*Journal of General Virology* (2001), **82**, 379–383. Printed in Great Britain SHORT COMMUNICATION

# Genetic analysis of full-length genomes and subgenomic sequences of TT virus-like mini virus human isolates

Philippe Biagini,[1] Pierre Gallian,[1] Houssam Attoui,[1] Mhammed Touinssi,[1] Jean-François Cantaloube,[1] Philippe de Micco[1,2] and Xavier de Lamballerie[2]

[1] Unité des Virus Emergents, EA 871, Laboratoire de Virologie Moléculaire, Établissement Français du Sang 'Alpes-Méditerranée', 149 Boulevard Baille, 13005 Marseille, France

[2] Unité des Virus Emergents, EA 871, Laboratoire de Virologie Moléculaire, Tropicale et Transfusionnelle, Faculté de Médecine, 27 Boulevard Jean Moulin, 13005 Marseille, France

**The phylogenetic relationship between the complete genomic sequences of ten Japanese and one French isolate of TT virus-like mini virus (TLMV) was investigated. Analysis of the variability of the nucleotide sequences and the detection of signature patterns for overlapping genes suggested that ORFs 1 and 2 are probably functional. However, this was not the case for a putative third ORF, ORF3. Throughout the viral genome, several nucleotide or amino acid motifs that are conserved in circoviruses such as TT virus (TTV) and chicken anaemia virus were identified. Phylogenetic analysis distinguished three main groups of TLMV and allowed the identification of putative recombination breakpoints in the untranslated region. TLMV genomes were detected by PCR in the plasma of 38/50 French blood donors tested and were also identified in peripheral blood mononuclear cells, faeces and saliva. A phylogenetic study of 37 TLMV strains originating from France, Japan and Brazil showed that groupings were not related to geographical origin.**

The genome of TT virus-like mini virus (TLMV), a newly recognized virus, was identified in the sera of three Japanese blood donors by Takahashi *et al.* (2000). This genome was shown to be a single-stranded DNA molecule of negative polarity, approximately 2·9 kb long, with three putative overlapping ORFs. The size of the virion was estimated, by filtration studies, to be less than 30 nm with an isopycnic density of 1·31–1·34 g/ml in CsCl (Takahashi *et al.*, 2000). These properties tend to affiliate TLMV with the family *Circoviridae*, and in particular, to the recently described TT virus (TTV) (Nishizawa *et al.*, 1997) and chicken anaemia virus (CAV) (Yuasa *et al.*, 1979).

**Author for correspondence:** Xavier de Lamballerie.
Fax +33 491 32 44 95. e-mail virophdm@gulliver.fr

The GenBank accession numbers of the sequences reported in this paper are AF291073–AF291091.

To date, the complete sequence of ten Japanese TLMV isolates has been determined by Takahashi *et al.* (2000) and deposited in databases. However, there have been no reports of complete sequence data from non-Japanese isolates, phylogenetic analysis of genomes or epidemiological investigations.

In this study, we report the first complete sequence of a European isolate (PB4TL) of TLMV. Viral DNA was extracted from the peripheral blood mononuclear cells (PBMCs) of a French haemodialysis patient using a commercial kit (High Pure viral nucleic acid kit, Roche), according to the manufacturer's instructions. Primers specific for the non-coding region (NCR) of the viral genome were designed according to alignment analysis of the available full-length Japanese TLMV sequences. These NCR-specific primers (TLMS, sense 5′ ATTWRMATTGCCGACCACAAAC 3′, and TLMS2INV, antisense 5′ GTTTCTTGCCCRKTCCGCYAG 3′) amplified a 288 nt product using standard PCR conditions with an annealing temperature of 55 °C. Determination of the sequence of this 288 nt PCR product allowed the design of two new sets of specific primers that could be used for two successive steps of 'inverted' PCR around the circular genome of TLMV (Fig. 1*a*). Primers used for first-round inverted PCR were SF1 (sense 5′ TGGCTGAGTTTATGCCGCTAGACG 3′) and RF1 (antisense 5′ TCCCCGCCTAGTTATGACGGTGTG 3′). For the nested inverted PCR, primers SF2 (sense 5′ GAAGACGGACAACGACTTCGGCTG 3′) and RF2 (antisense 5′ CATGAATAATAATGAGCAAAAGAG 3′) were used.

Because there was a GC-rich sequence localized in the NCR of the viral genome, amplification and sequencing were performed using the LA PCR kit (Takara Shuzo) as previously reported for TTV (Biagini *et al.*, 2000*a*). Direct sequencing of PCR products allowed the determination of the complete sequence of TLMV isolate PB4TL (GenBank accession number AF291073).

Sequence analysis revealed that the genome of the PB4TL isolate was 2910 nt in length and contained two large ORFs,

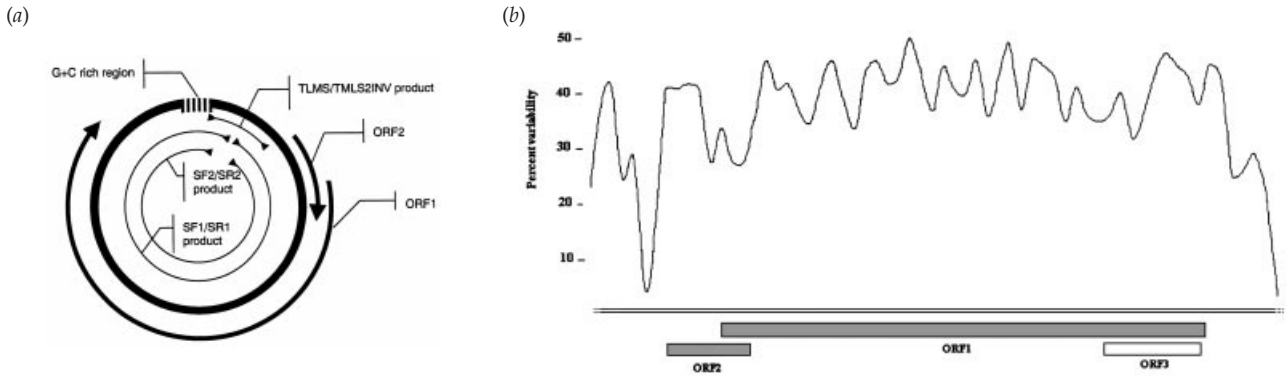


Fig. 1. (*a*) Genetic organization of TLMV. (*b*) Percentage variability plotted across the alignment of 11 full-length TLMV nucleotide sequences. Variability was determined using the software program MEGA and a sliding window of analysis of 50 positions. The relative positions of the ORFs are shown.

corresponding to ORF1 (nt 529–2556; 676 aa) and ORF2 (nt 341–640; 99 aa), similar to those present in the genomes of the Japanese isolates described by Takahashi *et al*. (2000). In contrast, no initiation codon was found for the proposed ORF3 reported by the same authors. It is noteworthy that this is also the case for the Japanese isolate CLC-156. To further investigate the probable functionality of the different ORFs, full-length TLMV sequences were aligned using the CLUSTAL W program, version 1.74 (Thompson *et al*., 1994). As expected, it was observed that the two first codon positions were less variable than the third position along the sequences of ORFs 1 and 2. In the region where the two genes overlap, the coding constraint applies to all codon positions, resulting in a lower global variability (Fig. 1*b*). In this region, we also detected signature sequences of overlapping genes (Pavesi, 2000): the end of the ORF2 sequence includes a large number of acidic amino acid residues, while the beginning of ORF1 includes a large number of basic amino acid residues. This situation is similar to that observed in the genome of another circovirus, CAV, in the region where the genes of the 24 and 52 kDa proteins overlap (Pavesi, 2000). Altogether, these observations indicate that both ORFs 1 and 2 of TLMV are functional. In contrast, we did not identify a significant drop in nucleotide variability at the end of ORF1, which is the region corresponding to the putative ORF3. Moreover, no signature sequence of overlapping genes could be detected in that part of the genome. Together with the absence of an initiation codon in that region for at least two TLMV isolates, these data suggest that the probability of ORF3 being functional is low.

In the NCR of the viral genomes analysed, the alignment of full-length sequences allowed the identification of two highly conserved patterns located between nucleotides 183 and 332 (positions are related to strain CBD279), downstream of the GC-rich zone. Interestingly, the nucleotide patterns spanning positions 183–198 and 247–298 correspond to sequences conserved among the most highly divergent TTV isolates. Another conserved pattern, common to both TTV and TLMV, has been previously identified in the NCR upstream of the GC-rich zone by Takahashi *et al*. (2000). Other circovirus-related motifs are present in the genome of TLMV: in particular, a CAV-like VP2 amino acid motif ($Wx_7Hx_3Cx_1Cx_5H$) is present in the TLMV ORF2, a motif which is also observed in TTV sequences (Hijikata *et al*., 1999; Biagini *et al*., 2000*b*; Takahashi *et al*., 2000). Finally, since circoviruses replicate using a rolling circle mechanism, we investigated the presence of motifs related to the Rep protein, which possesses up to four sequence motifs and is conserved among many plant and animal circoviruses (Niagro *et al*., 1998). Conserved motifs 1 (FTL/FxTL), 2 (HxH) and 3 (YxxK) were identified in ORF1 of numerous TLMV isolates, while motif 4 (GxxxxGKS) and the P-loop (putative ATP/GTP-binding motif) could not be found, as previously reported in the case of TTV and CAV.

Pairwise comparison of complete nucleotide sequences was performed using the software program MEGA (Kumar *et al*., 1993) to assess the genetic relationship between the 11 TLMV isolates tested. The p-distance algorithm for distance determination, the neighbour-joining method for tree-drawing and a bootstrap resampling of 500 replications were used. An unrooted phylogenetic tree was constructed in which three main groups, supported by a 100% bootstrap confidence level, could be distinguished (Fig. 2*a*). Group 1 includes isolates PB4TL, NLC030, CLC062, CLC156 and CBD203. Group 2 includes isolates CBD231, CBD279, CLC138 and CLC205 and group 3 includes isolates NLC023 and NLC026. The topology of this phylogram was compared with those observed using independent alignments of subgenomic sequences. Similar groupings were found for ORFs 1 and 2. A tree corresponding to the ORF1 region is shown in Fig. 2(*b*). Phylogenetic analysis of amino acid sequences corresponding to ORFs 1 and 2 provided similar groupings to those identified by nucleotide sequence analysis (data not shown). The untranslated region of the genome was also submitted to phylogenetic analysis. The region of the NCR located upstream of the GC-rich zone produced phylogenetic groupings similar to those obtained from complete or coding sequences. This was not the case for

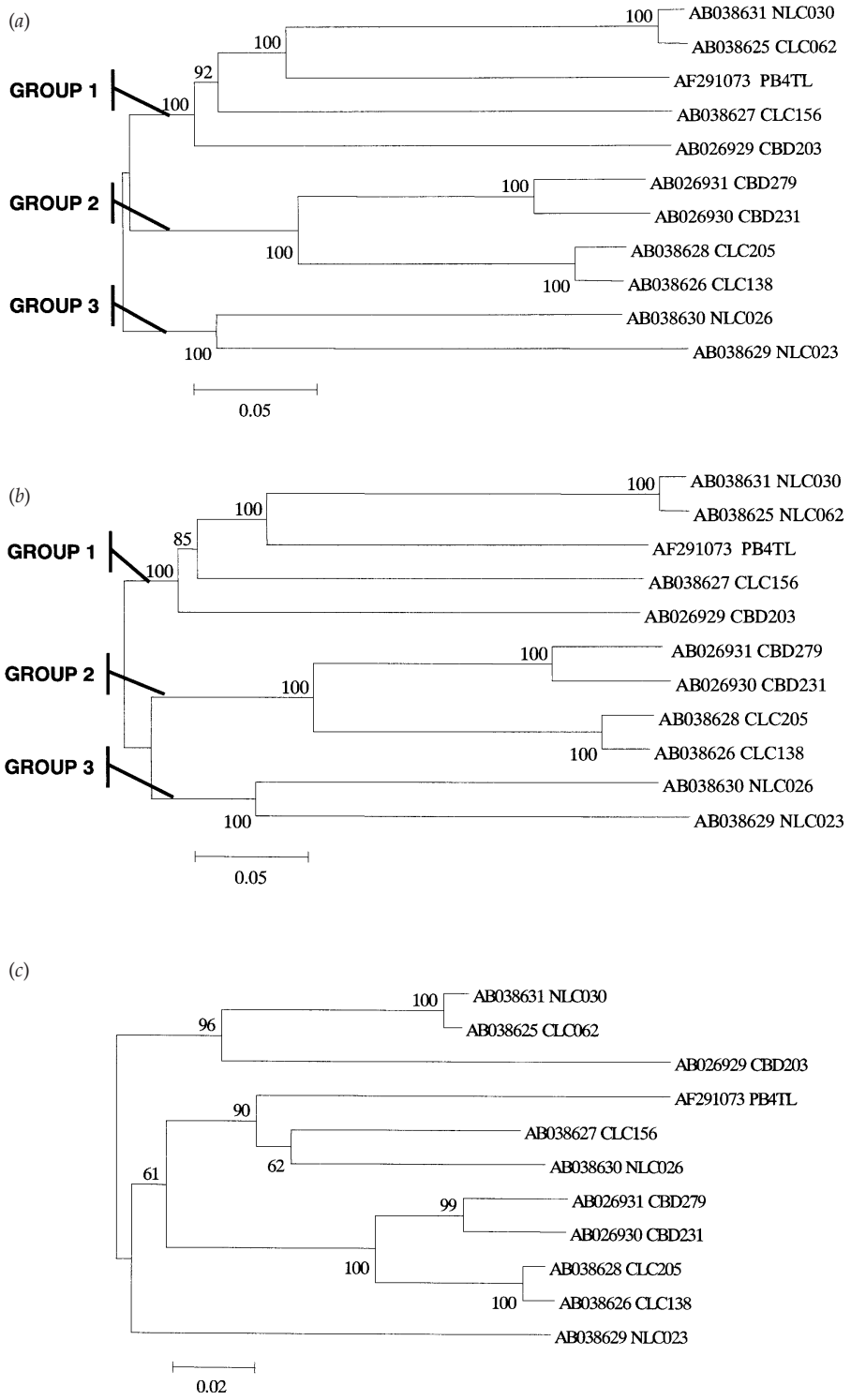


Fig. 2. Phylogenetic analysis of 11 complete genome sequences (*a*), ORF1 sequences (*b*) and TLMV nucleotide sequences in the untranslated region, downstream from the GC-rich zone (*c*). The name of the strains and the GenBank accession numbers are given. Numbers at the forks indicate the bootstrap confidence level supporting the observed phylogenetic topology.

the region of the NCR located downstream of the GC-rich zone. The topology of group 2 was identical to that observed when using complete sequences, but the positions of some isolates (PB4TL, NLC026 and CLC156) belonging to groups 1 and 3 within the tree showed important variations (Fig. 2*c*). One of the hypotheses that could take this observation into

**Table 1.** Pairwise comparisons of full-length nucleotide and ORF amino acid sequence divergence

| Genetic group | ORF | Sequence divergence (%) | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| Nucleotide | | | | |
| 1 | – | 2·3–39·0 | 41·1–44·2 | 42·3–46·9 |
| 2 | – | – | 4·0–28·2 | 40·1–44·0 |
| 3 | – | – | – | 35·7 |
| Amino acid | | | | |
| 1 | ORF1 | 4·7–62·2 | 65·2–68·3 | 64·4–68·5 |
| | ORF2 | 2·0–56·8 | 56·8–67·9 | 57·1–70·6 |
| | ORF3 | 1·7–60·9 | 62·1–70·9 | 59·8–62·0 |
| 2 | ORF1 | – | 4·8–44·7 | 59·5–65·0 |
| | ORF2 | – | 3·3–40·7 | 48·9–68·6 |
| | ORF3 | – | 3·1–50·0 | 58·5–64·8 |
| 3 | ORF1 | – | – | 54·1 |
| | ORF2 | – | – | 56·2 |
| | ORF3 | – | – | 52·8 |

account is the existence of recombination events between TLMV isolates. The location of such recombinations within the NCR could be favoured by the existence of several highly conserved nucleotide patterns. This hypothesis was tested by submitting the alignments of full-length sequences to analysis by the Recombination Detection Program (RDP) (Martin & Rybicki, 2000). Despite the low number of sequences available, RDP determined a high probability for recombination breakpoints, all of them located within the NCR. These findings are in accordance with recent results concerning TTV (Worobey, 2000). In particular, TTV isolate NLC026 was identified as a possible recombinant ($P = 2 \times 10^{-8}$) between a 'minor' parent belonging to the group 1 PB4TL lineage (region spanning nt 2852–330) and a 'major' parent belonging to the group 3 NLC023 lineage. This would explain the phylogenetic assignment of NLC026 to group 3 when complete or coding sequences are analysed and its apparent genetic relatedness to the PB4TL isolate when non-coding sequences are analysed. Similarly, within group 1, isolate CLC156 was identified as a possible recombinant between one parent belonging to the PB4TL/NLC026 lineage for the NCR (nt 14–213) and another parent belonging to the CLC062 lineage for the rest of the genome. The hypothesis that some of the complete sequences used for analysis are recombinant sequences derived from different viruses present in the original samples cannot be formally ruled out. In the case of isolate PB4TL, however, the fact that identical sequences were obtained using specific or degenerated/direct or reverse PCR systems renders this hypothesis improbable.

Very little is known about the natural history and molecular epidemiology of TLMV. This study provides evidence of the genetic variability of TLMV, which allows the delineation of three groups. Pairwise comparisons of sequence divergence between complete nucleotide sequences and those related to the putative ORF amino acid sequences are reported in Table 1. Using full-length nucleotide sequences, a cut-off value of 40% matches our phylogenetic findings, i.e. the delineation of groups 1, 2 and 3. Data analysis revealed a high degree of genetic diversity between TLMV isolates, with divergence values reaching up to 46·9% for full-length nucleotide sequences and up to 68·5, 70·6 and 70·9% for amino acid sequences of ORFs 1, 2 and 3, respectively. This high degree of genetic variability might be an obstacle to the molecular diagnosis of TLMV infection. However, the presence of conserved patterns in the NCR of the genome allows the design of PCR primers that can be used for the amplification of the most divergent isolates identified to date. We used such conserved primers for a study involving samples from 50 French blood donors. DNA that was extracted from plasma samples was amplified by PCR using the primers TLMS and TLMS2INV for first-round PCR and the reverse primer TLMRC (5′ CGGTGGTTTCACTCACCTTCG 3′) for semi-nested PCR amplification. TLMV DNA was identified in the plasma of 76% of the individuals tested, corresponding to the recently determined prevalence of TTV in French blood donors (Biagini *et al.*, 2000*b*). These data imply the existence of chronic forms of infection. TLMV DNA could be detected in various body fluids including not only plasma and PBMCs but also faeces and saliva. As previously proposed for TTV (Gallian *et al.*, 2000), an oral spread by saliva droplet is the probable route of transmission for TLMV; this would explain the high prevalence of TLMV in the general population of an industrialized country.

To investigate the possible relationship between the phylogenetic position of virus strains and their geographical origin, we analysed 37 partial NCR sequences (nt 66–271). These sequences corresponded to the ten completely sequenced Japanese isolates, PB4TL, eight Brazilian isolates (sequences retrieved from databases) and 18 additional sequences from French isolates (this study, GenBank accession numbers AF291074–AF291091). Phylogenetic analysis of these sequences identified three main groups (data not shown). The distribution of these different isolates was not related to their geographical origin, since each group included Brazilian, French and Japanese isolates, suggesting a long history of TLMV infection in human populations.

Finally, our analysis of TLMV sequences confirms that the general organization of the genome and the presence of conserved nucleotide or amino acid motifs show a clear relatedness with other circoviruses such as TTV or CAV. However, these virus species are only distantly related to each other as shown by genetic distances ($> 75\%$) or by analysis of their G+C content, which revealed a high degree of discrepancy (the calculated values from full-length sequences are 37·7, 51·7 and 56·4% for TLMV, TTV and CAV, respectively). Preliminary observations suggest that, as pre-

viously observed for TTV, the infection of humans by TLMV is extremely frequent and generally non-symptomatic. This reinforces the hypothesis of Simmonds *et al*. (1999) that circoviruses could be a part of the normal human flora and constitutes a new and important physiological concept, since the active replication of viruses that chronically infect humans (such as *Herpesviridae*) is generally considered to be associated with pathology. The presence of TTV or TLMV virus particles in the blood of a large proportion of healthy individuals demonstrates that viraemia (and thus active virus replication) is not always associated with disease. However, the example of the saprophytic bacterial flora of humans reminds us that this kind of co-habitation is the result of a subtle balance between the multiplication of micro-organisms and the host defences. Consequently, the high prevalence of TTV or TLMV infections should not be interpreted as proof that these viruses are never implicated in human pathologies.

The authors wish to emphasize the generosity of the scientists who made this work possible: Dr Takahashi and colleagues who deposited numerous TLMV sequences in databases prior to publication and Dr Martin who created and made available the RDP program. This work was supported by a grant from the Établissement Français du Sang (EFS).

## References

**Biagini, P., Attoui, H., Gallian, P., Touinssi, M., Cantaloube, J. F., de Micco, P. & de Lamballerie, X. (2000*a*).** Complete sequences of two highly divergent European isolates of TT virus. *Biochemical and Biophysical Research Communications* **271**, 837–841.

**Biagini, P., Gallian, P., Touinssi, M., Cantaloube, J. F., Zapitelli, J., de Lamballerie, X. & de Micco, P. (2000*b*).** High prevalence of TT virus infection in French blood donors revealed by the use of three PCR systems. *Transfusion* **40**, 590–595.

**Gallian, P., Biagini, P., Zhong, S., Touinssi, M., Yeo, W., Cantaloube, J. F., Attoui, H., de Micco, P., Johnson, P. J. & de Lamballerie, X. (2000).** TT virus: a study of molecular epidemiology and transmission of genotypes 1, 2 and 3. *Journal of Clinical Virology* **17**, 43–49.

**Hijikata, M., Takahashi, K. & Mishiro, S. (1999).** Complete circular DNA genome of a TT virus variant (isolate name SANBAN) and 44 partial ORF2 sequences implicating a great degree of diversity beyond genotypes. *Virology* **260**, 17–22.

**Kumar, S., Tamura, K. & Nei, M. (1993).** MEGA: Molecular evolutionary genetics analysis, version 1.02. Pensylvania State University, PA, USA.

**Martin, D. P. & Rybicki, E. P. (2000).** RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562–563.

**Miyata, H., Tsunoda, H., Kazi, A., Yamada, A., Khan, M. A., Murakami, J., Kamahora, T., Shiraki, K. & Hino, S. (1999).** Identification of a novel GC-rich 113-nucleotide region to complete the circular, single-stranded DNA genome of TT virus, the first human circovirus. *Journal of Virology* **73**, 3582–3586.

**Niagro, F. D., Forsthoefel, A. N., Lawther, R. P., Kamalanathan, L., Ritchie, B. W., Latimer, K. S. & Lukert, P. D. (1998).** Beak and feather disease virus and porcine circovirus genomes: intermediates between the geminiviruses and plant circoviruses. *Archives of Virology* **143**, 1723–1744.

**Nishizawa, T., Okamoto, H., Konishi, K., Yoshizawa, H., Miyakawa, Y. & Mayumi, M. (1997).** A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochemical and Biophysical Research Communications* **241**, 92–97.

**Pavesi, A. (2000).** Detection of signature sequences in overlapping genes and prediction of a novel overlapping gene in hepatitis G virus. *Journal of Molecular Evolution* **50**, 284–295.

**Simmonds, P., Prescott, L. E., Logue, C., Davidson, F., Thomas, A. E. & Ludlam, C. A. (1999).** TT virus – part of the normal human flora? *Journal of Infectious Diseases* **180**, 1748–1750.

**Takahashi, K., Iwasa, Y., Hijikata, M. & Mishiro, S. (2000).** Identification of a new human DNA virus (TTV-like mini virus, TLMV) intermediately related to TT virus and chicken anemia virus. *Archives of Virology* **145**, 979–993.

**Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994).** CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

**Worobey, M. (2000).** Extensive homologous recombination among widely divergent TT viruses. *Journal of Virology* **74**, 7666–7670.

**Yuasa, N., Taniguchi, T. & Yoshida, I. (1979).** Isolation and characteristics of an agent inducing anemia in chicks. *Avian Diseases* **23**, 366–385.

Received 9 August 2000; Accepted 4 October 2000