

Genomics and structure/function studies of Rhabdoviridae proteins involved in replication and transcription.

R. Assenberg, O Delmas, B Morin, C Graham, X de Lamballerie, C Laubert, B Coutard, J Grimes, J Neyts, R J Owens, et al.

▶ **To cite this version:**

R. Assenberg, O Delmas, B Morin, C Graham, X de Lamballerie, et al.. Genomics and structure/function studies of Rhabdoviridae proteins involved in replication and transcription.. Antiviral Research, 2010, 87 (2), pp.149-61. 10.1016/j.antiviral.2010.02.322 . pasteur-01492926

HAL Id: pasteur-01492926
https://pasteur.hal.science/pasteur-01492926

Submitted on 21 Apr 2017

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.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Genomics and structure/function studies of Rhabdoviridae proteins involved in replication and transcription

R. Assenberg[1], O. Delmas[2], B. Morin[3], S. C. Graham[1], X. De Lamballerie[4], C. Laubert[5], B. Coutard[3], J. M. Grimes[1], J. Neyts[6], R. J. Owens[1], B. W. Brandt[5], A. Gorbalenya[5] , P. Tucker[7], D. I. Stuart[1] , B. Canard[3] , H. Bourhy[2]

[1]Division of Structural Biology and Oxford Protein Production Facility, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK

[2]Institut Pasteur, Unité Dynamique des lyssavirus et adaptation à l'hôte, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France

[3]Architecture et Fonction des Macromolécules Biologiques, CNRS and Universités d'Aix-Marseille I et II, UMR 6098, ESIL Case 925, 13288 Marseille, France

[4]UMR190, Emergence des Pathologies Virales, Institut de Recherche pour le Développement - Université de la Méditerranée, Unité des Virus Emergents, Faculté de Médecine de Marseille, 27, Bd Jean Moulin, 13005 Marseille cedex 05

[5]Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands

[6]Rega Institute for Medical Research, KULeuven, Leuven, Belgium.

[7]EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D22603 Hamburg, Germany

Corresponding author:

Hervé Bourhy,

Institut Pasteur, Unité Dynamique des lyssavirus et adaptation à l'hôte, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France,

hbourhy@pasteur.fr

tel: 33.145688785

Fax: 33.140613020

- **Abstract**

Mammals, and in many cases humans, can be infected by viruses of most of the major mononegavirus lineages, representing a vast reservoir of human pathogens. Some mammalian rhabdoviruses may also infect invertebrates, dogs and bats which may work as vectors transmitting viruses among different host species. The VIZIER programme, an EU funded FP6 program has characterized viruses that belong to Vesiculovirus, Ephemerovirus and Lyssavirus genera of the *Rhabdoviridae* family to perform ground-breaking impact on the identification of potential new drug targets against these RNA viruses through comprehensive structural characterization of the replicative machinery. The contribution of VIZIER programme was of several orders. First, it contributed substantially to research aimed at understanding the origin, evolution and diversity of rhabdoviruses. This diversity was then used to obtain further structural information on the proteins involved in the replication of rhabdoviruses. Two strategies were used to produce recombinant proteins by expression of both full length or domain constructs in either *E. coli* or insect cells using the baculovirus system. In both cases parallel cloning and expression screening at small-scale of multiple constructs based on different viruses including the addition of fusion tags, was key to the rapid generation of expression data. As a result, some progress has been made in the VIZIER Programme towards dissecting the multi-functional L protein into components suitable for structural and functional studies. However, the phosphoprotein polymerase co-factor and the structural matrix protein which plays a number of roles during viral replication and drives viral assembly have both proved much more amenable to structural biology. Applying the multi-construct /multi-virus approach central to the protein production processes in VIZIER has yielded new structural information which may ultimately be exploitable in the derivation of novel ways of intervening in viral replication.

Keywords

Rhabdovirus, replication, evolution, diversity, structural studies.

1. Introduction

1.1 The VIZIER program

The aim of the VIZIER program, an EU funded FP6 programme, VIZIER, was to characterize RNA viruses, including strains of medical interest and the core enzymes/proteins of their replication machinery (Coutard et al., 2008). Among other RNA viruses, this program focused on rhabdoviruses. The final aim of the program was to allow identification of potential new drug targets against these RNA viruses through comprehensive structural characterization of the replicative machinery. In this paper we will review the recent developments in that topic with special attention on those obtained in the framework of the VIZIER program.

1.2. Taxonomy of the *Rhabdoviridae*

The Order *Mononegavirales* was the first created in the Virus Taxonomy. It formally recognized remarkable relationships that were documented over the years between three families of negative-stranded RNA viruses, including the *Rhabdoviridae, Paramyxoviridae* and *Filoviridae* (Pringle, 1997). Subsequently, the order has been expanded (and is still growing) to include the fourth family, *Bornaviridae*, and accommodate a number of newly described viruses that prototype deeply separated phylogenetic lineages mostly in the founding families (Figure 1).

Several dozens of species unevenly populate major lineages of the Mononegavirales. Two most profoundly separated lineages of the Paramyxoviridae were recognized as subfamilies, Paramyxovirinae and Pneumovirinae including, respectively, five and two genera. The *Rhabdoviridae* family seems to include the largest diversity of viruses that was partitioned into six genera. In contrast, the *Bornaviridae* and *Filoviridae* families are very compact and include one and two genera, respectively. Most recently, two closely related viruses were described that prototype a separate lineage most closely related to bornaviruses and provisionally dubbed the Nyavirus genus (Mihindukulasuriya et al., 2009). Mammals, and in many cases humans, can be infected by viruses of all major mononegavirus lineages excluding three Rhabdovirus genera, *Nucleorhabdovirus* and *Cytorhabdovirus* that infect plants, and *Novirhabdovirus* infecting fish. Some mammalian rhabdoviruses may also infect invertebrates which may work as vectors transmitting viruses among different host species

(Bourhy et al., 2008b). The VIZIER programme has characterized viruses that belong to *Vesiculovirus*, *Ephemerovirus* and *Lyssavirus* genera of the *Rhabdoviridae* family.

Rhabdoviruses infect a wide range of mammals, including man, and transmission is commonly vector-borne, particularly by haematophagous insects. The relationships between many of these unassigned rhabdoviruses have been determined based on serological cross reactions and, more recently, according to phylogenetic analysis performed on partial sequences of the polymerase and nucleoprotein genes. Six serogroups (Hart Park group, Le Dantec group, Bahia Grande group, Timbo group, Sawgrass group and Kern Canyon group) have been distinguished, three of them being supported as independent clades by the phylogenetic analysis. Furthermore, based on the phylogeny, 4 other groups can be proposed: Almpiwar group, Tibrogargan group, Mount Elgon bat group, and Kolongo and Sandjimba group (Figure 2) (Bourhy et al., 2005). It is proposed to name all vesiculoviruses, ephemeroviruses and several of the proposed groups infecting mammals and mosquitoes which share common phylogenetic relationships as the Dimarhabdovirus supergroup (Bourhy et al., 2005). There is a constantly growing list of rhabdoviruses (presently 85), isolated from a variety of vertebrate or invertebrate hosts, that are partially characterized and are still waiting for definitive species assignment (Bourhy et al., 2008a).

Currently, there are seven recognised genotypes (GT) or species of lyssavirus defined on the basis of their genetic similarity: rabies virus (RABV, GT1) responsible for classical rabies in terrestrial mammals globally and in bats on the American continent, as well as the cause of most rabies-related human deaths worldwide; Lagos bat virus (LBV, GT2); Mokola virus (MOKV, GT3); Duvenhage virus (DUVV, GT4); European bat lyssavirus type 1 (EBLV-1, GT5); European bat lyssavirus type 2 (EBLV-2, GT 6); and Australian bat lyssavirus (ABLV, GT7). Additionally, four new lyssavirus genotypes that infect bats in central and southeast Asia have been proposed: Aravan virus, Khujand virus, Irkut virus and West Caucasian Bat virus (Arai et al., 2003, Botvinkin et al., 2003, Kuzmin et al., 2005).

1.3. **Organization of the genome**

All mononegaviruses employ a monopartite genome whose size varies two-fold, from ~9 kb (Bornaviridae) to 19 kb (Filoviridae) excepted one recently characterized plant virus, the orchid fleck virus, which exhibits a bipartite genome (Kondo et al., 2006). The antigenome replica of mononegavirales encodes several positionally conserved genes located in separate open reading frames (ORFs) that may be interspersed by genes unique to distinct lineages.

The order of the backbone genes from the 3'- to 5'- ends of the genome is N-P-M-G-L, where N is nucleoprotein, P – phosphoprotein, M – matrix protein, G – glycoprotein and L – large protein also known as polymerase. The products of the first four genes produce major proteins forming enveloped virions and they may be known under other names in some viruses (see below). Additional ORFs are located between the phosphoprotein and the matrix protein genes and between the glycoprotein and polymerase genes (Figure 3). The L gene encodes a multidomain protein, including putative RNA-dependent RNA polymerase, that is virion-associated and mediates replication and expression of the virus genome.

1.4. Virion Structure

Rhabdoviruses, like other Mononegavirales, are enveloped viruses with a lipid bilayer envelope generated from the plasma membrane of the infected host cell. Rhabdoviruses virions are rod- or bullet-shaped particles approximately 100 to 430 nm long and 45 to 100 nm in diameter.

Extending from the surface of the membrane is one glycoprotein (G) inserted into the envelope. Under and associated to the membrane by hydrophobic and electrostatic interactions is a layer formed by the matrix protein (M) which shares conserved tertiary structure within Mononegavirales families but not between them (Dessen et al., 2000, Graham et al., 2008, Money et al., 2009, Ruigrok et al., 2000, Timmins et al., 2004). The M protein condenses the nucleocapsid and gives the lyssavirus virion its bullet-shaped appearance. In addition to interacting with nucleocapsid, M protein also associates with lipid bilayer (Dancho et al., 2009, Manie et al., 2000, Solon et al., 2005) and the glycoprotein (Mebatsion et al., 1999) suggesting M protein could make a link between nucleocapsid and the viral envelope containing glycoproteins. Inside the particle is the helical ribonucleoproteic core consisting of the negative single-stranded RNA wrapped by the nucleoprotein (N). In Rhabdoviruses, the N protein is composed of two main domains and two smaller extra domains extending from the N- and C-termini. The RNA wraps around the main N-terminal part of the N with each N protomer making contact with 9 nucleotides. The bound RNA is totally enclosed by the N with its C-terminal main domain covering the bound RNA. Additional contacts between neighboring N protomers participate in the polymerization of N leading to the formation of the nucleocapsid (Albertini et al., 2006).

The nucleocapsid is associated with lesser amounts of two others proteins: the phosphoprotein (P), and the large protein (L). These two proteins are responsible for the RNA polymerase

activity associated with the viral particles. L proteins carry all the enzymatic RNA polymerase activities whilst the P links L to the N-RNA complex (Bourhis et al., 2006, Gerard et al., 2009).

The P protein has a modular organization with three structured domains alternating with two disordered domains (Gerard et al., 2009). The N-terminal ordered domain contains a site of interaction with nude N, the first disordered region contains phosphorylation sites that may be involved in regulation of transcription, the central oligomerization domain, a second disordered region that in rabies at least is able to interact with cellular partners, and finally the C-terminal ordered region that interacts with N-RNA complexes.

The L protein is poorly characterized. It is thought to perform RNA synthesis as well as the mRNA capping and polyadenylation activities (Grdzelishvili et al., 2005, Li et al., 2006, Ogino et al., 2005). Bioinformatics studies show that the rhabdovirus L protein is about 220-250 kDa and has six highly conserved domains proposed to be individually responsible for each of the multiple L functions (Poch et al., 1990, Svenda et al., 1997). Expression and purification of L has been elusive so far and, consequently, very little information is available about its enzyme activities or structural organisation.

2. Life cycle

The replication cycle of rhabdovirus involves the same general steps. After the first steps of attachment, penetration and uncoating, the nucleocapsid and all the components necessary for the early transcription are released in the cytoplasm of the infected cell. The surface glycoprotein mediates the attachment of the virion. After attachment, viral particles are endocytosed. The pH decreasing along the endocytosis pathway induces a change in the conformation of the G protein that becomes a fusion protein (Roche et al., 2006, Roche et al., 2007). The G in its fusion conformation mediates the mixing of the endosome membrane and the virus membrane that allows the release of the nucleocapsid into the cytoplasm. Membrane fusion may occur with an internal vesicle of the multivesicular body. The release of the nucleocapsid in the cytoplasm then involves a back-fusion step between the internal vesicle and the membrane that surrounds the multivesicular body (Luyet et al., 2008). The negative sense ribonucleoproteic complex is then released into the cytoplasm and used as a template for the primary transcription. For transcription the viral polymerase does a stop-start mechanism with a single entry point at the 3' extremity of the genome. The viral polymerase produces first the short leader RNA that is not capped nor polyadenylated. It restarts with the

transcription of the first gene on the genome i.e. the nucleoprotein gene. This transcript is capped and polyadenylated by the viral polymerase as are all the other viral transcripts. At the end of the nucleoprotein gene the viral polymerase starts again and so on until the end of the last gene (Figure 4). There is a gradient in the quantity of each transcript depending on its order and its distance from the 3' end of the genome. Indeed, at each stop a fraction of the polymerase dissociates from the RNA template; in this way regulating the relative abundance of each viral protein. The necessity for efficient infection to regulate the relative abundance of viral proteins and the conservation of transcription mechanism among mononegavirales probably explains the conservation of the basic gene order (Figure 3).

After accumulation of neo-synthesized viral proteins, the RdRp switches to a processive mode and ignores gene junctions to synthesize a full-length positive, complementary copy called the antigenome. The levels of the N protein mediate the switch between transcription and replication. When the amount of N protein is sufficiently high to allow RNA encapsidation of the nascent RNA chain, the polymerase switches to the replication mode (Arnheiter et al., 1985, Fearns et al., 1997, Plumet et al., 2005, Vidal & Kolakofsky, 1989). The stochiometry of L and P proteins in the viral polymerase complex is modified (Gupta et al., 2003) consistent with the fact that the viral polymerase does not respond to the same signals during transcription and replication. The amounts of rhabdovirus genomes and antigenomes are equivalent and both can be encapsidated. By a mechanism similar to that of positive strand synthesis, the antigenome is used as a template to direct synthesis of encapsidated minus-strand genome (Leppert et al., 1979). This latter strand can be used for three purposes: as a template for synthesis of both new mRNAs (in a process referred to as secondary transcription) and antigenomic RNA. Finally, genomic RNA can be incorporated into progeny virions during the budding process.

After mRNA translation, cellular chaperones are used to transport the M protein to the plasma membrane and glycoproteins from the ER to the Golgi apparatus to the plasma membrane. Like other enveloped viruses, a budding process forms virions and involves cooperation of viral and cellular proteins. The viral components are assembled near the plasma membrane where new virus vesicles bud from sites on the host cell membrane.

3. **Pathogenesis**

3.1. Introduction

RNA viruses of the family *Rhabdoviridae* comprise arthropod-borne agents that infect plants, fish and mammals, as well as a variety of non-vector-borne mammalian viruses. The *Rhabdoviridae* family presently comprises six genera, and members of three of these genera – *Vesiculovirus*, *Lyssavirus* and *Ephemerovirus* – have been obtained from a variety of animal hosts and vectors, including mammals, fish and invertebrates. The remaining three rhabdovirus genera are more taxon-specific in their host preference. Novirhabdoviruses infect numerous species of fish, while cytorhabdoviruses and nucleorhabdoviruses are arthropod-borne and infect plants (Bourhy et al., 2005, Bourhy et al., 2008a). Although vesiculoviruses have often be used as a model to study experimentally the replication and transcription mechanisms in rhabdoviruses, we will here focus on the lyssaviruses which are the only etiological agents of human diseases in the rhabdovirus family.

3.2. Lyssavirus

Lyssaviruses cause rabies, an acute encephalomyelitis transmitted to humans by rabid animals. Although effective economical control measures are available and although WHO and OIE consider it a high priority zoonosis, rabies unfortunately remains a neglected disease in a large part of the world, especially in Africa and Asia (Knobel et al., 2005). Rabies is transmitted by a bite, scratch or lick on damaged skin or by projection of infectious material (i.e. saliva or lacrymal liquid) on mucosae. Except in some rare cases of transmission by organ or tissue transplantation, human to human transmission has never been described. Rabies virus is highly neurotropic, and after inoculation it is retrogradely transported by peripheral neurons before being passed on to second and higher-order neurons without being taken up by glia (Kelly & Strick, 2000). The glycoprotein G is known to be the only protein component of the viral envelope that mediates viral entry into host cells (Etessami et al., 2000). Several receptors have been shown to play a role in cell attachement and entry: NCAM (Thoulouze et al., 1998), P75NTR (Tuffereau et al., 1998, Tuffereau et al., 2001) and acetyl choline receptor (Lentz et al., 1982). However other mechanisms are probably responsible for the tropism of the virus for neuronal cells (Lafon, 2005, Langevin et al., 2002, Tuffereau et al., 2007). The infection is responsible for high neurological disorders and is invariably fatal

in the absence of timely administration of a post-exposure prophylaxis consisting of several doses of vaccines and of a passive immunoprophylaxis (Anonymous, 2005). However, when the symptoms appear post-exposure prophylaxis is no longer active and no specific therapy is available (Jackson, 2009). Each year, at least 15 million people receive treatment after being exposed to suspected rabid animals (Bourhy et al., 2009); however, 55,000 people still die in Asia and Africa, based on the WHO estimation (Knobel et al., 2005). The main reservoir of the disease is the dog but several other species of the Carnivora and Chiroptera orders act also as reservoir of different species (or genotypes) of lyssaviruses.

4. Molecular epidemiology

4.1. Introduction

Rapid evolution and consequently high diversity and fitness, are the major driving forces of virus virulence, tropism, host range, transmission, and thus perpetuation of mononegavirales. An important determinant of this genetic variability is the error-prone nature of RNA-dependent RNA polymerases which, during virus replication, may result in high rates of mutation, sequence deletion, insertion, and recombination between RNA genomes. Due to the absence of efficient proofreading and post-replicative repair activities associated with RNA replicases, mutation rates of RNA viruses have been estimated to vary between approximately 10^{-3} to 10^{-5} substitutions per nucleotide copy. As a results, viral population are constantly shaped by evolutionary driven forces and genetic diversification is a common process in virus evolution. In the long term this leads to the segregation of new variants, species or genus. The VIZIER programme contributed substantially to research aiming at understanding the origin, evolution and diversity of lyssaviruses.

4.2. Origin, evolution of lyssavirus

Phylogenetic analyses of lyssaviruses have determined the existence of 7 genotypes or species, although this number is likely to increase with more intensive sampling (Bourhy et al., 1993, Gould et al., 1998, Kuzmin et al., 2005). Genomic and evolutionary studies have most often utilized partial genome sequences, particularly of the nucleoprotein and glycoprotein genes, with little consideration of genome-scale evolution.

Collaborative work obtained in the framework of VIZIER reported the first genomic and evolutionary analysis using complete genome sequences of all recognised lyssavirus genotypes, including 14 new complete genomes of field isolates from 6 genotypes and one genotype that was completely sequenced for the first time (Figure 5) (Delmas et al., 2008). This analysis revealed that all lyssaviruses have the same genomic organization (Figure 3). It also identified strong geographical structuring, with the greatest genetic diversity in Africa, and an independent origin for the two known genotypes that infect European bats. It also suggested that multiple genotypes may exist within the diversity of viruses currently classified as 'Lagos Bat'. This rigorous phylogenetic analysis based on full length genome sequence provides the best discriminatory power for genotype classification within the lyssaviruses.

Of the mammalian RABV, those that circulate in dogs (*Canis lupus familiaris*) are responsible for more than 99% of the human cases worldwide (Knobel et al., 2005). However, despite its role as a vector for human disease, the extent and structure of viral biodiversity in this key vector species, as well as the mode and timescale of its evolution, have only been studied on a limited geographical scale. To address these questions on a global scale, a comparative analysis of RABV gene sequence data was performed in the framework of VIZIER. From this, we identified six clades of RABV in non-flying mammals, each of which has a distinct geographical distribution, most likely reflecting major physical barriers to gene flow. Indeed, a detailed analysis of phylogeographic structure revealed slow and only limited viral movement among continents, countries and geographical localities (Bourhy et al., 2009). The distribution of the different clades and their relationships are shown in Figure 6. Using Bayesian coalescent methods, this analysis revealed that the sampled lineages of canid RABV derived from a common ancestor that originated within the past 1500 years probably from the indian subcontinent (Figure 6). Additionally, no evidence was found for either positive selection or widespread population bottlenecks during the global expansion of canid RABV. further supporting that the stochastic processes of genetic drift and population subdivision are the most important factors shaping the global phylogeography of canid RABV.

4.3. Emergence of new variants adapted to new hosts

Determining the genetic basis of the traits that govern cross-species transmission also clearly represents a major goal for future research on rhabdovirus, although it is important to note that patterns of cross-species transmission may also be in part determined by the ecological factors that shape host contact rates.

During virus replication, high mutation frequencies result in the generation of virus populations in which the majority of genomes differ by more than a single nucleotide and can therefore be defined as a spectrum of mutants derived from dominant parent copies. The word quasispecies (also sometimes referred to as population polymorphism or intra-host variation) was introduced to describe the resulting heterogeneous population structure generated by RNA viruses (Eigen & Schuster, 1979). This intra-host variation provides the virus with the capacity to adapt immediately to changing environmental circumstances, including the changing of hosts as demonstrated for lyssaviruses (Benmansour et al., 1992, Bourhy et al., 1999, Morimoto et al., 1998).

Rabies viruses are able to establish productive infections in dogs but also in many other host species (Nadin-Davis et al., 1994). Rabies viruses sampled from other species of the *Canidae* family, such as foxes and raccoon dogs, as well as hosts belonging to other families within the *Carnivora* order – the *Herpestidae* in southern Africa and the *Mephitidae* (skunks) in America – are interspersed within the phylogenetic diversity of dog RABV. While we found no significant evidence for adaptive evolution, our observation strongly suggests that the dog has served as the main vector for inter-species RABV transmission, generating viral lineages that then spread to other taxa (Bourhy et al., 2009).

An important example of such a successful host switch involved the transfer of the virus from dogs to the red fox (*Vulpes vulpes*) in Northeast Europe during the 1930s (Bourhy et al., 1999). After the initial cross-species transmission event, rabies virus was able to spread rapidly westward and southward through European red fox populations in the subsequent 60 years.

5. Progress towards the control using antivirals

5.1. Introduction

Considering, the rapid evolution, periodic emergence, capacity of spread and change of host and tropism of *Mononegavirales*, and in particular of rhabdoviruses, there is a critical need for broad spectrum antiviral molecules. We will here review the molecules which have shown some effects together with the step of the viral multiplication that they are targeting, our aim being to focus more on the structure-based design of small molecule inhibitors of viral replication.

5.2. Current state of the art of antivirals against *Mononegavirales*

Concerning rabies and lyssaviruses, several regimens of post exposure prophylaxis (PEP) have been developed and validated by WHO. They mainly consist of several injections of vaccine associated with a passive immunotherapy in the case of the most severe wounds (Anonymous, 2005). The passive immunotherapy is based on human or equine immunoglobulins or F(ab')2 fragments. A cocktail of human monoclonal antibodies recently developed is under study (Bakker et al., 2005, Bakker et al., 2008, de Kruif et al., 2007) and will probably be on the market in 2013. Considering the availability of efficacious biologicals to fight against rabies, the reader might be surprised by the fact that rabies is re-emerging in many countries and that the incidence in humans remains globally high. Beside infrastructure and organisational reasons that will not be detailed here, several explanations may be discussed (Wilde, 2007). First, there is a global lack of accessibility to these products to populations at risk in developing countries. Second, this PEP is also becoming less and less effective as the delay in days between the start of the PEP and the exposure is increasing. Third, the PEP is no longer effective after the onset of the symptoms. Therefore, a lot of efforts have been made to develop a curative treatment active in rabid patients, but so far no effective therapy as been found despite some proposed protocols (Willoughby, 2007, Willoughby et al., 2005) which later became highly controversial (Hemachudha et al., 2006, Jackson et al., 2008). One other main reason is probably that very few molecules have been so far reported to exert activity against the rabies virus. Ribavirin (1-beta-D-ribofuranosyl-1,2,4-triazole-3-carboxamide) is a broad-spectrum RNA virus inhibitor that was discovered almost four decades ago. However, a lack of activity is reported so far for rhabdoviruses. Ammonium-5-tungsto-2-antimoniate (HPA23) and analogues were shown to result in a protective effect in rabies-infected mice; HPA23 was even able to delay mortality in rabies-infected foxes. The mechanism by which these compounds exert their anti-rabies activity has not been unravelled (Bussereau et al., 1988). Ketamine was one of the therapeutic agents used as a therapy for a human rabies survivor who did not receive rabies vaccine (Willoughby, 2007, Willoughby et al., 2005). Unfortunately, this ketamine therapy when re-examined in infected mouse primary neuron cultures and in adult ICR mice using the rabies challenge virus with ketamine vs. vehicle given intraperitoneally did not lead to any significant beneficial therapeutic effects (Jackson et al., 2008).

SAH hydrolase inhibitors are endowed with broad-spectrum antiviral activity against (-)ssRNA viruses (De Clercq, 2005). The antiviral activity spectrum of SAH hydrolase inhibitors against RNA viruses includes amongst other rhabdoviruses. Among the most potent

SAH hydrolase inhibitors and antiviral agents rank carbocyclic 3-deazaadenosine, neplanocin A and 3-deazaneplanocin A. Some of these compounds are particularly active against the vesicular stomatitis virus (a possible surrogate for rabies) and have proven to be effective against a filovirus, the Ebola virus (Huggins, 1989). Even when administered as a single dose of 1 mg/kg on the first or second day after an Ebola Zaire virus infection in mice, 3-deazaneplanocin A reduced peak viremia by more than 1000-fold compared with mock-treated controls, and most or all of the animals survived (Bray et al., 2000). This protective effect was likely not the result of a direct antiviral activity but probably mediated by the production of high concentration of IFN-α in virus-infected mice (Bray et al., 2002). A possible explanation for such mechanism may be the blocking of the 5′-capping of the nascent viral (+)RNA strands, of which the maturation is dependent on methylation. Hence these molecules would prevent the dissociation of these strands from the viral (−)RNA genome and thus lead to an accumulation of replicative intermediates containing double-stranded RNA stretches, that may in turn result in the high level production of interferon.

Besides the SAH hydrolase, various other enzymes, in particular involved in host cell nucleoside/nucleotide pathways have been reported to be good targets for inhibition of RNA virus replication. The OMP decarboxylase inhibitors pyrazofurin [3-(β-D-ribofuranosyl)-4-hydroxypyrazole-5-carboxamide, as prototype of the OMP decarboxylase inhibitors, is active against several mononegavirales including rhabdoviruses (vesicular stomatitis virus) (Andrei & De Clercq, 1993, Georges-Courbot et al., 2006). Carbocyclic analogues of the normal nucleoside cytidine, such as cyclopentylcytosine (C-Cyd, carbodine) and cyclopentenylcytosine (Ce-Cyd) inhibit a variety of viruses including paramyxoviruses and rhabdoviruses (vesicular stomatitis virus) (De Clercq et al., 1990). The putative target is CTP synthetase, the enzyme that catalyses the conversion of UTP to CTP, which is the final step in the de novo pyrimidine biosynthetic pathways. Although these compounds exert often potent *in vitro* antiviral activity; the fact that they target a cellular function renders them not very selective, despite the fact that this function is crucial for efficient viral replication. As outlined in this manuscript, rhabdoviruses offer many potential targets for inhibition by selective antiviral agents. The development of potent and selective antiviral agents against these viruses should be eminently feasible.

5.3. Structural analysis of the proteins involved in the replication of *Rhabdoviridae*

5.3.1. Introduction

The structure-based design of small molecule inhibitors of viral replication has a strong track record of success in the development of novel anti-viral drugs. The starting point of any such drug discovery programme is the determination of the structure of key components of the replication machinery and in particular the RNA-dependent RNA polymerase. However, the proteins involved in replication of *Rhabdoviridae* have proved largely refractory to recombinant expression and therefore structural data on individual components is very limited.

To date the crystal structure of N-(Albertini et al., 2006, Green et al., 2006) and two crystal- and one NMR structure for parts of the P protein (Ding et al., 2006, Mavrakis et al., 2004, Ribeiro et al., 2008) have been determined for two members of the *Rhabdoviridae,* VSV and RABV, whereas no high resolution structural data is available for the L protein. Although not an integral component of the replicative machinery of *Rhabdoviridae*, the M protein plays a number of key roles during the replication cycle of the virus (Connor et al., 2006, Finke & Conzelmann, 2003). However, prior to the start of the VIZIER project, the only structural information available on rhabdovirus M proteins was the structure of a thermolysin-stable M core (M^{th}) of VSV Indiana (VSV_{Ind}) which lacked approximately 50 residues from the N-terminus and the surface-exposed hydrophobic loop between residues 121–124 (Gaudier et al., 2002). In this section the approaches made in the VIZIER project to obtain further structural information on the proteins involved in the replication of rhabdoviruses will be reviewed.

5.3.2. Experimental methods

Two strategies were used to produce recombinant proteins by expression of both full length or domain constructs in either *E. coli* or insect cells using the baculovirus system. In both cases parallel cloning and expression screening at small-scale of multiple constructs including the addition of fusion tags, was key to the rapid generation of expression data. DNAs encoding either full length genes or domains designed from a combination of published analyses (Poch et al., 1990) and further work carried out in the VIZIER Programme were amplified by PCR from cDNAs generated by reverse transcription of viral RNA templates extracted from

infected cells (Delmas et al., 2008). Gene sequences were inserted into expression vectors by ligation independent cloning (Berrow et al., 2007, Walhout et al., 2000) to generate multiple constructs targeted at L, M and P proteins. All constructs incorporated a hexahistidine tag at either the N or C-terminus to enable detection and purification of expressed genes. The protein production platforms operated by OPPF (Berrow et al., 2007) and the AFMB (Vincentelli et al., 2003, Vincentelli et al., 2005) were used to screen for soluble protein expression and subsequent purification and crystallization.

5.3.3. Results and discussion

5.3.3.1 L proteins

Based on sequence analysis, the L protein of non segmented negative RNA viruses carries six highly conserved domains (Poch et al., 1990). The constructs that were designed for the L protein initially relied on these data to target a domain. Among the different activities thought to be carried out by the L protein of non-segmented negative RNA viruses, the capping enzymes are very attractive for structural analysis since an original capping process based on a GDP rather than GMP transfer is catalysed by the VSV L protein (Ogino & Banerjee, 2007). This unusual guanylyltransferase activity requires a non canonical GxxT[n]HR amino acids motif on Domain V of the VSV L protein (Li et al., 2008). Complementary to this process, the methyltransferase activity is carried by the Domain VI (Galloway et al., 2008, Grdzelishvili et al., 2005, Li et al., 2005). It was thus decided to focus on Domains V and VI. To that aim, three constructs encompassing either i) the domain VI alone (amino acids numbering 1675-1871 according to the ABLV L protein), ii) the Domain VI with the C-terminus part of the L protein (amino acids 1668-2128 ABLV L protein numbering), or iii) the combination of Domains V, VI and the C-terminus part of the L (amino acids 1087-2128, ABLV L protein numbering). The equivalent three constructs were made for the L proteins of four rhabdoviruses, Australian bat lyssavirus (ABLV), mokola virus (MOKV), VSV Indiana (VSV_{Ind}) and rabies virus (SADB) and evaluated for expression in *E. coli*. Low-level soluble expression was observed for only two constructs of ABLV (Table 1). These were scaled-up but the data suggested that they did not give homogenous samples following a IMAC purification step due to the low expression. Small changes to the amino acid start and end points of recombinant protein constructs can have a major effect on soluble expression as shown by Graslund and collaborators (Graslund et al., 2008). Therefore a series of 84 further constructs were designed and tested in which the boundaries of the domains V and VI of

ABLV were systematically varied by five to ten residues. Disappointingly none of these constructs showed any improvement in the level of soluble expression suggesting that either *E. coli* is not the appropriate expression system for L protein C-terminal domains or that the rational domain design did not target the correct boundaries of the domains to be expressed.

In parallel with the work on expression of the C-terminal region of the L protein in *E. coli*, the use of insect cells using the baculovirus expression system was evaluated for N-terminal region of L which encompasses the RNA-dependent RNA polymerase (RdRp). Baculoviruses have been used previously to produce the whole L protein of VSV for functional studies only (Chen et al., 2006). From a series of N-terminal constructs of the L protein of the MOKV we identified a number of large fragments which produced soluble proteins in insect cells (Figure 7). Interestingly, soluble expression was enhanced by both the presence of the native N-terminal region and by fusion to maltose binding protein (MBP) (Alexandrov et al., 2001). In pilot experiment, production of the construct encompassing amino acids 1- 880 fused to MBP was scaled up to 100 mL and the yield of pure, soluble protein following removal of the tag was 100µg, sufficient for initial crystallization trials in fluidigm chips (Hansen & Quake, 2003, Hansen et al., 2002). Small crystals have been obtained indicating that the protein is of sufficient quality for crystallization but further work is required to improve crystal growth. These findings have not yet undergone under peer review. Low protein expression remains a challenge, though from this preliminary work it appears that expression in insect cells will be the way forward to obtaining structural information for these complex and difficult to express proteins.

In another approach synthesised DNA has been used in an attempt to avoid potential problems associated with rare codons and transcription termination due to unwanted secondary structure formation, but a series of VSV N-terminal His-tagged constructs only produced insoluble polypeptide when expressed in *E. coli*.

5.3.3.2 P proteins

The P protein fulfils a crucial role during viral RNA synthesis as it is an essential cofactor for the RNA dependant RNA polymerase activity of L. A recent model proposes that during replication the L proteins from a complex with P, which in turns binds to the N-RNA polymer and acts as a bridge to allow access of L to the RNA. This model also suggests that the P may feed the the newly synthetised viral RNA strand with free N for immediate encapsidation (Albertini et al., 2008, Albertini et al., 2006). Expression of P proteins cloned from a number of different viruses was evaluated in both *E. coli* and insect cells. The results showed that full

length P can be expressed in both systems consistent with previous reports (Mavrakis et al., 2004, Schoehn et al., 2001). For ease of use, *E. coli* was selected for the production of protein for crystallization experiments (Table 1). Trials were set up for four different rhabdovirus P proteins from which crystals were obtained for MOKV P but only of a stable C-terminal domain. This was presumably due to proteolysis of the full length protein during crystallization as observed for RABV P (Mavrakis et al., 2004) and reflects the modular organization of the P protein. The C-terminal N-RNA binding domain forms an independently folded proteolytically stable fragment which follows a region of unstructured sequence interrupted by a putative dimerisation domain (Gerard et al., 2009). Using both site directed mutagenesis and yeast two hybrid experiments to measure P-N interaction in lyssaviruses, the relative roles of key amino-acids involved in the interaction of P to N and located in this C-terminal domain of P were precisely determined completing and for some of them confirming previous studies (Jacob et al., 2001). Four residues K212, K215, L225 and R261 were shown to be involved in P-N binding (Assenberg et al., 2009). The high degree of conservation of these positions despite despite the high variability of this region of the P further supports their importance. Disrupting this interaction may offer a novel way of blocking viral replication. At the N-terminus of the protein there is a hydrophobic region which is associated with N^0 binding (N^0 refers to the RNA free form of N)(Mavrakis et al., 2006). Reflecting a high level of amino acid sequence conservation (68%), the structure of the MOKV P is essentially identical to that of RABV P (Mavrakis et al., 2004) both in terms of conformation and electrostatic properties (Assenberg et al., 2009). By contrast the structure of the N-RNA binding domain of VSV P (Ribeiro et al., 2008) shows significant differences to the lyssavirus structures, in particular it lacks the equivalent of RABV/MOKV P α3 and α6 and has a shorter α1, making it a much more compact structure (Figure 8). Although the precise location of the residues involved in binding N-RNA are at topologically distinct sites it is clear that the C-terminal domains of Rhabdoviridae P proteins share a common fold (Assenberg et al., 2009). More largely, this study revealed that the N-RNA binding domains of the P proteins of the Rhabdoviridae and the Paramyxoviridae exhibit structural relationships and probably derived from a common ancestor despite very limited primary sequence conservation

5.3.3.3 M proteins

The matrix (M) proteins of rhabdoviruses are multifunctional proteins essential for virus maturation and budding (Bieniasz, 2006, Harty et al., 1999) that also regulate the expression of viral and host proteins. M has also been implicated in controlling the balance between

transcription and replication of the viral genome (Connor et al., 2006, Finke & Conzelmann, 2003) and in modulating host-cell transcription (Komarova et al., 2007) and -translation (von Kobbe et al., 2000). To date no structure of a full length matrix protein from *Rhabdoviridae* has been reported. The problem appears to be producing the protein in a non-aggregated form suitable for crystallization. To address this issue we have screened three lyssavirus M proteins for soluble expression in *E. coli* making use of different tags including fusion to the ubiquitin-like protein SUMO (Butt et al., 1989, Malakhov et al., 2004). Of the constructs produced, only the SUMO fusions proved to be soluble (Table 1) and enabled the subsequent crystallization and structure determination of the Lagos bat virus (LBV) and VSV serotype New Jersey (VSV_{NJ}) M proteins (Assenberg et al., 2008, Graham et al., 2008). The structures of M of LBV and VSV_{nj} and VSV_{ind} (Gaudier et al., 2002) are remarkably similar, despite sharing less that 10% sequence identity, when structurally aligned (138 residues are aligned out of 202 with an alpha carbon root mean square deviation of 3.1Å). However comparing the structures of the matrix proteins of *Rhabdoviridae* with those of Ebola virus (Dessen et al., 2000) and Borna disease virus (Dessen et al., 2000, Neumann et al., 2009) the only other mononegavirales whose matrix protein structures have been determined, reveals the remarkable structural diversity of matrix proteins between *Rhabdoviridae*, *Bornaviridae* and *Filoviridae* (Figure 9). As proposed by Timmins *et al.* (Timmins et al., 2004), this strongly suggests that matrix proteins of different viruses have nothing in common but rather have arisen independently to carry out a very similar structural role in the assembly of certain enveloped viruses

The structures of the rhabdovirus M proteins reveal that they share a similar overall fold and that both self-associate via a stretch of amino acids (in the otherwise-disordered N terminus) that bind to a similar region on the globular domain. Although the binding region is similar in both structures, the molecular details of the interaction interface differ dramatically between VSV_{NJ} and LBV M. In the case of VSV M, this involves docking of a conserved phenylalanine residue in the N-terminal region into a deep hydrophobic pocket, whilst association in the LBV M depends upon the binding of a type-II polyproline helix into a shallow groove on the surface of the protein (Figure 10). This inter-molecular interaction provides a plausible mechanism for the self-assembly of M, leading to enhanced affinity for membranes. Disrupting this interaction may offer a novel way of blocking viral assembly. Further, the differences in the interaction between VSV and LBV may provide a structural framework for understanding the distinct cytopathic effects of vesiculoviruses and lyssaviruses.

6. Conclusions and what is next in antiviral drug design against Mononegavirales

The list of rhabdoviruses and more largely of mononegavirales is not complete and more viruses will certainly be characterized in the near future and will emerge, illustrating the large genetic diversity of these viruses and their impact on human health. However, Antiviral drug discovery against Mononegavirales has been lagging behind that of other RNA viruses, in particular (+)RNA viruses. There are multiple reasons behind this observation, and it is certainly not the lack of important human pathogens that has failed to ignite clinical and scientific interest. Perhaps the most advocated reason for the failure to discover interesting antiviral molecules is the virus architecture itself. First, the protection of the replication/transcription machinery into the RNP may provide intrinsic resistance to the access of small inhibitors to active sites known to represent interesting targets. This is also reflected in the crystallization success rate in the VIZIER project, much lower for (-)RNA viruses than (+)RNA viruses. Hence, inhibitor screening produces notoriously low "hit" rates compared to those directed against other RNA viruses, an observation correlated with the low abundance of publications describing the discovery of small molecule inhibitors against Mononegavirales. Second, the vaccine field had shown impressive success with historic viruses, such as rabies, and current viruses robustly attacked by ongoing efficient vaccination campaigns, such as Measles.

Targeting the enzyme activities involved in the replication of mononegavirales and in particular rhabdoviruses using a structure-based design approach then remains a major goal of anti-viral research. Some progress has been made in the VIZIER Programme towards dissecting the multi-functional L protein into components suitable for structural and functional studies. The phosphoprotein polymerase co-factor (P) and the structural matrix protein (M) which plays a number of roles during viral replication drives viral assembly have both proved much more amenable to structural biology than L. Applying the multi-construct /multi-virus approach deduced from the studies on rhabdovirus diversity central to the protein production processes in VIZIER has yielded new structural information which may ultimately be exploitable in the derivation of novel ways of intervening in viral replication.

There are also several observations that might stimulate these researches soon. First, the plasticity and rapid rate of evolution of mononegavirales exemplified by Henipah outbreaks have shown that the emergence of new highly lethal agents is no fiction, reinforcing the necessity for preparedness against this virus order. Second, pharmaceutical feasibility has

combined with a growing medical need for RSV, another mononegavirales. This will undoubtedly foster scientific collaborative effort that will stimulate the Mononegavirales field as a whole. And third, the influenza research field has recently shown, by successful structural genomic approaches, that it is timely to launch novel drug discovery programs on extremely interesting targets, such as the influenza polymerase. Indeed, the influenza polymerase was considered until now as a very difficult challenge in structural biology, much like the mononegavirales. Recent crystal structures on the influenza polymerase subunits (Dias et al., 2009, Guilligay et al., 2008, He et al., 2008, Kuzuhara et al., 2009, Obayashi et al., 2008, Tarendeau et al., 2007, Tarendeau et al., 2008, Yuan et al., 2009) have shown that novel avenues are now offered to the structure-based drug design field , pointing to better conserved targets less prone to genetic variation than the neuraminidase. It is tempting to speculate that time is ripe for such significant advances in the mononegavirales field into which vaccination and antiviral approaches would meet against this vast reservoir of human pathogens.

Table 1. Summary of Rhabdovirus L, M, P expression

Virus	Protein	Construct (aa)	Soluble Expression - Coomassie		
			N-His	*C-His*	*N-sumo*
ABL	L	1675-1710	+	N/A	N/A
ABL	L	1668-2128	-	N/A	N/A
ABL	L	1087-2128	+	N/A	N/A
SAD	L	1674-1870	-	N/A	N/A
SAD	L	1667-2127	-	N/A	N/A
SAD	L	1086-2127	-	N/A	N/A
VSV	L	1640-1836	-	N/A	N/A
VSV	L	1633-2109	-	N/A	N/A
VSV	L	1072-2109	-	N/A	N/A
Mokola	L	1673-1869	-	N/A	N/A
Mokola	L	1666-2126	+	N/A	N/A
Mokola	L	1673-2126	-	N/A	N/A
Mokola	Matrix	1-end	-	-	++
Mokola	Matrix	46 - End	+++	-	++
Mokola	Matrix	46-110	-	-	tag only
Mokola	Matrix	111-202	-	-	-
Mokola	Matrix	46-130	-	-	N/A
Mokola	Matrix	131-end	-	-	N/A
Lagos	Matrix	1-end	-	-	++
Lagos	Matrix	46 - End	+++	-	++
Lagos	Matrix	46-110	-	-	tag only
Lagos	Matrix	111-202	-	-	-
Lagos	Matrix	46-130	-	-	N/A
Lagos	Matrix	131-end	-	-	N/A
Thailand	Matrix	1-end	-	-	++
Thailand	Matrix	46 - End	+++	-	+
Thailand	Matrix	46-110	-	-	tag only
Thailand	Matrix	111-202	-	-	-
Thailand	Matrix	46-130	-	-	N/A
Thailand	Matrix	131-end	-	-	N/A
VSV	PHOSPHO	full length	++	+	N/A
SAD	PHOSPHO	full length	+	+	N/A
ABL	PHOSPHO	full length	++	++	N/A
9147	PHOSPHO	full length	+	+++	N/A
THAI	PHOSPHO	full length	+	++	N/A
LAG	PHOSPHO	full length	+	++	N/A
MOK	PHOSPHO	full length	-	++	N/A

List of figures

Figure 1: **Phylogeny of the order *Mononegavirales***.

A multiple alignment of the L proteins region flanked by Block I (Poch et al., 1990), from the N-terminus, and ending in the middle of Block IV from a representative set of the Mononegavirales was produced using Muscle (Edgar, 2004) and ClustalX (Thompson et al., 1997) programs with manual adjustment. This alignment was used to construct a Bayesian posterior probability tree. Only informative blocks representing most conserved alignment regions (I.V. Antonov, A.M. Leontovich and A.E. Gorbalenya, in preparation) were included in the analysis (428 positions in total). The alignment construction and processing was assisted by the Viralis software (Gorbalenya et al., unpublished). Two independent Bayesian MCMC chains (4 million steps, 10% burn-in) were run under the WAG amino acid substitution model (Whelan & Goldman, 2001) and rate heterogeneity among sites (gamma distribution with 4 categories) as implemented in the BEAST software (Drummond & Rambaut, 2007). The uncorrelated relaxed molecular clock approach (lognormal distribution) (Drummond et al., 2006) with the exponential growth model was applied. Convergence of runs was verified using Tracer (Rambaut & Drummond, 2007). Numbers at branching points indicate posterior support values (values ~1.0 of bifurcations leading to extremely closely related viruses are not shown). The scale bar represents 0.5 amino acid substitutions per site on average. The following viruses were included in the analysis: NiV, Nipah virus (NC_002728); HeV, Hendra virus (NC_001906); PDPRV, Peste-des-petits-ruminants virus (NC_006383); MeV, Measles virus (NC_001498); RPV, Rinderpest virus strain Kabete O (NC_006296); dMV, Dolphin morbillivirus (NC_005283); cDV, Canine distemper virus (NC_001921); tPMV, Tupaia paramyxovirus (NC_002199); MossV, Mossman virus (NC_005339); JV, J-virus (NC_007454); BeiV, Beilong virus (NC_007803); hPIV-3, Human parainfluenza virus 3 (NC_001796); hPIV-1, Human parainfluenza virus 1 (NC_003461); FDLV, Fer-de-lance virus (NC_005084); NDV, Newcastle disease virus (NC_002617); aPMV-6, Avian paramyxovirus 6 (NC_003043); MuV, Mumps virus (NC_002200); sPIV-5, Simian parainfluenza virus 5 (NC_006430); hPIV-2, Human parainfluenza virus 2 (NC_003443); MenV, Menangle virus (NC_007620); TioV, Tioman virus (NC_004074); hMPV, Human metapneumovirus (NC_004148); aMPV Avian metapneumovirus (NC_007652); mPV, Pneumonia virus of mice J3666 (NC_006579); hRSV, Human respiratory syncytial virus (NC_001781); Ebola, Zaire ebolavirus (NC_002549); Marburg, Lake Victoria marburgvirus (NC_001608); scRV, Siniperca chuatsi rhabdovirus

(NC_008514); SVCV, Spring viremia of carp virus (NC_002803); tRV, Tupaia rhabdovirus (NC_007020); sYNV, Sonchus yellow net virus (NC_001615); mFSV, Maize fine streak virus (NC_005974); rYSV, Rice yellow stunt virus (NC_003746); taVCV, Taro vein chlorosis virus (NC_006942); mMV, Maize mosaic virus (NC_005975); NCMV, Northern cereal mosaic virus (NC_002251); lNYV, Lettuce necrotic yellows virus (NC_007642); BDV, Borna disease virus (NC_001607); MidV, Midway virus (FJ554525); NyaV, Nyamanini virus (FJ554526); VHSV, Viral hemorrhagic septicemia virus (NC_000855); shRV, Snakehead rhabdovirus (NC_000903); IHNV, Infectious hematopoietic necrosis virus (NC_001652); hiRV, Hirame rhabdovirus (NC_005093); VZ08MOKV, Mokola virus; VZ08RABV, Rabies virus; VZ08ABLV, Australian bat lyssavirus; VZ08BEFV, Bovine ephemeral fever virus; VZ08VSIV, Vesicular stomatitis Indiana virus. VIZIER sequences are indicated by names starting with VZxx where xx represents the number of the project that contributed the sequence. Virus names were extracted from Genbank/Refseq annotations using SNAD (Sidorov et al., 2009) http://veb.lumc.nl/SNAD/. Families, subfamilies and genera are indicated; for genera, the -virus suffix was excluded from the names. Nyavirus is a provisional floating genus in the Order (Mihindukulasuriya et al., 2009).

Figure 2: **Phylogenetic relationships of the *Rhabdoviridae* based on a maximum likelihood analysis of a 158 residue alignment of the L polymerase region (Bourhy et al., 2005).**

The established rhabdovirus genera as well as the new groups proposed here are indicated. Horizontal branches are drawn to scale and quartet puzzling frequencies are shown for key nodes (values in italics are for genera, groups and supergroups, while all other quartet puzzling frequencies are shown in normal font). The tree is mid-pointed rooted for purposes of clarity only and all potential outgroup sequences were deemed too divergent to include in the analysis.

Figure 3: **Genetic organization of Mononegavirales**

Large open reading frames are indicated by the colored boxes (adapted from (Bourhy et al., 2008a)).

Figure 4: **Replication cycle of rhabdoviruses**

Figure 5: **Phylogenetic relationships of 22 complete coding regions of lyssavirus genomes representatives of the 7 genotypes** (Delmas et al., 2008). The phylogeny was inferred using an ML procedure, and all horizontal branches are scaled according to the number of substitutions per site. Boot strap values (>95%) are shown for key nodes. The tree is mid-point rooted for purposes of clarity only.

Figure 6: **Origin and evolution of dog rabies virus**

Reconstruction of the spatial dynamics of dog rabies virus based on ML phylogeny of 151 sequences from the N-coding region of rabies virus (Bourhy et al., 2008b). The major clades with bootstrap support values (>90%) of rabies virus are indicated, in different colors. The arrows indicated the migration events as determined in (Bourhy et al., 2008b).

Figure 7. **Small scale insect expression screen of L domain constructs of a lyssavirus (MOKV)**.

Domain boundaries were determined using VaZyMolO (Ferron et al., 2005) and cloned into the appropriate pOPIN vectors using In-fusion (Berrow et al., 2007). Recombinant baculoviruses were prepared using the method developed by Jones & Co-workers (Zhao et al., 2003). Transfections and infections were all performed in 24-well culture dishes, using 500ul Sf9 cells and serum-free media in each well. Following one round of virus amplification, expression was tested by infecting Sf9 cells with different amounts of viruses and harvesting at different time points. Expression screening was performed by detergent lysis of the cells whilst still attached to the plate, followed by robotic Ni-NTA protein extraction of the soluble fraction (obtained by centrifugation). Expression was analysed using SDS-PAGE followed by Coomassie blue staining. The figure shows a summary of the best results obtained indicating that two factors appeared to improve the amount of soluble protein obtained: 1) the presence of the N-terminus of MOKV L and 2) the use of the MBP fusion tag.

Figure 8: **Side by side comparison of the structures of MOKV (left) and VSV (right, PDB Id 2K47) phosphoprotein**

For MOKV P the crystal showed in an interaction between a positively charged pocket implicated previously in N-RNA binding (Jacob et al., 2001)and Assenberg et al, submitted), and the last 4 C-terminal residues of a neighbouringMOKV P molecule in the crystal (shown in yellow in the left figure). The latter consists of 3 acidic amino acids which is therefore highly similar to the acidic region immediately following S389 in N, known to be important

in the interaction with P. This raises the possibility that the interaction seen in the crystal mimicks (at least part of) the interaction between N-P. This contrasts with VSV P for which the C-terminal ~10 amino acids (including VSV P alpha 5) have been implicated in binding N instead (Das et al., 1997, Takacs & Banerjee, 1995) suggesting that the N-binding sites on P are at topologically distinct sites between MOKV and VSVP. Both molecules are rainbow ramped from blue (N-terminus) to red (C-terminus).

Figure 9: **Mononegalevirales matrix protein structures**

The matrix proteins of Ebola virus (VP40), Borna disease virus, vesicular stomatitis virus and Lagos bat virus are shown as cartoons coloured from blue (N-terminus) to red (C-terminus). While the Borna disease virus matrix protein resembles the N-terminal domain of Ebola virus VP40, there is no structural similarity between these proteins and the matrix proteins of vesicular stomatitis virus and Lagos bat virus. For vesicular stomatitis virus and Lagos bat virus the stretch of residues from the otherwise-disordered N-terminus that bind to the main globular domain of the protein are shown as sticks.

Figure 10: **Comparison of M protein structures**

Self-association of rhabdovirus matrix proteins. Residues from the mostly-disordered N-terminal segment of rhabdovirus M proteins are observed binding to grooves on the surface of adjacent M proteins, thereby forming non-covalent linear polymers of M in the crystals. Insets highlight the different molecular details of the interactions between the N-terminal segments and the globular domain

References

Albertini, A. A., Schoehn, G., Weissenhorn, W. & Ruigrok, R. W. (2008). Structural aspects of rabies virus replication. *Cell Mol Life Sci* **65**, 282-94.

Albertini, A. A., Wernimont, A. K., Muziol, T., Ravelli, R. B., Clapier, C. R., Schoehn, G., Weissenhorn, W. & Ruigrok, R. W. (2006). Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* **313**, 360-3.

Alexandrov, A., Dutta, K. & Pascal, S. M. (2001). MBP fusion protein with a viral protease cleavage site: one-step cleavage/purification of insoluble proteins. *Biotechniques* **30**, 1194-8.

Andrei, G. & De Clercq, E. (1993). Molecular approaches for the treatment of hemorrhagic fever virus infections. *Antiviral Res* **22**, 45-75.

Anonymous (2005). WHO Expert Consultation on rabies. In *World Health Organ Tech Rep Ser*, pp. 1-88. Geneva: WHO.

Arai, Y. T., Kuzmin, I. V., Kameoka, Y. & Botvinkin, A. D. (2003). New lyssavirus genotype from the Lesser Mouse-eared Bat (Myotis blythi), Kyrghyzstan. *Emerg Infect Dis* **9**, 333-7.

Arnheiter, H., Davis, N. L., Wertz, G., Schubert, M. & Lazzarini, R. A. (1985). Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell* **41**, 259-67.

Assenberg, R., Delmas, O., Graham, S. C., Verma, A., Berrow, N., Stuart, D. I., Owens, R. J., Bourhy, H. & Grimes, J. M. (2008). Expression, purification and crystallization of a lyssavirus matrix (M) protein. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **64**, 258-62.

Assenberg, R., Delmas, O., Ren, J., Vidalain, P. O., Verma, A., Larrous, F., Graham, S. C., Tangy, F., Grimes, J. M. & Bourhy, H. (2009). The Structure of the N-RNA Binding Domain of the Mokola virus Phosphoprotein. *J Virol*.

Bakker, A. B., Marissen, W. E., Kramer, R. A., Rice, A. B., Weldon, W. C., Niezgoda, M., Hanlon, C. A., Thijsse, S., Backus, H. H., de Kruif, J., Dietzschold, B., Rupprecht, C. E. & Goudsmit, J. (2005). Novel human monoclonal antibody combination effectively neutralizing natural rabies virus variants and individual in vitro escape mutants. *J Virol* **79**, 9062-8.

Bakker, A. B., Python, C., Kissling, C. J., Pandya, P., Marissen, W. E., Brink, M. F., Lagerwerf, F., Worst, S., van Corven, E., Kostense, S., Hartmann, K., Weverling, G. J., Uytdehaag, F., Herzog, C., Briggs, D. J., Rupprecht, C. E., Grimaldi, R. & Goudsmit, J. (2008). First administration to humans of a monoclonal antibody cocktail against rabies virus: safety, tolerability, and neutralizing activity. *Vaccine* **26**, 5922-7.

Benmansour, A., Brahimi, M., Tuffereau, C., Coulon, P., Lafay, F. & Flamand, A. (1992). Rapid sequence evolution of street rabies glycoprotein is related to the highly heterogeneous nature of the viral population. *Virology* **187**, 33-45.

Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I. & Owens, R. J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* **35**, e45.

Bieniasz, P. D. (2006). Late budding domains and host proteins in enveloped virus release. *Virology* **344**, 55-63.

Botvinkin, A. D., Poleschuk, E. M., Kuzmin, I. V., Borisova, T. I., Gazaryan, S. V., Yager, P. & Rupprecht, C. E. (2003). Novel lyssaviruses isolated from bats in Russia. *Emerg Infect Dis* **9**, 1623-5.

Bourhis, J. M., Canard, B. & Longhi, S. (2006). Structural disorder within the replicative complex of measles virus: functional implications. *Virology* **344**, 94-110.

Bourhy, H., Cowley, J. A., Larrous, F., Holmes, E. C. & Walker, P. J. (2005). Phylogenetic relationships among rhabdoviruses inferred using the L polymerase gene. *J Gen Virol* **86**, 2849-58.

Bourhy, H., Goudal, M., Mailles, A., Sadkowska-Todys, M., Dacheux, L. & Zeller, H. (2009). Is there a need for anti-rabies vaccine and immunoglobulins rationing in Europe? *Euro Surveill* **14**.

Bourhy, H., Gubala, A. J., Weir, R. P. & Boyle, D. B. (2008a). Animal rhabdoviruses. In *Encyclopedia of Virology*, 3 edn, pp. 111-121. Edited by B. W. J. V. R. Mahy, M. H. V.

Bourhy, H., Kissi, B., Audry, L., Smreczak, M., Sadkowska-Todys, M., Kulonen, K., Tordo, N., Zmudzinski, J. F. & Holmes, E. C. (1999). Ecology and evolution of rabies virus in Europe. *J Gen Virol* **80 (Pt 10)**, 2545-57.

Bourhy, H., Kissi, B. & Tordo, N. (1993). Molecular diversity of the Lyssavirus genus. *Virology* **194**, 70-81.

Bourhy, H., Reynes, J. M., Dunham, E. J., Dacheux, L., Larrous, F., Huong, V. T., Xu, G., Yan, J., Miranda, M. E. & Holmes, E. C. (2008b). The origin and phylogeography of dog rabies virus. *J Gen Virol* **89**, 2673-81.

Bray, M., Driscoll, J. & Huggins, J. W. (2000). Treatment of lethal Ebola virus infection in mice with a single dose of an S-adenosyl-L-homocysteine hydrolase inhibitor. *Antiviral Res* **45**, 135-47.

Bray, M., Raymond, J. L., Geisbert, T. & Baker, R. O. (2002). 3-deazaneplanocin A induces massively increased interferon-alpha production in Ebola virus-infected mice. *Antiviral Res* **55**, 151-9.

Bussereau, F., Picard, M., Blancou, J. & Sureau, P. (1988). Treatment of rabies in mice and foxes with antiviral compounds. *Acta Virol* **32**, 33-49.

Butt, T. R., Jonnalagadda, S., Monia, B. P., Sternberg, E. J., Marsh, J. A., Stadel, J. M., Ecker, D. J. & Crooke, S. T. (1989). Ubiquitin fusion augments the yield of cloned gene products in Escherichia coli. *Proc Natl Acad Sci U S A* **86**, 2540-4.

Chen, M., Ogino, T. & Banerjee, A. K. (2006). Mapping and functional role of the self-association domain of vesicular stomatitis virus phosphoprotein. *J Virol* **80**, 9511-8.

Connor, J. H., McKenzie, M. O. & Lyles, D. S. (2006). Role of residues 121 to 124 of vesicular stomatitis virus matrix protein in virus assembly and virus-host interaction. *J Virol* **80**, 3701-11.

Coutard, B., Gorbalenya, A. E., Snijder, E. J., Leontovich, A. M., Poupon, A., De Lamballerie, X., Charrel, R., Gould, E. A., Gunther, S., Norder, H., Klempa, B., Bourhy, H., Rohayem, J., L'Hermite, E., Nordlund, P., Stuart, D. I., Owens, R. J., Grimes, J. M., Tucker, P. A., Bolognesi, M., Mattevi, A., Coll, M., Jones, T. A., Aqvist, J., Unge, T., Hilgenfeld, R., Bricogne, G., Neyts, J., La Colla, P., Puerstinger, G., Gonzalez, J. P., Leroy, E., Cambillau, C., Romette, J. L. & Canard, B. (2008). The VIZIER project: preparedness against pathogenic RNA viruses. *Antiviral Res* **78**, 37-46.

Dancho, B., McKenzie, M. O., Connor, J. H. & Lyles, D. S. (2009). Vesicular stomatitis virus matrix protein mutations that affect association with host membranes and viral nucleocapsids. *J Biol Chem* **284**, 4500-9.

Das, T., Pattnaik, A. K., Takacs, A. M., Li, T., Hwang, L. N. & Banerjee, A. K. (1997). Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA. *Virology* **238**, 103-14.

De Clercq, E. (2005). Antiviral drug discovery and development: where chemistry meets with biomedicine. *Antiviral Res* **67**, 56-75.

De Clercq, E., Bernaerts, R., Shealy, Y. F. & Montgomery, J. A. (1990). Broad-spectrum antiviral activity of carbodine, the carbocyclic analogue of cytidine. *Biochem Pharmacol* **39**, 319-25.

de Kruif, J., Bakker, A. B., Marissen, W. E., Kramer, R. A., Throsby, M., Rupprecht, C. E. & Goudsmit, J. (2007). A human monoclonal antibody cocktail as a novel component of rabies postexposure prophylaxis. *Annu Rev Med* **58**, 359-68.

Delmas, O., Holmes, E. C., Talbi, C., Larrous, F., Dacheux, L., Bouchier, C. & Bourhy, H. (2008). Genomic diversity and evolution of the lyssaviruses. *PLoS ONE* **3**, e2057.

Dessen, A., Volchkov, V., Dolnik, O., Klenk, H. D. & Weissenhorn, W. (2000). Crystal structure of the matrix protein VP40 from Ebola virus. *Embo J* **19**, 4228-36.

Dias, A., Bouvier, D., Crepin, T., McCarthy, A. A., Hart, D. J., Baudin, F., Cusack, S. & Ruigrok, R. W. (2009). The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458**, 914-8.

Ding, H., Green, T. J., Lu, S. & Luo, M. (2006). Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J Virol* **80**, 2808-14.

Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88.

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113.

Eigen, M. & Schuster, P. (1979). The hypercycle. A principle of natural self association. Berlin: Springer-Verlag.

Etessami, R., Conzelmann, K. K., Fadai-Ghotbi, B., Natelson, B., Tsiang, H. & Ceccaldi, P. E. (2000). Spread and pathogenic characteristics of a G-deficient rabies virus recombinant: an in vitro and in vivo study. *J Gen Virol* **81**, 2147-53.

Fearns, R., Peeples, M. E. & Collins, P. L. (1997). Increased expression of the N protein of respiratory syncytial virus stimulates minigenome replication but does not alter the balance between the synthesis of mRNA and antigenome. *Virology* **236**, 188-201.

Ferron, F., Rancurel, C., Longhi, S., Cambillau, C., Henrissat, B. & Canard, B. (2005). VaZyMolO: a tool to define and classify modularity in viral proteins. *J Gen Virol* **86**, 743-9.

Finke, S. & Conzelmann, K. K. (2003). Dissociation of rabies virus matrix protein functions in regulation of viral RNA synthesis and virus assembly. *J Virol* **77**, 12074-82.

Galloway, S. E., Richardson, P. E. & Wertz, G. W. (2008). Analysis of a structural homology model of the 2'-O-ribose methyltransferase domain within the vesicular stomatitis virus L protein. *Virology* **382**, 69-82.

Gaudier, M., Gaudin, Y. & Knossow, M. (2002). Crystal structure of vesicular stomatitis virus matrix protein. *Embo J* **21**, 2886-92.

Georges-Courbot, M. C., Contamin, H., Faure, C., Loth, P., Baize, S., Leyssen, P., Neyts, J. & Deubel, V. (2006). Poly(I)-poly(C12U) but not ribavirin prevents death in a hamster model of Nipah virus infection. *Antimicrob Agents Chemother* **50**, 1768-72.

Gerard, F. C., Ribeiro Ede, A., Jr., Leyrat, C., Ivanov, I., Blondel, D., Longhi, S., Ruigrok, R. W. & Jamin, M. (2009). Modular organization of rabies virus phosphoprotein. *J Mol Biol* **388**, 978-96.

Gould, A. R., Hyatt, A. D., Lunt, R., Kattenbelt, J. A., Hengstberger, S. & Blacksell, S. D. (1998). Characterisation of a novel lyssavirus isolated from Pteropid bats in Australia. *Virus Res* **54**, 165-87.

Graham, S. C., Assenberg, R., Delmas, O., Verma, A., Gholami, A., Talbi, C., Owens, R. J., Stuart, D. I., Grimes, J. M. & Bourhy, H. (2008). Rhabdovirus matrix protein structures reveal a novel mode of self-association. *PLoS Pathog* **4**, e1000251.

Graslund, S., Sagemark, J., Berglund, H., Dahlgren, L. G., Flores, A., Hammarstrom, M., Johansson, I., Kotenyova, T., Nilsson, M., Nordlund, P. & Weigelt, J. (2008). The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. *Protein Expr Purif* **58**, 210-21.

Grdzelishvili, V. Z., Smallwood, S., Tower, D., Hall, R. L., Hunt, D. M. & Moyer, S. A. (2005). A single amino acid change in the L-polymerase protein of vesicular stomatitis virus completely abolishes viral mRNA cap methylation. *J Virol* **79**, 7327-37.

Green, T. J., Zhang, X., Wertz, G. W. & Luo, M. (2006). Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* **313**, 357-60.

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R. W., Ortin, J., Hart, D. J. & Cusack, S. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol* **15**, 500-6.

Gupta, A. K., Shaji, D. & Banerjee, A. K. (2003). Identification of a novel tripartite complex involved in replication of vesicular stomatitis virus genome RNA. *J Virol* **77**, 732-8.

Hansen, C. & Quake, S. R. (2003). Microfluidics in structural biology: smaller, faster em leader better. *Curr Opin Struct Biol* **13**, 538-44.

Hansen, C. L., Skordalakes, E., Berger, J. M. & Quake, S. R. (2002). A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. *Proc Natl Acad Sci U S A* **99**, 16531-6.

Harty, R. N., Paragas, J., Sudol, M. & Palese, P. (1999). A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with WW domains of cellular proteins: implications for viral budding. *J Virol* **73**, 2921-9.

He, X., Zhou, J., Bartlam, M., Zhang, R., Ma, J., Lou, Z., Li, X., Li, J., Joachimiak, A., Zeng, Z., Ge, R., Rao, Z. & Liu, Y. (2008). Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature* **454**, 1123-6.

Hemachudha, T., Sunsaneewitayaku, l. B., Desudchit, T., Suankratay, C., Sittipunt, C., Wacharapluesadee, S., Khawplod, P., Wilde, H. & Jackson, A. C. (2006). Failure of therapeutic coma and ketamine for therapy of human rabies. *J Neurovirol* **12**, 407-9.

Huggins, J. W. (1989). Prospects for treatment of viral hemorrhagic fevers with ribavirin, a broad-spectrum antiviral drug. *Rev Infect Dis* **11 Suppl 4**, S750-61.

Jackson, A. C. (2009). Update on rabies diagnosis and treatment. *Curr Infect Dis Rep* **11**, 296-301.

Jackson, A. C., Scott, C. A., Owen, J., Weli, S. C. & Rossiter, J. P. (2008). Human rabies therapy: lessons learned from experimental studies in mouse models. *Dev Biol (Basel)* **131**, 377-85.

Jacob, Y., Real, E. & Tordo, N. (2001). Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. *J Virol* **75**, 9613-22.

Kelly, R. M. & Strick, P. L. (2000). Rabies as a transneuronal tracer of circuits in the central nervous system. *J Neurosci Methods* **103**, 63-71.

Knobel, D. L., Cleaveland, S., Coleman, P. G., Fevre, E. M., Meltzer, M. I., Miranda, M. E., Shaw, A., Zinsstag, J. & Meslin, F. X. (2005). Re-evaluating the burden of rabies in Africa and Asia. *Bull World Health Organ* **83**, 360-8.

Komarova, A. V., Real, E., Borman, A. M., Brocard, M., England, P., Tordo, N., Hershey, J. W., Kean, K. M. & Jacob, Y. (2007). Rabies virus matrix protein interplay with eIF3, new insights into rabies virus pathogenesis. *Nucleic Acids Res* **35**, 1522-32.

Kondo, H., Maeda, T., Shirako, Y. & Tamada, T. (2006). Orchid fleck virus is a rhabdovirus with an unusual bipartite genome. *J Gen Virol* **87**, 2413-21.

Kuzmin, I. V., Hughes, G. J., Botvinkin, A. D., Orciari, L. A. & Rupprecht, C. E. (2005). Phylogenetic relationships of Irkut and West Caucasian bat viruses within the Lyssavirus genus and suggested quantitative criteria based on the N gene sequence for lyssavirus genotype definition. *Virus Res* **111**, 28-43.

Kuzuhara, T., Kise, D., Yoshida, H., Horita, T., Murazaki, Y., Nishimura, A., Echigo, N., Utsunomiya, H. & Tsuge, H. (2009). Structural basis of the influenza A virus RNA polymerase PB2 RNA-binding domain containing the pathogenicity-determinant lysine 627 residue. *J Biol Chem* **284**, 6855-60.

Lafon, M. (2005). Rabies virus receptors. *J Neurovirol* **11**, 82-7.

Langevin, C., Jaaro, H., Bressanelli, S., Fainzilber, M. & Tuffereau, C. (2002). Rabies virus glycoprotein (RVG) is a trimeric ligand for the N-terminal cysteine-rich domain of the mammalian p75 neurotrophin receptor. *J Biol Chem* **277**, 37655-62.

Lentz, T. L., Burrage, T. G., Smith, A. L., Crick, J. & Tignor, G. H. (1982). Is the acetylcholine receptor a rabies virus receptor? *Science* **215**, 182-4.

Leppert, M., Rittenhouse, L., Perrault, J., Summers, D. F. & Kolakofsky, D. (1979). Plus and minus strand leader RNAs in negative strand virus-infected cells. *Cell* **18**, 735-47.

Li, J., Fontaine-Rodriguez, E. C. & Whelan, S. P. (2005). Amino acid residues within conserved domain VI of the vesicular stomatitis virus large polymerase protein essential for mRNA cap methyltransferase activity. *J Virol* **79**, 13373-84.

Li, J., Rahmeh, A., Morelli, M. & Whelan, S. P. (2008). A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense RNA viruses that is essential for mRNA capping. *J Virol* **82**, 775-84.

Li, J., Wang, J. T. & Whelan, S. P. (2006). A unique strategy for mRNA cap methylation used by vesicular stomatitis virus. *Proc Natl Acad Sci U S A* **103**, 8493-8.

Luyet, P. P., Falguieres, T., Pons, V., Pattnaik, A. K. & Gruenberg, J. (2008). The ESCRT-I subunit TSG101 controls endosome-to-cytosol release of viral RNA. *Traffic* **9**, 2279-90.

Malakhov, M. P., Mattern, M. R., Malakhova, O. A., Drinker, M., Weeks, S. D. & Butt, T. R. (2004). SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics* **5**, 75-86.

Manie, S. N., de Breyne, S., Vincent, S. & Gerlier, D. (2000). Measles virus structural components are enriched into lipid raft microdomains: a potential cellular location for virus assembly. *J Virol* **74**, 305-11.

Mavrakis, M., McCarthy, A. A., Roche, S., Blondel, D. & Ruigrok, R. W. (2004). Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. *J Mol Biol* **343**, 819-31.

Mavrakis, M., Mehouas, S., Real, E., Iseni, F., Blondel, D., Tordo, N. & Ruigrok, R. W. (2006). Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology* **349**, 422-9.

Mebatsion, T., Weiland, F. & Conzelmann, K. K. (1999). Matrix protein of rabies virus is responsible for the assembly and budding of bullet-shaped particles and interacts with the transmembrane spike glycoprotein G. *J Virol* **73**, 242-50.

Mihindukulasuriya, K. A., Nguyen, N. L., Wu, G., Huang, H. V., da Rosa, A. P., Popov, V. L., Tesh, R. B. & Wang, D. (2009). Nyamanini and midway viruses define a novel taxon of RNA viruses in the order Mononegavirales. *J Virol* **83**, 5109-16.

Money, V. A., McPhee, H. K., Mosely, J. A., Sanderson, J. M. & Yeo, R. P. (2009). Surface features of a Mononegavirales matrix protein indicate sites of membrane interaction. *Proc Natl Acad Sci U S A* **106**, 4441-6.

Morimoto, K., Hooper, D. C., Carbaugh, H., Fu, Z. F., Koprowski, H. & Dietzschold, B. (1998). Rabies virus quasispecies: implications for pathogenesis. *Proc Natl Acad Sci U S A* **95**, 3152-6.

Nadin-Davis, S. A., Casey, G. A. & Wandeler, A. I. (1994). A molecular epidemiological study of rabies virus in central Ontario and western Quebec. *J Gen Virol* **75 (Pt 10)**, 2575-83.

Neumann, P., Lieber, D., Meyer, S., Dautel, P., Kerth, A., Kraus, I., Garten, W. & Stubbs, M. T. (2009). Crystal structure of the Borna disease virus matrix protein (BDV-M) reveals ssRNA binding properties. *Proc Natl Acad Sci U S A* **106**, 3710-5.

Obayashi, E., Yoshida, H., Kawai, F., Shibayama, N., Kawaguchi, A., Nagata, K., Tame, J. R. & Park, S. Y. (2008). The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* **454**, 1127-31.

Ogino, T. & Banerjee, A. K. (2007). Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Mol Cell* **25**, 85-97.

Ogino, T., Kobayashi, M., Iwama, M. & Mizumoto, K. (2005). Sendai virus RNA-dependent RNA polymerase L protein catalyzes cap methylation of virus-specific mRNA. *J Biol Chem* **280**, 4429-35.

Plumet, S., Duprex, W. P. & Gerlier, D. (2005). Dynamics of viral RNA synthesis during measles virus infection. *J Virol* **79**, 6900-8.

Poch, O., Blumberg, B. M., Bougueleret, L. & Tordo, N. (1990). Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *J Gen Virol* **71 (Pt 5)**, 1153-62.

Pringle, C. R. (1997). The order Mononegavirales--current status. *Arch Virol* **142**, 2321-6.

Rambaut, A. & Drummond, A. J. (2007). Tracer v1.4.

Ribeiro, E. A., Jr., Favier, A., Gerard, F. C., Leyrat, C., Brutscher, B., Blondel, D., Ruigrok, R. W., Blackledge, M. & Jamin, M. (2008). Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. *J Mol Biol* **382**, 525-38.

Roche, S., Bressanelli, S., Rey, F. A. & Gaudin, Y. (2006). Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. *Science* **313**, 187-91.

Roche, S., Rey, F. A., Gaudin, Y. & Bressanelli, S. (2007). Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. *Science* **315**, 843-8.

Ruigrok, R. W., Schoehn, G., Dessen, A., Forest, E., Volchkov, V., Dolnik, O., Klenk, H. D. & Weissenhorn, W. (2000). Structural characterization and membrane binding properties of the matrix protein VP40 of Ebola virus. *J Mol Biol* **300**, 103-12.

Schoehn, G., Iseni, F., Mavrakis, M., Blondel, D. & Ruigrok, R. W. (2001). Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. *J Virol* **75**, 490-8.

Sidorov, I. A., Reshetov, D. A. & Gorbalenya, A. E. (2009). SNAD: Sequence Name Annotation-based Designer. *BMC Bioinformatics* **10**, 251.

Solon, J., Gareil, O., Bassereau, P. & Gaudin, Y. (2005). Membrane deformations induced by the matrix protein of vesicular stomatitis virus in a minimal system. *J Gen Virol* **86**, 3357-63.

Svenda, M., Berg, M., Moreno-Lopez, J. & Linne, T. (1997). Analysis of the large (L) protein gene of the porcine rubulavirus LPMV: identification of possible functional domains. *Virus Res* **48**, 57-70.

Takacs, A. M. & Banerjee, A. K. (1995). Efficient interaction of the vesicular stomatitis virus P protein with the L protein or the N protein in cells expressing the recombinant proteins. *Virology* **208**, 821-6.

Tarendeau, F., Boudet, J., Guilligay, D., Mas, P. J., Bougault, C. M., Boulo, S., Baudin, F., Ruigrok, R. W., Daigle, N., Ellenberg, J., Cusack, S., Simorre, J. P. & Hart, D. J. (2007). Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat Struct Mol Biol* **14**, 229-33.

Tarendeau, F., Crepin, T., Guilligay, D., Ruigrok, R. W., Cusack, S. & Hart, D. J. (2008). Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog* **4**, e1000136.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876-82.

Thoulouze, M. I., Lafage, M., Schachner, M., Hartmann, U., Cremer, H. & Lafon, M. (1998). The neural cell adhesion molecule is a receptor for rabies virus. *J Virol* **72**, 7181-90.

Timmins, J., Ruigrok, R. W. & Weissenhorn, W. (2004). Structural studies on the Ebola virus matrix protein VP40 indicate that matrix proteins of enveloped RNA viruses are analogues but not homologues. *FEMS Microbiol Lett* **233**, 179-86.

Tuffereau, C., Benejean, J., Blondel, D., Kieffer, B. & Flamand, A. (1998). Low-affinity nerve-growth factor receptor (P75NTR) can serve as a receptor for rabies virus. *Embo J* **17**, 7250-9.

Tuffereau, C., Desmezieres, E., Benejean, J., Jallet, C., Flamand, A., Tordo, N. & Perrin, P. (2001). Interaction of lyssaviruses with the low-affinity nerve-growth factor receptor p75NTR. *J Gen Virol* **82**, 2861-7.

Tuffereau, C., Schmidt, K., Langevin, C., Lafay, F., Dechant, G. & Koltzenburg, M. (2007). The rabies virus glycoprotein receptor p75NTR is not essential for rabies virus infection. *J Virol* **81**, 13622-30.

Vidal, S. & Kolakofsky, D. (1989). Modified model for the switch from Sendai virus transcription to replication. *J Virol* **63**, 1951-8.

Vincentelli, R., Bignon, C., Gruez, A., Canaan, S., Sulzenbacher, G., Tegoni, M., Campanacci, V. & Cambillau, C. (2003). Medium-scale structural genomics: strategies for protein expression and crystallization. *Acc Chem Res* **36**, 165-72.

Vincentelli, R., Canaan, S., Offant, J., Cambillau, C. & Bignon, C. (2005). Automated expression and solubility screening of His-tagged proteins in 96-well format. *Anal Biochem* **346**, 77-84.

von Kobbe, C., van Deursen, J. M., Rodrigues, J. P., Sitterlin, D., Bachi, A., Wu, X., Wilm, M., Carmo-Fonseca, M. & Izaurralde, E. (2000). Vesicular stomatitis virus matrix protein inhibits host cell gene expression by targeting the nucleoporin Nup98. *Mol Cell* **6**, 1243-52.

Walhout, A. J., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., van den Heuvel, S. & Vidal, M. (2000). GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol* **328**, 575-92.

Whelan, S. & Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* **18**, 691-9.

Wilde, H. (2007). Failures of post-exposure rabies prophylaxis. *Vaccine* **25**, 7605-9.

Willoughby, R. E., Jr. (2007). A cure for a rabies? *Sci Am* **296**, 88-95.

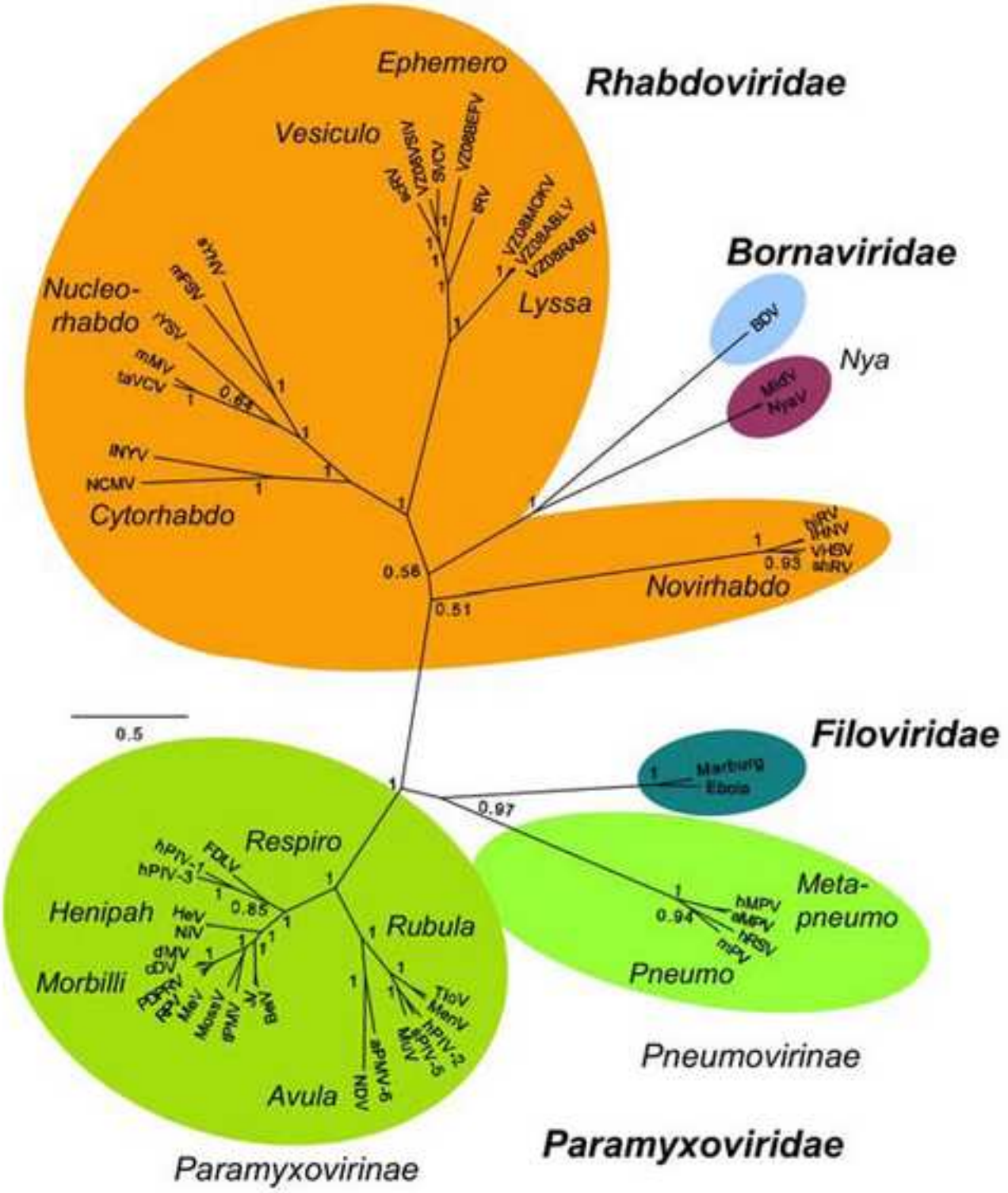
Willoughby, R. E., Jr., Tieves, K. S., Hoffman, G. M., Ghanayem, N. S., Amlie-Lefond, C. M., Schwabe, M. J., Chusid, M. J. & Rupprecht, C. E. (2005). Survival after treatment of rabies with induction of coma. *N Engl J Med* **352**, 2508-14.

Yuan, P., Bartlam, M., Lou, Z., Chen, S., Zhou, J., He, X., Lv, Z., Ge, R., Li, X., Deng, T., Fodor, E., Rao, Z. & Liu, Y. (2009). Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature* **458**, 909-13.

Zhao, Y., Chapman, D. A. & Jones, I. M. (2003). Improving baculovirus recombination. *Nucleic Acids Res* **31**, E6-6.

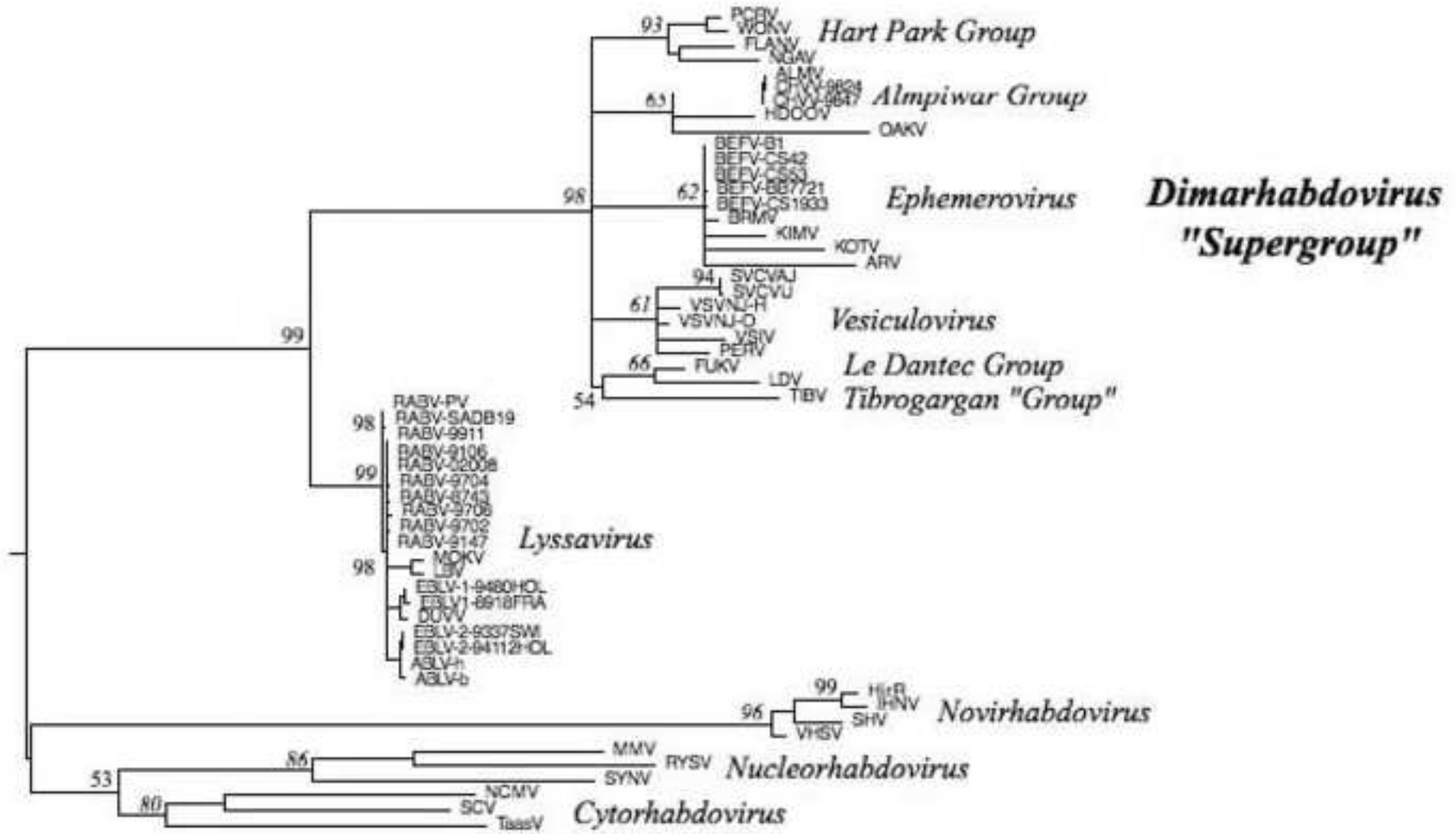
Figure 1

Click here to download high resolution image



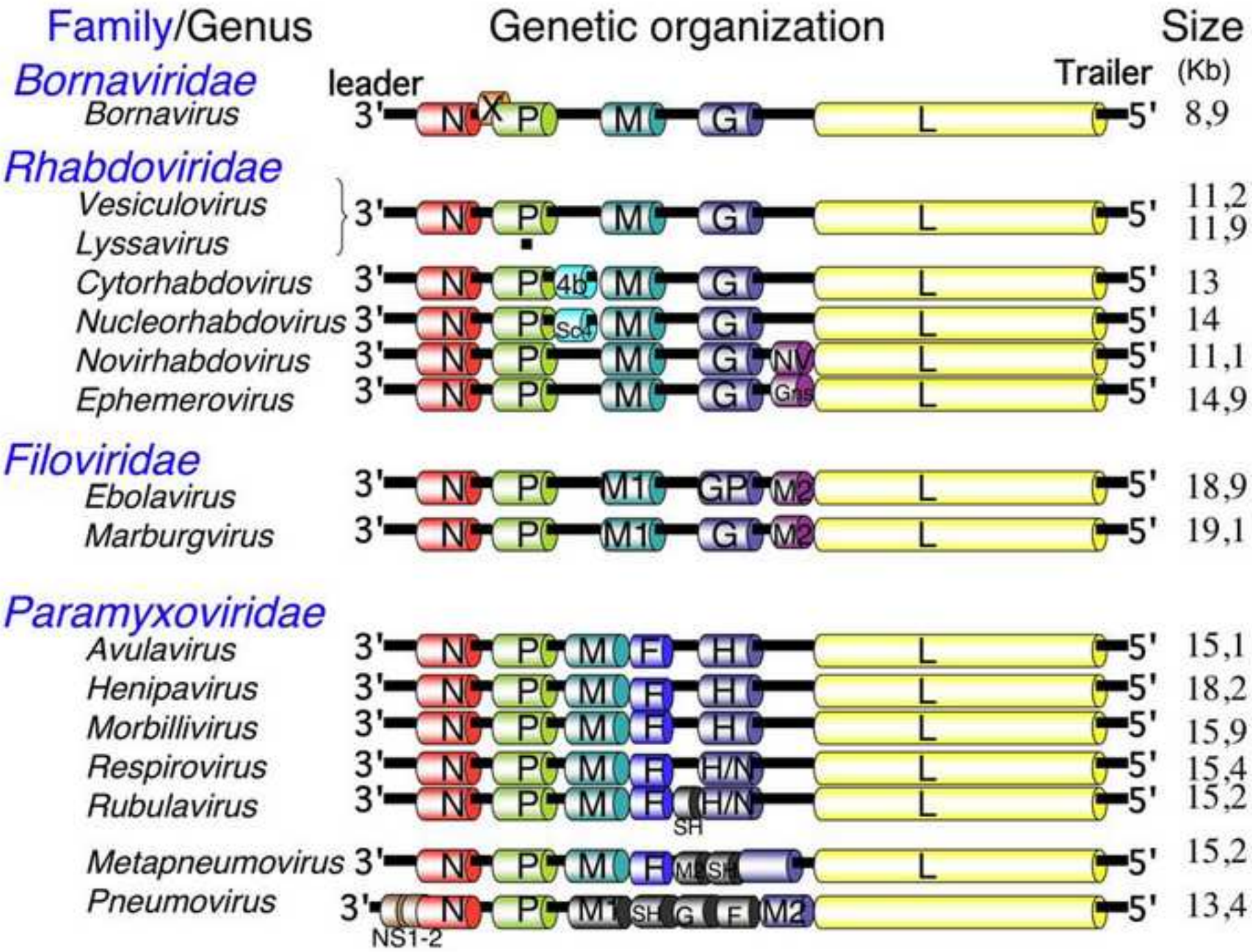
Figure(s) 2

Click here to download high resolution image



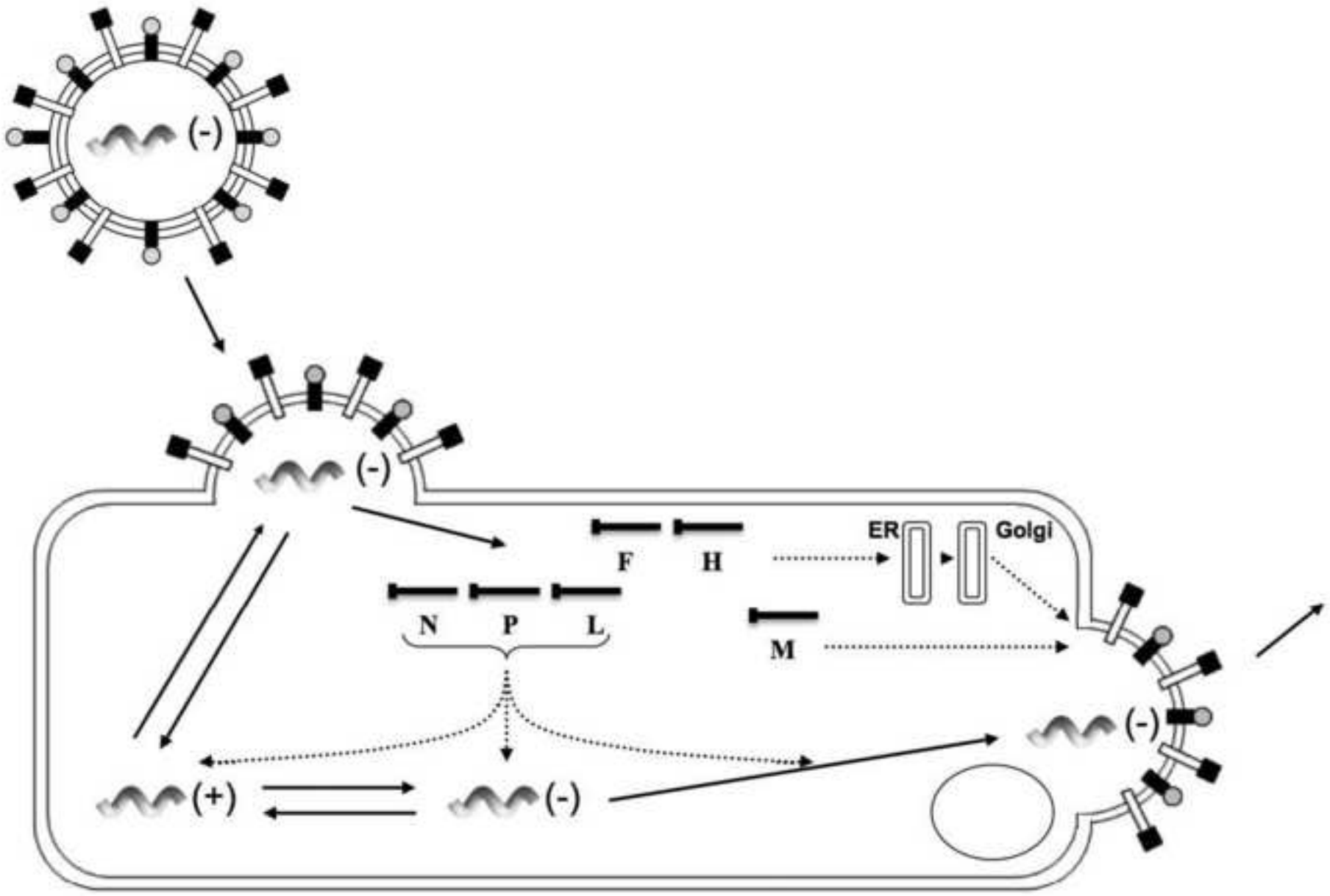
Figure(s) 3

Click here to download high resolution image



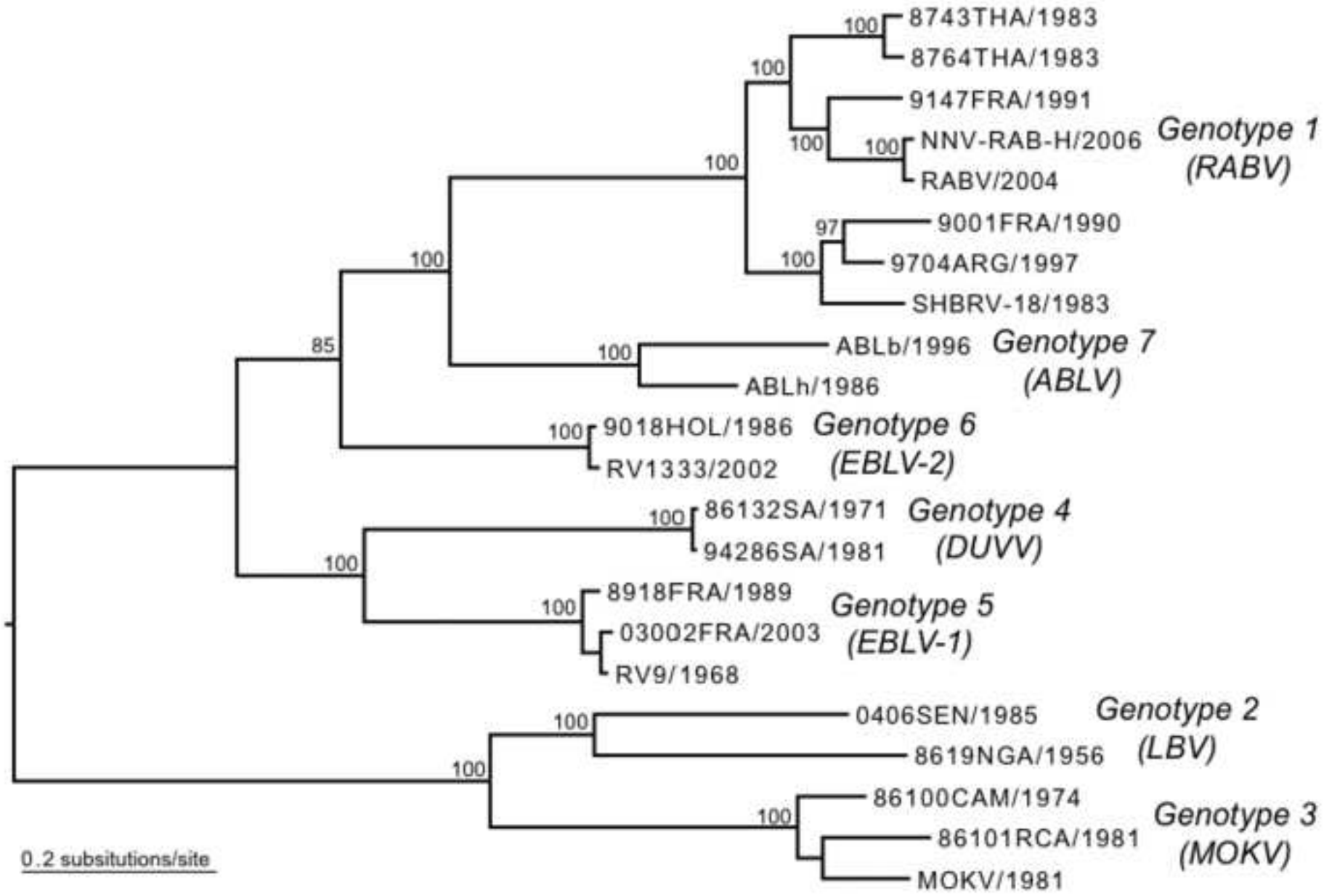
Figure(s) 4

Click here to download high resolution image



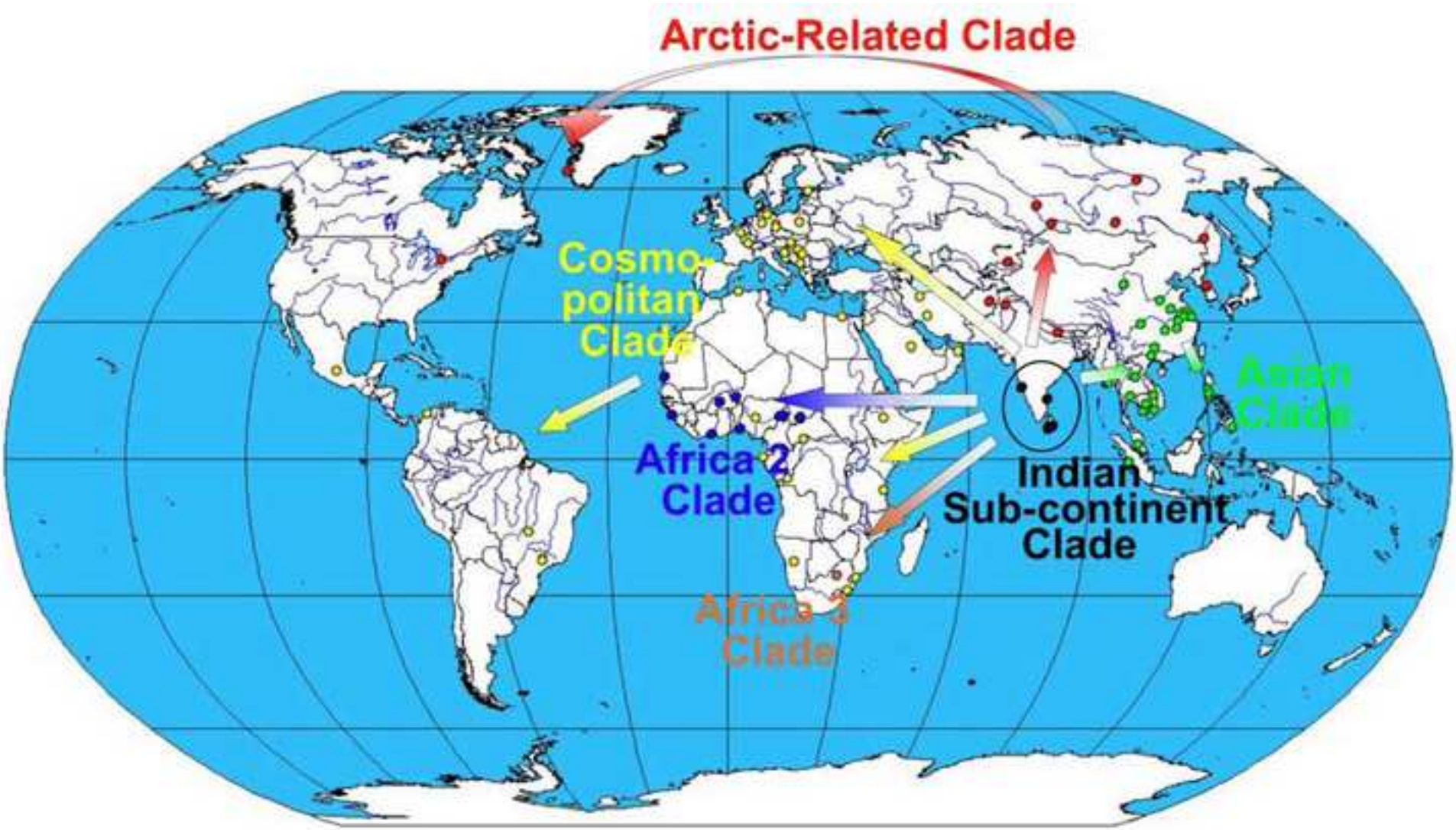
Figure(s) 5

Click here to download high resolution image



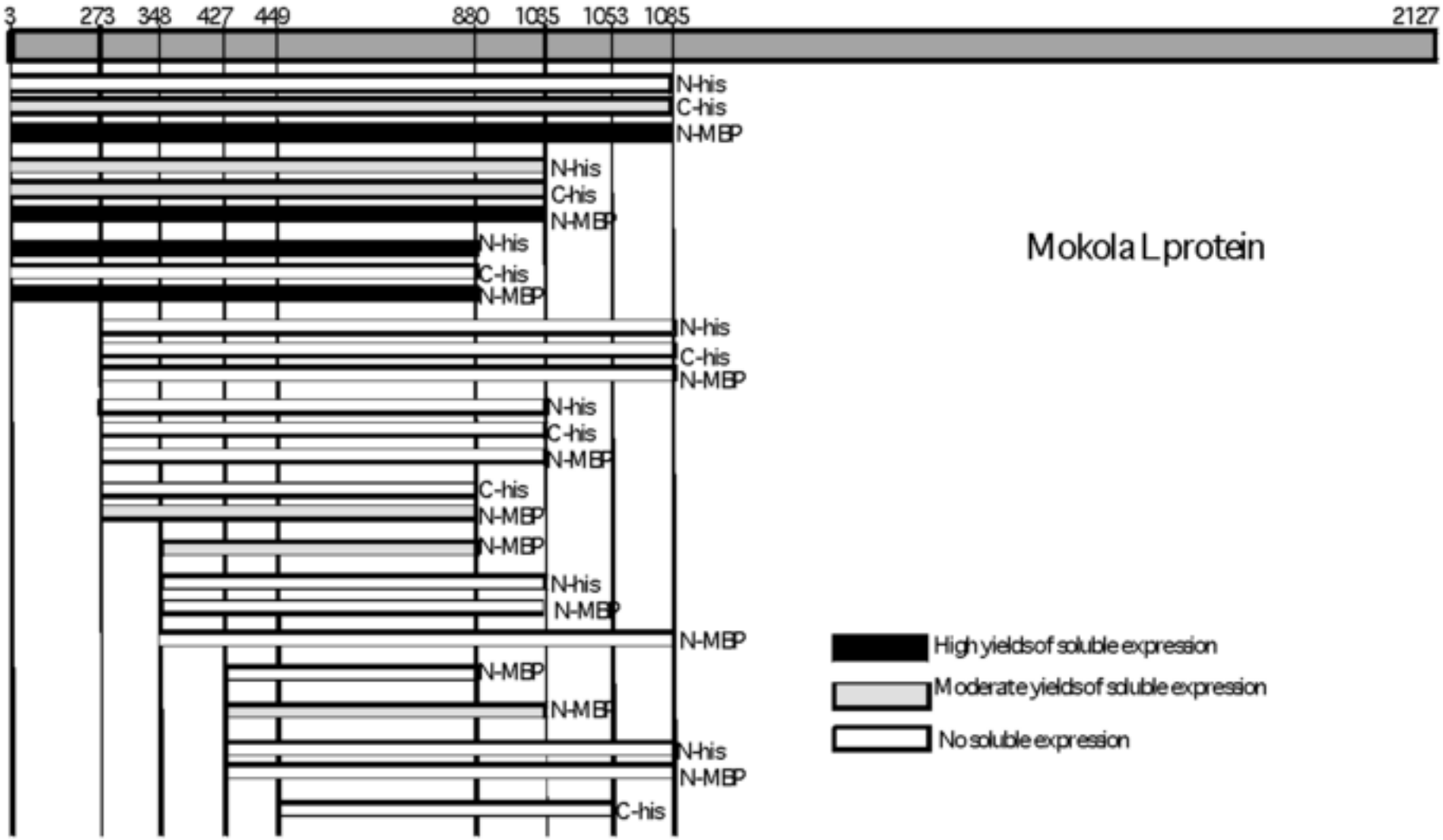
Figure(s) 6

Click here to download high resolution image



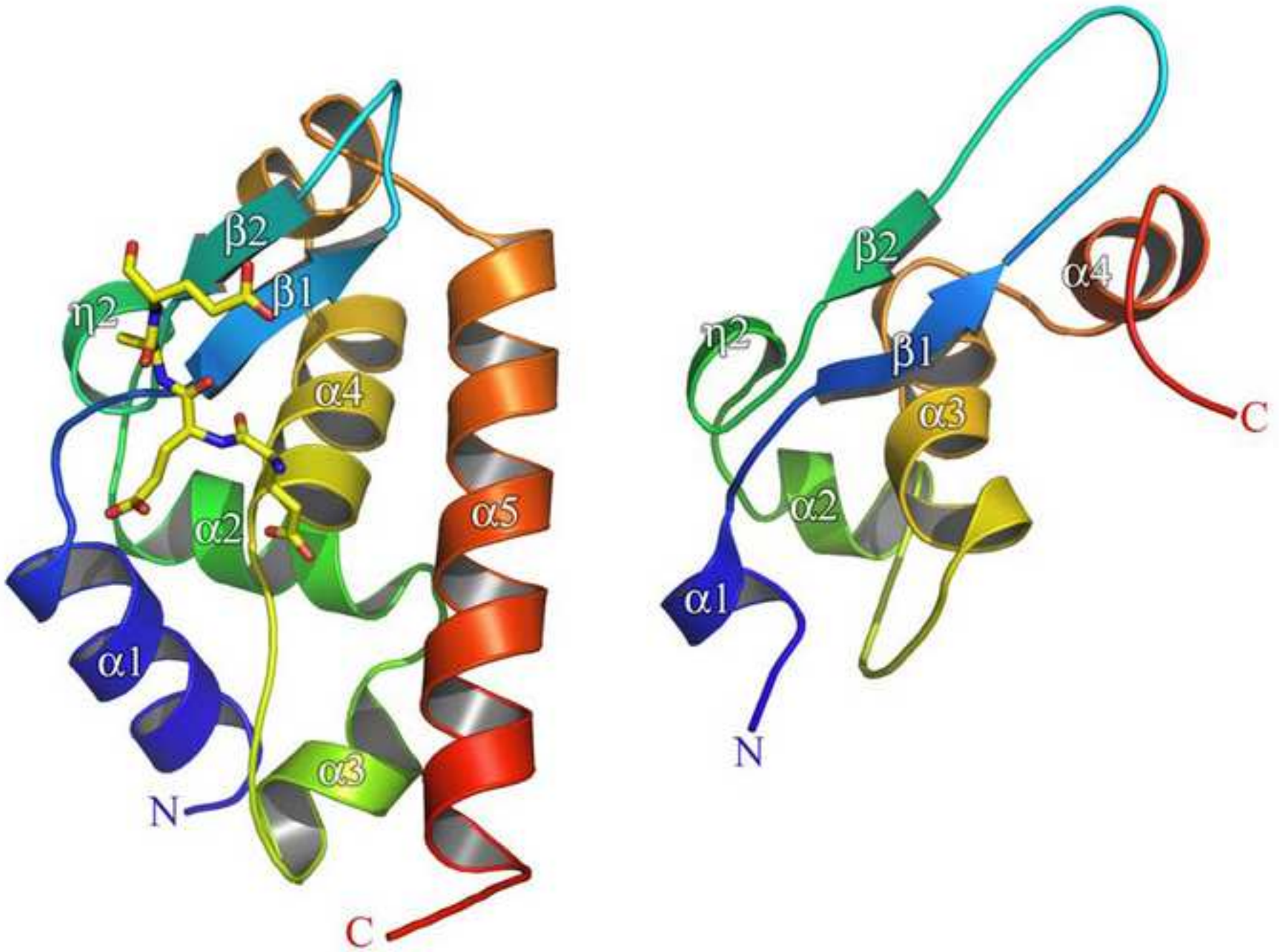
Figure(s) 7

Click here to download high resolution image



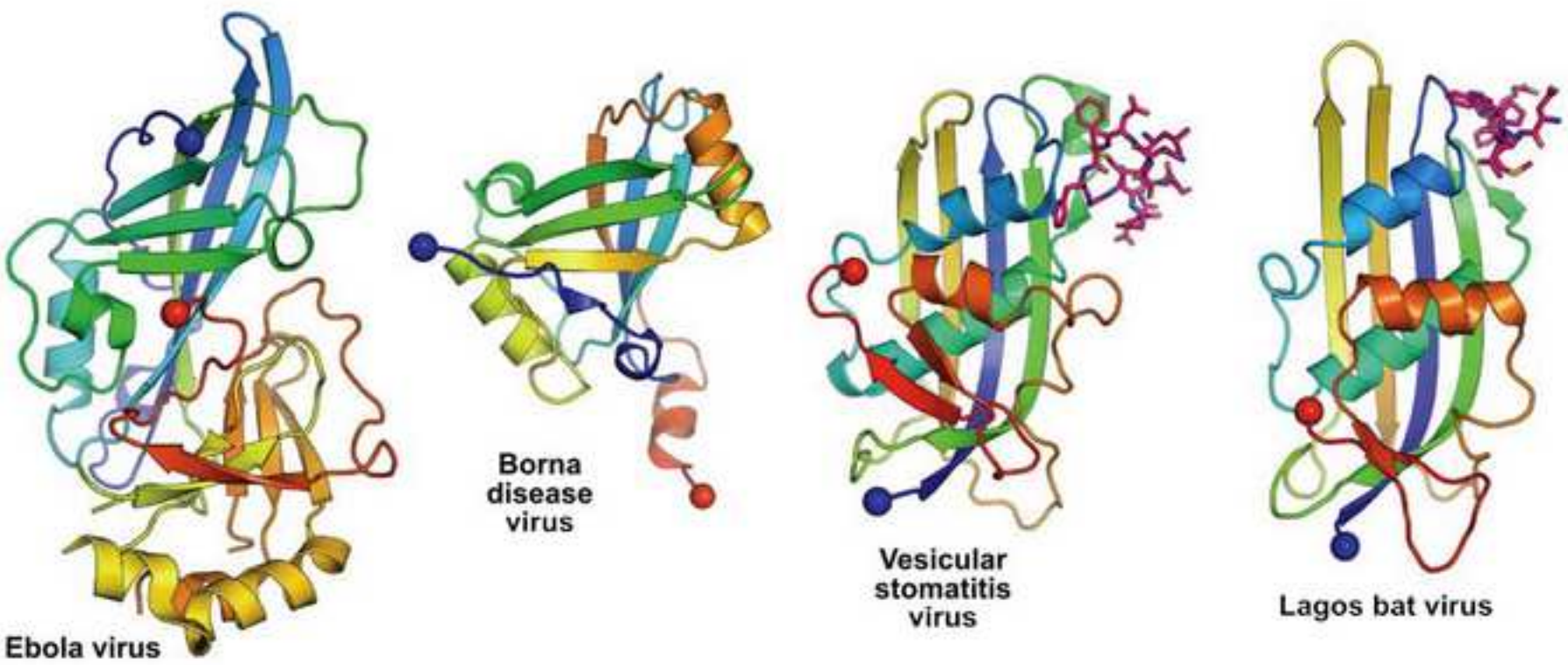
Figure(s) 8

Click here to download high resolution image



Figure(s) 9

Click here to download high resolution image



Figure(s) 10

Click here to download high resolution image

