Complete nucleotide sequence of Newcastle disease virus: evidence for the existence of a new genus within the subfamily Paramyxovirinae

Olav de Leeuw and Ben Peeters

Institute for Animal Science and Health (ID-DLO), Department of Avian Virology, PO Box 65, 8200 AB Lelystad, The Netherlands

We have completely sequenced the genome of Newcastle disease virus (NDV) vaccine strain LaSota. The sequences of the 3′- and 5′-terminal ends of the RNA genome were determined by sequencing cDNA fragments generated by rapid amplification of cDNA ends. The entire genomic sequence, which was established by sequencing cDNA fragments generated by high-fidelity RT–PCR, consists of 15 186 nt. Comparison of the 5′-terminal sequence of NDV LaSota with the 5′-terminal sequences of ten members of the Paramyxovirinae showed that NDV LaSota has an unusually long 5′ untranscribed region. Comparison of the entire genomic sequences showed that NDV is only distantly related to the other members of the genus Rubulavirus, to which NDV has been assigned. In this paper we present data which suggest that NDV should not be classified in the genus Rubulavirus, but instead should be considered as a member of a new genus within the subfamily Paramyxovirinae.

Newcastle disease is a serious disease of poultry that can cause severe economic losses in many countries. The causative agent of the disease is Newcastle disease virus (NDV), also called avian paramyxovirus type-1. NDV is a member of the genus Rubulavirus of the subfamily Paramyxovirinae (family Paramyxoviridae, order Mononegavirales) (Rima et al., 1995). NDV strains are classified on the basis of their pathogenicity for chickens into highly pathogenic (velogenic), intermediate (mesogenic) and apathogenic (lentogenic) strains (Hanson, 1988). Lentogenic strains are used as live vaccines to protect poultry against Newcastle disease.

NDV contains a non-segmented, single-stranded RNA genome of negative polarity. The genome length has been predicted to be 15 156 nt (Millar & Emmerson, 1988). The genome comprises six genes, which encode the nucleocapsid protein (NP), phosphoprotein (P), matrix protein (M), fusion protein (F), haemagglutinin–neuraminidase (HN) and large polymerase protein (L). In addition to these gene products, additional proteins (designated V and W protein) may be produced by an RNA-editing event that occurs during transcription of the P gene (Steward et al., 1993). NDV lacks the gene encoding the small hydrophobic (SH) protein, which is present in some members of the Rubulavirus genus.

Although the nucleotide sequence of all NDV genes is known, these sequences have been obtained from different NDV strains. Furthermore, the exact 5′-terminal sequence has never been unambiguously determined. Thus, in contrast to other members of the Paramyxovirinae, the entire genomic sequence of one and the same NDV strain has never been obtained. The purpose of this study was to establish the complete sequence of one and the same NDV strain and to compare it with the genomic sequences of other members of the Paramyxovirinae.

We sequenced the genome of NDV strain LaSota (ATCC VR-699), a lentogenic strain that is used worldwide as a live vaccine (Goldhaft, 1980). In order to avoid sequence variation, the virus was plaque-purified three times on primary chicken embryo fibroblasts (Harper, 1989) and multiplied in embryonated eggs. The plaque-purified virus had the same growth properties and showed the same intracerebral pathogenicity index in 1-day-old chickens as the original uncloned virus (data not shown).

The sequences of the 3′- and 5′-terminal ends of the viral RNA were determined by rapid amplification of cDNA ends (RACE). To determine the sequence of the 5′ end, primer p360 (5′ GGCGATGTAATCAGCCTAGTGCTT 3′), which was derived from the published sequence of the L-gene of NDV strain Beaudette C (Yusoff et al., 1987), was used to generate single-stranded cDNA with reverse transcriptase (RT) as described by the supplier (Life Technologies). The single-
stranded cDNA (2.5 µl of the RT mixture) was ligated to 8 pmol anchor-primer ALG3 (5’ CACGAATTCCACTATCGATTCTGATCCCTTTCGATAGTCAGTATTTGTGTATTTTGGTTGTTTGGTTGTT 3’) with T4 RNA ligase as described by Tessier et al. (1986). One µl of the ligation mixture was used in a PCR with Taq polymerase (Perkin Elmer) and primers ALG4 (5’ CACGAATTCCACTATCGATTCTGATCCCTTTCGATAGTCAGTATTTGTGTATTTTGGTTGTTTGGTTGTTTGGTTGTT 3’), which is complementary to ALG3, and p375 (5’ ATTATGACCTATCTCATTTCGGCGCCCATTGTTTGGTTGTTTGGTTGTT 3’). The PCR consisted of 40 cycles of 1 min at 94 °C, 1 min at 55 °C and 2 min at 72 °C. The PCR products were treated with Klenow DNA polymerase I to create blunt ends and cloned in the Hincll site of plasmid pGEM-4Z (Promega). Five independent clones were sequenced. To determine the sequence of the 3’-terminal end, primer ALG3 was ligated to the 3’-end of genomic NDV LaSota RNA with T4 RNA ligase as described by Schütze et al. (1995). The mixture was incubated overnight at room temperature and 5 µl of the ligation reaction was used as template in a reverse transcription reaction with ALG4 as primer. One µl of the RT reaction was used in a PCR reaction with primers ALG4 and p376 (5’ GAGCCTTAAGGAGCTGCTCGTACTGATC 3’). The latter primer was derived from the published sequence of the 3’-end of NDV strain D26 (Ishida et al., 1986). The RT–PCR and cloning conditions were as described above for the 5’ RACE. Four independent clones were sequenced.

To clone and sequence the entire NDV LaSota genome, three similar sets (A, B and C) of five overlapping subgenomic cDNA fragments (approximately 4 to 7 kbp in size) were generated by means of RT–PCR. To keep the number of PCR errors to a minimum, a proofreading enzyme mixture was used.

---

**Fig. 1.** Sequence comparison of NDV LaSota and four members of the Rubulavirus genus, three members of the Paramyxovirus genus and three members of the Morbillivirus genus. The sequences are presented as cDNA in the 5’–3’ direction (antigenomic strand). Conserved nucleotides are indicated by boxes from black (highly conserved) to grey (less conserved). The consensus is printed above the sequences. (A) Sequences of the first 65 nt of the 3’-end. The last 10 nt represent the sequence of the start box of the NP-gene mRNA (underlined). (B) Sequences from the region between the L-gene transcription end box to the 5’-terminal end.
Table 1. Properties of NDV LaSota in comparison to other members of the three genera of the Paramyxovirinae

<table>
<thead>
<tr>
<th>Genus and viruses*</th>
<th>Genome length (nt)†</th>
<th>Rule of six (length/6)</th>
<th>5′ UTR‡</th>
<th>Protein switch after P-gene editing</th>
<th>Subunit hexamer phasing position at P-gene editing site§</th>
<th>Subunit hexamer phasing position at the gene start site¶</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDV</td>
<td>15 186</td>
<td>2531</td>
<td>114</td>
<td>P → V</td>
<td>1</td>
<td>2 4 4 4 4 — 3 6</td>
</tr>
<tr>
<td>Rubulavirus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hPIV2</td>
<td>15 046</td>
<td>2607 + 4</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MuV</td>
<td>15 384</td>
<td>2564</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SV5</td>
<td>15 240</td>
<td>2541</td>
<td>32</td>
<td>V → P</td>
<td>3</td>
<td>2 1 1 1 — 1 6</td>
</tr>
<tr>
<td>SV41</td>
<td>15 450</td>
<td>2575</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parvovirus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bPIV3</td>
<td>15 480</td>
<td>2580</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hPIV3</td>
<td>15 462</td>
<td>2577</td>
<td>45</td>
<td>P → V</td>
<td>2</td>
<td>2 1 1 1 — 1 2</td>
</tr>
<tr>
<td>SeV</td>
<td>15 384</td>
<td>2564</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morbillivirus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDV</td>
<td>15 016</td>
<td>2602 + 4</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MeV</td>
<td>15 894</td>
<td>2649</td>
<td>41</td>
<td>P → V</td>
<td>6</td>
<td>2 2 4 3 — 3 2</td>
</tr>
<tr>
<td>RPV</td>
<td>15 882</td>
<td>2647</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* NDV, Newcastle disease virus strain LaSota; hPIV2, human parainfluenza virus 2; MuV, mumps virus; SV5 and SV41, simian virus 5 and 41, respectively; SeV, Sendai virus; bPIV3 and hPIV3, bovine and human parainfluenza virus, respectively; CDV, canine distemper virus; MeV, measles virus; RPV, rinderpest virus.
† Nucleotide sequences of the entire genomes were obtained as follows (accession no.): NDV LaSota (AF077761); hPIV2 (X57559); MuV (AB000388); SV5 (AF052755); SV41 (X64275); bPIV3 (D84095); hPIV3 (Z11575); CDV (L13194); MeV (X16565); RPV (Z30697).
‡ 5′ UTR: size (nt) of untranscribed region between the transcriptional end box of the L-gene and the 5′-terminal end.
§ Position 2285 of the NDV LaSota genome (AAAAA). The positions of the other viruses are described by Kolakofsky et al. (1998). Subunit hexamer phasing positions of hPIV2 and CDV were not determined because the genome length does not conform to the ‘rule-of-six’.
¶ NP, nucleoprotein; P, phosphoprotein; M, matrix protein; F, fusion protein; SH, small hydrophobic protein; HN, haemagglutinin–neuraminidase; L, large polymerase protein.

and the number of PCR cycles was limited to 15. First strand cDNA synthesis was performed with primer 3UT (5′ ACCAAAAAGAGATCCGTTGAG 3′) and 1 μl of the RT mixture (three independent reactions were performed) was used for PCR with the Expand Long Template PCR kit (Boehringer Mannheim). The PCR consisted of five cycles of 10 s at 94 °C, 30 s at 58 °C, and 6 min at 68 °C, followed by ten cycles of 10 s at 94 °C, 30 s at 58 °C, and 6 min at 68 °C, in which the elongation time at 68 °C was increased by 20 s per cycle. The cDNAs were cloned in pGEM-T (Promega). The complete nucleotide sequence of the cDNA clones of sets A and B was determined with primers flanking the inserts and with NDV-specific primers which were either derived from published NDV sequences or from NDV LaSota sequences which were established during this sequence project. After comparison of the complete nucleotide sequence of the A and B sets of clones, remaining ambiguities were resolved by sequencing the relevant regions of corresponding cDNAs of set C. All sequences were determined by using the PRISM Kit (Perkin Elmer) and an Applied Biosystems automatic sequencer (ABI 310).

The sequences of the 3′- and 5′-terminal ends of NDV LaSota were compared with the published 3′-terminal sequence of NDV strain D26 (Ishida et al., 1986) and the putative 5′-terminal sequence of NDV strain Beaudette C (Yusoff et al., 1987), respectively. The sequence of the 3′ end was almost identical to the sequence of the 3′ end of strain D26 (data not shown). However, the sequence of the 5′ end was found to contain an extra 64 nt compared to the putative 5′ sequence of strain Beaudette C. The entire genome of NDV LaSota consists of 15 186 nt, which is the smallest genome size of all Paramyxovirinae for which the entire genomic sequence has been established to date. The genome organization of NDV is described by Millar & Emmerson (1988).

The complete genomic sequence of NDV LaSota was compared with the genomic sequences of ten viruses from three different genera within the subfamily Paramyxovirinae for which the complete sequence is available. The comparison showed that the first 4 nt (ACCA) of the 3′-terminal end are identical in all viruses (Fig. 1A). However, nt 6–8 (ACA) of NDV LaSota are identical to nt 6–8 of the morbilliviruses and paramyxoviruses but differ from the conserved GGG of the rubulaviruses. Comparison of the sequence located between the transcription end box of the L gene and the 5′-terminal end
showed that in NDV LaSota this region is much larger (114 nt) than the same region in the other viruses, especially the rubulaviruses (Fig. 1B and Table 1). The 3’ and 5’ ends of NDV LaSota are partially complementary at the first 12 nt which seems to be a common feature of the Paramyxovirinae. The entire genomic RNA sequences as well as the protein sequences of NDV LaSota and ten members of the Paramyxovirinae were aligned using the program Clustal V (Higgins et al., 1991). The alignments were further analysed with the phylogeny program PHYLIP (Felsenstein, 1989). Fig. 2(A) shows the phylogenetic tree which was obtained after aligning the entire genomic RNA sequences of the viruses. The tree shows that NDV is only distantly related to the rubulaviruses. In fact, NDV is almost as distantly related to the rubulaviruses as the morbilliviruses are to the paramyxoviruses. This can also be seen in Fig. 2(B), which shows the percentage similarity and divergence between the viruses. This analysis shows that the members of the genera Rubulavirus, Paramyxovirus and Morbillivirus represent distinct groups (boxed in Fig. 2B). However, according to this analysis, NDV LaSota cannot be assigned to any of these genera. Similar results were obtained after aligning the protein sequences. According to these analyses, the NP, P and HN proteins were somewhat more related to their homologous rubulavirus proteins than the M, F and L proteins (data not shown).

That NDV is only distantly related to other members of the Rubulavirus genus is confirmed by several differences in other biological properties, which support our view that NDV does
not belong to this genus. First, NDV edits its P gene mRNA from P to V instead of V to P (Steward et al., 1993) (Table 1). In accordance with this finding is the observation that the nucleotide sequence surrounding the mRNA editing site is similar in NDV LaSota and the paramyxoviruses and morbilliviruses (TTTTAAGGGG) but differs from that of the rubulaviruses (TTTAAAGGGG). Second, Calain & Roux (1993) have suggested that paramyxovirus genomes are replicated efficiently only when they are a multiple of 6 nt in length. This has been dubbed the ‘rule-of-six’. The genome length of NDV LaSota (15,186 nt) is a multiple of 6 nt and seems to conform to this rule (Table 1). It has been shown that replication of simian virus 5 (SV5) is enhanced by, but not strictly dependent on, this rule (Murphy & Parks, 1997). Our preliminary data indicate that, similar to Sendai virus (SeV) but unlike SV5, replication of NDV LaSota minigenomes is strictly dependent on the rule-of-six (B. Peeters and others, unpublished data). Third, Kolakofsky et al. (1998) recently described the subunit hexamer phasing positions of the nucleotide start site for the initiation of the mRNA synthesis for several members of the Paramyxovirinae. These positions can be determined because each nucleocapsid protein is predicted to be associated with precisely 6 nucleotides (hexamer) (Engelman et al., 1989). Until now these hexamer phasing positions could not be investigated for NDV because the entire genome of a single strain of NDV was not known. The mRNA start sites for NDV LaSota are clustered over four subunit hexamer phasing positions (6, 2, 3, 4), whereas the mRNA start sites of the members of the Rubulavirus genus are clustered over only three positions (6, 1, 2). In addition, the position of the subunit hexamer phasing at the P gene mRNA editing site of NDV LaSota is 1, whereas its position for the members of the Rubulavirus genus is 3 (Table 1). Fourth, NDV and human parainfluenza virus type 2 (hPIV2) are not immunologically related. Tsurudome et al. (1989) used a panel of 128 monoclonal antibodies (MAbs) directed against the NP, P, M, F and HN proteins of the rubulavirus hPIV2, and showed that none of the MAbs reacted with NDV or the paramyxoviruses hPIV1 and hPIV3. However, some MAbs did react with SV5, SV41 and mumps virus, which are all members of the Rubulavirus genus. Finally, like SeV (Paramyxovirus genus) and unlike hPIV2 (Rubulavirus genus), the NP protein of NDV can bind to RNA in vivo to form nucleocapsid-like structures in the absence of the phosphoprotein (Errington & Emmerson, 1997).

NDV seems to have evolved separately from other rubulaviruses, probably as a result of an early switch of host. NDV is the only member of the Paramyxovirinae whose host is a bird and not a mammal. Therefore, we suggest that NDV (and probably all avian paramyxoviruses) should be assigned to a new genus. Yu et al. (1998) have recently found similar results for Hendra virus (equine morbillivirus), which also does not seem to fit into any of the genera of the subfamily Paramyxovirinae. This again shows that some members of the Paramyxovirdae cannot be assigned to the existing genera.

References


Received 3 August 1998; Accepted 8 September 1998