Characterization of the non-coding regions of the
1918 influenza A H1N1 virus

Ruixue Wang and Jeffery K. Taubenberger

Viral Pathogenesis and Evolution Section, Laboratory of Infectious Diseases, National Institutes
of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA 20892

Running title: 1918 influenza noncoding region sequences

Journal of Virology short format paper

Abstract: 135
Main text: 1549
References: 22

Address correspondence to:
Jeffery K. Taubenberger, M.D., Ph.D.,
Chief, Viral Pathogenesis and Evolution Section
Laboratory of Infectious Diseases
National Institute of Allergy and Infectious Diseases
National Institutes of Health
33 North Drive, Room 3E19A.2 MSC 3203
Bethesda, MD 20892-3203 USA
Tel. 1-301-443-5960; Fax. 1-301-480-1696
email: taubenbergerj@niaid.nih.gov
Abstract

The terminal non-coding region (NCR) sequences of the eight gene segments of the influenza A/Brevig Mission/1/1918 (H1N1) virus were determined by rapid amplification of cDNA ends (RACE). Chimeric viruses encoding the open reading frames of the 1918 virus but flanked by either the wild-type 1918 NCR sequences, or the NCR sequences of two other H1N1 virus strains, A/WSN/1933 and A/New York/312/2001 were produced. No growth differences between the NCR variant 1918 influenza viruses were noted.

Influenza A viruses are significant pathogens of humans and animals (1). In humans, influenza A viruses cause annual epidemics and occasional pandemics. The worst pandemic on record, the ‘Spanish’ influenza pandemic of 1918-1919 resulted in the deaths of approximately 50 million people (2). While attempts to characterize the causative agent of the pandemic were made in 1918, no successful viral isolates were made; consequently the virus was only characterized using an archaevirological approach by overlapping RT-PCR of small RNA fragments from 1918 post-mortem lung tissue samples (3). Using this approach, the complete 1918 viral genome coding sequences were determined, but the terminal non-coding region (NCR) sequences were not reported as consensus viral NCR primers were used in the amplification of the 5’ and 3’ ends of the coding regions of each segment (4-9). The complete coding sequence of the 1918 influenza A virus (1918 virus) hemagglutinin (HA) gene was determined from a formalin-fixed, paraffin-embedded 1918 post-mortem lung tissue sample [A/South Carolina/1/1918 (H1N1)] with partial HA sequence from several other cases (5, 10, 11). The coding sequences of the
remaining seven segments were determined from a frozen 1918 post-mortem lung tissue sample [A/Brevig Mission/1/1918 (H1N1)] (3, 9).

Once the complete coding sequences were determined, it became possible to re-construct the 1918 virus using reverse genetics in order to study its experimental pathogenesis (12). In order to produce infectious virus, however, complete vRNA segments were constructed by adding the NCR sequences from the mouse-adapted strain A/WSN/1933 (H1N1) (12). Studies performed using the 1918 virus sequences and infectious viruses containing one or more 1918 virus gene segments have yielded insights into influenza viral evolution, host adaptation and pathogenesis (13, 14).

The coding regions of each viral gene segment are terminated at both ends by non-coding regions (NCRs) (15). The biological significance of NCR sequence variation is not well understood. The NCRs have been previously demonstrated to be critical for viral polymerase binding, cap-snatching, transcription initiation, packaging, and replication (16-18). In this study, we determined the sequences of the 5’- and 3’-noncoding region (NCR) sequences of the influenza A/Brevig Mission/1/1918 (H1N1) virus along with the previously unsequenced portion of the hemagglutinin gene HA2 domain, thus generating the complete viral genome sequence of a 1918 pandemic influenza virus strain. 1918 viruses were constructed with different NCR sequences and evaluated for in vitro growth in MDCK cells.

1918 virus RNA was isolated from frozen post-mortem lung tissue (5), and the complete hemagglutinin gene HA2 domain sequence from the A/Brevig Mission/1/1918 was determined.
by overlapping RT-PCR (primers available upon request) using previously described methods (5). The HA2 domain sequence was identical to that of the previously reported A/South Carolina/1/1918 (H1N1) hemagglutinin complete coding sequence.

In order to determine the 16 NCR sequences of the 1918 influenza A virus (5'- and 3'-NCRs for each of the 8 viral gene segments), SMART™ RACE cDNA amplification (BD Biosciences, Palo Alto, CA) along with Poly(A) addition were employed to successfully amplify the full-length 1918 influenza A virus RNA NCRs (20). Briefly, the SMART™ RACE cDNA amplification system (BD Biosciences, Palo Alto, CA) provides a method for performing both 5'- and 3'-RACE. The cDNA for 5'-RACE is synthesized using a modified lock-docking oligo(dT) primer and the SMART II A oligo whose terminal stretch of dG residues can anneal to the dC-rich cDNA tails and serve as an extended template for Reverse Transcriptase (RT). For 1918 virus RNA, a Poly(A) tail was added first using E. coli poly(A) polymerase (Ambion, Inc., Austin, TX), and the Poly(A) tailed viral RNA was then used for synthesizing the RACE cDNA. RACE PCR reactions were performed using both 5'-end and 3'-end RACE CDS primers (Clontech, Mountain View, CA) and 1918 coding sequence gene specific primers (available upon request). At least four overlapping PCR reactions were performed for each NCR from each segment end. PCR was performed using Platinum® PCR SuperMix High Fidelity polymerase (Invitrogen, Carlsbad, CA) under the following conditions: initial denaturation at 94°C for 2 min, and then 40 cycles of 94°C/30s, 60°C/30s, and 72°C/1min. RACE PCR products were subcloned into the TOPO TA vector (Invitrogen, Carlsbad, CA), sequenced, and compared to other reference influenza viral sequences (19). The complete NCR sequences for the influenza
A/Brevig Mission/1/1918 (H1N1) virus are shown in Figure 1 aligned with NCR sequences from A/WSN/1933 (H1N1) and A/New York/312/2001 (H1N1) (20).

In the 3'-vRNA NCR (presented in sense orientation in Figure 1), variation has been reported at nucleotide position 4, and the 1918 virus encodes a cytidine at this position (guanidine in sense orientation) in the 3 polymerase gene segments and a uridine (adenine in sense orientation) in the remaining 5 gene segments. Additional quasispecies variations were noted in the first nucleotide position of five of the 1918 virus gene segments (PB1, PB2, PA, NP and NA), typically reported as a uridine (adenine in sense orientation). Using RACE, clones of the ends of these segments encoded uridine as expected, but clones encoding adenine and cytidine were also observed. Of the clones generated from the last overlapping RT-PCR RACE reaction, 43% reached to the first base of the 3’-end vRNA, the ratios of the first nucleotide base polymorphisms in the 1918 vRNA segments were 70% uridine, 14% adenine and 15% cytidines collectively (Table). Additionally, a cytidine, as opposed to the consensus uridine, was also identified at nucleotide position 3 in the 5’-vRNA NCR of the 1918 virus gene segment 7, which is located in the highly conserved terminal NCR sequence in influenza A viruses.

Many pathophysiologic studies have been performed with the reconstructed chimeric 1918 influenza A virus, containing the full-length coding genome of this extinct H1N1 influenza virus with the flanking NCRs of the mouse-adapted strain, A/WSN/1933 (H1N1). In order to test whether the newly determined sequence of the 1918 NCRs of A/Brevig Mission/1/1918 (H1N1) would alter the growth kinetics of the virus in vitro as compared to the 1918-WSN NCR chimeric viruses previously produced (12, 21), infectious 1918 virus variants were created by
reverse genetics with 1918 viral gene segments flanked by wild-type 1918 NCR sequences as determined in this study (1918 virus), the NCR sequences from A/WSN/1933, as used in prior experimental pathogenesis studies with the 1918 virus (1918-WSN33 NCR) (12) and as a control, the NCR sequences from a contemporary human strain, influenza A/New York/312/2001 (H1N1) (1918 NY312-NCR) (Figure 1). The variations in the NCRs (Figure 1) were introduced into reverse genetics plasmids using the QuikChange site-directed mutagenesis kit according to the manufacturer’s instructions (Stratagene, La Jolla, CA) (21). All clones were generated such that the first base of the 3’ NCR of each vRNA segment would be the consensus uridine (Figure 1). Fully reconstructed 1918 viruses with these NCR sequences were prepared using a standard reverse genetics system (12). Virus and infectious samples were handled under biosafety level 3 (BSL3) enhanced laboratory conditions in accordance with the select agent guidelines of the National Institutes of Health and the Centers for Disease Control and Prevention. Viruses were passaged 1 to 3 times on MDCK cells in the presence of 1μg/mL N-tosyl-L-phenylalanyl chloromethyl ketone (TPCK)-treated trypsin in Dulbecco’s Modification of Eagle’s Medium (DMEM). Infectious titer of virus stocks was determined by plaque assay performed in triplicate according to standard protocols.

Replication kinetics of the 1918 viruses were evaluated in MDCK cells (Figure 2). Triplicate sets of 80% confluent monolayers of MDCK cells were inoculated at a multiplicity of infection of 0.01, washed with sterile PBS, and overlaid with Opti-MEM (Invitrogen, Carlsbad, CA) supplemented with bovine serum albumin (0.2%) and 1μg/mL TPCK-treated trypsin. Cells were incubated at 37°C, and supernatants were collected at 12-hour intervals through 72 hours and every 24 hours thereafter until 120 hours. Infectious titers of supernatants were determined by
plaque assay. Student t tests were performed using the SAS 9.3 (SAS Institute, Cary, NC) to compare the difference at each time point, and graphical analysis utilized SigmaPlot software (Systat software, San Jose, CA). Error bars indicated the standard deviations. No significant differences were seen among the growth kinetics of the different 1918 viruses in vitro (Figure 2).

In order to test if the third position variation in the 5'-vRNA NCR of the 1918 virus segment 7 could affect viral replication, a 1918 virus construct that contained the consensus uridine at this position was also rescued along with other three NCR 1918 virus chimeras of 1918, but no significant changes for the replication kinetics were identified (data not shown). Unlike this variation, the polymorphism at the first nucleotide of 3'-vRNA NCRs in the 1918 virus segments (Figure 1) were independently identified in a recent deep-sequencing study of RNA from a 1918 influenza postmortem lung sample (22). Future studies will investigate possible replication changes with the identified first nucleotide variations.

In this study, the full wild-type genomic sequence of the prototype 1918 strain, A/Brevig Mission/1/1918 (H1N1) was completed, by determining the hemagglutinin HA2 domain sequence and performing RACE to determine the NCR sequences for each segment. Since previous studies with the 1918 virus used chimeras containing the A/WSN/1933 (H1N1) NCR sequences, we sought to evaluate whether this would affect growth kinetics of 1918 viruses in vitro. NCR sequence differences did not contribute to significant variation in growth, and thus the 1918-WSN NCR chimeric virus, previously used in pathogenicity studies, is likely representative of the wild-type 1918 virus.
Acknowledgements: This work was supported by the Intramural Research Program of the NIH and the NIAID.

References:


Table: 1918 NCR vRNA 3’ end first base sequence polymorphisms

<table>
<thead>
<tr>
<th>Influenza Gene Segment</th>
<th>No. RACE clones sequenced</th>
<th>No. (%) RACE clones reaching 3’ vRNA first base</th>
<th>vRNA 1st Base Sequence*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>PB2</td>
<td>16</td>
<td>5 (31%)</td>
<td>2</td>
</tr>
<tr>
<td>PB1</td>
<td>21</td>
<td>3 (14%)</td>
<td>1</td>
</tr>
<tr>
<td>PA</td>
<td>29</td>
<td>9 (31%)</td>
<td>3</td>
</tr>
<tr>
<td>HA</td>
<td>15</td>
<td>8 (53%)</td>
<td>8</td>
</tr>
<tr>
<td>NP</td>
<td>18</td>
<td>18 (100%)</td>
<td>10</td>
</tr>
<tr>
<td>NA</td>
<td>17</td>
<td>17 (100%)</td>
<td>17</td>
</tr>
<tr>
<td>MP</td>
<td>20</td>
<td>1 (5%)</td>
<td>1</td>
</tr>
<tr>
<td>NS</td>
<td>31</td>
<td>9 (29%)</td>
<td>8</td>
</tr>
</tbody>
</table>

First base vRNA sequence polymorphisms: 71/167 (43%) clones reached 1st base vRNA

71 (43%) total clones: 50/71 (70%), 11/71 (15%), 0 (0%), 10/71 (14%)

*Sequences presented in sense orientation
Figure Legends:

**Figure 1.** Noncoding region (NCR) sequences of influenza A/Brevig Mission/1/1918 (H1H1) [1918] aligned to NCRs from influenza A/WSN/1933 (H1N1) [WSN33], as used in the initial construction of the 1918 virus (12) and A/New York/312/2001 (H1N1) [NY312] (19). Sequences are presented in the sense orientation. Start and stop codons of the open reading frames are boxed. Sequence differences are marked with a dot. Sequence polymorphisms in the first base (3’ vRNA), as shown in sense orientation, presented with standard IUPAC mixed base code (R = A or G; D = A, G, or T; W = A or T).

**Figure 2.** 1918 influenza virus replication kinetics. Variant influenza A/Brevig Mission/1/1918 (H1H1) viruses were constructed containing the wild-type NCR sequences determined by RACE in this study (Figure 1) [1918] and growth kinetics was compared to chimeric 1918 viruses with the NCR sequences of A/WSN/1933 (H1N1) [1918 WSN33 NCR] and A/New York/312/2001 (H1N1) [1918 NY312 NCR].
Figure 1.

1918 PB2 5’-DGGCAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCTGAT...TAGGGTTGTTTCTACT-3’
WSN33 PB2 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCTGAT...TAGGGTTGTTTCTACT-3’
NY312 PB2 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCTGAT...TAGGGTTGTTTCTACT-3’

1918 PB1 5’-DGGCAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’
WSN33 PB1 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’
NY312 PB1 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’

1918 PA 5’-DGGCAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’
WSN33 PA 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’
NY312 PA 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’

1918 HA 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’
WSN33 HA 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’
NY312 HA 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’

1918 NP 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’
WSN33 NP 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’
NY312 NP 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’

1918 NA 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’
WSN33 NA 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’
NY312 NA 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’

1918 M 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’
WSN33 M 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’
NY312 M 5’-AGCGGAAAGCGGCAATTTCAATATATTCAATATG...TAGCTGGAAAGGCAATTTCAATATATTGCAAT...TAGGGTTGTTTCTACT-3’
1918 NS

5'-RGCAAAAGCAGGGTGACAAAAACATAATG..//..TAATAATAAAAAATACCCTTGTTTCTACT-3'

WSN33 NS

5'-AGCAAAAGCAGGGTGCAAAGACATAATG..//..TAATGATAAAAAACACCCTTGTTTCTACT-3'

NY312 NS

5'-AGCAAAAGCAGGGTGCAAAGACATAATG..//..TAATGATAAAAAACACCCTTGTTTCTACT-3'
Figure 2.

PFU/ml

Hours Post Inoculation

1918-WSN33 NCR
1918-NY312 NCR
1918