)isdifferent with 1 nt at the position 907 (A — G). The region between PBS and the gag start codon is 116 nt; the T content is 27% (A + *T* = 55%) and it does not contains any repeats. Compared to MMTV (X97044) 5'LTR, the BSHRV 5'LTR is shorter by 293 nt and has 39,1% nt identity to each other. There is also a high degree of nt identity (96% and 94%) between BSHRV 5'LTR and *Phodopus sungorus* (Djungarian hamster) endogenous retrovirus MRS-Ps (M63643) 5'LTR and BSHRV 5'LTR and the 5' LTR of dwarf hamster endogenous retro-virus MRS-X3 (S77169), respectively, available 5'LTR sequences in GenBank from years 1991 (Tikhonenko et al., 1995) and 1994 (Vasetskii, 1994).

The 3'LTR is 899 nt in length and is preceded by an 18 nt long polypurine tract (PPT, A5TA6G4A2 at nts 7885-7902) (Fig. 2) involved in plus-strand DNA synthesis (Bowman et al., 1996). The predicted length of the U5, the R and the U3 regions are 751 nts, 13nts and 135 nts within the 3'LTR, respectively (Fig. 2). Corresponding to the 5'LTR, conserved 3'LTR elements like TATA box (TATAAAAA) at nts 8629-8636 and AATAAA motif (AGTAAA) at nts 8650-8655 can be recognized in BSHRV 3'LTR, respectively. The 3'LTR of BSHRV does not contains any repeats. There is a high degree (93%) of nt identity between BSHRV 3'LTR and dwarf hamster endogenous retrovirus (S77150) 3'LTR.

Only at the express request of the manuscript Reviewer we completed the possible retroviral integration time based on the strongly hypothesis-based mathematical calculation (t = k/2 N; k = divergence; N = nucleotide substitution rate of host genome) (Martins and Villesen, 2011). A wide range of possible integration time was calculated, between 1639 years to 6.7 million years, depending on the available nu-cleotide substitution rate of rodent hosts. The authors do not draw conclusions from this result because of its unreliability.

3.5. Analysis of the coding region

The start codon for gag is present at nts 1029-1031 (A_{1029}UGG). The gag, pro/pol and env genes were encoded in three different reading frames (Fig. 2). The gag is 534 amino acids (aa) long and contains the structural betaretrovirus proteins, p10 and p24 (core nucleocapsid protein), a zinc knuckle and the gag-polyprotein viral zinc-finger (Fig. 2). The BSHRV gag protein is 57 aa shorter than the gag protein in MMTV and has 56% aa identity to the corresponding gag protein of MMTV (AF228552). The pro includes

the deoxyuridine 5'-triphosphate nucleotide hydrolase (dUTPase, trimeric dUTP bisphosphatase, dut gene) domain and the retropepsin (aspartate protease). Because the gag-pro-pol gene is translated by ribosomal frameshifting, exact N- and C-terminal ends of the pro proteins are difficult to assign. The pol gene is comprised from nt 3370 to nt 6036. The pol encodes the basic ret-roviral enzymes the reverse transcriptase (RT) including the RT thumb domain, which is composed of a four-helix bundle, the RNase H tran-scriptase and the integrase. The highly conserved aa sequence of tyr-osine (Y)-methionine (M)-aspartic acid (D)-aspartic acid (D) (YMDD motif) is present in the catalytic domain C of viral RT (Tipples et al., 1996) as well as the active residues ((H)-histidine (C)-cysteine and D,D-35(E)-glutamic acid) of the integrase (Fig. 2). The pro/pol polyprotein (~1099aa) of BSHRV has 63% aa identity to the corresponding protein of MMTV (AF228551). Env gene is comprised from 5867 nt to 7885 nt (672 aa) encoding the surface and transmembrane glycoproteins including the envelope glycoprotein of the viral envelope. The env polyprotein of BSHRV has 50% aa identity to the corresponding protein of MMTV (AF228552). No significant reading frame preceded by a start codon is found downstream of the env gene, suggesting that BSHRV does not encode a MMTV-like superantigen (sag) in 3'LTR.

The possible integration site (or sites) of the proviral BSHRV in the host genome is could not be determined. Identical betaretrovirus coding (intact) protein sequences were not identified in the available genome of the Syrian hamster *(Mesocricetus auratus)* (taxid: 10036) and the Chinese hamsters *(Cricetuluus griseus)* genomes (taxid: 10029) using blastx (Reference proteins, refseq_protein database) in GenBank.

BLASTn search of the full length BSHRV genome (MK304634) against MesAur1.0 - which contains genomic (non intact) scaffold sequences of a female Golden Hamster (ID: GCA_000349665.1) - showed only moderate sequence identity (65.81%) and low query coverage (40%) to a scaffold sequence (ID: 01962; NW_004803565.1) as the closest match. This result could suggest that BSHRV are not consistently present in all genomes of golden hamsters or the available hamster genome sequences.

3.6. Phylogenetic analysis

Phylogenetic analysis of the translated pol and env genes of BSHRV was performed using the neighbor-joining option in MEGA6 and reference sequences from betaretroviruses indicating that BSHRV is situated phylogenetically in a cluster with MMTV confirming the genetic relationship (Fig. 3), however, it could represent a novel betare-trovirus species.

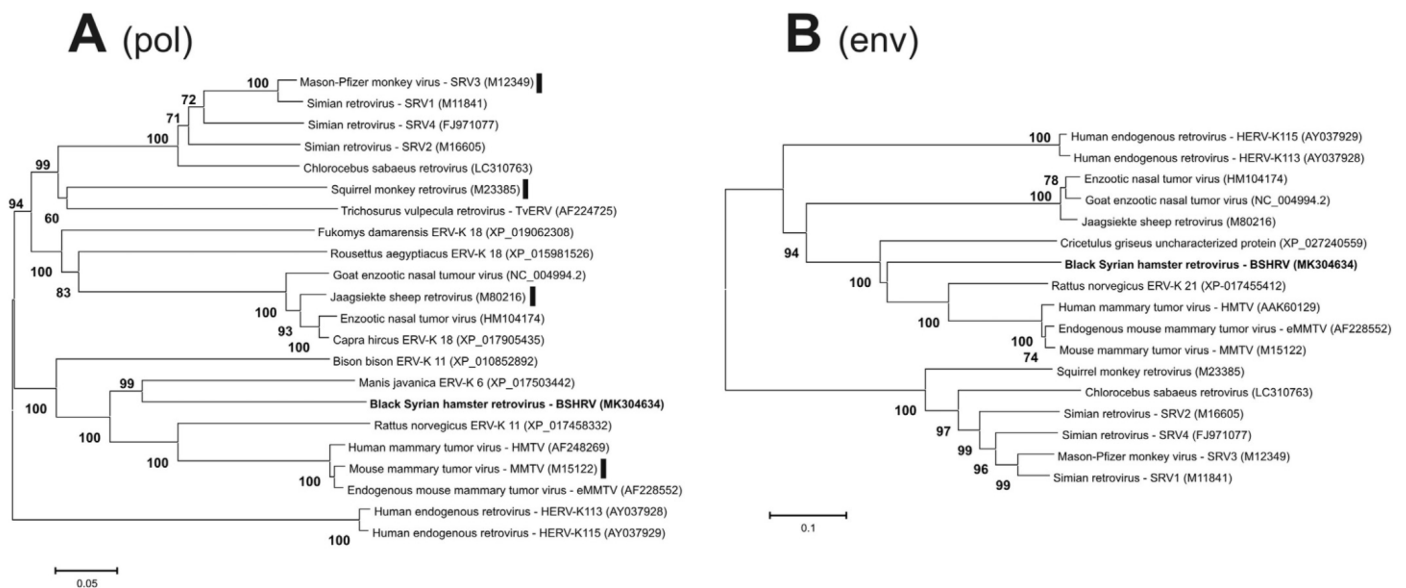

Fig. 3. Phylogenetic analysis of the translated (aa) pol (A) and env (B) genes of BSHRV/2013/HUN (GenBank accession number MK304634, bold letters) using the neighbor-joining option in MEGA6 and reference sequences from betaretroviruses. 1000 bootstrap replicates were analyzed. Bootstrap values>50 are shown. Sequences had been aligned using the Clustal X program. Vertical black lines indicate the prototype members of the known betaretrovirus species.

3.7. *Histological findings*

The haematoxylin and eosin stained histological sections of the tumor (Animal 1) demonstrate a well-differentiated squamous cell carcinoma. The fibroblastic neostroma is infiltrated with typical flamelike tongues of centrally keratinized tumor cell nests (Fig. 4A and B).

3.8. *In situ hybridization*

BSHRV-specific hybridization was predominantly restricted to the proliferating tumor cells of the squamous epithelium which formed numerous concentric keratin nodules (Fig. 4). The hybridizations were observable mainly in the nuclei of the tumor cells in sections hybridized with either sense or antisense probes (Fig. 4C and D). Furthermore, the localization of the hybridization signals were essentially the same in the sections labeled with the sense or the antisense probes (Fig. 4C and D). These results suggest that the probes were detecting the integrated proviral DNA of BSHRV not the viral RNA. There was no visible specific hybridization signal in the negative control sections (Supplementary
Fig. S2).

3.9. *Detection of BSHRV-like betaretrovirus sequence from further black Syrian hamsters*

One of the three DNA extracts from faecal specimens collected from further three animals were strong positive for BSHRV-like betare-trovirus by nested-PCR and sequencing. The amplified 538 nt long gag

gene sequence had 99.4% nt identity to the corresponding regions of BSHRV/2013/HUN. This animal (a 1.5-year-old female, Animal 2) died - because of the old age - on November 22, 2018. Different *post mortem* tissue specimens (skin, lung, ovaries and liver) were collected and all specimens tested also positive by nested PCR for the novel betare-trovirus.

3.10. Integrated BSHRV-like betaretrovirus genome copy number determination using qPCR

Based on the results of qPCR analysis the tumor, liver and lung samples of the black Syrian hamster with differentiated squamous cell skin cancer contained similar numbers of integrated viral copies (~1.18E + 04/ng DNA, Table 2), while the tail sample contained nearly twice as large a copy number of integrated viral genomes (2.03E+04/ ng DNA, Table 2). Interestingly, the integrated viral genome copy numbers of the tissue samples of the tumor-free animal (Animal 2) were *ca.* one order of magnitude higher (ranged between 1.11E + 05/ng DNA and 1.76E + 05/ng DNA) than the animal (Animal 1) with skin tumor (Table 2).

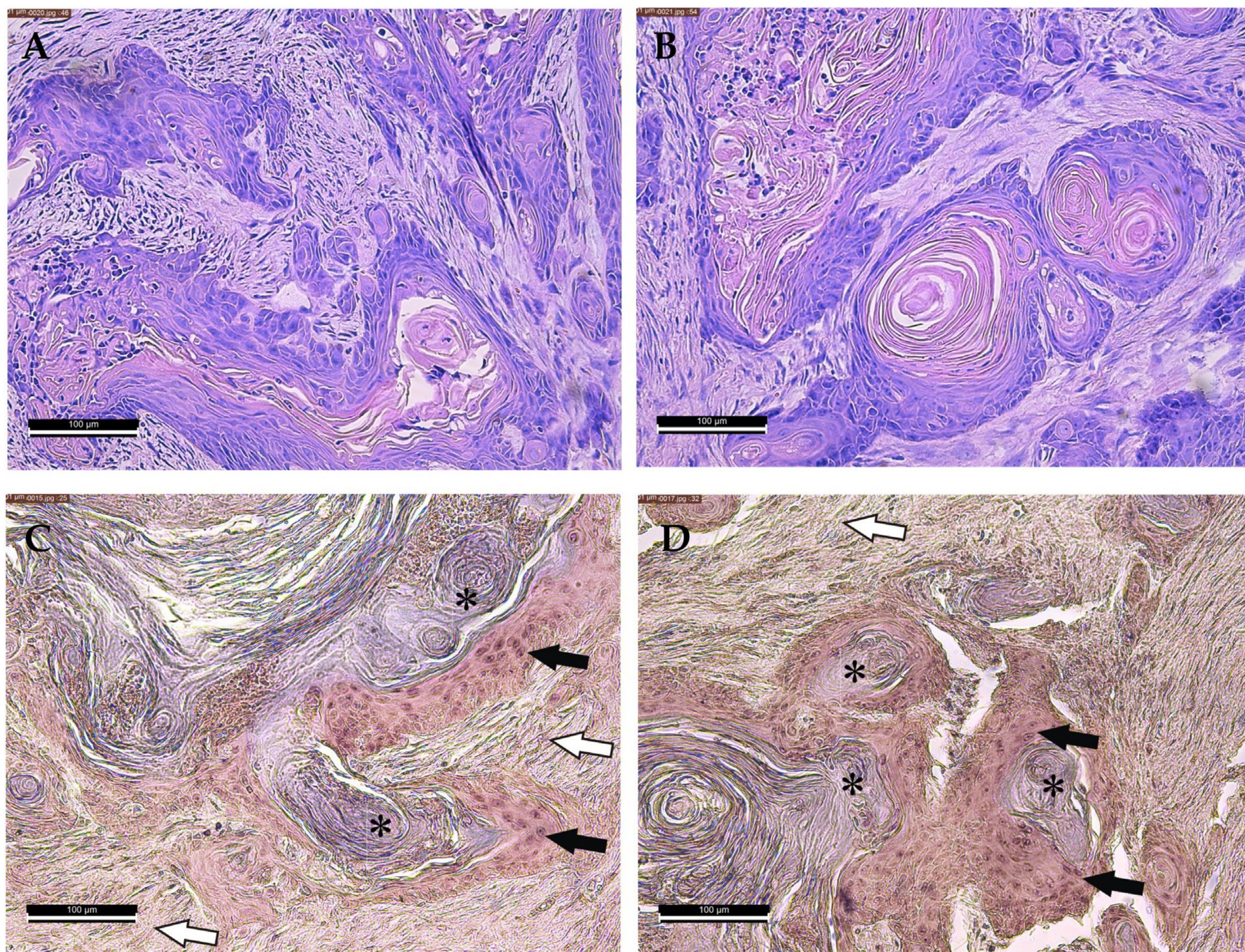

Fig. 4. Sections from squamous cell skin tumor of a black Syrian hamster (Mesocricetus auratus) (Animal 1) stained with Haematoxylin and eosin (A) and (B) and chromogenic ISH using sense (C) or antisense (D) probes. Black arrows indicate the stained nuclei of BSHRV-positive tumor cells surrounding the concentric keratin nodules (asterisks). White arrows indicate the BSHRV-negative, potentially tumor-free cells.

Table 2: Summary of the results of absolute quantification of integrated viral copy number in a hamster with a skin tumor (Animal 1) and a clinically healthy hamster (Animal 2) using SYBR-Green based qPCR. Ct: threshold cycle; SD: standard deviation.				
Tested animals	Sample type	Average Ct	SD (Ct)	Average copy number/ng DNA
Animal 1	skin tumor	20.3888	0.078	1.18E+04
	liver	20.4348	0.349	1.17E+04
	lung	20.3701	0.064	1.20E+04
	tail	19.5900	0.039	2.03E+04
Animal 2	skin	16.4016	0.057	1.76E+05
	liver	17.7126	0.052	1.11E+05
	lung	16.4566	0.026	1.70E+05

4. Discussion

In this study, using viral metagenomics, histological analysis, *in situ* hybridization, PCR, RT-PCR and qPCR approaches, we report the serendipitous identification, complete genome characterization and analysis of a novel endogenous retrovirus sequence, that was named black Syrian hamster retrovirus (BSHRV), in a black Syrian hamster *(Mesocricetus auratus)* that presented with differentiated squamous cell skin cancer. Taxonomically, the golden and Syrian hamsters are identical (M. aurateus)(https://www.ensembl.org/Mesocricetus_auratus/ Info/ Annotation). Black Syrian hamsters are color variations of *M. auratus.*

The BSHRV retroviral DNA sequence was detectable from the DNA extracts of the tumor tissue as well as the additional analyzed tissue samples of liver, lung, testes and tail collected from the same animal by conventional PCR and qPCR. These results suggested that the retroviral sequence(s) is possibly present in integrated form in this host germ-line, as a provirus. Interestingly, the BSHRV retroviral RNA was detectable only in the 2nd PCR round of RT nested-PCR in the DNase-treated RNA sample of the tumor but not in the other RNA samples of tumor-free organs which suggest a low level viral RNA expression in the tumor but not in the other analyzed tissues. The BSHRV retroviral DNA sequence was also identified in a faecal (most likely from host cells present in the faeces) and other tissue samples (skin, lung, ovaries and liver) -showing the same systemic distribution as in the diseased animal -collected from an additional black Syrian hamster that was clinically healthy at the time of the specimen's collection. Interestingly, the integrated viral genome copy numbers were > 10 times higher in tissue specimens in tumor-free animal than animal with skin cancer. The association between the BSHRV and the skin tumor in hamster remains unknown. A non-viral etiology for tumor formation in the presented case is also the possible. To prove that BSHRV, as do oncogenic retro-viruses (Braoudaki and Tzortzatou-Stathopoulou, 2011), influences or contributes to tumor genesis and therefore capable of transforming normal cells into malignant cells will require further studies. Further studies also need to investigate the origin of the BSHRV retroviral DNA

shedding in the hamster faecal sample. Interestingly, counterpart study investigating the faeces for the "milk-transmitted" MMTV is not known for mice - to our best knowledge.

The BSHRV retrovirus genome had full length and typical betaretroviral genome features and organization of 5'LTR-gag-pro-pol-env-3'LTR with the dUTPase domain but without the superantigen (sag) gene in 3'LTR (Hayward et al., 2013). The continuous and single pro-viral nucleotide sequence was verified by amplification of the full length betaretrovirus genome by Long Range PCR and Sanger sequencing. In the translated region, the gag gene encodes the characteristic betaretrovirus elements p10, p24 (capsid), zinc knuckle and the zinc finger. The pro gene encodes the dUTPase and the retropepsin. The pol gene includes the reverse transcriptase (RT), the RNase H and the integrase. The env gene encodes the envelope proteins. The complete and intact LTR/coding proviral genome regions of BSHRV are remarkable and one of the prerequisites for a viral replication and production of potential viable retrovirus particles.

By sequence- and phylogenetic analysis the proteins of the BSHRV had high aa sequence diversity from the known members of the be-taretrovirus species. BSHRV represents a potential novel betaretrovirus species but its proteins had the closest relationship to the corresponding proteins of mouse mammary tumor virus (MMTV) betaretrovirus. MMTV is an important prototype member of the betaretroviruses associated with tumors in mice. In this context, the determination of a novel sequence relative of MMTV-like betaretrovirus in a novel mammal host species increases the knowledge of the sequence diversity of these retroviruses. It would be worth considering this betaretrovirus sequence variant in screening primer design for PCR in human and animal prevalence studies in the future.

There are several important open questions related to BSHRV. Based on the methods used in this study the endogenous retrovirus nature, the DNA copy number(s) in the host genome and the integration site(s) remains unknown in either the tumor or the germ-line. The viral particle nature and the transmission mode(s) of the potential exogenous form of BSHRV virions are also undetermined. Functional analysis of BSHRV by RNA expression in different tissues, electron microscopy and Southern blotting with specific probes maybe brings us closer to the answers. BSHRV or BSHTV-like complete betaretrovirus sequences have not been reported from hamsters and could not be identified in the available assembled genome databases of hamsters. Further study is required to determine the prevalence and biological significance - including in tumorigenesis - of BSHRV and BSHRV-like betaretroviruses in different genetic lineages of hamsters and other mammals.

Supplementary data to this article can be found online at https:// doi.org/10.1016/j.meegid.2019.103995.

Funding

This work was financially supported by a grant from the Hungarian Scientific Research Fund (OTKA/NKFIH K111615).

Author agreement/declaration

All authors have seen and approved the final version of the manuscript being submitted. They warrant that the article is the authors' original work, hasn't received prior publication and isn't under consideration for publication elsewhere.

Acknowledgements

Á.B. and P.P. are supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

Axel, R., Schlom, J., Spiegelman, S., 1972. Presence in human breast cancer of RNA homologous to mouse mammary tumour virus RNA. Nature 235 (5332), 32-36.

Bassin, R.H., Simons, P.J., Chesterman, F.C., Harvey, J.J., 1968. Murine sarcoma virus (Harvey): characteristics of focus formation in mouse embryo cell cultures, and virus production by hamster tumor cells. Int. J. Cancer 3 (2), 265-272.

Bittner, J.J., 1936. Some possible effects of nursing on the mammary gland tumor incidence in mice. Science 84 (2172), 162.

Boros, A., Pankovics, P., Knowles, N.J., Reuter, G., 2012. Natural interspecies recombinant bovine/porcine enterovirus in sheep. J. Gen. Virol. 93 (9), 1941-1951 Pt.

Bowman, E.H., Pathak, V.K., Hu, W.S., 1996. Efficient initiation and strand transfer of polypurine tract-primed plus-strand DNA prevent strand transfer of internally initiated plus-strand DNA. J. Virol. 70 (3), 1687-1694.

Braoudaki, M., Tzortzatou-Stathopoulou, F., 2011. Tumorigenesis related to retrovirual infections. J Infect. Dev. Ctries. 5 (11), 751-758.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59-60.

Callahan, R., Smith, G.H., 2008. Common integration sites for MMTV in viral induced mouse mammary tumors. J. Mammary Gland Biol. Neoplasia 13 (3), 309-321.

Coffin, J.M., 1992. Superantigens and endogenous retroviruses: a confluence of puzzles. Science 255 (5043), 411-413.

Cohen, J.C., Varmus, H.E., 1979. Endogenous mammary tumour virus DNA varies among wild mice and segregates during inbreeding. Nature 278 (5703), 418-423.

Das, A.T., Klaver, B., Berkhout, B., 1995. Reduced replication of human immunodeficiency virus type 1 mutants that use reverse transcription primers other than the natural tRNS3 Lys. J. Virol. 69 (5), 3090-3097.

Deng, X., Naccache, S.N., Ng, T., Federman, S., Li, L., Chiu, C.Y., Delwart, E.L., 2015. An ensemble strategy that significantly improves de novo assembly of microbial genomes from metagenomic next-generation sequencing data. Nucleic Acids Res. 43, e46.

Dudley, J., 2011. Retroviruses and insights into cancer, 1 ed. Springer, New York.

Ellerman, V., Bang, O., 1908. Experimentelle Leukämie bei Hühnern. Zentralbl. Bakteriol. Parasitenkd. Infectionskr. Hyg. Abt. 46, 595-609.

Feller, W.F., Chopra, H.C., 1969. Studies of human milk in relation to the possible viral etiology of breast cancer. Cancer. 24 (6), 1250-1254.

Garcia-Montojo, M., Doucet-O'Hare, T., Henderson, L., Nath, A., 2018. Human endogenous retrovirus-K (HML-2): a comprehensive review. Crit. Rev. Microbiol. 44, 715-738.

Gonzalez-Cao, M., Iduma, P., Karachaliou, N., Santarpia, M., Blanco, J., Rosell, R., 2016. Human endogenous retroviruses and cancer. Cancer Biol. Med. 13 (4), 483-488.

Guntaka, R.V., 1993. Transcription termination and polyadenylation in retroviruses. Microbiol. Rev. 57 (3), 511-521.

Hayward, J.A., Tachedjian, M., Cui, J., Field, H., Holmes, E.C., Wang, L.F., Tachedjian, G., 2013. Identification of diverse full-length endogenous betaretroviruses in megabats and microbats. Retrovirology 10, 35. Hohn, O., Hanke, K., Bannert, N., 2013. HERV-K(HML-2), the best preserve family of HERVs: endogenization, expression, and implications in health and disease. Front. Oncol. 3, 246.

ICTV, 2011. 9th Report. Retroviridae. https://talk.ictvonline.org/ictv-reports/ictv_9th_report/reverse-transcribing-dna-and-rna-viruses-2011/w/rt_viruses/161/retroviridae (accessed on 18.12.2018).

Jelinek, F., Felsberg, J., Mestan, M., 2013. Familial incidence ofmammary gland tumours in the Djungarian hamster (Phodopus sungaros): a case report. Vet. Med. 58,442-448.

Kassiotis, G., 2014. Endogenous retroviruses and the development of cancer. J. Immunol. 192 (4), 1343-1349.

Kassiotis, G., Stoye, J.P., 2017. Making a virtue ofnecessity: the pleiotropic role ofhuman endogenous retroviruses in cancer. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 372 (1732).

Katz, E., Lareef, M.H., Rassa, J.C., Grande, S.M., King, L.B., Russo, J., Ross, S.R., Monroe, J.G., 2005. MMTV Env encodes an ITAM responsible for transformation of mammary epithelial cells in three-dimensional culture. J. Exp. Med. 201 (3), 431-439.

Kelloff, G., Huebner, R.J., Lee, Y.K., Toni, R., Gilden, R., 1970. Hamster-tropic sarco-magenic and nonsarcomagenic viruses derived from hamster tumors induced by the gross pseudotype of Moloney sarcoma virus. Proc. Natl. Acad. Sci. U. S. A. 65 (2), 310-317.

Keydar, I., Ohno, T., Nayak, R., Sweet, R., Simoni, F., Weiss, F., Karby, S., Mesa-Tejada, R., Spiegelman, S., 1984. Properties of retrovirus-like particles produced by a human breast carcinoma cell line: immunological relationship with mouse mammary tumor virus proteins. Proc. Natl. Acad. Sci. U. S. A. 81 (13), 4188-4192.

Kondo, H., Onuma, M., Shibuya, H., Sato, T., 2008. Spontaneous tumors in domestic hamsters. Vet. Pathol. 45 (5), 674-680.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al., 2001. Initial sequencing and analysis ofthe human genome. Nature 409 (6822), 860-921.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with bowtie 2. Nat. Methods 9, 357-359.

Lawson, J.S., Glenn, W.K., Salmons, B., Ye, Y., Heng, B., Moody, P., Johal, H., Rawlinson, W.D., Delprado, W., Lutze-Mann, L., et al., 2010. Mouse mammary tumor virus-like sequences in human breast cancer. Cancer Res. 70 (9), 3576-3585.

Liu, Y.G., Whittier, R.F., 1995. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25 (3), 674-681.

Liu, B., Wang, Y., Melana, S.M., Pelisson, I., Najfeld, V., Holland, J.F., Pogo, B.G., 2001. Identification of a proviral structure in human breast cancer. Cancer Res. 61 (4), 1754-1759.

Martins, H., Villesen, P., 2011. Improved integration time estimation of endogenous retroviruses with phylogenetic data. PLoS One 6, e14745.

Mesa-Tejada, R., Oster, M.W., Fenoglio, C.M., Magidson, J., Spiegelman, S., 1982. Diagnosis of primary breast carcinoma through immunohistochemical detection of antigen related to mouse mammary tumor virus in metastatic lesions: a report of two cases. Cancer 49 (2), 261-268.

Phan, T.G., Vo, N.P., Boros, A., Pankovics, P., Reuter, G., Li, O.T., Wang, C., Deng, X., Poon, L.L., Delwart, E., 2013. The viruses of wild pigeon droppings. PLoS One 8 (9), e72787.

Ross, S.R., 2008. MMTV infectious cycle and the contribution ofvirus-encoded proteins to transformation of mammary tissue. J. Mammary Gland Biol. Neoplasia 13 (3), 299-307.

Ross, S.R., 2010. Mouse mammary tumor virus molecular biology and oncogenesis. Viruses 2 (9), 2000-2012.

Rous, P., 1911. Asarcoma of the fowl transmissible by an agent separable from the tumor cells. J. Exp. Med. 13 (4), 397-411.

Russell, P., Gregerson, D.S., Albert, D.M., Reid, T.W., 1979. Characteristics of a retrovirus associated with a hamster melanoma. J. Gen. Virol. 43 (2), 317-326.

Seifarth, W., Frank, O., Zeilfelder, U., Spiess, B., Greenwood, A.D., Hehlmann, R., LeibMosch, C., 2005. Comprehensive analysis of human endogenous retrovirus tran-scriptional activity in human tissues with a retrovirus-specific microarray. J. Virol. 79 (1), 341-352.

Somers, K.D., May, J.T., Kit, S., McCormick, K.J., Hatch, G.G., Stenback, W.A., Trentin, J.J., 1973. Biochemical properties of a defective hamster C-type oncornavirus. Intervirology 1 (1), 11-18.

Spiegelman, S., Burny, A., Das, M.R., Keydar, J., Schlom, J., Travnicek, M., Watson, K., 1970. Characterization of the products of DNA-directed DNA polymerases in onco-genic RNA viruses. Nature 227 (5258), 563-567.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30 (12), 2725-2729.

Tikhonenko, A.T., Polotskaya, A.V., Vassetzky III, N.S., Golovkina, T.V., Gudkov, A.V., 1995. Long terminal repeats of dwarf hamster endogenous retrovirus are highly diverged and do not maintain efficient transcription. Virology 181 (1), 367-370.

Tipples, G.A., Ma, M.M., Fischer, K.P., Bain, V.G., Kneteman, N.M., Tyrrell, D.L., 1996. Mutation in HBV RNA-dependent DNA polymerase confers resistance to lamivudine in vivo. Hepatology 24 (3), 714-717.

Vasetskii, N.S., 1994. Analysis of the sequence of the endogenous Djungarian hamster provirus (MRS-Ps) related to the murine mammary cancer virus. Mol. Biol. (Mosk) 28 (4), 900-908.

Verma, I.M., Meuth, N.L., Fan, H., Baltimore, D., 1974. Hamster leukemia virus: lack of endogenous DNA synthesis and unique structure of its DNA polymerase. J. Virol. 13 (5), 1075-1082.

Victoria, J.G., Kapoor, A., Li, L., Blinkova, O., Slikas, B., Wang, C., Naeem, A., Zaidi, S., Delwart, E., 2009. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. J. Virol. 83 (9), 4642-4651.

Wang, Y., Holland, J.F., Bleiweiss, I.J., Melana, S., Liu, X., Pelisson, I., Cantarella, A., Stellrecht, K., Mani, S., Pogo, B.G., 1995. Detection of mammary tumor virus env gene-like sequences in human breast cancer. Cancer Res. 55 (22), 5173-5179.

Ye, J., McGinnis, S., Madden, T.L., 2006. BLAST: improvements for better sequence analysis. Nucleic Acids Res. 34, W6-W9.

Young, G.R., Stoye, J.P., Kassiotis, G., 2013. Are human endogenous retroviruses pathogenic? An approach to testing the hypothesis. Bioessays 35 (9), 794-803.