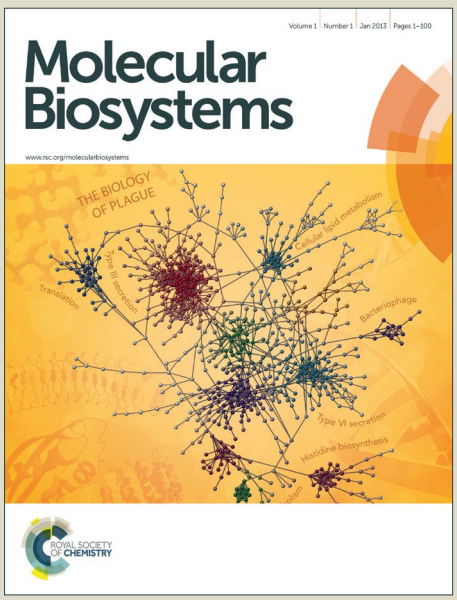
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Journal Name

ARTICLE

An updated evolutionary study of *Flaviviridae* NS3 helicase and NS5 RNA-dependent RNA polymerase, reveals novel invariable motifs as potential pharmacological targets.

Louis Papageorgiou[a,b], Styliani Loukatou[a], Kossida Sofia[c], Dimitrios Maroulis[b], Dimitrios Vlachakis[a,*]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

The rate of *Flaviviridae* family viruses infections worldwide has increased dramatically in the last few years. In addition, infections caused by arthropod vectors viruses including Hepatitis C, West Nile, Dengue fever, Yellow fever and Japanese encephalitis are emerging throughout the world. Based on a recent taxon update, *Flaviviridae* family comprises four main genera; *Flavivirus*, *Hepacivirus*, *Pestivirus* and a recent genus *Pegivirus*. Although the new scientific classification plays a key role with useful information about the relations between viruses, many new documented viruses remain unclassified. Furthermore, concentering the different results of several studies the classification is unclear. In an effort to provide more insights about the viruses classification, a holistic evolutionary study of the two viral enzymes NS3 helicase and NS5 RNA-depended RNA polymerase (RdRp) has been conducted in this study. These two viral enzymes are very crucial for viruses inhibition due to the fact they are involved in viruses survival, proliferation and transmission. The main goal of this study is the presentation of two novel updated phylogenetic trees of the enzymes NS3 helicase and NS5 RdRp as a reliable phylogeny "map" towards to correlate the information of the closely related viruses and identify new possible targets for the *Flaviviridae* family viruses inhibition. Despite the earliest trials for drugs against *Flaviviridae* related viruses, no antiviral drug vaccine is available to date. Therefore there is urgent need for research towards the development of efficient antiviral agents.

Introduction

The *Flaviviridae* is a large scale viral family, which has a worldwide distribution[1]. The viruses within this family fall within significant medical concern due to the variety of diseases that they inflict both humans and animals. The majority of these viruses is zoonotic and can therefore be transmitted from animals to humans through arthropod vectors such as ticks and mosquitoes[2]. These viruses can survive for long periods in their hosts by replicating without damaging the insect. Additionally, humans can be infected by contact with infected blood.

The genus *Flaviviridae* comprises four main genera *Flavivirus*, *Pestivirus*, *Hepacivirus* and *Pegivirus*, although some viruses remain unclassified[3]. The genus *Flavivirus* is the largest one and contains 67 identified human and animal viruses. Dengue Virus (DENV), Japanese Encephalitis Virus (JEV), Yellow Fever Virus (YFV), West Nile Virus (WNV), and Tick Borne Encephalitis virus (TBEV) constitute some important human pathogens of the genus *Flavivirus*. The Classical Swine Fever Virus (CFSV) and the Bovine Viral Diarrhea Virus (BVD) belong to the genus *Pestivirus* and they have not been classified as zoonotic diseases yet. However their impact on livestock is closely related with the economic and social well-being of many countries[4]. The genus *Hepacivirus* is the smallest group and contains the Hepatitis C Virus (HCV), one of the most important human pathogens in the *Flaviviridae* family. The last genus, *Pegivirus, was* recently introduced and includes three virus species that infect humans, primates, and bats[5].

All genera of the family *Flaviviridae* show similarities in the organization of the viral genome, the estimated life cycle, replication, and morphology of the viral particles[6-8]. Virions of *Flaviviridae* family are spherical, about 40-60 nm in diameter[9, 10] (Lindenbach 2001). Each virion contains a lipid envelope which is constructed by two or three virus-encoded membrane proteins, a membrane and a small capsid composed of a protein[3]. The viral genome is located inside the capsid. The genome contains a single-stranded positive sense RNA molecule of approximately 9.5-12.5 kb. It consists of a long Open Reading Frame (ORF) which is located between untranslated regions (UTRs) at 5' and 3' ends. All members of the family lack a 3' terminal polyadenylated tail[11]. The whole genome is translated into a polyprotein, which is processed co- and post-translationally by host and viral proteases. This polyprotein consists of minimum 10 different products, depending on the genus of the virus, that can be classified to structural and non-structural (NS) proteins.

[a] Computational Biology & Medicine Group, Biomedical Research Foundation, Academy of Athens, Soranou Efessiou 4, Athens 11527, Greece
[b] Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, University Campus, Athens, 15784, Greece
[c] IMGT®, the international ImMunoGeneTics information system®, Universite de Montpellier, Laboratoire d'ImmunoGenetique Moleculaire LIGM, UPR CNRS 1142, Institut de Genetique Humaine, 141 rue de la Cardonille, Montpellier, 34396 cedex 5, France.

† Electronic supplementary information (ESI) available: Phylogenetic Analyses results (Newick tree format) and four supplementary figures. See DOI: 10.1039/x0xx00000x

Table 1 The *Flaviviridae* family viruses were contained in the NS3 and NS5 protein sequences data set. All the viruses were extracted from the NIAID Virus Pathogen Database and Analysis Resource (ViPR).

NS5 protein sequences data set		NS3 protein sequences data set	
NS5|Aedes flavivirus	4	NS3|Aedes flavivirus	4
NS5|Apoi virus	2	NS3|Apoi virus	2
NS5|Border disease virus	12	NS3|Border disease virus	11
NS5|Bovine viral diarrheavirus	114	NS3|Bovine viral diarrhea virus	103
NS5|Cell fusing agent virus	1	NS3|Cell fusing agent virus	2
NS5|Chimeric Dengue virus	0	NS3|Chimeric Dengue virus	4
NS5|Classical swine fever virus	75	NS3|Classical swine fever virus	75
NS5|Culex flavivirus	23	NS3|Culex flavivirus	23
NS5|Dengue virus	3747	NS3|Dengue virus	3752
NS5|Flavivirus	2	NS3|Flavivirus	2
NS5|GB virus	66	NS3|GB virus	69
NS5|Greek goat encephalitis virus	1	NS3|Greek goat encephalitis virus	1
NS5|Hepatitis C virus	1497	NS3|Hepatitis C virus	1497
NS5|Hepatitis GB virus	11	NS3|Hepatitis GB virus	13
NS5|Japanese encephalitis virus	203	NS3|Japanese encephalitis virus	203
NS5|Kamiti River virus	3	NS3|Kamiti River virus	3
NS5|Kyasanur forest disease virus	22	NS3|Kyasanur forest disease virus	26
NS5|Langat virus	4	NS3|Langat virus	4
NS5|Louping ill virus	3	NS3|Louping ill virus	3
NS5|Modoc virus	2	NS3|Modoc virus	2
NS5|Montana myotis leukoencephalitis virus	2	NS3|Montana myotis leukoencephalitis virus	2
NS5|Murray Valley encephalitis virus	15	NS3|Murray Valley encephalitis virus	15
NS5|Omsk hemorrhagic fever virus	2	NS3|Omsk hemorrhagic fever virus	3
NS5|Pegivirus	3	NS3|Pegivirus	3
NS5|Pestivirus	3	NS3|Pestivirus	1
NS5|Powassan virus	15	NS3|Powassan virus	15
NS5|Rio Bravo virus	3	NS3|Rio Bravo virus	3
NS5|Rodent hepacivirus	3	NS3|Rodent hepacivirus	3
NS5|Rodent pegivirus	2	NS3|Rodent pegivirus	2
NS5|Royal Farm virus	3	NS3|Royal Farmvirus	3
NS5|Simian pegivirus	33	NS3|Simian pegivirus	33
NS5|Spanish sheep encephalitis virus	1	NS3|Spanish sheep encephalitis virus	1
NS5|St. Louis encephalitis virus	12	NS3|St.Louis encephalitis virus	12
NS5|Tamanabat virus	3	NS3|Tamanabat virus	3
NS5| Theiler's disease-associated virus	1	NS3|Theiler's disease-associated virus	1
NS5|Tick-borne encephalitis virus	127	NS3|Tick-borneencephalitisvirus	127
NS5|Usutu virus	15	NS3|Usutu virus	15
NS5|West Nile virus	1067	NS3|West Nile virus	1067
NS5|Yellow fever virus	72	NS3|Yellow fever virus	72
NS5|Yokose virus	1	NS3|Yokose virus	1
Total	**7175**	**Total**	**7181**

Recent studies have revealed that proteins from the NS region play a significant role in RNA replication[12]. Two of the most significant NS proteins are the viral helicase and the viral RdRp[13]. The viral helicase is coded in the NS3 region of the viral genome and the viral RdRp is coded in the NS5 region. Helicases are enzymes which can unwind double-stranded regions of DNA or RNA in an ATP-dependent reaction. Polymerases are enzymes that synthesize polymers of nucleic acids. Inhibition of these enzymes could be a useful tool for the suppression of the replication rate of the Flaviviridae viral proliferation. Despite Flaviviridae 's increased potential to cause severe diseases, there is no antiviral therapy available until now[14]. Consequently, there is urgent need for new antiviral strategies to be rooted with a forward looking potential, while drawing lessons from the past. In this direction, the non-structural proteins NS3 helicase and NS5 RdRp maybe constitute ideal targets[15-18]. In order to provide more insights about the viral enzymes inhibition, the NS3 helicase and NS5 RdRp were analyzed through a comprehensive phylogenetic analysis. The idea behind is that by sorting viruses evolutionarily will help correlate the knowledge that we have acquired about each one of them and

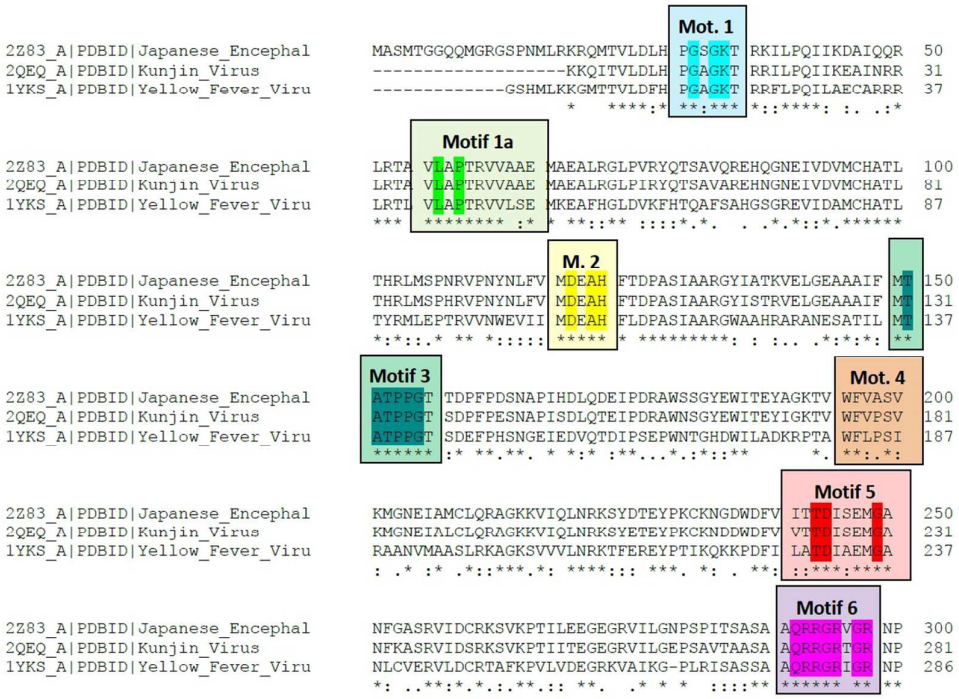


Fig. 1 Sequence alignment between three representative Flaviviruses viral NS3 helicases (Japanese Encephalitis virus, Kunjin virus and Yellow Fever virus). All seven major conserved motifs of the *Flavivirus* helicases have been marked (M1, M1a, M2-M6).

will eventually lead to more accurate and focussed research towards finding drugs that will suppress the activity of these enzymes.

The taxonomy of any virus is not easily decided upon. Although the most of the *Flaviviridae* viruses have been classified in one of the four family genera, many of them remain unclassified. Furthermore, many new *Flaviviridae* related viruses have been documented, but their overall relation with the other family viruses has not been determined[19-21]. Phylogenetic analysis classification of these viruses has been attempted before[22-24]. Since few sequence data were available from known viruses, in particular the viruses without known vectors, these phylogenetic trees had provided only partial information. To achieve a comprehensive phylogenetic analysis in the *Flaviviridae* family, we attempted to obtain the genomic sequence of the NS3 and NS5 genes from all the documented viruses. Based on two phylogenetic trees, we analyzed together the evolutionary relationships among the members of this group.

Methods

Database sequence search

Sequence data for the *Flaviviridae* enzymes NS3 helicase and NS5 RdRp were collected from the NIAID Virus Pathogen Database and Analysis Resource (ViPR)[25] (www.viprbrc.org), and the NCBI RefSeq database. In total, for NS3 helicase and NS5 RdRp were used 7181 and 7175 sequences respectively from species with fully sequenced genomes.

Sequence alignment

NS3 and NS5 protein sequences were aligned using CLUSTALW from MEGA version 6 (Molecular Evolutiona0ry Genetics Analysis Version 6)[26], Jalview[27] and Matlab Bioinformatics Toolbox[28]. In case of the NS3 and NS5 annotation was not available in RefSeq, the start and end positions of the NS3 and NS5 sequence within the whole genome polyprotein was inferred from CLUSTALW alignments with closely related annotated species.

Phylogenetic Analysis

Multiple sequence alignments of the NS3 helicase and NS5 RdRp sequences were performed using two different programs, MUSCLE[29] and Jalview[27]. Only unambiguous homologous regions were retained for phylogenetic analysis; manual masking, trimming and consensus multiple alignments were performed in Matlab[28]. The NS5 alignment was combined with data concerning NS3 helicase from our previous work[13, 17] to create the tree based on both sequences. For each species, the NS3 helicase and NS5 RdRp sequences were concatenated after alignment, masking and trimming. Afterwards, this concatenated alignment was checked with ProtTest[30] in order to estimate the appropriate model evolution. Phylogenetic analyses were performed by

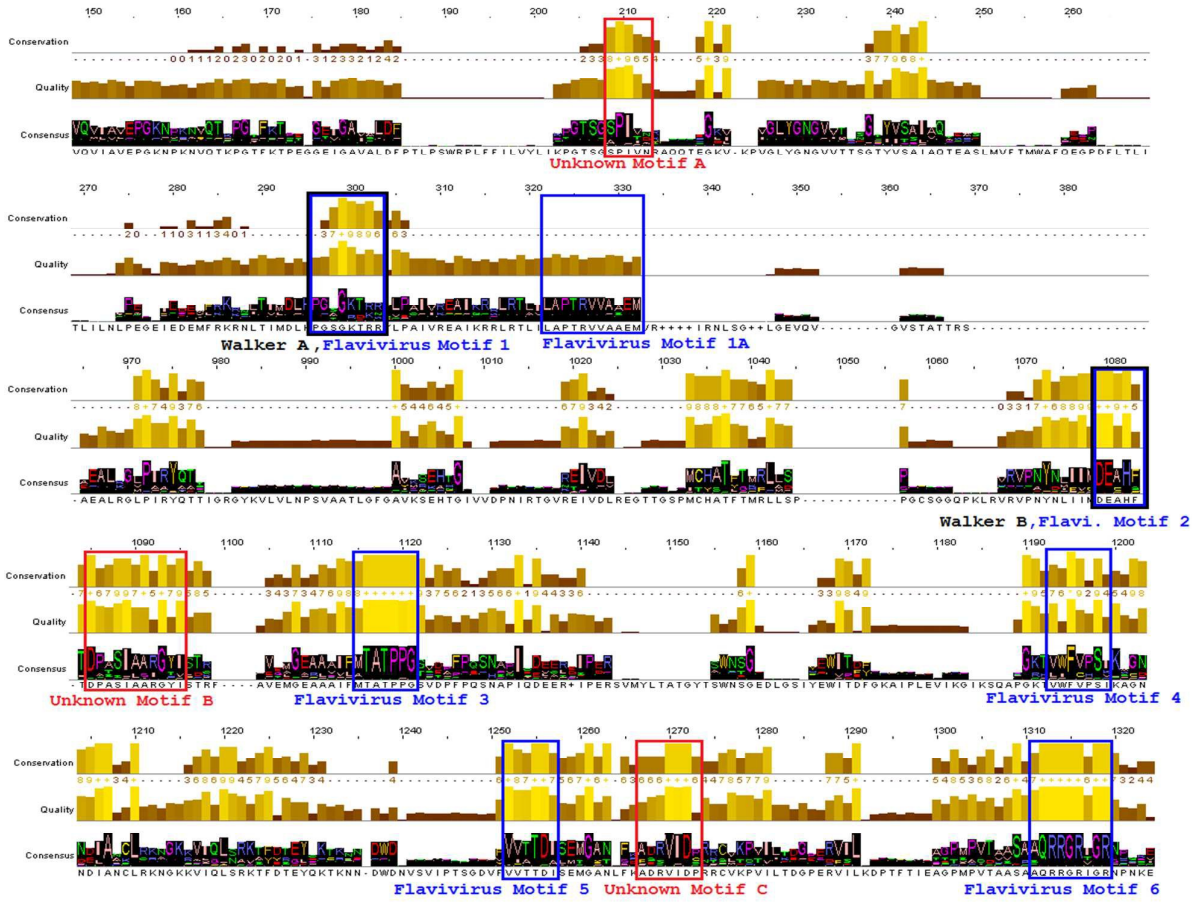


Fig. 2 Consensus sequence based on the NS3 multiple sequences alignment results and parameters such as amino acids quality and conservation. In the consensus consensus sequence they have been marked the seven major conserved motifs of the *Flavivirus* helicases (blue), the Walker motifs (black) and the five unknown suggested conserved regions. The consensus sequence was performed using the Jalview program.

three different ways and two representative phylogenetic trees were constructed for each NS3 helicase and NS5 RdRp data sets. The first set of phylogenetic trees were constructed using the MEGA[26] utilizing Bayesian and Maximum likelihood methods as described in Vlachakis et al., 2013[17] with 100 bootstrap replicates and visualized using MEGA radiation option. The second set of phylogenetic trees were constructed using the Matlab utilizing neighbour joining method as described in Cai et al., 2005 and Gascuel, 1997 [31, 32] and visualized using MEGA radiation option. The third set were constructed using the JALVIEW[27] utilizing the average distance statistical method and visualized using JALVIEW. The conservation score of the consensus sequence was estimated using the Matlab function "seqconsensus" from the bioinformatics toolbox. Scores are computed with the scoring matrix BLOSUM50 and are the average euclidean distance between the scored symbol and the M-dimensional consensus value. M is the size of the alphabet. The consensus value is the profile weighted by the scoring matrix.

Results

Sequence alignment

Multiple sequences alignments were performed using progressive methods and tested in several input parameters. The best sequences matching was, in both cases, detected using the "Gonnet" scoring matrix values, 10 for the gap open value (pairwise alignment score for the first residue in a gap) and 0.1 for the extend gap (pairwise alignment score for each additional residue in a gap). Sequence analyses of the NS3 and NS5 protein sequences show clear conservation in known and unknown important regions within all *Flaviviridae* viruses. According to the multiple sequence alignments they have been highlighted conserved regions in the *Flavivirus*, *Pestivirus*, *Hepacivirus* and *Pegivirus* genera.

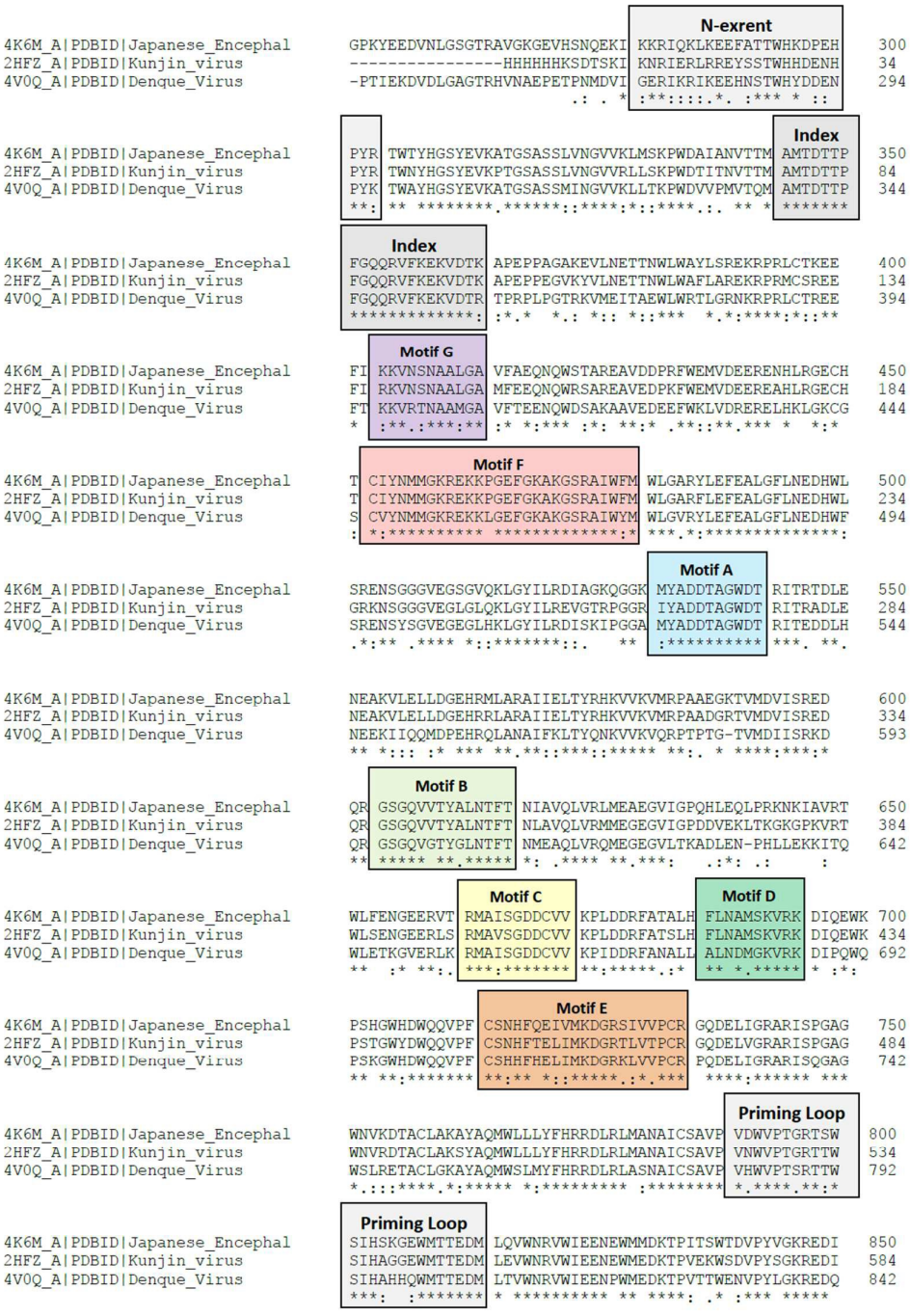


Fig. 3 Sequence alignment between three representative Flaviviruses viral RdRp (Japanese Encephalitis virus, Kunjin virus and Denque Fever virus). All eight major conserved motifs of the *Flavivirus* RdRp have been marked (Motifs A-G).

The first found protein of the comparable enzymatic function proteins among members of the *Flaviviridae* is the NS3 protein. In fact, this polypeptide is multifunctional and it encodes three different enzymes including the viral serine protease, the NTPase and the RNA helicase. The serine protease and nucleoside triphosphatase activity has been experimentally demonstrated for each genus in the *Flaviviridae*[33]. Through the multiple sequences alignment, NS3 proteins were estimated to be highly conserved within the *Flaviviridae*. According to the consensus sequence and sequences conservation from the NS3 multiple sequence alignment results, they have been identified and

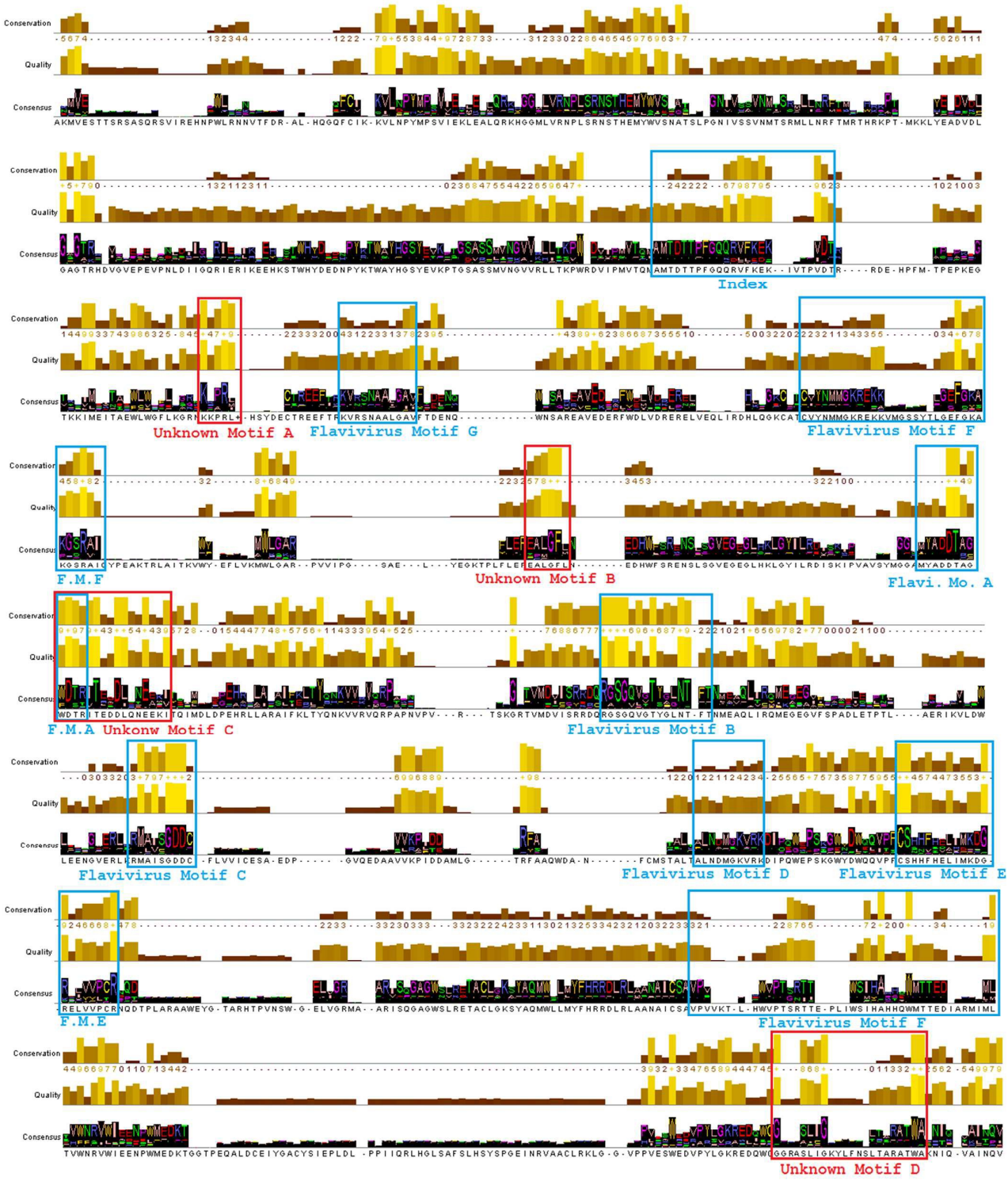


Fig. 4 Consensus sequence based on the NS5 multiple sequences alignment results and parameters such as amino acids quality and conservation. The four unknown suggested conserved regions and the eight major conserved motifs of the Flavivirus helicases have been marked in the consensus sequence (blue).The consensus sequence was performed using the Jalview program.

marked known conserved regions such as the serine protease active site were formed by the catalytic triad of histidine – aspartate –and the glycine residue, the conserved motifs of the DEAD box superfamily 2 of RNA helicases[34-36] (Fig. 1 & 2) and the seven conserved motifs of the genus *Flavivirus* where contained in the helicase active site[13, 24, 37] (Fig. 1, 2 & 5). Moreover, new conserved regions with unknown action they have been marked and proposed as new exploration regions that may play a key role in those viruses function towards to inhibit *Flaviviridae* family viruses (Fig. 2 & 5). The multiple sequence alignment of the NS3 protein sequences revealed several groups of residues in important regions of the enzyme that are highlight conserved too many closely related species.

The second protein with comparable enzymatic function among members of the *Flaviviridae* is the NS5 protein. Sequence similarity calculations through the multiple sequence alignment, revealed that the NS5 proteins are the most highly conserved proteins of the *Flaviviridae* nonstructural proteins. The multiple sequence alignment of the NS5 protein sequences revealed that NS5 region contained conserved motifs within all the *Flaviviridae* family viruse[38] (Fig. 4 & 6). A good conservation is presented throughout the full length of the NS5 protein sequence, especially between viruses' species that belong to the same *Flaviviridae* genus.

Moreover, in the NS5 domain there are two main regions identified, the N-terminal S-adenosyl-L-methyltransferase dependent methyltransferase (MTase) region and the C-terminal RdRP region. Based on the consensus sequence and sequences conservation from the NS5 multiple sequence alignment results, conserved regions including the two conserved motifs of the MTase region and the eight conserved motifs of the RdRP[13] (Fig. 3, 4 & 6) were identified. Furthermore, unknown conserved regions were highlighted and identified within all NS5 proteins samples that have never before been reported in literature.

Exploration of the NS3 and NS5 conserved motifs.

The physical positions and formations of the described NS3 and NS5 Flaviviridae conserved motifs (known and unknown) were further studied within solved proteins 3D structures in order to understand their properties for those molecular functions. To achieve this, Representative experimental protein 3D structures were retrieved from the Protein Data Bank (PDB). Specifically, the PDB crystal structures of 2WZQ, 4B6E and 4CBL (NS3 region related), and the PDB crystal structures of 4V0Q, 3QGD and 2CJQ (NS5 region related), were extracted for the *Flaviviridae* family genera *Flavivirus*, *Hepacivirus* and *Pestivirus* respectively. No available 3D structure was found for the genus *Pegivirus* for these regions. As predicted by the multiple sequences alignment of the non structural proteins 3 (NS3) (Fig. 2) and its representative protein structures (Fig. 5), the structural features of known *Flaviviridae* helicases are identified in the *Flaviviridae* helicases. The Walker motifs A and B otherwise well-known *Flavivirus* motifs 1 and 2 were the most conserved regions within all genera of the *Flaviviridae* family and they have been marked in each representative 3D structure of the NS3 (Fig. 5). Moreover, the "unknown motif B" appeared to be equally important conserved in all *Flaviviridae* genera. Based on the representative 3D structures, the "unknown motif B" is found to be involved in the formation of the helicases active site, as it directly interacts with the Walker motif B and in the case of *Flavivirus* also with the motif 3. In the same direction, the "Unknown motif C" may perform the same action as Flavivirus motifs 4 and 6 which participate in helicases active site. Furthermore, the "unknown motifs A" also found to be highly conserved within all *Flaviviridae* genera in helicases regions with no documented function. This motif may play a critical role in the binding mechanism of these molecules.

The eight conserved motifs[40] and four new descripted conserved motifs, the "unknown motifs A-D", were identified in the Flaviviridae non structural protein 5 (NS5) (Fig. 6). The three domains of representative RdRp were found to be structurally conserved, as well as the various motifs and regions in all *Flaviviridae* NS5 genera (Fig. 6). *Flaviviridae* RdRp was separated into the N-terminal extension, the main polymerase and the priming loop. The main polymerase adopts a shape analogous to a cupped right hand and contains the finger and thumb domains, which rise on the sides of the palm domain (Fig. 6). Based on results, the most conserved part of viral the RdRp is the palm domain, with the motifs A, B[40] and the new discovered motif named "unknown motif C" (Fig. 3 & 6). In motifs A and B there are highlighted the two conserved catalytic aspartic acid residues (D536 and D668) in all *Flaviviridae* RdRp dataset. Furthermore, in the finger domain were identified the motifs F and G[41], that were based on the multiple alignments results and representative proteins structures and exhibit high similarity in all *Flaviviridae* genera NS5 proteins. Moreover, the motif E is marked in the thumb domain of the viral RdRP. Based on the representative 3D structures, the "unknown motif D" has been found to be involved in the formation of the thumb domain, as it directly interacts with the motif E.

Phylogenetic analysis

Three representative set of phylogenetic trees (MEGA and Matlab/MEGA and JALVIEW) have been constructed for each data set (NS3 helicases and NS5 RdRp). Phylogenetic reconstructions of the 7181 NS3 and the 7175 NS5 protein sequences show clear separation between the *Flavivirus*, *Hepacivirus*, *Pegivirus* and *Pestivirus* genera. Moreover, new documented viruses and viruses with unclear classification/transmission have been clustered and closely related within the four *Flaviviridae* genera. Such examples are viruses including Apoi virus, Rio Bravo virus, Montana Myotis Leukoencephalitis virus, Aedes virus, Kamiti River virus and Culex virus, which have been classified in two smaller

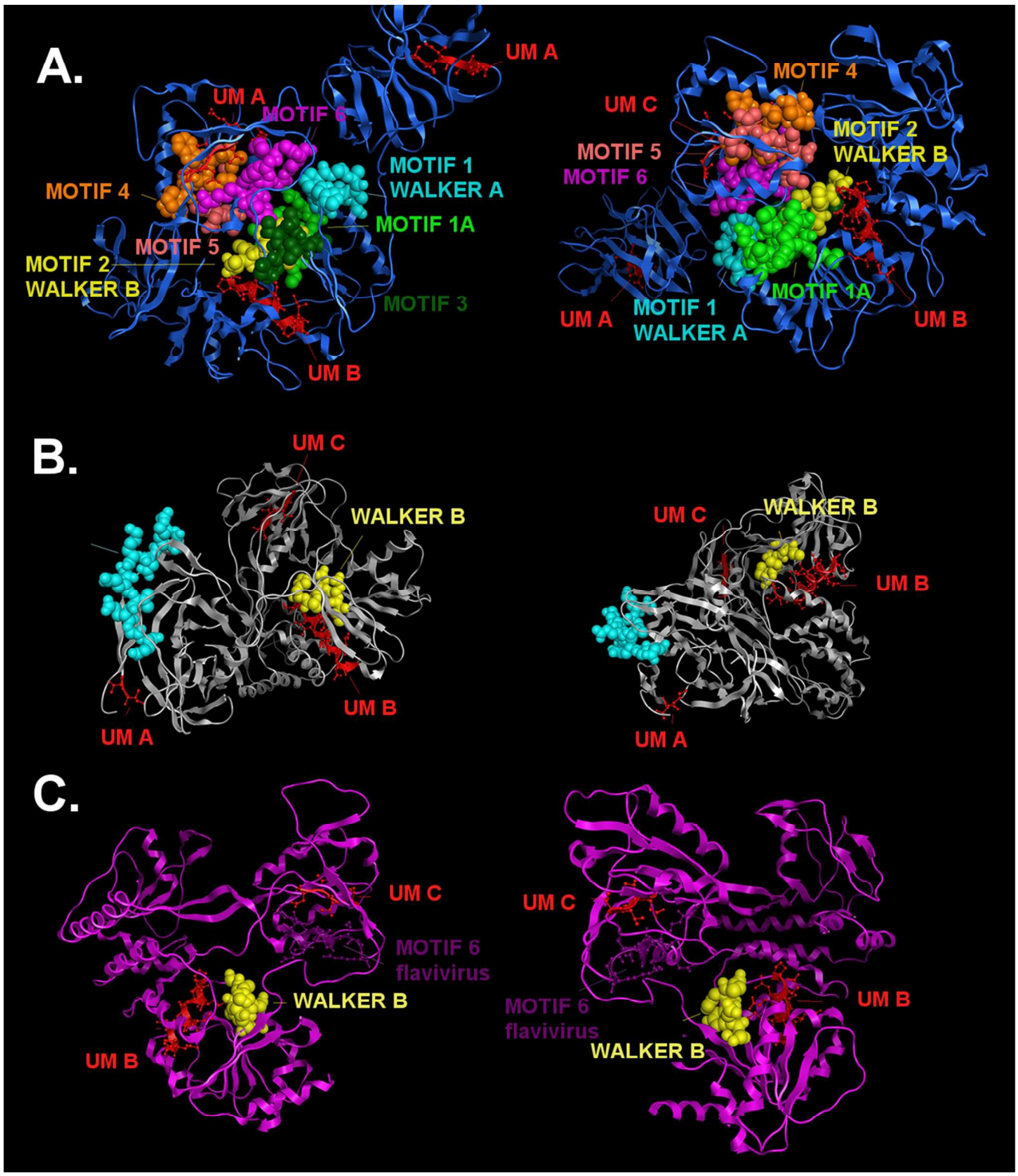


Fig. 5 Representative 3D protein structures of the *Flaviviridae* genera virus helicases. All the major motifs have been color-coded according to the conventions of Fig. 1 and 2, and are shown in CPK format (Usual space filling) along with the rest of the helicase motifs. A: Ribbon representation of the *Flavivirus* virus helicase protein structure (2WZQ). B: Ribbon representation of the *Hepacivirus* virus helicase protein structure (4B6E). C: Ribbon representation of the *Pestivirus* virus helicase protein structure (4CBL).

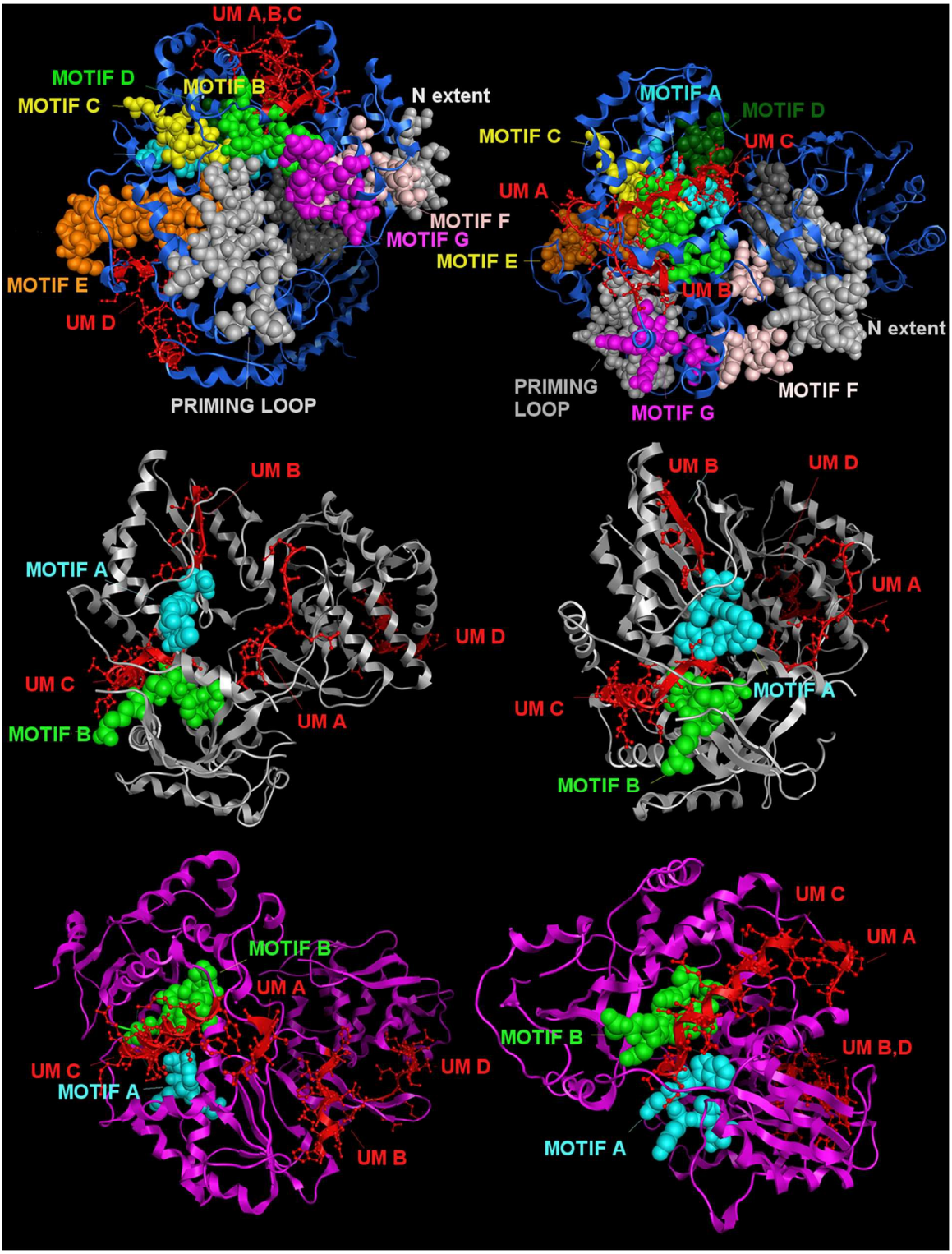


Fig. 6 Representative 3D protein structures of the *Flaviviridae* genera virus RdRp. All the major motifs have been color-coded according to the conventions of Fig. 3 and 4, and are shown in CPK format (Usual space filling) along with the rest of the helicase motifs. A: Ribbon representation of the *Flavivirus* virus helicase protein structure (4V0Q). B: Ribbon representation of the *Hepacivirus* virus helicase protein structure (3QGD). C: Ribbon representation of the *Pestivirus* virus helicase protein structure (2CJQ).

subgroups and were contained in the monophyletic clusters of the genus *Flavivirus*. Furthermore, the constructed phylogenetic trees of NS3 and NS5 protein sequences were found to share similar topology as well as subsequent branching patterns. The phylogenetic conservation scores can be found as raw data in the supplementary material section.

The NS3 phylogenetic trees

The two extracted phylogenetic trees of the NS3 dataset have many similarities (Fig. 7 & 9A, Supplementary Fig. 1 & 2). The genus *Flavivirus* comprises in four main monophyletic subgroups. In the subgroup A and subgroup B we clustered the mosquito-borne Flaviviruses including Dengue Fever virus, West Neil virus, St. Louis Encephalities, Japanese Encephalitis virus, Usutu virus, Murray Valley Encephalitis virus and Yellow fever. In the subgroup C were clustered the tick-borne related Flaviviruses including Tick-borne Encephalitis virus, Louping Ill virus, Powassan virus and Kyasanur Forest disease virus. In addition, in the subgroup D two smaller subgroups were formed within the genus *Flavivarus*. The first small subgroup contains viruses such as Apoi virus, Rio Bravo virus and Montana Myotis Leukoencephalitis virus, found to be closely clustered with the tick-borne Subgroup C Flaviviruses. The second small subgroup is generally related with mosquito and tick borne Flaviviruses and contains the Aedes virus, the Kamiti River virus and the Culex virus. Moreover, the genus *Hepacivirus* (Hepatitis C viruses) and the genus *Pegivirus* (GB viruses) are more closely related compared to the other two genas. Finally, the last genus *Pestivirus* is shown to be evolutionarily differentiated from all the other three genera in a clear separated branch which contained viruses including Classical Swine Fever virus, Bovine virus Diarrhea and Border disease virus.

The NS5 phylogenetic trees

NS5 phylogenetic trees reconstructions (Fig. 8 & 9B, Supplementary Fig. 3 & 4) are found to be sufficiently identical with the NS3 phylogenetic tree. Comparing the two phylogenetic trees of the NS5, a highly clear separation of the Flavivirus, *Hepacivirus*, *Pegivirus* and *Pestivirus* genera was formed. Flaviviruses monophyletic trees are formed by four main clusters which consist of the subgroups A and B with the mosquito-borne flaviviruses, the subgroup C with the tick-borne flaviviruses and the smaller subgroup D with the *Flavivirus* associated flaviviruses. The *Flavivirus* genera monophyletic tree clusters contained the same viruses such as NS3 trees. Moreover, the *Flavivirus* subgroup D with the unclear virus transmission including Apoi virus, Rio Bravo virus, Montana Myotis Leukoencephaliti, Aedes virus, Kamiti River virus and Culex virus is clustered more closely to the tick-borne flaviviruses subgroup C. Furthermore, the *Hepacivirus* (Hepatitis C viruses) and the *Pegivirus* (GB viruses) genera seem to be evolutionary distance more related both in NS3 and NS5 constructed trees. Last but not least, the *Pestivirus* genus is shown to be oriented between Flavivirus and *Hepacivirus* in a clear separated branch.

The phylogenetic trees of *Flaviviridae* family viruses that were created in the past were based on the sequences of a small sample of the family members and thus provided only partial information. Nevertheless, the *Flavivirus* genus dichotomy between tick-borne and mosquito-borne viruses has been recognized by previous studies and was confirmed again in our NS3 and NS5 study. As shown in the NS3 and NS5 phylogenetic trees, the genus *Flavivirus* presented in four main monophyletic trees, the genus *Hepacivirus* presented in one monophyletic tree, the genus *Pegivirus* in one monophyletic tree and the genus Pestivirus in one monophyletic tree. Unlike previous studies, this study reveals that the putative ancestor of the genus *Flavivirus* contained four major branches. According to the phylogenetic analyses in the *Flavivirus* contains the subgroup A with mosquito-borne viruses clusters (Dengue Fever related viruses), the subgroup B with mosquito-borne viruses clusters (Japanese Encephalitis and West Nile related viruses), the subgroup C with tick-borne viruses cluster (Tick-borne Encephalitis, Powassan and Kyasanur forest disease related viruses) and the subgroup D with tick-borne viruses cluster (Apoi, Rio Bravo, Montana Myotis, Leukoencephaliti, Aedes, Kamiti River, Tamana Bat and Culex related viruses. It is worth mentioning that the Yellow fever virus seems to be more correlated with the tick-borne viruses than mosquito-borne viruses. The above phylogenetic trees topologies as well as subsequent branching patterns in each cluster were found to be identical between the trees based on NS3 and NS5 protein sequences data sets.

The phylogenetic segregation of the Flaviviridae family viruses into four major genera clusters was not surprising because of a clear distinction in the size of the sequences. Both NS3 and NS5 protein sequences have clear distinction in the sequence size in the monophyletic trees of each genus. The annotated Flavivirus (~903 aa) NS5 sequences are significantly longer than the annotated NS5 sequences from Pestivurs (~719 aa), Hepacivirus (~590 aa) and Pegivirus (~574 aa). Moreover, the annotated Pestivurs (~1137 aa) NS3 sequences are significantly longer than the annotated NS3 sequences from Hepacivirus (~631 aa), Pegivirus (~627 aa) and Flavivirus (~620 aa). Although the Flaviviridae family phylogenetic trees that were produced in the past were primarily based on envelope gene sequences, it has been reported that the trees were based on NS3 and NS5 nonstructural proteins showed perfect agreement. The envelope gene of Flaviviridae related viruses is less conserved than the NS3 and NS5 proteins, and this difference is reflected in greater differences in the amino acid sequence. The envelope gene of Flaviviridae related viruses is less conserved than the NS3 and NS5 proteins, and this difference is reflected in greater differences in the amino acid sequence.

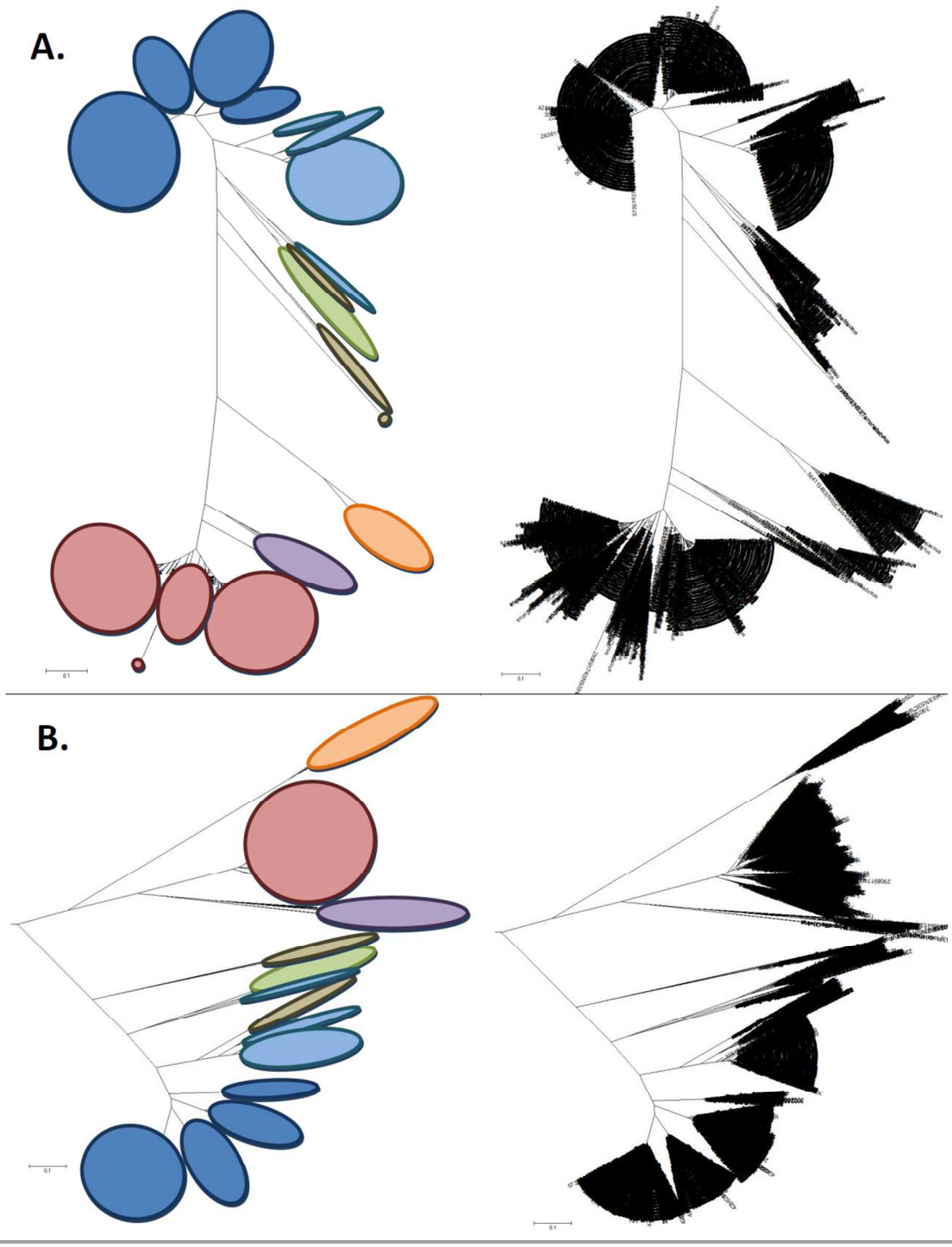


Fig. 7 Phylogenetic trees of the *Flaviviridae* family viruses, using the NS3 non-structural protein sequences dataset. The phylogenetic trees confidently separates the *Hepacivurs* (branch colored Red), *Pestivirus* (branch colored Orange), *Pegivirus* (branch colored Violet), *Flavivirus* / Mosquito-borne (branches colored dark and light Blue) and *Flavivirus* / Tick-borne (branches colored Brown and Green) genera. (A) The tree was constructed by MEGA utilizing Bayesian and Maximum likelihood method and visualized using MEGA radiation option. (B) The tree was constructed by Matlab Bioinformatics Toolbox utilizing neighbour joining method and visualized using MEGA radiation option.

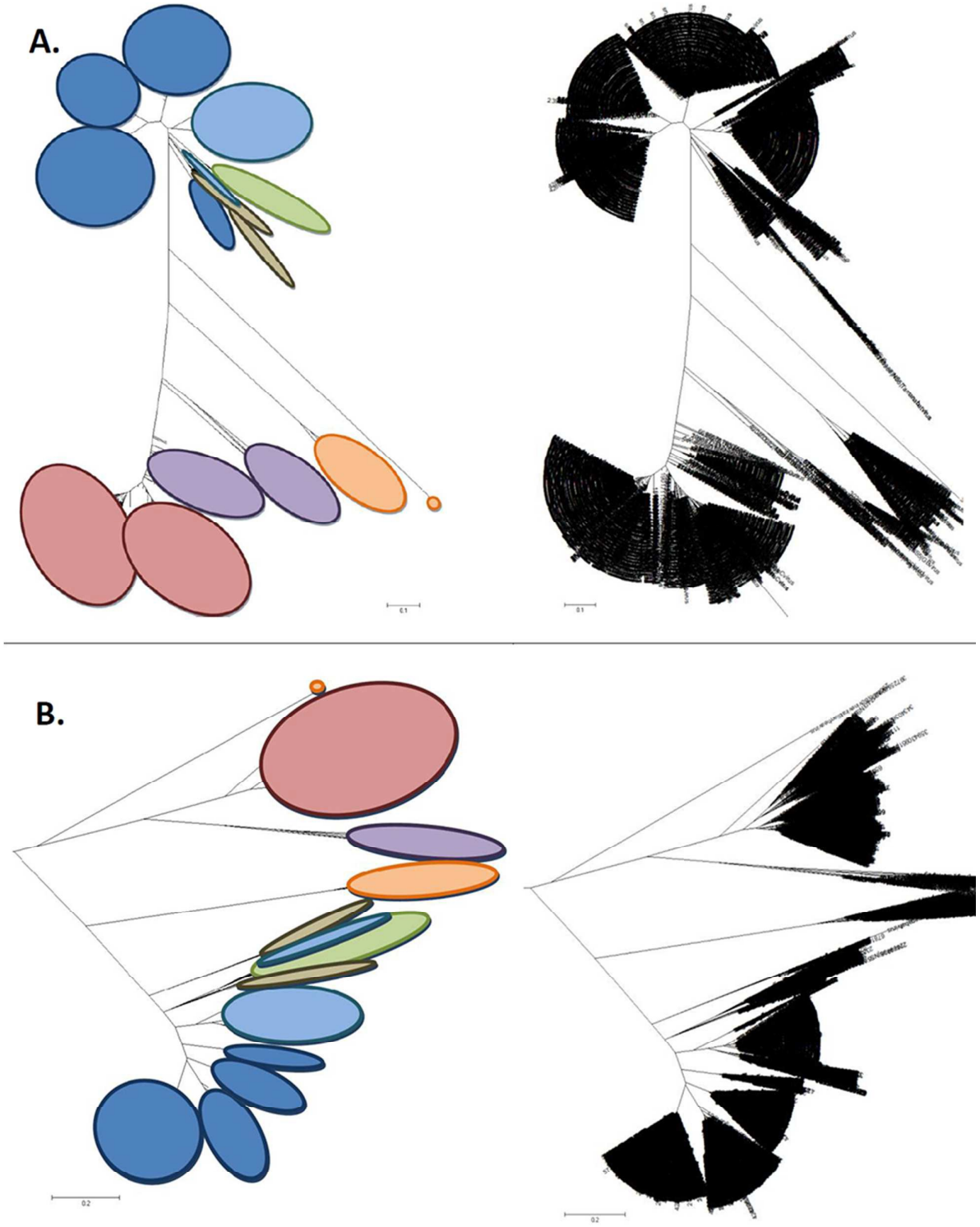


Fig. 8 Phylogenetic trees of the *Flaviviridae* family viruses, using the NS5 non-structural protein sequences dataset. The phylogenetic trees confidently separates the *Hepacivurs* (branch colored Red), *Pestivirus* (branch colored Orange), *Pegivirus* (branch colored Violet), *Flavivirus* / Mosquito-borne (branches colored dark and light Blue) and *Flavivirus* / Tick-borne (branches colored Brown and Green) genera. (A) The tree was constructed by MEGA utilizing Bayesian and Maximum likelihood method and visualized using MEGA radiation option. (B) The tree was constructed by Matlab Bioinformatics Toolbox utilizing neighbour joining method and visualized using MEGA radiation option.

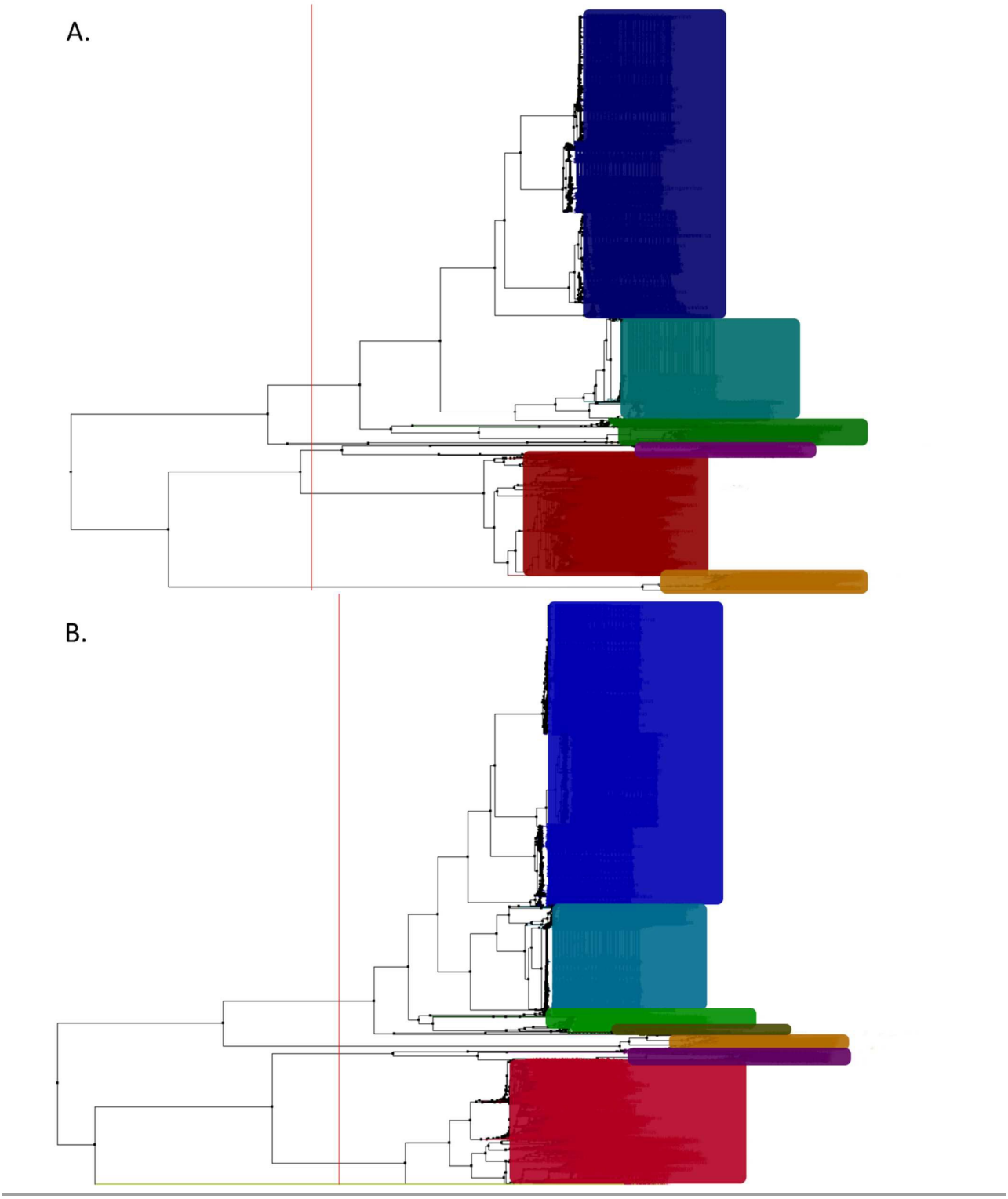


Fig. 9 Phylogenetic trees of the *Flaviviridae* family viruses, using the NS3 non-structural protein sequences (A) and NS5 non-structural protein sequences (B). The trees were constructed using the average distance statistical method and the Jalview software. The phylogenetic trees confidently separates the *Hepacivurs* (branch colored Red), *Pestivirus* (branch colored Orange), *Pegivirus* (branch colored Violet), *Flavivirus* / Mosquito-borne (branches colored dark and light Blue) and *Flavivirus* / Tick-borne (branches colored Brown and Green) genera.

Conclusions

The *Flaviviridae* viral NS3 helicase and NS5 RdRp protein sequences data sets were evolutionary studied using phylogenetic analyses techniques. The phylogenetic trees were performed and displayed high conservation in conserved motifs and functional domains previously characterized among the Flaviviridae species. Moreover, new conserved regions were identified within all *Flaviviridae* viruses and suggested as novel conserved motifs for furthermore analyses. We therefore nominate our *Flaviviridae* enzymes phylogenetic trees to be suitable for a classification scheme and system of taxonomy for newly documented and unclassified viruses. Last but not least, herein we provide results from a sequence-based approach that may proof invaluable when studied on the 3D structure or model of either of those enzymes. Novel, previously unreported regions of invariable residues has been discovered. These regions may be exploited by structural-based drug design techniques, towards the development of new anti-viral agents in the battle for efficient *Flaviviridae* family enzyme inhibition.

Abbreviations

NS3	Non Structural Protein 3
NS5	Non Structural Protein 5
RdRp	RNA-depended RNA Polymerase
NTPase	Nucleoside Triphosphatase

References

1. Daep, C.A., J.L. Munoz-Jordan, and E.A. Eugenin, *Flaviviruses, an expanding threat in public health: focus on dengue, West Nile, and Japanese encephalitis virus.* J Neurovirol, 2014. **20**(6): p. 539-60.
2. Weissenbock, H., et al., *Zoonotic mosquito-borne flaviviruses: worldwide presence of agents with proven pathogenicity and potential candidates of future emerging diseases.* Vet Microbiol, 2010. **140**(3-4): p. 271-80.
3. International Committee on Taxonomy of Viruses. and A.M.Q. King, *Virus taxonomy : classification and nomenclature of viruses : ninth report of the International Committee on Taxonomy of Viruses*. 2012, London ; Waltham, MA: Academic Press. x, 1327 p.
4. Behrens, S.E., et al., *Characterization of an autonomous subgenomic pestivirus RNA replicon.* J Virol, 1998. **72**(3): p. 2364-72.
5. Kapoor, A., et al., *Identification of a pegivirus (GB virus-like virus) that infects horses.* J Virol, 2013. **87**(12): p. 7185-90.
6. Venugopal, K. and E.A. Gould, *Towards a new generation of flavivirus vaccines.* Vaccine, 1994. **12**(11): p. 966-75.
7. Walker, M.A., *Hepatitis C virus: an overview of current approaches and progress.* Drug Discov Today, 1999. **4**(11): p. 518-529.
8. Sudo, K., et al., *Novel hepatitis C virus protease inhibitors: thiazolidine derivatives.* Biochem Biophys Res Commun, 1997. **238**(2): p. 643-7.
9. Guzman, M.G. and G. Kouri, *Dengue diagnosis, advances and challenges.* Int J Infect Dis, 2004. **8**(2): p. 69-80.
10. Neyts, J., P. Leyssen, and E. De Clercq, *Infections with flaviviridae.* Verh K Acad Geneeskd Belg, 1999. **61**(6): p. 661-97; discussion 697-9.
11. Chambers, T.J., et al., *Flavivirus genome organization, expression, and replication.* Annu Rev Microbiol, 1990. **44**: p. 649-88.
12. Nulf, C.J. and D. Corey, *Intracellular inhibition of hepatitis C virus (HCV) internal ribosomal entry site (IRES)-dependent translation by peptide nucleic acids (PNAs) and locked nucleic acids (LNAs).* Nucleic Acids Res, 2004. **32**(13): p. 3792-8.
13. Papageorgiou, L., et al., *Structural models for the design of novel antiviral agents against Greek Goat Encephalitis.* PeerJ, 2014. **2**: p. e664.
14. Heinz, F.X. and K. Stiasny, *Flaviviruses and flavivirus vaccines.* Vaccine, 2012. **30**(29): p. 4301-6.
15. Paula, T., et al., *New drug targets for hepatitis C and other Flaviviridae viruses.* Infect Disord Drug Targets, 2009. **9**(2): p. 133-47.
16. Caillet-Saguy, C., et al., *Polymerases of hepatitis C viruses and flaviviruses: structural and mechanistic insights and drug development.* Antiviral Res, 2014. **105**: p. 8-16.
17. Papageorgiou, L., et al., *Computer-Aided Drug Design and Biological Evaluation of Novel Anti-Greek Goat Encephalitis Agents.* International Journal of Systems Biology and Biomedical Technologies, 2013. **2**(4): p. 1-16.
18. Vlachakis, D. and S. Kossida, *Molecular modeling and pharmacophore elucidation study of the Classical Swine Fever virus helicase as a promising pharmacological target.* PeerJ, 2013. **1**: p. e85.
19. Huhtamo, E., et al., *Novel flaviviruses from mosquitoes: mosquito-specific evolutionary lineages within the phylogenetic group of mosquito-borne flaviviruses.* Virology, 2014. **464-465**: p. 320-9.
20. Kolodziejek, J., et al., *Barkedji virus, a novel mosquito-borne flavivirus identified in Culex perexiguus mosquitoes, Israel, 2011.* J Gen Virol, 2013. **94**(Pt 11): p. 2449-57.
21. Davidson, I., et al., *Development of a reliable dual-gene amplification RT-PCR assay for the detection of Turkey Meningoencephalitis virus in Turkey brain tissues.* J Virol Methods, 2012. **185**(2): p. 239-43.
22. Kuno, G., et al., *Phylogeny of the genus Flavivirus.* J Virol, 1998. **72**(1): p. 73-83.
23. Chandriani, S., et al., *Identification of a previously undescribed divergent virus from the Flaviviridae family in an outbreak of equine serum hepatitis.* Proc Natl Acad Sci U S A, 2013. **110**(15): p. E1407-15.
24. Vlachakis, D., V.L. Koumandou, and S. Kossida, *A holistic evolutionary and structural study of flaviviridae provides insights into the function and inhibition of HCV helicase.* PeerJ, 2013. **1**: p. e74.
25. Pickett, B.E., et al., *ViPR: an open bioinformatics database and analysis resource for virology research.* Nucleic Acids Res, 2012. **40**(Database issue): p. D593-8.

26. Tamura, K., et al., *MEGA6: Molecular Evolutionary Genetics Analysis version 6.0.* Mol Biol Evol, 2013. **30**(12): p. 2725-9.
27. Waterhouse, A.M., et al., *Jalview Version 2--a multiple sequence alignment editor and analysis workbench.* Bioinformatics, 2009. **25**(9): p. 1189-91.
28. Sobie, E.A., *An introduction to MATLAB.* Sci Signal, 2011. **4**(191): p. tr7.
29. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
30. Abascal, F., R. Zardoya, and D. Posada, *ProtTest: selection of best-fit models of protein evolution.* Bioinformatics, 2005. **21**(9): p. 2104-5.
31. Gascuel, O., *BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data.* Mol Biol Evol, 1997. **14**(7): p. 685-95.
32. Cai, J.J., et al., *MBEToolbox: a MATLAB toolbox for sequence data analysis in molecular biology and evolution.* BMC Bioinformatics, 2005. **6**: p. 64.
33. Warrener, P. and M.S. Collett, *Pestivirus NS3 (p80) protein possesses RNA helicase activity.* J Virol, 1995. **69**(3): p. 1720-6.
34. Rho, J., et al., *The arginine-1493 residue in QRRGRTGR1493G motif IV of the hepatitis C virus NS3 helicase domain is essential for NS3 protein methylation by the protein arginine methyltransferase 1.* J Virol, 2001. **75**(17): p. 8031-44.
35. Li, C., et al., *The DEAD-box RNA helicase DDX5 acts as a positive regulator of Japanese encephalitis virus replication by binding to viral 3' UTR.* Antiviral Res, 2013. **100**(2): p. 487-99.
36. Koonin, E.V. and V.V. Dolja, *Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences.* Crit Rev Biochem Mol Biol, 1993. **28**(5): p. 375-430.
37. Wu, J., et al., *Structure of the Flavivirus helicase: implications for catalytic activity, protein interactions, and proteolytic processing.* J Virol, 2005. **79**(16): p. 10268-77.
38. Iacono-Connors, L.C. and C.S. Schmaljohn, *Cloning and sequence analysis of the genes encoding the nonstructural proteins of Langat virus and comparative analysis with other flaviviruses.* Virology, 1992. **188**(2): p. 875-80.
39. Koonin, E.V., *Computer-assisted identification of a putative methyltransferase domain in NS5 protein of flaviviruses and lambda 2 protein of reovirus.* J Gen Virol, 1993. **74 (Pt 4)**: p. 733-40.
40. Lu, G. and P. Gong, *Crystal Structure of the full-length Japanese encephalitis virus NS5 reveals a conserved methyltransferase-polymerase interface.* PLoS Pathog, 2013. **9**(8): p. e1003549.
41. Gong, P. and O.B. Peersen, *Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase.* Proc Natl Acad Sci U S A, 2010. **107**(52): p. 22505-10.