Complete nucleotide sequence of Middelburg virus, isolated from the spleen of a horse with severe clinical disease in Zimbabwe

Houssam Attoui,[1] Corinne Sailleau,[2] Fauziah Mohd Jaafar,[1] Mourad Belhouchet,[3] Philippe Biagini,[3] Jean François Cantaloube,[3] Philippe de Micco,[3] Peter Mertens[1] and Stephan Zientara[2]

Correspondence
Houssam Attoui
houssam.attoui@bbsrc.ac.uk

[1]Department of Arbovirology, Institute for Animal Health, Pirbright, Woking, Surrey GU24 0NF, UK

[2]Agence Française de Sécurité Sanitaire des Aliments, 22 rue Pierre Curie, 94703 Maisons-Alfort Cedex 07, France

[3]Unité de Virologie Moléculaire, Etablissement Français du Sang Alpes-Méditerranée, 149 Boulevard Baille, 13005 Marseille, France

The complete nucleotide sequence of Middelburg virus (MIDV) was determined for strain MIDV-857 from Zimbabwe. The isolation of this virus in 1993 from a horse that died showing severe clinical signs represents the first indication that MIDV can cause severe disease in equids. Full-length cDNA copies of the viral genome were successfully synthesized by an innovative RT-PCR amplification approach using an 'anchor primer' combined with the SMART methodology described previously for the synthesis of full-length cDNA copies from genome segments of dsRNA viruses. The MIDV-857 genome is 11 674 nt, excluding the 5′-terminal cap structure and poly(A) tail (which varies in length from approximately 180 to approximately 220 residues). The organization of the genome is like that of other alphaviruses, including a read-through stop codon between the nsP3 and nsP4 genes. However, phylogenetic analyses of the structural protein amino acid sequences suggested that the MIDV E1 gene was generated by recombination with a Semliki Forest virus-like virus. This hypothesis was supported by bootscanning analysis using a recombination-detection program. The 3′ untranslated region of MIDV-857 also contains a 112 nt duplication. This study reports the first full-length sequence of MIDV, which was obtained from a single RT-PCR product.

INTRODUCTION

The genus *Alphavirus* belongs to the family *Togaviridae* and includes 29 virus species that are recognized by the International Committee for the Taxonomy of Viruses (ICTV) (Table 1). Alphaviruses are important pathogens of livestock animals and humans worldwide. The alphavirus particle contains a single copy of the virus genome, which is a linear, positive-sense, single-stranded RNA. The genome length of those alphaviruses that have been sequenced ranges from 11 442 nt for Semliki Forest virus to 11 851 nt for Ross river virus (Pfeffer *et al.*, 1998). The genome has a 5′-terminal cap and the 3′ end has a poly(A) tract. The 5′ two-thirds of the genome encodes the non-structural proteins nsP1 to nsP4, followed by the 26S junction region, which promotes transcription of the intracellular subgenomic 26S RNA. This latter mRNA

contains the structural genes that encode the capsid (C), E3, E2, 6K and E1 proteins and represents the remaining 3′ one-third of the genome (Pfeffer *et al.*, 1998).

Most alphaviruses are transmitted by haematophagous arthropods and can be grouped into eight antigenic complexes based on their serological cross-reactions. These are the Barmah Forest, Eastern equine encephalitis, Middleburg, Ndumu, Semliki Forest, Trocara, Venezuelan equine encephalitis and Western equine encephalitis complexes (Griffin, 2001; Weaver *et al.*, 2005; Table 1).

Antigenic cross-reactivities between alphaviruses reflect sequence conservation in the C protein and E1 glycoprotein, whilst antibodies directed against the E2 protein are usually virus specific (Griffin, 2001).

Within a given antigenic serocomplex, viruses exhibit more than 57 % nucleic acid and 56 % amino acid identities. Between members of distinct serocomplexes, these values are more than 62 and 60 %, respectively. Middelburg virus (MIDV, the only member of the MIDV serocomplex) is the

Table 1. List of the 29 species of the genus *Alphavirus*, together with their antigenic relationships and the host that they infect

The GenBank accession numbers of the sequences used in phylogenetic analysis are also provided. The current classification is from Weaver *et al.* (2005).

Virus species (abbreviation)	Known isolates	Complex	GenBank accession no.	Host*
1. *Aura virus* (AURAV)	Aura	WEE	AF126284	Human
2. *Barmah Forest virus* (BFV)	Barmah Forest	BF	NC_001786	Human
3. *Bebaru virus* (BEBV)	Bebaru	SF	AF339480	Human
4. *Cabassou virus* (CABV)	Cabassou	VEE		Marsupials, bats
5. *Chikungunya virus* (CHIKV)	Chikungunya	SF	NC_004162	Human
6. *Eastern equine encephalitis virus* (EEEV)	Eastern equine encephalitis	EEE	NC_003899	Human, equid, birds
7. *Everglades virus* (EVEV)	Everglades	VEE	AF075251	Human
8. *Fort Morgan virus* (FMV)	Fort Morgan	WEE	AF339475	Birds
	Buggy Creek	WEE	AF339474	Birds
9. *Getah virus* (GETV)	Getah	SF	NC_006558	Human, equid, pigs
10. *Highlands J virus* (HJV)	Highlands J	WEE	AF339476	Primates, perissodactyls, rodents
11. *Mayaro virus* (MAYV)	Mayaro	SF	DQ001069	Human
12. *Middelburg virus* (MIDV)	Middelburg	MID	AF023284/K00699/AF339486/ U94599/J02246	*Aedes* mosquitoes
13. *Mosso das Pedras virus* (MDPV)	Mosso das Pedras			Human, bats
14. *Mucambo virus* (MUCV)	Mucambo	VEE		Primates, perissodactyls, rodents
15. *Ndumu virus* (NDUV)	Ndumu	NDU	AF339487	Rodents
16. *O'nyong-nyong virus* (ONNV)	O'nyong-nyong	SF	NC_001512	Human
	Igbo Ora		AF079457	Human
17. *Pixuna virus* (PIXV)	Pixuna	VEE		Human, equid
18. *Rio Negro virus* (RNV)	Rio Negro			Human
19. *Ross River virus* (RRV)	Ross River	SF	DQ226993	Human, birds
	Sagiyama	SF	AB032553	Mammals
20. *Salmon pancreas disease virus* (SPDV)	Salmon pancreas disease		NC_003930	Atlantic salmon
	Sleeping disease		NC_003433	Rainbow trout
21. *Semliki Forest virus* (SFV)	Semliki Forest	SF	NC_003215	Human, equid
	Me Tri	SF	AF398380	Human, domestic animals
22. *Sindbis virus* (SINV)	Sindbis	WEE	NC_001547	Human
	Babanki	WEE	AF339477	Human
	Kyzylagach	WEE	AF398392	Mammals
	Ockelbo	WEE		Human
23. *Southern elephant seal virus* (SESV)	Southern elephant seal			Elephant seal
24. *Tonate virus* (TONV)	Tonate		AF398384	Primates, perissodactyls, rodents
25. *Trocara virus* (TROV)	Trocara	TRO	AF252265	Not known
26. *Una virus* (UNAV)	Una	SF	AF339481	Equid, other mammals
27. *Venezuelan equine encephalitis virus* (VEEV)	Venezuelan equine encephalitis	VEE	NC_001449	Human, equid
28. *Western equine encephalitis virus* (WEEV)	Western equine encephalitis	WEE	NC_003908	Human, equid
29. *Whataroa virus* (WHAV)	Whataroa	WEE	AF339479	Rodents?

*The host range of these viruses was obtained from information available at the ICTV database (http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/).

least divergent of all antigenic complexes, exhibiting more than 67 % nucleic acid and 69 % amino acid sequence identity when compared with other alphaviruses. In contrast, Trocara virus is the most divergent from all other alphaviruses, exhibiting only 57 % nucleic acid and 53 % amino acid identity (Powers *et al.*, 2001).

The type species of the genus *Alphavirus* is *Sindbis virus*, which has a very wide distribution, with isolates from Europe, Asia (India, the Philippines and China), Australia and many parts of Africa. Viruses related to Sindbis virus have also been isolated from New Zealand (Whataroa virus) and South America (Aura virus) (Griffin, 2001). However, many of the alphaviruses that are transmitted by mosquitoes are geographically restricted in their distribution. These viruses circulate primarily between their mosquito vectors and small mammals or birds, whilst infection of larger mammals (such as humans or horses, which are regarded as dead-end hosts) can result in severe or fatal forms of disease (Griffin, 2001). More recently, three alphaviruses have been isolated from fish (salmon pancreas disease virus and rainbow trout sleeping disease virus, classified as a single species) and from the elephant seal louse *Lepidophthirus macrorihini* (classified as a separate species) (Weaver *et al.*, 2005).

The pathogenic alphaviruses can be divided into those that cause a rash and arthritis (mainly the Old World alphaviruses) and those that cause encephalitis (New World alphaviruses), although some have not yet been shown to cause any disease. An example is MIDV, which has never been isolated from mammals or from marine animals. MIDV was originally isolated in South Africa during the summer of 1957. Two isolates of MIDV were reported from *Aedes caballus* (isolate AR749) and other *Aedes* mosquitoes (isolate AR747) (Kokernot *et al.*, 1957). Partial sequence information is available only for isolate AR749.

We isolated a new strain of MIDV (isolate MIDV-857) from the spleen of a horse that had died in Zimbabwe in 1993 with severe clinical signs similar to those caused by African horsesickness virus (AHSV; genus *Orbivirus*, family *Reoviridae*) (Mertens, 1994). This paper reports the sequence analysis of the complete MIDV-857 genome, using full-length amplification of cDNA (FLAC; Maan *et al.*, 2007) and the SMART methodology described previously (Attoui *et al.*, 2000).

METHODS

Cell lines and virus titration. MIDV-857 replication was tested in mammalian cells (including BHK-21 and Vero cells) and in C6/36 *Aedes albopictus* mosquito cells. BHK-21 and Vero cells were incubated at 37 °C in Eagle's minimum essential medium (EMEM) containing 10 % fetal bovine serum (FBS) with 5 % CO_2. C6/36 cells were grown in Leibovitz's L-15 medium at 27 °C. The virus titre ($TCID_{50}$) was assayed in 96-well microtitre plates containing BHK-21 and Vero cells at 6, 12, 18, 24 and 36 h post-infection.

Virus isolation and virus propagation. Fourteen paired samples of whole blood and spleen were obtained from the Veterinary Research

Laboratory in Causeway, Zimbabwe, during 1993. Thirteen of these tested positive for African AHSV in diagnostic RT-PCR assays (using first-round and nested PCR primers targeting AHSV genome segment 7; Zientara *et al.*, 1995). However, one spleen sample, which tested negative for AHSV by PCR, contained a virus that lysed BHK-21 cells 18 h after infection. The unidentified virus was subsequently plaque-purified three times on Vero cells using SeaPlaque agarose and plaques were identified by staining with trypan blue. The virus was further propagated for two passages in BHK-21 cells for biochemical and molecular biology studies. The unidentified virus was also screened by PCR for equine infectious anemia virus, equine arteritis virus and equine herpesviruses.

Determination of the nature of the virus. Clarified supernatants from infected cell cultures were treated with organic solvents (Freon 113 and Vertrel XF) to see whether the virus contained lipids (i.e. was enveloped or non-enveloped). Briefly, 10 ml of the clarified supernatant was mixed with an equal volume of the solvent and shaken vigorously. The solution was spun at 2000 *g* for 10 min at 4 °C, the supernatant was recovered and its infectivity was assessed by virus titration on BHK-21 cells. Virus preparations treated with either solvent were subsequently used in lipofection assays using Fugene-6 reagent (Roche) as described by the manufacturer. Briefly, BHK-21 cells were grown to subconfluency in six-well plates. In a separate tube, 7 µl Fugene-6 was added to 100 µl of the treated culture supernatant and mixed by pipetting. The supernatant was removed from the wells of the culture plates and replaced with fresh serum-free EMEM. The transfection mixture was added to the wells and mixed by gently shaking the plates. The plates were incubated for 6 h at 37 °C and the supernatant was then replaced with EMEM containing 5 % FBS.

Similarly, other aliquots of the virus suspension were treated with 1 % sodium deoxycholate. Briefly, 1 ml of the clarified supernatant was treated with an equal volume of 2 % sodium deoxycholate (final concentration 1 %). After incubation for 1 h at room temperature (Auletta & Marlowe, 1968), the mixture was diluted with an equal volume of fresh serum-free EMEM and infectivity was assessed by titration on BHK-21 cells.

To determine whether the virus genome was DNA or RNA, virus replication in all three cell types was also assessed by measurement of virus titre in the presence of 1 or 5 µg actinomycin D ml^{-1}.

Extraction of viral RNA, cDNA synthesis and cloning. RNA was extracted from cell culture supernatants or cell pellets at 72 h post-inoculation using RNA Now (Biogentex). The supernatant was clarified by low-speed centrifugation at 2000 *g* for 10 min and then concentrated from 200 ml down to 1 ml by ultrafiltration (molecular mass cut-off of 5000 Da; Sartorius). Five fractions of 200 µl of the resuspended material were extracted using RNA Now, as described by the manufacturer. The extracted RNA was dissolved in a total volume of 50 µl. Fractions of 16 µl were then used in ligation reactions with the anchor primer RTC12-Spacer: 5′-AGGTCTCGTAGACCGTG-CACC(12)TCCAGGTGCACGGTC-3′ (the number 12 in parentheses denotes 12 CH_2 residues constituting a spacer arm). Ligation was carried out using 20 U T4 RNA ligase (NE Biolabs) at 16 °C overnight. Unligated primer was subsequently removed by filtration through concentrators (molecular mass cut-off of 10 000 Da; Sartorius) and the RNA purified using an RNaid kit (Qbiogen).

The anchor primer molecule can fold, so that complementary 3′ and 5′ sequences base pair, forming a hairpin-like structure, thus preventing internal mispriming. This provides a reverse transcriptase priming site, which can be used to copy the RNA molecules after the anchor primer has been ligated to their 3′ termini. The purified and ligated RNA of MIDV was reverse transcribed using Superscript II reverse transcriptase at 40 °C in the presence of the SMART II

oligonucleotide (5′-AAGCAGTGGTATCAACGCAGAGTGGCCATT-ACGGCCGGG-3′) as described previously (Attoui *et al.*, 2000). Superscript II reverse transcriptase, like other Moloney murine leukemia virus reverse transcriptases, possesses a terminal cytidine transferase activity, thus adding poly(C) at the 3′ end of the cDNA (Chenchik *et al.*, 1998; Attoui *et al.*, 2000). The SMART II oligonucleotide has a poly(G) sequence at its 3′ end that can hybridize to the poly(C) at the 3′ end of the nascent cDNA, inducing template switching and therefore providing a new template for the reverse transcriptase to copy. This approach introduces two distinct target sequences at the 3′ and 5′ ends, respectively, of the first cDNA copy. The resulting cDNA was PCR amplified using the 5′ PCR primer 5′-AAGCAGTGGTATC-AACGCAGAGT-3′, the PCR-Spacer primer 5′-GTCCAGGTGCACGG-TCTACGAGACCT-3′ and a TripleMaster PCR kit (Eppendorf).

For rapid sequencing, the resulting full-length amplicon was sonicated and the DNA ends repaired using a cloned *Pfu* DNA polymerase (Clontech) in presence of dNTPs. Briefly, the DNA, in 500 μl TE buffer [10 mM Tris/HCl (pH 7.5), 1 mM EDTA] was sonicated using a microtip sonication probe, at an output capacity of 20 %, for 1 min. The smaller DNA fragments in the sonicated product were removed by ultrafiltration using Vivaspin 500 concentrators (PES membrane, molecular mass cut-off of 300 kDa; Sartorius). Only products that were larger than 800 bp were retained. The concentrate was diluted in TE buffer (pH 8.0) and reconcentrated again in the same column until the volume reached 40 μl. Five microlitres of $10 \times$ *Pfu* DNA polymerase buffer (Clontech) was added and mixed with dNTPs (0.25 mM final concentration) and incubated at 72 °C for 30 min to repair the DNA ends. The reaction was cleaned using a MinElute Reaction Cleanup kit (Qiagen). The DNA was eluted from the column using 15 μl water preheated to 70 °C. Two microlitres of $10 \times$ *Taq* DNA polymerase buffer (Invitrogen) was added and mixed with 0.2 μl 100 mM dATP (final concentration 1 mM) and 1 U *Taq* DNA polymerase for end tailing of the DNA. The mixture was incubated at 72 °C for 30 min and the DNA was purified using a MinElute Reaction Cleanup kit. The DNA was eluted from the column using 8 μl water preheated to 70 °C. The purified products were ligated into a pGEM-T plasmid (Promega) and transformed into JM109 competent bacteria as described by the manufacturer. The resulting clones were screened by PCR using M13 universal primers. Clones with PCR inserts larger than 800 bp were sequenced using a dRhodamine sequencing kit (Applied Biosystems).

Two primers were designed from the sequence of the 3′ untranslated region (UTR) of the of MIDV-857 genome to confirm the sequence of this region and to estimate the length of the poly(A) tract. These primers were Ds1 (5′-GTAGGCACTAGATATAGTAGAACGG-3′, nt 11101–11125) and Ds2 (5′-GGTAGGCAAAGGCATCATTAATCA-TC-3′, nt 11152–11177). These primers were used with the PCR-Spacer primer in PCR.

Sequence analysis. The contig sequence of the various cDNA clones was assembled using CONTIGEXPRESS (Vector NTI advance 10, version 10.1.1). The sequence of cDNA was compared with viral sequences deposited in GenBank/EMBL/DDBJ using the NCBI's BLAST program (http://www.ncbi.nlm.nih.gov/BLAST/).

These sequences were aligned using CLUSTAL W (version 1.83; Thompson *et al.*, 1994). Phylogenetic analyses were performed using the neighbour-joining method (Saitou & Nei, 1987) implemented in the MEGA 3.1 software (Kumar *et al.*, 2004). The p-distance model and the Kimura two-parameter or the Poisson correction model were used for tree building. The alpha shape parameter used for the gamma distribution analysis was calculated using the PAML package (Yang, 1997). The shape parameter measures how variable the rates are among sites. With a value of $\alpha > 1$, most sites have rates around 1 (similar rates) and only a few sites have either very low or very high rates. When $\alpha \leqslant 1$ (there is a relatively large amount of rate variation)

and most sites have very low rates, but there are evolutionary hot spots with higher rates.

Zuker's algorithm, as implemented in version 1.5 of the Vienna RNA Package RNAFOLD (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi), was used to draw the theoretical secondary structures of the 3′ UTRs of the RNA genomes.

Detection of recombination events between RNAs of the alphaviruses was carried out using the recombination detection program RDP (Martin & Rybicki, 2000), which compiles, alongside RDP itself, other recombination analysis programs including CHIMAERA, GENECOV, SISCAN and MAXCHI.

The GenBank accession numbers of the sequences used in this study are provided in Table 1.

RESULTS

Virus replication and titration

The MIDV-857 virus replicated in each of the three cell lines that were tested (Vero, BHK-21 and C6/36 cells). The virus consistently lysed BHK-21 and Vero cells by 18 h post-inoculation. However, it did not lyse the mosquito cells, which is consistent with previous findings concerning alphaviruses (Yunker & Cory, 1975). Assays in BHK-21 cells gave a virus titre in the supernatant from infected BHK-21 or Vero cell cultures of approximately 10^8 TCID$_{50}$ ml^{-1} at 18 h post-infection. A growth curve showing a time course for the titre determination is shown in Fig. 1.

MIDV-857 produced plaques of 3–4 mm in diameter.

Nature of the virus

MIDV-857 was isolated from a horse that died after showing severe clinical signs of disease. The cause of death was initially thought to be African horse sickness. However,
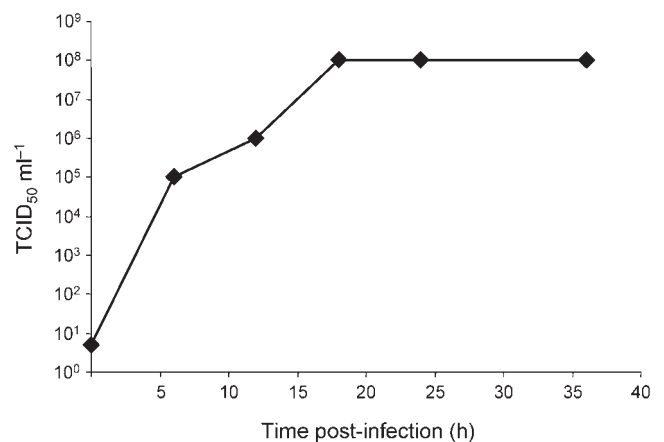
Fig. 1. Growth curve for MIDV-857. Time-course determination of the TCID$_{50}$ titre of MIDV-857 following inoculation of 5 TCID$_{50}$ ml^{-1} into 25 cm^2 culture flasks containing confluent monolayers of BHK-21 cells. The maximal titre of 10^8 TCID$_{50}$ ml^{-1} was reached by 18 h post-infection.

AHSV-specific RT-PCR assays gave negative results with tissue samples (blood and spleen; data not shown). Agarose gel electrophoresis of the RNA extracted from infected cell cultures also failed to show a dsRNA-segment migration pattern typical of AHSV isolates, indicating the absence of this virus (or any other member of the family Reoviridae) (data not shown).

The sample also tested negative for equine infectious anemia virus, equine arteritis virus and equine herpesviruses.

Treatment of the infected tissue culture supernatants with organic solvents (Freon 113 or Vertrel XF) or deoxycholate abolished infectivity. Following lipofection, high levels of infection were observed, indicating that lipofection could restore virus infectivity. The lipofected virus lysed the cells and lysis was maintained for six subsequent passages (this was to confirm that new virus progeny was produced). Mock-lipofected cells were not lysed, indicating that lysis only occurred when the replication-competent virus capsid was present.

The presence of actinomycin D did not prevent virus replication in BHK-21, Vero or C6/36 cells (as determined by subsequent titration in BHK-21 cells), indicating that the virus had an RNA genome. Titres of 10^8 TCID$_{50}$ (at 18 h post-infection) were consistently obtained, whether cells were treated with actinomycin D or not.

Cloning and sequencing of the MIDV-857 genome

The genome of isolate MIDV-857 was amplified as a single, full-length RT-PCR product, using a combination of a modified anchor primer (FLAC) and the SMART methodology (Fig. 2a, b). Cloning and sequencing of the genome showed it to be 11 674 nt, excluding a poly(A) tract that varied in length between approximately 180 and approximately 220 nt. The estimated length of the poly(A) tail is in agreement with the activity of eukaryotic poly(A) polymerase, which adds between 200 and 250 adenylate residues to the 3′ end of RNA (Bienroth et al., 1993). The complete nucleotide sequence of the amplified cDNA was determined and has been deposited in GenBank under accession number EF536323.

Primers DS1 and DS2 designed from the 3′ UTR of the viral genome were used with the PCR-Spacer primer to amplify a region of 574 bp (Ds1) or 523 bp (Ds2) from the end of the 3′ UTR region. The longest length of PCR product obtained with the Ds2 primer was estimated from the gel to be 750 bp (Fig. 2c), indicating that the maximum length of the poly(A) tail was approximately 220 nt. The organization of the non-structural and structural polyproteins is shown in Fig. 2(d) (individual 'cleaved-protein' start and end sites are also shown).

Sequence comparisons

Sequence analysis of the genome of MIDV-857 and comparisons with the GenBank/EMBL/DDBJ sequence databases clearly identified it as an isolate of MIDV, an alphavirus that has previously been isolated from mosquitoes but has never been associated with disease in mammals. The sequence databases only contained partial sequences for the structural and non-structural protein genes (GenBank accession numbers given in Table 1) of MIDV (isolate MIDV-SAAR749). MIDV-857 and MIDV-SAAR749 were 98 % identical at the nucleotide level and 99.5 % identical at the amino acid level in the structural polyprotein, and 100 % identical in the partial sequence of nsP4 that was available from GenBank. Although the 5′ UTR of MIDV-857 was identical to that of the earlier isolate (Ou et al., 1982), there were significant differences in the 3′ UTR. The sequence alignment not only identified six base changes between the two strains in this region (T/C or A/G transitions), but also identified a 112 nt duplication in the MIDV-857 3′ UTR (Fig. 3b). The downstream section of the repeated sequence had lost three nucleotides at two distinct positions and deletions were detected at two other positions in the 3′ UTR of MIDV-857 compared with that of MIDV-SAAR749 (Fig. 3a). The insertion in the MIDV-857 sequence changed the predicted secondary structure of the 3′ UTR as shown in Fig. 4 (RNAFOLD program).

The G+C content of the complete genome sequence of MIDV-857 was found to be 53 mol%, a value similar to that of the full-length genomes of other alphaviruses (48–56 mol%, the highest value being from the fish alphaviruses).

Phylogenetic comparison

Phylogenetic trees were built with the coding sequences of different alphaviruses. The values for the alpha shape parameter (used for the gamma distribution analysis) were calculated for different sequence sets using the PAML package and are provided in Table 2.

Neighbour-joining trees were constructed for the non-structural plus the structural coding sequence using the Kimura two-parameter model with gamma distance (Fig. 5a) and for the non-structural amino acid sequences using the p-distance model (Fig. 5b). The topologies showed that Western equine encephalitis virus (WEEV) clustered with Eastern equine encephalitis virus (EEV), whilst MIDV was clearly separate from any of the other serocomplexes.

In contrast, the trees built using the structural proteins (or the structural-gene nucleotide sequences) produced a topology where WEEV clustered with the Sindbis serocomplex viruses. They also showed that MIDV clustered within the Semliki Forest serocomplex when the structural polyprotein (using p-distance; Fig. 5c), the non-structural gene nucleotide sequences (using either the p-distance or Kimura two-parameter model; data not shown) or the E1 amino acid sequences were compared (Fig. 6a, b). However, when the structural polyprotein sequences were used in conjunction with a Poisson correction (using the
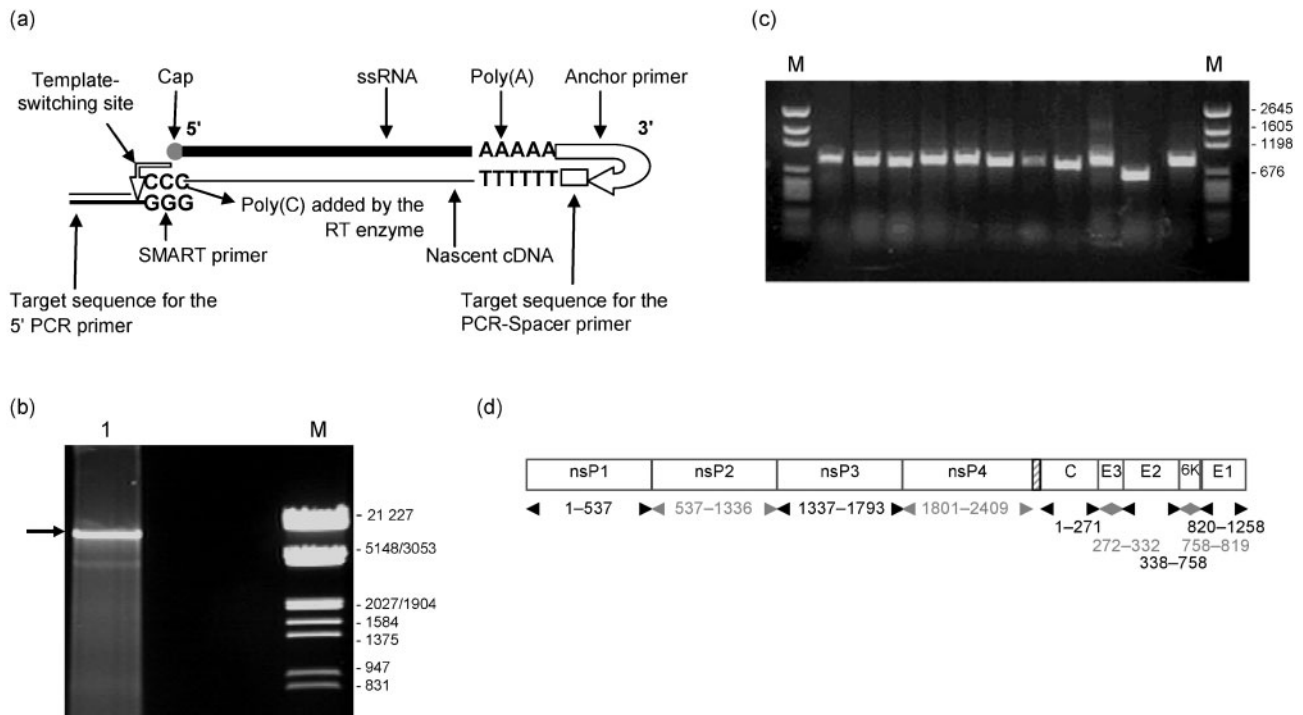
Fig. 2. Molecular cloning strategy, RT-PCR amplification and organization of the MIDV-857 genome. (a) Molecular cloning strategy. The anchor primer RTC12-Spacer [5′-AGGTCTCGTA<u>GACCGTGCACC</u>(12)TCCA<u>GGTGCACGGTC</u>-3′] folds so that its 3′ distal sequence hybridizes to a complementary sequence located immediately before the C12 bridge (both sequences are underlined). The reverse transcriptase adds a poly(C) sequence at the 3′ end of the nascent cDNA. The SMART II oligonucleotide primer has a poly(G) sequence at its 3′ end, which will hybridize to the poly(C) of the cDNA, inducing template switching and therefore providing a new template for the reverse transcriptase to copy. The newly synthesized cDNA now possesses a target sequence for the PCR-Spacer primer at its 5′ end and a target sequence for the 5′ PCR primer at its 3′ end. (b) RT-PCR amplification of the genome of MIDV-857. Lane 1, full-length PCR product of the MIDV-857 genome including the poly(A) tail (amplicon indicated by arrow). Lane M, size markers (bp). (c) PCR amplicons of the 3′ UTR, including the poly(A) tail, of MIDV-857. The length of the poly(A) tail is between approximately 180 and approximately 220 bases. Amplification was carried out using primer Ds1 or Ds2 and the PCR-Spacer primer. Direct sequencing using primers Ds1 and Ds2 confirmed the existence of the direct repeat in the 3′ UTR shown in Fig. 3(b). The variation in the length of the poly(A) tail is in agreement with the activity of the cellular poly(A) polymerase. This enzyme adds 200–250 adenylate residues at the 3′ end of the RNA (Bienroth et al., 1993), which explains the variation in the length of the tail. (d) Organization of the non-structural and structural polyproteins of MIDV-857. The positions of the beginning and end of each protein are indicated (with respect to the length of the non-structural or the structural polyproteins).

calculated shape parameter alpha), the topology of the tree (Fig. 5d) was identical to that obtained with non-structural amino acid sequences (Fig. 5b; MIDV clustered separately from the Semliki Forest serocomplex).

It has been suggested previously that WEEV was generated as a result of a recombination event between a member of the Sindbis virus group and EEV (Hahn et al., 1988). In our hands, bootscanning within the RDP program package identified potential recombination events involving the E1 region. These data indicated that MIDV was generated by recombination events involving members of the Semliki Forest serocomplex as sequence donors. This could explain the discrepancies between the trees obtained by analyses of structural and non-structural genes. The recombination region included nt 2430–2980 (551 nt) in MIDV-857

($P > 0.01$), where two separate events were identified. The first event concerned nt 2430–2585 where the parental donors were Semliki Forest virus and Mayaro virus. The second event concerned nt 2850–2980 where again the parental donors were Semliki Forest virus and Mayaro virus. These events were identified by RDP and the parallel programs run within the RDP package including CHIMAERA, GENECOV and MAXCHI.

DISCUSSION

MIDV was first isolated in 1957 from *Aedes* mosquitoes caught during a disease outbreak in sheep in the Eastern Cape Province in South Africa (Kokernot *et al.*, 1957). Field investigations showed that Wesselsbron virus (genus

(a)

```
MIDV-SAAR749          TAGATATAGTAGAACGGAAAATGGAAAACTAGAAAAATAGTAGGGTAGGCAAAGGCATCA
MIDV-857        11109 TAGATATAGTAGAACGGAAAATGGAAAACTAGAAAAATAGTAGGGTAGGCAAAGGCATCA 11168
                      ************************************************************

MIDV-SAAR749          TTAATCATCATACTAAGCAGCATACACTTATATAACCTCAAATACAACCTACAATTACAC
MIDV-857        11169 TTAATCATCATACTAAGCAGTATACACTTATATAACCTCAAATACAACCTATAATTACAC 11228
                      ********************  ****************************** ********

MIDV-SAAR749          TAACTGAGTAGAAAAACAAGAAAACCCATAAAAAGTAGAGTAAGCAATGACATTAACCCT
MIDV-857        11229 TAACTGAGTAGAAAAACAAGAAAACCCATAAAAAGTAGAGTAAGCAATGACATTAACCCT 11288
                      ************************************************************

MIDV-SAAR749          ATAATTACTAACACGTATACTAACAGACTAGAAAAAGAGAAAATTACAGAAAAAATAGAG
MIDV-857        11289 ATAATTACTAACACGTATACTAACAGACTAGAAAAGAGAAAATTACAGAAAAAGTAGAG 11348
                      *********************************** ****************** *****

MIDV-SAAR749          TAGGCAATGGCATTATTACATATTAACATATATAACACTACA------------------
MIDV-857        11349 TAGGCAATGGCATTATTACATATTAACATGTATAGCACTACAGTAGTAAGCAATGACATT 11408
                      *****************************  **** *******

MIDV-SAAR749          ------------------------------------------------------------
MIDV-857        11409 AACCCTATAATTACTAACACGTATACTAACAGACTAGAAAAGAGAAAATTACAGAAAAG 11468

MIDV-SAAR749          ------------------------------ATATAAC-------------TTACAA
MIDV-857        11469 TAGAGTAGGCAATGGCATTATTACATATTAACATATATAACACTACAATATAACTTACAA 11528
                      ******                               *    *******

MIDV-SAAR749          CACCATTGGATACTAATAATTGGTATAGAAGGGCCTTGCCACAGGCCGCCATAATTGTAA
MIDV-857        11529 CACCAT------------------ATAGAAGGGCCTTGCCACAGGCCGCCATAATT---- 11566
                      ******                  ***************************** 

MIDV-SAAR749          ACAGAACTAATTGGCAGAACTAAATTAAACAGAACTAATTGGCAGAACTAATTGGATACT
MIDV-857        11567 --------------------AT-AAACAGAACTAATTGGCAGAATTAATTGGATACT 1160
                      ** *********************** **********

MIDV-SAAR749          AATAATTGGATTATTTATGATTTGGAAATAATATCGCGACGATTGGCTAGTCGCAATTGG
MIDV-857        11603 AATAATTGGATTATTTATGATTTGGAAATAATATCGCGACGATTGGCTAGTCGCAATTGG 11662
                      ************************************************************

MIDV-SAAR749          TTTTAATATTCC
MIDV-857        11663 TTTTAATATTCC 11674
                      ************
```

(b)

```
MIDV-857        11263 GTAGAGTAAGCAATGACATTAACCCTATAATTACTAACACGTATACTAACAGACTAGAAA 11322
MIDV-857        11391 GTAG--TAAGCAATGACATTAACCCTATAATTACTAACACGTATACTAACAGACTAGAAA 11448
                      ****  ******************************************************

MIDV-857        11323 AAGAGAAAATTACAGAAAAAATAGAGTAGGCAATGGCATTATTACATATTAACAT 11377
MIDV-857        11449 AAGAGAAAATTACAGAAA-GTAGAGTAGGCAATGGCATTATTACATATTAACAT 11502
                      ******************  **********************************
```

Fig. 3. Alignment of the 3′ UTRs of MIDV-857 and MIDV-SAAR749, and sequence of the MIDV-857 direct repeat. (a) The sequence in bold indicates the direct repeat sequence (which is shown highlighted in a black box in the original MIDV-SAAR749), whilst sequences shaded in grey indicate deletions not found in the MIDV-857 3′ UTR. As the sequence of MIDV-SAAR749 is partial, numbering was provided for isolate MIDV-857 only. (b) The direct repeat sequence in MIDV-857 was confirmed by direct RT-PCR using primer Ds1 or Ds2. This direct repeat sequence (in bold) had a deletion of three nucleotides. *, Identical nucleotide.

Flavivirus) was responsible for that outbreak. Pools of mosquitoes collected from the same region were screened for other infectious agents and two strains of MIDV were isolated. These viruses produced viraemia and pyrexia in experimentally inoculated sheep. The intracerebral inoculation of monkeys (*Cercopithecus aethiops pygerythrus*) failed to produce viraemia or disease. However, chicks were found to become viraemic after inoculation, without any clinical signs. The virus was also found to be lethal to newborn mice by either the intracerebral or intraperitoneal route. Mosquitoes fed with virus-contaminated brain-blood suspension transmitted the virus upon feeding on a naïve lamb. As a consequence, this lamb developed viraemia and pyrexia (Kokernot *et al.*, 1957).

Following this first isolation from mosquitoes in South Africa, the virus has subsequently been isolated from *Aedes* mosquitoes on a number of occasions, especially in Senegal from *Aedes cumminsi* (Robin *et al.*, 1969). Attempts to isolate the virus from mammalian or avian hosts were unsuccessful. Interestingly, haemagglutination inhibition antibodies to MIDV were identified in Romania from migratory wild birds (Antipa *et al.*, 1984).

This study reports the first full-length sequence of MIDV-857 isolated from of a horse, which died showing severe clinical signs of disease. These signs were highly similar to those of African horse sickness (including a rise in body temperature, tachycardia, pulmonary affection and generalized oedema, particularly of the head and neck). The biological samples obtained from the horse tested negative for several horse pathogens, particularly AHSV, equine infectious anemia virus, equine arteritis virus and equine herpesviruses. These results prompted attempts to characterize and identify the virus that had been isolated in cell culture from the spleen sample. After plaque purification, the virus was found to be sensitive to organic solvent treatment, suggesting that it was enveloped. However, lipofection of the purified Vertrel XF-treated virus re-established infectivity, indicating the presence of 'capsid' that was replication competent (after cell entry) but incapable of cell attachment and penetration by itself. Virus replication was not inhibited by the use of actinomycin D, suggesting that the virus had an RNA genome.

The anchor primer/SMART approach that we developed permitted PCR amplification of the MIDV-857 genome as a single PCR product of 11 674 nt. The genome organization was typical of alphaviruses and contained a poly(A) tail, which varied in length between approximately 180 and approximately 220 adenylate residues. Poly(A) tails are added by a cellular poly(A) polymerase, which incorporates between 200 and 250 adenylate residues into a

(a)

(b)
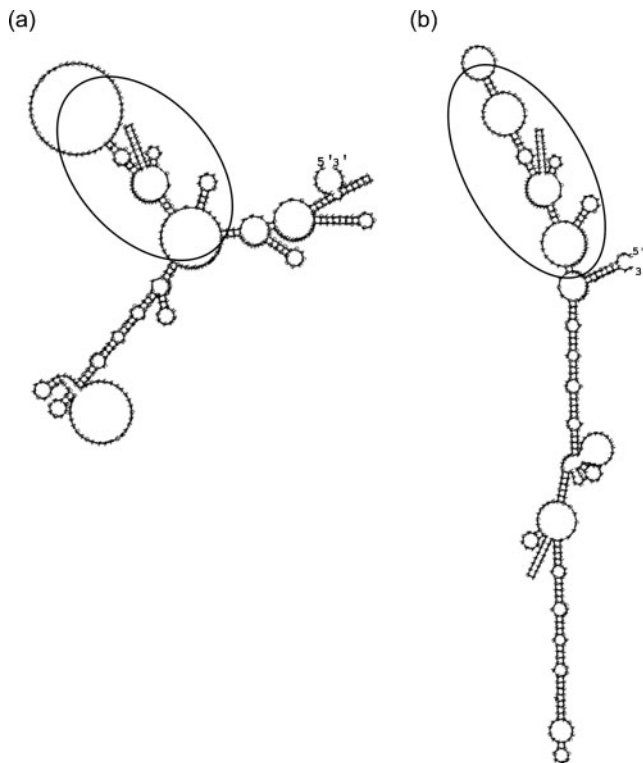
Fig. 4. Secondary structures of the 3′ UTRs of MIDV. Theoretical secondary structures formed by the 3′ UTR of the original South African isolate MIDV-SAAR749 (a) and the Zimbabwe 1993 isolate MIDV-857 (b) (sequences shown in Fig. 3a). Zuker's algorithm, as implemented in version 1.5 of the Vienna RNA package RNAFOLD, was used to draw the secondary structures. Shaded areas represent the region where the fold was not affected by the insertion.

Table 2. Values of the α shape parameter of the various sequence sets used in the phylogenetic reconstructions

The shape parameter measures how variable the rates are among sites. With a value of $\alpha > 1$, most sites have rates around 1 (similar rates) and only a few sites have either very low or very high rates. When $\alpha \leqslant 1$ (there is a relatively large amount of rate variation), most sites have very low rates, but there are evolutionary hot spots with higher rates.
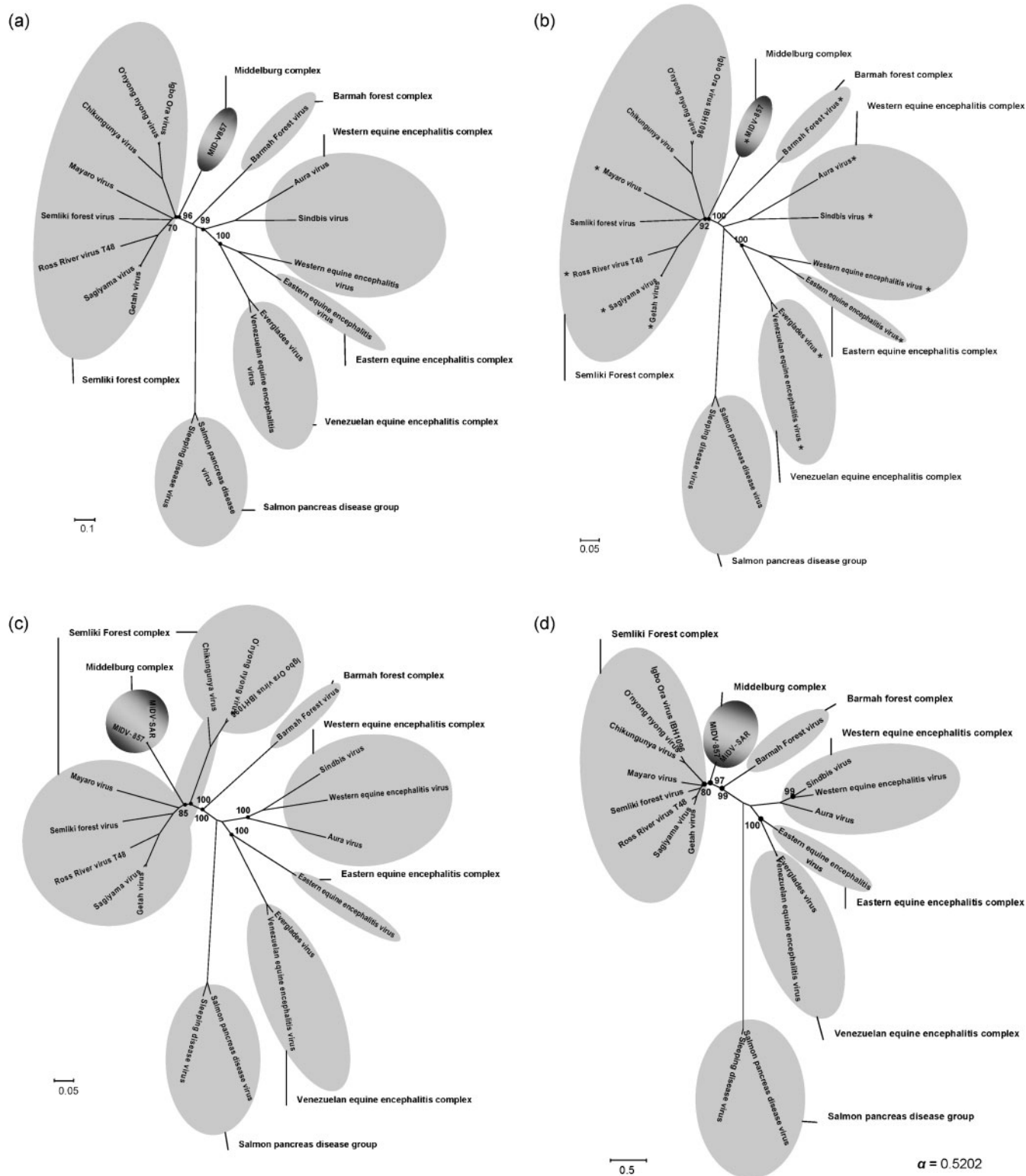
Sequence	Alpha value
Complete coding sequence	0.52320
Non-structural polyproteins	0.55722
Structural coding sequence	0.23625
Structural polyproteins	0.52020
E1 protein	1.13573

Semliki Forest virus complex clade (Powers *et al.*, 2001). However, it appears likely that a recombination event (within the E1 gene) between MIDV and other viruses of this serocomplex has occurred, leading to inconsistent phylogenetic results depending on the region of the alphavirus genome that is being compared. This hypothesis was subjected to analysis using bootscanning, implemented in the RDP program, which identified possible recombination events between MIDV and viruses of the Semliki Forest serocomplex. The mechanisms that resulted in the proposed recombination event are not known.

The process of genome replication for plus-sense RNA viruses begins with the generation of a full-length complementary copy of the genomic RNA (Hardy, 2006). The synthesis of this anti-genomic minus-strand RNA by the RNA-dependent RNA polymerase (RdRp) must start at the 3′ end of the genome in order to produce a full-length copy of the virus genome. Alphaviruses possess a highly conserved 3′ sequence element (3′ CSE; approximately 19 nt), which immediately precedes the poly(A) tail (Pfeffer *et al.*, 1998). Both the poly(A) tail and the 3′ CSE are required for virus replication and, more specifically, for efficient minus-strand RNA synthesis (Hardy & Rice, 2005; Kuhn *et al.*, 1991, 1992; Raju *et al.*, 1999). An analysis of the 3′ UTR of MIDV-857 and MIDV-SAAR749 identified an insertion of 112 nt within the 3′ UTR of MIDV-857. This insertion represents a direct repetition of the sequence located immediately upstream of the insertion site, but does not modify the 3′ CSE region and therefore is not thought to have any deleterious effect on replication (as confirmed by the high titres of virus produced in cell cultures).

It appears likely that the insertion sequence modifies the secondary structure of the 3′ UTR, although the effects of these changes remain unclear. However, some features of the predicted secondary structure fold are maintained between the UTRs of MIDV-857 and MIDV-SAAR749, and there are similarities to Semliki Forest virus and O'nyong nyong virus of the SFV serocomplex (data not shown).

tail (Bienroth *et al.*, 1993). This explains the variation in the MIDV poly(A) tail lengths, as individual encapsidated genomes would have tails of distinct lengths.

Comparison of the complete MIDV-857 genome sequence with the GenBank/EMBL/DDBJ databases identified it as an isolate of the species *Middelburg virus*, which forms a distinct antigenic complex within the genus *Alphavirus*. Phylogenetic comparisons of the MIDV-857 non-structural sequence with those of other alphaviruses identified MIDV as a member of a phylogenetic group on its own (which correlates with MIDV being a member of a separate serocomplex). However, when E1 amino acid sequences were used for neighbour-joining phylogenetic reconstruction (using p-distances or a Poisson correction), a major shift was observed in the tree, with MIDV clustering within the Semliki Forest serocomplex.

This inconsistency between the inter-relationships of established *Alphavirus* serocomplexes and the evolutionary similarities based on sequence analysis has led previous authors to suggest that MIDV may be a member of the

(a)

(b)

(c)

(d)

Direct repeats have been described previously in the 3′ UTRs of the flaviviruses (Gritsun & Gould, 2007a, b). It was considered likely that the 3′ UTR of the flaviviruses (and possibly the open reading frames) had evolved through multiple duplication of a single RNA domain,

and that short direct repeats appeared to represent an evolutionary remnant of these domains. Inspection of the UTRs of various alphaviruses has revealed stretches of 18–102 bases that occur at least twice in the viral RNA (Pfeffer et al., 1998). A mechanism by which such duplications
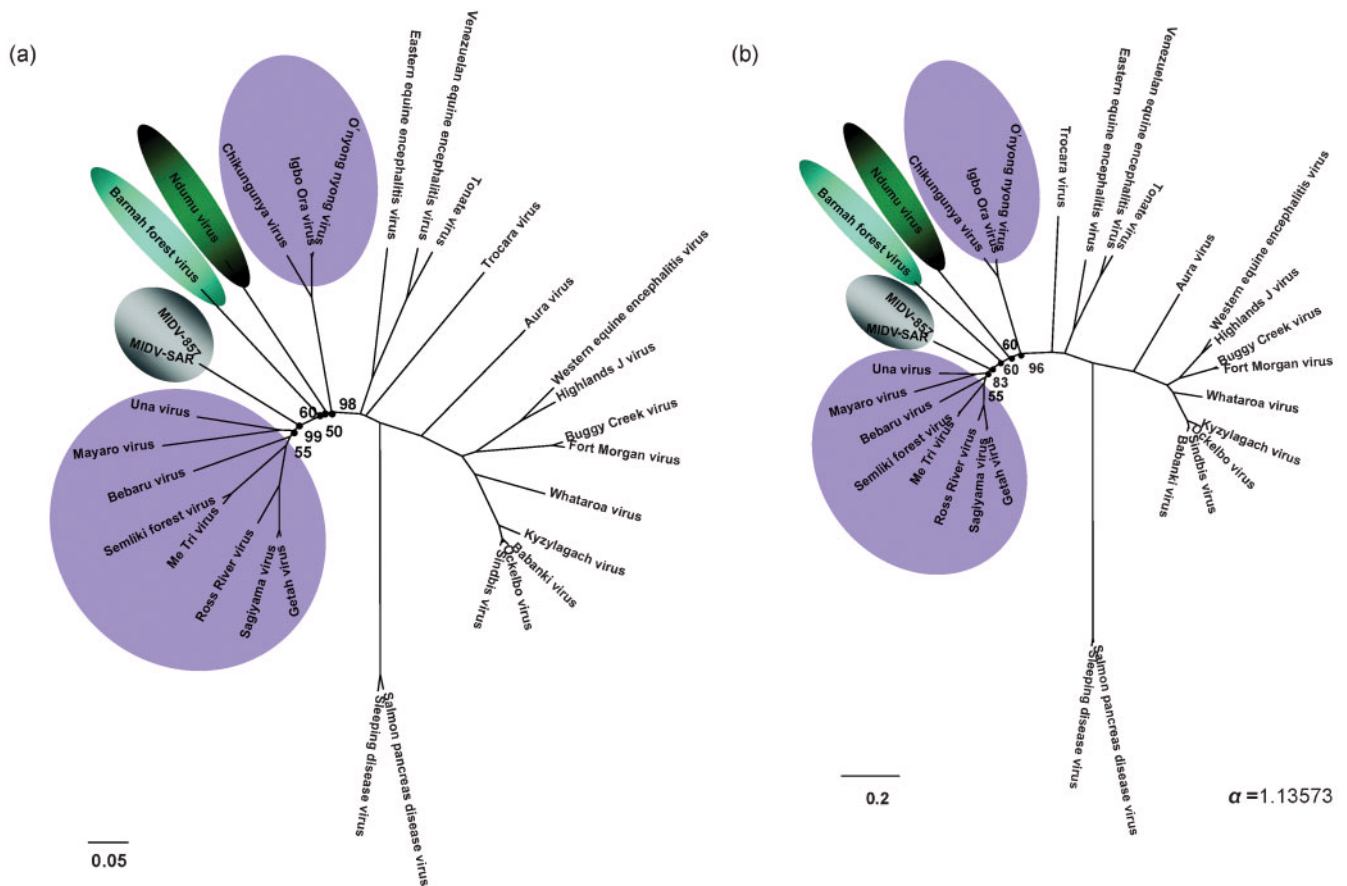
Fig. 6. Phylogenetic relationships of the alphaviruses, based on partial amino acid sequences of the E1 protein (aa 910–1244 of the structural polyprotein, corresponding to aa 91–425 of the E1 protein) with respect to the MIDV-857 sequence. Neighbour-joining trees using the p-distance model (a) and the Poisson correction model (b; the value of alpha is given at the lower right of the tree) are shown. Bootstrap values are shown at the nodes. Bars indicate the number of nucleotide substitutions per site. Lilac, Semliki Forest serocomplex; grey, Middelburg serocomplex; turquoise, Barmah Forest serocomplex; green, Ndumu serocomplex.

could occur has been proposed for the RdRp of a dsRNA virus, with a loop mechanism at the origin of the partial duplication (Matthijnssens *et al.*, 2006).

In conclusion, the isolation of MIDV from a diseased horse provides an important first indication that MIDV can cause disease in equids. The methodology used in the

Fig. 5. Phylogenetic relationships between members of the genus *Alphavirus*. (a) Phylogenetic relationships of the alphaviruses, based on complete coding nucleotide sequences (structural and non-structural). A neighbour-joining tree was constructed using the Kimura two-parameter model ($\alpha=0.5232$). (b) Phylogenetic relationships between the non-structural polyprotein sequences of the alphaviruses. A neighbour-joining tree was constructed using the p-distance model. A tree constructed using the Kimura two-parameter model had an identical topology (not shown). Viruses where read-through translation of nsP4 occurs are indicated with an asterisk. (c, d) Phylogenetic relationships between the structural polyprotein sequences (amino acids) of alphaviruses, shown by neighbour-joining trees built using the p-distance model (c) or the Poisson correction model with a value of $\alpha=0.5202$ calculated from the sequence set (d). Note the change in the clustering of WEEV and MIDV compared with the complete nucleotide coding sequence and the tree based on amino acids of the non-structural genes. Recombination events between WEEV and the Sindbis virus group have been described previously. In this analysis of the coding sequence of the structural genes using the RDP recombination detection program, it was possible to detect a potential recombination. This recombination event identified MIDV as a daughter sequence with parental donors being members of the Semliki Forest serocomplex. Antigenic complexes are indicated by the grey ellipses. Bootstrap values are indicated at the nodes. Bars indicate the number of nucleotide substitutions per site.

RT-PCR amplification of MIDV RNA avoids false priming, making it particularly important for the determination of the full sequence of the 3′ end of the RNA. Finally, the sequence analysis showed that WEEV is not the only recombinant virus within the genus *Alphavirus*, as MIDV also has the same status. The E1 gene of MIDV is probably a recombinant product from viruses of the Semliki Forest virus serocomplex.

ACKNOWLEDGEMENTS

REFERENCES

Antipa, C., Girjabu, E., Iftimovici, R. & Drăgănescu, N. (1984). Serological investigations concerning the presence of antibodies to arboviruses in wild birds. *Virologie* **35**, 5–9.

Attoui, H., Billoir, F., Cantaloube, J. F., Biagini, P., de Micco, P. & de Lamballerie, X. (2000). Strategies for the sequence determination of viral dsRNA genomes. *J Virol Methods* **89**, 147–158.

Auletta, A. E. & Marlowe, M. L. (1968). Effect of sodium deoxycholate on rubella virus. *Appl Microbiol* **16**, 1224.

Bienroth, S., Keller, W. & Wahle, E. (1993). Assembly of a processive messenger RNA polyadenylation complex. *EMBO J* **12**, 585–594.

Chenchik, A., Zhu, Y. Y., Diatchenko, L., Li, R., Hill, J. & Siebert, P. D. (1998). Generation and use of high-quality cDNA from small amounts of total RNA by SMART™ PCR. In *Gene Cloning and Analysis by RT-PCR*, pp. 305–319. Edited by P. Siebert & J. Larrick. Natick, MA: Biotechniques Books.

Griffin, D. E. (2001). Alphaviruses. In *Fields Virology*, 4th edn, pp. 917–962. Edited by D. M. Knipe & P. M. Howley. Philadelphia, PA: Lippincott Williams & Wilkins.

Gritsun, T. S. & Gould, E. A. (2007a). Direct repeats in the flavivirus 3′ untranslated region; a strategy for survival in the environment? *Virology* **358**, 258–265.

Gritsun, T. S. & Gould, E. A. (2007b). Origin and evolution of 3′UTR of flaviviruses: long direct repeats as a basis for the formation of secondary structures and their significance for virus transmission. *Adv Virus Res* **69**, 203–248.

Hahn, C. S., Lustig, S., Strauss, E. G. & Strauss, J. H. (1988). Western equine encephalitis virus is a recombinant virus. *Proc Natl Acad Sci U S A* **85**, 5997–6001.

Hardy, R. W. (2006). The role of the 3′ terminus of the Sindbis virus genome in minus-strand initiation site selection. *Virology* **345**, 520–531.

Hardy, R. W. & Rice, C. M. (2005). Requirements at the 3′ end of the Sindbis virus genome for efficient synthesis of minus-strand RNA. *J Virol* **79**, 4630–4639.

Kokernot, R. H., de Meillon, B., Paterson, H. E., Heymann, C. S. & Smithburn, K. C. (1957). Middelburg virus; a hitherto unknown agent isolated from *Aedes* mosquitoes during an epizootic in sheep in the eastern Cape Province. *S Afr J Med Sci* **22**, 145–153.

Kuhn, R. J., Niesters, H. G., Hong, Z. & Strauss, J. H. (1991). Infectious RNA transcripts from Ross River virus cDNA clones and the construction and characterization of defined chimeras with Sindbis virus. *Virology* **182**, 430–441.

Kuhn, R. J., Griffin, D. E., Zhang, H., Niesters, H. G. & Strauss, J. H. (1992). Attenuation of Sindbis virus neurovirulence by using defined mutations in nontranslated regions of the genome RNA. *J Virol* **66**, 7121–7127.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**, 150–163.

Maan, S., Rao, S., Maan, N. S., Anthony, S. J., Attoui, H., Samuel, A. R. & Mertens, P. P. C. (2007). Rapid cDNA synthesis and sequencing techniques for the genetic study of bluetongue and other dsRNA viruses. *J Virol Methods* **143**, 132–139.

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

Matthijnssens, J., Rahman, M. & Van Ranst, M. (2006). Loop model: mechanism to explain partial gene duplications in segmented dsRNA viruses. *Biochem Biophys Res Commun* **340**, 140–144.

Mertens, P. P. C. (1994). Orbiviruses and coltiviruses. In *Encyclopedia of Virology*, pp. 941–956. Edited by R. G. Webster & A. Granoff. London: Academic Press.

Ou, J. H., Trent, D. W. & Strauss, J. H. (1982). The 3′-non-coding regions of alphavirus RNAs contain repeating sequences. *J Mol Biol* **156**, 719–730.

Pfeffer, M., Kinney, R. M. & Kaaden, R. O. (1998). The alphavirus 3′-nontranslated region: size heterogeneity and arrangement of repeated sequence elements. *Virology* **240**, 100–108.

Powers, A. M., Brault, A. C., Shirako, Y., Strauss, E. G., Kang, W., Strauss, J. H. & Weaver, S. C. (2001). Evolutionary relationships and systematics of the alphaviruses. *J Virol* **75**, 10118–10131.

Raju, R., Hajjou, M., Hill, K. R., Botta, V. & Botta, S. (1999). In vivo addition of poly(A) tail and AU-rich sequences to the 3′ terminus of the Sindbis virus RNA genome: a novel 3′-end repair pathway. *J Virol* **73**, 2410–2419.

Robin, Y., Cornet, M., Bres, P., Hery, G. & Château, R. (1969). Isolation of a strain of Middelburg virus from a batch of *Aedes (A.) cumminsi* gathered in Bandia (Senegal). *Bull Soc Pathol Exot Filiales* **62**, 112–118.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

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

Weaver, S. C., Frey, T. K., Huang, H. V., Kinney, R. M., Rice, C. M., Roehrig, J. T., Shope, R. E. & Strauss, E. G. (2005). *Togavirigae*. In *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*, pp. 999–1008. Edited by C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. London: Elsevier/Academic Press.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**, 555–556.

Yunker, C. E. & Cory, J. (1975). Plaque production by arboviruses in Singh's *Aedes albopictus* cells. *Appl Microbiol* **29**, 81–89.

Zientara, S., Sailleau, C., Moulay, S., Wade-Evans, A. & Cruciere, C. (1995). Application of the polymerase chain reaction to the detection of African horse sickness viruses. *J Virol Methods* **53**, 47–54.