Virology 537 (2019) 121–129

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

Molecular characterization of Equine Infectious Anemia Viruses using targeted sequence enrichment and next generation sequencing

Alexandre Deshiere[a], Nicolas Berthet[b,c], Fanny Lecouturier[a], Delphine Gaudaire[a], Aymeric Hans[a,*]

[a] ANSES- Laboratory for Animal Health in Normandy, Physiopathology and Epidemiology of Equine Diseases Unit, Goustranville, France
[b] Institut Pasteur, Unité Environnement et Risques Infectieux, Cellule d'Intervention Biologique d'Urgence, Paris, France
[c] Centre National de Recherche Scientifique (CNRS) UMR3569, Paris, France

ARTICLE INFO

Keywords:
Equine infectious anemia
Horse
Retrovirus
NGS
Phylogeny
Targeted enrichment system

ABSTRACT

Equine infectious anemia virus (EIAV) is responsible of acute disease episodes characterized by fever, anemia, thrombocytopenia and anorexia in equids. The high mutation rate in EIAV genome limited the number of full genome sequences availability. In the present study, we used the SureSelect target enrichment system with Illumina Next Generation Sequencing to characterize the proviral DNA of Equine Infectious Anemia Virus (EIAV) from asymptomatic horses. This approach allows a direct sequencing of the EIAV whole genome without cloning or amplification steps and we could obtain for the first time the complete genomic DNA sequences of French EIAV strains. We analyzed their phylogenetic relationship and genetic variability by comparison with 17 whole EIAV genome sequences from different parts of the world. The results obtained provide new insights into the molecular detection of EIAV and genetic diversity of European viral strains.

1. Introduction

Equine Infectious Anemia Virus (EIAV) is a macrophage-tropic lentivirus of the Retrovirus family, with an almost worldwide distribution. The disease is usually characterized by an acute phase during the first days following contamination and can evolve in a persistent infection characterized by recurring febrile episodes associating viremia, fever, thrombocytopenia, and wasting symptoms. Finally, the chronic phase is followed by a final asymptomatic stage associated with an immunologic control of virus replication (Leroux et al., 2004; Howe et al., 2005; Hammond et al., 2000; Cook et al., 2013). Nevertheless, asymptomatic carriers never eliminate the virus and remain a reservoir of infection for other horses (Issel et al., 1982). EIAV is transmitted mainly by infected blood through contaminated needles or haematophagous vectors such as horse flies and stable flies (Hawkins et al., 1973). Since there is a very low viral load in blood of infected animals, identification of EIA positive horses is essentially based on detection of antibodies directed against the p26 viral epitopes using the agar gel immunodiffusion assay (AGID) (Coggins and Norcross, 1970). The seroconversion of infected horses can take up to several months. Moreover, the stud farm where confirmation of EIA infection is declared is place under quarantine with a testing of all equids living in the vicinity of the positive horse until the risk of new asymptomatic cases is cleared out. Since the beginning of the 21st century, numerous outbreaks of EIA have been reported on all continents and have a financial impact on the equine industry (Issel and Cook, 1993; Issel and Foil, 2015; Bolfa et al., 2016). Therefore, the development of a rapid and efficient molecular diagnosis assay independent from seroconversion is of critical importance. EIAV genome is a single-stranded RNA of approximately 8200 nucleotides and is present in two identical copies in each viral particle. Like other lentiviruses, EIAV RNA genome is reverse transcribed into a cDNA copy following entry and integrates as a provirus into the genome of the host-cell (Liu et al., 2015). The major part of the sequence consists of 3 structural genes named Gag, Pol and Env (Fig. 1). The Gag (Group-specific antigen) gene encodes the precursor gag polyprotein, which is post-translationally maturated by the viral protease into 4 structural proteins: MA (p15 matrix protein), CA (p26 capsid protein), NC (p11 nucleocapsid protein) and P9 protein. The Pol (Polymerisation) gene sequence also corresponds to a polyprotein containing viral enzymes such as the p66 reverse transcriptase/ribonuclease H (RT/RNAseH), the p30 deoxyuridine-triphosphatase (dUTPase) and the p30 integrase. Finally, the Env (Envelope) gene codes for 2 proteins: the surface unit gp90 (SU) and the gp45 transmembrane (TM) glycoproteins that bind to the ELR1-TNFRSF14 receptor at the

* Corresponding author.
E-mail address: aymeric.hans@anses.fr (A. Hans).

https://doi.org/10.1016/j.virol.2019.08.016
Received 3 May 2019; Received in revised form 13 August 2019; Accepted 16 August 2019
Available online 22 August 2019

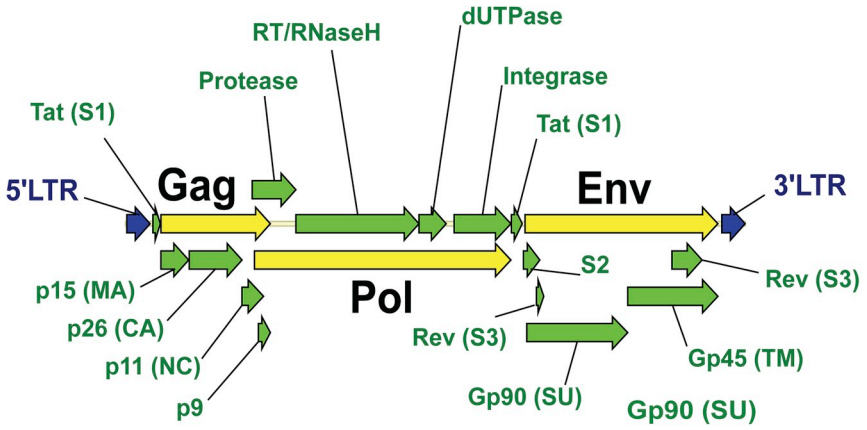

Fig. 1. Detailed map of the EIAV genome. The Gag (Group-specific antigen) gene encodes the precursor gag polyprotein containing 4 structural proteins: MA (p15 matrix protein), CA (p26 capsid protein), NC (p11 nucleocapsid protein) and the P9 protein. The Pol (Polymerase) gene codes for viral enzymes: a protease, a reverse transcriptase/ribonuclease H (RT/RNAse H), a deoxyuridine-triphosphatase (dUTPase) and the integrase. The Env (Envelope) region contains the sequence of the surface unit gp90 (SU) and the gp45 transmembrane (TM) glycoproteins. EIAV genome contains 3 regulatory peptides: S1 (Tat), S2 and S3 (Rev). The 8200 pb sequence is flanked in 5′ and 3′ positions by two identical long terminal repeats (LTR) containing promotor and regulatory elements.

Gag = Group-specific antigen	CA = Capsid	Gp = Glycoprotein	SU = Surface Unit
Pol = Polymerase	MA = Matrix	p = Protein	TM = Transmembrane
Env = Envelope	NC = Nucleocapsid	RT = Reverse Transcriptase	LTR = Long terminal Repeat

surface of macrophages (Zhang et al., 2005; Leroux et al., 2004). Moreover, EIAV encodes 3 regulatory peptides: S1, S2 and S3. Similar to other lentiviruses, S1 (Tat) binds the 5′-LTR to activate the transcription and S3 (Rev) binds to Rev responsive element (RRE) in the Env gene to control the transfer of non-spliced and partially spliced RNA from nucleus to cytoplasm (Cook et al., 2013; Ball et al., 1988). The S2 gene is unique in EIAV and its precise role in viral cycle remains unclear. However, S2 has a cytoplasmic expression and is not required for infection in horses but seems to increase virulence (Cook et al., 1998; Yoon et al., 2000; Covaleda et al., 2010; Chande et al., 2016). The complete sequence of the integrated virus consists of, the essential genes gag, pol, and env and the regulatory genes S1 (Tat), S2 and S3 (Rev) and is flanked by two identical long terminal repeats (LTR) of approximately 330 nt in 5′ en 3′ positions. Previous studies based on the EIAV Gag gene sequence have demonstrated that EIAV phylogeny has a more complex distribution than the classical model describing 4 monophyletic groups (Europe, USA, China and Japan) (Gaudaire et al., 2018; Leroux et al., 2004; Cook et al., 2013). We recently started surveillance and epidemiology studies on EIAV outbreaks in France. In one of these studies, we identified two French EIAV asymptomatic cases, designated as EIAV-FR-15 and EIAV-FR-16, respectively. Importantly, these strains could not be detected using a commercial real-time PCR kit targeting the EIAV pol gene, but both were successfully characterized using the AGID test and ELISA assays (unpublished data). Molecular detection of EIAV by conventional Sanger sequencing depend on sensitive PCR techniques since genomic DNA extracted from EIAV-infected animals contains a large proportion of the host cell genome. Moreover, several cloning steps are also required in order to amplify these partial sequences and eliminate equine nucleic acid sequences (Cook et al., 1998; Cappelli et al., 2011; Dong et al., 2012; Tu et al., 2007). Another approach has been described to amplify viral sequences using cell lines susceptible to EIAV such as equine dermal fibroblastic cells (ED cells). Nevertheless, after several passages, the virus is adapted to these cells, leading to a significant bias in the phylogenetic analysis with these amplified genome sequences (Maury et al., 1998, 2005). In order to selectively capture targeted regions of genomic DNA, Agilent Technologies (Santa Clara, CA) and the Broad Institute of MIT and Harvard have developed the SureSelect target enrichment method (Agilent SureSelectXT) using custom-designed streptavidin bead-bound biotinylated RNA probes (baits) to enrich exomes for variant calling by deep sequencing (Gnirke et al., 2009). DNA libraries are generated from total genomic DNA and sequences of interest are then captured and amplified to be determined by next generation sequencing (NGS). The applications of this method have been recently extended to the enrichment of viral sequences (Depledge et al., 2011; Brown et al., 2016; Li et al., 2017; Dao et al., 2018). Here we used this approach to characterize the whole-genome sequence of the proviral genome from two EIAV strains of French origin. This is the first study to report the use of nucleic acid capture associated with NGS for the detection of EIAV and to provide a rapid sensitive method allowing full-length genome sequencing of this virus. Direct sequencing of enriched EIAV genomes strongly limits technical bias thanks to the absence of cloning steps or the use of sequence specific primers and can then be applied to all EIAV variants, including from asymptomatic horses. These results provide new insight into the genetic variability of European EIAV strains, for which obtaining complete genome sequences is of great interest for molecular diagnosis. Finally, the French EIAV sequences obtained were compare with those obtained from different regions of the world, reinforcing the complex distribution model of the phylogenetic diversity of EIAV strains.

2. Materials and methods

2.1. Collection of experimental samples and extraction of genomic DNA

EIA-positive horses were identified using the EIAV antibody AGID test kit (VMRD Inc., Leipzig, Germany). Spleens from two EIA-positive horses were collected and stored at - 80 °C. Approximately 1 g of spleen tissue was grinded into Gibco minimum essential medium (Thermo Fisher Scientific, Waltham, MA, USA) and treated as described by Hans et al. (2015) (Hans et al., 2015). Genomic DNA was then purified from 300 µL of homogenate extract supernatant using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. Genomic DNA extracted from spleen was then kept at - 80 °C until use. DNA concentration was assessed using using NanoDrop spectrophotometer (Thermo Fisher Scientific) and Qubit 2.0 with the dsDNA BR Assay Kit (Life Technologies, Carlsbad, CA, USA).

2.2. EIAV whole-genome library preparation

DNA libraries were prepared using the SureSelect XT Library preparation kit for Illumina Multiplexed sequencing Kit (Agilent Technologies, Santa Clara, CA, USA). 3 µg of each sample was sheared with a Covaris instrument (Covaris, Woburn, MA, USA) and processed for end repair and addition of adaptors. The adaptor-tagged library fragments were purified using Agencourt AMPure XP beads (Beckman Coulter Inc., Brea, CA, USA) according to the manual. Next, each purified adaptor-tagged library was PCR amplified with the Herculase II

Fusion DNA Polymerase (Agilent Technologies) and captured with 1 μL of the SureSelect target enrichment custom-designed biotinylated probes (Agilent reference ID: 3115791) specific to all known EIAV full-length genomes (Genbank accession numbers: AF247394; M87581; AF016316; AB008196; JX480631; JX480632; JX480633; JX480634; KM247554; KM247555; AF327877; JX003263) and Dynabeads MyOne Streptavidin T1 (Life Technologies). Captured libraries were then PCR-amplified (16 cycles) with pre-capture primers targeting adapter sequences. After purification with AMPure XP beads, a second round of capture was performed followed by PCR-amplification (16 cycles) with post-capture and indexing primers targeting 5′- and 3- linked adapters and allowing to barcode the reads corresponding to the different samples (AGCAGGAA and GAGCTGAA for EIAV-FR-15 and EIAV-FR-16, respectively). The quality and quantity of amplified libraries were assessed using the Agilent 2200 Tapestation (Agilent Technologies) and the Qubit dsDNA HS Assay Kit (Life Technologies). Indexed library fragments had an average size of 350 bp. Indexed libraries were pooled at equimolar concentrations (12 picoMolar) with PhiX sequencing control V3 (Illumina, San Diego, CA, USA) and subjected to multiplex deep sequencing on the MiSeq platform using the MiSeq Reagent Kit V2 (Illumina).

2.3. NGS sequence data analysis

After multiplex deep sequencing on the MiSeq platform, we obtained a cluster density of 1126 ± 16 K/mm^2 with 84.5 ± 1.37% of clusters passing filters (CPF). Fastq format files were generated as raw data sequences by the Illumina MiSeq sequencer, with 2 x 150 cycles of paired-end read. A total of 12 samples were introduced in the pool, the reads corresponding to EIAV-FR-15 and EIAV-FR-16 corresponded to 8.16% and 6.79% of the total reads, respectively. Validation of data was performed by evaluation of the distribution of quality scores. Validated fastq files from each viral genome were trimmed and paired reads were merged using the Geneious Prime sequencing data analysis software (Biomatters LTD, Auckland, New Zealand) (Kearse et al., 2012). After removing duplicated reads, contigs were obtained by alignment to the reference genome EIAV F2 (GenBank accession number: JX480631.1), following the GeneiousReadMapper algorithm method (http://assets.geneious.com.s3-website-us-east-1.amazonaws.com). After 5 iterations with medium sensitivity, we obtained a HQ score (percentage of untrimmed bases that are high quality of 98.9 and 98.6% for EIAV-FR-15 and EIAV-FR-16, respectively).

2.4. Sanger sequencing

Whole EIAV 8.2 kb proviral genomes were amplified from genomic DNA with the PHUSION Taq polymerase using 3 sets of primers generating 3 overlapping fragments of approximately 4 kb. Primers were designed based on sequences obtained from NGS analysis. To ensure the purity of PCR amplicons, a second round of PCR was performed with 3 additional sets of internal primers (see Supplementary Fig. S1). Sense and antisense strands of PCR products were sequenced (Eurofins GATC Biotech, Constance, Germany). Sequences were assembled using MEGA 7.0. The obtained sequences of EIAV-FR-15 and EIAV-FR-16 are registered in GenBank under the accession numbers MK593463 and MK593462, respectively.

2.5. Phylogenetic analysis

Nucleotide and translated amino acid sequences were aligned in MEGA 7.0 alignment software with the Clustal W algorithm. Phylogenetic trees were constructed using the maximum-likelihood method. The statistical robustness and evolutionary distances were assessed by bootstrap resampling (1000 replicates). The evolutionary history and calculation of evolutionary distances was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model. The tree with the highest log likelihood (-36819.7216) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (2 categories (+G, parameter = 0.2693)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 19 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. All positions containing gaps and missing data were eliminated. There was a total of 7756 positions in the final dataset (Kumar et al., 2016; Kimura, 1980). Nucleotide and amino acid identities were calculated using the Sequence Identity And Similarity algorithm (SIAS: http://imed.med.ucm.es/Tools/sias.html) with the Gonnet PAM 250 matrix (Gonnet et al., 1992).

2.6. Semi-quantitative PCR

PCR amplification was conducted in triplicate starting with 250 ng of genomic DNA or 5 ng of first and second round captured libraries. Oligonucleotides used for the specific amplification were EIAV-FR-15-F (GGTGACGGTGCAAGGGTCTC), EIAV-FR-15-R (ATTGCCCACCATGTTCTTTC), EIAV-FR-16-F (CAATCCCAATGACAGCAAGA) and EIAV-FR-16-R (GGTTCTTTTGGCCCTTGTCT) for EIAV Gag gene and GAPDH-F (AGAAGGAGAAAGGCCCTCAG), GAPDH-R (GGAAACTGTGGAGGTCAGGA) for equine GAPDH gene. Measurements were performed with Fast SYBR Green PCR master mix (Applied Biosystems) and the Roche LightCycler 480 Instrument II (Roche Applied Science, Penzberg, Germany). Amplification of target genes was normalized to the equine GAPDH standard curve and expressed as relative expression using the absolute quantification fit point method.

3. Results

3.1. Validation of custom EAIV capture probes designed for EIAV by semi-quantitative PCR

In order to enrich the sample for sequencing of whole EIAV genomes, we designed custom biotinylated RNA probes specific to all known EIAV genomes. Fig. 2A describes the general flow chart used for enrichment of EIAV genome for EIAV sequencing. Genomic DNA samples were quantified with a Qubit fluorometer and randomly fragmented by Covaris sonication to generate DNA libraries. The resulting adapter-attached DNA libraries were hybridized to the SureSelectXT custom-designed capture probes. In order to maximize the enrichment of EIAV sequences, we performed two successive rounds of capture and amplification (capture#1 and capture#2). For the two samples, we assessed the quality of the enrichment using specific qPCR primers and SyBR green quantification for the GAPDH equine gene and the EIAV Gag gene. As shown in Fig. 2B, initial genomic DNA extracts harbored a rapid amplification of the GAPDH gene for both samples (15 cycles) whereas detection of the EIAV Gag gene was only observed after 35 to 38 cycles for isolates EIAV-FR-16 and EIAV-FR-15, respectively. After a first round of capture, the detection of both genes was observed at similar levels with an amplification starting at an average of 25 cycles. The second capture provided an even better enrichment of EIAV sequences with an earlier detection of the Gag gene after 10 to 12 cycles and a later detection of the GAPDH gene at 27 to 28 cycles. In order to quantify this enrichment, data were normalized to a standard curve of the GAPDH gene from an equimolar pool of both EIAV-FR-15 and EIAV-FR-16 samples (Fig. 2C). We observed a ratio of 1/10^7 to 1/10^6 for EIAV-FR-15 and EIAV-FR-16, respectively between expression of the GAPDH equine gene and the EIAV Gag gene in genomic DNA non-

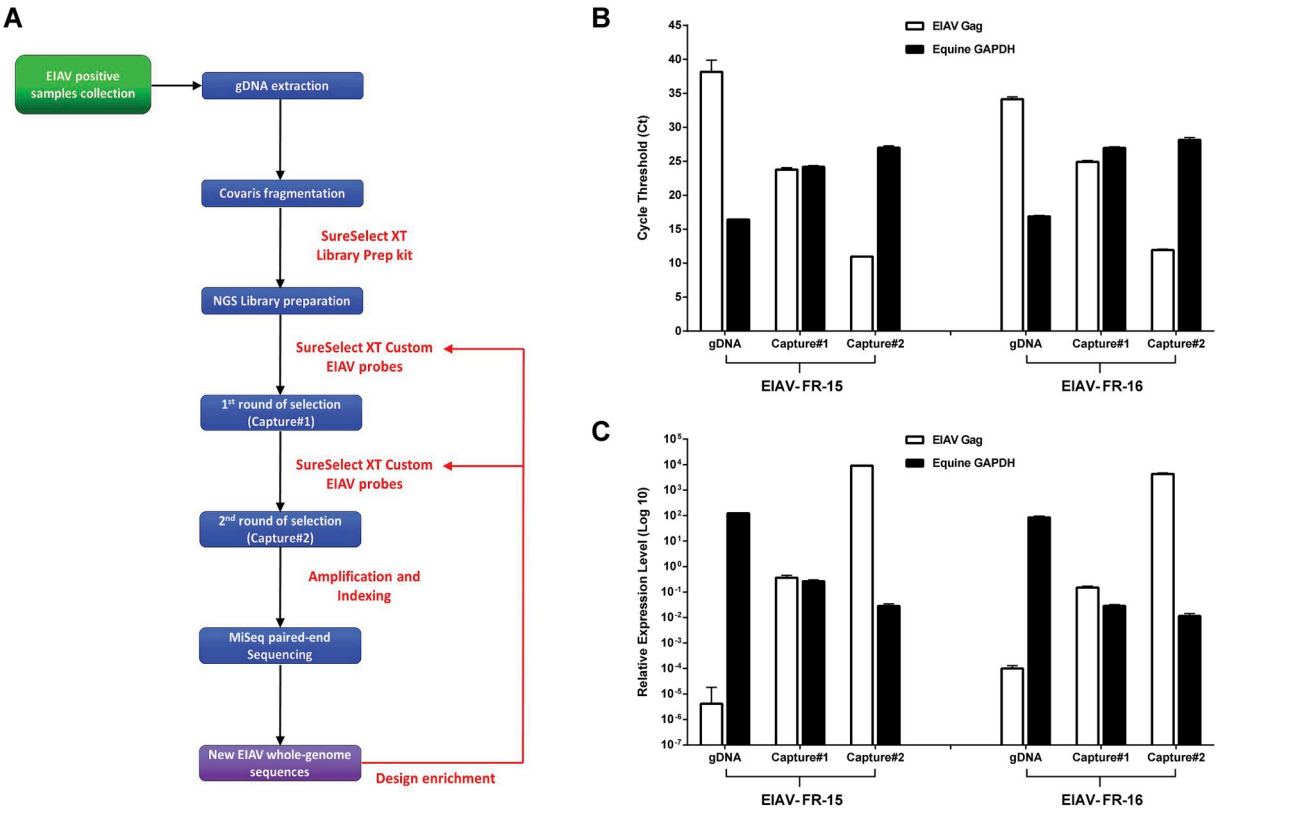

Fig. 2. Enrichment of EIAV proviral sequences from equine genomic DNA. (A) Flowchart of EIAV DNA capture. The equine genome containing the proviral sequence is sheared into approximately 300 bp fragments and processed for NGS library preparation. Indexed fragments are then captured into two rounds of enrichment using custom-designed biotinylated probes. Amplified enriched libraries are finally sequenced using the Illumina MiSeq platform. (B) Enrichment of EIAV sequences and depletion of equine sequences are evaluated using the cycle threshold (Ct) of real time PCR amplification targeting the GAPDH and Gag EIAV genes. (C) The relative expression level of each gene was quantified using a standard curve of the GAPDH gene amplification ranging from 0.32 pg to 250 ng of initial gDNA introduced into the reaction. Data are presented as the mean ± SD on a 10-fold logarithmic scale.

enriched samples, witnessing the low level of infected cells in spleen tissue extracts. After the first capture, we observed an increase of 10^3 to 10^5 fold for the EIAV Gag gene and a decrease of 10^3 fold of the GAPDH equine gene expression compared to initial samples, leading to a similar quantity of both equine and EIAV sequences. After the second capture, we could detect an inversion of the initial ratio with 10^5 to 10^6 more EIAV sequences compared to equine gene sequences. Taken together, these results demonstrate a successful increase of the ratio between EIAV sequences and equine genomic sequences using two rounds of capture with our custom-designed set of enrichment probes. Interestingly, we could observe that a second enrichment dramatically increased the coverage of the sequence after alignment to the reference genome.

3.2. Analysis of sequencing data

After demultiplexing and trimming of paired-end reads, alignment was performed to the reference genome EIAV F2 (GenBank accession number: JX480631). For the EIAV-FR-15 sample, 119 192 reads were mapped among a total of 2 791 162 reads (4.27%) with an average read depth of 1190.14 reads per nucleotide position (Fig. 3A). 17 708 reads remained after removing duplicates and covered 100% of the EIAV sequence. For the second sample (EIAV-FR-16), 7060 reads could be aligned among the 2 323 002 reads sequenced (0.31%) corresponding to 1876 unique reads of 150 bp, covering 92.2% of the EIAV whole genome. The read depth was lower for this sample with an average of 127.75 ranging from 0 to 599 reads per nucleotide position (Fig. 3B). Interestingly, the positions with a low coverage corresponding to the 7.8% gap in the sequence were essentially located in the hypervariable region of the Env gene harboring poor similarities with the reference sequences used for the design of the capture probe set. Similar results were obtained when using other European strains as reference sequences, as well as Chinese, Japanese and North American genomes (data not shown). To complete these gaps, the EIAV-FR-16 sample was fully re-sequenced by conventional Sanger sequencing. The EIAV-FR-15 sample was also verified to validate the sequence obtained by read alignment.

3.3. Whole-genome phylogenetic analysis

The validated sequences from EIAV-FR-15 and EIAV-FR-16 samples were aligned using MEGA 7.0 and the ClustalW algorithm with previously published EIAV whole genome sequences from Europe (F2, F3, F4, H3, DE, SA), China (Liaoning and vaccine strains LN1, FDDV-2, DLA, DLV2-6, DV3-5), Japan (Miyazaki) and North America (V26, V70, Wyo, UK). The V26 and V70 strains have been initially characterized as Japanese viruses but are genetically derived from the Wyoming strain (Zheng et al., 2000; Dong et al., 2013). We performed a maximum-likelihood phylogenetic analysis of these sequences represented in Fig. 4. Interestingly, both EIAV-FR-15 and EIAV-FR-16 sequences were not related and formed phylogenetic tree branches independent from other viral genomes. As shown in Supplemental Fig. S2, the full-length sequence nucleotide identity (N.Id.) between EIAV-FR-15 and EIAV-FR-16 was only 77.11%. Similar scores when obtained when comparing these 2 sequences with other strains from Europe, Japan, China, USA suggesting that they belong to 2 different phylogenic groups. A detailed analysis of protein-coding genes (Gag, Pol, Env, S1, S2 and S3) shows that the diversity within the nucleotide sequence is also reflected by in

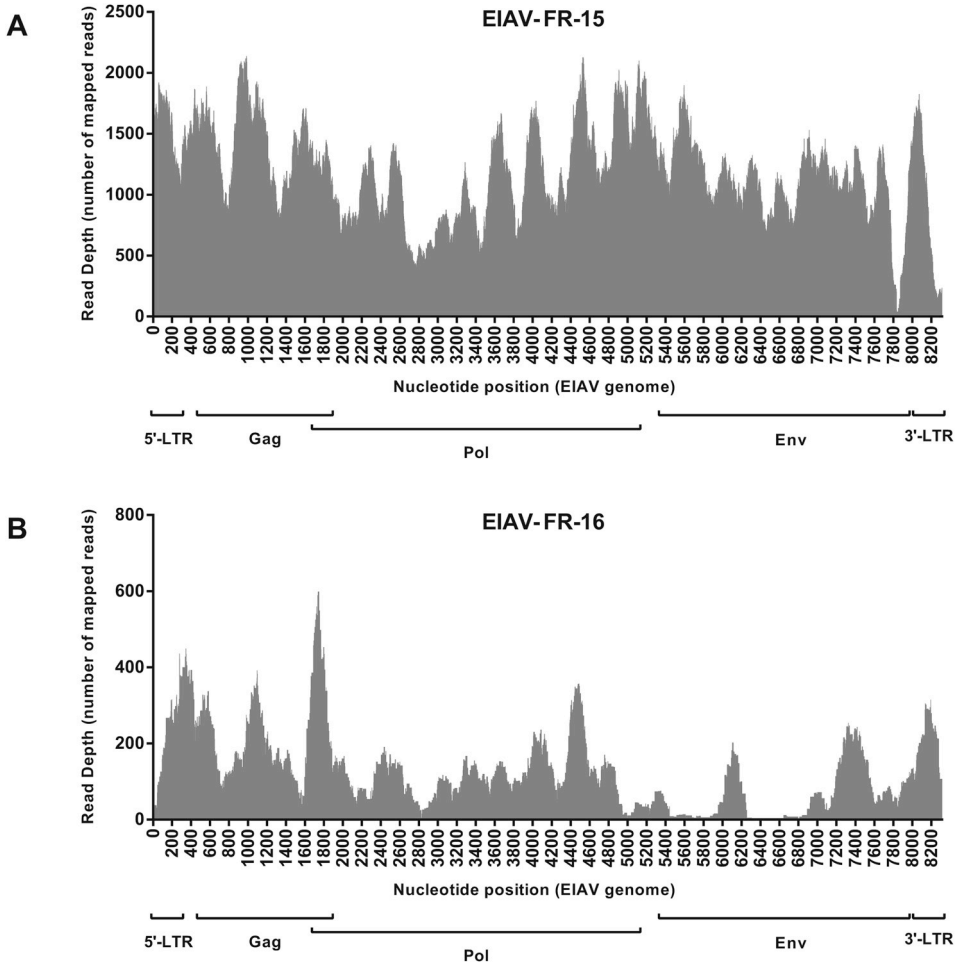

Fig. 3. Read depth diagram after Next Generation Sequencing of the enriched EIAV-FR-15. (A) and EIAV-FR-16 (B) sequences. Data were generated from the Geneious software after alignment to the EIAV F2 European reference genome (Genbank accession number: JX480631) which was performed using the GeneiousReadMapper algorithm.

the amino acid sequence, suggesting non-silent substitutions within the coding sequence. Considering the Gag and Pol sequences, the N.Id., as well as the amino acid identity (A.A.Id.) of EIAV-FR-15 and EIAV-FR-16 vary between 78% and 85% when compared to other sequences (Fig. 5 and Fig. S3). The lowest identity of both genomes was with the Miyasaki Japanese strain. Surprisingly, EAIV-FR-15 harbored similar N.Id. and AA.Id. Scores when compared to the Chinese strain (Liaoning), American strains (UK and Wyoming) and European strains (F2 and SA). For the Tat gene (S1), the identity scores of EIAV-FR-15 were even stronger with EIAV-Liao and EIAV-Wyo compared to EIAV-F2 and EIAV-SA. Those findings highlight the complexity of EIAV phylogeny which cannot be limited to geographical clusters. As expected, the highest diversity at both nucleotide and amino acid levels between all strains was found in the 3' region of the genome containing Env and S2 genes. Except for strains originating from the same outbreaks (EIAV-Wyo and EIAV-Uk vs EIAV-F2 and EIAV-SA), N.Id. Scores and A.A.Id. Scores ranged from 50% to 70%. The lowest conservation was found in the S2 gene (Rev) with A.A.Id. Scores of 39.5% and 38.3 comparing EIAV-Miya with EIAV-FR-15 and EIAV-FR-16, respectively.

3.4. Gag gene phylogenetic analysis

To assess the homology of these two French genomes to an extended number of EIAV strains, we also performed a phylogenetic analysis with available partial sequences of the Gag gene. As shown in Fig. 6, the EIAV-FR-15 Gag sequence harbored a strong homology with the viral strains of the 2009 French outbreak (Gaudaire et al., 2018). The N.Id. Score between EIAV-FR-15 and the 2009 phylogenetic group ranged between 96.76% for EIAV-FR-3 and 98.24% for EIAV-FR-12. These strains are related with the Romanian strain ROM-1 (GenBank #GQ229581) and with the Warsage strains (GenBank #JX193071 and #JX193072) the latest was isolated in Belgium from horses imported from Romania (Caij and Tignon, 2014) and harbored an average N.Id. Score of 89% with EIAV-FR-1 to EIAV-FR-15 sequences (Fig. S4). On the other hand, the EIAV-FR-16 sample was more related to two Italian strains described in 2011 by Felicetti et al.: EIAV-ITA-87 (GenBank #HQ888861) and EIAV-72 (GenBank #HM177436) (Felicetti et al., 2011). Using a comparison of nucleotide identity of the Gag partial sequence of EIAV-FR-16, we could find a N.Id. Score of 86.14% and 83.56% with EIAV-ITA-87 and EIAV-72, respectively. These findings indicate that despite the two viruses from the 2014 French outbreak occurred at the same period and within the same geographical area, they are not related, and might emerge from different sources of contamination.

4. Discussion

In 2014, two asymptomatic cases of EIAV infection were reported in the south of France and were confirmed by the European reference laboratory for EIA. The AGID test is based on the detection of equine antibodies directed against EIAV, and requires a seroconversion that can take several months. In most situations, horses seroconvert within

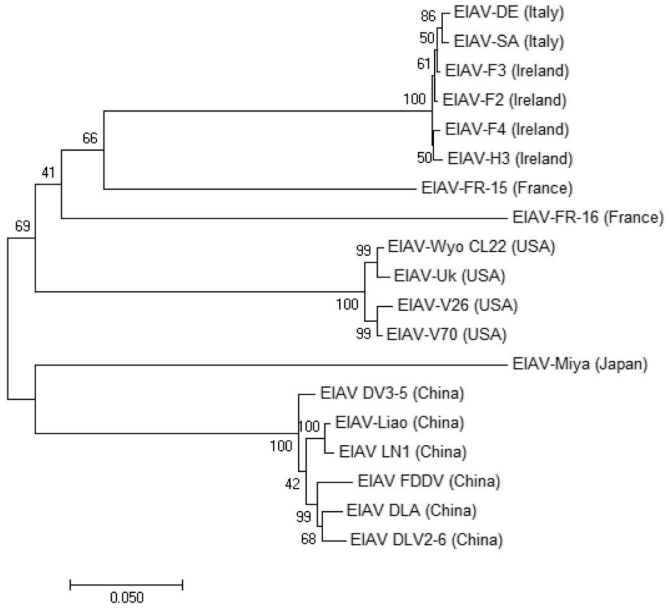

Fig. 4. Phylogenetic analysis of the whole-genome sequences. The analysis was conducted with MEGA 7 and involved 19 nucleotide sequences inferred with the Maximum Likelihood method based on the Kimura 2-parameter model. The percentage of trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The tree with the highest log likelihood (-36819.7216) is shown.

45 days of exposure with the virus but detection of antibodies can be observed up to 157 days after infection (Issel and Cook, 1993; Cullinane et al., 2007; Ricotti et al., 2016). The viruses associated with the 2014 French outbreak were characterized using an in-house semi-nested PCR assay but the full-length genome of the provirus could not be sequenced due to several technical limitations and a low viral loads in those asymptomatic horses. Indeed, the high mutation rate of this virus strongly interferes with the possibility of finding universal primers and genomic DNA extracted from infected horses provides a very low rate of EIAV sequences compared to the host cell genome, particularly in asymptomatic cases. Thus, several rounds of PCR with different sets of primers are required to characterize the molecular features of each infections with a limited success (Dong et al., 2012, 2013; Cappelli et al., 2011). In consequence, this is of great interest to develop molecular assays that can efficiently detect EIAV strains shortly after viral transmission and allow phylogenetic tracing of the different outbreaks. In this study, we used custom-designed capture probes to efficiently enrich EIAV proviral sequences from equine genomic DNA and successfully sequenced the whole genome of the 2014 EIAV strains by NGS. This method does not require the use of specific primers, and thus can be extended to all strains of EIAV independently of the presence of known conserved sequences. Moreover, the enrichment of EIAV sequences with the capture probes allows the detection of low amount of virus and can then be applied to asymptomatic horses. Using semi-quantitative PCR, we could achieve an enrichment of the EIAV Gag gene ranging from 10^3 to 10^5 fold after the first capture and from 10^7 to 10^9 fold after two rounds of capture. The full genome sequence of the EIAV-FR-15 isolate was obtained by sequencing the captured libraries with a coverage of 100% and a high read depth. For EIAV-FR-16 isolate, read alignment to a European reference genome provided 92.2% of the full-length sequence with gaps in the variable regions of the Env gene. This region is strongly divergent within the EIAV subtypes and displays low similarities with the reference sequences used for the design of the capture probe set (Craigo et al., 2009; Craigo and Montelaro, 2013; Leroux et al., 1997; Quinlivan et al., 2013). For the

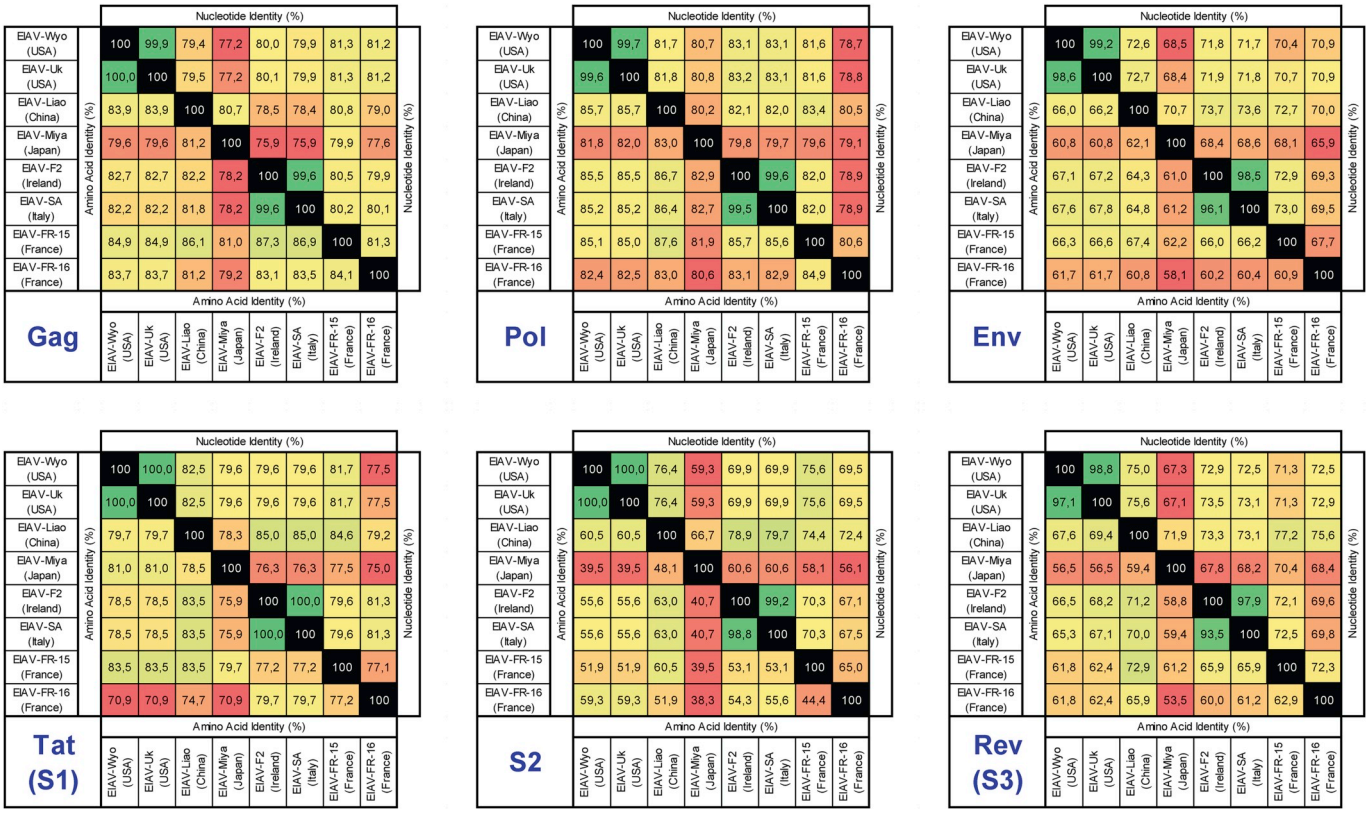

Fig. 5. Sequence identity between gene coding sequences from EIAV known genomes. Nucleotide and amino acid identities were calculated using the Sequence Identity And Similarity algorithm (SIAS) with the Gonnet PAM 250 matrix. Data are presented as percentage of identity. Complete datasets are listed in Supplemental Fig. S4 for each gene and protein.

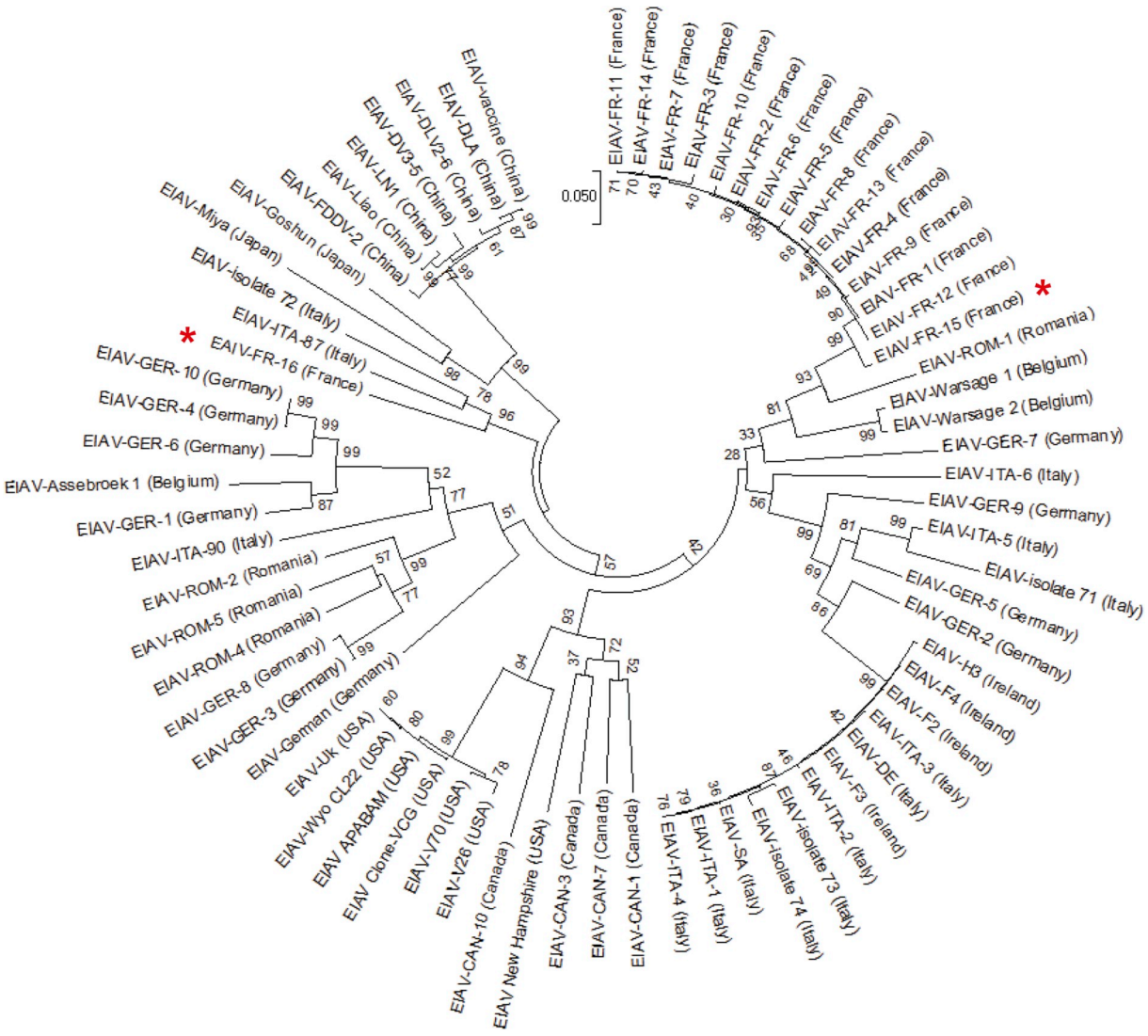

Fig. 6. Phylogenetic analysis of the gag gene sequence. The maximum likelihood tree was created using MEGA 7, involving 72 nucleotide sequences. The percentage of trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model. The tree with the highest log likelihood (-13527.3254) is shown. French EIAV sequences EIAV-FR-15 and EIAV-FR-16 are labelled with a red asterisk.

two samples, we compared the enrichment after a single or a double round of capture. For both EIAV-FR-15 and EIAV-FR-16, we could observed that the coverage significantly increased after the second enrichment due to the higher ratio between viral and host genome reads. Another approach consists in increasing the number or the temperature of washing steps, and/or reducing the probe hybridization time recommended by Agilent SureSelect protocol to increase the proportion of viral reads. We tested all these parameters and despite they actually increased the virus/host sequences ratio up to 95%, they were deleterious for the coverage of the hypervariable env region (data not shown). Since there is a low conservation of some sequences, and therefore a low homology with capture probes, covering the largest amount of the EIAV sequence seems to require a fine balance in the enrichment parameters to increase the virus/host sequences ratio without depleting low conserved regions. Moreover, we have observed that adding new reference sequences to the design of these probes drastically increased the coverage of such low conserved sequences (data not shown). Therefore, this system can be exponentially improved with the sequencing of new full-length genomes of EIAV strains from phylogenetically divergent outbreaks and origins. The two sequences were fully re-sequenced using conventional Sanger sequencing and the gaps in the EIAV-FR-16 Env gene were therefore fulfilled. A phylogenetic analysis of these new sequences was performed at the whole genome level with previously described strains from Europe, North America, China and Japan. We observed that both samples could not be related to any of the monophyletic groups and formed independent branches within the phylogenetic tree. This is consistent with previous reports suggesting that the monophyletic model for EIAV does not represent the vast diversity of viral strains from the different outbreaks (Craigo et al., 2009; Gaudaire et al., 2018). Since a larger number of sequences is available for the partial Gag gene sequence, we also conducted a comparative analysis of the homology between these two French genomes and available EIAV strains. Surprisingly, the two viruses seem to have emerged from the different origins. The EIAV-FR-15 Gag sequence appeared to be related with a strain isolated in 2009 in France (Gaudaire et al., 2018). These strains harbor a strong homology with the Romanian strain ROM-1 (GenBank #GQ229581) and with the Warsage strains (GenBank #JX193071 and #JX193072) the latest was isolated from Belgium equid that were actually imported from Romania (Caij and Tignon, 2014). On the other hand, the EIAV-FR-16 sequence displayed common features with the Italian strains EIAV-ITA-87 (GenBank #HQ888861) and EIAV-72 (GenBank #HM177436) isolated in

2011 (Felicetti et al., 2011). These results highlight the importance of phylogenetic tracing of all EIAV strains from common outbreaks since different pathways of contamination can occur even within a restricted time-lapse and area. Our findings provide new insight into the genetic variability of French EIAV outbreaks and reinforce the complex distribution model of the phylogenetic diversity of EIAV around the world. This is the first study to report the complete genome sequences of EIAV of French origin using custom-designed capture probes associated with deep sequencing appears to be highly sensitive method for detecting with a strong precision and coverage proviral sequences into the horse genome. Indeed, direct sequencing without enrichment leads to a very low number of viral reads, particularly in asymptomatic horses therefore limiting the number of samples from which a sequence can be obtained in a single experiment. Using the capture system, we could obtain a significant number of EIAV reads from all horses that were diagnosed positive and pool up to 16 samples in a single MiSeq sequencing run. Therefore, this approach appears to be cost-effective compared to non-enriched sequencing. Interestingly, new sequences obtained after each run can be included into the probe design, leading to a positive feedback loop exponentially increasing the diversity of viral strains that can be captured. Taken together, our data show that targeted sequence enrichment and next generation sequencing may constitute an innovative molecular tool for diagnosis of EIAV infected horses, including asymptomatic cases.

Conflicts of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be considered as a potential conflict of interest.

Aknowledgments

The authors would like to thank the high-throughput sequencing platform SeSAME (Sequencing for Health, Agronomy, the Sea and the Environment) of the Centre François Baclesse for the access to core facilities and advice with sample processing and analysis. We also thank official veterinary services for conducting the epidemiological investigations and tissue collection on infected horses. We are also grateful to Gaël Amelot for his technical assistance on the project as well as Arnaud Prehu, Sabrina Hauvel and Pascal Saussey for administrative support. Alexandre Deshiere and “CENTAURE project” was supported by a grant awarded by the Regional Council of Normandy and the French Ministry of Higher Education, within the framework of CPER 2015–2020 and FEDER/FSE 2014–2020. Study's financial support was obtained from the « Institut Français du cheval et de l’équitation (IFCE) », the European Reference Laboratory for Equine Diseases other than African Horse Sickness, Anses's own institutional resources. Anses-Laboratory for animal health in Normandy is a member of the GIS Centaure equine research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2019.08.016.

References

Ball, j. m., Payne, s. l., Issel, c. j., Montelaro, r. c., 1988. EIAV genomic organization: further characterization by sequencing of purified glycoproteins and cDNA. Virology 165, 601–605.
Bolfa, p., Barbuceanu, f., Leau, s. e., Leroux, c., 2016. Equine infectious anaemia in Europe: time to re-examine the efficacy of monitoring and control protocols? Equine Vet. J. 48, 140–142.
Brown, j. r., Roy, s., Ruis, c., Yara romero, e., Shah, d., Williams, r., Breuer, j., 2016. Norovirus whole-genome sequencing by SureSelect target enrichment: a robust and sensitive method. J. Clin. Microbiol. 54, 2530–2537.
Caij, a. b., Tignon, m., 2014. Epidemiology and genetic characterization of equine infectious anaemia virus strains isolated in Belgium in 2010. Transbound Emerg Dis 61, 464–468.
Cappelli, k., Capomaccio, s., Cook, f. r., Felicetti, m., Marenzoni, m. l., Coppola, g., Verini-supplizi, a., Coletti, m., Passamonti, f., 2011. Molecular detection, epidemiology, and genetic characterization of novel European field isolates of equine infectious anemia virus. J. Clin. Microbiol. 49, 27–33.
Chande, a., Cuccurullo, e. c., Rosa, a., Ziglio, s., Carpenter, s., Pizzato, m., 2016. S2 from equine infectious anemia virus is an infectivity factor which counteracts the retroviral inhibitors SERINC5 and SERINC3. Proc. Natl. Acad. Sci. U. S. A. 113, 13197–13202.
Coggins, l., Norcross, n. l., 1970. Immunodiffusion reaction in equine infectious anemia. Cornell Vet. 60, 330–335.
Cook, r. f., Leroux, c., Cook, s. j., Berger, s. l., Lichtenstein, d. l., Ghabrial, n. n., Montelaro, r. c., Issel, c. j., 1998. Development and characterization of an in vivo pathogenic molecular clone of equine infectious anemia virus. J. Virol. 72, 1383–1393.
Cook, r. f., Leroux, c., Issel, c. j., 2013. Equine infectious anemia and equine infectious anemia virus in 2013: a review. Vet. Microbiol. 167, 181–204.
Covaleda, l., Fuller, f. j., Payne, s. l., 2010. EIAV S2 enhances pro-inflammatory cytokine and chemokine response in infected macrophages. Virology 397, 217–223.
Craigo, j. k., Barnes, s., Zhang, b., Cook, s. j., Howe, l., Issel, c. j., Montelaro, r. c., 2009. An EIAV field isolate reveals much higher levels of subtype variability than currently reported for the equine lentivirus family. Retrovirology 6, 95.
Craigo, j. k., Montelaro, r. c., 2013. Lessons in AIDS vaccine development learned from studies of equine infectious, anemia virus infection and immunity. Viruses 5, 2963–2976.
Cullinane, A., Quinlivan, M., Nelly, M., Patterson, H., Kenna, R., Garvey, M., Gildea, S., Lyons, P., Flynn, M., Galvin, P., Neylon, M., Jankowska, K., 2007. Diagnosis of equine infectious anaemia during the 2006 outbreak in Ireland. Vet. Rec. 161, 647–652.
Dao, t.D., Bui, v.N., Omatsu, t., Katayama, y., Mizutani, t., Ogawa, h., Imai, k., 2018. Application of the SureSelect target enrichment system for next-generation sequencing to obtain the complete genome sequence of bovine leukemia virus. Arch. Virol. 163, 3155–3159.
Depledge, D.P., Palser, A.L., Watson, S.J., Lai, I.Y., Gray, E.R., Grant, P., Kanda, R.K., Leproust, E., Kellam, P., Breuer, J., 2011. Specific capture and whole-genome sequencing of viruses from clinical samples. PLoS One 6, e27805.
Dong, J.B., Zhu, W., Cook, F.R., Goto, Y., Horii, Y., Haga, T., 2012. Development of a nested PCR assay to detect equine infectious anemia proviral DNA from peripheral blood of naturally infected horses. Arch. Virol. 157, 2105–2111.
Dong, J.B., Zhu, W., Cook, F.R., Goto, Y., Horii, Y., Haga, T., 2013. Identification of a novel equine infectious anemia virus field strain isolated from feral horses in southern Japan. J. Gen. Virol. 94, 360–365.
Felicetti, M., Cappelli, K., Capomaccio, S., Coletti, M., Marenzoni, M.L., Passamonti, F., 2011. World-wide Equine Infectious Anemia Virus (EIAV) Gag Variability. (Unpublished).
Gaudaire, D., Lecouturier, F., Poncon, N., Morilland, E., Laugier, C., Zientara, S., Hans, A., 2018. Molecular characterization of equine infectious anaemia virus from a major outbreak in southeastern France. Transbound Emerg Dis 65, e7–e13.
Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., Leproust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D.B., lander, E.S., Nusbaum, C., 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat. Biotechnol. 27, 182–189.
Gonnet, G.H., Cohen, M.A., Benner, S.A., 1992. Exhaustive matching of the entire protein sequence database. Science 256, 1443–1445.
Hammond, S.A., Li, F., Mckeon, B.M., Sr, Cook, S.J., Issel, C.J., Montelaro, R.C., 2000. Immune responses and viral replication in long-term inapparent carrier ponies inoculated with equine infectious anemia virus. J. Virol. 74, 5968–5981.
Hans, A., Gaudaire, D., Manuguerra, J.C., Leon, A., Gessain, A., Laugier, C., Berthet, N., Zientara, S., 2015. Combination of an unbiased amplification method and a re-sequencing microarray for detecting and genotyping equine arteritis virus. J. Clin. Microbiol. 53, 287–291.
Hawkins, J.A., Adams, W.V., Cook, L., Wilson, B.H., Roth, E.E., 1973. Role of horse fly (Tabanus fuscicostatus Hine) and stable fly (Stomoxys calcitrans L.) in transmission of equine infectious anemia to ponies in Louisiana. Am. J. Vet. Res. 34, 1583–1586.
Howe, L., Craigo, J.K., Issel, C.J., Montelaro, R.C., 2005. Specificity of serum neutralizing antibodies induced by transient immune suppression of inapparent carrier ponies infected with a neutralization-resistant equine infectious anemia virus envelope strain. J. Gen. Virol. 86, 139–149.
Issel, C.J., Adams JR., W.V., Meek, L., Ochoa, R., 1982. Transmission of equine infectious anemia virus from horses without clinical signs of disease. J. Am. Vet. Med. Assoc. 180, 272–275.
Issel, C.J., Cook, R.F., 1993. A review of techniques for the serologic diagnosis of equine infectious anemia. J. Vet. Diagn. Investig. 5, 137–141.
Issel, C.J., Foil, L.D., 2015. Equine infectious anaemia and mechanical transmission: man and the wee beasties. Rev Sci Tech 34, 513–523.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Leroux, C., Cadore, J.L., Montelaro, R.C., 2004. Equine Infectious Anemia Virus (EIAV): what has HIV's country cousin got to tell us? Vet. Res. 35, 485–512.
Leroux, c., Issel, c.J., Montelaro, r.C., 1997. Novel and dynamic evolution of equine infectious anemia virus genomic quasispecies associated with sequential disease cycles

in an experimentally infected pony. J. Virol. 71, 9627–9639.
Li, T., Unger, E.R., Batra, D., Sheth, M., Steinau, M., Jasinski, J., Jones, J., Rajeevan, M.S., 2017. Universal human papillomavirus typing assay: whole-genome sequencing following target enrichment. J. Clin. Microbiol. 55, 811–823.
Liu, Q., Wang, X.F., Ma, J., He, X.J., Wang, X.J., Zhou, J.H., 2015. Characterization of equine infectious anemia virus integration in the horse genome. Viruses 7, 3241–3260.
Maury, W., Oaks, J.L., Bradley, S., 1998. Equine endothelial cells support productive infection of equine infectious anemia virus. J. Virol. 72, 9291–9297.
Maury, W., Thompson, R.J., Jones, Q., Bradley, S., Denke, T., Baccam, P., Smazik, M., Oaks, J.L., 2005. Evolution of the equine infectious anemia virus long terminal repeat during the alteration of cell tropism. J. Virol. 79, 5653–5664.
Quinlivan, M., Cook, F., Kenna, R., Callinan, J.J., Cullinane, A., 2013. Genetic characterization by composite sequence analysis of a new pathogenic field strain of equine infectious anemia virus from the 2006 outbreak in Ireland. J. Gen. Virol. 94, 612–622.
Ricotti, S., Garcia, M.I., Veaute, C., Bailat, A., Lucca, E., Cook, R.F., Cook, S.J., Soutullo, A., 2016. Serologically silent, occult equine infectious anemia virus (EIAV) infections in horses. Vet. Microbiol. 187, 41–49.
Tu, Y.B., Zhou, T., Yuan, X.F., Qiu, H.J., Xue, F., Sun, C.Q., Wang, L., Wu, D.L., Peng, J.M., Kong, X.G., Tong, G.Z., 2007. Long terminal repeats are not the sole determinants of virulence for equine infectious anemia virus. Arch. Virol. 152, 209–218.
Yoon, S., Kingsman, S.M., Kingsman, A.J., Wilson, S.A., Mitrophanous, K.A., 2000. Characterization of the equine infectious anaemia virus S2 protein. J. Gen. Virol. 81, 2189–2194.
Zhang, B., Jin, S., Jin, J., Li, F., Montelaro, R.C., 2005. A tumor necrosis factor receptor family protein serves as a cellular receptor for the macrophage-tropic equine lentivirus. Proc. Natl. Acad. Sci. U. S. A. 102, 9918–9923.
Zheng, Y.H., Sentsui, H., Kono, Y., Ikuta, K., 2000. Mutations occurring during serial passage of Japanese equine infectious anemia virus in primary horse macrophages. Virus Res. 68, 93–98.