



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Re-emergence of a genetic outlier strain of equine arteritis virus: Impact on phylogeny



F. Steinbach, D.G. Westcott, S.L. McGowan, S.S. Grierson, J.P. Frossard, B. Choudhury\*

Department of Virology, Animal and Plant Health Agency, Weybridge, Surrey KT15 3NB, United Kingdom

### ARTICLE INFO

#### Article history:

Available online 17 December 2014

#### Keywords:

Equine arteritis virus  
Phylogeny  
Classification  
Full genome sequencing

### ABSTRACT

Equine arteritis virus (EAV) is the causative agent of equine viral arteritis (EVA), a respiratory and reproductive disease of equids, which is notifiable in some countries including the Great Britain (GB) and to the OIE. Herein, we present the case of a persistently infected stallion and the phylogenetic tracing of the virus strain isolated. Discussing EAV occurrence and phylogenetic analysis we review features, which may aid to harmonise and enhance the classification of EAV.

Crown Copyright © 2015 Published by Elsevier B.V. All rights reserved.

### 1. The EAV genome

EAV is an enveloped, linear, positive sense, single stranded RNA virus representing the prototype virus of the *Arteriviridae* family, in the order *Nidovirales*. This family also includes simian haemorrhagic fever virus, porcine reproductive and respiratory syndrome virus (PRRSV) and lactate dehydrogenase-elevating virus (Cavanagh, 1997).

The genome of EAV (Fig. 1) is approximately 12.7 kb in length and consists of 10 overlapping open reading frames (ORFs), which are flanked by 5' and 3' untranslated regions (UTRs) (Snijder and Meulenberg, 1998; Firth et al., 2011; Balasuriya et al., 2014). ORFs 1a and b occupy approximately 75% of the genome and both are translated to produce polyproteins, which are then extensively processed into 13 non-structural proteins (Table 1) (Snijder and Meulenberg, 1998). At the 3' end, occupying approximately 25% of the genome, the remaining eight ORFs, 2–7 are nested and encode seven envelope proteins and the viral nucleocapsid (Table 1) (Snijder and Meulenberg, 1998; Firth et al., 2011).

Although there is only one serotype of EAV, variation exists between field strains of the virus and their neutralising properties, i.e. neutralising epitopes are not shared across all strains as evident from the use of monoclonal antibodies (Glaser et al., 1995; Balasuriya et al., 1997). ORF5 encodes the major envelope protein, GP5, which contains the best characterised neutralisation determinants. This is also a most variable region within the genome and hence is commonly used for phylogenetic analysis (Balasuriya et al.,

1995; Stadejek et al., 1999; Hornyak et al., 2005; Mittelholzer et al., 2006; Metz et al., 2011). Phylogenetic analysis has also been conducted using ORFs 3, 6 and 7 (Lepage et al., 1996; Hedges et al., 2001; Chirnside et al., 1994).

Despite variations in gene, phylogenetic analysis models and number of sequences used the general consensus of such phylogenetic studies show a segregation of sequences into European and North American lineages of which the European lineage can be further divided into two clusters. Unsurprisingly, some European lineage viruses have been documented in North America and vice versa, highlighting the transport of viruses between the two continents via trade of horses and semen.

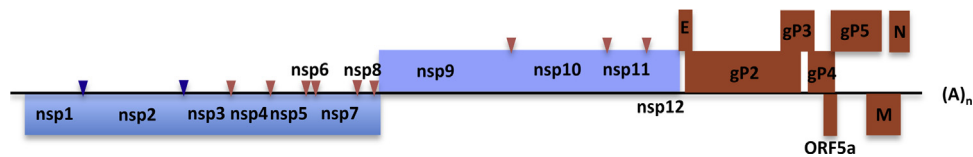
Excluding patent applications and molecular clones there are 800 plus EAV sequences available in GenBank (as of June 2014). Thirty-two of these are full genomes: two (accession numbers: NC002532 and DQ846750) are different iterations of Bucyrus, sharing 99.6% homology at the nucleotide level; five are from the French outbreak in 2007 and the other 25 are from the USA.

Of the remaining sequences, around 25% of sequences deposited are of North American origin, another 25% were reported in Western Europe and a further 25% from Eastern Europe. There is only one Asian sequence, from China (JX868590) and three from the Middle East (Qatar; accession no. AY453329-32). The African continent is also poorly represented with only 16 sequences, all reported from South Africa. There are around sixty sequences for which the location was not stated in the GenBank record or associated manuscripts.

As mentioned the majority of phylogenetic analysis so far has been conducted on ORF5. Of the other segments used for phylogenetic analysis, ORF3 offers over 250 sequences, ORF6 over 210 sequences and ORF7 over 280 sequences, compared to nearly 600

\* Corresponding author. Tel.: +44 01932 357559.

E-mail address: [Bhudipa.Choudhury@apha.gsi.gov.uk](mailto:Bhudipa.Choudhury@apha.gsi.gov.uk) (B. Choudhury).



**Fig. 1.** EAV genome organisation (adapted from [Snijder and Kikkert, 2013](#) and GenBank's genome section on EAV). The ORF1ab (blue) is expressed as a fusion protein, from the genomic RNA, despite the  $-1$  programmed ribosomal frameshift required for ORF1b (light blue). The genes encoding the structural proteins (red) are translated from subgenomic RNAs. Polyprotein cleavage sites are depicted above ORF1ab. Blue arrowheads: sites cleaved by PLP domains of nsp1/nsp2; red arrowheads: sites cleaved by the nsp4. The graph displays the relative sizes and locations of the resulting proteins. Proteins expressed in frame to each other are displayed on the same axis. NSP: non-structural protein; GP: glycoprotein; M: membrane protein; E: envelope protein; N: nucleocapsid protein. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sequences which will allow use of ORF5 e.g. including those which cover ORFs 2–7. However, where the focus is singularly on ORF5 only, the majority (188/278) sequences deposited in GenBank are partial sequences which are 518 nucleotides in length and cover genome positions 11296–11813 (Bucyrus) representing the most variable region within EAV ORF5.

## 2. EAV distribution and occurrence

EAV was first isolated in Bucyrus, OH, USA in 1953, from foetal lung tissue during an outbreak of respiratory disease and abortions in horses ([Doil et al., 1957](#); reviewed in [Balasuriya et al., 2014](#)). The virus is thought to have an almost global distribution: via serological surveys EAV has been noted in equine populations in Europe, Africa, Asia, the Americas and Australia ([Timoney and McCollum, 1993](#)). The lack of unified control measures, poor biosecurity and the increased international trade in horses has assisted in the spread of EAV. The only countries considered to be disease free are Iceland and Japan. Recently, [McFadden et al. \(2013\)](#) demonstrated a lack of EAV prevalence in the general horse population in New Zealand.

A hallmark of EAV is its persistence in and spread through stallions. Larger clinical outbreaks occur only periodically and genetic markers for pathogenicity have been suggested, but lack further characterisation ([Balasuriya et al., 2014](#)). The first clinical appearance of EAV in the United Kingdom was in 1993 ([Wood et al., 1995](#)). The source of the outbreak was a stallion imported from Poland. The outbreak began on a non-thoroughbred stud and spread to five other premises via chilled semen used for artificial insemination. Around 100 animals became infected and the outbreak was

contained by means of voluntary movement restrictions. Subsequently EAV was made notifiable in the UK (EAV order 1995). Since then, there have been sporadic cases up to and including the case of the imported Spanish stallion in 2012, which is described in this manuscript.

During the last decade clinical outbreaks of EAV occurred in several countries on both sides of the Atlantic. In the summer of 2007, the largest EVA outbreak in France occurred causing mortality and economic loss due to disruption of race schedules ([Pronost et al., 2010](#)). The outbreak affected 18 premises in 5 counties in western France, which represented the index, 8 primary and 9 secondary premises. Artificial insemination in draught horses was responsible for the index case.

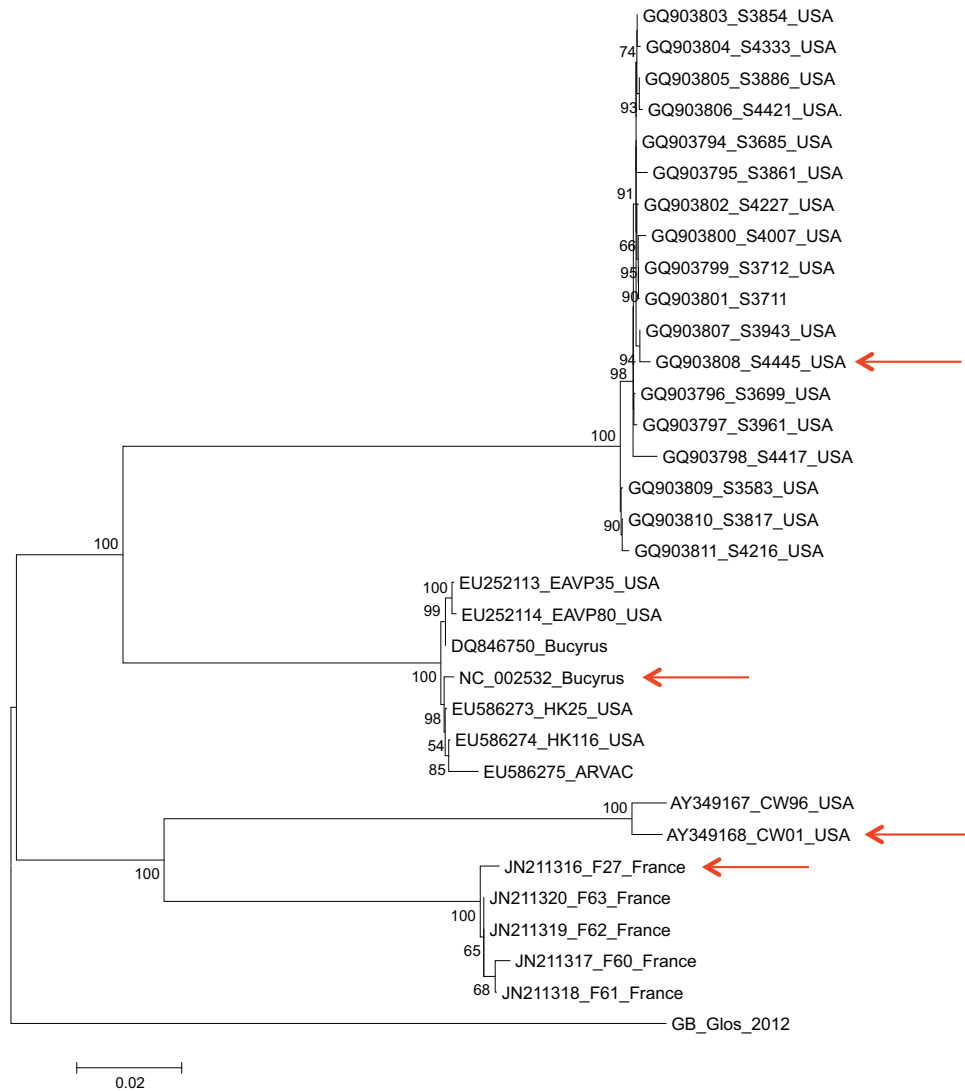
In a similar timeframe (2006–2007), a multistate US outbreak occurred and for the first time disease was observed in Quarter Horses ([Powell and Timoney, 2006](#)). Due to the nature of EAV dissemination through infectious semen, EVA prevalence can show disparity between horse breeds e.g. prior to that outbreak seroprevalence in Quarter Horses was recorded at 0.6% ([Anonymous, 2000](#)).

There is limited information regarding EAV prevalence in South America. [Echeverría et al. \(2003\)](#) reported the first isolation of EAV in Argentina, and perhaps South America, in 2001. The virus was isolated from the semen of an imported seropositive stallion held in isolation at a breeding farm in the Buenos Aires Province.

Three additional isolates were obtained in 2007 with phylogenetic analysis suggesting an European lineage ([Echeverría et al., 2007](#)). Shortly thereafter in 2010 another significant outbreak occurred in Argentina, which was linked to a semen import from the Netherlands ([Barrandeguy, 2010](#)).

**Table 1**  
EAV genes, proteins and their functions.

ORF	Protein	Function
1a	nsp1	Accessory protease: important role in virion biogenesis and crucial in processing of the replicase polyprotein and the production of subgenomic RNA ( <a href="#">Tijms et al., 2007</a> )
	nsp2	Accessory-protease: auto-protease activity between nsp2 and 3. Co-factor for the successful processing of nsp4/5 ( <a href="#">Snijder et al., 1994</a> ). Deubiquitinating enzyme activity ( <a href="#">van Kasteren et al., 2013</a> ).
	nsp3	Transmembrane non-structural protein thought to be involved in double-membrane vesicle formation ( <a href="#">Posthuma et al., 2008</a> ).
	nsp4	Main protease: responsible for cleaving nsp3–8 and nsp3–12 polypeptides ( <a href="#">Snijder et al., 1994</a> ).
	nsp5	Suggested role in the membrane association of the EAV replication complex. ( <a href="#">Snijder et al., 1994</a> ).
	nsp6	Function not determined.
	nsp7	Further cleaved into nsp7 $\alpha$ and nsp7 $\beta$ , function not fully determined, critical role in RNA synthesis ( <a href="#">Manolaridis et al., 2011</a> ).
	nsp8	Function not determined.
	nsp9	RNA-dependent RNA polymerase ( <a href="#">Snijder et al., 1994</a> ).
	nsp10	Helicase ( <a href="#">van Dinten et al., 1997</a> ).
	nsp11	Multifunctional including endoribonuclease activity, critical role in RNA synthesis ( <a href="#">Posthuma et al., 2006</a> ).
	nsp12	Function not determined.
2a	E	Predicted to be a membrane protein which is essential for the production of infectious progeny ( <a href="#">Snijder et al., 1999</a> ).
2b	GP2	Minor envelope protein: membrane protein which is essential for the production of infectious progeny ( <a href="#">Snijder et al., 1999</a> ).
3	GP3	Membrane associated protein: function not fully determined but thought essential for viral replication ( <a href="#">Wieringa et al., 2002</a> ).
4	GP4	Membrane glycoprotein: function not fully determined but thought essential for viral replication ( <a href="#">Wieringa et al., 2002</a> ).
5a	ORF5a	Membrane protein: function not fully determined but thought important for viral replication ( <a href="#">Firth et al., 2011</a> ).
5	GP5	Major envelope protein, location of neutralisation domains ( <a href="#">Balasuriya et al., 1997</a> )
6	M	Envelope protein suggested to be involved in virus budding ( <a href="#">de Vries et al., 1995</a> )
7	N	Nucleocapsid protein suggested to bind to envelope protein domains during budding ( <a href="#">Tijms et al., 2002</a> ).



**Fig. 2.** Phylogenetic comparison of EAV full length genome sequences. GB.Glos.2012 was compared against all of the full genome sequences ( $n=32$ ; excluding molecular clones) available in GenBank. The arrows indicate the strains selected from each cluster for genomic comparison (Table 2) with GB.Glos.2012. The Maximum-likelihood phylogeny was constructed using the GTR +G+ I model. Sequences are listed by their accession numbers followed by strain and country of isolation. Percentage of a 1000 bootstrap replicates are cited at nodes for values  $\geq 50\%$ .

Scant information is only available regarding the disease situation in Africa. Paweska et al. (1996) reported high seroprevalence of an asinine strain of EAV in South African donkeys. Guthrie et al. (2003) reported lateral transmission among Lipizzaner stallions caused by an imported stallion from Yugoslavia in 1981. A few years later, Stadejek et al. (2006) reported the isolation of a highly divergent EAV strain from a South African donkey.

No internationally published information appears to be available for remainder of Africa, Asia and the Middle East; however the fact there are sequences from these regions available on GenBank shows the disease to be prevalent there.

### 3. Methods

#### 3.1. Virus isolation

In September 2012 we analysed a semen sample from a 9-year-old Andalusian stallion which was imported to Gloucestershire, UK from Spain. Virus was isolated following the protocol described in chapter 2.5.10 of the OIE Terrestrial Manual 2013 (Anonymous,

2013). The resultant isolate was named GB.Glos.2012, reflecting the country, province and year of isolation.

#### 3.2. Molecular testing

RNA was extracted from the semen using the QIAamp Viral RNA Mini Kit (Qiagen Ltd., Manchester, UK) according to the manufacturer's instructions. A 564 nucleotide portion of the ORF5 gene was amplified using the CR2 and EAV32 primers described by Stadejek et al. (1999). In parallel, a 395 nucleotide portion of the ORF7 gene was amplified using the 14a and 15 primers described by Belak et al. (1994). Following purification, the PCR products were Sanger sequenced using the same primer pair on ABI PRISM 3130xl Genetic Analyser machine (Applied Biosystems Inc.).

For phylogenetic analysis sequences which represent all described EAV genetic groups were downloaded from GenBank and compared to the sequence (GB.Glos.2012) generated in this study. The sequences were aligned and maximum likelihood and neighbour joining phylogenies constructed using the MEGA5 programme (Tamura et al., 2011).

**Table 2**

The percentage sequence homology between GB.Glos.2012 and the other selected full genomes.

		Percent Identity					
		1	2	3	4	5	
Divergence	1	■	85.5	84.4	84.7	84.6	1 GB_Glos_2012
	2	16.5	■	85.6	86.0	87.2	2 NC_002532_Bucyrus
	3	18.0	16.4	■	87.0	84.8	3 AY349168_CW01_USA
	4	17.6	15.9	14.6	■	84.9	4 JN211316_F27_France
	5	17.5	14.1	17.1	17.0	■	5 GQ903808_S4445_USA
		1	2	3	4	5	

The GB.Glos.2012 isolate was also used as the starting material for next generation sequencing via the Roche 454 FLX platform following protocols described by Rasmussen et al. (2008, 2010).

#### 4. Results

The semen sample tested positive for EAV both by virus isolation and PCRs targeting ORF5 and ORF7.

##### 4.1. Comparison of GB.Glos.2012 isolate to other full length sequences

It was possible to recover the full genome of GB.Glos.2012 (GenBank: LC000003) via NGS. All of the full genome sequences available on GenBank ( $n=32$ , excluding molecular clones) were compared phylogenetically. As expected comparing these genomes did not elucidate any details regarding origins as the limited geographic distribution of the available data hinders any meaningful discovery. However, GB.Glos.2012 fell outside all existing full genome sequences and by using the existing clusters it was possible to take a reference sequence from each to compare genome features against GB.Glos.2012 (Fig. 2). The GB.Glos.2012 isolate is 12,702 nucleotides in length, slightly shorter than Bucyrus (12,704 nucleotides), CW01 (12,708 nucleotides), F27 (12,710 nucleotides) and S4445 (12,731 nucleotides). The isolate shares greatest homology, 85.5% at the nucleotide level with Bucyrus and 84.4%, 84.7% and 84.6% and homology with CW01, F27 and S4445 respectively (Table 2). Analysing homology at the amino acid level for specific ORFs least similarity was noted across ORF3 at 78.5% with F27 and greatest similarity was across ORF7 at 99.1% with Bucyrus, CW01 and S4445. GB.Glos.2012 had a three nucleotide deletion in nsp2 (codon 459 of protein 1ab) apart from which no other indels were noted.

After a thorough analysis of GenBank entries, phylogenies were initially constructed using 211 sequences. Sequences were selected to cover greatest geographic and time differences, thus molecular clones, Bucyrus derivatives and laboratory strains were not selected for the analysis. Reference sequences from each of the manuscripts cited in Fig. 3 were used to inform clusters. Phylogenies were constructed in both neighbour-joining and maximum likelihood methods, which made little difference to topology (data not shown). For ease of viewing Fig. 3 contains an abridged version of the phylogenetic tree containing 101 sequences on display.

Analysis of the GB.Glos.2012 ORF5 PCR product sequence (Fig. 3) revealed the isolate to be an outlier, only clustering with isolates KY63 (GenBank: U81014), H10 (GenBank: AY453306) and H21 (GenBank: AY453306) supported by bootstrap values of 70% and 99% respectively. GB.Glos.2012 shared 87%, 94.6% and 95% nucleotide homology across ORF5 GL segment with KY63, H10 and H21 respectively. The next closest neighbours were strains isolated in South Africa (GenBank: AY453337 and AY4533379) which were

thought to be of Yugoslavian origin and Hungarian strains (GenBank: JN314862 and JN314866).

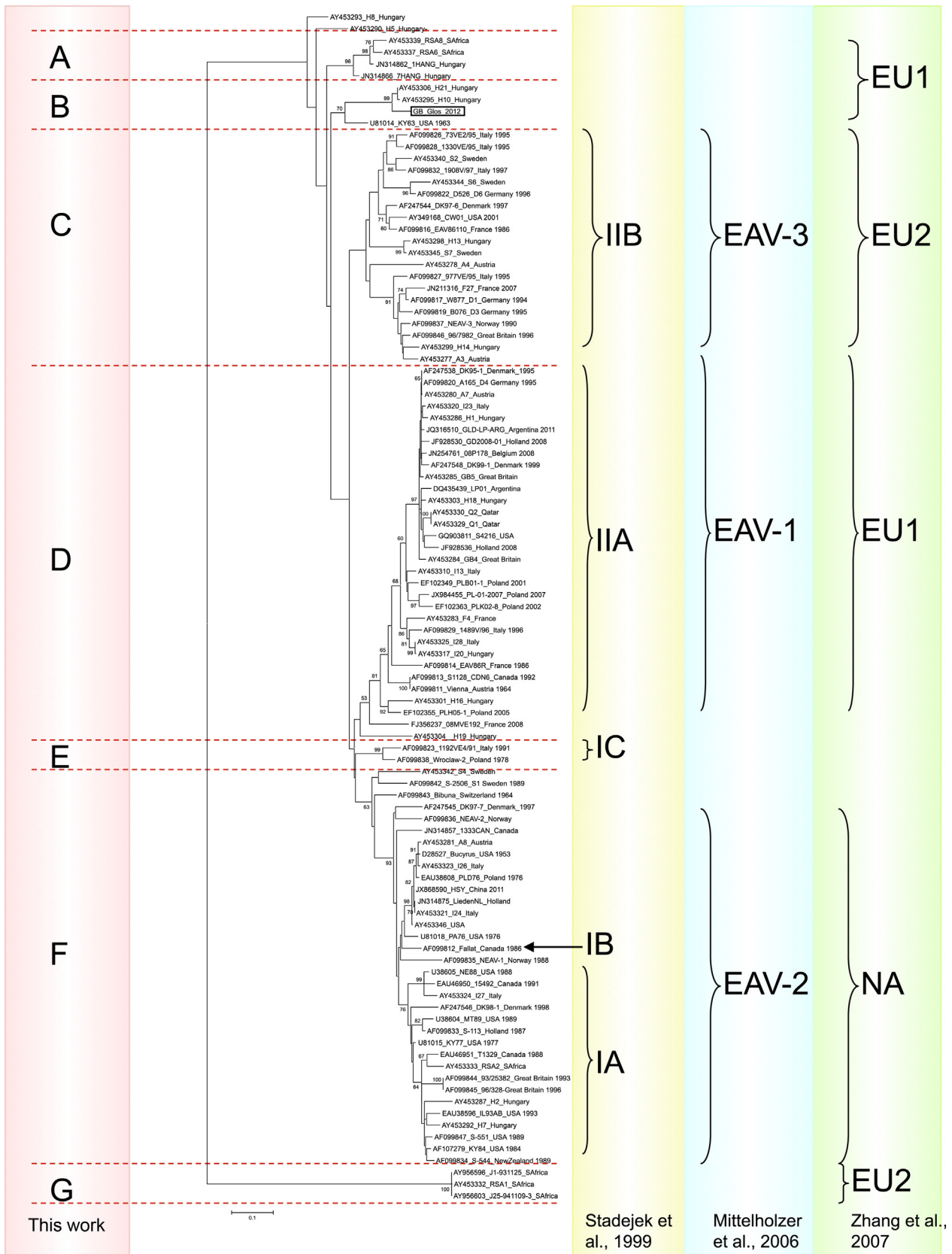
Following tree construction 7 clusters alphabetically labelled A–G, became apparent (Fig. 3). In our hands three new clusters A, B and G were visible, which were supported by bootstrap values of 96%, 70% and 100% respectively. Clusters A and B have previously been included in group EU1 and cluster G within group EU2 (Zhang et al., 2007). However taking the bootstrap values and the within between group variation (shown in Table 3) into account separate clusters appear warranted.

There were three clusters, C, D, and F in which the majority of isolates segregated. The existence of one cluster, C, however, was not supported via bootstrap analysis, and regardless of the phylogenetic scheme used it has the greatest within group variation with 9.9%/10.10%/9.9% and 13.41% respectively (Table 3). This, along with the lack of bootstrap support suggests this cluster could be further broken down. Comparing the previously published phylogenies the IA and IB groupings cited in Stadejek et al. (1999) were not sustainable with current information and thus merged into one group as already suggested by Mittelholzer et al. (2006) and Zhang et al. (2007), here labelled group F in Fig. 3. Regardless of the phylogenetic scheme used there are isolates which did not fit into any of the defined clusters (Fig. 3): the use of full genome data and/or further sampling may aid to resolve this, either by generating further clusters or including the viruses into existing ones.

#### 5. Discussion

Reviewing the relation of GB.Glos.2012 with other EAV strains, we focused our phylogeny on ORF5, as did previous authors (Balasuriya et al., 1995; Stadejek et al., 1999; Hornyak et al., 2005; Mittelholzer et al., 2006; Metz et al., 2011). Whilst it would have been interesting to compare further gene segments appropriate geographic and time point distribution were observed across ORF5 sequences only. As seen from Fig. 2 limited information could be gleaned from phylogeny of full genome data as these data are limited in number and location. It is anticipated that with the development of full genome sequencing into a routine technology, more viruses from all currently existing clades will be added soon to allow for a wider comparison. With regards to GB.Glos.2012 it would be particularly interesting to compare it against the South African and Hungarian strains with which the isolate clustered as here we were limited to the 518 nucleotides of the ORF5 available for the other strains. As with PRRSV, variability of nsp2 is also noted for EAV. We reported a three nucleotide deletion at codon 459 of protein 1ab and comparison with other full genome sequences highlighted numerous alterations in this region e.g. strain S4445 contains a 5 amino acid insertion between codons 448 and 489 of protein 1ab. As with the phylogeny, we were again limited in our ability to collate such observations as the publicly available sequence data for the nsp2 region is restricted to data from the USA or France with the majority of these sequences being linked to a single outbreak in each of these countries.

The re-appearance of a rare genotype of EAV provided us with reason to re-visit the relationship of EAV strains and their nomenclature. GB.Glos.2012 demonstrates the re-emergence of a genotype across decades and around the globe. KY63 is a field strain isolated in Kentucky, USA in 1963; this isolate has always clustered separately from the other US isolates of the same time period suggesting a different origin. Strains H10 and H21 appear to have been isolated in the early 2000s in Hungary and the “South African” strains were isolated in the mid-1990s with a history of Eastern European imports, whereas the isolated characterised here (GB.Glos.2012) is from a Spanish stallion. Even considering a potential lack of sampling the case demonstrates that established



**Fig. 3.** Phylogenetic analysis of ORF5. Maximum-likelihood tree using GTR+G+I model with 101 EAV sequences shown. Phylogenetic grouping as suggested by [Stadejek et al. \(1999\)](#), [Mittelholzer et al. \(2006\)](#) and [Zhang et al. \(2007\)](#) are labelled on the right hand side of the tree. Phylogenetic scheme suggested from this work is displayed on the left of the tree. Sequences are listed by their accession numbers followed by strain name, country of isolation and where available year of isolation. Percentage of a 1000 bootstrap replicates are cited at nodes for values  $\geq 50\%$ .

**Table 3a**

The average percentage pairwise nucleotide distances within clades.

Stadejek et al. (1999)		Mittelholzer et al. (2006)		Zhang et al. (2007)		This work	
IA	7.04%	EAV-1	6.93%	NA	8.23%	A	4.76%
IB	NA	EAV-2	8.25%	EU1	9.87%	B	7.85%
IC	5.98%	EAV-3	9.90%	EU2	13.41%	C	9.9%
IIA	6.97%					D	6.93%
IIB	10.10%					E	5.6%
						F	8.96%
						G	0.13%

**Table 3b**

The average percentage pairwise nucleotide distances between clades.

Stadejek et al. (1999)			Mittelholzer et al. (2006)			Zhang et al. (2007)			This work		
G1	G2	Dist	G1	G2	Dist	G1	G2	Dist	G1	G2	Dist
IA	IB	9.4	EAV-1	EAV-3	14.35	NA	EU1	15.84	B	A	15.44
IA	IC	15.34	EAV-2	EAV-1	15.48	NA	EU2	17.84	B	C	15.44
IA	IIB	16.77	EAV-2	EAV-3	16.44	EU1	EU2	16.2	B	E	15.37
IC	IB	14.29							B	G	26.93
IIA	IA	15.64							C	A	16.08
IIA	IIB	14.56							C	E	13.34
IIA	IB	14.32							C	G	25.18
IIA	IC	14.15							D	A	15.46
IIB	IB	16.25							D	B	14.93
IIB	IC	13.58							D	C	14.35
									D	E	13.95
									D	G	26.60
									E	A	16.31
									E	G	27.80
									F	A	16.88
									F	B	17.43
									F	C	16.26
									F	D	15.37
									F	E	14.32
									F	G	27.06
									G	A	26.45

NA: not applicable; G1: group 1; G2 group 2; Dist: distance expressed as a percentage.

strains do not become extinct and through global trade may re-appear unexpectedly. This is surprising since Arteriviruses, such as EAV or PRRSV are known for their rapid evolution, linked to the lack of a proofreading enzyme such as the exoribonuclease ExoN, which Coronaviruses for example contain and which is vital for their genetic integrity (Minskaia et al., 2006; Nga et al., 2011).

With viruses being subjected to host pressure the general notion was that they are constantly changing, perhaps as is reflected in the phylogenetic trees for EAV and PRRSV getting ever more complicated and seemingly highlighting new clades and genotypes. However, this case as well as the recent analysis of PRRSV and Bovine viral diarrhoea virus (BVDV; another persistent RNA virus of ruminants) in the United Kingdom demonstrates the persistence of strain variations in animal populations over decades (Frossard et al., 2013; Strong et al., 2013). This significantly broadens our understanding of the evolution of such viruses: in the absence of appropriate control measure to eradicate certain viruses (strains), host pressure will lead to a continuous divergence of viruses, but with the parallel persistence of “old” variants.

If phylogenies are to be used for epidemiological tracing further data points to fill in sampling gaps need to be included. That will require a more active sampling and control policy than currently established in most countries. The fact that this case (as some recent cases) was detected in a horse of some economic value, which was supposed to be used for breeding purposes, emphasizes the need for further vigilance, in particular in relation to the health horse concept advocated by the OIE.

Additionally, while ORF5 represents a good starting point with its gene product GP5 being directly under host-pressure from the antibody response other genes are important for the pathogenicity too and recombination has been described for Arteriviruses, thus

full genome data certainly represents the state of the art for the description of index cases in outbreaks. The available data needs to be labelled appropriately with the sequence record containing the location and date of isolation: as mentioned above there are multiple sequences at present in GenBank without any such information. Similarly manuscripts citing sequences should state accession numbers and particularly if presenting novel sequences provide details in the manuscript.

Following the emergence of the Asian lineage of highly pathogenic H5N1 influenza, the WHO along with OIE and FAO created a working group to implement a unified nomenclature system (Anonymous, 2012). The system set out the minimum requirements to define a new clade including minimum bootstrap value after 1000 bootstrap replicates ( $\geq 60$ ) and the average percentage pairwise nucleotide distances between and within clades ( $>1.5\%$  and  $<1.5\%$ , respectively). A similar approach could be undertaken for EAV and thus aid in resolving the differences observed between phylogenies. Such a system would allow a standard set of reference sequences to be available representing each clade, which would in turn allow future users to accurately define their data, harmonising the approach for all users.

From the analysis conducted within, we propose a novel nomenclature system for EAV and identify 3 novel clades labelled A, B and G, which are supported via bootstrap analysis. As all such systems the approach is not static: sequences that currently remaining outside the clades may form new clusters or existing clusters expand to include such current outliers. Existing clusters may divide into two or more new clusters (not least depending on criteria to be defined further) or even that clusters collapse (cluster C does not have strong bootstrap support at present for example). Importantly, we think that previous approaches by Mittelholzer et al. (2006) and

Zhang et al. (2007) made reference to the nomenclature for PRRSV, which is confusing. The numerical nomenclature system (e.g. EAV-1) conveys the impression of subtypes, sub-species or alike. For EAV no biological criteria have been identified to define such and serologically there is only one group. However, the strains identified in South Africa (Fig. 3, cluster G) are the furthest away from all other viruses; this highlights that a second serogroup or (sub)-species may arise in the future. Similarly reference to and EU or NA origin is not warranted for EAV. More than with PRRSV, EAV strains are spreading globally and the historic origins are not traceable: since clinical outbreaks of EAV appear only periodically, it is assumed that EAV circulated already long before its first isolation.

## Acknowledgement

The work was funded by the Department of Environment Food and Rural Affairs (Defra) (SV3300).

## References

- Anonymous, 2000. NAHMS in Equine Viral Arteritis and the US Horse Industry. USDA-APHIS VS CEAH, National Health Monitoring System, Fort Collins, CO.
- Anonymous, 2012. WHO/OIE/FAO H5N1 Evolution Working Group: continued evolution of highly pathogenic avian influenza A (H5N1): updated nomenclature. *Influenza Other Respir. Viruses* 6, 1–5.
- Anonymous, 2013. Equine viral arteritis. In: OIE Manual of Diagnostic Tests and Vaccines for Terrestrial Animals (Chapter 2.5.10).
- Balasuriya, U.B., Timoney, P.J., McCollum, W.H., et al., 1995. Phylogenetic analysis of open reading frame 5 of field isolates of equine arteritis virus and identification of conserved and nonconserved regions in the GL envelope glycoprotein. *Virology* 214, 690–697.
- Balasuriya, U.B., Patton, J.F., Rossitto, P.V., et al., 1997. Neutralisation determinants of laboratory strains and field isolates of equine arteritis virus: identification of four neutralisation sites in the amino-terminal ectodomain of the G(L) envelope glycoprotein. *Virology* 232, 114–128.
- Balasuriya, U.B., Zhang, J., Go, Y.Y., MacLachlan, N.J., 2014. Experiences with infectious cDNA clones of equine arteritis virus: lessons learned and insights gained. *Virology* 462–463, 388–403.
- Barrandeguy, M., 2010. EVA outbreak in Argentina. *Equine Dis. Q.* 19, 3–4.
- Belak, S., Ballagi-Pordany, A., Timoney, P., et al., 1994. Evaluation of a nested PCR assay for the detection of equine arteritis virus. In: Proceedings of the Seventh International Conference on Equine Conference on Equine Infectious Diseases, Tokyo, Japan. R&W Publications Newmarket, UK.
- Cavanagh, D., 1997. Nidovirales: a new order comprising Coronaviridae and Arteriviridae. *Arch. Virol.* 142, 629–633.
- Chirnside, E.D., Wearing, C.M., Binns, M.M., et al., 1994. Comparison of M and N gene sequences distinguishes variation amongst equine arteritis virus isolates. *J. Gen. Virol.* 75, 1491–1497.
- de Vries, A.A., Post, S.M., Raamsman, M.J., et al., 1995. The two major envelope proteins of equine arteritis virus associate into disulfide-linked heterodimers. *J. Virol.* 66, 6294–6303.
- Doll, E.R., Bryans, J.T., McCollum, W.H., et al., 1957. Isolation of a filterable agent causing arteritis of horses and abortions in mares. Its differentiation from equine abortion (influenza) virus. *Cornell Vet.* 47, 3–41.
- Echeverría, M.G., Pecoraro, M.R., Galosi, C.M., et al., 2003. The first isolation of equine arteritis virus in Argentina. *Rev. Sci. Technol.* 22 (3), 1029–1033.
- Echeverría, M.G., Díaz, S., Metz, G.E., et al., 2007. Genetic typing of equine arteritis virus isolates from Argentina. *Virus Genes* 35 (2), 313–320.
- Firth, A.E., Zevenhoven-Dobbe, J.C., Wills, N.M., et al., 2011. Discovery of a small arterivirus gene that overlaps the GP5 coding sequence and is important for virus production. *J. Gen. Virol.* 92, 1097–1106.
- Frossard, J.P., Hughes, G.J., Westcott, D.G., et al., 2013. Porcine reproductive and respiratory syndrome virus: genetic diversity of recent British isolates. *Vet. Microbiol.* 162, 507L 18.
- Glaser, A.L., de Vries, A.A., Dubovi, E.J., 1995. Comparison of equine arteritis virus isolates using neutralizing monoclonal antibodies and identification of sequence changes in GL associated with neutralization resistance. *J. Gen. Virol.* 76 (Pt 9), 2223–2233.
- Guthrie, A.J., Howell, P.G., Hedges, J.F., et al., 2003. Lateral transmission of equine arteritis virus among Lipizzaner stallions in South Africa. *Equine Vet. J.* 35 (6), 596–600.
- Hedges, J.F., Balasuriya, U.B., Topol, J.B., et al., 2001. Genetic variation of ORFs 3 and 4 of equine arteritis virus. *Adv. Exp. Med. Biol.* 494, 69–72.
- Hornyak, A., Bakonyi, T., Tejes, G., et al., 2005. A novel subgroup among genotypes of equine arteritis virus: genetic comparison of 40 strains. *J. Vet. Med. B* 52, 112–118.
- Lepage, N., St-Laurent, G., Carman, S., et al., 1996. Comparison of nucleic and amino acid sequences and phylogenetic analysis of the Gs protein of various equine arteritis virus isolates. *Virus Genes* 13 (1), 87–91.
- Manolaridis, I., Gaudin, C., Posthuma, C.C., et al., 2011. Structure and genetic analysis of the arterivirus non-structural protein 7alpha. *J. Virol.* 85, 7449–7453.
- McFadden 1, A.M., Pearce, P.V., Orr, D., et al., 2013. Evidence for absence of equine arteritis virus in the horse population of New Zealand. *N. Z. Vet. J.* 61 (5), 300–304.
- Metz, G.E., Ocampos, G.P.M., Serena, M.S., et al., 2011. Extended phylogeny of the equine arteritis virus sequences including South American sequences. *Intervirology* 54, 30–36.
- Minskaia, E., Hertzog, T., Gorbalenya, A.E., Campanacci, V., Cambillau, C., Canard, B., Ziebuhr, J., 2006. Discovery of an RNA virus 3' → 5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5108–5113.
- Mittelholzer, C., Stadejek, T., Johansson, I., et al., 2006. Extended phylogeny of equine arteritis virus: division into new subgroups. *J. Vet. Med. B* 53, 55–58.
- Nga, P.T., Parquet Mdel, C., Lauber, C., Parida, M., Nabeshima, T., Yu, F., Thuy, N.T., Inoue, S., Ito, T., Okamoto, K., Ichinose, A., Snijder, E.J., Morita, K., Gorbalenya, A.E., 2011. Discovery of the first insect nidovirus, a missing evolutionary link in the emergence of the largest RNA virus genomes. *PLoS Pathog.* 7, e1002215.
- Paweska, J.T., Aitchison, H., Chirnside, E.D., Barnard, B.J., 1996. Transmission of the South African asinine strain of equine arteritis virus (EAV) among horses and between donkeys and horses. *Onderstepoort J. Vet. Res.* 63, 189–196.
- Posthuma, C.C., Nedialkova, D.D., Zevenhoven-Dobbe, J.C., et al., 2006. Site-directed mutagenesis of the Nidovirus replicative endoribonuclease NendoU exerts pleiotropic effects on the arterivirus life cycle. *J. Virol.* 80, 1653–1661.
- Posthuma, C.C., Pedersen, K.W., Lu, Z., et al., 2008. Formation of the arterivirus replication/transcription complex: a key role for non-structural protein 3 in the remodelling of intracellular membranes. *J. Virol.* 82, 4480–4491.
- Powell, D.G., Timoney, P.J., 2006. EVA outbreak in New Mexico. *Equine Dis. Q.* 15, 2–3.
- Pronost, S., Pitel, P.H., Miszczak, F., et al., 2010. Description of the first recorded major occurrence of equine viral arteritis in France. *Equine Vet. J.* 42 (8), 713–720.
- Rasmussen, T.B., Reimann, L., Hoffmann, B., et al., 2008. Direct recovery of infectious pestivirus from a full-length RT-PCR amplicon. *J. Virol. Methods* 149 (2), 330–333.
- Rasmussen, T.B., Reimann, L., Uttenthal, A., et al., 2010. Generation of recombinant pestiviruses using a full-genome amplification strategy. *Vet. Microbiol.* 142 (1–2), 13–17.
- Snijder, E.J., Kikkert, M., 2013. Arteriviruses. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Lippincott Williams & Wilkins, Philadelphia, PA, pp. 859–879.
- Snijder, E.J., Meulenber, J.J., 1998. The molecular biology of arteriviruses. *J. Gen. Virol.* 79, 961–979.
- Snijder, E.J., Wassenaar, A.L., Spaan, W.J., 1994. Proteolytic processing of the replicase ORF1a protein of equine arteritis virus. *J. Virol.* 68, 5755–5764.
- Snijder, E.J., van Tol, H., Pedersen, K.W., et al., 1999. Identification of a novel structural protein of arteriviruses. *J. Virol.* 73, 6335–6345.
- Stadejek, T., Bjorklund, H., Ros Bascunana, C., et al., 1999. Genetic diversity of equine arteritis virus. *J. Gen. Virol.* 80, 691–699.
- Stadejek, T., Mittelholzer, C., Oleksiewicz, M.B., et al., 2006. Highly diverse type of equine arteritis virus (EAV) from the semen of a South African donkey: short communication. *Acta Vet. Hung.* 54 (2), 263–270.
- Strong, R., Errington, J., Cook, R., Ross-Smith, N., Wakeley, P., Steinbach, F., 2013. Increased phylogenetic diversity of bovine viral diarrhoea virus type 1 isolates in England and Wales since 2001. *Vet. Microbiol.* 162, 315–320.
- Tamura, K., Peterson, D., Peterson, N., et al., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739.
- Tijms, M.A., van der Meer, Y., Snijder, E.J., 2002. Nuclear localization of non-structural protein 1 and nucleocapsid protein of equine arteritis virus. *J. Gen. Virol.* 83, 795–800.
- Tijms, M.A., Nedialkova, D.D., Zevenhoven-Dobbe, J.C., et al., 2007. Arterivirus subgenomic mRNA synthesis and virion biogenesis depend on the multifunctional nsp1 autoprotease. *J. Virol.* 81, 10496–10505.
- Timoney, P.J., McCollum, W.H., 1993. Equine viral arteritis. *Vet. Clin. N. Am. Equine Pract.* 9, 295–309.
- van Dinten, L.C., den Boon, J.A., Wassenaar, A.L.M., et al., 1997. An infectious arterivirus cDNA clone: identification of a replicase point mutation that abolishes discontinuous mRNA transcription. *Proc. Natl. Acad. Sci. U. S. A.* 94, 991–996.
- van Kasteren, P.B., Bailey-Elkin, B.A., James, T.W., et al., 2013. Deubiquitinase function of arterivirus papain-like protease 2 suppresses the innate immune response in infected host cells. *Proc. Natl. Acad. Sci. U. S. A.* E838–E847.
- Wieringa, R., de Vries, A.A., Raamsman, M.J., et al., 2002. Characterisation of two new structural glycoproteins GP(3) and GP(4), of equine arteritis virus. *J. Virol.* 76, 10829–10840.
- Wood, J.L., Chirnside, E.D., Mumford, J.A., et al., 1995. First recorded outbreak of equine viral arteritis in the United Kingdom. *Vet. Rec.* 136 (15), 381–385.
- Zhang, J., Miszczak, F., Pronost, S., et al., 2007. Genetic variation and phylogenetic analysis of 22 French isolates of equine arteritis virus. *Arch. Virol.* 152 (11), 1977–1994.