

Application of Broad-Spectrum Resequencing Microarray for Genotyping Rhabdoviruses.

Laurent Dacheux, Nicolas Berthet, Gabriel Dissard, Edward C Holmes, Olivier Delmas, Florence Larrous, Ghislaine Guigon, Philip Dickinson, Ousmane Faye, Amadou A Sall, et al.

▶ **To cite this version:**

Laurent Dacheux, Nicolas Berthet, Gabriel Dissard, Edward C Holmes, Olivier Delmas, et al.. Application of Broad-Spectrum Resequencing Microarray for Genotyping Rhabdoviruses.. Journal of Virology, 2010, epub ahead of print. 10.1128/JVI.00771-10 . pasteur-00507428

HAL Id: pasteur-00507428

https://pasteur.hal.science/pasteur-00507428

Submitted on 30 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

JVI Accepts, published online ahead of print on 7 July 2010
J. Virol. doi:10.1128/JVI.00771-10
Copyright © 2010, American Society for Microbiology and/or the Listed Authors/Institutions. All Rights Reserved.

JVI Accepts published online ahead of print

Application of Broad-Spectrum Resequencing Microarray for Genotyping Rhabdoviruses

Running title: Rhabdovirus Identification by Resequencing Microarray

Laurent Dacheux[1‡], Nicolas Berthet[2‡], Gabriel Dissard[3], Edward C. Holmes[4], Olivier Delmas[1], Florence Larrous[1], Ghislaine Guigon[3], Philip Dickinson[5], Ousmane Faye[6], Amadou A. Sall[6], Iain G. Old[7], Katherine Kong[5], Giulia C. Kennedy[5], Jean-Claude Manuguerra[8], Stewart T. Cole[9], Valérie Caro[3], Antoine Gessain[2], and Hervé Bourhy[1*]

[1]Institut Pasteur, Lyssavirus dynamics and host adaptation Unit, Paris, France; [2]Institut Pasteur, Epidemiology and Pathophysiology Oncogenic Virus Unit, CNRS URA3015, Paris, France; [3]Institut Pasteur, Genotyping of Pathogens and Public Health Technological Platform, Paris, France; [4]Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, USA; [5]Affymetrix, Santa Clara, CA, USA [6]Institut Pasteur de Dakar, Arbovirology Laboratory, Dakar, Senegal; [7]Institut Pasteur, European Office, Paris, France; [8]Institut Pasteur, Laboratory for Urgent Responses to Biological Threats, Paris, France; [9]Institut Pasteur, Bacterial Molecular Genetics Unit, Paris, France

[‡] Had contributed in equal part in this work.

* Corresponding author: Hervé Bourhy, Unité Dynamique des lyssavirus et adaptation à l'hôte, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France, Phone: +33 1 45 68 87 85, Fax: +33 1 40 61 30 20, Email: herve.bourhy@pasteur.fr

Abstract word count: 217
Text word count: 5756

JVI Accepts published online ahead of print

ABSTRACT

The rapid and accurate identification of pathogens is critical in the control of infectious disease. To this end we analysed the capacity for viral detection and identification of a newly described high density resequencing microarray (RMA), termed PathogenID, which was designed for multiple pathogen detection using database similarity searching. We focused on one of the largest and most diverse viral families described to date, the family *Rhabdoviridae*. We demonstrate that this approach has the potential to identify both known and related viruses for which precise sequence information is unavailable. In particular, we demonstrate that a strategy based on consensus sequence determination for analysis of RMA output data enabled successful detection of viruses exhibiting up to 26% nucleotide divergence with the closest sequence tiled on the array. Using clinical specimens obtained from rabid patients and animals, this method also shows a high species-level concordance with standard reference assays, indicating that it is amenable for the development of diagnostic assays. Finally, twelve animal rhabdoviruses which were currently unclassified, unassigned, or assigned as tentative species within the family *Rhabdoviridae* were successfully detected. These new data allowed an unprecedented phylogenetic analysis of 106 rhabdoviruses, and further suggests that the principles and methodology developed here may be used for the broad-spectrum surveillance and the broader-scale investigation of biodiversity in the viral world.

JVI Accepts published online ahead of print

INTRODUCTION

The ability to simultaneously screen for a large panel of pathogens in clinical samples, especially viruses, will represent a major development in the diagnosis of infectious diseases and in surveillance programs for emerging pathogens. Currently, most diagnostic methods are based on species-specific viral nucleic acid amplification. Although rapid and extremely sensitive, these methods are suboptimal when testing for a large number of known pathogens, when viral sequence divergence is high, when new but related viruses are anticipated, or when no clear viral etiologic agent is suspected. To overcome these technical difficulties, newer technologies have been employed, especially microarrays dedicated to pathogen detection. Indeed, DNA-microarrays have been shown to be a powerful platform for the highly multiplexed differential diagnosis of infectious diseases. For example, pathogen microarrays can be simultaneously used to screen various viral or bacterial families, and have been successfully used in the detection of microbial agents from different clinical samples (10-12, 19, 32, 35, 41, 42, 48).

The 'classical' DNA-microarrays developed so far are based on use of long oligonucleotide pathogen specific probes (≥ 50 nt). Although powerful in terms of sensitivity, these diagnostic tools have the disadvantage of decreased specificity, making it necessary to target multiple markers, and rely on hybridization patterns for pathogen identifications, leading to unquantifiable errors (4). Moreover, these methods lack comprehensive information about the pathogen at the single nucleotide level, which could represent a major problem when the sequences in question share a high degree of similarity (21). The microarray-based pathogen resequencing assay represents a promising alternative tool to overcome these limitations. This method identifies each specific pathogen and is capable of resequencing, or "fingerprinting", multiple pathogens in a single test. Indeed, this technology uses tiled sets of 10^5 to 10^6 probes of 25mers, which contain one perfectly matched and three mismatched probes per base for both strands of the target genes (16). This technology also offers the potential for a single test that detects and discriminates between a target pathogen and its closest phylogenetic near neighbours, which expands the repertoire of identifiable organisms far beyond those that are initially included on the array. Successful results have been obtained using this technology, especially for the detection of broad-spectrum respiratory tract pathogens using respiratory pathogen microarrays (2, 25, 26) or the detection of a broad range of biothreat agents (1, 23, 36, 45). The amplification step, which is more often limiting for this technology, has also

JVI Accepts published online ahead of print

benefited from recent developments. Phi29 polymerase-based amplification methods provide amplified DNA with minimal changes in sequence and relative abundance for many biomedical applications (3, 31, 40). The amplification factor varied between 10^6 to 10^9, and it was also demonstrated that co-amplification occurred when viral RNA was mixed with bacterial DNA (3). This whole transcriptome amplification (WTA) approach can also be successfully applied to viral genomic RNA of all sizes. Amplifying viral RNA by WTA provides considerably better sensitivity and accuracy of detection than random RT-PCR in the context of resequencing microarrays (RMAs) (3).

The rhabdoviruses are single-stranded, negative-sense RNA genome viruses classified into six genera, three of which – *Vesiculovirus*, *Lyssavirus* and *Ephemerovirus* – include arthropod-borne agents that infect birds, reptiles and mammals, as well as a variety of non-vector-borne mammalian or fish viruses (International Committee on Taxonomy of Viruses database, ICTVdb) (reviewed in (7)). These rhabdoviruses are the etiological agents of human diseases, such as rabies, that cause serious public health problems. Some rhabdoviruses also cause important economic losses in livestock. The three others genera include *Nucleorhabdovirus* and *Cytorhabdovirus* with are arthropod-borne viruses infecting plants, and *Novirhabdovirus* which comprises fish viruses. Other than the well characterized rhabdoviruses that are known to be important for agriculture and public health, there is also a constantly growing list of rhabdoviruses, isolated from a variety of vertebrate and invertebrate hosts, that are partially characterized and are still waiting for definitive genus or species assignment. Considering the large spectrum of potential animal reservoirs of these viruses compared to the few identified virus species, it is highly likely that the number of uncharacterised rhabdoviruses is immense. Unclassified or unassigned viruses have been tentatively identified as members of the *Rhabdoviridae* by electron microscopy, based on their bullet-shaped morphology – a characteristic trait of members of this family – or using their antigenic relationships based on serological tests (9, 38). Gene sequencing and phylogenetic relationships have then been progressively applied to complete this initial virus taxonomy (6, 22, 27). Importantly, a strongly conserved domain in the rhabdovirus genome, within the polymerase gene, is a useful target for the exploration of the distant evolutionary relationships among these diverse viruses (6). This region corresponds to block III of the viral polymerase, a region predicted to be essential for RNA polymerase function as it is highly conserved among most of the RNA-dependant RNA polymerases (14, 33, 46). A direct application using this sequence region was recently described for lyssavirus RNA detection in human rabies diagnosis (13). Taking

JVI Accepts published online ahead of print

advantage of these characteristics, this polymerase region was also used to design probes for high-density RMAs, also called PathogenID arrays (Affymetrix), which are optimized for the detection and sequence determination of several RNA viruses, particularly rhabdoviruses (1).

In the present study, PathogenID microarrays containing probes for the detection of up to 126 viruses were tested using a consensus sequence determination strategy for analysis of output RMA data. We demonstrate that this approach has the potential to identify, in experimentally infected and clinical specimens, known but also phylogenetically related rhabdoviruses for which precise sequence information was not available.

JVI Accepts published online ahead of print

MATERIALS AND METHODS

Design of the PathogenID microarray for rhabdovirus detection

Two generations of PathogenID arrays were used in this study: PathogenID v1.0 containing probes for the detection of 42 viruses (including 3 prototype rhabdoviruses), 50 bacteria and 619 toxin or antibiotic resistance genes (previously described in (1)), and PathogenID v2.0 able to detect 126 viruses (including 30 different rhabdoviruses), 124 bacteria, 673 toxin or antibiotic resistance genes and two human genes as controls. These arrays include prototype sequences of all species (or genotypes) of the genus *Lyssavirus*, of the other major genera defined in the family *Rhabdoviridae*, such as *Ephemero*virus and *Vesiculovirus,* and of 13 rhabdoviruses awaiting classification or tentatively classified among minor groups such as Le Dantec or Hark park groups (6). For all the selected probes tiled on the two versions of PathogenID array, the same conserved region of the viral polymerase gene was used (block III). However, the size of the target region tiled on the array was longer in the second version (up to 937 nucleotides in length for some sequences compared to roughly 500 nucleotides in the first version) (Tables 1 and 2).

Virus strains and biological samples analyzed

Detailed descriptions of all prototype and field virus strains used in this study and their sources are listed in Tables 1 and 2. Briefly, a total of 16 and 31 different viruses were tested using PathogenID v1.0 (15 lyssaviruses and one vesiculovirus) and PathogenID v2.0 (14 lyssaviruses, one vesiculovirus, 12 unassigned and four tentative species of animal rhabdoviruses according to ICTVdb) respectively. Samples tested included *in vitro* infected cells, synthetic nucleotide target (when the corresponding virus strain was not available), brain biopsies obtained from experimentally infected mice, and biological specimens from various animals (brains from bat, cat, dog and fox) and humans (brain, saliva and skin biopsies).

Extraction and amplification of viral RNA

RNA extraction from biological samples was processed with TRI Reagent (Molecular Research Center) according to the manufacturer's recommendations. After extraction, viral RNA were reverse-transcribed, then amplified using the whole transcriptome amplification (WTA) protocol (QuantiTect Whole Transcriptome Kit, Qiagen) as described previously (3).

JVI Accepts published online ahead of print

Microarrays assay

All amplification products obtained from viral RNA were quantified by Quantit BR (Invitrogen) according to manufacturer's instructions or by the NanoDrop ND-1000 spectrophotometer instrument (Thermo Scientific). A recommended amount of DNA target was fragmented and labelled according to GeneChip Resequencing Assay Manual (Affymetrix). The microarray hybridization process was carried out according to the protocol recommended by the manufacturer (Affymetrix). All details and parameter settings for the data analysis (essentially conversion of raw image files obtained from scanning the microarrays into FASTA files containing the sequences of base calls made for each tiled region of the microarray) have been described previously (1). The base call rate refers to the percentage of base calls generated from the full-length tiled sequence.

Data analysis

In the first approach, resequencing data obtained by the PathogenID v1.0 microarray were manually submitted to the nr/nt database of NCBI for BLASTN query. The default BLAST options were modified. The word size was set to 7 nucleotides. The expected threshold was increased from its default value of 10 up to 100,000 to reduce the filtering of short sequences and sequences rich in undetermined calls, which can assist the correct taxonomic identification. To avoid false negative results induced by high number of undetermined nucleotides in the sequences, the 'low complexity level filter' (*-F*) was also turned off. BLAST sorts the resulting hits according to their bit scores so that the sequence that is the most similar to the entry sequence appears first. Identification of virus strains tested was considered successful only when the best hit was unique and corresponded to the expected species or isolate (according to the nucleotide sequences of these viruses already available in the nr/nt database of NCBI).

In the second approach, an automatic bioinformatic-based analysis of RMA data provided by PathogenID v2.0 was developed, including a consensus sequence determination strategy completed with a systematic BLAST strategy. The general workflow of this strategy is represented in Figure 2. A Perl script reads the input data, which consists of one FASTA file per sample that contains all the sequences read by the GSEQ software from the hybridization. A modified version of the filtering process described by Malanoski *et al.* (29) is applied to the sequences. The retained sequences contain stretches of nucleotides that are ascertained according to the following algorithm. Briefly, sequences that do not contain subsequences

JVI Accepts published online ahead of print

fulfilling specific parameters (minimum *m* nucleotide length and maximum undetermined nucleotides '*N*' content) defined by the user are discarded. These parameters differ from those described in the original filtering process where *m* was fixed to 20 and '*N*' was a value depending on *m*, leading to filter out all short subsequences even with a high base call rate. For subsequence determination, the program starts from the first base call of the considered sequence and searches for the first *m*-bases window area that scores the elongation threshold defined by the user, which represents another difference with the filtering process described by Malanoski *et al.* where this elongation threshold was fixed to 60% (29). The subsequence is extended by one base (*m* + 1) if the percentage of '*N*' remains inferior to the elongation threshold. When this threshold is exceeded, the elongation is stopped and the subsequence is conserved. This process is reiterated until the end of the sequence is reached to generate as many informative sequences as possible. All our analysis were performed using filtering parameters as follow: *m* = 12, *N* = 10 and elongation threshold = 10%.

A systematic BLAST strategy to search for sequence homologues was then performed with the filtered sequences containing subsequences. These sequences individually undergo a BLAST analysis based on a local viral and bacterial database (sequences obtained after filtering from the nr/nt database of NCBI, updated and used for BLAST queries on December, 2009), and the taxonomy of the best BLAST hits are retrieved (Fig. 2A). The default BLAST options were modified as previously described. In the case where several hits obtain the highest bit score, the script automatically retrieves the taxonomies of the 10 first BLAST hits. The final taxonomic identification for each virus strain tested was done by the user as follows: i) identification at the species or isolate level when an unique best hit corresponds to expected the species or isolate, ii) identification at the genus level (if available) when multiple best viral hits exist and correspond to different species within the same genus among the *Rhabdoviridae*, iii) identification at the family level when multiple best viral hits exist and correspond to different rhabdoviruses genera, or iv) negative or inaccurate identification when BLAST query is not possible or when multiple best hits correspond to other viral families, respectively.

For the consensus sequence determination strategy, resequencing data obtained from rhabdoviral tiled sequences are filtered as previously described then submitted to a multiple alignment with Clustal W (39) from which a consensus sequence is determined (Fig. 2B). For each sequence in the alignment, if a called base has undetermined calls on both sides, it is

JVI Accepts published online ahead of print

replaced by an undetermined call. If different calls appear in the sequences for a given position, the majority base call is added to the consensus. The positions that contain an undetermined call or a gap are not considered in the majority base call computation. If multiple base calls tie for the majority, an undetermined call appears at this position in the consensus sequence. This procedure generally increases the length and accuracy of the query sequence for subsequent analysis. Homology searching of the consensus sequences is performed with BLAST using the parameters previously described, and the taxonomy of the best hit is retrieved as for the systematic homology searching approach. We tested if the resulting consensus sequences had higher identification accuracy than any individual sequence or could be used to design PCR primers for a characterization of a potential novel isolate.

Sequencing confirmation

Conventional sequencing was undertaken after the PCR amplification of viral targets directly from biological samples (after RNA extraction and reverse transcription) or from 10- to 100-fold water-diluted WTA products. Primer design was first based on consensus sequences obtained using the consensus sequence determination strategy previously described and/or on rhabdovirus nucleotide sequences available on GenBank. Depending on the results obtained and the virus strain tested, primer design, the set of primers used and PCR conditions for partial polymerase gene amplification were then adjusted (list of primers and PCR conditions available on request from the corresponding author). All PCR products were obtained using the proofreading DNA polymerase ExtTaq (Takara). Sequence assembly and consensus sequences were obtained using Sequencher 4.7 (Gene Codes). GenBank accession numbers for the sequences newly acquired are designated GU815994-GU816024 and are indicated in Tables 1 and 2.

Phylogenetic analysis

The data set of 15 newly sequenced rhabdoviruses from this study (including the Sandjimba and Kolongo viruses previously only identified on partial nucleoprotein gene sequences, as well as *Piry virus* for which the nucleotide sequences of different genes were available) was compared with the corresponding block III polymerase amino acid sequences of 91 other rhabdoviruses collected from GenBank (Table 6). DNA translation was performed with BioEdit software (17) and sequence alignment was performed using Clustal W programme (39) and then checked for accuracy by eye. This resulted in a final alignment of 106

sequences of 160 amino acid residues in length. Phylogenetic analysis of these sequences was then undertaken using the Bayesian method available in the MrBayes package (18). This analysis utilized the WAG model of amino acid replacement with a gamma distribution of among-site rate variation. Chains were run for 10 million generations (with a 10% burnin) at which point all parameter estimates had converged. The level of support for each node is provided by Bayesian Posterior Probability (BPP) values.

JVI Accepts published online ahead of print

JVI Accepts published online ahead of print

RESULTS

Identification of lyssaviruses based on two successive PathogenID microarray generations using a systematic BLAST strategy

To test whether PathogenID microarrays, and specifically the prototype tiled regions, could be used for the identification of a broad number of viral variants without relying on predetermined hybridization patterns, representative animal viruses from the family *Rhabdoviridae* (including unassigned or tentatively classified rhabdoviruses according to ICTVdb) were studied. The capability of these RMAs to identify and discriminate between near phylogenetic neighbours was first tested using one sequence of the genus *Lyssavirus* (PV strain, genotype or species 1) tiled on the first generation of PathogenID microarray (Table 1). It was possible to use BLAST to successfully identify virus strains with approximately 18% nucleotide divergence compared to the prototype (Fig. 1). The hybridization of 15 virus strains representative of the genetic diversity found in this species indicated that a single tiled sequence was able to detect all variant strains belonging to the same species.

In addition, we evaluated the spectrum of detection of the second generation of PathogenID microarray, which included one prototype sequence representative of each of the seven species described in the genus *Lyssavirus* (Table 2). All the tested isolates led to the correct species identification using a systematic BLAST strategy when hybridizing target belonging to the same species that is tiled on the array (Table 3). Moreover, all tested isolates of known genotype were also recognized by heterospecific tiled sequences (Table 3). We also investigated the capacity of this RMA to detect more distantly related viruses not yet classified into a species. Isolates 0406SEN and WCBV, which have been proposed to represent new species in the genus *Lyssavirus* (5, 15), were surprisingly recognized by almost all of the seven species sequences tiled on the PathogenID v2.0 microarray (Table 3). This recognition indicates that each sequence tiled on the array has the ability to identify strains that are more than 18% divergent, and up to 25.9% in some cases (Table 3). This analysis also reveals that information on a strain hybridized on PathogenID v2.0 can be obtained from distinct species or isolates tiled on the array. Evaluation of the spectrum of detection of this RMA was further extended to two other genera of the *Rhabdoviridae* - *Ephemerovirus* and *Vesiculovirus* (Table 4). Here again a successful identification was achieved using homospecific sequences tiled on the array confirming the reliability of the identification.

JVI Accepts published online ahead of print

In both experiments (Tables 3 and 4), low base call rate values were obtained for several combinations of hybridized and tiled sequences. These values were sufficient for a viral identification by BLAST, despite the presence of sequence reads as short as 14 nucleotides. This indicates that most of these short sequences corresponded to highly conserved sequence domains. The accuracy of these short sequences was checked by comparison with those obtained by classical sequencing (data not shown).

Identification of lyssaviruses based on the consensus sequence determination strategy

A bioinformatic workflow was developed to gather stretches of sequence reads obtained with more or less distantly related sequences tiled on PathogenID v2.0. The aim of this strategy was to enlarge the length of the determined sequence in order to improve the sensitivity of the BLAST analysis compared to previously described methodologies (29). All sequence reads obtained from prototype sequences of the genus *Lyssavirus* (at least 12 nucleotides long with no more than one undetermined base, whether or not they initially led to a positive BLAST identification), were used to generate a contiguous sequence. When overlapping fragments were identified, a consensus sequence was generated to remove ambiguous or undetermined base calls. The methodology used to obtain consensus sequences confirmed the species identification after BLAST analysis in the case of the seven lyssavirus nucleotide sequences used for hybridization (Table 5). Moreover, these consensus sequences were found to be more powerful in identifying unclassified or new species of lyssaviruses not tiled on the RMA than resequencing data collected individually from each tiled sequences, as shown for strains 0406SEN and WCBV. In both cases, an increase of the base call rate was observed using this consensus sequence strategy, from 63.5% (best base call rate obtained from individual prototype sequences) to 75.9% for 0406SEN strain and from 32.7% to 60.9% for WCBV (Tables 3 and 5). Once again, this increase of nucleotide base determination was associated with a relatively high accuracy (91.8% and 97.3% concordance between the consensus sequences and the reference sequences of 0406SEN and WCBV isolates, respectively (Table 5). To further demonstrate the ability of this strategy to detect and identify novel virus species, consensus sequences were generated based only on six of the seven prototype tiled sequences (excluding the homospecific sequence of the same species tiled on the array). All strains from the seven species tested were accurately and specifically identified using this restricted approach (Table 5). These results indicate that the consensus sequences obtained

JVI Accepts published online ahead of print

could improve the detection of novel domain(s) not identified using only the closest prototype sequence tiled on the RMA.

Assessment of clinical specimens

A total of 17 brain biopsy samples originating from experimentally infected mice as well as various clinical samples (n=8) obtained from the National Reference Centre for Rabies at Institut Pasteur were tested for lyssavirus detection and identification using the two versions of PathogenID microarray (Tables 1 and 2). These specimens were previously collected from humans and animals with clinically documented encephalitis and suspected of rabies. They were used to compare RMA results with conventional methods of diagnosis, including the RT-hemi-nested PCR technique for the *intra-vitam* diagnosis of rabies in humans (13), the fluorescent antibody test, the rabies tissue culture inoculation test, and the ELISA test for the *post-mortem* diagnosis in humans and animals (8, 47). Among the eight clinical samples, most were brain biopsies collected from different rabid mammals including bat (n=1), cat (n=1), dog (n=1), fox (n=2) and in one case human (n=1). The two other samples comprised a saliva specimen and a skin biopsy sample collected from two different rabid human patients (Tables 1 and 2). Except for the skin biopsy case which was not recognized, this comparison demonstrated a complete concordance between our method and conventional methods for all samples tested. Hence, the accuracy of the sequences provided with PathogenID microarray was close to that obtained using classical sequencing (data not shown). The failure to detect lyssaviruses in the skin biopsy samples was probably due to insufficient sensitivity of the current RMA method, as viral RNA was only weakly detected after RT-hnPCR.

In sum, these results demonstrated that the newly developed amplification process by WTA coupled to hybridization to the PathogenID microarray allowed the detection of a large range of viral variants from various complex biological samples, including clinical samples (Tables 1 and 2).

Application of the RMA strategy to characterize new rhabdoviruses

Broad spectrum detection was demonstrated using the consensus sequences-based analysis strategy amongst viruses of the family *Rhabdoviridae*, and the more distantly related viruses examined included many viruses that are not yet classified as species. Accordingly, 17 different rhabdoviruses were tested, using brain samples from experimentally infected mice (n=16) or infected cell suspension. These viruses included four strains belonging to the genus

JVI Accepts published online ahead of print

Vesiculovirus, with *Vesicular stomatitis Indiana virus* (VSIV), Boteke (BOTK), Jurona (JURV) and Porton's (PORV) viruses, currently classified as tentative species for the three latter; two strains belonging to the genus *Ephemerovirus* with Kimberley (KIMV) and kotonkan (KOTV) viruses, corresponding to a tentative and unassigned species respectively; and 11 presently unassigned rhabdoviruses, namely Kamese (KAMV), Mossuril (MOSV), Sandjimba (SAJV), Keuraliba (KEUV), Nkolbisson (NKOV), Garba (GARV), Nasoule (NASV), Ouango (OUAV), Bimbo (BBOV), Bangoran (BGNV) and Gossas (GOSV) viruses (virus taxonomy according to ICTVdb) (Table 2).

In the first step, successful detection and identification of these viruses using the PathogenID v2.0 microarray was obtained for 12 out of 17 (70.5%) viruses; an accurate taxonomic positioning – that is, within the family *Rhabdoviridae*– was also achieved, and for some the corresponding genus (when available) was also matched accurately (data not shown). In the second step, specific and consensus primers were designed based on the stretches of sequences identified by the microarray using the consensus sequence determination strategy, then subsequently used for PCR and classical sequencing of the amplified target nucleotide sequences. For four (GARV, NASV, OUAV and BBOV) of the five rhabdoviruses that failed to be detected by the microarray, a region of 1000 nucleotides of the polymerase gene encompassing that tiled on the array was successfully amplified by PCR and sequenced using the primers described above. The only exception was the GOSV isolate which remained undetected either using the microarray or by PCR. Further, two other rhabdoviruses not previously tested with the PathogenID v2.0 microarray – Kolongo virus (KOLV, unclassified species) and *Piry virus* (PIRYV, vesiculovirus) – were also amplified and sequenced using these primers.

All the newly sequenced nucleotide regions of the polymerase gene were further translated into protein sequences and aligned with 88 sequences of animal or plant rhabdoviruses obtained from GenBank, producing a total data set of 106 sequences, 160 amino acid residues in length. A Bayesian phylogenetic analysis of these sequences tentatively distinguished 15 groups of viruses based on their strongly supported monophyly (Table 6 and Fig. 3). The members of the six genera – *Ephemerovirus*, *Lyssavirus, Vesiculovirus*, *Cytorhabdovirus, Nucleorhabdovirus*, and *Novirhabdovirus* – fall into well-supported monophyletic groups (BPP value ≥0.97) (Fig. 3). Interestingly, this analysis suggested the existence of at least nine more groups of currently unclassified rhabdoviruses,

JVI Accepts published online ahead of print

which reflect important biological characteristics of the viruses in question. Five of these groups have been proposed previously and were further supported by our analysis (data available at the CRORA database website: http://www.pasteur.fr/recherche/banques/CRORA/) (6, 27), reviewed in (7). The first group, tentatively named Hart Park group, contains the previously described Parry Creek (PCRV), Wongabel (WONV), Flanders (FLANV) and Ngaingan (NGAV) viruses added to the newly identified BGNV, KAMV, MOSV and PORV viruses. This group has a large distribution that encompasses Africa, Australia, Malaysia and the USA. These viruses have a wide host range, as they have been found to infect dipterans, birds and mammals. The second group is the Almpiwar group containing four members – two strains of Charleville (CHVV) virus with CHVV_Ch9824 and CHVV_Ch9847 – and the Almpiwar (ALMV) and Humpty Doo (HDOOV) viruses. Viruses of this group were isolated from Australia and are associated with infection in dipterans, lizards but also in birds and mammals including humans. Another group, herein referred to as the Le Dantec group, was also seen to form a distinct cluster with Le Dantec virus (LDV), Fukuoka virus (FUKV) and the two newly molecularly identified KEUV and NKOV viruses. Members were isolated in Japan and in Africa where they were shown to infect dipterans and mammals including humans. The fourth group has been tentatively named the Tibrogargan group and includes the Tupaia (TUPV) and Tibrogargan (TIBV) viruses. These viruses were isolated in South-East Asia, Australia and New Guinea from dipterans and mammals. Finally, we observed the Sigma group as previously described (27). It includes *Drosophila affinis* (DAffSV), *Drosophila obscura* (DObsSV) and two strains of *Drosophila melanogaster* (SIGMAV_AP30 and SIGMAV_HAP23) sigma viruses, infecting *Drosophila* which were found in USA and in Europe.

In addition, four other tentative groups of viruses are newly described in this study. The Sandjimba group includes the first molecularly classified BBOV, BTKV, NASV, GARV and OUAV viruses and the previously described Oak-Vale (OVRV), SJAV and KOLV viruses (identification based only on a limited region of the nucleoprotein gene for the two latter). These viruses were isolated from birds and dipterans from the Central African Republic and Australia (data available at http://www.pasteur.fr/recherche/banques/CRORA/) (6, 9). Interestingly, all the African members of this group clustered closely, whereas the sole Australian virus was more divergent, suggesting a potential geographical segregation. Second, the Sinistar group includes the *Siniperca chuatsi* rhabdovirus (SCRV) isolated from mandarin fish in China (37) and the starry flounder rhabdovirus (SFRV) identified from starry flounder in USA (30). These two viruses appear to be more closely related to Le Dantec group than to

JVI Accepts published online ahead of print

viruses in the genus *Vesiculovirus* in which several other fish rhabdoviruses are classified. The third one is the Moussa group including two isolates of Moussa virus (MOUV_D24 and MOUV_C23) collected from mosquitoes in Ivory Coast (34). Finally, a phylogenetic analysis suggests the presence of another group within the plant rhabdoviruses: the Taastrup group which comprises the single isolate Taastrup virus (TV) isolated from leafhoppers (*Psammotettix alienus*) originally collected in France (28). All these groups were strongly supported by the Bayesian analysis (BPP ≥ 0.98), with the exception of the Sigma group which exhibits a BPP of 0.88.

In addition, classification of some uncharacterised rhabdoviruses from our phylogenetic analysis diverged from that previously suggested by serology (according to ICTVdb), and will probably need further investigation to determine their precise taxonomic position within the family *Rhabdoviridae* (Table 4) (9, 38). In particular, the PORV and BTKV viruses previously identified as vesiculoviruses were included within the Hart Park and Sandjimba groups, respectively, and NKOV was classified into the Le Dantec group instead of the Kern Canyon group. Moreover, in contrast to a previous phylogenetic study (22), TUPV was found to be more closely related to TIBV than to any other isolates of the Sandjimba group. Finally, our study confirmed the previous serologically-based classification of JURV and the recently identified *Scophthalmus maximus* rhabdovirus (SMRV) within the *Vesiculovirus* genus (38, 49).

JVI Accepts published online ahead of print

DISCUSSION

We have analysed the capacity of viral detection and identification of two versions of a newly described RMA, termed PathogenID, which was designed specifically for multiple pathogen detection using database similarity searching (1). To evaluate this microarray we focused on one of the largest and most diverse viral families described to date, the *Rhabdoviridae* (ITCVdb, reviewed in (7)). All virus strains tested (except WCBV) were extracted from biological samples and amplified using a non-specific and unbiased WTA step as previously described (3). Rhabdovirus targeted sequences were selected among blocks of conservation within the polymerase gene (6). This region was chosen so as to encompass a sufficient number of homologous but also polymorphic sites. The key advantage of this RMA strategy is that it does not require a specific match between the tested samples and tiled sequences; indeed, mismatches add value as they allow a precise typing of the unknown genetic re-sequenced element. In our case, the conserved nature of the target region in the polymerase gene (block III) and the capability of detection of the RMA allows a precise taxonomic identification (i.e. family, genus, species) and also provides key information on phylogenetic relationships for some unclassified, unassigned or tentative species of rhabdoviruses. For example, results obtained by the PathogenID v1.0 microarray evaluation demonstrated that most of the intra-species nucleotide diversity found in the genus *Lyssavirus* can be covered by a single prototype sequence tiled on the microarray. Using the second version of PathogenID, which included one prototype sequence of each of the seven species recognized thus far within the genus *Lyssavirus*, we extended the spectrum of detection of the RMA to potentially all known or unknown lyssaviruses (i.e. positive detection of virus isolates presenting up to 25.9% nucleotide divergence with the considered tiled sequence) which is higher than previous reports (24-26, 43, 44).

This study also indicates that accurate viral identification may still be possible even when only shorter sequences are obtained from individual tiled prototype sequences. Indeed, taken individually, these short-length stretches of nucleotide sequence could not give positive results during the initial BLAST query. However, when used in the consensus sequence determination strategy employed here, they improved the identification of virus strains distantly related with that tiled on the RMA. For example, we were able to test and detect rhabdoviruses based on sequence data obtained with tiled sequences that originated from other viral genera.

JVI Accepts published online ahead of print

The strategy developed here also allowed the potential detection of genetically diverse rhabdoviruses, previously identified or unknown, using a limited number of sequences tiled on the microarray. Using the PathogenID v2.0 microarray we were able to identify 30 rhabdoviruses in total. This included 12 viruses currently unclassified, unassigned or assigned as tentative species within the family *Rhabdoviridae* (according to ICTVdb). Moreover, the consensus sequence based-analysis of RMA results was shown to be accurate when compared to sequences obtained through classical sequencing (Table 5 and data not shown). Sequence data provided by the PathogenID v2.0 microarray were also extremely helpful in the design of specific primers to further sequence the targeted region of the viral polymerase gene of some other rhabdoviruses. Finally, this approach allowed us to undertake the largest phylogenetic analysis of the family *Rhabdoviridae* (Table 6 and Fig. 3), even though it is important to note that the list of viruses and potential taxa described here is still incomplete and more viruses will clearly be characterized in the near future. Despite these phylogenetic divisions, all the viruses included in these proposed groups are closely related to vesiculoviruses and ephemeroviruses and were found to infect a large spectrum of animals, included dipterans and mammals (and previously referred to as the dimarhabdovirus supergroup (6) but also lizards (Almpiwar group), birds (especially with Sandjimba group but also with Hart Park group) and fish (Sinistar group) (Table 6).

Although promising, inadequate sequence selection for the design of the RMA, and consequently a lack of coverage of the viral sequence space, represents an important limitation. A proper selection of blocks of conserved sequence across taxonomic subdivisions in the viral world could be similarly defined and targeted by the RMA assay, and in doing so improve the detection power of this tool and therein greatly aid viral identification among the *Rhabdoviridae,* or even in other viral families. The results presented here validated the usefulness of the design methodology. It emphasizes the gain in identification using a consensus sequence strategy determination compared to systematic BLAST strategy (29). Indeed, this strategy allows us to use and accurately analyze the RMA output data, even if only short subsequences with a high base call rate are obtained. It provides an informative alternative to current molecular methods, such as classical or multiplex PCR, for the rapid identification of viral pathogens. It is currently being applied to assist in a new generation of RMA aimed at the detection and identification of genetically diverse and unknown viral pathogens, and more broadly of *any* virus present in a clinical specimen. In contrast to

JVI Accepts published online ahead of print

conventional microarrays, it is not limited by the requirement for prior knowledge on the identities of viruses present in biological samples, and it is not restricted to the detection of a limited number of candidate viruses. As such, this strategy has a great potential for being implemented as a high throughput platform to identify more divergent viral organisms. This technology could be especially useful in clinical diagnosis or in surveillance programs for detecting uncharacterized viral pathogens or highly variable virus strains among a same taxonomic genus or family, which is frequently the case for RNA viruses (2). The potential applications of such a methodology therefore appear to be numerous: differential diagnostics for illnesses with multiple potential causes (for example, central nervous diseases like encephalitis and meningitis), tracking of emergent pathogens, the distinction of biological threats from harmless phylogenetic neighbours, and the broader-scale investigation of biodiversity in the viral world.

JVI Accepts published online ahead of print

ACKNOWLEDGMENTS

This work was supported by Grant No. UC1 AI062613 (Kennedy) from the US National Institute of Allergy and Infectious Diseases, National Institute of Health; Programme Transversal de Recherche (PTR DEVA n°246) from Institut Pasteur, Paris, France, the European Commission through the "VIZIER" Integrated Project (LSHG-CT-2004-511966) and by the Institut Pasteur International Network Actions Concertées InterPasteuriennes (2003/687). We thank the sponsorship of Total-Institut Pasteur for financial support.

We are grateful to D. Blondel, H. Zeller and CRORA database for having provided some of the rhabdovirus isolates tested in this study. We are also grateful to the technical staff of the Genotyping of Pathogens and Public Health Technological Platform for their patience and their excellent work realized for sequencing of the different rhabdoviruses.

JVI Accepts published online ahead of print

REFERENCES

1. **Berthet, N., P. Dickinson, I. Filliol, A. K. Reinhardt, C. Batejat, T. Vallaeys, K. A. Kong, C. Davies, W. Lee, S. Zhang, Y. Turpaz, B. Heym, G. Coralie, L. Dacheux, A. M. Burguière, H. Bourhy, I. G. Old, J. M. Manuguerra, S. T. Cole, and G. C. Kennedy.** 2007. Massively parallel pathogen identification using high-density microarrays. Microbial Biotechnology **1:**79-86.
2. **Berthet, N., I. Leclercq, A. Dublineau, S. Shigematsu, A. M. Burguiere, C. Filippone, A. Gessain, and J. C. Manuguerra.** 2010. High-density resequencing DNA microarrays in public health emergencies. Nat Biotechnol **28:**25-27.
3. **Berthet, N., A. K. Reinhardt, I. Leclercq, S. van Ooyen, C. Batejat, P. Dickinson, R. Stamboliyska, I. G. Old, K. A. Kong, L. Dacheux, H. Bourhy, G. C. Kennedy, C. Korfhage, S. T. Cole, and J. C. Manuguerra.** 2008. Phi29 polymerase based random amplification of viral RNA as an alternative to random RT-PCR. BMC Mol Biol **9:**77.
4. **Bodrossy, L., and A. Sessitsch.** 2004. Oligonucleotide microarrays in microbial diagnostics. Curr Opin Microbiol **7:**245-54.
5. **Botvinkin, A. D., E. M. Poleschuk, I. V. Kuzmin, T. I. Borisova, S. V. Gazaryan, P. Yager, and C. E. Rupprecht.** 2003. Novel lyssaviruses isolated from bats in Russia. Emerg Infect Dis **9:**1623-5.
6. **Bourhy, H., J. A. Cowley, F. Larrous, E. C. Holmes, and P. J. Walker.** 2005. Phylogenetic relationships among rhabdoviruses inferred using the L polymerase gene. J Gen Virol **86:**2849-58.
7. **Bourhy, H., A. Gubala, R. P. Weir, and D. Boyle.** 2008. Animal Rhabdoviruses, p. 111-121. *In* B. W. J. Mahy and M. H. V. Van Regenmortel (ed.), Encyclopedia of Virology, vol. 1. Elsevier, Oxford.
8. **Bourhy, H., P. E. Rollin, J. Vincent, and P. Sureau.** 1989. Comparative field evaluation of the fluorescent-antibody test, virus isolation from tissue culture, and enzyme immunodiagnosis for rapid laboratory diagnosis of rabies. J Clin Microbiol **27:**519-23.
9. **Calisher, C. H., N. Karabatsos, H. Zeller, J. P. Digoutte, R. B. Tesh, R. E. Shope, A. P. Travassos da Rosa, and T. D. St George.** 1989. Antigenic relationships among rhabdoviruses from vertebrates and hematophagous arthropods. Intervirology **30:**241-57.
10. **Chiu, C. Y., A. A. Alizadeh, S. Rouskin, J. D. Merker, E. Yeh, S. Yagi, D. Schnurr, B. K. Patterson, D. Ganem, and J. L. DeRisi.** 2007. Diagnosis of a critical respiratory illness caused by human metapneumovirus by use of a pan-virus microarray. J Clin Microbiol **45:**2340-3.
11. **Chiu, C. Y., A. L. Greninger, K. Kanada, T. Kwok, K. F. Fischer, C. Runckel, J. K. Louie, C. A. Glaser, S. Yagi, D. P. Schnurr, T. D. Haggerty, J. Parsonnet, D. Ganem, and J. L. DeRisi.** 2008. Identification of cardioviruses related to Theiler's murine encephalomyelitis virus in human infections. Proc Natl Acad Sci U S A **105:**14124-9.
12. **Chiu, C. Y., A. Urisman, T. L. Greenhow, S. Rouskin, S. Yagi, D. Schnurr, C. Wright, W. L. Drew, D. Wang, P. S. Weintrub, J. L. Derisi, and D. Ganem.** 2008. Utility of DNA microarrays for detection of viruses in acute respiratory tract infections in children. J Pediatr **153:**76-83.
13. **Dacheux, L., J. M. Reynes, P. Buchy, O. Sivuth, B. M. Diop, D. Rousset, C. Rathat, N. Jolly, J. B. Dufourcq, C. Nareth, S. Diop, C. Iehle, R. Rajerison, C.**

JVI Accepts published online ahead of print

Sadorge, and H. Bourhy. 2008. A reliable diagnosis of human rabies based on analysis of skin biopsy specimens. Clin Infect Dis **47:**1410-7.

14. **Delarue, M., O. Poch, N. Tordo, D. Moras, and P. Argos.** 1990. An attempt to unify the structure of polymerases. Protein Eng **3:**461-7.
15. **Delmas, O., E. C. Holmes, C. Talbi, F. Larrous, L. Dacheux, C. Bouchier, and H. Bourhy.** 2008. Genomic diversity and evolution of the lyssaviruses. PLoS One **3:**e2057.
16. **Hacia, J. G.** 1999. Resequencing and mutational analysis using oligonucleotide microarrays. Nat Genet **21:**42-7.
17. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser **41:**95-98.
18. **Huelsenbeck, J. P., and F. Ronquist.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754-5.
19. **Kistler, A., P. C. Avila, S. Rouskin, D. Wang, T. Ward, S. Yagi, D. Schnurr, D. Ganem, J. L. DeRisi, and H. A. Boushey.** 2007. Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. J Infect Dis **196:**817-25.
20. **Kondo, H., T. Maeda, Y. Shirako, and T. Tamada.** 2006. Orchid fleck virus is a rhabdovirus with an unusual bipartite genome. J Gen Virol **87:**2413-21.
21. **Kothapalli, R., S. J. Yoder, S. Mane, and T. P. Loughran, Jr.** 2002. Microarray results: how accurate are they? BMC Bioinformatics **3:**22.
22. **Kuzmin, I. V., G. J. Hughes, and C. E. Rupprecht.** 2006. Phylogenetic relationships of seven previously unclassified viruses within the family Rhabdoviridae using partial nucleoprotein gene sequences. J Gen Virol **87:**2323-31.
23. **Leski, T. A., B. Lin, A. P. Malanoski, Z. Wang, N. C. Long, C. E. Meador, B. Barrows, S. Ibrahim, J. P. Hardick, M. Aitichou, J. M. Schnur, C. Tibbetts, and D. A. Stenger.** 2009. Testing and validation of high density resequencing microarray for broad range biothreat agents detection. PLoS One **4:**e6569.
24. **Lin, B., K. M. Blaney, A. P. Malanoski, A. G. Ligler, J. M. Schnur, D. Metzgar, K. L. Russell, and D. A. Stenger.** 2007. Using a resequencing microarray as a multiple respiratory pathogen detection assay. J Clin Microbiol **45:**443-52.
25. **Lin, B., A. P. Malanoski, Z. Wang, K. M. Blaney, A. G. Ligler, R. K. Rowley, E. H. Hanson, E. von Rosenvinge, F. S. Ligler, A. W. Kusterbeck, D. Metzgar, C. P. Barrozo, K. L. Russell, C. Tibbetts, J. M. Schnur, and D. A. Stenger.** 2007. Application of broad-spectrum, sequence-based pathogen identification in an urban population. PLoS One **2:**e419.
26. **Lin, B., Z. Wang, G. J. Vora, J. A. Thornton, J. M. Schnur, D. C. Thach, K. M. Blaney, A. G. Ligler, A. P. Malanoski, J. Santiago, E. A. Walter, B. K. Agan, D. Metzgar, D. Seto, L. T. Daum, R. Kruzelock, R. K. Rowley, E. H. Hanson, C. Tibbetts, and D. A. Stenger.** 2006. Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. Genome Res **16:**527-35.
27. **Longdon, B., D. J. Obbard, and F. M. Jiggins.** 2010. Sigma viruses from three species of Drosophila form a major new clade in the rhabdovirus phylogeny. Proc Biol Sci **277:**35-44.
28. **Lundsgaard, T.** 1997. Filovirus-like particles detected in the leafhopper Psammotettix alienus. Virus Res **48:**35-40.
29. **Malanoski, A. P., B. Lin, Z. Wang, J. M. Schnur, and D. A. Stenger.** 2006. Automated identification of multiple micro-organisms from resequencing DNA microarrays. Nucleic Acids Res **34:**5300-11.

JVI Accepts published online ahead of print

30. **Mork, C., P. Hershberger, R. Kocan, W. Batts, and J. Winton.** 2004. Isolation and characterization of a rhabdovirus from starry flounder (Platichthys stellatus) collected from the northern portion of Puget Sound, Washington, USA. J Gen Virol **85:**495-505.
31. **Paez, J. G., M. Lin, R. Beroukhim, J. C. Lee, X. Zhao, D. J. Richter, S. Gabriel, P. Herman, H. Sasaki, D. Altshuler, C. Li, M. Meyerson, and W. R. Sellers.** 2004. Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. Nucleic Acids Res **32:**e71.
32. **Palacios, G., P. L. Quan, O. J. Jabado, S. Conlan, D. L. Hirschberg, Y. Liu, J. Zhai, N. Renwick, J. Hui, H. Hegyi, A. Grolla, J. E. Strong, J. S. Towner, T. W. Geisbert, P. B. Jahrling, C. Buchen-Osmond, H. Ellerbrok, M. P. Sanchez-Seco, Y. Lussier, P. Formenty, M. S. Nichol, H. Feldmann, T. Briese, and W. I. Lipkin.** 2007. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. Emerg Infect Dis **13:**73-81.
33. **Poch, O., I. Sauvaget, M. Delarue, and N. Tordo.** 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. Embo J **8:**3867-74.
34. **Quan, P. L., S. Junglen, A. Tashmukhamedova, S. Conlan, S. K. Hutchison, A. Kurth, H. Ellerbrok, M. Egholm, T. Briese, F. H. Leendertz, and W. I. Lipkin.** 2010. Moussa virus: a new member of the Rhabdoviridae family isolated from Culex decens mosquitoes in Cote d'Ivoire. Virus Res **147:**17-24.
35. **Quan, P. L., G. Palacios, O. J. Jabado, S. Conlan, D. L. Hirschberg, F. Pozo, P. J. Jack, D. Cisterna, N. Renwick, J. Hui, A. Drysdale, R. Amos-Ritchie, E. Baumeister, V. Savy, K. M. Lager, J. A. Richt, D. B. Boyle, A. Garcia-Sastre, I. Casas, P. Perez-Brena, T. Briese, and W. I. Lipkin.** 2007. Detection of respiratory viruses and subtype identification of influenza A viruses by GreeneChipResp oligonucleotide microarray. J Clin Microbiol **45:**2359-64.
36. **Taitt, C. R., A. P. Malanoski, B. Lin, D. A. Stenger, F. S. Ligler, A. W. Kusterbeck, G. P. Anderson, S. E. Harmon, L. C. Shriver-Lake, S. K. Pollack, D. M. Lennon, F. Lobo-Menendez, Z. Wang, and J. M. Schnur.** 2008. Discrimination between biothreat agents and 'near neighbor' species using a resequencing array. FEMS Immunol Med Microbiol **54:**356-64.
37. **Tao, J. J., G. Z. Zhou, J. F. Gui, and Q. Y. Zhang.** 2008. Genomic sequence of mandarin fish rhabdovirus with an unusual small non-transcriptional ORF. Virus Res **132:**86-96.
38. **Tesh, R. B., A. P. Travassos Da Rosa, and J. S. Travassos Da Rosa.** 1983. Antigenic relationship among rhabdoviruses infecting terrestrial vertebrates. J Gen Virol **64 (Pt 1):**169-76.
39. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 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:**4673-80.
40. **Vora, G. J., C. E. Meador, D. A. Stenger, and J. D. Andreadis.** 2004. Nucleic acid amplification strategies for DNA microarray-based pathogen detection. Appl Environ Microbiol **70:**3047-54.
41. **Wang, D., L. Coscoy, M. Zylberberg, P. C. Avila, H. A. Boushey, D. Ganem, and J. L. DeRisi.** 2002. Microarray-based detection and genotyping of viral pathogens. Proc Natl Acad Sci U S A **99:**15687-92.
42. **Wang, D., A. Urisman, Y. T. Liu, M. Springer, T. G. Ksiazek, D. D. Erdman, E. R. Mardis, M. Hickenbotham, V. Magrini, J. Eldred, J. P. Latreille, R. K.**

JVI Accepts published online ahead of print

Wilson, D. Ganem, and J. L. DeRisi. 2003. Viral discovery and sequence recovery using DNA microarrays. PLoS Biol **1:**E2.

43. **Wang, Z., L. T. Daum, G. J. Vora, D. Metzgar, E. A. Walter, L. C. Canas, A. P. Malanoski, B. Lin, and D. A. Stenger.** 2006. Identifying influenza viruses with resequencing microarrays. Emerg Infect Dis **12:**638-46.
44. **Wang, Z., A. P. Malanoski, B. Lin, C. Kidd, N. C. Long, K. M. Blaney, D. C. Thach, C. Tibbetts, and D. A. Stenger.** 2008. Resequencing microarray probe design for typing genetically diverse viruses: human rhinoviruses and enteroviruses. BMC Genomics **9:**577.
45. **Wilson, W. J., C. L. Strout, T. Z. DeSantis, J. L. Stilwell, A. V. Carrano, and G. L. Andersen.** 2002. Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. Mol Cell Probes **16:**119-27.
46. **Xiong, Y., and T. H. Eickbush.** 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. Embo J **9:**3353-62.
47. **Xu, G., P. Weber, Q. Hu, H. Xue, L. Audry, C. Li, J. Wu, and H. Bourhy.** 2007. A simple sandwich ELISA (WELYSSA) for the detection of lyssavirus nucleocapsid in rabies suspected specimens using mouse monoclonal antibodies. Biologicals **35:**297-302.
48. **Yoo, S. M., J. Y. Choi, J. K. Yun, J. K. Choi, S. Y. Shin, K. Lee, J. M. Kim, and S. Y. Lee.** 2009. DNA microarray-based identification of bacterial and fungal pathogens in bloodstream infections. Mol Cell Probes.
49. **Zhang, Q. Y., J. J. Tao, L. Gui, G. Z. Zhou, H. M. Ruan, Z. Q. Li, and J. F. Gui.** 2007. Isolation and characterization of Scophthalmus maximus rhabdovirus. Dis Aquat Organ **74:**95-105.

JVI Accepts published online ahead of print

FIGURES

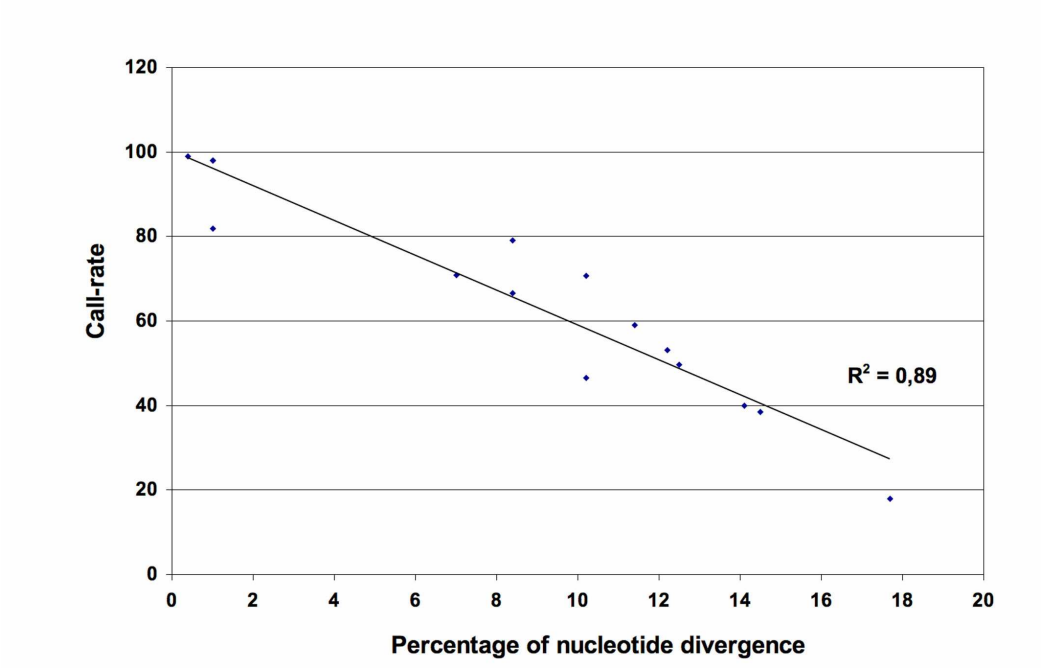


Figure 1: Spectrum of detection of the PathogenID v1.0 microarray among the genus *Lyssavirus* according to the natural nucleotide variation of the virus strains tested. For each lyssavirus strain tested (n=15, indicated with black diamonds), results are indicated by the percentage nucleotide divergence (compared to the single lyssavirus prototype sequence tiled on the microarray, x-axis) according to the percentage of nucleotide bases determined (call-rate, y-axis). The linear correlation curve between these two values is presented, demonstrating a high correlation between these two parameters (correlation coefficient value of 0.89). All these 15 virus strains belonged to the same species as the tiled prototype sequence (genotype 1) and were accurately identified after BLAST analysis (at the species level). Other species (or genotypes) of lyssaviruses were not successfully detected with PathogenID v1.0 microarray (nucleotide divergence over 20%, data not shown). For further details concerning the lyssavirus strains used, see Table 1.

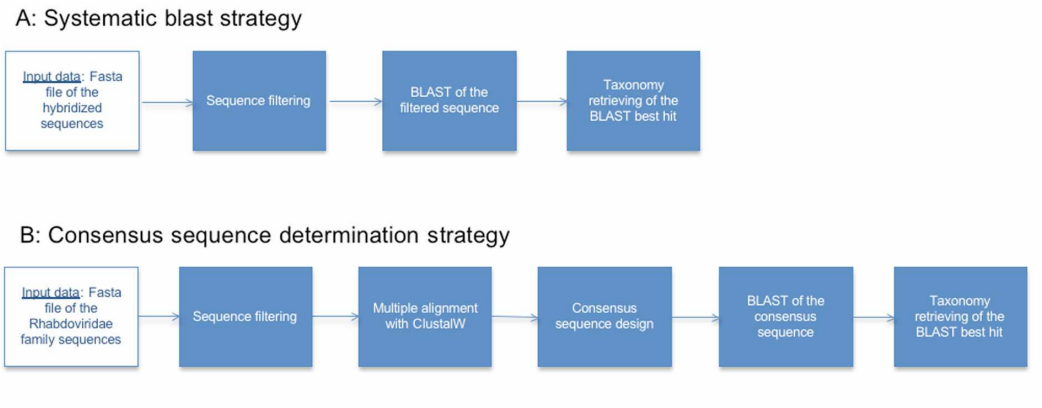


Figure 2: Descriptive workflow of the automatic Perl bioinformatic-based analysis of the PathogenID v2.0 data.

A: Systematic BLAST strategy

This strategy consists of filtering the sequences obtained from the output data of the RMA with filter parameters defined by the user (see Methods section for further details), followed by the systematic researching of homologues using a local BLAST viral and bacterial database, and finally retrieve taxonomy of the best BLAST hits.

B: Consensus sequence determination strategy

A consensus sequence is generated using a multiple alignment with Clustal W based on the sequences obtained from prototype rhabdovirus sequences tiled on the microarray. With this process, the length and the accuracy of the query sequence can be increased. Homology searching of the consensus sequences is performed with BLAST using the previously described parameters and database. The taxonomy of the best BLAST hit is retrieved as for the systematic homology searching approach.

JVI Accepts published online ahead of print

JVI Accepts published online ahead of print

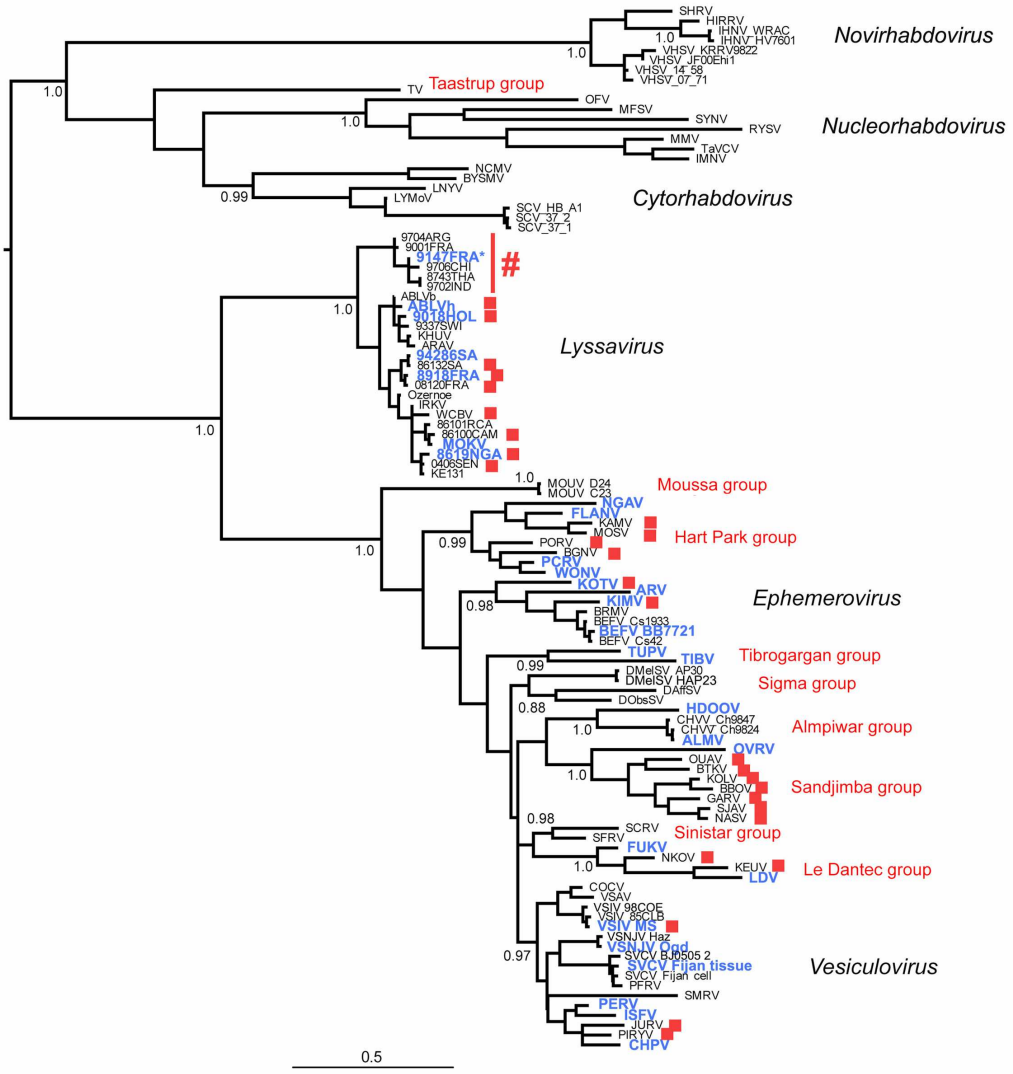


Figure 3: Phylogenetic relationships of the *Rhabdoviridae* based on a 160 amino acid alignment of the polymerase gene.

Phylogenetic analysis of 106 amino-acid sequences of the block III of the polymerase (160 amino acid residues in length) of rhabdoviruses using a Bayesian method. This analysis utilized the WAG model of amino acid replacement with a gamma distribution of among-site rate variation. Chains were run for 10 million generations (with a 10% burnin) at which point all parameter estimates had converged. The level of support for each node is provided by Bayesian Posterior Probability (BPP) values. The genera (black font) and groups (red font) of the *Rhabdoviridae* are indicated, along with their associated BPP values. All horizontal branch lengths are drawn to a scale of amino acid replacements per residue. The tree is mid-point rooted for purposes of clarity only. Sequence tiled on the array or their closely related sequence (* 9147FRA instead of PV) are indicated in blue font. Sequences corresponding to lyssavirus species 1 and positively detected by PathogenID v1.0 are indicated by a red line (#). Sequences detected by PathogenID v2.0 are indicated by red squares**.**

JVI Accepts published online ahead of print

JVI Accepts published online ahead of print

TABLES

Table 1 :
Description of virus species belonging to the *Rhabdoviridae* used for selection of tiled sequences and for validation of the PathogenID v1.0 microarray.

Genus and specie name[a] (abbreviation)	Reference no.	Host species/ vector	Origin	Year of first isolation	Tiled region[b]	Length (bp)	Biological samples tested	GenBank accession no.
Origin of tiled sequences								
Lyssavirus								
Rabies virus (RABV)	PV	Vaccine	--	--	7452-7953	502	--	NC_001542
Vesiculovirus								
Vesicular stomatitis Indiana virus (VSIV)	VSVLMS	--	--	--	7453-7953	497	--	K02378
Ephemerovirus								
Bovine ephemeral fever virus (BEFV)	BB7721	*Bos taurus*	Australia	1968	7454-7952	498	--	NC_002526
Tested rhabdovirus species								
Lyssavirus								
Genotype 1								
Rabies virus (RABV)	8764THA	Human	Thailand	1983	--	--	Human brain	EU293111
Rabies virus (RABV)	9147FRA	Red fox	France	1991	--	--	Fox brain	EU293115
Rabies virus (RABV)	93128MAR	Fixed strain	Morroco	?	--	--	Mouse brain	GU815994
Rabies virus (RABV)	9811CHI	Dog	China	1998	--	--	Mouse brain	GU815995

JVI Accepts published online ahead of print

Rabies virus (RABV)	0435AFG	Dog	Afghanistan	2004	--	--	Mouse brain	GU815996
Rabies virus (RABV)	9001FRA	Dog bitten by a bat	French Guyana	1990	--	--	Mouse brain	EU293113
Rabies virus (RABV)	9026CI	Dog	Ivory Coast	1990	--	--	Mouse brain	GU815997
Rabies virus (RABV)	9105USA	Fox	USA	1991	--	--	Fox brain	GU815998
Rabies virus (RABV)	9233GAB	Dog	Gabon	1992	--	--	Dog brain	GU815999
Rabies virus (RABV)	93127FRA	Fixed strain	France	?	--	--	Mouse brain	GU816000
Rabies virus (RABV)	9503TCH	Fixed strain (Vnukovo, SAD)	Tchekoslovakia	?	--	--	Mouse brain	GU816001
Rabies virus (RABV)	9737POL	Raccoon dog	Poland	1997	--	--	Mouse brain	GU816002
Rabies virus (RABV)	Challenge Virus Strain (CVS_IP13)	Fixed strain	--	--	--	--	Mouse brain	GU816003
Rabies virus (RABV)	ERA	Fixed strain	--	--	--	--	Mouse brain	GU816005
Rabies virus (RABV)	LEP	Fixed strain	--	--	--	--	Chicken embryo fibroblasts	GU816004
Vesiculovirus								
Vesicular stomatitis Indiana virus (VSIV)	Orsay (0503FRA)	Fixed strain	--	--	--	--	BSR cells[c]	GU816006

[a] Classification and names of viruses correspond to approved virus taxomony according to the International Committee on Taxonomy of Viruses database (ICTVdb). Names of viruses in italic correspond to validated virus species.
[b] Position according to the reference Pasteur virus genome (NC_001542), after alignment of all the tiled sequences with the reference sequence.
[c] A clone of baby hamster kidney cell line (BHK-21).

JVI Accepts published online ahead of print

Table 2 : Description of virus species belonging to the *Rhabdoviridae* used for selection of tiled sequences and for validation of the PathogenID v2.0 microarray.

Genus or group[a]	Species and name[a]	UA/TS/ UC[b]	Reference no.	Host species/vector	Tiled region[c]	Length (bp)	Biological samples tested	Origin of samples	Year of first isolation	GenBank accession no.
Origin of tiled sequences										
Lyssavirus										
	Genotype 1 *Rabies virus* (RABV)		PV	Vaccine	7040-7977	937	--	--	--	NC_001542
	Genotype 2 *Lagos bat virus* (LBV)		8619NGA	Bat : *Eidolon helvum*	7040-7977	937	--	Nigeria	1956	EU293110
	Genotype 3 *Mokola virus* (MOKV)		MOKV	Cat	7040-7977	937	--	Zimbabwe	1981	NC_006429
	Genotype 4 *Duvenhage virus* (DUVV)		94286SA	Bat : *Minopterus* species	7040-7977	937	--	South Africa	1981	EU293120
	Genotype 5 *European*		8918FRA	Bat : *Eptesicus*	7040-	937	--	France	1989	EU293112

JVI Accepts published online ahead of print

	bat lyssavirus 1 (EBLV-1*)*		*serotinus*	7977					
	Genotype 6 *European bat lyssavirus 2* (EBLV-2)	9018HOL	Bat : *Myotis dasycneme*	7040-7977	937	--	The Netherlands	1986	EU293114
	Genotype 7 *Australian bat lyssavirus* (ABLV)	ABLh	Human	7040-7977	937	--	Australia	1986	AF418014
Vesiculovirus									
	Chandipura virus (CHPV)	I 653514	Human	7040-7981	935	--	India	1965	AJ810083
	Isfahan virus (ISFV)	91026-167	*Phlebotomus papatasi*	7040-7981	935	--	Iran	1975	AJ810084
	Vesicular stomatitis New Jersey virus (VSNJV)	VSV NJ-O	*Bos taurus*, equine / *Culex nigripalpus*, *Culicoides* species, *Mansonia indubitans*	7040-7981	935	--	USA	1949	AY074804
	Vesicular stomatitis	VSVLMS	--	7040-7981	935	--	--	--	K02378

JVI Accepts published online ahead of print

	Indiana virus (VSIV)									
	Perinet virus (PERV)	TS	Ar Mg 802	*Anopheles coustani*, *Culex antennatus*, *Culex* gr. *pipiens*, *Mansonia uniformis*, *Phlebotomus berentens*is	7089-7502	405	--	Madagascar	1978	AY854652
	Spring viremia of carp virus (SVCV)	TS	VR-1390	*Cyprinus carpio*	7040-7981	935	--	Yougoslavia	1971	U18101
Ephemerovirus										
	Adelaide River virus (ARV)		DPP 61	*Bos taurus*	7089-7502	408	--	Australia	1981	AY854635
	Bovine ephemeral fever virus (BEFV)		BB7721	*Bos taurus*	7089-7502	408	--	?	1968	AY854642
	Kimberley virus (KIMV)	TS	CS 368	*Bos taurus*	7089-7502	408	--	Australia	1980	AY854637
	Kotonkan virus[d] (KOTV)	UA	Ib Ar23380	*Culicoides* species	7089-7502	408	--	Nigeria	1967	AY854638

JVI Accepts published online ahead of print

Other dimarhabodviruses[d]										
	Almpiwar group									
	Almpiwar virus (ALMV)	UA	MRM4059	*Ablepharus boutonii virgatus*	7089-7502	411	--	Australia	1966	AY854645
	Humpty doo virus (HDOOV)	UA	CS 79	*Lasiohelea* species	7089-7502	411	--	Australia	1975	AY854643
	Oak-Vale virus (OVRV)	UA	CS 1342	*Culex* species	7089-7502	408	--	Australia	1981	AY854670
	Hart Park group									
	Flanders virus (FLANV)	UA	61-7484	*Culiseta melanura, Culex* species / *Seiurus aurocapillus*	7089-7502	410	--	USA	1961	AF523199
	Ngaingan virus (NGAV)	UA	NRM14556	*Culicoides brevitarsis* / wallabies, kangaroos, cattle	7089-7502	408	--	Australia	1970	AY854649
	Parry Creek virus (PCRV)	UA	OR 189	*Culex annulirostris*	7089-7502	408	--	Australia	1972	AY854647
	Wongabel	UA	CS 264	*Culicoides*	7089-	408	--	Australia	1979	AY854648

JVI Accepts published online ahead of print

	virus (WONV)			*austropalpalis*	7502					
	Le Dantec and Kern Canyon group									
	Fukuoka virus (FUKV)	UA	FUK-11	*Culicoides punctatus*	7089-7502	408	--	Japan	1982	AY854651
	Le Dantec virus (LDV)	UA	DakHD 763	Human	7089-7502	408	--	Senegal	1965	AY854650
	Tibrogargan group									
	Tibrogargan virus (TIBV)	UA	CS 132	*Culicoides brevitarsis,* water buffaloes, cattle	7089-7502	408	--	Australia	1976	AY854646
Other animal rhabdoviruses										
	Tupaia rhabdovirus (TUPV)	UA	TRV 1591	*Tupaia belangeri*	7089-7502	408	--	Thailand	?	NC_007020
	Sigma virus (SIGMAV)	UA	234HRC	*Drosophila melanogaster*	6220-6642	408	--	?	?	X91062
	Sea trout rhabdovirus (STRV)	UC	28/97	*Salmo trutta trutta*	7108-7576	415		Sweden	1996	AF434992

JVI Accepts published online ahead of print

Tested rhabdovirus species									
Lyssavirus									
	Genotype 1 *Rabies virus* (RABV)	93127FRA	Fixed strain	--	--	Mouse brain	France	?	GU816000
	Rabies virus (RABV)	8764THA	Human	--	--	Human brain	Thailand	1983	EU293111
	Rabies virus (RABV)	08339FRA	Human (probably contamined by a bat)	--	--	Human saliva	France (French Guyana)	2008	GU816007
	Rabies virus[h] (RABV)	07029SEN	Human	--	--	Skin biopsy	Senegal	2006	--
	Genotype 2 *Lagos bat virus* (LBV)	8619NGA	Bat : *Eidolon helvum*	--	--	Mouse brain	Nigeria	1956	EU293110
	Genotype 3 *Mokola virus* (MOKV)	86100CAM	Shrew	--	--	Mouse brain	Cameroon	1981	NC_006429
	Genotype 4 *Duvenhage virus* (DUVV)	86132SA	Human	--	--	Mouse brain	South Africa	1971	EU293119

JVI Accepts published online ahead of print

Genotype 5 *European bat lyssavirus 1* subtype a (EBLV-1a)	07240FRA	Cat (contaminated by a bat)	--	--	Cat brain	France	2007	EU626552
European bat lyssavirus 1 subtype a (EBLV-1b)	08341FRA	Bat : *Eptesicus serotinus*	--	--	Bat brain	France	2008	GU816009
European bat lyssavirus 1 subtype b (EBLV-1b)	8918FRA	Bat : *Eptesicus serotinus*	--	--	Mouse brain	France	1989	EU293112
Genotype 6 *European bat lyssavirus 2* (EBLV-2)	9018HOL	Bat : *Myotis dasycneme*	--	--	Mouse brain	Holland	1986	EU293114
Genotype 7 *Australian bat lyssavirus* (ABLV)	9810AUS	Bat	--	--	Mouse brain	Australia	?	GU816008
Genotype 8 (tentative species)								

JVI Accepts published online ahead of print

	Dakar bat lyssavirus (DBLV)	UC	0406SEN (AnD 42443)	Bat : *Eidolon helvum*	--	--	Mouse brain	Senegal	1985	EU293108
	Not assigned West Caucasian Bat Virus (WCBV)	UC	--	Bat : *Myotis schreibersi*	--	--	Plasmid[e]	Russia	2002	EF614258
Vesiculovirus										
	Vesicular stomatitis Indiana virus (VSIV)		Orsay (0503FRA)	?	--	--	BSR cells[f]	?	?	GU816006
	Boteke virus (BTKV)	TS	DakArB 1077 (0417RCA)	*Coquillettidia maculipennis*	--	--	Mouse brain	Central African Republic	1968	GU816014
	Jurona virus (JURV)	TS	BeAr 40578 (0414BRE)	*Haemagogus spegazzinii*	--	--	Mouse brain	Brazil	1962	GU816024
	Porton's virus (PORV)	TS	1643 (0416MAL)	*Mansonia uniformis*	--	--	Mouse brain	Malaysia (Sarawak)	?	GU816013
Ephemerovirus										
	Kotonkan virus[d]	UA	Ib Ar23380	*Culicoides* species	--	--	Mouse brain	Nigeria	1967	AY854638

JVI Accepts published online ahead of print

	(KOTV)		(9145NIG)							
	Kimberley virus (KIMV)	TS	CS 368	*Bos taurus*	--	--	Mouse brain	Australia	1980	AY854637
Other animal rhabodviruses										
	Hart Park group									
	Kamese virus (KAMV)	UA	MP 6186 (08343RCA)	*Aedes africanus, Culex* species	--	--	Mouse brain	Central African Republic	1967	GU816011
	Mossuril virus (MOSV)	UA	SA Ar 1995 (0418SA)	*Aedes abnormalis, Culex* species / *Andropadus virens , Coliuspasser macrourus*	--	--	Mouse brain	Central African Republic , Gambia	1959	GU816012
	Kolongo and Sandjimba group									
	Sandjimba virus (SJAV)	UA	DakAnB 373d (07244RCA)	*Acrocephalus schoenobaenus*	--	--	Mouse brain	Central African Republic	1970	GU816019
	Le Dantec									

JVI Accepts published online ahead of print

and Kern Canyon group									
Keuraliba virus (KEUV)	UA	DakAnD 5314 (9715SEN, 0420SEN)	*Tatera kempi*, *Taterillus* species	--	--	Mouse brain	Senegal	1968	GU816021
Nkolbisson virus (NKOV)	UA	Ar Y 31/65 (0425CAM)	*Aedes* species, *Eretmapodites* species, *Culex telesilla*	--	--	Mouse brain	Ivory Coast, Cameroon	1965	GU816022
Ungrouped									
Garba virus[g] (GARV)	UA	DakAnB 439a (0422RCA	*Corythornis cristata, Nectarina pulchellabelangeri*	--	--	Mouse brain	Central African Republic	1970	GU816018
Nasoule virus[g] (NASV)	UA	DakAnB 4289a (0410RCA)	*Andropadus virens*	--	--	Mouse brain	Central African Republic	1973	GU816017
Ouango virus[g] (OUAV)	UA	DakAnB 1582a (9718RCA)	*Ploceus melanocephalus*	--	--	Mouse brain	Central African Republic	1970	GU816015
Bimbo virus[g] (BBOV)	UA	DakAnB 1054d (9716RCA)	*Euplectes afra*	--	--	Mouse brain	Central African Republic	1970	GU816016
Bangoran	UA	DakArB	*Turdus*	--	--	Mouse	Central	1969	GU816010

JVI Accepts published online ahead of print

virus (BGNV)		2053 (0424RCA)	*libonyanus* / *Culex perfuscus*			brain	African Republic		
Gossas virus[h] (GOSV)	UA	DakAnD 401 (08344SEN)	*Tadarida* species	--	--	Mouse brain	Senegal	1964	NA[i]

[a] Unless stated, the classification and names of viruses correspond to approved virus taxomony according to the International Committee on Taxonomy of Viruses database (ICTVdb). Names in italic correspond to validated virus species.
[b] UA : unassigned, TS : tentative species, UC : unclassified (not found in the ICTVdb).
[c] Position according to the reference Pasteur virus genome (NC_001542), after alignment of all the tiled sequences with the reference sequence (except for tiled sequences from ungrouped rhabdovirus TUPV and SIGMAV which where where aligned independently with the reference sequence).
[d] Taxonomical classification according to (6).
[e] A 977 nucleotides fragment of the polymerase gene (from 7020 to 7997 according to the reference Pasteur virus genome (NC_001542)) was synthetized in vitro then cloned into pCR2.1 plasmid (Operon).
[f] A clone of baby hamster kidney cell line (BHK-21).
[g] Not detected using PathogenID v2.0 microarray but amplified by PCR or nested PCR using consensus or specific primers.
[h] Not detected using PathogenID v2.0 microarray neither amplified by PCR or nested PCR using consensus or specific primers.
[i] NA : Not applicable.

JVI Accepts published online ahead of print

Table 3: Level of taxonomic identification of virus species among the genus *Lyssavirus* based on lyssavirus sequences tiled on the PathogenID_v2.0 microarray.

Strains of lyssavirus tested			Results from lyssavirus sequences tiled						
Genotype (abbreviation)			1 (RABV)	2 (LBV)	3 (MOKV)	4 (DUVV)	5 (EBLV-1)	6 (EBLV-2)	7 (ABLV)
Isolate			PV	8619NGA	MOKV	94286SA	8918FRA	9018HOL	ABLV
1 (RABV)	93127FRA	Base call rate[a]	95.0	3.8	4.6	6.2	8.0	9.0	6.7
		Identification[b]	**A**	**C**	**B**	**A**	**A**	**A**	**B**
		Divergence[c]	0.2	25.6	24.8	22.8	23.2	21.0	22.0
	8764THA	Base call rate	32.6	5.4	5.7	7.0	6.0	3.3	9.0
		Identification	**A**	**A**	**B**	**C**	**A**	**A**	**B**
		Divergence	13.7	24.3	24.9	22.7	22.9	21.2	20.7
2 (LBV)	8619NIG	Base call rate	2.7	96.6	11.1	6.9	7.1	5.3	6.7
		Identification	**Neg**	**A**	**A**	**A**	**Neg**	**Neg**	**B**
		Divergence	25.8	0.0	22.3	22.8	25.2	23.8	23.3
3 (MOKV)	86100CAM	Base call rate	2.7	7.2	56.3	8.1	7.0	7.7	3.6
		Identification	**A**	**A**	**A**	**A**	**A**	**A**	**B**
		Divergence	24.9	22.5	10.2	22.0	23.6	22.2	24.5
4 (DUVV)	86132SA	Base call rate	3.9	1.1	1.4	97.3	0.6	2.5	5.6
		Identification	**A**	**Neg**	**Neg**	**A**	**A**	**A**	**Neg**
		Divergence	23.2	22.8	22.2	6.0	20.6	21.6	21.9
5 (EBLV-1)	8918FRA	Base call rate	8.1	6.9	15.2	13.3	93.8	7.8	4.8
		Identification	**B**	**A**	**A**	**A**	**A**	**A**	**Neg**

JVI Accepts published online ahead of print

		Divergence	23.8	25.5	23.8	20.7	0.6[d]	23.4	22.1
6 (EBLV-2)	9018HOL	Base call rate	5.7	2.1	3.5	6.5	4.1	98.4	8.7
		Identification	**Neg**	**Neg**	**B**	**A**	**B**	**A**	**A**
		Divergence	21.2	23.8	23.5	21.7	23.3	0.0	22.3
7 (ABLV)	9810AUS	Base call rate	8.4	8.3	1.4	12.9	3.8	11.3	94.9
		Identification	**B**	**A**	**Neg**	**A**	**B**	**B**	**A**
		Divergence	22.5	23.7	24.4	21.6	22.1	22.4	1.6
8[e] (DBLV)	0406SEN	Base call rate	19.3	63.5	29.4	16.3	22.8	18.3	19.5
		Identification	**A**	**A**	**A**	**A**	**A**	**A**	**A**
		Divergence	25.0	20.1	21.5	23.4	23.5	22.8	23.8
Non classified	WCBV	Base call rate	25.3	28.3	32.7	26.9	26.5	23.8	24.5
		Identification	**C**	**A**	**A**	**A**	**A**	**A**	**A**
		Divergence	25.7	23.8	24.2	24.9	24.6	24.8	25.9

[a] Percentage of the base calls generated from full-length tiled sequences.
[b] Taxonomic identification according to:
A = identification at the species or isolate level when a unique best hit corresponds to expected species or isolate
B = identification at the genus level when multiple best viral hits exist and correspond to the genus *Lyssavirus*
C = identification at the family level when multiple best viral hits exist and correspond to genera of the family *Rhabdoviridae*
Neg = negative or inaccurate identification when BLAST query is not possible or when multiple best hits and some or all of them correspond to other viral families, respectively
[c] Percentage of nucleotide divergence (based on a 937 nucleotide region of the polymerase gene, position 7040-7977 according to the reference Pasteur virus genome (NC_001542).
[d] The tiled sequence of 8918FRA correspond to a preliminary result of sequencing and the complete genome of this virus strain was obtained latter (EU293112), which may explain the nucleotide differences (n=7) between those two sequences.
[e] Tentative genotype.
In grey: results obtained using the sequence belonging to the same species tiled on the array (homonymous sequence).

JVI Accepts published online ahead of print

Table 4: Level of taxonomic identification of virus species among the genera *Vesiculovirus* and *Ephemerovirus* based on vesiculovirus and ephemerovirus sequences tiled on the PathogenID_v2.0 microarray.

Strains of rhabdovirus tested			Results from specific rhabdovirus sequences tiled										
Genus			*Vesiculovirus*						*Ephemerovirus*				*Lyssavirus*
	Isolate (reference)		CHPV	ISFV	PERV	SVCV	VSIV	VSNJV	ARV	BEFV	KIMV	KOTV	RABV (PV)
Vesiculovirus	VSIV (0503FRA)	Base call rate[a]	1.2	4.1	1.0	1.0	98.6	2.9	0	0	0	0	0
		Score[b]	**Neg**	**Neg**	**Neg**	**Neg**	**A**	**A**	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**
Ephemerovirus	KIMV[c] (CS 368)	Base call rate	1.9	1.1	0	0	0.3	0.3	9.4	7.3	70.6	9.1	1.4
		Score	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**	**A**	**Neg**	**Neg**
	KOTV[d] (Ib Ar23380, 9145NIG)	Base call rate	6.6	3.8	5.7	3.2	3.7	7.2	8.8	5.2	3.4	100	2.1
		Score	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**	**C**	**Neg**	**Neg**	**Neg**	**A**	**Neg**
Lyssavirus	RABV (93127FRA)	Base call rate	0.3	1.2	2.6	1.5	0	0	0.1	0	0.1	2.3	95.0
		Score	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**	**Neg**	**A**

[a] Percentage of the base calls generated from full-length tiled sequences.
[b] Taxonomic identification according to:
A = identification at the species or isolate level when an unique best hit corresponds to expected species or isolate
C = identification at the family level when multiple best viral hits exist and correspond to genera of the family *Rhabdoviridae*
Neg = negative or inaccurate identification when BLAST query is not possible or when multiple best hits exist and some or all of them correspond to other viral families.
In grey: results obtained using the sequence belonging to the same species or isolate tiled on the array (homonymous sequence).

JVI Accepts published online ahead of print

[c] TS : tentative species, according to the International Committee on Taxonomy of Viruses database (ICTVdb).
[d] Taxonomical classification according to (6).

JVI Accepts published online ahead of print

Table 5: Identification of virus species among the genus *Lyssavirus* based on lyssavirus sequences tiled on the PathogenID_v2.0 microarray and using the consensus sequence determination strategy.

Strains of lyssavirus tested			Strategy of analysis		
Genotype (Abbreviation)	**Strain tested**		**Use of the prototype sequence**	**Use of the consensus sequence (based all tiled sequences)**	**Used of consensus sequence (excluding the prototype sequence)**
1 (RABV)	93127FRA	Base call rate[a]	95.0	96.3	32.7
		BLAST score[b]	791	801	38
		Accuracy[c]	100	99.9	95.9
	8764THA	Base call rate	32.6	47.4	26.7
		BLAST score	46	64	31
		Accuracy	94.8	99.1	98.4
2 (LBV)	8619NIG	Base call rate	96.6	96.4	28.1
		BLAST score	816	814	39
		Accuracy	99.9	99.9	97.7
3 (MOKV)	86100CAM	Base call rate	56.3	67.4	28.4
		BLAST score	66	112	64
		Accuracy	98.2	99.8	98.5
4 (DUVV)	86132SA	Base call rate	97.3	97.3	18.1
		BLAST score	843	833	20
		Accuracy	99.9	99.8	96.4
5 (EBLV-1)	8918FRA	Base call rate	93.8	96.0	41.1
		BLAST score	757	807	83
		Accuracy	100	100	97.9
6 (EBLV-2)	9018HOL	Base call rate	98.4	98.8	26.8

JVI Accepts published online ahead of print

		BLAST score	871	879	44
		Accuracy	100	99.9	99.6
		Base call rate	94.9	95.6	29.7
7 (ABLV)	ABLV	BLAST score	749	741	40
		Accuracy	100	99.9	94.5
		Base call rate	NA	75.9	NA
8[e] (DBLV)	0406SEN	BLAST score	NA	82	NA
		Accuracy	NA	91.8	NA
		Base call rate	NA	60.9	NA
?	WCBV	BLAST score	NA	56	NA
		Accuracy	NA	97.3	NA

[a] Percentage of the base calls generated from full-length tiled sequences.
[b] BLAST score (bit score), obtained after BLAST query on a local viral and bacterial database using the consensus sequence determination strategy with $m = 12$ (minimum nucleotide length), $N = 10$ (maximum undetermined nucleotides content). Default BLAST parameters, except for the minimum length word length (7 nucleotides), the expect threshold (increased from the default of 10 to 100,000), and the 'low complexity level filter' (-F, turned off). All the BLAST scores indicate a correct identification at the species or isolate level (i.e., unique best hit corresponds to the expected species or isolate).
[c] Percentage of correct nucleotide identification, compared to the sequence obtained after classical sequencing of the corresponding lyssavirus species tested.
[d] NA: Not applicable.
[e] Tentative species.

JVI Accepts published online ahead of print

Table 6 : Description and final classification of rhabdovirus species used for phylogenetic analysis.

Genus and name[a]	UA/TS/UC[b]	Abbreviation	Reference no. (strain)	Princpal host species/vector[c]	Origin of samples	Year of first isolation	GenBank accession no.
Lyssavirus (Genotype)							
Rabies virus (Gt 1)		RABV	9001FRA	Dog bitten by a bat	French Guyana	1990	EU293113
Rabies virus (Gt 1)		RABV	9147FRA	Fox	France	1991	EU293115
Rabies virus (Gt 1)		RABV	8743THA	Human	Thailand	1983	EU293121
Rabies virus (Gt 1)		RABV	9704ARG	Bat : *Tadarida brasiliensis*	Argentina	1997	EU293116
Rabies virus (Gt 1)		RABV	9706CHI	Vaccine AG	China		AY854663
Rabies virus (Gt 1)		RABV	9702IND	Human	India	1997	AY854665
Lagos bat virus (Gt 2)		LBV	8619NGA	Bat : *Eidolon helvum*	Nigeria	1956	EU293110
Mokola virus (Gt 3)		MOKV	MOKV	Cat	Zimbabwe	1981	NC_006429
Mokola virus (Gt 3)		MOKV	86100CAM	Shrew	Cameroon	1974	EU293117
Mokola virus (Gt 3)		MOKV	86101RCA	Rodent	Republic of Central Africa	1981	EU293118
Duvenhage virus (Gt 4)		DUVV	94286SA	Bat : *Miniopterus* species	South Africa	1981	EU293120
Duvenhage virus (Gt 4)		DUVV	86132SA	Human	South Africa	1971	EU293119
European bat lyssavirus 1 (Gt 5)		EBLV-1	8918FRA	Bat : *Eptesicus serotinus*	France	1989	EU293112
European bat lyssavirus 1 (Gt 5)		EBLV-1	08120FRA	Bat : *Eptesicus serotinus*	France	2008	EU626551
European bat lyssavirus 2 (Gt 6)		EBLV-2	9018HOL	Bat : *Myotis dasycneme*	The Netherlands	1986	EU293114
European bat lyssavirus 2 (Gt 6)		EBLV-2	9337SWI	Bat : *Myotis*	Switzerland	1993	AY854657

JVI Accepts published online ahead of print

				daubentonii			
Australian bat lyssavirus (Gt 7)		ABLV	ABLh	Human	Australia	1986	AF418014
Australian bat lyssavirus (Gt 7)		ABLV	ABLb	Bat : *Pteropus* species	Australia	1996	NC_003243
Dakar bat lyssavirus (proposed Gt 8)	UC	DBLV	0406SEN (AnD 42443)	Bat : *Eidolon helvum*	Senegal	1985	EU293108
Dakar bat lyssavirus (proposed Gt 8)	UC	DBLV	KE131	Bat : *Eidolon helvum*	Kenya	2007	EU259198
Irkut virus (proposed Gt 9)	UC	IRKV		Bat : *Murina leucogaster*	Russia	2002	EF614260
Ozernoe virus (proposed Gt 9)	UC	IRKV		Human	Russia	2007	FJ905105
Aravan virus	UC	ARAV		Bat : *Myotis blythi*	Kyrgyzstan	1991	EF614259
Khujand virus	UC	KHUV		Bat : *Myotis mystacinus*	Tajikistan	2001	EF614261
West Caucasian bat virus	UC	WCBV		Bat : *Miniopterus schreibersii*	Russia	2002	EF614258
Vesiculovirus							
Chandipura virus		CHPV	I 653514	**Human** ; Domestic animals[d] ; Hedgehog : *Atelerix* species ; Dipterian : *Phlebotomus* species	India	1965	AJ810083
Cocal virus		COCV	TRVL 40233	Livestock : equine, bovine ; Mites : ***Gigantolaelaps*** species	Trinidad and Tobago: Trinidad	1961	EU373657
Isfahan virus		ISFV	91026-167	Dipteran : ***Phlebotomus papatasi***	Iran	1975	AJ810084
Piry virus		PIRYV	BeAn 24232 (0413BRE)	Human ; Opossum : ***Philander opossum***	Brazil	1960	GU816023

JVI Accepts published online ahead of print

Vesicular stomatitis New Jersey virus		VSNJV	VSV NJ-O	Livestock : Several species including *Bos taurus*, equine ; Dipterans : Several species including *Culex nigripalpus*, *Culicoides* species, *Mansonia indubitans*	Utah, USA	1949	AY074804
Vesicular stomatitis New Jersey virus		VSNJV	VSV NJ-H	Livestock : Several species including *Sus scrofa* ; Dipterans : Several species including *Culex nigripalpus*, *Culicoides* species, *Mansonia indubitans*	Georgia, USA	1952	AY074803
Vesicular stomatitis Indiana virus		VSIV	Mudd-Summers (MS)	Bovine : ***Bos taurus***	Indiana, USA	1925	EU849003
Vesicular stomatitis Indiana virus		VSIV	85CLB	**Bovine**	Colombia	1985	AF473865
Vesicular stomatitis Indiana virus		VSIV	98COE	**Equine**	Colorado, USA	1998	AF473864
Vesicular stomatitis Alagoas virus		VSAV	Indiana 3	Livestock : equine (**mule**), *Bos taurus* ; Dipterans : *Phlebotomus* species	Brazil	1964	EU373658
Jurona virus[e]	TS	JURV	BeAr 40578 (0414BRE)	Dipteran : ***Haemagogus spegazzinii***	Brazil	1962	GU816024

JVI Accepts published online ahead of print

Perinet virus	TS	PERV	Ar Mg 802	Dipterans : *Anopheles coustani*, ***Culex antennatus***, *Culex* gr. *pipiens*, *Mansonia uniformis*, *Phlebotomus berentens*is	Madagascar	1978	AY854652
Pike fry rhabdovirus	TS	PFRV	F4	Fish : ***Esox lucius***	The Netherlands	1972	FJ872827
Scophthalmus maximus rhabdovirus	UC	SMRV	QZ-2005	Fish : ***Scophthalmus maximus***	China	?	AY895167
Spring viremia of carp virus	TS	SVCV	Fijan_cell (VR-1390, isolated from fat head minnow cells)	Fish : ***Cyprinus carpio***	Yougoslavia	1971	AJ318079
Spring viremia of carp virus	TS	SVCV	Fijan_tissue (VR-1390, isolated from tissues of diseased common carp)	Fish : ***Cyprinus carpio***	Yougoslavia	1971	U18101
Spring viremia of carp virus	TS	SVCV	BJ0505-2	Fish : ***Cyprinus carpio***	China	2005	EU177782
Ephemerovirus							
Adelaide River virus		ARV	DPP 61	Bovine : ***Bos taurus***	Australia	1981	AY854635
Berrimah virus		BRMV	DPP 63	Bovine : ***Bos taurus***	Australia	1981	AY854636
Bovine ephemeral fever virus		BEFV	Cs 1933	Bovine : ***Bos taurus***	Australia	1973	AY854641
Bovine ephemeral fever virus		BEFV	Cs 42	Dipteran : ***Anopheles bancrofti***	Australia	1975	AY854639
Bovine ephemeral fever virus		BEFV	BB7721	Bovine : ***Bos taurus***	Australia	1968	NC_002526
Kimberley virus	TS	KIMV	CS 368	Bovine : ***Bos taurus***	Australia	1980	AY854637
Kotonkan virus	UA	KOTV	IbAr 23380	Dipteran : ***Culicoides***	Nigeria	1967	AY854638

JVI Accepts published online ahead of print

				species			
Almpiwar group							
Almpiwar virus	UA	ALMV	MRM 4059	Mammals[d] : bovine, equine, ovine, kangaroo, bandicoot, human ; Birds[d] ; Lizard : ***Ablepharus boutonii virgatus*** and other shink[d]	Australia	1966	AY854645
Charleville virus	UA	CHVV	Ch 9824	Human[d] ; Dipteran : ***Phlebotomus*** and *Lasiohelea* species	Australia	1969	AY854644
Charleville virus	UA	CHVV	Ch 9847	Human[d] ; Dipteran : ***Phlebotomus*** and *Lasiohelea* species	Australia	1969	AY854672
Humpty Doo virus	UA	HDOOV	CS 79	Dipteran : ***Lasiohelea*** species, *Culicoides marksi*	Australia	1975	AY854643
Hart Park group							
Bangoran virus[e]	UA	BGNV	DakArB 2053 (0424RCA)	Bird : *Turdus libonyanus* ; Dipteran : ***Culex perfuscus***	Central African Republic	1969	GU816010
Flanders virus	UA	FLANV	61-7484	Birds : *Seiurus aurocapillus, Agelaius phoeniceus* ; Dipterans : ***Culiseta***	New York, USA	1961	AF523199

JVI Accepts published online ahead of print

				melanura, *Culex* species			
Kamese virus[e]	UA	KAMV	MP 6186 (08343RCA)	Dipterans : *Aedes africanus, Culex* species including ***Culex annulioris***	Uganda, Republic of Central Africa	1967	GU816011
Mossuril virus[e]	UA	MOSV	SA Ar 1995 (0418SA)	Birds : *Andropadus virens, Coliuspasser macrourus* ; Dipterans : *Aedes abnormalis, Culex* species including ***Culex sitiens***	Mozambique, Republic of Central Africa, South Africa, Bostwana, Guinea	1959	GU816012
Ngaingan virus	UA	NGAV	MRM 14556	Mammals[d] : wallabies, kangaroo, bovine ; Dipteran : ***Culicoides brevitarsis***	Australia	1970	AY854649
Parry Creek virus	UA	PCRV	OR 189	Dipteran : ***Culex annulirostris***	Australia	1972	AY854647
Porton's virus[e]	TS (VSV)	PORV	1643 (0416MAL)	Dipteran : ***Mansonia uniformis***	Malaysia (Sarawak)	?	GU816013
Wongabel virus	UA	WONV	CS 264	Sea birds[d] ; Dipteran : ***Culicoides austropalpalis***	Australia	1979	AY854648
Le Dantec group							
Fukuoka virus	UA (Kern Canyon Group)	FUKV	FUK-11	Bovine ; Dipteran : ***Culicoides punctatus***, *Culex tritaeniorhynchus*	Japan	1982	AY854651
Keuraliba virus[e]	UA	KEUV	DakAnD 5314	Rodents : *Tatera*	Senegal	1968	GU816021

JVI Accepts published online ahead of print

			(9715SEN, 0420SEN)	species including ***Tatera kempi***, *Taterillus* species			
Le Dantec virus	UA	LDV	DakHD 763	**Human**	Senegal	1965	AY854650
Nkolbisson virus[e]	UA (Kern Canyon Group)	NKOV	Ar YM 31/65 (0425CAM)	Dipteran : *Aedes* species, *Eretmapodites* species including ***Eretmapodites leucopus***, *Culex telesilla*)	Ivory Coast, Cameroon	1965	GU816022
Moussa group							
Moussa virus	UC	MOUSV	C23	Dipteran : ***Culex decens***	Ivory Coast	2004	FJ985748
Moussa virus	UC	MOUSV	D24	Dipteran : ***Culex*** species	Ivory Coast	2004	FJ985749
Sandjimba group							
Bimbo virus[e]	UA	BBOV	DakAnB 1054d (9716RCA)	Bird : ***Euplectes afra***	Central African Republic	1970	GU816016
Boteke virus[e]	TS (VSV)	BTKV	DakArB 1077 (0417RCA)	Dipteran : ***Coquillettidia maculipennis***	Central African Republic	1968	GU816014
Garba virus[e]	UA	GARV	DakAnB 439a (0422RCA)	Birds : ***Corythornis cristata***, *Nectarina pulchella*	Central African Republic	1970	GU816018
Kolongo virus	UA	KOLV	DakAnB 1094d (9717RCA)	Birds : ***Euplectes afra***, *Ploceus cucullatus*	Central African Republic	1970	GU816020
Nasoule virus[e]	UA	NASV	DakAnB 4289a (0410RCA)	Bird : ***Andropadus virens***	Central African	1973	GU816017

JVI Accepts published online ahead of print

Oak-Vale virus	UA	OVRV	CS 1342	Ferral pigs[d] ; Dipteran : *Aedes vigilax, Culex* species including ***Culex edwardsi***)	Republic Australia	1981	AY854670
Ouango virus[e]	UA	OUAV	DakAnB 1582a (9718RCA)	Bird : ***Ploceus melanocephalus***	Central African Republic	1970	GU816015
Sandjimba virus	UA	SJAV	DakAnB 373d (07244RCA)	Bird : ***Acrocephalus schoenobaenus***	Central African Republic	1970	GU816019
Sigma group							
Drosophila affinis sigma virus	UC	DAffSV	10	Dipterian : ***Drosophila affinis***	New Connecticut, USA	2007	GQ410980
Drosophila melanogaster sigma virus	UA	SIGMA V (DMelS V)	AP30	Dipterian : ***Drosophila melanogaster***	Florida, USA	2005	NC_013135
Drosophila melanogaster sigma virus	UA	SIGMA V (DMelS V)	HAP23	Dipterian : ***Drosophila melanogaster***	France	?	GQ375258
Drosophila obscura sigma virus	UC	DObsSV	10A	Dipterian : ***Drosophila obscura***	UK	2007	GQ410979
Sinistar group							
Siniperca chuatsi rhabdovirus	UC	SCRV		Fish : ***Siniperca chuatsi***	China	?	NC_008514
Starry flounder rhabdovirus	UC	SFRV		Fish : ***Platichthys***	Washington,	2000	AY450644

JVI Accepts published online ahead of print

				stellatus	USA		
Tibrogargan group							
Tibrogargan virus	UA	TIBV	CS 132	Bovine[d] : water buffaloes, cattle ; Dipteran : ***Culicoides brevitarsis***	Australia, New Guinea	1976	AY854646
Tupaia virus	TS (VSV)	TUPV	TRV 1591	Tree shrew *:* ***Tupaia belangeri***	Thailand	?	NC_007020
Novirhabdovirus							
Hirame rhabdovirus		HIRRV	CA 9703	Fish : including **Cultured korean flounders**, *Paralichthys olivaceus*, *Plecoglossus altivelis*, *Milio macrocephalus*, *Sebastes inermis*	Korea	1997	NC_005093
Infectious hematopoietic necrosis virus		IHNV	HV7601				AB231660
Infectious hematopoietic necrosis virus		IHNV	WRAC strain	Fish : including salmonid, ***Oncorhynchus tschawytscha***	Idaho, USA		NC_001652
Snakehead rhabdovirus		SHRV		Fish : including ***Ophicephalus striatus***, *Clarias bratachus*, *Oxyeleotis marmoratus*	Thailand		NC_000903
Viral hemorrhagic septicemia virus		VHSV	KRRV9822	Fish : **Japanese flounder**	Japan		AB179621
Viral hemorrhagic septicemia virus		VHSV	07-71	Fish : ***Oncorhynchus mykiss***	France		AJ233396

JVI Accepts published online ahead of print

Viral hemorrhagic septicemia virus		VHSV	JF00Ehi1	Fish : ***Paralichthys olivaceus***	Japan	2000	AB490792
Viral hemorrhagic septicemia virus		VHSV	14-58	Fish : ***Oncorhynchus mykiss***	France		AF143863
Nucleorhabdovirus							
Maize mosaic virus		MMV		Plant (host) : *Graminae* including ***Zea mays*** ; Hemipterans (vector) : *Delphacidae*	USA		NC_005975
Rice yellow stunt virus		RYSV		Plant (host) : ***Oryza sativa*** ; Homopterans (vector) : *Cicadellidae*			NC_003746
Sonchus yellow net virus		SYNV		Plant (host) : *Asteraceae* including ***Sonchus oleraceus*** ; Hemipterans (vector) : *Aphididae*			NC_001615
Iranian maize mosaic nucleorhabdovirus	UC	IMMNV		Plant (host) : *Graminae* including ***Zea mays*** ; Hemipterans (vector) : *Delphacidae*	Iran		NC_011542
Maize fine streak virus	UC	MFSV		Plant (host) : *Graminae* including ***Zea mays*** ; Homopterans (vector) : *Cicadellidae*	Georgia, USA	1999	NC_005974
Orchid fleck virus[f]	UC	OFV	So	Plant (host) : *Orchidaceae* including ***Cymbidium* species** ; Acaris (vector) : *Brevipalpus*	Japan		NC_009609

JVI Accepts published online ahead of print

				californicus Plant (host) : ***Colocasia esculenta***	Fiji Islands	NC_006942
Cytorhabdovirus						
Barley yellow striate mosaic		BYSMV	Zanjan-1	Plant (host) : *Graminae* including ***Triticum* species** ; Hemipterans (vector) : *Delphacidae*	Iran	FJ665628
Lettuce necrotic yellows virus		LNYV	318	Plant (host) : several plant families and species including *Allium sativum,* ***Lactuca sativa*** ; Hemipterans (vector) : *Aphididae*	Australia	NC_007642
Northern cereal mosaic virus		NCMV		Plant (host) : *Graminae* including ***Hordeum vulgare*** ; Hemipterans (vector) : *Delphacidae*	Japan	NC_002251
Strawberry crinkle virus		SCV	HB-A1	Plant (host) : ***Fragaria* species** ; Hemipterans (vector) : *Aphididae*		AY331389
Strawberry crinkle virus		SCV	37-2	Plant (host) : ***Fragaria* species** ; Hemipterans (vector) : *Aphididae*		AY331388
Strawberry crinkle virus		SCV	37-1	Plant (host) : ***Fragaria* species** ;		AY331387

JVI Accepts published online ahead of print

Lettuce yellow mottle virus	UC	LYMoV	Hemipterans (vector) : *Aphididae* Plant (host) : ***Lactuca sativa***	France	1998	NC_011532
Taastrup group						
Taastrup virus	UC	TV	Hemipteran (potential vector) : ***Psammotettix alienus***	France	1996	AY423355

[a] Names of viruses in italic correspond to approved virus species by the eighth report of the International Committee on Taxonomy of Viruses database (ICTVdb).
[b] UA : unassigned, TS : tentative and UC : unclassified (not found in the ICTV) species.
[c] In bold : identification of the host species where the first viral isolation was obtained when data are available.
[d] Serological detection only.
[e] First identification based on nucleic acid determination and classification based on phylogenic analysis (this study).
[f] Also tentatively classified into the new genus *Dichorhabdovirus* according to its unusual bipartite genome (20).