This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.
An updated evolutionary study of Flaviviridae NS3 helicase and NS5 RNA-dependent RNA polymerase, reveals novel invariant motifs as potential pharmacological targets.

Louis Papageorgiou\textsuperscript{a,b}, Styliani Loukatou\textsuperscript{c}, Kosida Sofia\textsuperscript{d}, Dimitrios Maroulis\textsuperscript{b}, Dimitrios Vlachakis\textsuperscript{a,}\textsuperscript{*}

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, centering 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-dependent 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.

Encephalitis virus (TBEV) constitute some important human pathogens of the genus Flavivirus. The Classical Swine Fever Virus (CSFV) 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\textsuperscript{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\textsuperscript{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\textsuperscript{6-8}. Virions of Flaviviridae family are spherical, about 40-60 nm in diameter\textsuperscript{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\textsuperscript{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\textsuperscript{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.

* Computational Biology & Medicine Group, Biomedical Research Foundation, Academy of Athens, Sarouou Efessiou 4, Athens 11527, Greece
\textsuperscript{a} Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, University Campus, Athens, 15784, Greece
\textsuperscript{b} IMGT*, the international ImMunoGeneTics information system*, Universite de Montpellier, Laboratoire d’Immunogenetique Moleculaire LIGIM, UPR CNRS 1142, Institut de Genetique Humaine, 141 rue de la Cardanille, Montpellier, 34396 cedex 5, France.
\textsuperscript{c} Electronic supplementary information (ESI) available: Phylogenetic Analyses results (Newick tree format) and four supplementary figures. See DOI: 10.1039/k0xx00000x
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

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).

<table>
<thead>
<tr>
<th>NS5 protein sequences data set</th>
<th>NS3 protein sequences data set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aedes flavivirus</td>
<td>4</td>
</tr>
<tr>
<td>Apoi virus</td>
<td>2</td>
</tr>
<tr>
<td>Border disease virus</td>
<td>12</td>
</tr>
<tr>
<td>Bovine viral diarrhea virus</td>
<td>114</td>
</tr>
<tr>
<td>Cell fusing agent virus</td>
<td>1</td>
</tr>
<tr>
<td>Chimeric Dengue virus</td>
<td>0</td>
</tr>
<tr>
<td>Classical swine fever virus</td>
<td>75</td>
</tr>
<tr>
<td>Culex flavivirus</td>
<td>23</td>
</tr>
<tr>
<td>Dengue virus</td>
<td>3747</td>
</tr>
<tr>
<td>Flavivirus</td>
<td>2</td>
</tr>
<tr>
<td>GB virus</td>
<td>66</td>
</tr>
<tr>
<td>Greek goat encephalitis virus</td>
<td>1</td>
</tr>
<tr>
<td>Hepatitis C virus</td>
<td>1497</td>
</tr>
<tr>
<td>Hepatitis GB virus</td>
<td>11</td>
</tr>
<tr>
<td>Japanese encephalitis virus</td>
<td>203</td>
</tr>
<tr>
<td>Kamiti River virus</td>
<td>3</td>
</tr>
<tr>
<td>Kysanur forest disease virus</td>
<td>22</td>
</tr>
<tr>
<td>Langat virus</td>
<td>4</td>
</tr>
<tr>
<td>Louping ill virus</td>
<td>3</td>
</tr>
<tr>
<td>Modoc virus</td>
<td>2</td>
</tr>
<tr>
<td>Montana myotis leukoencephalitis virus</td>
<td>2</td>
</tr>
<tr>
<td>Murray Valley encephalitis virus</td>
<td>15</td>
</tr>
<tr>
<td>Omsk hemorrhagic fever virus</td>
<td>2</td>
</tr>
<tr>
<td>Pegivirus</td>
<td>3</td>
</tr>
<tr>
<td>Pestivirus</td>
<td>3</td>
</tr>
<tr>
<td>Powassan virus</td>
<td>15</td>
</tr>
<tr>
<td>Rio Bravo virus</td>
<td>3</td>
</tr>
<tr>
<td>Rodent hepacivirus</td>
<td>3</td>
</tr>
<tr>
<td>Rodent pegivirus</td>
<td>2</td>
</tr>
<tr>
<td>Royal Farm virus</td>
<td>3</td>
</tr>
<tr>
<td>Simian pegivirus</td>
<td>3</td>
</tr>
<tr>
<td>Spanish sheep encephalitis virus</td>
<td>1</td>
</tr>
<tr>
<td>St. Louis encephalitis virus</td>
<td>12</td>
</tr>
<tr>
<td>Tamanabat virus</td>
<td>3</td>
</tr>
<tr>
<td>Theiler’s disease-associated virus</td>
<td>1</td>
</tr>
<tr>
<td>Tick-borne encephalitis virus</td>
<td>127</td>
</tr>
<tr>
<td>Usutu virus</td>
<td>15</td>
</tr>
<tr>
<td>West Nile virus</td>
<td>1067</td>
</tr>
<tr>
<td>Yellow fever virus</td>
<td>72</td>
</tr>
<tr>
<td>Yokose virus</td>
<td>1</td>
</tr>
</tbody>
</table>

Total 7175  Total 7181
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 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 Evolutionary 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.

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 highlighted 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 non structural proteins. The multiple sequence alignment of the NS5 protein sequences revealed that NS5 region contained conserved motifs within all the Flaviviridae family virus [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-methionine transferase 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. 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, 4BE6 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 described 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 Aopii 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 *Hepadivirus* 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 Nile virus, St. Louis Encephalitis, Japanese Encephalitis virus, Usutu virus, Murray Valley Encephalitis virus and Yellow fever. In the tick-borne Cluster C were clustered the tick-borne related Flaviviruses including Tick-borne Encephalitis virus, Louping Ill virus, Powassan virus and Kysasanur Forest disease virus. In addition, in the subgroup D two smaller subgroups were formed within the genus *Flavivirus*. 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 includes 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 genera. Finally, the last genus *Pestivirus* is shown to be evolutionarily differentiated from all the other three genera in a clear separated branch which contains 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 as NS3 trees. Moreover, the *Flavivirus* subgroup D with the unclear virus transmission including Apoi virus, Rio Bravo virus, Montana Myotis Leukoencephalitis, 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 *Hepacivirus* presented in four main monophyletic trees, the genus *Pegivirus* and *Pestivirus* 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 Kysasanur Forest disease related viruses) and the subgroup D with tick-borne viruses cluster (Apoi, Rio Bravo, Montana Myotis, Leukoencephalitis, 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 Pestivirus (~719 aa), Hepacivirus (~590 aa) and Pegivirus (~574 aa). Moreover, the annotated Pestivirus (~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 *Hepacivirs* (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 *Hepacivirus* (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


