Title: Computationally-inferred structural landscape of the complete genomes of Dengue serotypes and other viral hemorrhagic fevers

Authors: Riccardo Delli-Ponti[1*], Marek Mutwil[1*]

[1]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore, 637551, Singapore

*Corresponding author:

Riccardo Delli-Ponti

School of Biological Sciences,

Nanyang Technological University,

60 Nanyang Drive,

637551, Singapore,

Singapore

Marek Mutwil

School of Biological Sciences,

Nanyang Technological University,

60 Nanyang Drive,

637551, Singapore,

Singapore

ABSTRACT

With more than 300 million potentially infected people every year, and with the expanded habitat of mosquitoes due to climate change, dengue cannot be considered anymore only a tropical disease. The RNA secondary structure is a functional characteristic of RNA viruses, and together with the accumulated high-throughput sequencing data could provide general insights towards understanding virus biology. Here, we profiled the RNA secondary structure of >7500 complete viral genomes from 11 different families of viral hemorrhagic fevers, including dengue serotypes, ebola, and yellow fever. Our results suggest not only an interesting lack of secondary structure for very aggressive and virulent viruses such as DENV-2 and ebola but also a different correlation between secondary structure and the number of interaction sites with human proteins, for example, with an anti-correlation of -0.84 for chikungunya. We demonstrate that the secondary structure and presence of protein-binding domains in the genomes to build similarity trees can be used to classify the viruses. We also used structural data to study the geographical distribution of dengue, finding a significant difference between DENV-3 from Asia and South-America, which could imply different evolutionary routes of this subtype.

INTRODUCTION

Dengue is a mosquito-borne virus that can potentially infect more than 300 million people a year in more than 120 countries [1,2]. Dengue infection can further evolve into a severe hemorrhagic fever, which could lead to shock and death. Due to climate change, the disease is now threatening an increasing number of countries, with cases reported in Europe [2]. The existence of four different serotypes (DENV-1, 2, 3, 4), with also a fifth recently reported [3], complicates the development of an effective vaccine [4]. The four serotypes show not only significant differences in sequence similarity [5] but also distinctive infection dynamics. For example, DENV-1 is the most wide-spread serotype, followed by DENV-2 [6,7], which is also more often associated with severe cases [8]. However, the mechanisms behind Dengue infections and the complete set of differences between the serotypes are still unclear.

Dengue is only one of the different Viral Haemorrhagic Fevers (VHF), comprising single-stranded RNA viruses from different families, such as flavivirus and filovirus. These viruses can be extremely lethal in humans, as in the case of Ebola. Other VHFs show not only similarities to Dengue in terms of symptoms, but also in the transmission vector. For example, Yellow Fever is also a mosquito-borne VHFs, with higher mortality but a slower rate of evolutionary change compared to Dengue [9]. The mild VHF Chikungunya shares the same vector with Dengue, mosquito *Aedes aegypti*, and the two viruses can even coexist in the same mosquito [10]. However, even with thousands of viral genomes available, the most common approach to understand the similarities between those viruses is still based only on the sequence.

The secondary structure of RNA viruses is fundamental for many viral functions, from encapsidation to egression from the cell and host defence [11–13]. Specific structures in the UTRs were found to be functional, for example, in Dengue, but also in HIV and coronaviruses [11,14,15]. Other structural regions, including the 3' UTR, were found conserved not only in Dengue serotypes but also between Dengue and Zika [16]. Moreover, the single-stranded RNA (ssRNA) viruses preserve their structure (folding) even if their sequence mutates rapidly [17,18]. Thus, the folding shows the potential to be used to classify different

viral species and subspecies.

However, while the RNA secondary structure is an informative element to characterise viruses, the secondary structure of only a few viral genomes has been experimentally characterized [16,19]. Consequently, while thousands of viral genomes have been sequenced, we can only rely on in silico data to study their secondary structure. Furthermore, predicting the RNA secondary structure of entire viral genomes can be challenging, due to usually large sizes of >10000 nucleotides (nt), where most thermodynamic algorithms used to model the secondary structure drop in performance after 700 nt [20]. In our work, we computationally profiled the RNA secondary structure of >7500 viral genomes (prioritising Dengue serotypes and in general VHFs) using Computational Recognition of Secondary Structure (CROSS), a neural network trained on experimental data, which was successfully applied to predict HIV genome structure [21]. We mapped the secondary structure properties of the viruses on the world map, to study the genome interaction with proteins, and to further classify and understand the viruses.

MATERIAL AND METHODS

Source of viral genomes

The viral genomes were downloaded from NCBI, selecting for each specific virus the fasta sequences containing the keyword *complete genome*. NCBI data was also used to extract the geographical information of Dengue viruses as well as the serotypes. Fragmented or incomplete genomes passing the filter were also removed by looking at the sequence length.

RNA Secondary structure

The secondary structure profiles were computed using the CROSS algorithm. CROSS is a neural network-based machine learning approach trained on experimental data, able to quickly profile large and complex molecules such as viral genomes without length restrictions. CROSS was already used to profile the complete HIV genome, also showing an AUC of 0.75 with experimental SHAPE data [21]. For a comprehensive analysis, we used the *Global Score* model, considering nucleotides with a score >0 as double-stranded, and <0 as single-stranded. For the purpose of computing the structural content of a complete genome (i.e., % double-stranded nucleotides), the total number of nucleotides with a score >0 was averaged for the total length of each genome. For the plots showing the complete secondary structure of dengue regions, we used the MFE structure computed using RNAfold [22].

Protein-RNA interactions

To analyse the protein binding motifs in the viral genomes, we selected 5-mer motifs from the table S3 of Dominguez et al. [23]. All of the 520 possible redundant motifs (270 non-redundant) were selected for further analysis. We scanned for the motifs on the complete genomes of the different strains, selecting only perfect matches. The number of motifs normalized by the average genome length of the different families of viruses was used to define a score for the number of potential interactions with proteins, according to the following formula:

$$\frac{m}{avg(n)}$$

where m is the number of exact motifs found in a genome, and avg(n) is the average length of the genome of a specific species.

Hierarchical clustering

The structural content, the averaged number of binding domains, and their correlation were employed to build dendrograms. To this end, we computed the Euclidean distance between the values associated with each virus using statistical software R. We then used the *hclust* function based on a *ward.D2* module to build the dendrograms according to the hierarchical clustering.

RESULTS

Structural properties of the Dengue genomes

Here, we analysed, for the first time to our knowledge, the secondary structure profiles of the complete genomes of more than 7500 ss-RNA viruses (Table 1; Materials and Methods: Source of viral genomes). The structural profiles were generated using the CROSS algorithm, a fast and comprehensive alternative to profile the structural content (i.e., % of double-stranded nucleotides) of long and complex RNA molecules, such as viruses ([21]; see Materials and Methods: RNA secondary structure).

To analyse Dengue secondary structure, we selected one strain for each serotype, selecting strains that were widely used in previous publications [24]. In general, the four serotypes show significant differences in sequence, with around 65-70% sequence similarity [5]. Their secondary structure also shows notable differences (Figure 1). For example, the 3' UTR of DENV-1 shows a peculiar structural peak, compared to the others. Interestingly, DENV-1 and DENV-2 share the highest structural peak around 6000 nucleotides, while DENV-3 and DENV-4 also have the highest structural peak in common, but at position 4000.

We further expanded our analysis to cover the 4 different serotypes of the dengue virus, comprising ~4000 genomes (Figure 2; Table 1). The analysis revealed that DENV-2 and DENV-3 are less structured than DENV-1 and DENV-4. A lower amount of complex secondary structure can improve the translation efficiency [25], suggesting that DENV-2 and DENV-3 could be more efficiently translated in the host cell.

Comparison of structural properties of the VHF genomes

Interestingly, dengue serotypes tend to be less structured than other Flaviviruses, such as West Nile Fever, Yellow Fever, Tick-borne Encephalitis, and Japanese Encephalitis (Figure 2). Even if not properly a VHF, we also used as comparison Zika, due to the similarities with Dengue not only in the vector (*Aedes aegypti*), but also in terms of secondary structure domains [16]. Interestingly, while Tick-borne Encephalitis and Zika genomes are more structured (average double-stranded nucleotides > 56%), Nile Fever and Japanese Encephalitis have a similar structural distribution, especially since they are also close in the species tree [26]. To further compare and classify the secondary structure of other VHFs outside of flavivirus, we also included >500 genomes of Ebola and Chikungunya (Figure 2, Table 1). The analysis revealed that the other viruses are significantly more structured than Dengue (mean structural content for Flaviviruses and Dengue serotypes is 0.55, 0.51, respectively; Kolmogorov-Smirnov < 2.2e-16), with the exception of Ebola, which is not only the most deadly virus in our dataset (50% mortality rate according to WHO; [27]), but also one of the less structured (mean structural content 0.50).

Structural properties of the untranslated regions of VHF genomes

To further study the secondary structure content for the >7500 viral genomes, we also analysed the 5' and 3' UTRs (first 1000 nt considered 5' UTR; last 1000 nt considered 3' UTR; Figure 3a, b). DENV-3 is the only serotype with both UTRs more structured than the entire genome (5' UTR= 0.55 and 3' UTR=0.53; Figure 2), while DENV-1 has a highly structured 5' UTR (structural content = 0.53). This result is in line with the experimental Parallel Analysis of RNA Structure (PARS) data coming from human RNAs, where the UTRs were more structured than the CDS [28]. This suggests that some viruses tend to mimic the secondary structure of human mRNAs to be efficiently translated by the cellular machinery [29]. This is also further supported in Dengue, where a complex structure at the 3' UTR was shown to mimic the absent polyA, to enhance translation [30]. Interestingly, Ebola has the least structured UTRs (structural content 5' UTR = 0.46; 3' UTR = 0.41). In Zika, the 3' UTR is more structured than the 5' (Figure 3c; 3' UTR=0.56, 5' UTR = 0.50). Chikungunya shows not only the highest structural variability in the 3' UTR (3' UTR = 0.43, Figure 3b), but also a very structured 5' UTR (Figure 3a). Finally, Ebola, DENV-1, and DENV-3 exhibit a more structured 5' UTR, especially when compared with DENV-2, DENV-4 and Japanese Encephalitis, which tend to be more structured (Figure 3c).

Structural content can be used to classify VHFs

The overall similarities and differences in structure are an additional feature that could be employed to characterise the different viruses. For the next step, the structural content (mean of the % of double-stranded nucleotides for all the viral genomes) in a specific species was used to hierarchically group the 11 different viruses (Materials and Methods: Hierarchical clustering; Table 1). The resulting dendrogram clustered the Dengue serotypes, showing that they are more structurally similar compared to other viruses (Figure 4). The structural similarities of Dengue serotypes, together with the similarity between West Nile Fever and Japanese Encephalitis, are in agreement with the Phylogenetic Tree of Viral Hemorrhagic Fever [26].

The structural content revealed interesting clustering of the viruses. For example, while having a different genome sequence, Dengue viruses also cluster together with Ebola, since they share a less structured genome. Interestingly, DENV-2, the most virulent and spread serotype [7,8], is similar in structural content to Ebola, probably the most lethal viral hemorrhagic fever [27]. According to its structural content, Zika is also part of the sub-cluster, together with West Nile Fever and Japanese Encephalitis. The mosquito-transmitted Yellow Fever and Chikungunya form a cluster, indicating that their structural content is similar. This is partially in agreement with the VHFs tree [26]. Interestingly, since Tick-borne Encephalitis is more structured than any of these viruses, it forms an outlier. To conclude, these results indicate that the level of secondary structure inside a viral genome can be used as a metric to build a tree of similarities, which could be further employed to classify viruses.

Interaction between viral genomes and human host proteins can be used to classify VHFs

During translation and replication, ss-RNA viruses are naked RNA molecules inside human host cells. Previous studies already showed that genomes of the Dengue viruses interact with multiple human proteins during the infection and that the protein binding can enhance or inhibit the virulence [31]. Furthermore, RNA binding proteins tend to exhibit an altered activity during viral infection, in some cases due to the presence of highly abundant viral RNA, which

can compete for the interaction with cellular RNA [32]. To study the relationship between human proteins and the viral RNA structures, we selected binding motifs from RNA Bind-n-Seq (RBNS) data of 78 human RNA-binding proteins [23], and searched the complete viral genomes for these motifs (Materials and Methods: Protein-RNA interactions). We observed that the 4 dengue serotypes have a different presence of protein binding domains, with DENV-2 showing the highest number of motifs, followed by DENV-4 (Supplementary Figure 1).

Similarly to the structural content analysis above, we used the number of protein binding domains to classify the viruses. To further understand how the connection between structure and interaction with proteins can classify viruses, we compared the resulting trees (Figure 4). Interestingly, the Dengue cluster is almost perfectly maintained, except that, for the number of protein interactions, DENV-2 is more similar to Chikungunya than Ebola, which in turn is more related to Yellow Fever. Furthermore, clustering of West Nile Fever, Japanese Encephalitis and Zika is partially maintained when using structure and interaction with proteins. To conclude, by analyzing thousands of different viral genomes, we identified specific clusters, conserved both in terms of secondary structure content and the potential number of interactions with proteins.

Relationship between the structural content and number of protein interaction motifs

Since both the structural content and the number of binding motifs could be used to classify the viruses, we hypothesized that there is a correlation between these two features. We found an overall high anti-correlation (r=-0.74; p-value < 2.2e-16) between the number of protein-binding motifs and the structural content in Dengue, meaning that less structured Dengue viruses tend to bind more proteins (Figure 5a). Interestingly, the different dengue serotypes cluster together according to their structure and the interaction with proteins (Figure 5a). Also, the serotypes show a different trend when independently analysed, with DENV-3 and DENV-4 exhibiting the highest influence of the structure on the number of possible interacting proteins (r is -0.31 and -0.36; p-value < 2.2e-16 and 7.177e-07, respectively; Figure 5b).

Next, we compared the secondary structure and protein binding motifs of the other VHFs. The general picture is quite complex, with some viruses showing a propensity to interact with proteins if they are more or less structured. Similarly to dengue, the mild hemorrhagic fever Chikungunya shows a high anticorrelation (r=-0.84; p-value < 2.2e-16, Figure 6). Conversely, Tick-borne Encephalitis and Zika show a positive correlation (Pearson correlation of 0.23 and 0.89; p-value 0.01 and <2.2e-16 respectively; Figure 6). The high positive correlation, especially for Zika, suggests that specific highly-structured viruses are also prone to interact with human proteins. DENV-3 and DENV-4 act similarly to Chikungunya, having unstructured genomes, allowing potentially more interactions with proteins (Figure 5b). Finally, Japanese Encephalitis and West Nile Fever show almost no correlation of the structure and on the possible interaction with proteins (r~0). These results indicate a complex relationship between the structural content and protein-binding motifs.

To further understand this phenomenon, we used the correlation between structure and number of motifs to cluster the different viruses (Supplementary Figure 2). Similarly to what we observed above, Japanese Encephalitis and West Nile Fever are grouped together, but this time also close to DENV-2 and Ebola. While Zika is the outlier together with Chikungunya, the first is a Flavivirus but not a VHF, while the second is a mild VHF but also

a Togavirus (Supplementary Figure 2).

Geographical distribution of Dengue serotypes

We also studied the geographical connection between Dengue serotypes and the secondary structure content. We found that the African DENV-1 is predicted to be more structured than the other serotypes, even when compared with the Asian strains (Supplementary Figure 23 Kolmogorov-Smirnov = 0.001), while on the contrary, Asian DENV-3 is more structured than the African strains (Kolmogorov-Smirnov = 0.07; Supplementary Figure 3).

The correlation between binding motifs and secondary structure at the geographical level portrays a quite complex scenario (Figure 7). The clearest clusters are identifiable for DENV-3, where the South-American strains are not only less structured, but also highly interacting with proteins. DENV-3 from South-America is also more distant from the Asian strains (Euclidean distance centroids x1000=5), compared, for example, with DENV-1 (Euclidean distance centroids x1000=1).

DISCUSSION

The genomes of viral hemorrhagic fevers show a high level of secondary structure, especially in the UTR regions [14,16]. This secondary structure is thought to be needed for different viral mechanisms, such as packaging and egression [11–13]. However, a comprehensive secondary structural landscape for their genomes was lacking. In our work, we computationally profiled and analysed the secondary structure profiles of more than 7500 complete viral genomes, including almost 4000 dengue samples, and 3500 other viral hemorrhagic fever-causing viruses. By studying the structural profiles, we observed that Dengue viruses are predicted to be less structured compared to viruses such as Zika, Yellow Fever, and West Nile Fever (Figure 2). Conversely, Dengue serotypes still tend to retain structured UTRs, probably to be efficiently translated by the cellular machinery, similarly to human mRNAs [25,28].

We also identified a correlation between the secondary structure and the number of protein binding domains, implying that the secondary structure is employed to regulate potential binding with proteins, as observed for human RNAs [33]. For viruses, the situation is more complex, as we observed a significantly positive (Tick-borne Encephalitis and Zika) and negative (Chikungunya, DENV-3, and DENV-4) relationship between secondary structure and the potential interaction with proteins. We speculate that the opposite behaviors of the viruses are caused by the fundamental differences between viruses. For example, Zika (positive RNA genome, positive correlation between secondary structure and interaction with proteins) and Chikungunya (negative genome, negative correlation) belong to different viral families (Flavivirus and Togavirus respectively), and have a completely different capsid.

We also analysed the secondary structure at a geographical level, showing that DENV-3 strains from South-America and Asia have different patterns in their structure and the potential interaction with proteins, especially when compared with DENV-1 and DENV-2 (Figure 7). Interestingly, this is in line with DENV-3 being the youngest serotype and the only one with a proposed origin not in Asia but in America [34]. This could explain the niched behaviour of DENV-3 in terms of structure and protein interactions, as well as supporting a possible independent origin of Asia and and American DENV-3. Interestingly, DENV-4 shows a similar trend, but there are too few samples available to explain its evolution.

To our knowledge, this is the first study that employed secondary structural content and the presence of protein-binding domains to build similarity trees between VHFs. The secondary structure and interaction with proteins can be used to cluster the viruses in agreement with previous phylogenetic trees, such as Dengue serotypes, and Japanese Encephalitis with West Nile Fever. Conversely, some relationships are surprising, as for example, DENV-2 is closer to Ebola when the secondary structure is used to establish similarity, but not when using the interaction with proteins. This result suggests how different measures, especially the secondary structure content, could be used to classify further and characterise different classes of viruses.

Our massive computational analysis provided novel results regarding the secondary structure and the interaction with human proteins, not only for Dengue serotypes but also for other viral hemorrhagic fevers. We envision that these approaches can be used by the scientific community to classify further and characterise these complex viruses.

Figures

Figure 1. Secondary structure of the four dengue serotypes represented as propensity profiles. Nucleotides with a score >0 are double-stranded, while <0 indicates single-stranded nucleotides. The profile is normalised using the same formula reported in the original paper of CROSS methodology. For each profile, the highest (+) and lowest (-) structural peaks are highlighted. The structures of 200 nt regions, including the most high-propensity double- (red) and single-stranded (gray) regions for each serotype, were computed using *RNAfold.*

Figure 2. Structural content (% double-stranded nucleotides) for all the genomes for the 11 families. The number above each violin plot indicates the number of genomes used in each family.

Figure 3. Structural content of the UTRs of the 11 viral species. (A) Structural content (% double-stranded nucleotides) for all the genomes for the 11 species for the 5' UTR. To have an equal comparison between the different species, we considered the 5' UTR as the first 1000nt. The abbreviations used for the viruses: Jap (Japanese encephalitis), Nile (West Nile Fever), Chiku (Chikungunya), Yellow (Yellow Fever) and Tick (Tick-borne encephalitis). (B) Structural content (% double-stranded nucleotides) for all the genomes for the 11 families for the 3' UTR. To have an equal comparison between the different species, we considered the 3' UTR as the last 1000 nt. (C) The difference for each individual genome between the structural content of the 5' and the 3' UTR. Viruses with more structured 5' UTR are >0 (blue area), while <0 indicates more structured 3' UTRs (green area).

Figure 4. Comparison between the dendrograms obtained using the structural content (left side figure) and the number of binding motifs normalised by the length of the genome (right side figure). The gray lines connect the same virus species and serotypes. The abbreviations used for the viruses: Jap (Japanese encephalitis), Nile (West Nile Fever), Chiku (Chikungunya), Yellow (Yellow Fever) and Tick (Tick-borne encephalitis).

Figure 5. Correlation between the secondary structure and the interaction with proteins for all the dengue genomes. (A) Correlation between the structural content and the averaged number of binding domains for all the dengue genomes. DENV1 is indicated by a blue box, DENV2 by an orange box, DENV3 by a green box, and DENV4 by a golden box. (B) Correlation between the structural content and the averaged number of binding domains independently for the 4 dengue serotypes. The correlation is different when considering each serotype individually.

Figure 6. Correlation between the structural content and the averaged number of binding domains for the remaining 7 viral species. Each point represents a single genome, while x- and y-axes indicate its structural content and the number of protein binding domains divided by the averaged size of the genome, respectively.

Figure 7. Correlation between the structural content and the averaged number of binding domains for the dengue serotypes, according to their geographical distribution. Samples coming from Africa (AF), Asia (AS), Oceania (OC), South America (SA), and North America (NA) were marked using different colours. Each point represents a single genome, while x- and y-axes indicate its structural content and the number of protein binding domains divided by the averaged size of the genome, respectively.

Tables

Table 1. The information regarding the number of genomes available, the family, and the average size of each family for all the viruses used in our analysis.

Supplementary Figures

Supplementary Figure 1. Violin plot showing the interaction with proteins for each dengue serotype, computed as the presence of RNA binding motifs on their genome, averaged for the mean of the length of each serotype.

Supplementary Figure 2. Dendrograms classifying the different viruses using as metric the correlation between structural content and the presence of protein binding domains.

Supplementary Figure 3. Barplot showing for each dengue serotype the differences in structural content (% double-stranded nucleotides) in different geographical samples coming from Africa (AF), Asia (AS), Oceania (OC), South America (SA) and North America (NA).

Bibliography

1. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496,** 504–507 (2013).
2. Brady, O. J. *et al.* Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl. Trop. Dis.* **6,** e1760 (2012).

3. Normile, D. Tropical medicine. Surprising new dengue virus throws a spanner in disease control efforts. *Science* **342,** 415 (2013).
4. The Lancet Infectious Diseases. The dengue vaccine dilemma. *Lancet Infect. Dis.* **18,** 123 (2018).
5. Anoop, M. *et al.* Complete genome sequencing and evolutionary analysis of dengue virus serotype 1 isolates from an outbreak in Kerala, South India. *Virus Genes* **45,** 1–13 (2012).
6. Messina, J. P. *et al.* Global spread of dengue virus types: mapping the 70 year history. *Trends Microbiol.* **22,** 138–146 (2014).
7. Bäck, A. T. & Lundkvist, A. Dengue viruses - an overview. *Infect. Ecol. Epidemiol.* **3,** (2013).
8. Fried, J. R. *et al.* Serotype-specific differences in the risk of dengue hemorrhagic fever: an analysis of data collected in Bangkok, Thailand from 1994 to 2006. *PLoS Negl. Trop. Dis.* **4,** e617 (2010).
9. Sall, A. A. *et al.* Yellow fever virus exhibits slower evolutionary dynamics than dengue virus. *J. Virol.* **84,** 765–772 (2010).
10. Leroy, E. M. *et al.* Concurrent chikungunya and dengue virus infections during simultaneous outbreaks, Gabon, 2007. *Emerging Infect. Dis.* **15,** 591–593 (2009).
11. Lu, K., Heng, X. & Summers, M. F. Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol.* **410,** 609–633 (2011).
12. Stewart, H. *et al.* Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome packaging. *Sci. Rep.* **6,** 22952 (2016).
13. Witteveldt, J. *et al.* The influence of viral RNA secondary structure on interactions with innate host cell defences. *Nucleic Acids Res.* **42,** 3314–3329 (2014).
14. de Borba, L. *et al.* RNA structure duplication in the dengue virus 3’ UTR: redundancy or host specificity? *MBio* **10,** (2019).
15. Yang, D. & Leibowitz, J. L. The structure and functions of coronavirus genomic 3’ and 5’

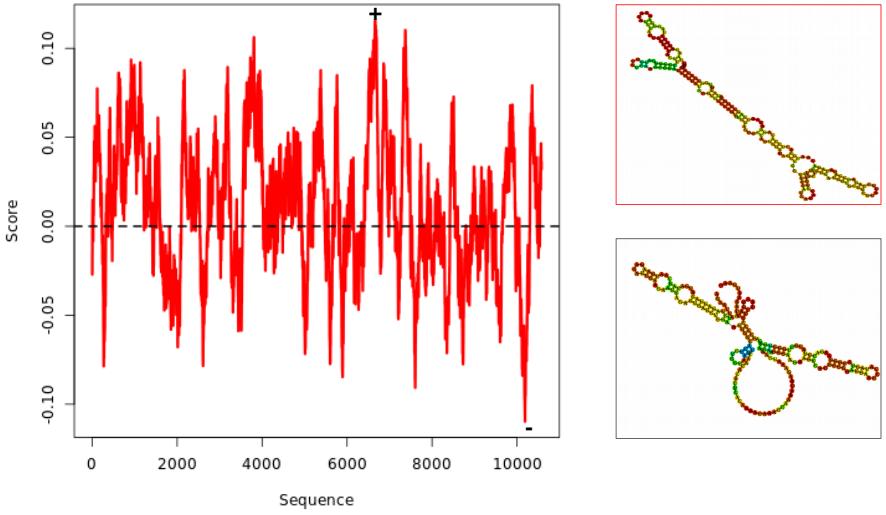
ends. *Virus Res.* **206,** 120–133 (2015).

16. Huber, R. G. *et al.* Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Nat. Commun.* **10,** 1408 (2019).
17. Chursov, A., Frishman, D. & Shneider, A. Conservation of mRNA secondary structures may filter out mutations in Escherichia coli evolution. *Nucleic Acids Res.* **41,** 7854–7860 (2013).
18. Rizvi, T. A. & Panganiban, A. T. Simian immunodeficiency virus RNA is efficiently encapsidated by human immunodeficiency virus type 1 particles. *J. Virol.* **67,** 2681–2688 (1993).
19. Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460,** 711–716 (2009).
20. Hajiaghayi, M., Condon, A. & Hoos, H. H. Analysis of energy-based algorithms for RNA secondary structure prediction. *BMC Bioinformatics* **13,** 22 (2012).
21. Delli Ponti, R., Marti, S., Armaos, A. & Tartaglia, G. G. A high-throughput approach to profile RNA structure. *Nucleic Acids Res.* **45,** e35 (2017).
22. Lorenz, R., Luntzer, D., Hofacker, I. L., Stadler, P. F. & Wolfinger, M. T. SHAPE directed RNA folding. *Bioinformatics* **32,** 145–147 (2016).
23. Dominguez, D. *et al.* Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell* **70,** 854-867.e9 (2018).
24. Chin-Inmanu, K. *et al.* Complete genome sequences of four serotypes of dengue virus prototype continuously maintained in the laboratory. *Microbiol. Resour. Announc.* **8,** (2019).
25. Kramer, M. C. & Gregory, B. D. Does RNA secondary structure drive translation or vice versa? *Nat. Struct. Mol. Biol.* **25,** 641–643 (2018).
26. Paul, B. & Tham, W. L. Controlling dengue: effectiveness of biological control and vaccine in reducing the prevalence of dengue infection in endemic areas. *Health* **08,** 64–74 (2016).
27. Keita, M. *et al.* Subsequent mortality in survivors of Ebola virus disease in Guinea: a

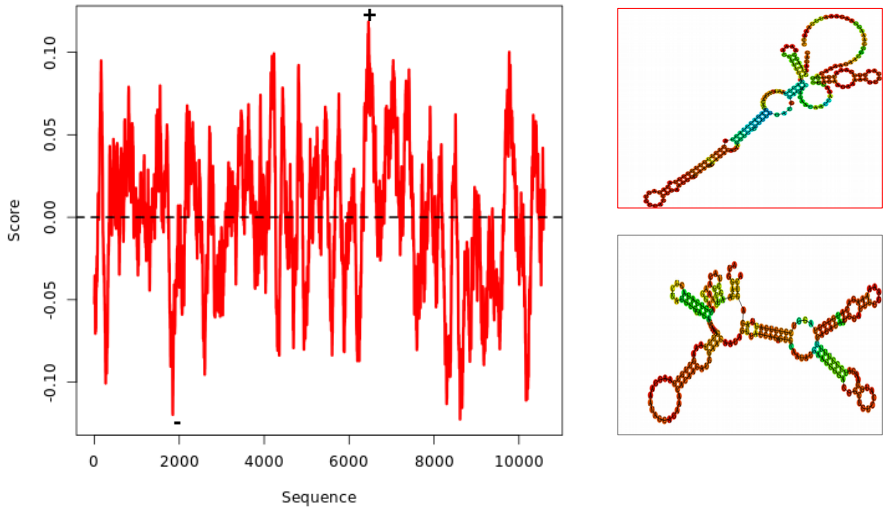
nationwide retrospective cohort study. *Lancet Infect. Dis.* **19,** 1202–1208 (2019).

28. Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* **15,** 469–479 (2014).
29. Reid, D. W. *et al.* Dengue Virus Selectively Annexes Endoplasmic Reticulum-Associated Translation Machinery as a Strategy for Co-opting Host Cell Protein Synthesis. *J. Virol.* **92,** (2018).
30. Clyde, K., Kyle, J. L. & Harris, E. Recent advances in deciphering viral and host determinants of dengue virus replication and pathogenesis. *J. Virol.* **80,** 11418–11431 (2006).
31. Viktorovskaya, O. V., Greco, T. M., Cristea, I. M. & Thompson, S. R. Identification of RNA Binding Proteins Associated with Dengue Virus RNA in Infected Cells Reveals Temporally Distinct Host Factor Requirements. *PLoS Negl. Trop. Dis.* **10,** e0004921 (2016).
32. Garcia-Moreno, M. *et al.* System-wide Profiling of RNA-Binding Proteins Uncovers Key Regulators of Virus Infection. *Mol. Cell* **74,** 196-211.e11 (2019).
33. Cid-Samper, F. *et al.* An Integrative Study of Protein-RNA Condensates Identifies Scaffolding RNAs and Reveals Players in Fragile X-Associated Tremor/Ataxia Syndrome. *Cell Rep.* **25,** 3422-3434.e7 (2018).
34. Costa, R. L., Voloch, C. M. & Schrago, C. G. Comparative evolutionary epidemiology of dengue virus serotypes. *Infect. Genet. Evol.* **12,** 309–314 (2012).

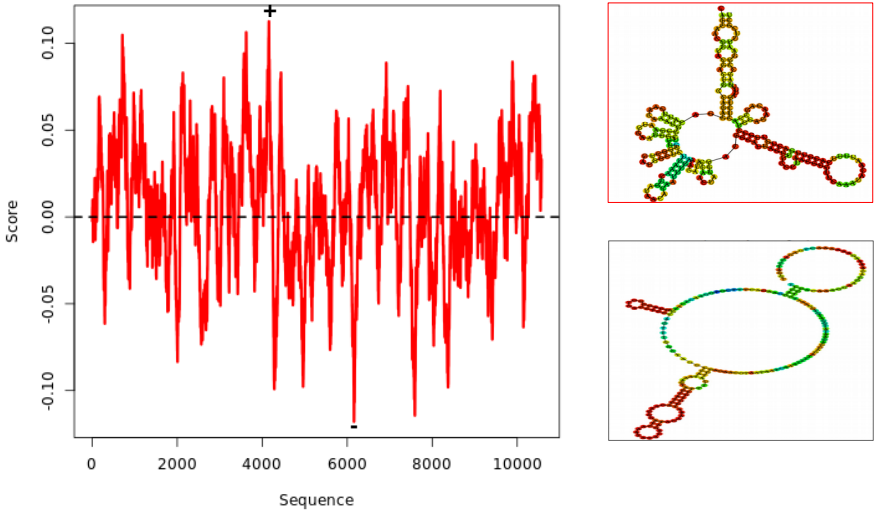
denv-1



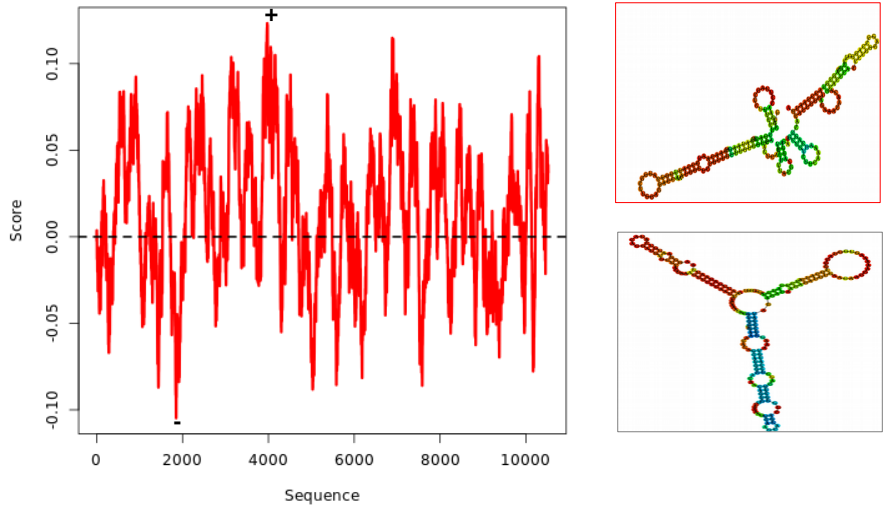
denv-2

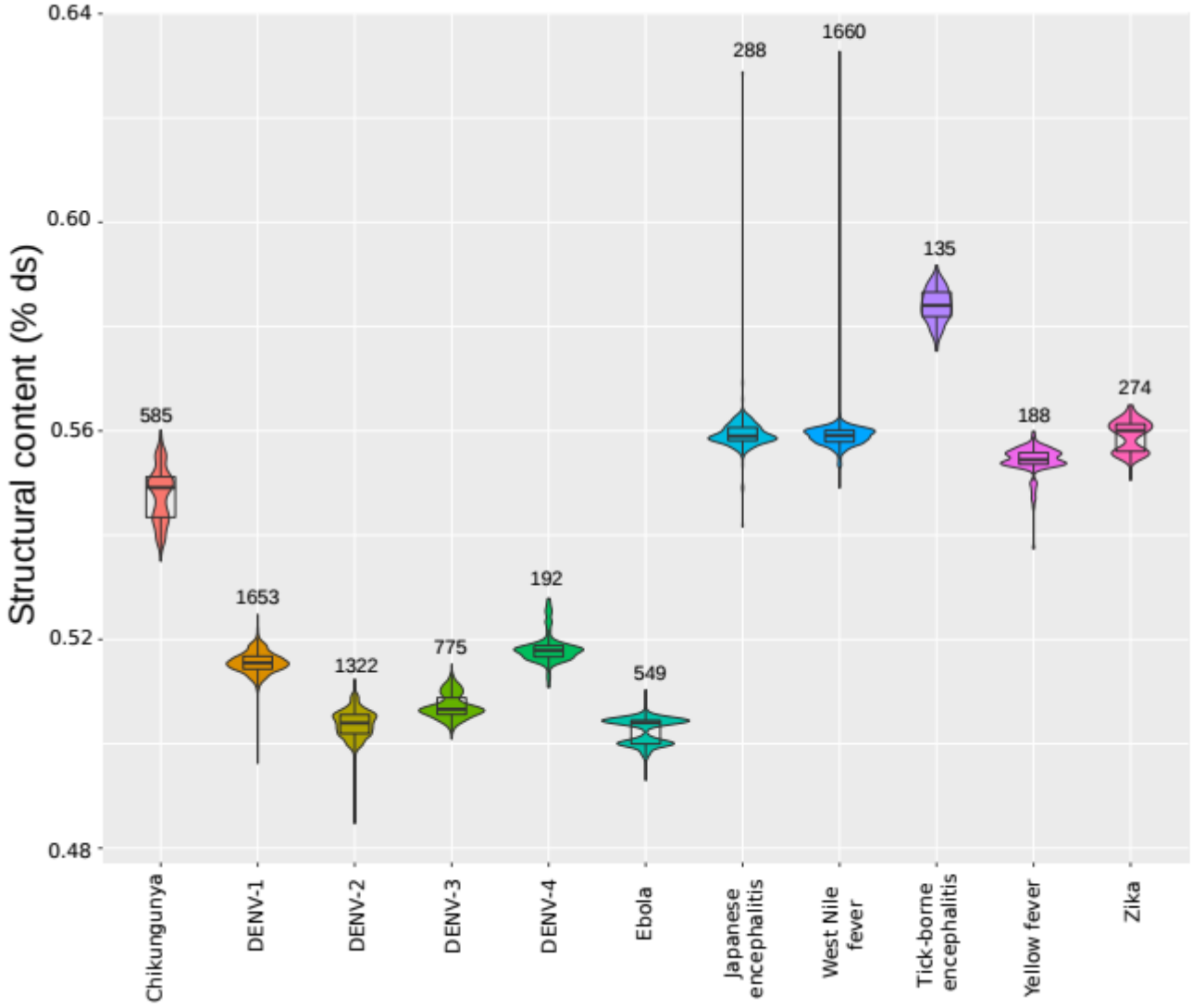


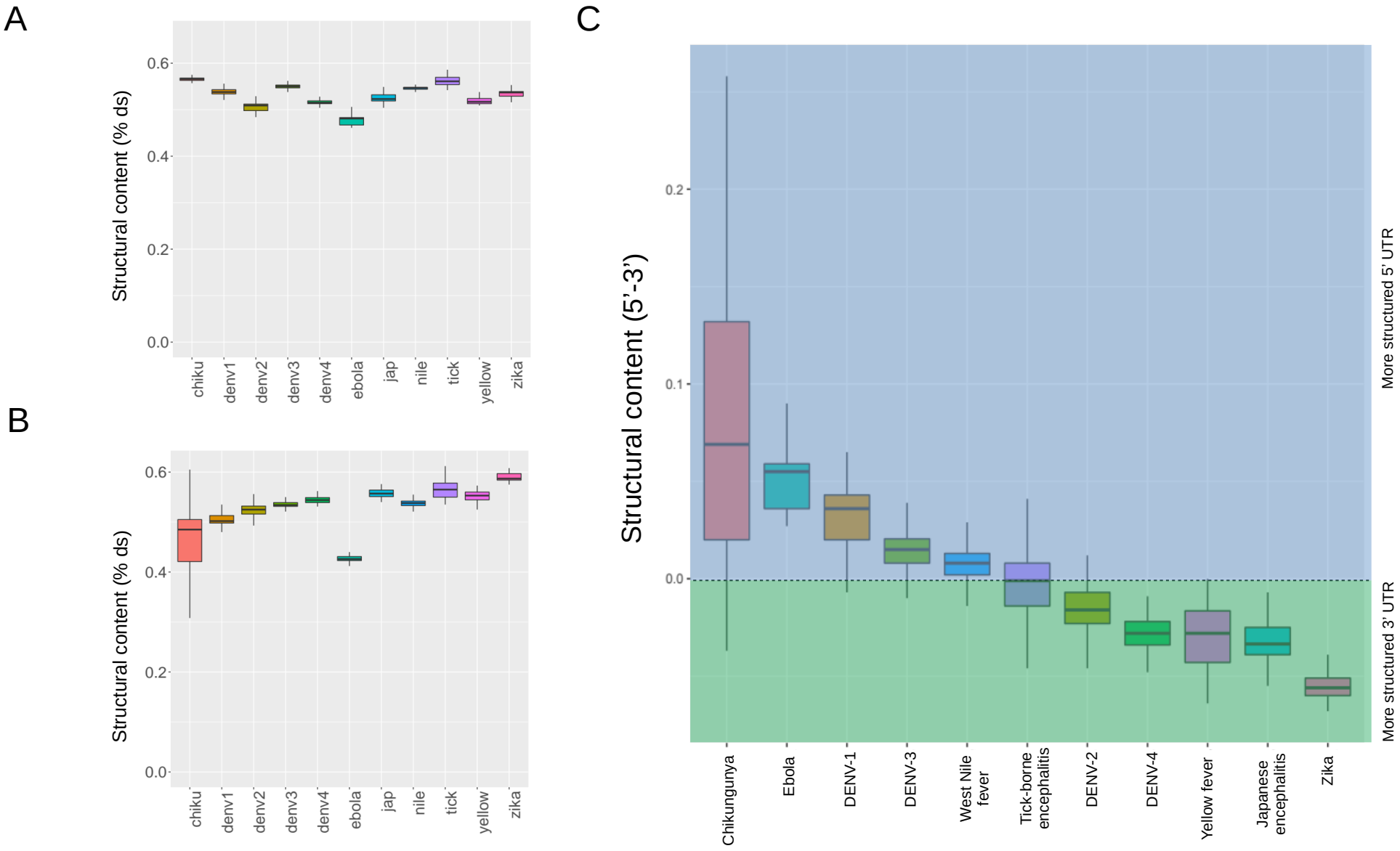
denv-3

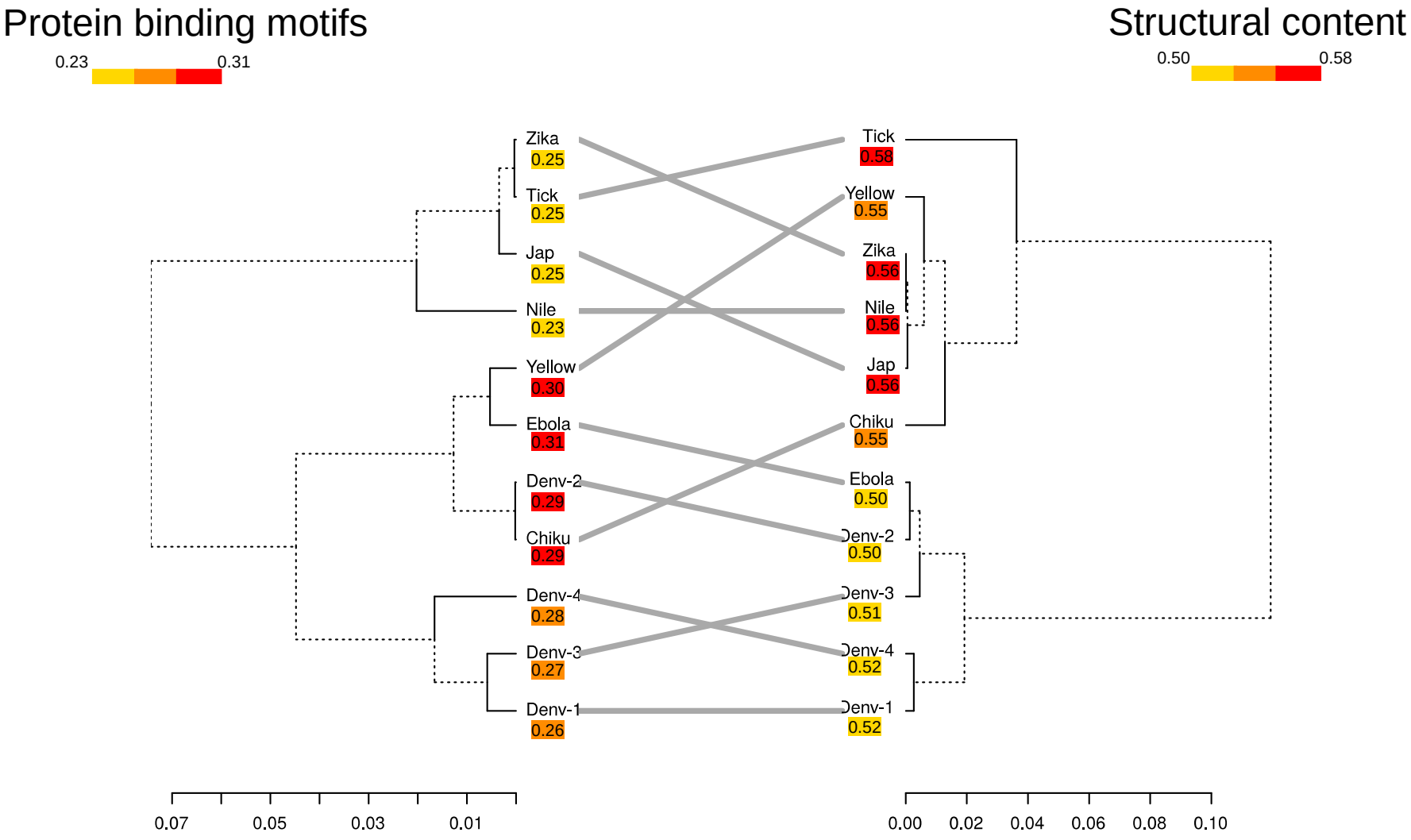


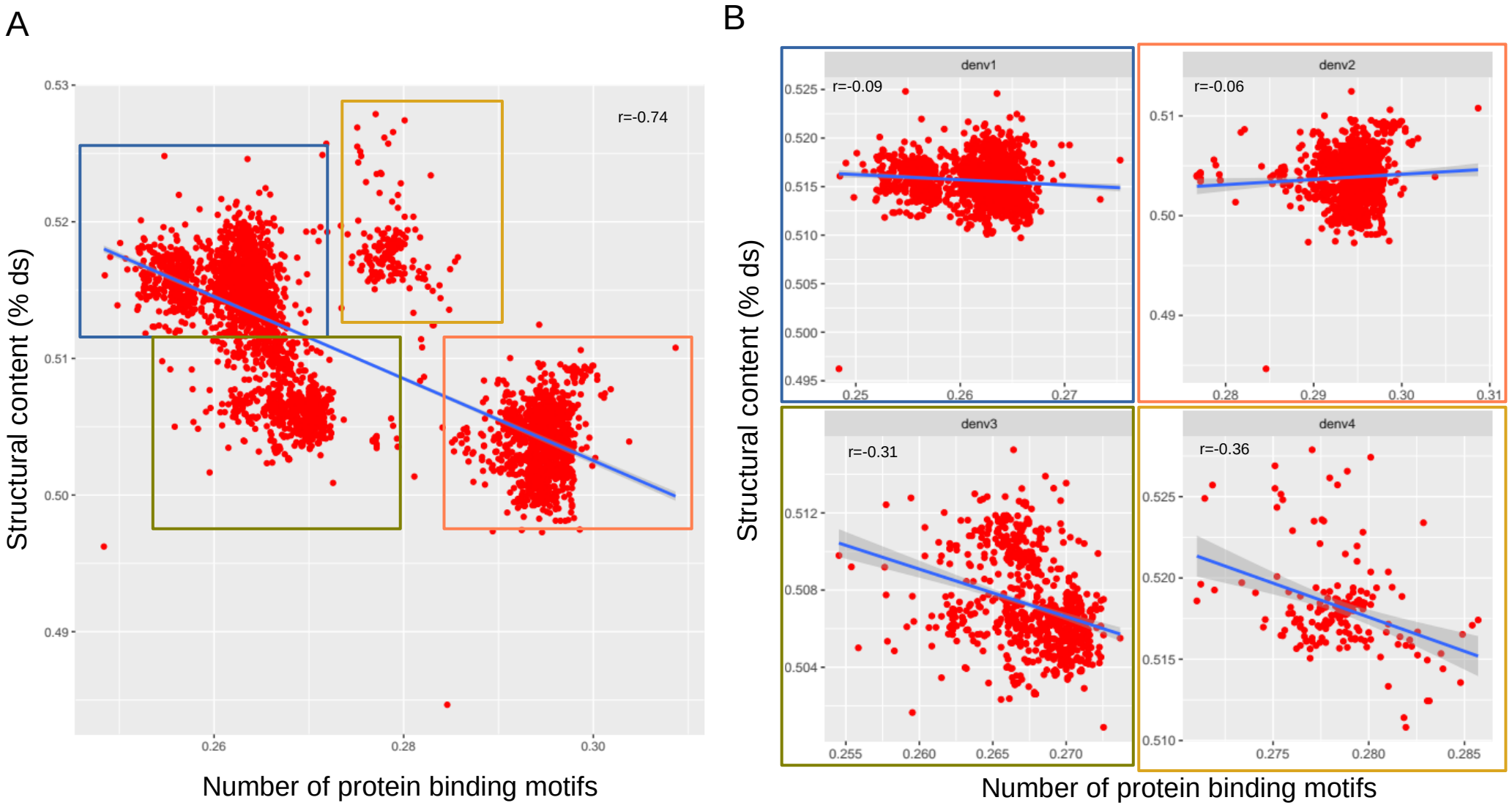
denv-4

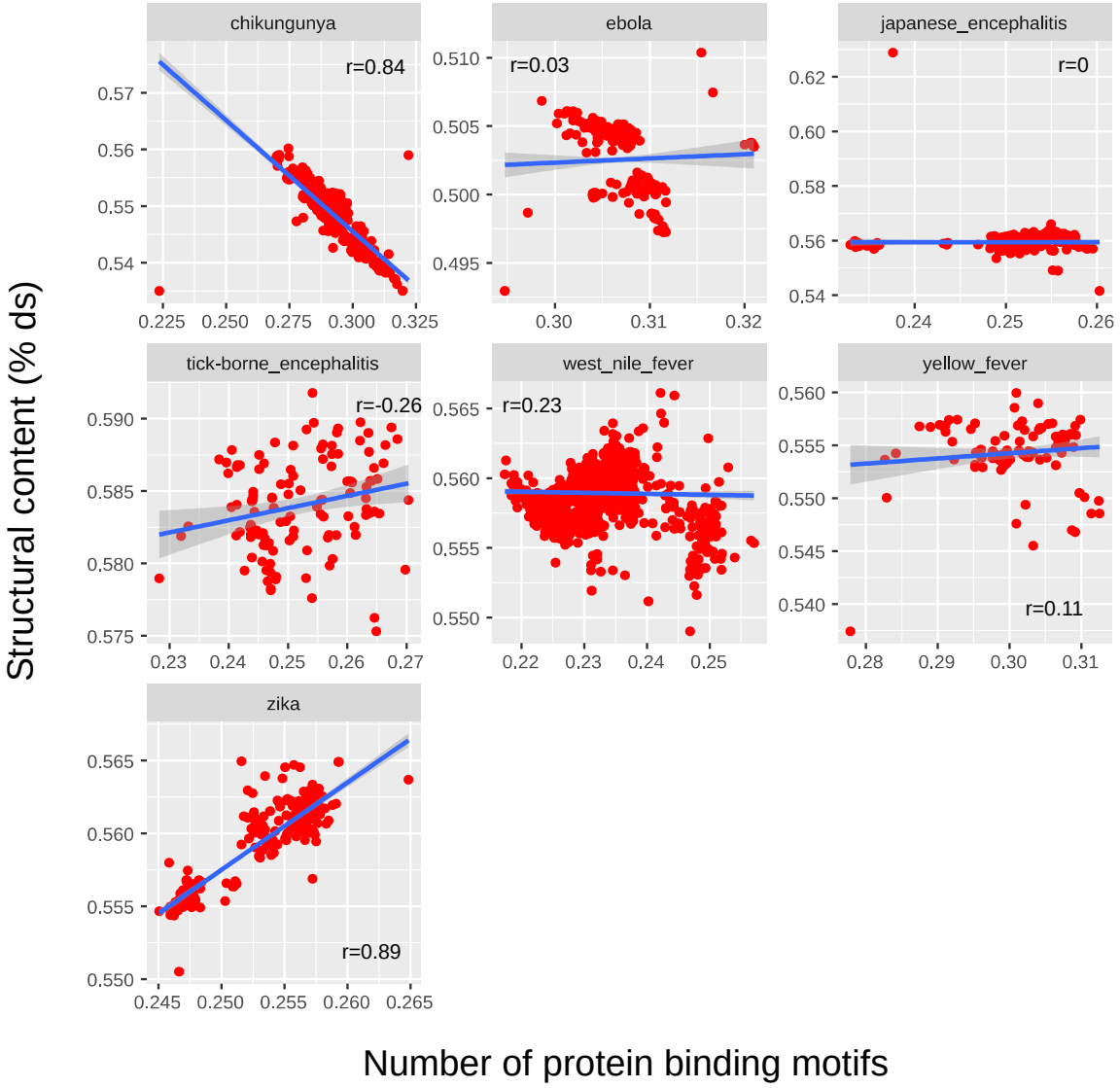


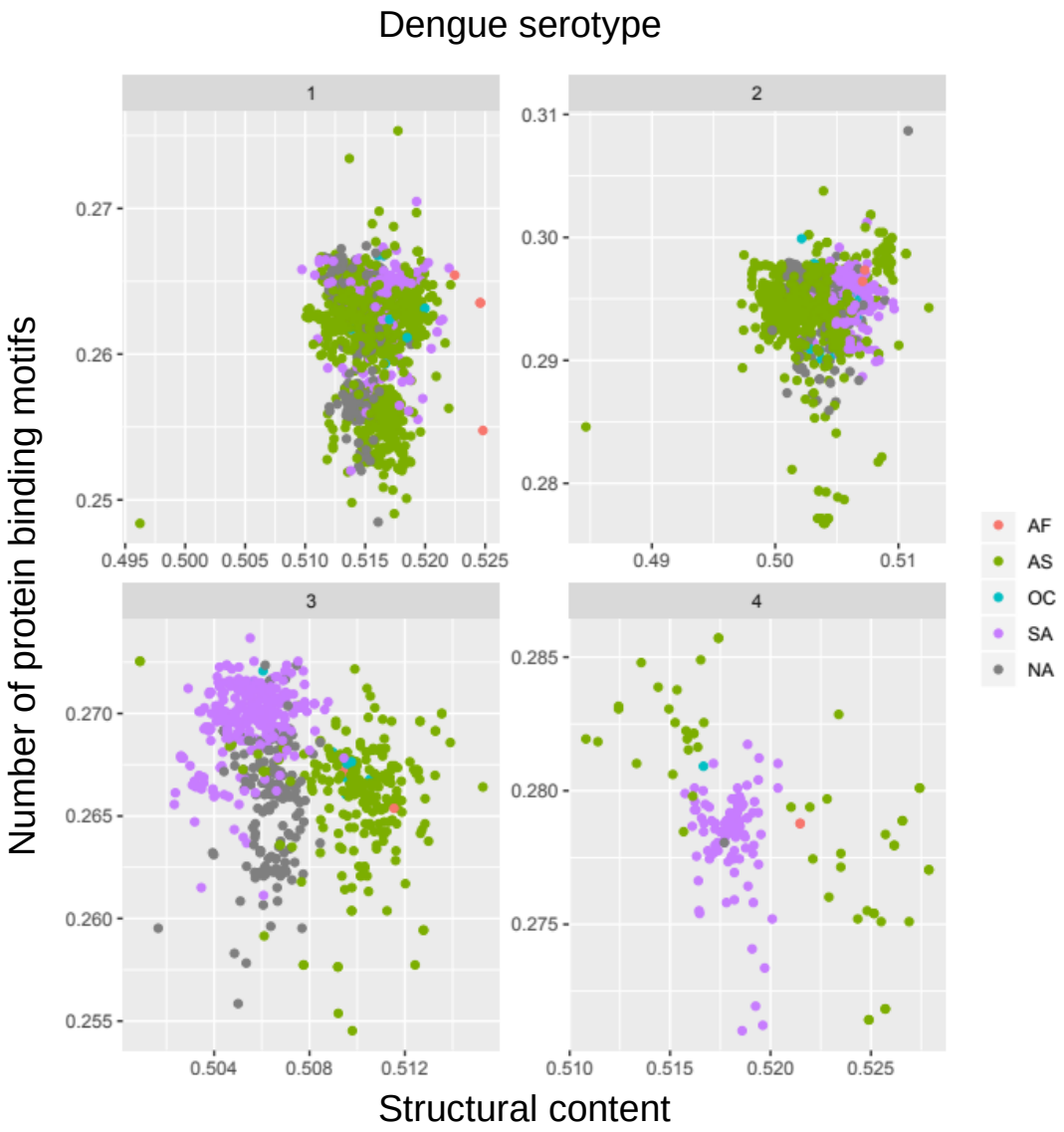

Structural content (% ds)
0.64
0.60
0.56
0.52
0.48
585
1653
1322
775
192
549
288
1660
135
188
274
Chikungunya
DENV-1
DENV-2
DENV-3
DENV-4
Ebola
Japanese encephalitis
West Nile fever
Tick-borne encephalitis
Yellow fever
Zika


A
Structural content (% ds)
0.0
0.2
0.4
0.6
chiku
denv1
denv2
denv3
denv4
ebola
jap
nile
tick
yellow
zika
B
Structural content (% ds)
0.0
0.2
0.4
0.6
chiku
denv1
denv2
denv3
denv4
ebola
jap
nile
tick
yellow
zika
C
Structural content (5'-3')
0.0
0.1
0.2
More structured 5' UTR
More structured 3' UTR
Chikungunya
Ebola
DENV-1
DENV-3
West Nile fever
Tick-borne encephalitis
DENV-2
DENV-4
Yellow fever
Japanese encephalitis
Zika


Protein binding motifs
0.23
0.31
Structural content
0.50
0.58
Zika
0.25
Tick
0.25
Jap
0.25
Nile
0.23
Yellow
0.30
Ebola
0.31
Denv-2
0.29
Chiku
0.29
Denv-4
0.28
Denv-3
0.27
Denv-1
0.26
Tick
0.58
Yellow
0.55
Zika
0.56
Nile
0.56
Jap
0.56
Chiku
0.55
Ebola
0.50
Denv-2
0.50
Denv-3
0.51
Denv-4
0.52
Denv-1
0.52
0.07
0.05
0.03
0.01
0.00
0.02
0.04
0.06
0.08
0.10


A
r=-0.74
Structural content (% ds)
Number of protein binding motifs
B
denv1
r=-0.09
denv2
r=-0.06
denv3
r=-0.31
denv4
r=-0.36
Structural content (% ds)
Number of protein binding motifs


chikungunya
r=0.84
ebola
r=0.03
japanese_encephalitis
r=0
tick-borne_encephalitis
r=-0.26
west_nile_fever
r=0.23
yellow_fever
r=0.11
zika
r=0.89
Structural content (% ds)
Number of protein binding motifs


Dengue serotype
1
2
3
4
Number of protein binding motifs
Structural content
0.27
0.26
0.25
0.495 0.500 0.505 0.510 0.515 0.520 0.525
0.31
0.30
0.29
0.28
0.49 0.50 0.51
0.270
0.265
0.260
0.255
0.504 0.508 0.512
0.285
0.280
0.275
0.510 0.515 0.520 0.525
AF
AS
OC
SA
NA

Genome	Family	Hemorrhagic	Number of genomes	Avg Size
Denv-1	Flavivirus	Yes	1653	10500
Denv-2	Flavivirus	Yes	1322	9750
Denv-3	Flavivirus	Yes	775	10590
Denv-4	Flavivirus	Yes	192	9730
Zika	Flavivirus	No	274	10120
Chikungunya	Togavirus	Yes* (mild)	585	10500
Japanese encephalitis	Flavivirus	Yes* (rare cases)	288	10500
Yellow fever	Flavivirus	Yes	158	8530
West Nile fever	Flavivirus	Yes	1660	10390
Tick encephalitis	Flavivirus	Yes	135	9770
Ebola	Filovirus	Yes	549	18200