Virus Genes (2015) 51:12–24
DOI 10.1007/s11262-015-1204-y

# Generation and propagation of recombinant mumps viruses exhibiting an additional U residue in the homopolymeric U tract of the F gene-end signal

**Christian J. Sauder[1] · Laurie Ngo[1] · Vahan Simonyan[2] · Yu Cong[1,3] · Cheryl Zhang[1] · Malen Link[1,4] · Tahir Malik[1] · Steven A. Rubin[1]**

Received: 17 December 2014 / Accepted: 27 April 2015 / Published online: 12 May 2015


**Abstract** As a member of the family paramyxoviridae, subfamily paramyxovirinae, the genome of mumps virus (MuV) is postulated to be polyhexameric in length in order to be able to replicate efficiently. While all natural MuV strains sequenced so far obey to this "rule of six," we describe here the isolation of recombinant MuVs that appeared to contain an additional U residue in the homopolymeric tract of the F gene-end signal, resulting in a genome length of 6n + 1. Sequencing of several plaque-purified viruses from these preparations did not reveal the existence of length-correcting mutations, suggesting that they are violators of the rule of six. Employing high-throughput sequencing technology, we provide evidence that the insertion of an additional U residue is mainly the result of the rescue system used that relies on T7 RNA polymerase. Limited in vitro and in vivo testing of the viruses did not reveal any significant impact of the longer genome on virus replication or virulence, suggesting that the rule of six is not a strict requirement for MuV replication.

**Keywords** Mumps virus · Rule of six · F gene-end signal · Homopolymeric tract · High-throughput sequencing

## Introduction

Mumps virus (MuV) is a neurotropic enveloped, non-segmented negative-strand RNA virus belonging to the family paramyxoviridae, subfamily paramyxovirinae, genus rubulavirus. The genome is 15,384 bases in length and contains seven genes that encode nine proteins. The N gene codes for the nucleoprotein (N), the P/V/I gene codes for the non-structural protein V (an interferon antagonist) as well as for the phosphoprotein (P), and the putative non-structural protein I with thus far unknown function. The P and I mRNAs are generated through RNA editing by co-transcriptional insertion of two and four, respectively, non-templated G nucleotides into the V mRNA editing site [1]. The M, F, SH, and HN genes code for the matrix (M), fusion (F), short hydrophobic (SH), and hemagglutinin/neuraminidase (HN) proteins, respectively. The M protein is attached to the inner side of the viral envelope and mediates virus assembly and budding. Anchored in the envelope are the HN and F proteins that mediate attachment to and fusion with the target cell membrane. The SH protein, being largely composed of a hydrophobic stretch of amino acids, is likely embedded in the envelope and, while shown to exhibit anti-apoptotic activities, is not essential for viral replication in vitro and in vivo [2–4]. The L gene encodes the large (L) protein which together with the P protein forms the RNA-dependent RNA polymerase

Cheryl Zhang—Deceased.

Edited by Hartmut Hengel.

✉ Christian J. Sauder
christian.sauder@fda.hhs.gov

[1] Food and Drug Administration (FDA), Center for Biologics Evaluation and Research (CBER), OVRR, DVP, 10903 New Hampshire Avenue, Bldg 52/72, Silver Spring, MD 20993, USA

[2] FDA, CBER, OD, 10903 New Hampshire Avenue, Bldg 75, Silver Spring, MD 20993, USA

[3] Present Address: Integrated Research Facility, Ft. Detrick, Frederick, MD, USA

[4] Present Address: Schafer Corporation, 3811 Fairfax Dr. Ste #400, Arlington, VA, USA

(RdRp). Viral RNA encapsidated with the N protein serves as template for the RdRp which mediates transcription and replication of the genome. The genome is flanked at the 3′ and 5′ end by an extragenic leader and an extragenic trailer, respectively, which are essential for transcription and replication. The open reading frames (ORFs) of the genes are separated from each other by the untranslated regions that contain transcriptional gene-end, intergenic and transcriptional gene-start sequences. The intergenic sequences are comprised of one to seven nucleotides.

MuV is the etiological agent of mumps and before the development of live-attenuated MuV vaccines, mumps was a common childhood disease, manifesting mainly as parotitis, but more severe complications, such as meningitis and orchitis, are not uncommon (see [5] for review). In countries where a two-dose regimen of MuV-containing vaccines has been implemented, disease incidence decreased drastically. However, outbreaks continue to occur, even in highly vaccinated populations [6, 7]. This has generated interest in development of new, more effective mumps vaccines, an endeavor complicated by our lack of understanding of the basis of virus attenuation, as manifested by the use of some vaccine strains that have caused meningitis in recipients [8]. Towards a better understanding of the genetic basis for MuV attenuation we previously generated a series of recombinant MuVs consisting of genes derived from virulent and attenuated strains and tested these viruses in an in vivo rat model [9, 10]. In the course of those studies we noted the occasional rescue of recombinant MuVs containing an additional uracil (U) residue in poly-U tracts (genomic orientation) [9, 10]. The most frequent site of this nucleotide insertion was observed in the poly-U tract in the F gene-end stop signal, extending its length from 7 to 8 U residues. Single nucleotide insertions were also sporadically observed in the poly-U tract in the P gene 3′UTR and in the M gene and L gene-end stop signals. The identification of the insertion of an additional nucleotide suggested that the MuV genome length might not strictly obey the "rule of six," a postulated requirement that all members of the *Paramyxovirinae* have genome lengths that are multiples of six, presumably for efficient replication [11–13]. However, the observed nucleotide insertions were apparent only in a portion of the amplified genomic RNA molecules, rarely exceeding 50 % of all PCR fragments sequenced. Therefore, it was unclear whether these insertions existed in replication-competent genomes, or were only present in non-infectious particles or were from non-virion-associated genomic RNA molecules. Further, because we did not fully sequence these MuV genomes, we could not rule out compensatory nucleotide deletions elsewhere. In order to better investigate this issue, we completely sequenced the genomes of two recombinant MuVs that appeared to contain a nucleotide insertion in the F gene-end signal, as well as several plaque-purified clones derived from these viruses. Our data indicate that MuVs with an additional nucleotide insertion in the F gene-end signal do exist and are replication competent, suggesting that MuV does not strictly obey the rule of six. Limited in vitro and in vivo testing did not reveal a significant impact of the longer genome on replication or virulence.

## Materials and methods

### Cell lines and viruses

Vero (monkey kidney) cells (passages 180–200) and BHK-BSR-T7/5 cells [14] were cultivated in Dulbecco's modified Eagle's medium (DMEM, Quality Biological (QB), Gaithersburg, MD) supplemented with 2 mM L-glutamine and 9 % fetal calf serum (FCS, QB). BHK-BSR-T7/5 cells were cultivated in the presence of 1 mg/ml geneticin (Invitrogen) at every other passage. The Urabe P-AM9 vaccine virus stock and the 88-1961 wild-type clinical isolate have been described previously [15, 16].

### Construction of plasmids

Plasmids pUrabe1750 and p88-1750 (used for in vitro transcription and/or DNA control for high-throughput sequencing (HTP) were constructed by PCR amplifying a 1750 bp fragment encompassing nucleotide positions 5035–6784 from MuV plasmids pURABE$_{HN\text{-}K335}$ [9] and p88-1961modΔNhe [10], respectively, using primers FMIDfwd (5′ cggtacaagcaatacaagacca 3′) and 6784rev (5′ cctcacaagctcacctaaagtg 3′). The PCR-amplified fragments were subcloned into plasmid pCR2.1-TOPO (Life Technologies) as described previously [10] and sequenced to confirm identity and correct orientation.

#### *Expression plasmids*

The construction of the plasmids expressing the Urabe AM9 virus N, P, and L gene products (pTM1-NUr, pTM1-PUr, and pTM1-LUr) has been described elsewhere [9].

#### *Full-length cDNA clones*

Plasmids pMuV(MPBS), pURABE$_{HN\text{-}K335}$, pMuVUrabe$_{NheI}$ and p88-1961modΔNhe, encoding the full-length cDNAs of viruses Jeryl Lynn (JL, Major component), Urabe AM9 and 88-1961, respectively, have been described elsewhere [9, 10, 17]. To construct pUrabe + JL(F), the *Mlu*I–*Sgf*I fragment of pURABE$_{HN\text{-}K335}$ was first cloned into plasmid pCR2.1-TOPO (Invitrogen), that was modified to harbor

*Mlu*I and *Sgf*I restriction sites. From the resulting plasmid pTopoUrabeMluI-SgfI, the Urabe AM9 F gene was removed with *PmeI* and *BmgBI* and was replaced with the JL F gene that was excised from pMuV(MPBS) with *PmeI* and *BmgBI*. From the resulting plasmid pTopoUrabeMluI-SgfIJLF, the *PmeI–SgfI* fragment containing the JL F and the Urabe AM9 SH gene was excised and ligated into $pURABE_{HN\text{-}K335}$ (from which the corresponding *PmeI* and *SgfI* fragment had been removed) to yield pUrabe + JL(F). To construct pUrabe + JL(F/HN), first the Urabe AM9 HN gene of plasmid $pMuVUrabe_{NheI}$ was excised with *SgfI* and *NheI* and replaced with the JL HN gene containing *SgfI–NheI* fragment that was excised from pMuV(MPBS). From the resulting plasmid pUrabe + JL(HN), the *PmeI-SgfI* fragment was removed and replaced with the corresponding fragment from pTopoUrabeMluI-SgfIJLF to yield plasmid pUrabe + JL(F/HN).

### Rescue of recombinant viruses from cDNA

For all virus rescues, BHK-BSRT7/5 cells were grown to 95 % confluence in 6-well dishes in the presence of 1 mg/ml Geneticin and a total volume of 2 ml DMEM/9 % (vol/vol) FCS. Medium was replaced with 2 ml of fresh medium lacking Geneticin at least 2 h before transfection. For viruses to be recovered from plasmids pUrabe + JL(F) and pUrabe + JL(F/HN), cells were transfected with a mixture containing 225 ng of pTM1-NUr, 37 ng pTM1-PUr, 172 ng of pTM1-LUr, and 6.95 μg of the respective full-length cDNA plasmids. For all transfections, DNA was mixed with 30 μl of Lipofectamine 2000 (Invitrogen) in a total volume of 530 μl with Opti-MEM I medium (Invitrogen) as diluent. All cells were transferred first to a 25-$cm^2$ flask 48 h post-transfection and subsequently into a 75-$cm^2$ flask 2–4 days later. Supernatants were collected 7–11 days post-transfection. Supernatants were clarified by centrifugation at 1200 rpm for 5 min and aliquots stored at −70 °C.

### Plaque purification of viruses

The two viruses rUrabe + JL(F) and rUrabe + JL(F/HN) that contained a high percentage of genomes with the insertion in the F gene-end stop signal were subjected to the same procedure employed to determine the potency of a virus stock (plaque assay) as described previously [15]. Five days following visualization of plaques using neutral red/agar overlay, individual plaques were picked only from wells that displayed less than five plaques. Plaques were picked using a 10-μl pipet tip and were resuspended in 400 μl of EMEM medium and stored at −70 °C until propagation in Vero cells grown to 95 % confluency on 25 $cm^2$ flasks. Supernatants were harvested when widespread cytopathic effect or syncytia formation was visible. Supernatants were clarified by centrifugation at 1200 rpm for 5 min.

### In vitro multicycle growth kinetics

Vero cells grown to confluency on six-well plates were infected at a multiplicity of infection of 0.05 in a total volume of 500 μl DMEM/FBS. Following incubation with virus for 1 h at 37 °C, cells were washed with PBS and 2 ml of fresh medium each was added. For each time point, 10 % of the cell culture supernatant was removed and immediately stored at −70 °C until it was assayed for virus titer. Removed medium was replaced with an equal volume of fresh medium. This procedure was repeated every 24 h for a period of 6 days. Samples additionally were taken at 39 h post-infection (p.i.). All time points were sampled in duplicate from independent wells.

### Serial passaging of viruses in Vero cells

For generation of the first passage in Vero cells (P1 Vero) of the viruses recovered from BHK-BSRT7/5 cells, either 1.7 or 5 ml, respectively, of supernatants collected from these cells were inoculated onto Vero cells grown to 95 % confluency in 25 or 75 $cm^2$ flasks, respectively, and supernatants harvested from cells 4–7 or 3–6 days, respectively, post-infection. Supernatants were clarified by centrifugation at 1200 rpm for 5 min and aliquots stored at −70 °C. Virus titer was determined by plaque assay as described previously [15]. For generation of passage two in Vero cells, 25 $cm^2$ flasks were inoculated with virus from the previous passage at an m.o.i. of 0.05. Cell culture supernatants were harvested when cytopathic effects were evident 4–7 days post-infection.

### Rat neurovirulence test

The rat neurovirulence test was carried out as described previously [9, 10]. The number of rats infected with viruses rUrabe + JL(F), rUrabe + JL(F) plaque purified (PPF)(7A), rUrabe + JL(F/HN), and rUrabe + JL(F/HN)-PPF(7A) was 19, 19, 23, and 32, respectively.

### In vitro transcription

To generate RNA transcripts of the Urabe AM9 mumps virus genomic region encompassing nucleotides 5035–6784, plasmid pUrabe1750 was linearized using *Spe*I. Following gel purification of the linearized DNA, RNA was transcribed using the MEGAscript kit (Ambion, Austin, TX). Briefly, RNA was transcribed using T7 RNA polymerase in a total volume of 20 μl, containing 1 μg of

linearized DNA resuspended in nuclease-free water, 2 µl each of 75 mM rATP, rCTP, rGTP, and rUTP solutions, 2 µl 10 × reaction buffer, 2 µl of enzyme mix (T7), and 10 units of RNaseOUT Ribonuclease Inhibitor (Invitrogen, Carlsbad, CA). Transcription was carried out at 37 °C for 4 h followed by degradation of the DNA template at 37 °C for 20 min in the presence of 1 unit of DNase. In vitro transcribed RNA was extracted with Tris–EDTA-saturated phenol/chloroform in the presence of 10 mM EDTA and 0.5 M ammonium acetate stop solution. RNA was precipitated with isopropanol for 1 h at −20 °C. RNA was pelleted at 10,000×*g* and washed with 70 % ethanol. Air-dried RNA was resuspended in 20 µl of DEPC treated water/0.5 mM EDTA and stored at −70 °C. Stock RNA was quantified based on OD260 and diluted in 0.5 mM EDTA.

### RNA extractions

RNA from virus supernatants clarified by centrifugation was extracted using the QIAamp Minelute virus spin kit (Qiagen, Valencia, CA) as recommended by the manufacturer. RNA that was extracted from supernatants from Vero cells that were infected with supernatants recovered from transfected BHK-BSR-T7/5 cells (=P1 Vero) was incubated with DNase for 30 min at 37 °C, followed by heat inactivation for 10 min at 75 °C in the presence of 3.5 mM EDTA. Total RNA extraction from virus-infected rat brains was done using the RNeasy Plus Mini kit (Qiagen). First, frozen brains were subjected to dounce homogenization in the presence of 400 µl of buffer RLT plus (Qiagen). Following transfer to 15 ml tubes and adding of buffer RLT plus at a concentration of 600 µl per 30 mg of brain tissue, homogenates were further subjected to three pulses (10 s each) of ultrasonic treatment. RNA was extracted from 600 µl of homogenate according to the instructions of the manufacturer.

### RT-PCR, PCR

RNA was subjected to reverse transcription (RT) using the Superscript II reverse transcriptase kit (Invitrogen) and virus sequence-specific genomic primers. If not stated otherwise, all PCR amplifications were carried out using Expand high-fidelity polymerase (Roche Biochemicals, Burlington, NC). To verify the RT dependence of RT-PCRs, parallel RT reactions were carried out omitting the reverse transcriptase enzyme.

### Determination of viral genomic termini

Viral genomic 3′ and 5′ termini were amplified using 3′RACE and 5′RACE kits, respectively (Invitrogen). To amplify the viral 3′ termini, RNA was extracted from 100 µl of virus stock, DNase treated and subjected to tailing with A residues using poly-A polymerase (Affymetrix, Inc., Cleveland, OH). A-tailed RNA was reverse transcribed as recommended by the manufacturer. PCR was performed using 2.5 µl of cDNA, 10 pmol each of abridged universal amplification primer and MuV-specific primer NP437 [15]. To amplify the viral 5′ termini, RNA was extracted from 100 µl of virus stock and DNase treated RNA was reverse transcribed with Superscript II (5′RACE kit, Invitrogen) and 2.5 pmol of MuV-specific primer 14.75 [15]. Following RNase incubation, cDNA was column purified and 50 % of the recovered cDNA was tailed with C residues using terminal deoxynucleotidyl transferase (5′RACE kit) as recommended by the manufacturer. First round and second round PCR was carried out as described previously [15]. GoTaq hot start polymerase (Promega, Madison, WI) was used.

### Capillary sequencing of cDNA-derived viruses

Viruses recovered from cDNAs were either partially or fully sequenced as indicated in the results section. To determine the entire viral genomic sequence, RT-PCR was carried out using up to 15 different primer pairs that produced partially overlapping amplicons ranging in size from 629 to 1884 bp. For most viruses, only the regions encompassing the viral untranslated regions (UTR) as well as those parts of open reading frames (ORF) that were adjacent to the UTRs were determined (Table 1). Sequence information for all primers used for RT-PCR and for sequencing can be obtained from the authors upon request. All PCR fragments were gel purified using QIAquick gel extraction kit (Qiagen). Sequencing was carried out using an ABI 3730xl automated capillary sequencer (Life Technologies Corporation, Carlsbad, CA) by Macrogen USA Corp. (Rockville, MD). Sequence data were analyzed with the Chromas (Technelysium, Tewantin, Australia) and

**Table 1** Sequencing of UTRs and partial ORFs

| UTR[a] | Genomic UTR positions | Genomic area sequenced |
|---|---|---|
| N-5′ UTR | 55–146 | 22–380 |
| N-P UTR | 1795–1978 | 1311–2080 |
| P-M UTR | 3151–3263 | 3000–3903 |
| M-F UTR | 4391–4545 | 4110–5090 |
| F-SH UTR | 6163–6268 | 6047–6765 |
| SH-HN UTR | 6442–6613 | 6016–6739 |
| HN-L UTR | 8363–8437 | 8200–9060 |
| L-3′ UTR | 15,224–15,360 | 14,830–15,362 |

[a] Except for the N-5′ UTR and the L-3′ UTR, UTRs denote areas between the stop codon of the upstream ORF and the start codon of the downstream ORF including the intergenic regions

Jellyfish (LabVelocity, Los Angeles, CA) software packages.

### High-throughput sequencing (HTS)

1750 bp PCR samples (Table 2) amplified from viral RNA-derived cDNA or from plasmid DNA using primers FMIDfwd and 6784rev (see above) and encompassing viral genomic nucleotides 5035–6784 were used. In addition, as control for background that is intrinsic to HTS technology, a 1765 bp fragment each containing nucleotides 5035–6784 of virus strains Urabe AM9 and 88-1961 was excised from plasmids pUrabe1750 and p88-1750, respectively, using *Eco*RI. HTS was carried out by our in-house sequencing core facility. All samples were fragmented using a transposon-mediated method according to the recommendations of the manufacturer of the kit (Nextera kit, Illumina, San Diego, CA). Desired sample fragmentation was verified using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). High-throughput sequencing was carried out on a Miseq sequencer (Illumina) according to the manufacturer's instructions. Data analysis was done using the specialized HIVE platform available for FDA at https://minihive.fda.gov and for the public domain at the George Washington University hive.biochemistry.gwu.edu. HIVE is a multicomponent cloud infrastructure architected specifically for HTS applications including big-data storage, management, security, and computations. Two newly-developed HIVE algorithms were used to analyze the data presented herein. The first algorithm used, HIVE-hexagon [18], is a novel, highly customizable sequence aligner that exploits sequence characteristics, CPU, RAM, and Input/Output (I/O) architecture to efficiently and rapidly compute reference-based alignments. Once alignments were computed, HIVE-hexagon results were used as inputs for the HIVE-heptagon profiler and variant-calling tool. The second algorithm, HIVE-heptagon, computes the reference-based SNP profile by computing the occurrences of each nucleotide base for every reference genomic position. Frequencies of each called variant are then computed relative to the reference genome or to the accumulated consensus genome of the sample alignments. HIVE-heptagon also enables a number of post-alignment quality control procedures including noise filtration, entropic bias, and others.

## Results

### Rescue and characterization of recombinant MuV with extragenomic insertion of an additional U residue in the F gene-end signal

Within the scope of a project aimed to identify genetic determinants of neurovirulence of the Urabe AM9 vaccine strain of MuV, two recombinant chimeric cDNA plasmids encoding Urabe AM9-based viruses were generated: one in which the Urabe AM9 F gene was replaced with the corresponding gene from the Jeryl Lynn vaccine strain [pUrabe + JL(F)], and one in which the Urabe AM9 F and HN genes were replaced with corresponding genes from the Jeryl Lynn vaccine strain [pUrabe + JL(F/HN)]. Virus was rescued from these two cDNA plasmids three times each. The genomic region encompassing the F gene stop signal was sequenced for all 6 viruses. The presence of an additional U residue in the poly-U tract in the F gene stop signal was clearly evident in a significant portion of genomes in one of the three virus rescue preparations from each of the two cDNAs (Fig. 1a, b). By estimating the area under the chromatogram peaks obtained for these two virus preparations, approximately 80 % of rUrabe + JL(F) genomes and approximately 60 % of rUrabe + JL(F/HN) genomes contained the insertion. While there was no evidence for this insertion in one of the two other viruses rescued from pUrabe + JL(F), about 30 % of the genomes of the third virus were estimated to contain this insertion. Only about 10 % of the genomes of the two additional viruses rescued from pUrabe + JL(F/HN) contained the insertion (data not shown). The two viruses with the highest percentage of insertions were then fully sequenced, including the 3′ and 5′ genomic termini, to determine if compensatory deletions existed elsewhere to restore the virus genome to a polyhexameric length. No deletions were identified, suggesting that these virus rescues did not comply with the rule of six. Of note, both fully sequenced viruses also displayed an insertion of an additional U

**Table 2** Additional genomic changes detected in four fully sequenced plaque-purified viruses containing the insertion of the additional U residue at the F gene-end signal

| Virus | Nucleotide change | Amino acid change | Gene |
|---|---|---|---|
| rUrabe + JL(F)-PPF(8A)-1 | A14214G | N1926S | L |
| rUrabe + JL(F)-PPF(8A)-2 | A1184U | M347L | N |
| rUrabe + JL(F/HN)-PPF(8A)-1 | C6185U | Non-coding | F 3′ UTR |
| | C14198A | V1921L | L |
| rUrabe + JL(F/HN)-PPF(8A)-2 | A6413C | F49C | SH |

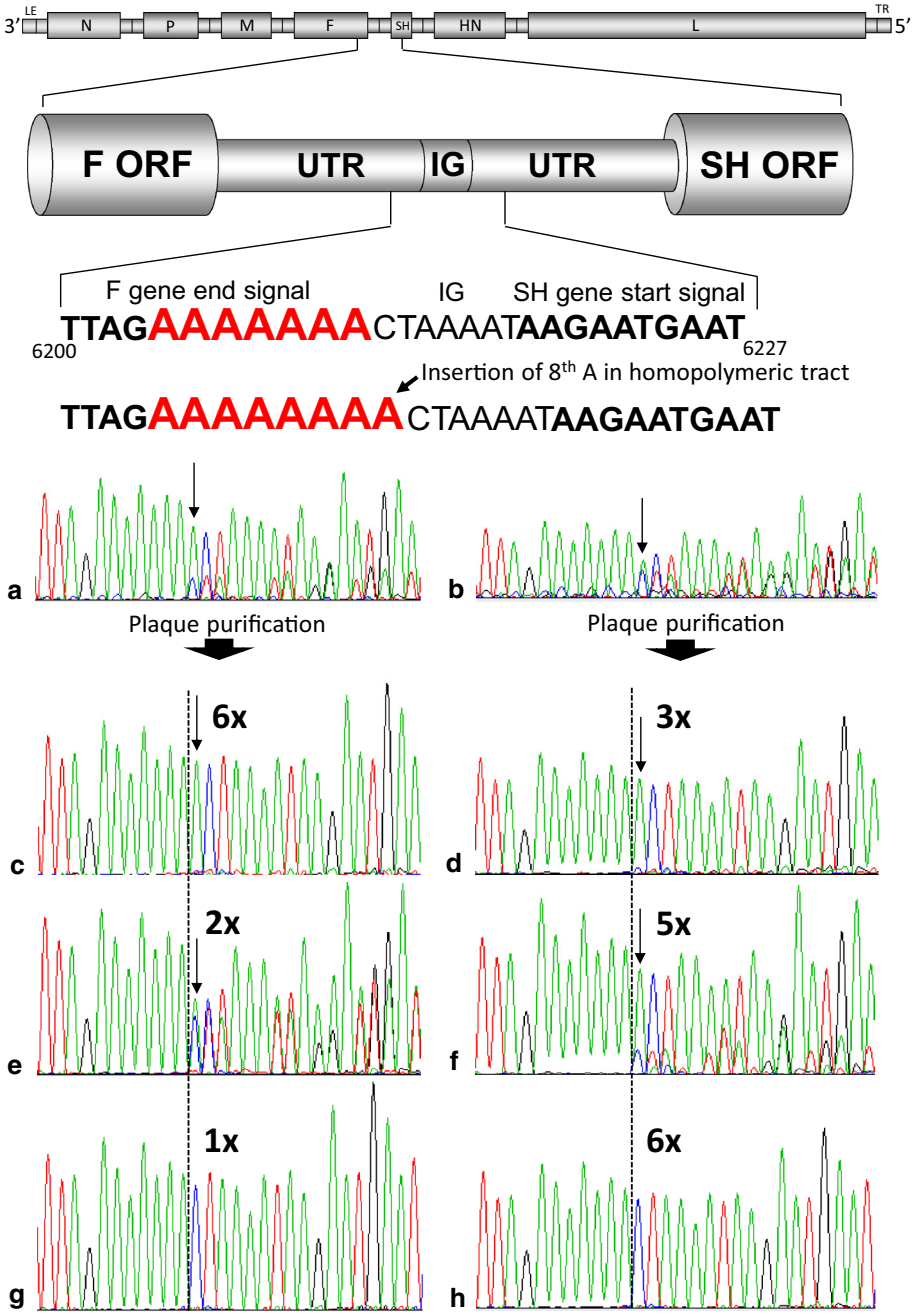


**Fig. 1** Recovery of recombinant MuVes with insertions of an additional U residue in the F gene-end homopolymeric tract. The genomic structure of MuV is depicted at the *top* of the figure. The untranslated (UTR) region between the F and SH open reading frames (ORF) is shown in more detail, highlighting the antigenomic nucleotide sequence of the F gene-end signal, the intergenic (IG) region, and the SH gene start region (MuV strain Jeryl Lynn (major variant, gene bank accession no. AF338106). The homopolymeric tract of 7 A residues (=U in genomic orientation) in the F gene-end signal is shown in *red* text. The insertion of an 8th A in the homopolymeric tract is indicated. *Numbers* indicate the nucleotide position in the genome. Chromatograms shown in **a** and **b** show, respectively, the corresponding sequences of RT-PCR fragments derived from the one rUrabe + JL(F) and the one rUrabe + JL(F/HN) rescue containing the insertion (*arrow*) at a high percentage, leading to a stretch of 8 instead of 7 A residues. Note a shift in downstream sequence caused by the insertion. Sequences representative of nine plaque-purified viruses derived from rUrabe + JL(F) and of 14 plaques derived from rUrabe + JL(F/HN), are shown below (**a**) and (**b**), respectively. These include plaques that were homogeneous for the insertion (**c**, **d**), plaques that were heterogeneous at this position (**e**, **f**), and plaques that lacked the insertion (**g**, **h**). *Numbers* in **c**–**h** indicate the number of plaques with the respective genotypes. *LE* leader, *TR* trailer

residue in the poly-U tract of the L gene-end signal (tract of 6 U residues) in about 5–10 % of genomes.

### Identification and propagation of plaque-purified viruses that harbor an additional U residue in the F gene-end signal

Given that not all of the amplified genomes in the two rUrabe + JL(F) and rUrabe + JL(F/HN) virus rescues displayed the nucleotide insertion in the F gene-end signal (Fig. 1a, b), it was unclear whether the genomes containing the inserts were associated with infectious particles, or if these represented non-viable particles or non-virion-associated genomic RNA molecules. In order to address this question, nine rUrabe + JL(F) viruses and 14 rUrabe + JL(F/HN) viruses were plaque picked and separately passaged on Vero cells once to generate material for titration and sequencing. Of the nine plaque-purified rUrabe + J(F) viruses, six (67 %) appeared to contain only genomes with the insertion in the F gene-end signal, two (22 %) contained mixed populations (some with and some without the insertion), and one (11 %) lacked any evidence of the insertion (Fig. 1c, e, g). For the 14 plaque-purified rUrabe + JL(F/HN) viruses, three (21 %) appeared to contain only genomes with the insertion in the F gene-end signal, five (36 %) contained mixed populations, and six (43 %) lacked any evidence of the insertion (Fig. 1d, f, h). These results are in agreement with the sequence heterogeneity observed with the parental viruses. The fact that viruses harboring 8 U residues in the F gene-end signal could readily be plaque purified indicates that these viruses are replication competent and were derived from infectious particles rather from non-infectious particles present in the parental virus or from non-virion-associated genomic RNA molecules.

With regard to a single picked plaque yielding genomes with and without the insertion, while uncommon, we have observed this phenomenon previously. It is unlikely to be due to the plaque purification technique, since plaques are only picked in regions of generous separation between neighboring plaques. Whether such a phenomenon represents polyploid viruses (i.e., particles that carry genomes with both seven U and eight U residues in the F gene-end signal) or whether they are indicative of loss of the additional U during the propagation of the purified plaque in cell culture is unknown. Alternatively, the picked plaques might have arisen from settling of an aggregate of two or more viruses of which some carried seven U's and others eight U's.

To analyze whether genomes with the insertion acquired compensatory mutations elsewhere in the genome that would correct the genome length, the six rUrabe + JL(F) and three rUrabe + JL(F/HN) plaque-purified viruses that homogeneously expressed the insertion were sequenced. We initially sequenced the viruses only at sites where a deletion would likely be tolerated, i.e., the untranslated regions and the hypervariable, dispensable SH gene. Being in close vicinity to the UTRs, the 3′ and 5′ end regions of the remaining ORFs also were covered by sequencing. Two of the three partially sequenced rUrabe + JL(F) viruses exhibited single deletions, one being deletion of a U residue in the SH gene-end signal at position 6523 (in an estimated 50–60 % of cDNAs; data not shown) and the other being a deletion of a cytosine (C) residue at genomic position 15,097, located in the L gene downstream of a homopolymeric tract of 6 U residues (genomic) and upstream of 2 U residues (in an estimated 10 % of cDNAs; data not shown). The deletion would result in predicted changes in amino acids 2221–2228 of the L protein including a stop codon leading to a premature stop of the L protein shortening it by the 33 most C-terminal amino acids.

The third partially sequenced virus showed no evidence of a deletion. The remaining three plaque-picked rUrabe + JL(F) viruses were fully sequenced and revealed no deletions. Among the three plaque-purified rUrabe + JL(F/HN) viruses that homogeneously expressed the F-gene-end insertion, two were fully sequenced and one was partially sequenced. Only in one of the two fully sequenced viruses [named rUrabe + JL(F/HN)-PPF(8A)-1] did we find evidence for deletion of a single residue. The deletion was identified in a homopolymeric tract of five A residues at position 14,327 in the L gene. However, this deletion was barely detectable in the chromatogram amounting to at most 10 % of all cDNAs amplified (data not shown).

All nine plaque-purified viruses that were homogenous for the insertion grew to typical titers, ranging from $4.48 \times 10^6$ to $6.92 \times 10^7$ pfu/ml, which were not significantly different from the titers measured for the parental viruses from which they were derived. To further evaluate the ability of viruses with an additional U residue to replicate efficiently, two fully sequenced plaque-purified rUrabe + JL(F) and rUrabe + JL(F/HN) viruses with an additional U residue and one partially sequenced plaque-purified rUrabe + JL(F) and rUrabe + JL(F/HN) virus without the insertion were tested in multicycle growth kinetics in Vero cells together with the parental viruses from which they were derived (Fig. 2). All viruses reached maximum titers around 2–3 days post-infection, by which time all cells were infected as evidenced by syncytium formation throughout the entire cell layer. Due to subsequent lysis of all cells, virus titers gradually decreased thereafter. One of the two plaque-purified rUrabe + JL(F) viruses containing the insertion [rUrabe + JL(F)-PPF(8A)-2] grew to the same maximum titer with similar kinetics as the plaque-purified virus without the insertion [rUrabe + JL(F)-PPF(7A)], while the other plaque-purified rUrabe + JL(F) virus containing the insertion [rUrabe + JL(F)-PPF(8A)-1] grew to a titer that was nearly one log higher (Fig. 2a). The opposite pattern was

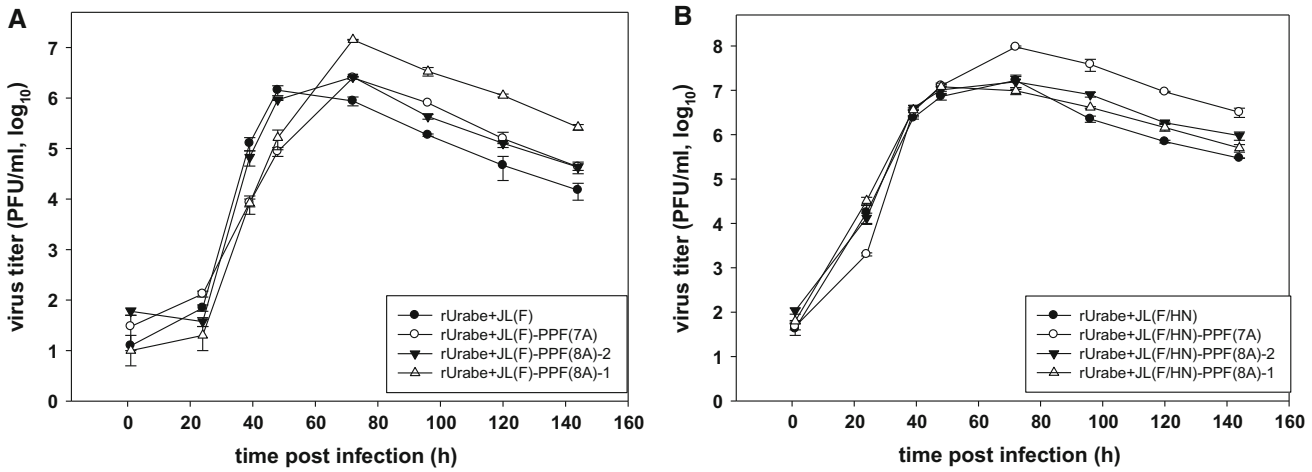


**Fig. 2** Multicycle growth kinetics of recombinant and plaque-purified viruses. Each two plaque-purified viruses containing eight U residues [rUrabe + JL(F)-PPF(8A)-1, rUrabe + JL(F)-PPF(8A)-2, rUrabe + JL(F/HN)-PPF(8A)-1, rUrabe + JL(F/HN)-PPF(8A)-2] and each one plaque-purified virus containing seven U residues in the F gene-end signal (rUrabe + JL(F)-PPF(7A), rUrabe + JL(F/HN)-PPF(7A), derived from the two parental recombinant viruses rUrabe + JL(F) and rUrabe + JL(F/HN), together with their respective non-plaque-purified parental viruses were analyzed in multicycle growth kinetics on Vero cells (m.o.i. = 0.05). Each virus was tested in duplicate. *Error bars* depict SEM values

seen with the rUrabe + JL(F/HN) viruses where the plaque-purified virus without the insertion [rUrabe + JL(F/HN)-PPF(7A)] grew to a maximum titer that was nearly one log higher compared to the two plaque-purified viruses containing the insertion. Of note, all four of the plaque-purified viruses containing the insertion also contained mutations elsewhere in the genome (Table 2). Although these were not compensatory changes to correct genome length, we cannot rule out that these may have influenced virus replication properties. The two plaque-purified 7A viruses were only sequenced in the region encompassing the F gene-end signal. Thus, we cannot exclude that these viruses also displayed mutations in their genome, which is quite likely, given that RdRps exhibit high mutation rates and mutations become fixed in plaque-purified viruses.

These data indicate that MuVs that violate the rule of six are not compromised in their ability to replicate in vitro. Furthermore, there was no evidence of a strong pressure to correct the F gene-end sequence, i.e., to revert from 8 to 7 U residues. This was demonstrated by additional passage in Vero cells of two representative rUrabe + JL(F) and two representative rUrabe + JL(F/HN) plaque-picked viruses homogeneous for the insertion. After passage, these four viruses were sequenced at the region encompassing the F gene-end signal. All still contained the insertion. Whether these particular viruses acquired corrective deletions elsewhere in the genome was not evaluated, but complete sequencing of other plaque-picked viruses homogeneous for the insertion showed no evidence of deletions elsewhere, as described above.

In order to further investigate any replicative disadvantage of possessing an additional U in the F gene-end signal, the original (non-plaque purified) rescued viruses rUrabe + JL(F) and rUrabe + JL(F/HN) that were composed of a mixture of genomes harboring and lacking the insertion were subjected to an additional passage in Vero cells. The F gene-end signal region in virus-containing supernatants from five independent cultures per virus was sequenced. All five of the rUrabe + JL(F) passaged cultures displayed a similar level of heterogeneity at the insertion site compared to the parental virus. In contrast, whereas the original rescue of rUrabe + JL(F/HN) displayed a slight preponderance of genomes containing the insertion (Fig. 1b), only a minority of the genomes in each of the five rUrabe + JL(F/HN) passaged cultures contained the insertion (approximately 30 %, data not shown). This result is in agreement with the growth kinetics shown in Fig. 2 indicating different preferences for different viruses, but no unified trend suggesting a replicative disadvantage of possessing an additional U in the F gene-end signal.

## Analysis of rescued and plaque-purified viruses in the rat neurovirulence test

Although there did not appear to be an effect of non-compliance with the rule of six on virus growth in vitro, we sought to determine if such an effect might exist in vivo. To this end, we tested in rats the original, non-plaque-purified rUrabe + JL(F) and rUrabe + JL(F/HN) viruses as well as one representative plaque-purified rUrabe + JL(F) and rUrabe + JL(F/HN) virus that lacked the

insertion [viruses rUrabe + JL(F)-PPF(7A) and rUrabe + JL(F/HN)-PPF(7A)]. As shown in Fig. 3, no statistically significant differences in neurovirulence scores were observed in rats [Mann–Whitney Rank Sum Test; $P = 0.540$ for rUrabe + JL(F) vs rUrabe + JL(F)-PPF(7A) and $P = 0.753$ for rUrabe + JL(F/HN) vs rUrabe + JL(F/HN)-PPF(7A)]. In order to determine whether the level of heterogeneity at the F gene end insertion site observed in the parental rUrabe + JL(F) and rUrabe + JL(F/HN) viruses changed during replication in rat brains, which would be indicative of a replicative advantage of one of the two insertion variants in vivo, brains from three rats each infected with one of the two viruses were harvested at day three post-infection, a time point of maximum viral replication in newborn MuV-infected rats [10, 15]. Virus-specific RNA was amplified by RT-PCR from RNA prepared from the brains and a 500 bp or 800 bp, respectively, PCR fragment encompassing the F gene-end signal was sequenced. Based on evaluation of chromatograms, we estimated that 60, 70, and 80 %, respectively, of genomes of the three viruses sequenced from brains infected with rUrabe + JL(F) contained the additional U residue, whereas the percentage of these genomes in the three viruses sequenced from brains infected with rUrabe + JL(F/HN) was estimated to be 60, 60, and 70 %, respectively (data not shown). Thus, viruses with genomes containing the insertion did not appear to have a replicative disadvantage in the rat brain. Given that the majority of viruses replicating in brains harbored the insertion, these data suggest that these viruses do not exhibit neurotoxicity significantly different from those obtained with viruses without the additional U residue, since this should have been detectable in this experimental setting.

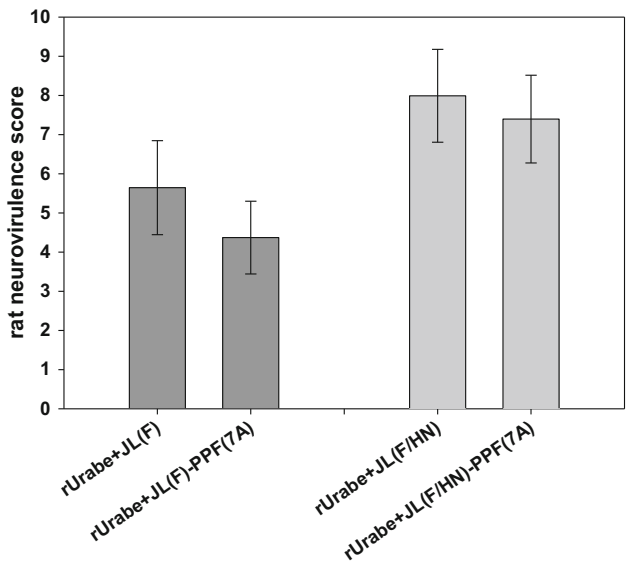


**Fig. 3** Analysis of recombinant and plaque-purified viruses in the rat assay. Parental recombinant viruses rUrabe + JL(F) and rUrabe + JL(F/HN) and a plaque-purified virus derived from each [rUrabe + JL(F)-PPF(7A) and rUrabe + JL(F/HN)-PPF(7A)] containing only seven U residues in the F gene-end signal were analyzed in newborn rats for neurovirulence. *Error bars* depict SEM values

## Insertions of additional nucleotides in the F gene-end signal by T7 RNA polymerase

Nucleotide insertions occur as a function of errant polymerase activity, most often as a consequence of polymerase "stuttering" at homopolymeric tracts [19]. To determine if the insertions we observed in poly-U tracts were due to the activity of the viral RNA-dependent RNA polymerase or the T7 DNA-dependent RNA polymerase used in our rescue system to generate antigenomic viral RNA, we subcloned a 1750-bp fragment corresponding to genomic positions 5058 through 6762 of the Urabe AM9 MuV strain, which includes several homopolymeric U tracts, including the tract of seven U residues in the F gene-end signal (positions 6204–6210), a tract of six U residues in the SH gene-end signal (positions 6528–6533); a tract of four U residues in the F ORF (positions 5580–5583), and a tract of four U residues in the F 3′UTR (positions 6193–6196). The subcloned fragment from the resulting plasmid pUrabe1750 was in vitro transcribed into RNA using T7 DNA-dependent RNA polymerase and then subjected to RT-PCR to amplify the 1750 bp sequence, which was analyzed by HTS. To monitor for background noise inherent to the HTS technology, plasmid DNA encompassing the same genomic positions (5058–6762) but derived from two different MuV cDNA clones (p88-1961modΔNhe and $pURABE_{HN\text{-}K335}$) was included in the analysis. To account for mutations that are introduced by the DNA polymerase during the PCR step, PCR fragments amplified using plasmids p88-1961modΔNhe as well as $pURABE_{HN\text{-}K335}$ as template were included as well.

Using algorithms developed by Simonyan and others [18, 20], we identified and evaluated each position in the 1750 bp amplified region that displayed insertions or deletions of one or more nucleotides. The frequency of reads (in percent of total reads covering the respective position) that display insertions and/or deletions at the four selected homopolymeric tracts is presented in Table 3. Additional insertions and deletions identified beyond these four sites are included in the table if they occurred at a frequency considerably above background noise ($\geq$0.25 %) of all reads at the respective specific position. For plasmid DNA (control samples 1 and 2) the frequency of single nucleotide insertions and deletions in each of the four homopolymeric tracts was very low, all being $\leq$0.035 %. Since the material sequenced was plasmid DNA, this level represents errors introduced during the HTS procedure or background noise. Similarly low levels

were also observed in single nucleotide insertion and deletion frequencies in plasmid-derived PCR fragments (control samples 3 and 4), with the exception of slightly higher deletion frequencies being apparent in the homopolymeric tract in the F gene-end signal (0.125 % for samples 3 and 0.157 % for sample 4). In contrast to these low levels, single nucleotide insertion frequencies were significantly higher in T7 RNA polymerase in vitro transcribed material (Sample 5), with values of 5.16 % in the F gene-end signal and 1.38 % in the SH gene-end signal. Interestingly, T7 RNA polymerase generated single nucleotide deletions of one U residue in the F gene-end signal with a frequency of 1.66 %. Moreover, there was a small but significant increase in the frequency of insertions of U residues in the homopolymeric tracts in the F ORF and the F 3′-UTR (0.136 and 0.093 %, respectively). Finally, we identified the additional presence of insertions of one adenine (A) residue in homopolymeric tracts of four A residues at genomic positions 6020 (F gene) and 6423 (SH gene) in 0.32 and 0.27 %, respectively, of reads.

Intriguingly, insertions in the homopolymeric tract in the F gene-end signal in the RNA that was in vitro transcribed with T7 RNA polymerase were not limited to single nucleotide insertions. While the latter represented the most frequent insertion event, we also identified, with decreasing frequency, insertions of two (0.53 %), three (0.09 %), four (0.02 %), and five U residues (one read only), as shown in Fig. 4.

Taken together these data suggest that the frequent observation of a significant percentage of genomes with insertions in the homopolymeric tract of the F gene-end signal following recovery of recombinant viruses from BHK-BSRT7/5 cells is at least in part attributable to the stuttering of the T7 RNA polymerase at this site.

To analyze whether the viral RdRp also is contributing to the insertion of additional nucleotides at the F gene-end signal, we sequenced the genomic regions encompassing positions 5058 through 6762 of non-recombinant MuV strains 88-1961 and Urabe P-AM9. As shown in Table 3 (Samples 6 and 7), viruses Urabe P-AM9 and 88-1961 also displayed the insertion of an additional U residue in the F gene-end signal albeit at low frequencies (0.59 and 0.82 %, respectively), yet significantly higher than background levels. The same was true for the frequency of insertions of an extra U residue at the SH gene-end signal found in the Urabe P-AM9 and 88-1961 viruses (0.39 and 1.06 %, respectively). In contrast, the frequencies of deletions in the F and SH gene-end signals in the two viruses were only marginally increased in comparison to the background values. Likewise, the frequencies of insertions/deletions in

**Table 3** Evaluation by HTS of the frequency of single nucleotide (U) insertions/deletions at selected homopolymeric tracts in Urabe AM9 and 88-1961 derived genomes as well as in RNA transcribed by T7 RNA polymerase

| Sample# | Origin of DNA or cDNA[a] | Homopolymeric tract in F gene-end signal (6204–6210)[b] | | Homopolymeric tract in SH gene-end signal (6528–6533)[b] | | Homopolymeric tract in F ORF (5580–5583)[b] | | Homopolymeric tract in F 3′ UTR (6193–6196)[b] | |
|---|---|---|---|---|---|---|---|---|---|
| | | Ins | Del | Ins | Del | Ins | Del | Ins | Del |
| 1 | Plasmid DNA (derived from p88-1961modΔNhe) | 0.011 | 0.015 | 0 | 0 | 0 | 0.009 | 0 | 0.019 |
| 2 | Plasmid DNA (derived from $pURABE_{HN\text{-}K335}$) | 0.014 | 0.035 | 0.003 | 0.007 | 0 | 0.008 | 0 | 0 |
| 3 | PCR using p88-1961modΔNhe as template | 0.004 | **0.125** | 0 | 0.027 | 0 | 0.005 | 0 | 0 |
| 4 | PCR using $pURABE_{HN\text{-}K335}$ as template | 0.013 | **0.157** | 0.006 | 0.035 | 0.001 | 0.001 | 0 | 0.01 |
| 5 | RT-PCR; RNA in vitro transcribed with T7 RNA-dependent RNA polymerase from plasmid pUrabe1750[c] | **5.16** | **1.66** | **1.38** | 0.035 | **0.136** | 0.021 | 0.093 | 0.011 |
| 6 | RT-PCR; RNA from Urabe P-AM9 vaccine virus (not recombinant) | **0.59** | **0.31** | **0.39** | 0.086 | 0 | 0.024 | 0.002 | 0.015 |
| 7 | RT-PCR; RNA from 88-1961 clinical isolate (not recombinant) | **0.82** | **0.27** | **1.06** | **0.134** | 0 | 0.024 | 0 | 0.017 |

[a] All sequenced samples encompassed genomic positions 5035–6784

[b] Values are expressed in percentage of all reads at the respective positions

[c] Additional presence of insertion of one residue (A) in homopolymeric tracts of four A's at genomic positions 6020 (F gene) and 6423 (SH gene) in 0.32 and 0.27 %, respectively, of reads

All values ≥0.1 % are highlighted in bold letters; Insertions/deletions identified in positions in addition to the four that were listed in the table are only mentioned if the percentage values at the respective positions amounted to ≥0.25 % of reads


**Fig. 4** Determination by HTS of the frequency of insertions of one to six U residues in the F gene-end sequence in RT-PCR fragments derived from RNA that was in vitro transcribed with T7 RNA polymerase. Depicted on the *left* are the various species of sequences identified by HTS in the F gene-end and F-SH intergenic region (IG) in RT-PCR products derived from in vitro transcribed RNA from plasmid pUrabe1750. The number of reads identified by HTS that harbor the respective species are indicated on the right. *Numbers in parentheses* indicate the percentage of the reads compared to the total number of reads over this stretch of sequence. The wild-type "7 U" genomic sequence is shown at the *top*, being the most prevalent species. Below that sequence, the same region is depicted with a C6203U mutation observed in some reads, resulting in a homopolymeric tract of 9 Us ("9 U"). Insertions of one to six uridine residues identified in the 7 U sequence and of one or two uridine residues in the 9 U sequence are highlighted in *red*. All sequences with the C6203U mutation are marked with an *asterisk*

the homopolymeric tracts in the F ORF and the F 3′UTR were not higher than background levels. Taken together, these data indicate that the frequently observed insertion of an additional genomic U residue in the F gene-end signal following recovery of recombinant MuV from BHK-BSR T7/5 cells is primarily a consequence of T7 polymerase (i.e., introduced as errors during the rescue process). However, such insertions can also be induced by the viral RdRp, albeit at a much reduced frequency.

## Discussion

The "rule of six" is based on the observation that all members of the *Paramyxovirinae* have genome lengths that are multiples of six [11–13]. This ostensibly applies to MuV as well, a member of the paramyxovirinae subfamily, given that all MuV sequences published to date display a genome length of 15,384 nucleotides, a polyhexameric number. However, our identification of recombinant MuVs that exhibited a mixture of genomes displaying seven (wild type) or eight (insertion) U residues in the F gene-end signal suggested that MuV, while clearly preferring a polyhexameric gene length based on sequence data from natural isolates, may not absolutely require it. To examine this issue, we characterized several plaque-purified viruses derived from parental viruses containing F gene-end signals composed of seven or eight U residues.

The wild-type MuV F gene-end signal encompasses a homopolymeric tract of seven U residues and is postulated to serve as the F gene polyadenylation signal. The length and integrity of the poly-U tract is known to influence termination of upstream gene transcription and re-initiation for downstream gene transcription [21, 22]. While we did not investigate if lengthening of the F polyadenylation signal from 7 to 8 U residues impacts F-gene termination and/or enhances re-initiation of transcription of the downstream gene (SH), we did not observe any significant impact of the additional nucleotide on the ability of the virus to replicate in vitro or to affect virus virulence in vivo. Despite this phenotypic indifference, there was some evidence that these viruses attempted to restore their genome length to a polyhexameric length, either via an effort to eliminate the additional U residue, as evident upon additional passage of the original rescued rUrabe + JL(F/HN), or via compensatory deletions elsewhere in the genome, as was evident in some plaque-purified cultures of rUrabe + JL(F) and rUrabe + JL(F/HN). However, such corrections were not universally observed, suggesting that MuV does not require a polyhexameric genome.

For those fully sequenced viruses that appear to be violators of the rule of six (i.e., contain an insertion and no observable compensatory deletion) we cannot exclude the possibility that such deletions exist at a level not detectable by capillary sequencing. For example, within the pool of virus particles it is possible that every genome exhibits a deletion somewhere to regain compliance with the rule of six, but such single-genome variants would not be detectable by standard capillary sequencing. While such a possibility is highly unlikely, studies employing high-throughput sequencing technology to examine this prospect are planned.

That the rule of six is not as strict a requirement for all members of the paramyxovirinae was also suggested by Murphy and Parks [23] in their studies of replication of defective interfering (DI) genomes of parainfluenza virus 5 and by Durbin et al. [24] in minigenome analyses of human parainfluenza virus type 3 transcription and replication.

Since the technology employed to rescue recombinant MuV involves generation of the primary antisense genomes from transfected cDNA using T7 RNA polymerase, it was conceivable that the nucleotide insertion frequencies (less than 10–80 %) observed in the F gene-end signal in recombinant viruses could be due to the action of the T7 RNA polymerase. This hypothesis is strongly supported by the fact that RNA in vitro transcribed using T7 RNA polymerase exhibited insertions of an additional nucleotide at a frequency of about 5 %. Thus, the frequently observed insertions in the F gene-end signal could be in fact due to the action of the T7 RNA polymerase. Nevertheless, it cannot be ruled out that the viral RdRp also is capable of inserting an additional nucleotide in the F gene-end signal, since high-throughput sequencing of the region encompassing the F gene-end signal of non-recombinant MuV 88-1961 and Urabe P-AM9 also revealed existence of virus subpopulations that display the additional nucleotide insertion-albeit at an extremely low percentage. The latter might be due to the preference of MuV genomes to adhere to the rule of six. However, it also should be mentioned that we cannot exclude the possibility that the reverse transcriptase used to convert RNA into cDNA during the RT-PCR also erroneously inserts additional U residues at homopolymeric tracts and thus might account for the low percentage of insertions seen in non-recombinant viruses MuV 88-1961 and Urabe P-AM9. Analysis of the contribution of the RT enzyme to the observed frequency of insertions in the F gene-end signal in the non-recombinant viruses Urabe P-AM9 and 88-1961 requires chemical synthesis of an RNA fragment that encompasses the F gene-end signal, followed by RT-PCR analysis of such a fragment. Synthesis of such an RNA fragment with a length that is amenable to RT-PCR is technically not feasible at present, precluding the assessment of the frequency of insertions at the F gene-end signal attributable to the RT enzyme. However, within the scope of a separate, ongoing project to characterize recombinant MuV, we have observed frequencies of insertions of additional U residues at the F and SH gene-end signals that were below those observed for viruses Urabe P-AM9 or 88-1961 (0.32 % for the F gene-end signal and 0.14 % for the SH gene-end signal). Thus, while it is unknown whether insertion frequencies at the F or SH gene-end signal of 0.32 and 0.14 %, respectively, are due to errors introduced by the RT enzyme, we reason that insertion frequencies above these values, as observed with rUrabe P-AM9 and r88-1961 (0.59 and 0.82 %, respectively, for the F gene-end signal and 0.39 and 1.06 %, respectively, for the SH gene-end signal), cannot be solely, or even substantially, be attributable to the activity of the RT enzyme.

That homopolymeric tracts are targets for insertions is also evident in viruses for which the rule of six does not apply, i.e., viruses that are not members of the *Paramyxovirinae,* such as Ebola virus and human metapneumovirus (HMPV). In the case of Ebola virus, the native homopolymeric tract of 7 U residues in the glycoprotein gene is used to generate the secreted form of the glycoprotein, whereas addition of a non-templated 8$^{th}$ A residue into this region of the mRNA gives rise to the full-length glycoprotein [25]. Intriguingly, expression of the glycoprotein mRNA by an in vitro transcription system employing T7 RNA polymerase or vaccinia virus RNA polymerase and using non-edited glycoprotein cDNA as template revealed that 1–5 % of the mRNA produced contained insertion of an 8th A at the homopolymeric tract of seven A residues. This percentage is very similar to our results with in vitro transcribed RNA using T7 RNA polymerase. It was concluded by the authors of the work that the editing site in the Ebola virus glycoprotein gene is recognized not only by Ebola virus polymerase but also by DNA-dependent RNA polymerases of different origin. Similarly, we show here evidence that insertion of U residues into a homopolymeric tract in the F gene-end signal of the MuV is mediated not only by the viral RdRp, but also by the DNA-dependent T7 RNA polymerase.

In addition to cotranscriptional editing displayed by the Ebola virus RdRp that gives rise to the 8 A glycoprotein mRNAs species, it was shown that the 7 U editing site in the genome is also prone to genomic RNA editing in that passage of Ebola virus with 7 genomic U residues in Vero E6 cells results in the appearance and rapid accumulation of a variant containing an additional U at the editing site in the viral genome [26]. In contrast to our studies, where the 8 U MuVs did not display a replicative advantage over 7 U MuVs, the 8 U Ebola virus variant outgrew and eventually replaced the wild-type 7 U genotype during 4–5 passages [26]. It should also be mentioned that in contrast to the in vitro situation, infection of guinea pigs with the 8 U Ebola virus variant led to a rapid reversion to the 7 U wild-type genotype [26].

Insertions in HMPV homopolymeric U tracts also occur frequently for viruses rescued using a T7 RNA polymerase-based system [27]. Most insertion events occurred in the SH gene which exhibits 12 different homopolymeric tracts of four or more U residues. Insertions were also observed in homopolymeric U tracts [putative poly(A) signal] in the HMPV M2 and F gene-end signals, although less frequently. For MuV, the fact that most insertions were observed in the F gene-end signal likely reflects the fact that

this is the region of the genome with the longest run of U residues. All other gene-end signals are composed of six or less consecutive U residues. Based on the work described herein, it appears that the larger the homopolymeric tract, the higher the frequency of RdRp stuttering, and the greater the chance of erroneous insertions. While the MuV genome contains several other homopolymeric tracts of as many as six consecutive U residues, these are within coding regions where insertions would not be tolerated due to the resulting frameshift which would likely compromise virus viability. Such genomes would be exceedingly difficult to detect. In summary, our data suggest that insertions of extra nucleotides at gene-end non-coding homopolymeric tracts of four or more U residues can occur and are tolerated. These "mutations" appear mainly to be the result of use of virus rescue systems that rely on T7 RNA polymerase. While these insertions are predicted to result in a non-polyhexameric genome length, suggesting that MuVs do not strictly obey the rule of six, confirmation of such a conclusion will require more sophisticated sequencing and bioinformatics methods. As to whether rule of six violations are generalizable for other MuV strains, we believe this to be the case, but only for strains generated in the laboratory, given our conclusion that these insertions arise mainly as a consequence of T7 RNA polymerase activity.

**Acknowledgments** Salary support for L.N., C.Z, Y.C., and M.L. was provided by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and the US Food and Drug Administration. We wish to thank Shen Rongfong and Rong Wang (FDA/CBER) for helpful discussion about HTS; U. Buchholz (NIH, Bethesda, MD) for providing the BHK BSR-T7/5 cells, P. Duprex for providing pMuV(MPBS). We are grateful to Ewan Plant and Konstantin Chumakov (FDA/CBER) for critically reading the manuscript.

## References

1. R.G. Paterson, R.A. Lamb, J. Virol. **64**, 4137–4145 (1990)
2. K. Takeuchi, K. Tanabayashi, M. Hishiyama, A. Yamada, Virology **225**, 156–162 (1996)
3. R.L. Wilson, S.M. Fuentes, P. Wang, E.C. Taddeo, A. Klatt, A.J. Henderson, B. He, J. Virol. **80**, 1700–1709 (2006)
4. T. Malik, C.W. Shegogue, K. Werner, L. Ngo, C. Sauder, C. Zhang, W.P. Duprex, S. Rubin, J. Virol. **85**, 6082–6085 (2011)
5. S.A. Rubin, C.J. Sauder, K.M. Carbone, in *Fields Virology*, ed. by D.M. Knipe, P.M. Howley (Lippincott Williams & Wilkins, Philadelphia, 2013), pp. 1024–1041
6. S.A. Plotkin, Pediatr. Infect. Dis. J. **32**, 381–382 (2013)
7. S. Rubin, J. Beeler, Pediatr. Infect. Dis. J. **32**, 1156–1157 (2013)
8. M.C. Bonnet, A. Dutta, C. Weinberger, S.A. Plotkin, Vaccine **24**, 7037–7045 (2006)
9. C.J. Sauder, C.X. Zhang, M.A. Link, W.P. Duprex, K.M. Carbone, S.A. Rubin, Vaccine **27**, 5822–5829 (2009)
10. C.J. Sauder, C.X. Zhang, L. Ngo, K. Werner, K. Lemon, W.P. Duprex, T. Malik, K. Carbone, S.A. Rubin, J. Virol. **85**, 7059–7069 (2011)
11. P. Calain, L. Roux, J. Virol. **67**, 4822–4830 (1993)
12. S. Hausmann, J.P. Jacques, D. Kolakofsky, RNA **2**, 1033–1045 (1996)
13. D. Kolakofsky, L. Roux, D. Garcin, R.W. Ruigrok, J. Gen. Virol. **86**, 1869–1877 (2005)
14. U.J. Buchholz, S. Finke, K.K. Conzelmann, J. Virol. **73**, 251–259 (1999)
15. C.J. Sauder, K. Vandenburgh, R.C. Iskow, T. Malik, K.M. Carbone, S.A. Rubin, Virology **350**, 48–57 (2006)
16. G. Amexis, S. Rubin, N. Chatterjee, K. Carbone, K. Chumakov, J. Med. Virol. **70**, 284–286 (2003)
17. K. Lemon, B.K. Rima, S. McQuaid, I.V. Allen, W.P. Duprex, J. Virol. **81**, 8293–8302 (2007)
18. L. Santana-Quintero, H. Dingerdissen, J. Thierry-Mieg, R. Mazumder, V. Simonyan, PLoS One **9**, e99033 (2014)
19. S. Hausmann, D. Garcin, C. Delenda, D. Kolakofsky, J. Virol. **73**, 5568–5576 (1999)
20. V. Simonyan, R. Mazumder, Genes **5**, 957–981 (2014)
21. J.C. Rassa, G.M. Wilson, G.A. Brewer, G.D. Parks, Virology **274**, 438–449 (2000)
22. E.E. Hinzman, J.N. Barr, G.W. Wertz, J. Virol. **76**, 7632–7641 (2002)
23. S.K. Murphy, G.D. Parks, Virology **232**, 145–157 (1997)
24. A.P. Durbin, J.W. Siew, B.R. Murphy, P.L. Collins, Virology **234**, 74–83 (1997)
25. V.E. Volchkov, S. Becker, V.A. Volchkova, V.A. Ternovoj, A.N. Kotov, S.V. Netesov, H.D. Klenk, Virology **214**, 421–430 (1995)
26. V.A. Volchkova, O. Dolnik, M.J. Martinez, O. Reynard, V.E. Volchkov, J. Infect. Dis. **204**(Suppl 3), S941–S946 (2011)
27. S. Biacchesi, B.R. Murphy, P.L. Collins, U.J. Buchholz, J. Virol. **81**, 6057–6067 (2007)