Highly Conserved Core Residues Define Old-World Alphaviruses and Trace Early Evolutionary Divergence

Wendy Carolina Piña-Ruiz,[1,2,#], Luis Rubén Jaime-Rocha[1,2,#], Andrea Castorena-Robles[1,3], Roger Castells-Graells[4], Roya Zandi [5], Juan Carlos Muñoz-Escalante[6,7*], Mauricio Comas-Garcia[1,2,9*].

[1]High-Resolution Section, Research Center for Health Sciences and Biomedicine, Autonomous University of San Luis Potosi, Mexico

[2]Department of Sciences, Autonomous University of San Luis Potosi, Mexico

[3]Life Sciences Graduate Program, Department of Sciences, Autonomous University of San Luis Potosi, Mexico

[4]Structural Biology Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

[5]Department of Physics, University of California, Riverside

[6]Bioinformatic Unit, Research Center for Health Sciences and Biomedicine, Autonomous University of San Luis Potosi, Mexico

[7]Infectious Diseases Unit, Research Center for Health Sciences and Biomedicine, Autonomous University of San Luis Potosi, Mexico

[8]Translational and Molecular Medicine Section, Research Center for Health Sciences and Biomedicine, Autonomous University of San Luis Potosi, Mexico

[#]These authors contributed equally

[*]To whom send any correspondence: carlos.escalante@uaslp.mx and mauricio.comas@uaslp.mx

Abstract

Alphaviruses are positive-sense, single-stranded RNA viruses that assemble into striking double-icosahedral particles. During budding, their nucleocapsid core forms in the cytoplasm and adopts a T=4 icosahedral symmetry, a hallmark of Alphaviruses. Here, we combine structural and phylogenetic analyses to identify the amino acids most likely to govern capsomer formation (pentamers and hexamers) and core organization. We find that these residues are highly conserved in present-day Old-World alphaviruses but are divergent in New-World lineages. This suggests that the common ancestor of both groups likely assembled cores using the same molecular interactions seen in present-day Old-World viruses. We propose that early divergence in these interactions weakened core assembly efficiency, potentially contributing to the attenuation observed

in encephalitic New-World alphaviruses. This attenuation may reflect an adaptive trade-off, in which reduced assembly efficiency lowers viral replication and virulence, supporting long-term persistence in enzootic cycles. Revealing how specific residues control capsid architecture and tracing their evolutionary history, this study provides key insights into alphavirus assembly mechanisms, opening new avenues for antiviral strategies and rational vaccine design.

Introduction

Alphaviruses belong to the *Togaviridae* family, have a positive-sense single-stranded RNA genome of around 11-12 kilobases, and two open reading frames (ORFs) that code for two polyproteins (1). The ORF1 and ORF2 code the non-structural and structural proteins, respectively. Alphaviruses infect vertebrates and are transmitted by invertebrate vectors (2). These viruses have a double-layered virion consisting of two T = 4 icosahedral shells separated by a lipid bilayer; the outer shell is composed of 80 trimeric spikes, each composed of heterodimers of the glycoproteins E1 and E2, and the inner shell or core is made of 240 copies of the capsid protein (3). These spikes are not all placed at the same position in the icosahedron; 20 and 60 spikes at the 3- and quasi-3-fold symmetry axes, respectively. The core contains the full-length viral genome and assembles into an amorphous shell in the cytoplasm (4-7). At the plasma membrane, the core interacts with the heterodimers E1/E2, resulting in the assembly and budding of the viral particle (4). Without the core, E1 and E2 bud as amorphous particles (8).

The Alphavirus virion has a unique architecture; the viral particle consists of two concentric T = 4 icosahedral shells separated by a lipid bilayer. Both shells interact through hydrophobic amino acids in the cytoplasmic domain of E2 and the capsid protein (CP) (3). These virions have structural features unique for an icosahedral viral particle(9-15): i) there are large “holes” at the center of the pentamer and hexamers (capsomers) of the core, ii) the capsomers do not establish large contact surface areas between them, and iii) the torsion angles and distances between capsomers can significantly differ between viral species.

Alphaviruses can be grouped by antigenic complexes (*e.g.,* Semliki Forest and Western equine encephalitis complexes) (16), by the location where they are endemic (*i.e.,* Old- and New-World Alphaviruses) (17), or by some characteristic of the host (*e.g.,* if they infect aquatic animals or have no known vertebrate hosts) (18). Most Alphaviruses are classified as either Old- or New-World; the latter is restricted only to the American continent, and the former to the rest of the world. On the one hand, Old-World Alphaviruses mostly cause a febrile and arthrogenic dengue-like illness with a low mortality rate (19). On the other hand, New-World Alphaviruses affect the central nervous system, resulting in an encephalitic disease that, in some cases, has high mortality rates (up to 90% depending on the virus and the age of the patient) (20). Forrester and coworkers proposed that Alphaviruses have a marine origin and jump to terrestrial hosts, with birds being key to their global spread (18). Interestingly, some Alphaviruses are endemic to the New World (*i.e.,* Mayaro and Una viruses) but are phylogenetically part of the Old-World Alphaviruses. The opposite is true for the Sindbis virus, which is endemic to the Old World but is genetically related to New-World Alphaviruses (17, 18).

We first focused on identifying some structural differences in the cores of Alphaviruses using known structures to determine putative interactions that regulate CP-CP and capsomer-capsomer interactions. Then, we predicted the structure of the CP for all Alphaviruses and performed bioinformatic analysis of the consensus and representative amino acid sequences of the CPs. We found at least eight highly conserved amino acids in Old-World Alphaviruses, which are likely to mediate core assembly through electrostatic interactions (salt bridges). These amino acids are not conserved in other Alphaviruses. Finally, through a series of evolutionary analyses, we found that the core of the common ancestor most likely had almost the same electrostatic interactions as in Old-World Alphaviruses. These results suggest that a strong selective pressure resulted in the divergence of Old- and New-World Alphaviruses and altered the interactions of the core's CP.

Materials and Methods

Viral sequences for the prediction of the capsid protein structure.

The CP sequences for 42 *Alphavirus* species were obtained from the NCBI Taxonomy browser (Viruses; *Riboviria, Orthonovirae; Kitrinoviricota; Alsuviricetes; Martellivirales; Togaviridae*). For the E2 protein sequences, we used only 29 species because these were the only ones with sufficient data to generate a representative sequence. The consensus sequences were generated using the MUSCLE algorithm (21).

Structural data and structure modeling.

The cryo-electron microscopy (cryo-EM) structures Sindbis (SINV) (9), Getah (GETV) (10), Chikungunya (CHIKV) (11), Mayaro (MAYV) (12), Western equine encephalitis (WEEV) (13), Venezuelan equine encephalitis (VEEV) (14), Eastern equine encephalitis (EEEV) (15), Barmah forest (BFV) (22), and Semliki forest (SFV) (23) viruses were obtained from the RCSB PDB website (with PDB IDs 3J0F, 7FD2, 3J2W, 7KO8, 8DEC, 7SFV, 6XO4, 2YEW, and 1DYl, respectively). The experimental resolution of these structures is 7.0, 2.8, 5.0, 4.4, 4.7, 4.0, 4.2, 5.0, and 9.0 Å, respectively. Models for the structures of the capsid protein and E2 were generated using AlphaFold2 (24); some sequences were randomly selected to be predicted using AlphaFold3 (25), and we obtained the same results in all cases as with AlphaFold2. The structures were analyzed and visualized using ChimeraX (26). All CP and E2 structures were compared to CHIKV (PDB ID: 3J2W) (11); in cases where multiple structures were obtained (*i.e.,* when using AlphaFold3), we reported the structure with the lowest RMSD when overlapped to CHIKV PDB ID 3J2W. All distances between side chains were calculated by measuring the centroid-to-centroid distances by generating spheres with 2 Å radius.

Alphavirus sequences datasets curation for phylogenetic analysis.

The complete nucleotide sequences corresponding to the 32 *Alphavirus* species were downloaded from GenBank using their taxonomy IDs. We used a different number of

viruses for the structural prediction because the number of available sequences for some was not ideal for phylogenetic analyses. Sequences for each species were individually aligned using MAFFT v7.450 (27). The respective annotated RefSeq sequence for each species was used to identify the core coding sequence (CDS) region and to trim the multiple sequence alignments accordingly using BioEdit v7.2.5. To improve the quality of subsequent analyses, sequences were removed if they lacked more than 5% of the total length of the core CDS, contained more than 5% ambiguous nucleotides, or exhibited atypical indels not present in other sequences from the same species.

Consensus and representative sequences for each *Alphavirus* species were generated using CD-HIT-EST at a 90% similarity threshold. These representative core CDS sequences for each species were included in a single dataset and aligned using MEGA v10.2.6 with the MUSCLE algorithm. Amino acid sequences were inferred from the aligned nucleotide sequences, and manual curation of the MSA was performed using BioEdit v7.2.5.

Phylogenetic analysis and reconstruction of ancestral sequences

Phylogenetic inference was conducted in MEGA v10.2.6 using the best-fitting model, and tree robustness was assessed with 1,000 bootstrap iterations. The most probable ancestral sequences for each internal node of the resulting phylogenetic tree were reconstructed using the ExtAncSeqMEGA extension.

Selective pressure, Entropy, and Recombination analysis

To assess potential recombination events within *Alphavirus* CP coding sequences, alignments were analyzed using RDP v5.66, using RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan, and 3Seq as assessment methods. Potential events supported by at least three different methods with statistically significant *p*-values (<0.05) were considered true recombination events. Selective pressure acting on codon

sites was evaluated on either all *Alphavirus* or only Old-World Alphaviruses nucleotide MSA using the Datamonkey Adaptive Evolution server employing SLAC, FEL, FUBAR and MEME methods; sites were classified as being under pervasive or episodic positive selection if detected by MEME ($p<0.1$), FUBAR (posterior probability >0.9), FEL ($p<0.1$), or SLAC ($p<0.1$), respectively.

Site-wise variability (either all *Alphavirus* or only Old-World Alphaviruses) was further examined using Shannon entropy calculation performed in BioEdit v7.2.5 on the MSA. Entropy scores were used to identify highly variable positions across the core protein sequences, supporting the detection of sites associated with adaptive or host-specific changes.

Results

The interactions between capsid proteins in Old-World viruses are highly conserved.

First, we analyzed the cryo-electron microscopy (cryo-EM) structures of the core of Sindbis (SINV) (9), Getah (GETV) (10), Chikungunya (CHIKV) (11), Mayaro (MAYV) (12), Western equine encephalitis (WEEV) (13), Venezuelan equine encephalitis (VEEV) (14), Eastern equine encephalitis (EEEV) (15), Barmah forest (BFV) (22), and Semliki forest (SFV) (23) viruses (see Figure 1). In Figure 1, the different colors in a structure represent the four proteins constituting the core asymmetric unit. These cores have a T = 4 symmetry; they are composed of 240 capsid proteins arranged in 12 pentamers and 30 hexamers. Despite having conserved overall architecture, close inspection of the hexamers, pentamers, and the three- and quasi-threefold symmetry axes shows interesting species-specific features.

Figure 2A shows no noticeable differences in the pentamers of these three Old-World Alphaviruses (*i.e.,* CHIKV, MAYV, and GETV). In contrast, Figure 2B shows the core of three New-World Alphaviruses; in EEEV, the capsid proteins (CP) in the pentamer establish fewer inter-subunit contacts than in VEEV and WEEV (see the asterisks). The structures of the VEEV and WEEV pentamers exhibit no noticeable

differences compared to those of CHIKV, MAYV, and GETV. However, the quasi-threefold symmetry axis (q3) shows some changes (supplementary figure 1); the space between two adjacent hexamers and one pentamer for CHIKV, MAYV, and GETV is approximately 23 Å, 19 Å, and 25 Å wide, respectively. Interestingly, while the general features of the VEEV pentamer are like those of the Old-World Alphaviruses, the space at the q3 is much smaller (≈ 13 Å) than for the other five viruses in this figure (≥ 15 Å) (see supplementary figure 1). The three-fold symmetry axis (*i.e.,* between three hexamers) shows a pattern (data not shown).

These results suggest that the physicochemical properties of the amino acids, which should control capsomer assembly and capsomer-capsomer interactions, may be similar for these three Old-World viruses and vary in the New-World Alphaviruses. To explore this idea, we focused on the CP contact surfaces using the CHIKV core cryo-EM structure as a reference (PDB ID 3J2W). Figure 3A suggests electrostatic interactions (salt bridges) between CHIKV CP K172 and E234 (outer), and K177 and E184 (inner). It should be noted that the resolution of these structures does not permit the precise localization of the side chain positions. Nonetheless, the distances of these amino acids and their chemical environment suggest electrostatic interactions between the CPs in the pentamers. These putative interactions are also present in the hexamers (data not shown); however, because of the asymmetry of the hexamer, the distances between K177 and E184 vary between adjacent CPs and are greater than in the pentamers (data not shown). Figure 3B shows the i3 axis of CHIKV, suggesting that the capsomer-capsomer interactions depend on two sets of electrostatic interactions: E120/K151 and K122/D148. The position of these amino acids resembles a dipole-dipole interaction; thus, the strength of these interactions should depend on the distances and angles between capsomers. To further understand the structures shown in Figure 2, we overlapped the structure of the monomeric CHIKV CP (Figure 3C) with those of Alphaviruses, MAYV and EEEV, with a known cryo-EM structure. Figures 3D and 3E show the overlap of CHIKV CP with MAYV and EEEV CPs, respectively. On the one hand, Figure 3D shows that MAYV CP has a perfect conservation of charge and identity at the position of the eight amino acids of interest. The only difference is that CHIKV CP K151 has an arginine in MAYV (R148). On the other hand, Figure 3E shows

key differences in the EEEV CP; the amino acids that should be in the same position as in CHIKV CP K122, K151, D148, K177, and E184 are instead Q, Q, A, L, and P, respectively. The overlap of the structures of GETV, SFV, and BFV with CHIKV shows that the amino acids in these eight positions are 100% conserved in charge, while the degree of conservation for SINV, WEEV, and VEEV is much lower (data not shown).

Figure 4 shows a closer look at the interfaces in a pentamer; Figure 4A (top row) contains only Old-World Alphaviruses, which exhibit the same salt bridges as CHIKV. Figure 4B shows the equivalent interface for three New-World Alphaviruses; the amino acids in positions equivalent to CHIKV CP K177 and E184 (inner interactions) are very different between EEEV, VEEV, and WEEV. Nonetheless, the charge of the amino acids equivalent to CHIKV CP K172 and E234 (outer interactions) is the same in all cases. The “outer interactions” in the pentamer seem to be conserved in Old- and New-World Alphaviruses, while the “inner interactions” are only conserved in Old-World Alphaviruses.

To better understand the interactions between capsomers (Figure 3B), we analyzed the three-fold symmetry (i3) axis (see Figure 5). The leftmost panel in Figure 5A shows the centroid-to-centroid distance (2-Å centroid) and torsion angle between three CHIKV CP K122 amino acids. The letters in parentheses indicate the chains being compared in the Cryo-EM structure. The distances between equivalent amino acids are asymmetric, albeit the exact positions of the side chains are unknown. This analysis was also conducted for MAYV, GETV, EEEV, VEEV, and WEEV. The interactions in CHIKV, MAYV, and GETV (Figure 5A) are the same in these viruses, and the distances and torsion angles between equivalent amino acids do not differ between these viruses. Figure 5B shows that at the i3 the interactions in EEEV and WEEV are between polar uncharged amino acids (Q and N) instead of charged amino acids.

Additionally, the middle panel in Figure 5B shows that in VEEV, a cluster of charged amino acids (K135 and E135 from three capsid proteins) does not appear to have a corresponding pair of amino acids that could establish a dipole-dipole-like interaction, as observed in CHIKV, MAYV, and GETV (see Figure 5A). The distances between Q121 and Q119 in EEEV and WEEV are similar to the basic amino acids in the

i3 of CHIKV, MAYV, and GETV. The only significant difference in WEEV is that the torsion angle of the amino acid is greater than that of Old-World Alphaviruses. The interactions in VEEV are very different from those of the other viruses in this Figure; for example, the charged amino acids are about half the distance as in CHIKV (≈ 12.8-15.8 Å vs ≈ 27.2-30.6 Å, respectively). There is no pair of amino acids with an opposite charge that could help to establish a dipole-dipole-like interaction. The torsion angles between these amino acids are indicated in the center of the i3 site; these angles are very similar for these three Old-World Alphaviruses and EEEV (between 51-57 °). However, VEEV and WEEV have torsion angles of 64 °, suggesting that while EEEV and WEEV have similar amino acids at this site, the way they interact differs between these two viruses.

The electrostatic interactions between capsid proteins are a unique feature of Old-World viruses.

The results in the previous section were limited to Alphaviruses with known Cryo-EM structures; however, we wanted to identify these interactions for all Alphaviruses with known CP sequences. First, we generated a consensus sequence for the CP of all Alphaviruses (including those in the previous section) and predicted the structure of the monomeric capsid protein using Alphafold2 (24). We also used Alphafold3 (25) for a few randomly selected structures, and the results were consistent with those obtained with Alphafold2. We modeled the capsid proteins as monomers because Alphafold2 and 3 failed to assemble these proteins into pentamers or hexamers. For this analysis, we used consensus sequences because, for some viruses, there is not enough data to generate a representative sequence. Then, we overlapped the predicted structures with the CHIKV CP (PDB 3J2W) as was done in Figures 3D and 3E to identify the amino acids that are in equivalent positions to those of CHIKV CP, specifically E120, K122, D148, K151, K172, K177, E184, and E234 (see supplementary files). From here on, we will use this numbering to indicate the amino acids of reference, and CHIKV will be the virus to which we will compare all interactions. The results of these analyses are

presented in Tables 1, 2, 3, and 4. The numbers in green indicate that the percentage of amino acid identity and/or charge conservation is greater than or equal to 80%.

Additionally, we performed a multiple-sequence alignment (MUSCLE) of the consensus sequences (see Supplementary Figures 2 and 3). The comparison between the two methods revealed a few differences. The lavender, turquoise, and maroon amino acids in Tables 2, 3, and 4 indicate that the positions of the amino acids predicted through Alphafold2 and MUSCLE differ. The lavender color indicates that the amino acid that overlaps with the reference structures comes from an insertion. The amino acids in turquoise overlap at a -1 position in the aligned sequence, and the maroon color refers to amino acids that overlap with the reference CHIKV amino acids because of a deletion.

Table 1 shows almost 100% charge conservation of these amino acids in all Old-World Alphaviruses; the Ndumu and Middelburg viruses have a charge inversion in K122 and a Q in K177, respectively. Also, except for the positions equivalent to E120 and K172, the identities of the other six amino acids are conserved between 80% and 100%. These results suggest that for Old-World Alphaviruses, there is strong evolutionary pressure on the charge of amino acids that might control the intra- and inter-capsomer interactions and the length of the side chains of these amino acids. These findings are consistent with the results presented in the previous section.

In New-World Alphaviruses, except for the amino acids equivalent to K172/E224 (*i.e.,* the “outer interactions”), the charge is not conserved in the rest of the amino acids of interest (Table 2). However, there seems to be some degree of conservation within antigenic complexes. In most cases, the charged amino acids are replaced by polar uncharged amino acids, except in the positions equivalent to D148 and K151, where there is a significant amount of hydrophobic side chains. Interestingly, E120 and K122 are conserved in the VEEV antigenic complex. In contrast, in the WEEV complex, E120 can be a polar uncharged, or acidic side chain, and K122 is either an acidic or a polar uncharged amino acid. In WEEV and VEEV, D148 is mostly substituted for Q or V. In the WEEV complex, K177 and E184 are Q and P, respectively, while in the VEEV complex, the charge of K177 is conserved, and E184 is almost always a Q. The

identities of five of the eight amino acids of interest within the VEEV complex are more conserved than within viruses of the WEEV complex. These results suggest a common ancestor for the CP of Aura virus, SINV, and Whataroa virus and a distinct ancestor for the other viruses within this antigenic complex. Furthermore, based on the results of Table 2, it is likely that the CP of the Madariaga virus may have had a common ancestor with the cores of WEEV, Fort Morgan Virus, and Highland J virus.

The lack of conservation of intra- and inter-capsomeric interactions is more evident for Alphaviruses found in mosquitoes with no known vertebrate host that are not included in the Old-/New-World classification and on aquatic Alphaviruses (see Tables 3 and 4). Table 3 shows that, on the one hand, like in New-World Alphaviruses, a hydrophobic side chain replaces D148. On the other hand, K151 is replaced by a polar uncharged amino acid, which is unique for this group. Interestingly, the charge conservation of the viruses in Table 3 is closer to that of Old-World than New-World Alphaviruses.

The degree of charge conservation of the amino acids of interest in aquatic Alphaviruses is low compared to CHIKV. However, in aquatic Alphaviruses that infect mammals, six of the eight amino acids of interest are conserved in these viruses (see shaded rows in Table 4). While the number of viruses in this category is too small to make a conjecture, it could suggest that, at least based on the core and/or CPs, these two viruses might be related to Old-World Alphaviruses. In aquatic Alphaviruses that infect fish, there are two positions with charge inversion, K151 and K177. This charge inversion should result in intra- and inter-capsomeric interactions that differ from the rest of the Alphaviruses.

The data in Tables 1-4 show a wide range of interactions for non-Old-World Alphaviruses. Nonetheless, it is clear that the “outer intracapsomeric interactions” (E234/K172) are highly conserved in almost all Alphaviruses; only in the Trocara virus and the Wenling fish alphavirus, the amino acids in the equivalent positions to CHIKV CP E234 and K172 are a Q and an H, respectively. These results suggest that there is a strong evolutionary pressure on the salt bridge between CHIKV CP E234 and K172,

and that this interaction has been conserved during evolution, possibly playing a crucial role in core assembly.

The electrostatic interactions between capsid proteins at the core might be related to glycosylation patterns in E2.

The results presented so far suggest that the organization and type of interactions in the core of Alphaviruses could have changed when Old- and New-World Alphaviruses diverged. But these are not the only changes. The glycosylation profiles of E2 and E1 for Old-World Alphaviruses differ from New-World ones (28). For example, CHIKV E2 has only one glycosylation site in N623. A bioinformatic and structural analysis identical to the one described in the previous section, aiming at finding an N-X-S/T motif at the same position as CHIKV E2 N263, shows that this glycosylation site is conserved in almost all Old-World Alphaviruses (see Figure 6A). Only the Bebaru virus appears to have an N-to-G mutation at this site. We could not find an N-X-T/S motif in or around the E2 region of the Bebaru virus that is structurally equivalent to CHIKV. Figure 6B shows the Cryo-EM density of the E2 glycosylation site for MAYV, which has been proposed to act as a “molecular handshake” between trimers that stabilize the viral particle (12). It should be noted that the analyses E2 and CP do not have the same number of species, as there are fewer sequences for E2 than for CP. Therefore, the number of viruses in Tables 1-4 and Figures 6A, 6B, and 6C differs.

Figure 6C shows no N-glycosylation motif on or around CHIKV E2 N263 in New-World Alphaviruses. This figure also suggests that the Fort Morgan virus may have an N-glycosylation site (N260) in a position close to CHIKV E2 (N263). However, E2 N260 in the Fort Morgan virus is not equivalent to CHIKV E2 N263. In the Old-World Alphaviruses, the glycosylation site equivalent to CHIKV N263 is in a β-sheet, while the Fort Morgan N260 (N-V-T motif) is in a loop before this β-sheet (Figure 6D). The position of this possible glycosylation site makes it unlikely to result in the same trimer-trimer interactions as the “molecular handshake” proposed for MAYV (12).

Finally, the data in Figure 6E suggest that Harbor porpoise alphavirus, which infects marine mammals, has the same glycosylation site as CHIKV. Interestingly, this marine Alphavirus has seven of the eight amino acids studied in the previous sections conserved in charge compared to CHIKV. Therefore, there may be a relationship between the identity of the amino acids that control assembly and the position of the glycosylation sites in E2. It is possible that in non-Old-World Alphavirus cores, the interactions between CPs are weak, and glycosylation in E2 N263 could introduce rigidity that inhibits the assembly of the icosahedral core during budding.

The common ancestor of Old- and New-World Alphaviruses may have had intra- and inter-capsomer interactions like those of CHIKV.

Phylogenetic inference analysis of Alphaviruses is usually done using E1, E2, and TF, or a non-structural protein (17, 18, 29). However, using E1 and E2 gives information about the evolution of viruses in the context of their interaction with their hosts. Therefore, using these proteins is extremely useful in understanding the evolution of a pathogen from a perspective that incorporates the contributions of both the host and vectors. Nonetheless, we wanted to understand the evolution of the core and determine what kind of core was first: was the core with strong electrostatic interactions (Old-World Alphaviruses) first, and then the core with weak interactions appeared (New-World Alphaviruses), or vice versa? First, we generated representative viral sequences with sufficient CP data and performed phylogenetic inference using MEGA v10.2.6 and ExtAncSeqMEGA (see Methods and Materials).

Figure 7 shows the result of the phylogenetic analysis using CP. The structural analysis was conducted using 42 viral species. In comparison, the phylogenetic analysis was performed only with 32 species due to the lack of sufficient genomes to generate representative sequences for this analysis. As expected from the results shown in Tables 1 and 2, there is a clear separation between Old- and New-World Alphaviruses; however, this analysis shows that the CP of Sindbis, Aura, and Whataroa viruses diverged separately from the rest of the New-World clades. This separation is similar but not identical to those reported using E1, E2, and TF (17, 18, 29). Tables 4 and Figure

6D suggest that Harbor porpoise alphavirus did not jump directly from fish to a porpoise, but it appeared when the Old-World group separated from the Aura virus clade (blue clade in Figure 7). As shown by Weaver and Forrester (17, 18), Middelburg, Bebaru, Semliki, Getah, and Ross River viruses belong to a distinct clade, separate from CHIKV, O'nyong nyong, Una, and Barmah Forest viruses. However, our results have some differences. By only looking at the CP, Una and Mayaro viruses are not part of the same clade as proposed using E1, E2, and TF (17).

Analysis of the CP results reveals that within the WEE antigenic complex, Aura, SINV, and Whataroa viruses belong to a different clade than the Highlands J and WEE viruses. Interestingly, this analysis suggests that the evolution of the CP of Fort Morgan virus differs significantly from that reported when using other structural proteins. New-World Alphaviruses appear to have given rise to three main clades of the viral core. One clade is for Aura, SINV, and Whataroa viruses, which seem to be a closer ancestor to Old-World Alphaviruses than the rest of the New-World Alphaviruses. This is consistent with SINV and Whataroa being found in the Old World. The second clade includes Fort Morgan, Madariaga, Highlands J, EEEV, and WEE viruses, while the third clade encompasses the rest.

Additionally, we performed a selective pressure and entropic analysis on the Old- and New-World Alphavirus CP. When these sequences were analyzed, we did not find a strong selective pressure to maintain the identity of CP E120, K122, D148, K151, K172, K177, E184, and E234 (supplementary files). However, when this analysis was conducted with only Old-World Alphavirus CP, 31 sites within the CP MSA were under purifying selection, as detected by FEL or FUBAR analyses (supplementary files). Interestingly, 16 amino acids also exhibited low entropy values (<0.9), indicating high evolutionary conservation. These conserved residues are likely to be functionally or structurally important, potentially contributing to the proper functioning of these proteins. The overlap between strong negative selective pressure and low sequence variability supports the idea that these sites are subject to evolutionary constraints due to their essential roles in the viral cycle.

Notably, a general trend of low entropy was observed starting at position 130 of the 295-amino-acid alignment (of the Old-World Alphavirus consensus sequence), suggesting a region of sustained conservation across multiple species, potentially associated with a structurally constrained domain (the first 100 to 130 amino acids are disordered and interact with the genomic RNA). Furthermore, most of the sites subjected to purifying selection were located between positions 96-145 and 187-206 (supplementary files), highlighting two discrete regions of potential functional importance that may be critical for the structural integrity of the viral capsid.

A closer look at CHIKV CP in positions E120, K122, D148, K151, K172, K177, E184, and E234 showed that the Shannon entropy of these amino acids is 0.7, 0.3, 0.3, 0.3, 0.3, 0, and 0.3, respectively (see Table 5). These results are consistent with the hypothesis that these eight amino acids are key in Old-World Alphavirus CP. The consequences of these observations will be discussed later, but they may be related to the pathogenesis and type of disease they cause.

Then we used the ExtAncSeqMEGA to infer a few most probable ancestor sequences (see the three asterisks and letters A-G in the insert in Figure 7). This analysis strongly suggests that the common ancestor of Alphaviruses that infect mammals, but not fish, had seven of the eight amino acids we propose to mediate CP-CP and capsomer-capsomer interactions as Old-World viruses. This data implies that after the appearance of the Old-World Alphavirus ancestor in node C, the selective pressure in the eight amino acids of interest became so strong that it made them invariant. The only difference between this common ancestor's amino acids of interest and Old-World Alphaviruses is that the equivalent amino acid to CHIKV D148 is V158.

The event that led to the divergence of these two main groups resulted in a significant diversification of the New-World Alphavirus CP. However, on the one hand, the common ancestor of the Sindbis clade has the same eight amino acids as the common ancestor for all Alphaviruses that do not infect fish (*** in Figure 7). On the other hand, the common ancestor for the VEE and EEE complexes already had three mutations in the CP: one and two mutations in the inter- and intracapsomeric

interactions, respectively. The most divergent node is the one that gave rise to the EEE complex.

Discussions

The results presented here strongly suggest that the CP-CP interactions that mediate core assembly of New- and Old-World Alphaviruses are distinct. On the one hand, based on the analysis of Cryo-EM structures and the predicted structure of the CPs, it is likely that the core of CHIKV and all other Old-World Alphaviruses is controlled through highly conserved electrostatic interactions. Most likely, salt bridges control the assembly of capsomers (*i.e.,* pentamers and hexamers), and dipole-dipole-like interactions at the i3 and q3 axes are responsible for capsomer-capsomer interactions. On the other hand, the amino acids that control the assembly of capsomers and capsomer-capsomer interactions in New-World and other Alphaviruses are not conserved and are weaker than in the Old-World cores. Based on the proposed amino acids, it is likely expected that the assembly of the core of Old-World Alphaviruses is mediated by highly conserved and strong electrostatic interactions unique to this group of viruses.

Kaelber *et al*. showed that CHIKV and VEEV have T = 4 virions, with a minority population of T = 3 and abnormal particles (30). They also showed that the percentage of T = 3 and abnormal virions is larger in VEEV and EEEV than in CHIKV. While there is no hypothesis to explain this observation, our results are consistent with a scenario where the assembly of EEEV and VEEV virions results in a smaller population of (correct) T = 4 virions than CHIKV, because the intra- and intercapsomere interactions of EEEV and VEEV CP are weaker than those of CHIKV and hence are more prone to make mistakes during assembly. Unpublished results from our group show that point mutations in the CHIKV CP at any of the four amino acids proposed here to control the intracapsomeric interactions increase the percentage of abnormal virions. While the biological implications of Kaelber and co-workers' observations are unclear, the fact that EEEV and VEEV have a higher rate of misassembled virions could be an attenuation mechanism. It is essential to note that EEEV and VEEV are lethal viruses, whereas CHIKV is rarely deadly (31). Therefore, encephalitic Alphaviruses may be more likely to

make more mistakes during assembly to achieve attenuation, thereby increasing the probability of infecting new hosts before the original host dies.

Alphaviruses generate defective interfering RNAs (diRNAs) during sequential passages in cultured cells, resulting in short genomes (32-34). It is believed that the generation of diRNAs serves as an attenuation mechanism (35). However, to the best of our knowledge, the generation of abnormal virions in this type of virus has not been explored as an attenuation mechanism; thus, our results open the door to a novel attenuation mechanism that is more prominent in encephalitic than in arthrogenic Alphaviruses.

Mukhopadhyay and coworkers cryo-EM reconstructed Ross River virus and WEEV *in vitro* assembled core-like particles (CLPs) and could not achieve a resolution higher than 30 Å (7). They also attempted to reconstruct cores produced in mammalian cells, in which the lipid bilayer and the E1/E2 layer were removed, obtaining similar results to those with the CLPs. Therefore, they proposed that their inability to achieve the same resolution as intact virions suggested that CLPs deviate from an icosahedron and/or are heterogeneous (7). Recently, Chmielewski and co-workers used cryo-ET to reconstruct CHIKV particles as they budded from cultured cells (4). In this work, they demonstrated that the cores in the cytoplasm are not icosahedral, and the double icosahedral symmetry of the virion is assembled during budding. Although Chmielewski and coworkers did not investigate the molecular assembly mechanism of CHIKV virions, their results can be correlated with the hypothesis presented here. The transition from a disordered to an icosahedral core could be thought as an annealing process in which the cytoplasmic core is not in its lowest free energy state and that it is upon interaction with the glycoproteins at the plasma membrane that multiple protein-protein interactions overcome the loss of entropy (because of the transition from disordered to icosahedral core). Weak interactions would facilitate this mechanism; for example, the proposed dipole-dipole-like interactions between capsomers depend on the distance between capsomers and the angle between them. An angle dependency would facilitate reordering a heterogeneous core into an icosahedral structure. Furthermore, our results suggest that the differences in the type of interactions between Old- and New-World

Alphaviruses result in a more efficient ordering process for the Old-World Alphaviruses, which could be correlated with the results from Kaelber et al. (30)

We investigated the evolution of the CP, and by only focusing on the CP, we could infer that the common ancestor of non-fish-infecting Alphaviruses had almost the same putative CP-CP and capsomer-capsomer interactions as Old-World Alphaviruses. On the one hand, New-World Alphaviruses are mostly restricted to the American continent, and yet there is a significant variability in these interactions. On the other hand, Old-World Alphaviruses can be found in the rest of the world, and these interactions are highly conserved.

Our analysis of the Shannon entropy and selective pressure clearly shows that the CPs of these two groups of Alphaviruses are very different; the amino acids proposed to control core assembly are invariant in Old-World viruses. Altogether, our results strongly suggest that the event that caused the divergence of Old- and New-World Alphaviruses likely resulted in changes of viral tropism, pathogenesis, and the type of diseases, increasing the lethality of New-World alphaviruses. This resulted in a rapid evolution of the CP on the American continent, leading to a possible viral attenuation mechanism through the formation of misassembled virions. This hypothesis is consistent with the fact that the glycosylation profiles of E2 and E1 differ between these two groups of Alphaviruses. Therefore, it is possible that glycosylation changes affected tropism and contributed to stabilizing cores with weak interactions. In other words, it is possible that the “molecular handshake” proposed by Ribeiro-Filho for MAYV (12) had to be removed to allow for assembly of the core when the interactions became weaker, and that this event might have happened when the divergence of Old- and New-World viruses occurred.

It has been proposed that Alphaviruses have a marine origin (18). However, a phylogenetic analysis of the CP alone suggests that marine alphaviruses infecting mammals appeared later in the evolution of these viruses. Our analysis also indicates that Harbor porpoise virus was reintroduced to marine animals after the divergence of Old- and New-World viruses. We are unclear when the Southern elephant seal virus (SEAV) was introduced. However, given that SEAV CP has six of the eight amino acids

of interest conserved with CHIKV, it suggests that it may have been reintroduced to a marine environment at the same time as Harbor porpoise virus.

The inference of the CP common ancestor shows that it might have had seven out of the eight interactions we are proposing to control CHIKV assembly. This indicates that the event that separated Old- and New-World Alphaviruses resulted in homogenous, and most likely more stable, Old-World cores, while this event favored an increase in heterogeneity of New-World Alphaviruses. The increase in variability in the intra- and intercapsomere interactions could likely be linked to the development of a lethal encephalitic disease.

Conclusions

The interactions that control core assembly in Old-World Alphaviruses differ significantly from those of other Alphaviruses. The amino acids studied here are likely to play a crucial role in the assembly of capsomers and the core in the cytoplasm, so that they favor the transition from an irregular to icosahedral core during assembly. In the case of Old-World Alphaviruses, these interactions are highly conserved and are likely to have emerged early in the evolution of these viruses. The divergence of Old- and New-World Alphaviruses resulted not only in different tropism and diseases but also in structural differences. These differences could have resulted in a novel attenuation mechanism in which lethal Alphaviruses are more prone to make mistakes during assembly. However, the results presented here require experimental confirmation, primarily due to the resolution of some of the Cryo-EM data and because modeling using AlphaFold2 was performed with the monomeric capsid protein; therefore, we may be missing some protein surfaces that appear once the core is assembled. It is essential to note that the structural analysis was conducted using consensus sequences, which could introduce folding errors; however, the phylogenetic analysis was performed with representative sequences. When the results from both analyses were compared, only 12 out of 293 amino acids (4.1%) showed discordance in position. This suggests that combining both approaches are likely to yield a highly accurate prediction of these interactions of interest. Our group is currently testing these scenarios to investigate the contribution of

these amino acids by performing site-directed mutagenesis in our CHIKV system (36) with a reporter gene to validate so we can further validate these hypothesis on core assembly, as well as the correlation between the E2 N263 glycosylation site characteristic of Old-World Alphaviruses and the identity of the amino acids that control core assembly.

Funding

M.C-G. thanks the Mexican Government for the SEIHCITI CBF-2023-2024-1145 grant, the San Luis Potosi State Government for the COPOCYT 2024-03-M07 grant, and the CICSaB/UASLP for intramural support. A.R-C. thanks the SEIHCITI for the Ph.D. fellowship 4032286. We also thank Jose Alberto Campillo Balderas for providing his helpful comments during the manuscript's writing.

References

1. Ou JH, Rice CM, Dalgarno L, Strauss EG, Strauss JH. 1982. Sequence studies of several alphavirus genomic RNAs in the region containing the start of the subgenomic RNA. Proc Natl Acad Sci U S A 79:5235-9.
2. Lambrechts L, Scott TW. 2009. Mode of transmission and the evolution of arbovirus virulence in mosquito vectors. Proc Biol Sci 276:1369-78.
3. Button JM, Qazi SA, Wang JC, Mukhopadhyay S. 2020. Revisiting an old friend: new findings in alphavirus structure and assembly. Curr Opin Virol 45:25-33.
4. Chmielewski D, Schmid MF, Simmons G, Jin J, Chiu W. 2022. Chikungunya virus assembly and budding visualized in situ using cryogenic electron tomography. Nat Microbiol 7:1270-1279.
5. Cheng F, Mukhopadhyay S. 2011. Generating enveloped virus-like particles with in vitro assembled cores. Virology 413:153-60.
6. Cheng F, Tsvetkova IB, Khuong YL, Moore AW, Arnold RJ, Goicochea NL, Dragnea B, Mukhopadhyay S. 2013. The packaging of different cargo into enveloped viral nanoparticles. Mol Pharm 10:51-8.
7. Mukhopadhyay S, Chipman PR, Hong EM, Kuhn RJ, Rossmann MG. 2002. In vitro-assembled alphavirus core-like particles maintain a structure similar to that of nucleocapsid cores in mature virus. J Virol 76:11128-32.
8. Zhang YN, Deng CL, Li JQ, Li N, Zhang QY, Ye HQ, Yuan ZM, Zhang B. 2019. Infectious Chikungunya Virus (CHIKV) with a Complete Capsid Deletion: a New Approach for a CHIKV Vaccine. J Virol 93.

9. Tang J, Jose J, Chipman P, Zhang W, Kuhn RJ, Baker TS. 2011. Molecular links between the E2 envelope glycoprotein and nucleocapsid core in Sindbis virus. J Mol Biol 414:442-59.
10. Wang A, Zhou F, Liu C, Gao D, Qi R, Yin Y, Liu S, Gao Y, Fu L, Xia Y, Xu Y, Wang C, Liu Z. 2022. Structure of infective Getah virus at 2.8 A resolution determined by cryo-electron microscopy. Cell Discov 8:12.
11. Sun S, Xiang Y, Akahata W, Holdaway H, Pal P, Zhang X, Diamond MS, Nabel GJ, Rossmann MG. 2013. Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. Elife 2:e00435.
12. Ribeiro-Filho HV, Coimbra LD, Cassago A, Rocha RPF, Guerra J, de Felicio R, Carnieli CM, Leme L, Padilha AC, Paes Leme AF, Trivella DBB, Portugal RV, Lopes-de-Oliveira PS, Marques RE. 2021. Cryo-EM structure of the mature and infective Mayaro virus at 4.4 A resolution reveals features of arthritogenic alphaviruses. Nat Commun 12:3038.
13. Sutton MS, Pletnev S, Callahan V, Ko S, Tsybovsky Y, Bylund T, Casner RG, Cerutti G, Gardner CL, Guirguis V, Verardi R, Zhang B, Ambrozak D, Beddall M, Lei H, Yang ES, Liu T, Henry AR, Rawi R, Schon A, Schramm CA, Shen CH, Shi W, Stephens T, Yang Y, Florez MB, Ledgerwood JE, Burke CW, Shapiro L, Fox JM, Kwong PD, Roederer M. 2023. Vaccine elicitation and structural basis for antibody protection against alphaviruses. Cell 186:2672-2689 e25.
14. Kafai NM, Williamson LE, Binshtein E, Sukupolvi-Petty S, Gardner CL, Liu J, Mackin S, Kim AS, Kose N, Carnahan RH, Jung A, Droit L, Reed DS, Handley SA, Klimstra WB, Crowe JE, Diamond MS. 2022. Neutralizing antibodies protect mice against Venezuelan equine encephalitis virus aerosol challenge. J Exp Med 219.
15. Williamson LE, Gilliland T, Jr., Yadav PK, Binshtein E, Bombardi R, Kose N, Nargi RS, Sutton RE, Durie CL, Armstrong E, Carnahan RH, Walker LM, Kim AS, Fox JM, Diamond MS, Ohi MD, Klimstra WB, Crowe JE, Jr. 2020. Human Antibodies Protect against Aerosolized Eastern Equine Encephalitis Virus Infection. Cell 183:1884-1900 e23.
16. Calisher CH, Karabatsos N. 1988. Arbovirus serogroups: definition and geographic

distribution. CRC Press,, Boca Raton, FL.
17. Weaver SC, Winegar R, Manger ID, Forrester NL. 2012. Alphaviruses: population genetics and determinants of emergence. Antiviral Res 94:242-57.
18. Forrester NL, Palacios G, Tesh RB, Savji N, Guzman H, Sherman M, Weaver SC, Lipkin WI. 2012. Genome-scale phylogeny of the alphavirus genus suggests a marine origin. J Virol 86:2729-38.
19. Suhrbier A, Jaffar-Bandjee MC, Gasque P. 2012. Arthritogenic alphaviruses--an overview. Nat Rev Rheumatol 8:420-9.
20. Zacks MA, Paessler S. 2010. Encephalitic alphaviruses. Vet Microbiol 140:281-6.
21. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113.
22. Kostyuchenko VA, Jakana J, Liu X, Haddow AD, Aung M, Weaver SC, Chiu W, Lok SM. 2011. The structure of barmah forest virus as revealed by cryo-electron

microscopy at a 6-angstrom resolution has detailed transmembrane protein architecture and interactions. J Virol 85:9327-33.

23. Mancini EJ, Clarke M, Gowen BE, Rutten T, Fuller SD. 2000. Cryo-electron microscopy reveals the functional organization of an enveloped virus, Semliki Forest virus. Mol Cell 5:255-66.

24. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596:583-589.

25. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung CC, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Zemgulyte A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Zidek A, Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630:493-500.

26. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE. 2023. UCSF ChimeraX: Tools for structure building and analysis. Protein Sci 32:e4792.

27. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-80.

28. Pletnev SV, Zhang W, Mukhopadhyay S, Fisher BR, Hernandez R, Brown DT, Baker TS, Rossmann MG, Kuhn RJ. 2001. Locations of carbohydrate sites on alphavirus glycoproteins show that E1 forms an icosahedral scaffold. Cell 105:127-136.

29. Tsetsarkin KA, Chen R, Sherman MB, Weaver SC. 2011. Chikungunya virus: evolution and genetic determinants of emergence. Curr Opin Virol 1:310-7.

30. Kaelber JT, Chmielewski D, Chiu W, Auguste AJ. 2022. Alphavirus Particles Can Assemble with an Alternate Triangulation Number. Viruses 14.

31. Baxter VK, Heise MT. 2018. Genetic control of alphavirus pathogenesis. Mamm Genome 29:408-424.

32. Poirier EZ, Mounce BC, Rozen-Gagnon K, Hooikaas PJ, Stapleford KA, Moratorio G, Vignuzzi M. 2015. Low-Fidelity Polymerases of Alphaviruses Recombine at Higher Rates To Overproduce Defective Interfering Particles. J Virol 90:2446-54.

33. Monroe SS, Schlesinger S. 1984. Common and distinct regions of defective-interfering RNAs of Sindbis virus. J Virol 49:865-72.

34. Kennedy SI, Bruton CJ, Weiss B, Schlesinger S. 1976. Defective interfering passages of Sindbis virus: nature of the defective virion RNA. J Virol 19:1034-43.

35. Langsjoen RM, Muruato AE, Kunkel SR, Jaworski E, Routh A. 2020. Differential Alphavirus Defective RNA Diversity between Intracellular and Extracellular Compartments Is Driven by Subgenomic Recombination Events. mBio 11.
36. Colunga-Saucedo M, Rubio-Hernandez EI, Coronado-Ipina MA, Rosales-Mendoza S, Castillo CG, Comas-Garcia M. 2022. Construction of a Chikungunya Virus, Replicon, and Helper Plasmids for Transfection of Mammalian Cells. Viruses 15.

Figure legends

Figure 1. A) Comparison of the cryo-EM structures of the cores of CHIKV (PDB ID: 3J2W), MAYV (7KO8), and GETV (7FD2). B) Comparison of cryo-EM structures of the cores EEEV (6XO4), VEEV (7SFV), and WEEV (8DEC). Each structure has four different colors, representing that in a T = 4 icosahedron, the capsid protein (CP) in the asymmetric unit has four different chemical environments.

Figure 2. Comparison of the structure of the pentamers with the structures in Figure 1. The proteins in the pentamer have the same color because they share the same chemical environment, unlike the proteins in the hexamer, which have three distinct chemical environments. A) The structures of the pentamers of these three Old-World Alphaviruses are very similar. B) The structure of the EEEV pentamer has fewer contact angles (asterisks) than the other New-World Alphaviruses. The contacts between CPs in VEEV and WEEV are similar to those of CHIKV, MAYV, and GETV.

Figure 3. A) Putative electrostatic interactions at the surface between three proteins in the CHIKV pentamer, between E184 and E234 of one protein and K172 and K177 of the other protein. B) The interactions at the quasi-threefold symmetry axis resemble dipole-dipole interactions between the pair of amino acids E120/K122 and D148/K151. C) CHIKV CP (PDB ID: 3J2W) monomer displaying E120, K122, D148, K151, K172, K177, E184, E234. D) Overlap of CHIKV CP with MAYV CP (PDB ID: 7KO8), where, although the numbering and chemical identity of some amino acids differ, the charges are conserved in all amino acids. E) Overlap of CHIKV CP with EEEV CP (PDB ID: 8DEC). This overlap indicates that some amino acids are not conserved in EEEV, suggesting that the intra- and inter-capsomere interactions in EEEV differ significantly from those in MAYV and CHIVK.

Figure 4. Comparison of the interfaces between the three proteins at i5 (pentamer). A) The "outer interactions" for CHIKV are between K172 and E234, and the "inner interactions" are between K177 and E184. It also shows that the possible interactions between amino acids at equivalent positions at the same interfaces for MAYV and GETV are the same as in CHIKV. B) The "outer interactions" for EEEV, VEEV, and WEEV are the same as in the Old-World Alphaviruses from the top row. However, the "inner interactions" are different. The inner interactions for EEEV involve a P and L, which could explain the large holes observed in Figure 2. In VEEV, the amino acids at the "inner position" are Gln and Lys, while for WEEV, they are P and Q.

Figure 5. A) The leftmost panel shows the geometry of the i3 axis for CHIVK, MAYV, and GETV, which was determined by measuring the centroid-to-centroid distance (2-Å centroid) and torsion angle between K122 of the three proteins at this axis. B) The distances between these three viruses and EEEV and WEEV, as well as those in the top row, are very similar. Only VEEV has a distance that is much smaller than that of the other viruses. Also, the distances for EEEV and WEEV were calculated by using Q121 and Q119, respectively, which are in equivalent positions to CHIKV CP K122. Interestingly, in VEEV, K134 and E133 do not have a pair of charged amino acids, as seen in the top row, that could help establish a dipole-dipole-like interaction. The torsion angles between three equivalent amino acids are indicated at the center of the i3 axes. The torsion angles for CHIKV, MAYV, and EEEV are very similar; however, VEEV and WEEV have larger torsion angles than the other four viruses.

Figure 6. A) Except for the Bebaru virus, all other Old-World Alphaviruses in this table have a predicted N-glycosylation site at a position that is equivalent to CHIKV N263. B) Cryo-EM structure of MAYV virus (PDB ID: 7KO8) showing the "molecular handshake" between glycosylation sites. This interaction occurs between two trimers of heterodimers. C) New-World Alphaviruses do not have a predicted N-glycosylation at an equivalent position to CHIKV N263. The only exception is the Fort Morgan virus, which has a possible glycosylation

site at N260, which is not in the same position as CHIKV E2 N263. D) Overlap of CHIKV E2 (green) with the predicted structure for the Fort Morgan virus E2 protein (salmon). CHIKV E2 N263 is in red and indicated by a red # and is in a β-sheet. The amino acid in the Fort Morgan virus E2 protein equivalent to CHIKV E2 N263 is in black and indicated by a black asterisk (Proline), while the Asp in an N-L-T motif is in a loop and is indicated with a red asterisk. If this amino acid were glycosylated, it is unlikely to be able to establish the same interactions as the glycosylation site shown in panel B. E) From the Alphaviruses that infect aquatic animals with available E2 sequences, only the Harbor Alphavirus has the same predicted E2 glycosylation site as CHIKVK. Interestingly, the core of this virus has seven of the eight interactions conserved in the CPs of Old-World Alphaviruses, as shown in Figures 4, 5, and 6 and Tables 1 and 4. This small correlation could suggest that the nature of the core and the interactions between trimers are related to each other.

Figure 7. Phylogenetic tree for the Alphaviruses in Tables 1-4. The predictions of possible common ancestors made from the data in Tables 1-4 are supported by this tree. The letters refer to the nodes of interest: *** is the common ancestor of the CP for Old- and New-World Alphaviruses (OWA and NWA, respectively). The table shows the identity of the amino acids of interest in nodes *** and A-G; the green and red colors indicate whether the charge of these amino acids is conserved or not with respect to CHIVK, respectively. The colors of the branches indicate the different clades. The prediction of the sequence of this node suggests that the seven amino acids of interest, which are highly conserved in Old-World Alphaviruses, were present in this ancestor; however, the origin of the amino acids observed in Old-World Alphaviruses today arose early in their evolution (see B). The node that gave origin to Sindbis is shared with Old-World Alphaviruses. At the same time, the early beginning of the rest of the New-World Alphaviruses started with a fast diversification of the intra- and intracapsomeric interactions. As these viruses evolved, the diversity of these interactions increased significantly.

Table legends

Table 1. Comparison of the identity of amino acids in equivalent positions to CHIKV CP E120, K122, D148, K151, K177, E184, and E234 for Old-World Alphaviruses. These amino acids were identified by predicting the structure of the consensus CP sequence of all viruses using AlphaFold2. The green color indicates that the charge and/or identity of the amino acids are conserved at least 80%.

Table 2. Comparison of the identity of amino acids in equivalent positions to CHIKV CP E120, K122, D148, K151, K177, E184, and E234 for New-World Alphaviruses. These amino acids were identified, as explained in Table 1. The lavender and turquoise colors indicate that these amino acids are at an equivalent position to CHIKV due to an insertion or a -1 shift, respectively. Only the positions equivalent to K172 and E234 are conserved at 94% and 100%, respectively.

Table 3. Comparison of the identity of amino acids in equivalent positions to CHIKV CP E120, K122, D148, K151, K177, E184, and E234 for Alphaviruses with no known host are not part of the Old-/New-World classifications These amino acids were identified, as explained in Table 1. The lavender color indicates that those amino acids are at an equivalent position to CHIKV because of insertion. Only the position equivalent to K172 and E234 are conserved at 94% and 100%, respectively. In this group, positions K172 and E234 are conserved, as well as E120 and K172.

Table 4. Comparison of the identity of amino acids in equivalent positions to CHIKV CP E120, K122, D148, K151, K177, E184, and E234 for aquatic Alphaviruses. These amino acids were identified, as explained in Table 1. The lavender, turquoise, and maroon colors indicate that those amino acids are at an equivalent position to CHIKV because of an insertion, a -1 shift,

or a deletion, respectively. Positions K172, E234, E120, and K172 are conserved. The gray color indicates that those viruses infect mammals. The rest infect fish.

Table 5. Comparison of the Shannon entropy and selective pressure of amino acids in equivalent positions to CHIKV CP E120, K122, D148, K151, K177, E184, and E234 Old-World Alphaviruses. This analysis was done employing SLAC, FEL, FUBAR, and MEME methods; sites were classified as being under pervasive or episodic positive selection if detected by MEME ($p<0.1$), FUBAR (posterior probability >0.9), FEL ($p<0.1$), or SLAC ($p<0.1$),

Figures

Figure 1

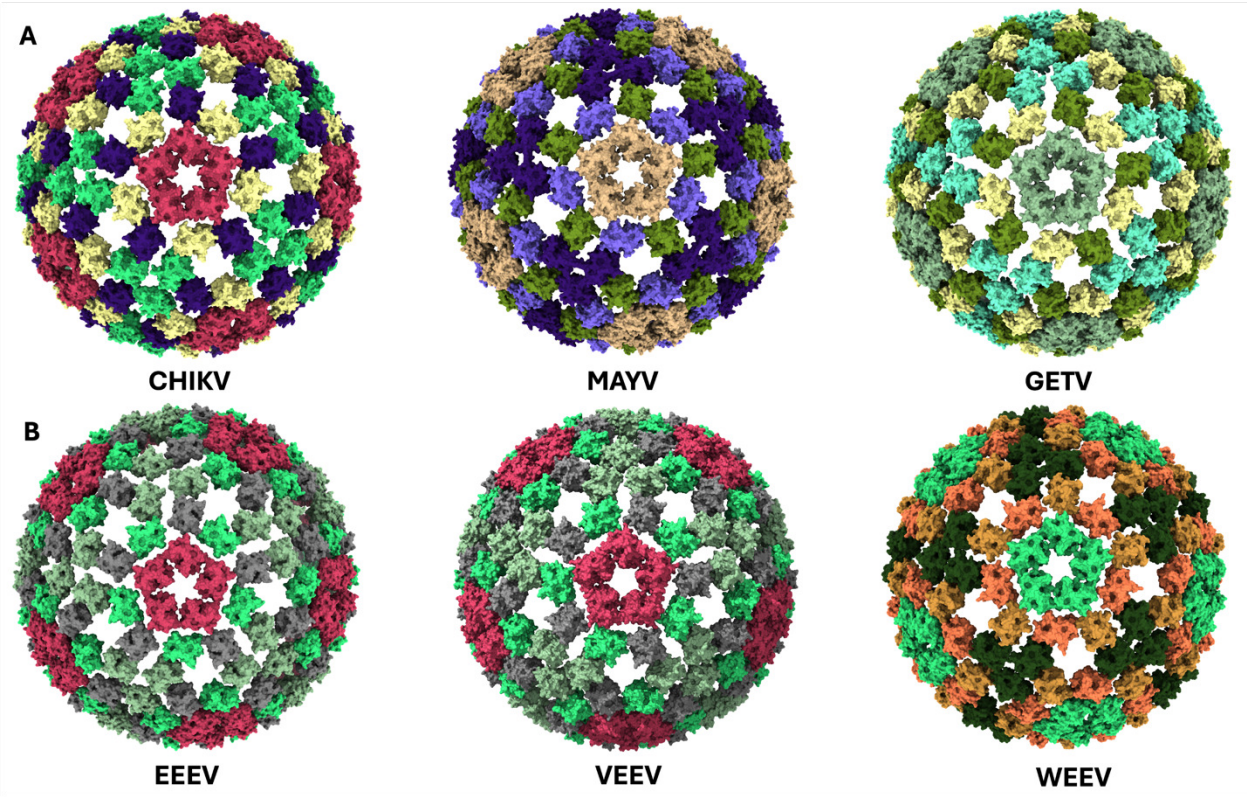


Figure 2

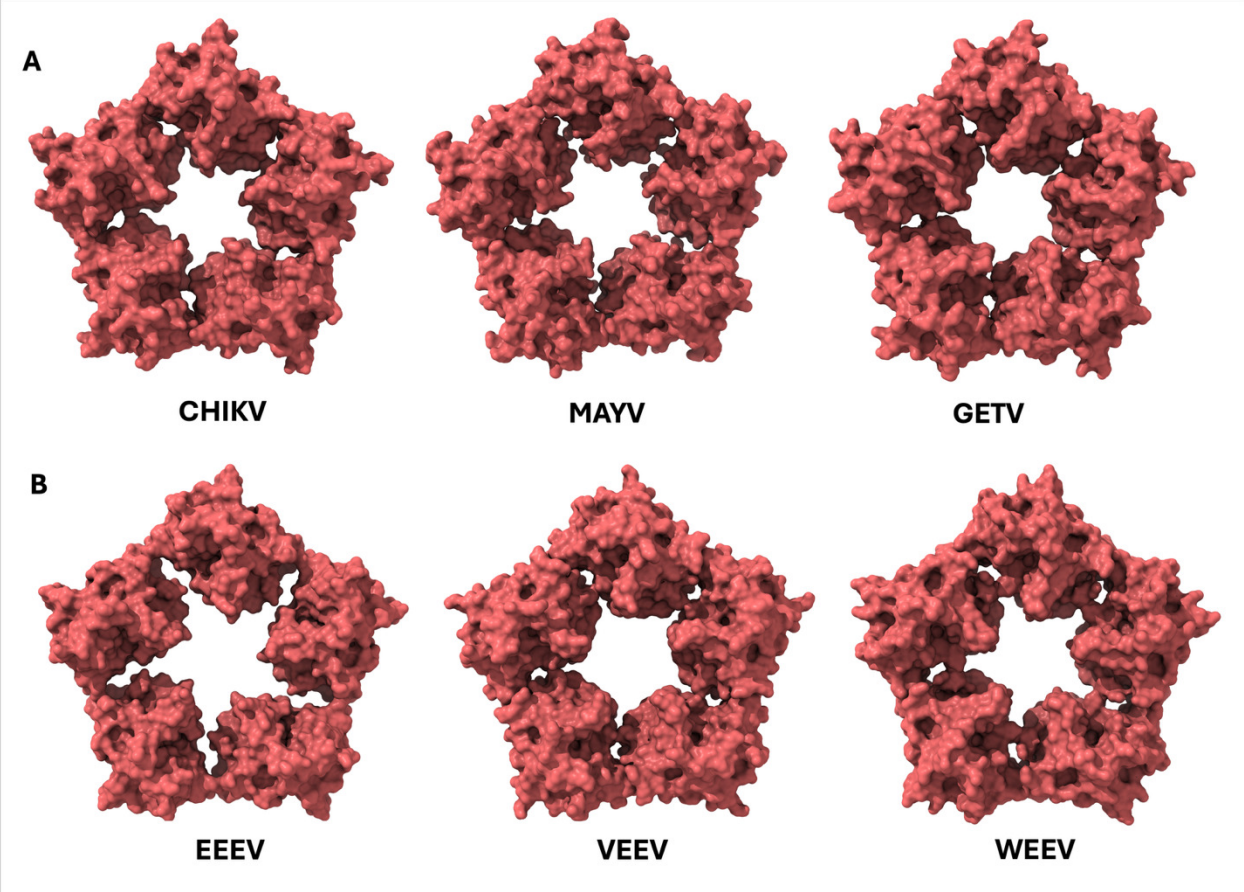


Figure 3

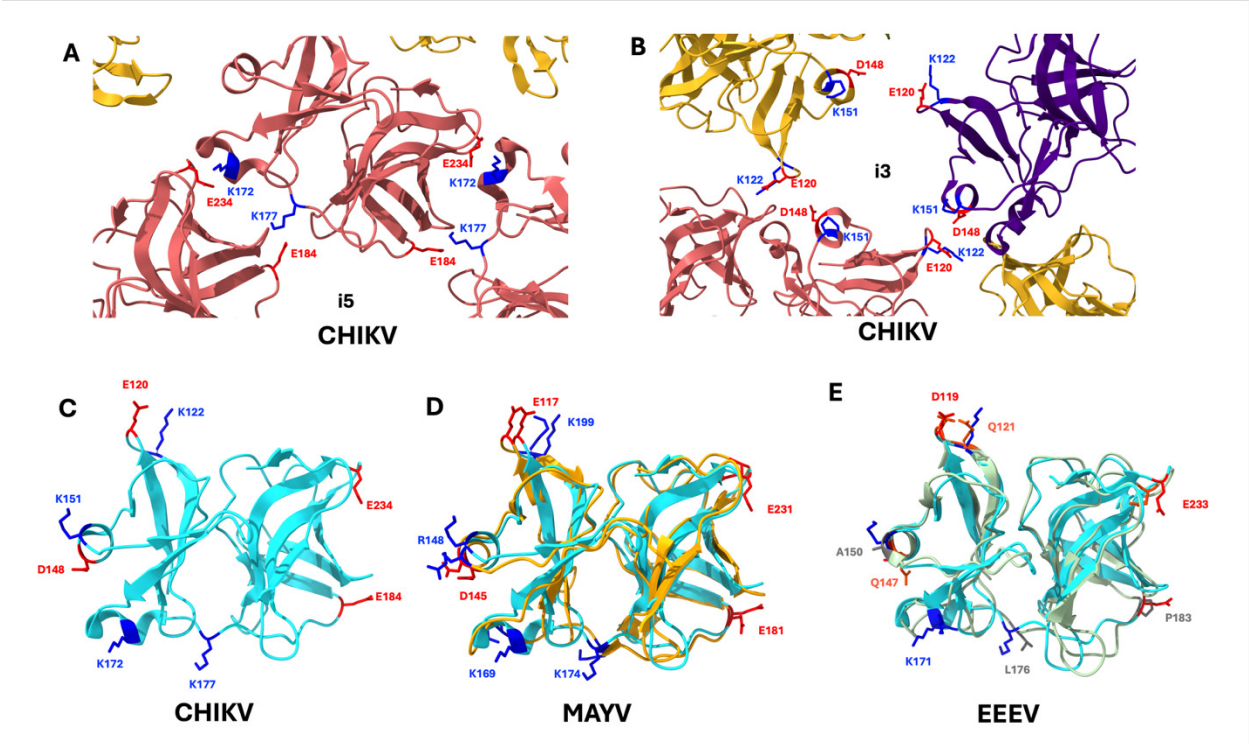

A
K172
E234
K177
E184
E234
K172
K177
E184
i5
CHIKV
B
D148
K151
K122
E120
i3
K122
E120
D148
K151
K151
D148
K122
E120
CHIKV
C
E120
K122
K151
D148
E234
E184
K172
K177
CHIKV
D
E117
K199
R148
D145
E231
E181
K169
K174
MAYV
E
D119
Q121
E233
A150
Q147
P183
K171
L176
EEEV

Figure 4

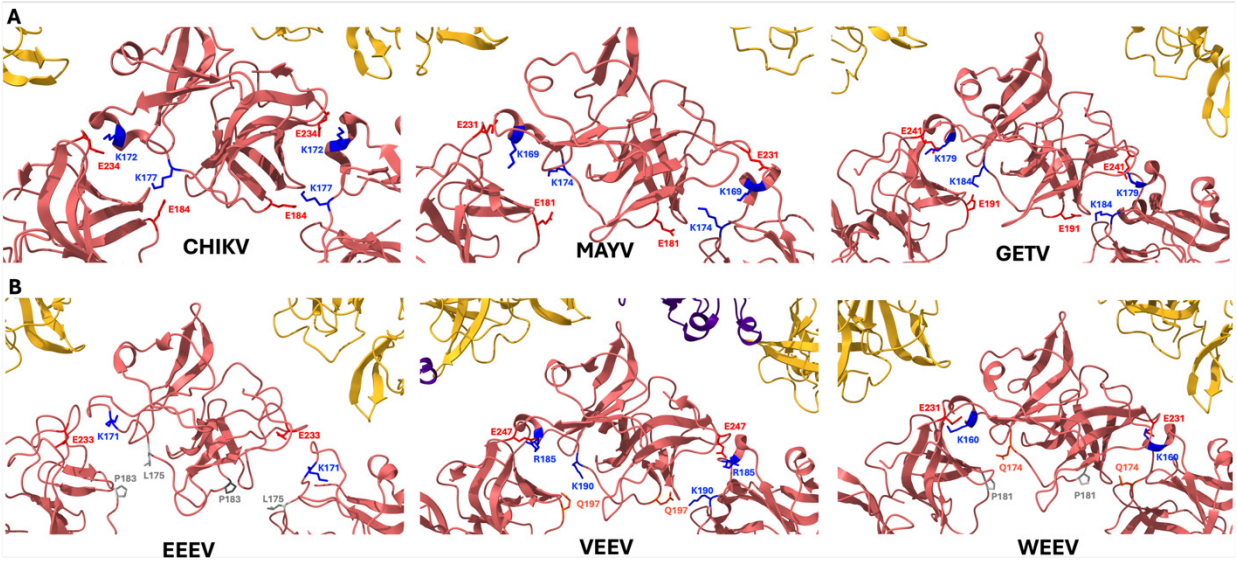

A
K172
E234
K177
E184
E234
K172
K177
E184
CHIKV
E231
K169
K174
E181
E231
K169
K174
E181
MAYV
E241
K179
K184
E191
E241
K179
K184
E191
GETV
B
E233
K171
P183
L175
E233
K171
P183
L175
EEEV
E247
R185
K190
Q197
E247
R185
K190
Q197
VEEV
E231
K160
Q174
P181
E231
K160
Q174
P181
WEEV

Figure 5

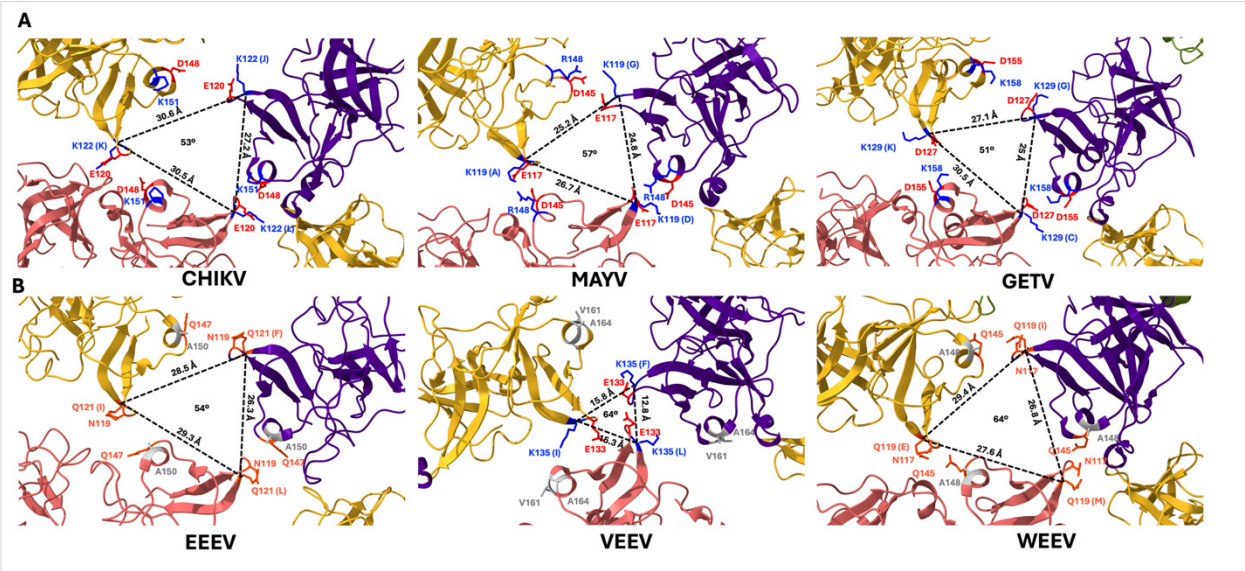
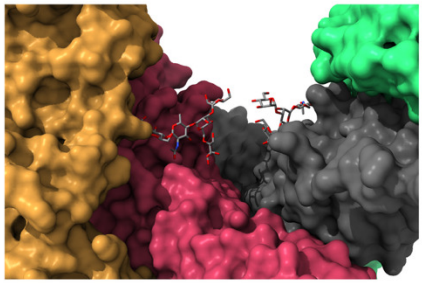


Figure 6

A

Virus	Table 5. Conservation of the amino acids of interest in the Old World Alphaviruses with respect to CHIKV			
CHIKV (Reference)	N263	+1	+2	Possible N-Gly
CHIKV	N	V	T	Yes
Getah virus	N	S	T	Yes
Semliki Forest virus	N	I	T	Yes
Mayaro virus	N	V	T	Yes
Ross River	N	T	T	Yes
Bebaru virus	G	V	T	No
Onyong-nyong virus	N	T	T	Yes
Una virus	N	S	T	Yes
Middelburg virus	N	S	T	Yes
Ndumu virus	N	T	T	Yes
% ID conservation with respect to CHIKV	90		100	

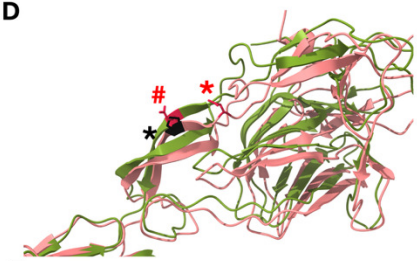
B



C

Virus	Table 6. Conservation of the amino acids of interest in the New World Alphaviruses with respect to CHIKV			
CHIKV (Reference)	N263	+1	+2	Possible N-Gly
Sindbis virus	P	S	T	No
Fort Morgan virus	P	T	T	No
	N	L	T	Yes
Aura virus	Q	A	E	No
Highlands J virus	A	T	A	No
Western equine encephalitis virus	P	T	V	No
Whataroa virus	P	G	Q	No
Eastern equine encephalitis virus	K	A	K	No
Venezuelan equine encephalitis virus	D	G	K	No
Everglades virus	D	G	K	No
Cabassou virus	E	A	K	No
Mucambo virus	E	A	K	No
Pixuna virus	D	A	R	No
RioNegro virus	G	A	K	No
Tonate virus	D	A	K	No
Madariaga virus	T	S	K	No
Mosso das Pedras virus	E	A	K	No
% ID conservation with respect to CHIKV	6		18	

D



E

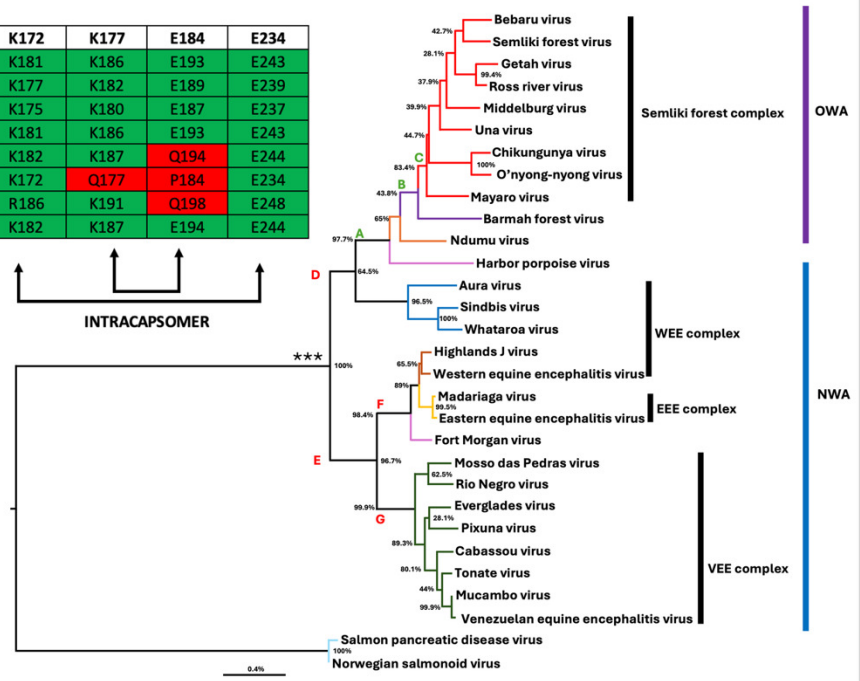
Virus	Table 7. Conservation of the amino acids of interest in aquatic Alphaviruses with respect to CHIKV			
CHIKV (Reference)	N263	+1	+2	Possible N-Gly
Salmon pancreatic disease virus	A	A	T	No
Harbor porpoise alphavirus	N	V	T	Yes
Norwegian salmonid	T	A	T	No
% ID conservation with respect to CHIKV	33		33	

Figure 7

	E120	K122	D148	K151	K172	K177	E184	E234
A	D129	K131	V157	K160	K181	K186	E193	E243
B	D125	K127	D153	K156	K177	K182	E189	E239
C	D123	K125	D151	K154	K175	K180	E187	E237
D	D129	K131	V157	K160	K181	K186	E193	E243
E	D130	K132	V158	A161	K182	K187	Q194	E244
F	N120	Q122	Q148	A151	K172	Q177	P184	E234
G	D134	K136	V162	A165	R186	K191	Q198	E248
***	D130	K132	V158	K161	K182	K187	E194	E244

INTERCAPSOMER

INTRACAPSOMER



Tables

Table 1

	Virus	Table 1. Conservation of the amino acids of interest in the Old World Alphaviruses with respect to CHIKV							
	CHIKV (Reference)	E120	K122	D148	K151	K172	K177	E184	E234
SF complex	CHIKV	E	K	D	K	K	K	E	E
	Getah virus	D	K	D	K	K	K	E	E
	Semliki Forest virus	E	K	D	K	R	K	E	E
	Mayaro virus	E	K	D	R	K	K	E	E
	Ross River	D	K	D	K	K	K	E	E
	Bebaru virus	D	K	D	K	R	K	E	E
	Onyong-nyong virus	E	K	D	K	K	K	E	E
	Una virus	D	K	D	K	R	K	E	E
BF complex	Barmah forest virus	D	K	E	K	K	K	E	E
MIID	Middelburg virus	E	K	D	K	K	Q	E	E
Ndumu complex	Ndumu virus	D	E	E	K	R	K	E	E
	% ID conservation with respect to CHIKV	45	91	82	91	64	91	100	100
	% Charge conservation with respect to CHIKV	100	91	100	100	100	91	100	100

Table 2

	Virus	Table 2. Conservation of the amino acids of interest in the New World Alphaviruses with respect to CHIKV							
	CHIKV (Reference)	E120	K122	D148	K151	K172	K177	E184	E234
WEE complex	Western equine encephalitis virus	N	Q	Q	A	K	Q	P	E
	Sindbis virus	D	D	V	K	R	T	E	E
	Fort Morgan virus	N	Q	E	A	K	Q	P	E
	Aura virus	D	K	A	K	K	G	E	E
	Highlands J virus	N	Q	Q	A	K	Q	P	E
	Whataroa virus	E	D	V	K	K	N	E	E
EEE complex	Eastern equine encephalitis virus	N	Q	Q	A	K	L	P	E
VEE complex	Venezuelan equine encephalitis virus	E	K	V	A	R	K	Q	E
	Everglades virus	E	K	V	A	R	K	Q	E
	Cabassou virus	D	K	T	A	R	R	Q	E
	Mucambo virus	D	K	T	S	R	K	Q	E
	Pixuna virus	E	R	A	A	R	K	Q	E
	RioNegro virus	D	R	V	T	R	K	Q	E
	Tonate virus	D	K	T	A	R	K	Q	E
Caaingua complex	Caaingua virus	D	K	E	N	I	K	M	E
Madariaga complex	Madariaga virus	N	Q	Q	A	K	Q	P	E
MdP complex	Mosso das Pedras virus	D	K	I	A	R	K	R	E
	% ID conservation with respect to CHIKV	24	47	0	18	41	47	18	100
	% Charge conservation with respect to CHIKV	71	59	12	18	94	53	18	100

Table 3

	Virus	Table 3. Conservation of the amino acids of interest in Alphaviruses found in Mosquitoes with no vertebrate host with respect to CHIKV							
	CHIKV (Reference)	E120	K122	D148	K151	K172	K177	E184	E234
Eilat complex	Eilat virus	K	V	L	S	R	A	D	E
Trocara complex	Trocara virus	E	K	A	K	R	K	E	Q
Unclassified	AguaSalud alphavirus	D	Q	I	K	K	R	E	E
Unclassified	Mwinilunga alphavirus	D	A	L	T	R	N	D	E
Unclassified	Pirahy virus	D	K	T	S	R	K	Q	E
Unclassified	Tai Forest alphavirus	D	V	L	S	R	S	D	E
Unclassified	Yada yada virus	E	R	T	S	R	A	D	E
	% ID conservation with respect to CHIKV	29	29	0	29	14	29	29	86
	% Charge conservation with respect to CHIKV	86	43	0	29	100	43	86	86

Table 4

	Virus	Table 4. Conservation of the amino acids of interest in aquatic Alphaviruses with respect to CHIKV							
	CHIKV (Reference)	E120	K122	D148	K151	K172	K177	E184	E234
SPDV	Salmon pancreatic disease virus	D	A	E	D	R	E	D	D
SES complex	Southern Elephant Seal virus	D	L	I	K	R	K	E	E
Trocara complex	Comber alphavirus	E	T	E	K	K	K	D	D
Unclassified	Harbor alphavirus	D	K	V	K	K	K	D	E
Unclassified	Norwegian salmonid	D	A	E	D	R	E	D	D
Unclassified	Wenling fish alphavirus	K	T	L	E	H	A	E	E
Unclassified	Wenling hagfish alphavirus	E	G	D	N	R	K	D	D
	% ID conservation with respect to CHIKV	29	14	14	43	29	57	29	43
	% Charge conservation with respect to CHIKV	86	14	57	43	86	57	100	100

Table 5

Table 5. Selective pressue and entropy of the amino acids of Old-Worl Alphaviruses

V irus	Selection						
	-	-	-	-	Purifying*	-	
	Entropy						
	0.689	0.305	0.305	0.305	0.655	0.000	0.305
Chikungunya Virus/MK286896	120	122	148	151	172	184	234
	E	K	D	K	K	E	E
Barmah Forest Virus/MW556197	112	114	140	143	164	176	226
	D	K	E	K	K	E	E
Bebaru Virus/NC 016962	122	124	150	153	174	186	236
	D	K	D	K	R	E	E
Semliki Virus/KF283988	126	128	154	157	178	190	240
	E	K	D	K	R	E	E
Getah Virus/PP236766	127	129	155	158	179	191	241
	D	K	D	K	K	E	E
Ross River Virus/NC 001544	129	131	157	160	181	193	243
	D	K	D	K	K	E	E
Una Virus/AF339481	138	140	166	169	190	202	252
	D	K	D	K	R	E	E
Middleburg Virus/AF398374	130	132	158	161	182	194	244
	E	K	D	K	K	E	D
Mayaro Virus/KP842808	117	119	145	148	169	181	231
	E	K	D	R	K	E	E
O nyongnyong Virus/NC 075006	119	121	147	150	171	183	233
	E	K	D	K	K	E	E
Ndumu Virus/MZ091531	128	130	156	159	180	192	242
	D	E	D	K	R	E	E

*Detected by two different algorithms

*Detected by one or no algorithms