Article

Accurate Recapitulation of Chikungunya Virus Complete Coding Sequence Phylogeny Using Variable Genome Regions for Genomic Surveillance

Eduardo D. Rodríguez-Aguilar [ID], Everardo Gutiérrez-Millán and Mario H. Rodríguez *[ID]

Center for Infectious Disease Research, National Institute of Public Health of Mexico, Av. Universidad 655, Cuernavaca 62100, Mexico; rodriguezaed@gmail.com (E.D.R.-A.); ever.gmillan@gmail.com (E.G.-M.)
* Correspondence: mhenry@insp.mx; Tel.: +52-777-3293087 (ext. 1109)

Abstract: Chikungunya virus (CHIKV) is transmitted by mosquito bites and causes chikungunya fever (CHIKF). CHIKV has a single-stranded RNA genome and belongs to a single serotype with three genotypes. The Asian lineage has recently emerged in the Western Hemisphere, likely due to travel-associated introduction. Genetic variation accumulates in the CHIKV genome as the virus replicates, creating new lineages. Whole genome sequencing is ideal for studying virus evolution and spread but is expensive and complex. This study investigated whether specific, highly variable regions of the CHIKV genome could recapitulate the phylogeny obtained with a complete coding sequence (CDS). Our results revealed that concatenated highly variable regions accurately reconstructed CHIKV phylogeny, exhibiting statistically indistinguishable branch lengths and tree confidence compared to CDS. In addition, these regions adequately inferred the evolutionary relationships among CHIKV isolates from the American outbreak with similar results to the CDS. This finding suggests that highly variable regions can effectively capture the evolutionary relationships among CHIKV isolates, offering a simpler approach for future studies. This approach could be particularly valuable for large-scale surveillance efforts.

Keywords: Chikungunya virus; phylogeny; genomic surveillance

1. Introduction

Chikungunya virus (CHIKV) is an alphavirus transmitted by the bite of infected *Aedes aegypti* and *Aedes albopictus* female mosquitoes. It causes chikungunya fever (CHIKF), an acute febrile disease with four clinical forms: acute, atypical acute, severe acute and chronic [1]. The CHIKV genome consists of a single stranded positive sense 12 kB RNA. The viral mRNA includes a $5'$ methylguanylate cap and a $3'$ polyadenylated tail, encoding two open reading frames (ORFs) for non-structural and structural polyproteins [2]. The non-structural ORF encodes four proteins (nsp1, nsp2, nsp3, and nsp4) essential for transcription, replication and caping of the viral RNA, and the viral polyprotein cleavage; the structural ORF encodes five proteins (C, E3, E2, 6K, y E1) [3].

CHIKV comprises a single serotype [4], with three genotypes distinguished by their historical geographic distributions: West African (WAf), East/Central/South African (ECSA), and Asian lineages [5,6]. In 2005–2006, the Indian Ocean Lineage (IOL), a strain of the ECSA lineage, sparked a significant epidemic in the Indian Ocean islands and the Indian subcontinent, marking the first major outbreak after 32 years of inactivity [5,7]. Isolated cases, associated with travelers returning from this region, led to small outbreaks in Italy in 2007 [8] and France in 2010 [9].

It has been suggested that the emergence of CHIKV in the Western hemisphere dates back to the 19th century, with cases potentially misdiagnosed as dengue fever (DENF) [10]. This speculation is plausible due to the overlapping signs and symptoms of CHIKF and

DENF [11]. Nonetheless, before December 2013, when a CHIKF outbreak was reported in the Caribbean Island of Saint Martin, there was no substantial evidence of local transmission [12].

The immunologically naive population in the Americas, the widespread *Ae. aegypti* mosquitoes, and the frequent travel between islands facilitated the rapid spread of the virus throughout the Caribbean and eventually to the American mainland [13]. By February 2014, autochthonous transmission had been confirmed in ten territories, leading to 2582 confirmed cases of CHIKF [14]. By December 2014, transmission had been confirmed in 33 territories, with a total of 1,071,696 suspected and 22,796 confirmed cases [14]. Certain strains of the IOL undergo adaptive mutations in the E1 and E2 envelope glycoproteins, enhancing its infectivity for *Ae. albopictus* [15]. This adaptation opened opportunities for CHIKV transmission in regions where this mosquito is prevalent and its extension beyond the tropical and subtropical areas where *Ae. aegypti* is endemic [16].

The significant number of infected travelers between 2006 and 2008 raised expectations that the IOL strain would be the first to emerge in the Western hemisphere [17]. However, sequence analysis has revealed that CHIKV isolated from Saint Martin at the onset of the outbreak belonged to the Asian lineage [12]. Subsequent analysis of CHIKV from the Caribbean islands, Martinique [18], the British Virgin Islands [19], and Trinidad [20], as well as from the American mainland, including Mexico [21], Panama [22], Colombia [23], and Brazil [24], has confirmed the presence of the Asian lineage. On the other hand, alongside the circulation of the Asian genotype, an outbreak caused by a newly introduced ECSA strain occurred in the east-central region of Brazil [24]. As sequence data were limited to a small subset of CHIKV-affected countries, it remains uncertain whether the ECSA genotype was more prevalent, other genotypes were circulating, or if the outbreak stemmed from a single Asian genotype introduction. Analyzing genomic sequences from a broader region could enhance accuracy in estimating evolutionary rates and identifying the selective pressures influencing viral evolution.

The spread of CHIKV is accompanied by the accumulation of genetic variation, leading to the emergence of distinct lineages with potential virulence and transmission differences [25]. The accumulation of genetic variations is mainly produced by errors generated by RNA polymerase during viral replication (10^3 to 10^5 substitutions per nucleotide copied per round of replication) [26]. Although mutations arise stochastically within the genome, viral genes contain conserved regions, critical to preserve structure and function, combined with regions permissive to multiple mutations, which allow immune evasion through antigenic variation [27].

Advances in DNA sequencing and bioinformatics tools have improved viral genomic surveillance [28]. These allow thorough examination of the genome and dissemination patterns of emerging and reemerging pathogens [29]. Despite this progress and the significant number of reported CHIKF cases in recent years, public genomic data primarily consists of genomes from localized outbreaks, limiting their representativeness [30]. Extensive genomic surveillance efforts can provide insights into the genetic diversity of the circulating viral populations during outbreaks and periods of reduced transmission. However, epidemiological surveillance based on whole genomes is limited in many endemic countries due to the technical and analytical complexity and high costs [31]. The inherent variability within the CHIKV genome presents an opportunity to identify hypervariable regions. If these regions demonstrate sufficient phylogenetic signal to recapitulate results obtained from complete genome analysis, they could serve as useful alternatives for evolutionary studies.

Previously, we demonstrated that highly variable regions of the dengue virus genome can be useful in recapitulating the phylogeny obtained using the whole genome [32]. In this study, we examined genetic variability across the complete coding sequence (CDS) of CHIKV of a selection of sequences from around the world. We examined the performance of various genomic regions with differing variability levels for phylogenetic analysis, compared to that obtained with the whole genome. We propose that hypervariable ge-

nomic regions could be useful for monitoring CHIKV evolution through ongoing genomic surveillance efforts.

2. Materials and Methods

2.1. Genetic Variability Analysis

A total of 1114 CHIKV complete genome sequences from the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) of the National Institute of Allergy and Infectious Diseases (NIH/DHHS), available at https://www.bv-brc.org/ (accessed on 5 February 2024), were included in this study.

We conducted a genetic variability analysis to explore whether specific regions of the CHIKV genome (conserved or variable regions) could yield useful phylogenetic signals [32]. We identified candidate regions and assessed their ability to replicate the phylogeny compared to that using the complete coding sequence (CDS). To identify conserved and highly variable regions, we aligned 45 nt non-overlapping sliding windows of the CDS sequences using Clustal Omega software (v. 1.2.1) [33]. For each window, we calculated the number of nucleotide changes and the ratio of synonymous to non-synonymous mutations for every codon. Three regions exhibiting the highest number of mutations and the highest ratios of synonymous to non-synonymous mutations, as well as three regions displaying the lowest number of mutations and the lowest rates of synonymous to non-synonymous mutations, were selected for subsequent phylogenetic analysis.

2.2. Phylogenetic Analysis

To assess the sensitivity of various regions in discriminating between sequences from different lineages and geographic origins, datasets were constructed with sequences from Africa, Asia, and America. These datasets incorporated sequences belonging to the three well-defined genotypes (Asian and Caribbean, ECSA, and West Africa) and lineages such as IOL and the Caribbean. Redundant sequences were removed from the initial datasets to prevent overrepresentation caused by oversampling. The remaining sequences were then randomly chosen to create six different alignments, each consisting of 30 sequences (detailed sequence data from the six alignments are presented in Table S1).

Potential sequence recombination events were screened using Genetic Algorithm for Recombination Detection (GARD), accessible via the Datamonkey web server, available at https://www.datamonkey.org/gard (accessed on 26 February 2024). Multiple sequence alignments for each alignment, based on the Percent Nucleotide Identities (PNI) calculated using p-distances, were generated using Clustal Omega (Version 1.2.1) [33]. From these alignments, three highly variable regions (Hi) of 100, 300, 500, 700, 900, 1200, 1500, and 2100 nt each, corresponding to regions of genes E2, NSP1, and NSP3 (Hi-E2, Hi-NSP1, and Hi-NSP3), and three regions with a low rate of variability (Lo) of 100, 300, 500, 700, 900, 1200, 1500, and 2100 nt each, corresponding to regions of genes C, NSP1, and NSP2 (Lo-C, Lo-NSP1, and Lo-NSP2) were extracted.

Separate phylogenetic analyses were performed using the three highly variable regions of identical lengths, as well as concatenated sequences constructed with the three high (Hi-Concatenated) or low (Lo-Concatenated) variable regions, respectively (3×100, 3×300, 3×500, and 3×700 nt). The best-fit model of nucleotide substitution for each alignment was determined based on the Bayesian information criterion (BIC) using ModelTest 3.5 [34]. The GTR + G + I model (general time-reversible model with gamma-distributed rates of variation among sites and a proportion of invariable sites) was identified as the best-fit model for all alignments. Phylogenetic trees were constructed with these sequences using Mr. Bayes software v.3.2 [35].

2.3. Recapitulating the Phylogeny of the CDS

Using trees constructed with the complete coding region (CDS) of each alignment as references, we assessed the performance of each candidate sequence (identified from the genetic variability analysis) in reproducing the phylogeny obtained with the entire

coding sequence. For each tree constructed with the different regions under evaluation, we computed relative branch length, topological incongruence, and tree confidence. Relative branch length was determined using the branch length distance (BLD). The BLD metrics quantify the similarity in branch lengths between phylogenetic trees. They calculate a scaling factor that minimizes the total evolutionary divergence between a reference tree and a compared tree. A BLD value of 1 indicates identical total branch length in both trees. Values greater than 1 suggest that the compared tree has proportionally shorter branches compared to the reference one, while values less than 1 indicate proportionally longer branches, thus quantifying the scaling factor to approximate the overall divergence between the reference and compared trees [36]. Topological incongruences were evaluated using the K-tree score, measuring the minimum branch length distance between trees after scaling. Lower K-tree scores indicate greater similarity to the reference tree. This calculation was performed using the Perl program Ktreedist Version 1.0 [37]. Additionally, tree confidence was computed as the mean posterior probability (mean PP) [38]. This parameter was normalized based on the number of nodes in each group of trees resulting from summing the posterior probability of all nodes in each compared tree divided by the number of nodes in the corresponding reference tree.

3. Results

3.1. Genetic Variability Analysis

The analysis of CHIKV genome variability identified regions of high variability and highly conserved regions (Figure 1).
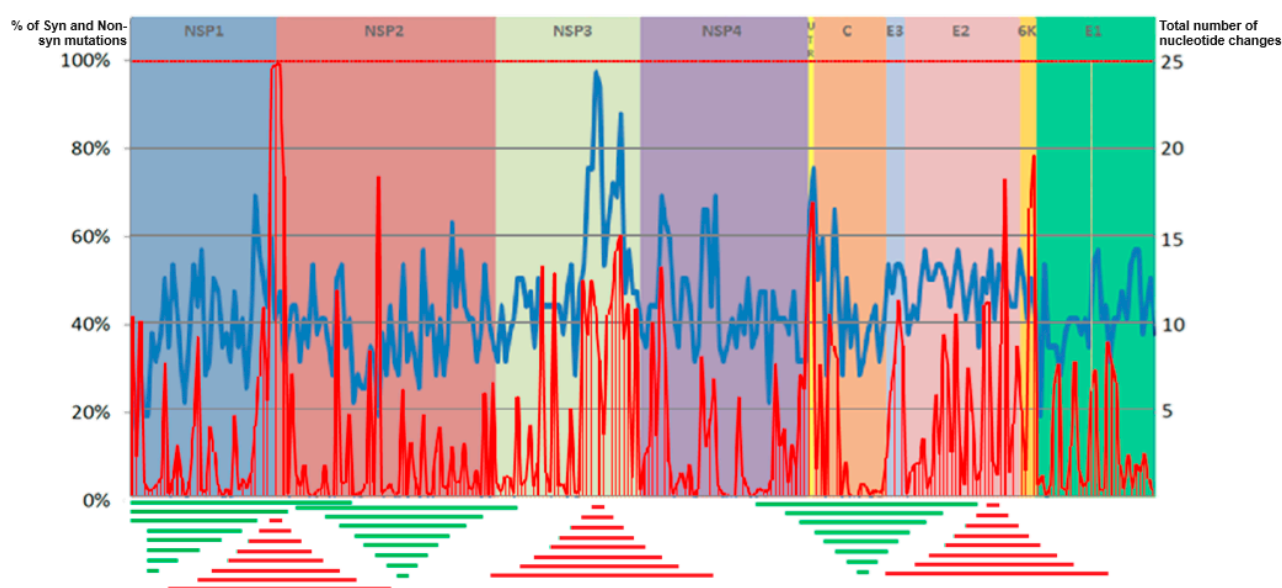
Figure 1. Genetic variability analysis of the CHIKV complete coding sequence (CSD). The analysis utilized 924 complete CHIKV genome sequences from the Bacterial and Viral Bioinformatics Resource Center (BV-BRC). A recurring pattern of peaks and troughs across the CDS reflect variations in the mutation rate and the relative proportions of synonymous and non-synonymous nucleotide substitutions. The *x*-axis corresponds to the nucleotide position within the CDS. The continuous blue line (right *y*-axis) represents the total number of nucleotide changes observed within a 45-nucleotide sliding window. The red line depicts the percentage of synonymous and non-synonymous mutations (left *y*-axis). The lengths of the regions used in subsequent analyses are depicted below the figure. High-variability regions (red lines) correspond to nucleotide numbers 100, 300, 500, 700, 900, 1200, and 1500 nt. Low-variability regions (green lines) correspond to positions 2100, 1500, 900, 700, 500, 300, and 100 nt.

Analysis of the hypervariable regions and the proportion of synonymous and non-synonymous mutations showed that the region with the highest number of mutations occurred in the NSP3 gene, followed by regions in the NSP1, NSP4, and E2 genes. Analysis of synonymous and non-synonymous mutations across the CDS showed a pattern of peaks with a high proportion of non-synonymous mutations alternating with peaks with a low proportion of non-synonymous mutations. The highest peak of non-synonymous mutations (nearly 100% of the mutations found were non-synonymous) occurred in the NSP1 gene. The NSP3 and E2 genes showed regions with peaks over 50% of non-synonymous mutations. The regions with the lowest number of mutations and the lowest rate of synonymous and non-synonymous mutations were in NSP1, NSP2, and C (including regions where no non-synonymous mutations were found) (Figure 1).

To assess the capacity of regions with genetic variation in recapitulating the CDS phylogeny, we used the three regions of the genome with the highest numbers of mutations and the highest proportions of non-synonymous mutations, separately, and a sequence formed by concatenating the same three regions. These included regions of the NSP1, NSP3, and E2 genes. The three regions of the genome exhibiting minimal mutational burden and the lowest proportions of non-synonymous mutations, isolated and concatenated, were used as controls.

### 3.2. Evaluation of the Suitability of Hypervariable Regions for Recapitulating CDS Phylogeny

3.2.1. Branch Length Distance

Analysis of BLD revealed that regions of both high and low variability up to 700 nucleotides (100, 300, 500, and 700 nt) resulted in scaling factors surpassing 1. This trend remained consistent across all regions, except for Hi-NSP3, which consistently generated scaling factors below 1. These results imply that phylogenetic trees constructed from these variable regions display relatively short branches compared to trees reconstructed using the CDS (Figure 2). For regions of 900 nt or longer (1200, 1500, and 2100 nt), BLD values progressively approached 1, reaching a point where no statistically significant differences were observed compared to an ideal scaling factor. This suggests that longer highly variable regions may more accurately reflect viral phylogeny compared to the shorter regions. Concatenated highly variable regions maintained values below 1, except for the 2100 nt region. The low-variability regions did not follow this trend, consistently exhibiting BLD values above 1 across all analyzed lengths (Figure 2).
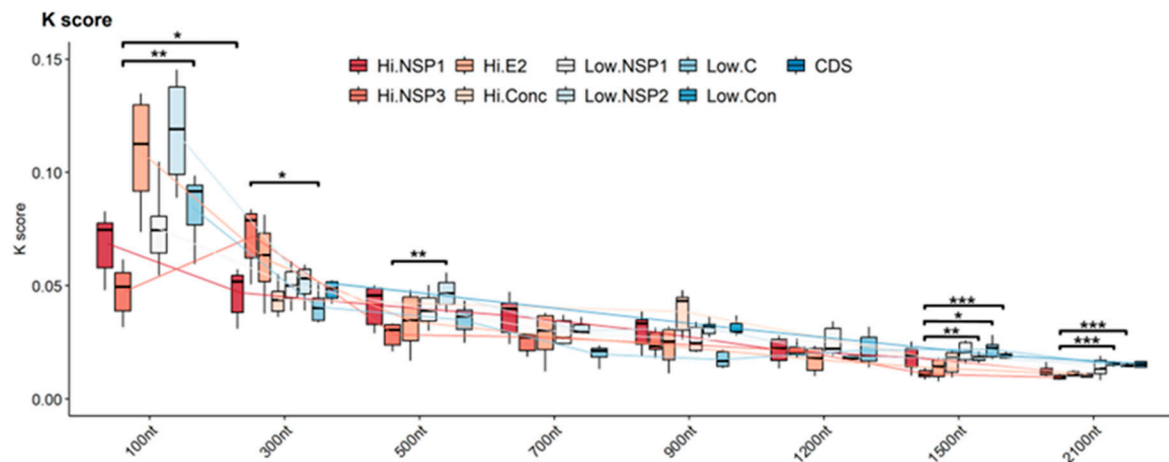
3.2.2. Topological Incongruence

Topological incongruence between phylogenetic trees was assessed using the K-tree score. This metric calculates the minimum branch length distance required to transform one tree into another after scaling them to the CDS reference tree length. Across all analyzed regions, there was a consistent decline in K-tree scores as the length of the analyzed region increased. For 100 nt regions, K-tree scores varied from 0.03 to 0.15, indicating a high level of uncertainty in the tree's reconstruction. This variability slightly decreased for 300 nt regions, with scores ranging from 0.03 to 0.08. A notable enhancement was observed as the analyzed region size increased. For 2100 nt regions, k-tree scores exhibited a much narrower range, with NSP3 being the best-performing region, with scores between 0.008 and 0.01. In contrast, concatenated low-variability regions showed the lowest performance, having scores ranging from 0.009 to 0.03. In contrast, the high-variability and the low-variability 100, 300, 1500, and 2100 nt regions of NSP3 presented statistically significant differences. In addition, reduced dispersion in K scores was observed with increasing region size, indicating that larger regions provide a more robust signal for reconstructing the evolutionary relationships represented by the topology of CSD trees (Figure 3).
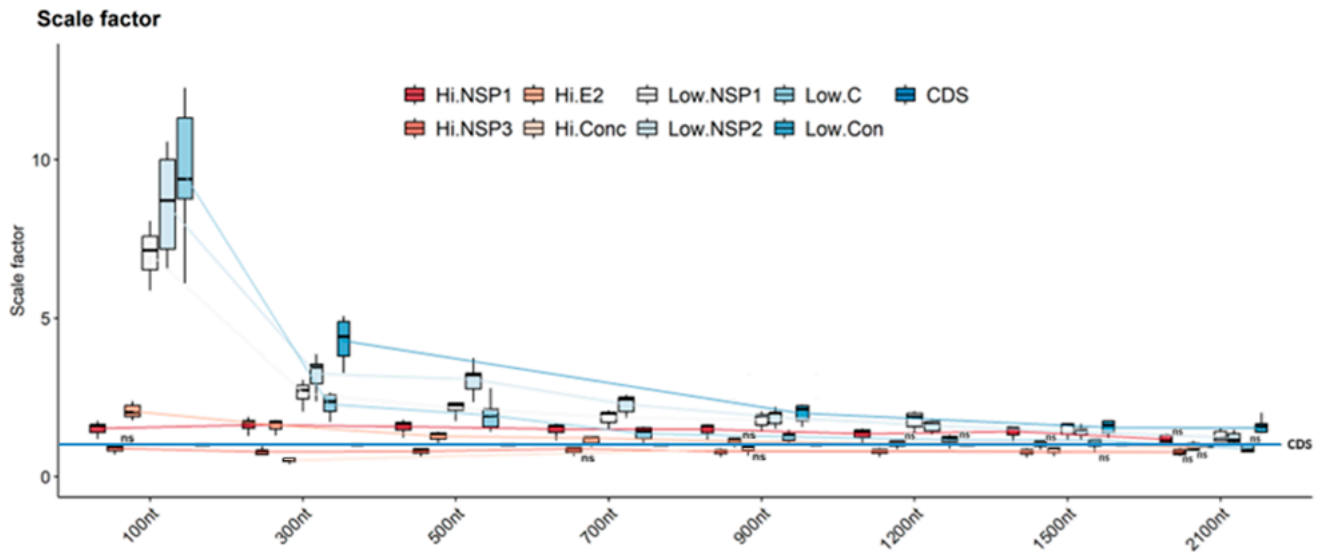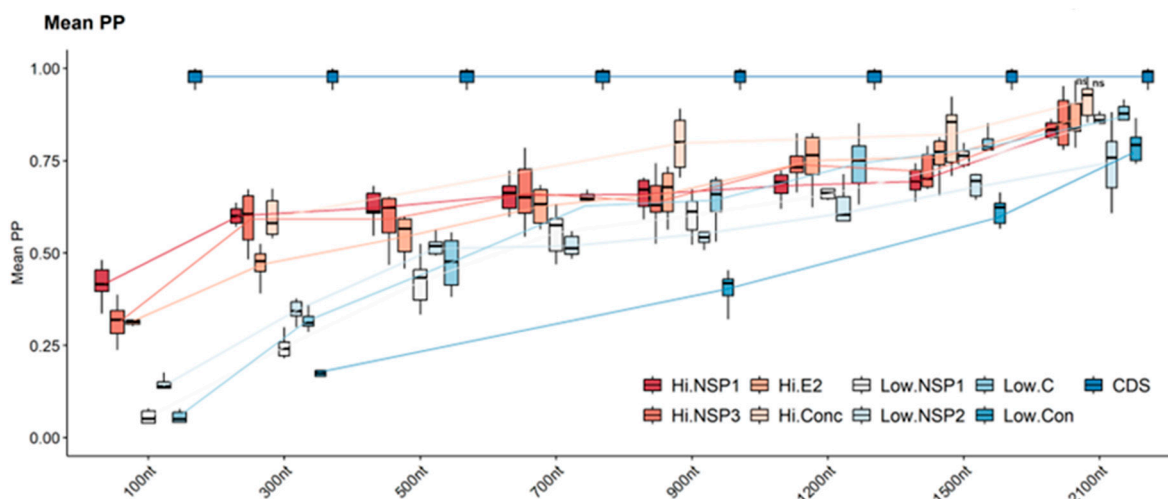
Figure 2. Comparison of relative branch length between CDS tree and trees constructed with high- and low-variation sequences. The trees constructed with the different regions (Hi- and Low-variables) and different lengths (100, 300, 500, 700, 900, 1200, 1500, and 2100 nt) were compared with the one constructed with their respective CDS to obtain the scaling factor. For each column, the box represents the second (Q2) and third (Q3) quartiles of the data distribution, the line within the box marks the median, whiskers denote the most extreme data. One-way ANOVA analysis of variance and the *p*-value of Tukey's test were used to identify specific regions with non-significant differences compared to the CDS phylogeny. Regions with *p*-values "ns" (not significant) denote *p* values greater than 0.05.

Figure 3. Comparison of the topological incongruence between CDS tree and trees constructed with high- and low-variation sequences. The trees constructed with the different regions (Hi- and Low-variables) and different lengths (100, 300, 500, 700, 900, 1200, 1500, and 2100 nt) were compared with the one constructed with their respective CDS to obtain the K score. For each column, the box represents the second (Q2) and third (Q3) quartiles of the data distribution, the line within the box marks the median, whiskers denote the most extreme data. ANOVA one-way analysis of variance and the *p*-value of Tukey's test were used to verify the existence of statistical differences between the K-tree scores of the different regions evaluated. * *p*-values < 0.05, ** *p*-values between 0.01 and 0.001, *** *p*-values < 0.001.

3.2.3. Tree Confidence

Tree confidence, assessed using mean posterior probability (mean PP), required normalization due to variations in the number of nodes recovered across different tree sets. The CDS phylogenetic tree exhibited high confidence in its branching patterns. This is reflected in the mean PP values for the nodes within the tree, which ranged from 0.94 to 1. Mean PP values from the high- and low-variability regions exhibited a consistent increase as the analyzed region length increased and regions with high variability performed better than regions with low variability. Analysis of nucleotide region length revealed a consistent positive correlation between region length and mean posterior probability (PP) values for both high- and low-variability regions. High-variability regions exhibited consistently higher mean PP values compared to their low-variability counterparts. For instance, at a region length of 100 nucleotides, the mean PP values in high-variability regions ranged from 0.23 to 0.48, while those in low-variability regions spanned only 0.04 to 0.17. Despite initially exhibiting very low mean PP values, high-variability regions demonstrated a significant increase in tree confidence, reaching moderate levels (mean PP values greater than 0.75) by 1200 nucleotides (mean PP: 0.62–0.82). Also, very high confidence consistency (mean PP values greater than 0.9) was only achieved with high-variability regions Hi-NSP3 and HiE2, and the concatenated high-variability region, all analyzed at region lengths of 2100 nucleotides. This result suggests that larger regions provide a more robust phylogenetic signal, ultimately enhancing the confidence of reconstructed trees. In contrast, concatenated low-variability regions exhibited statistically significant decreases in mean PP values compared to all other regions (regardless of length). Conversely, concatenated highly variable regions demonstrated better mean PP values compared to other regions, ultimately reaching a point where no statistically significant differences were observed from the values obtained using the reference tree (CSD) (Figure 4). (Detailed pairwise statistical comparisons of scaling factors, K-scores, and mean PP values between all analyzed regions are provided in Figure S1).

Figure 4. Comparison of tree confidence between CDS tree and trees constructed with high- and low-variation sequences. The trees constructed with the different regions (Hi- and Low-variability) and different lengths (100, 300, 500, 700, 900, 1200, 1500, and 2100 nt) were compared with those constructed with their respective CDS to obtain the mean PP. For each column, the box represents the second (Q2) and third (Q3) quartiles of the data distribution, the line within the box marks the median, whiskers denote the most extreme data. ANOVA one-way analysis of variance and the *p*-value of Tukey's test were used to identify specific regions with non-significant differences compared to the CDS phylogeny. Regions with *p*-values greater than 0.05 are indicated by "ns".

3.3. Inferring the Evolutionary Relationships among CHIKV Isolates from the American Outbreak Using Genomic Regions of Different Variability

To evaluate the suitability of highly variable genome regions of CHIKV for outbreak tracking through genomic surveillance, we employed a comparative phylogenetic approach. We reconstructed phylogenetic trees using the CDS of CHIKV isolates from the three distinct genotypes (Asian and Caribbean, ECSA, and West Africa) previously determined using the complete genome and sequences collected during the American outbreak (detailed sequence data employed for this analysis can be found in Table S2). Subsequently, we generated phylogenetic trees for the same isolates using the selected genomic regions exhibiting varying degrees of variability. By comparing the resulting trees to the reference trees derived from the complete CDS, we assessed the congruence between the evolutionary relationships inferred from these distinct genomic regions. The phylogenetic tree constructed using the CDS effectively distinguished the three distinct genotypes (Asian and Caribbean, ECSA, and West Africa). Within the ECSA genotype, it further differentiated between the lineage circulating in Asia and the one that circulated in Brazil during the 2014 outbreak. Additionally, within the Asian and Caribbean genotype, it distinguished between the lineages circulating before the main inland American outbreak. The Asian sequences from 2007 and 2012 were correctly separated from the 2013 sequences, which are hypothesized to carry the adaptations responsible for the American outbreak. The sequences of American isolates also formed a distinct clade, including the sequence isolated in St. Martin at the beginning of the outbreak (Figure 5a).
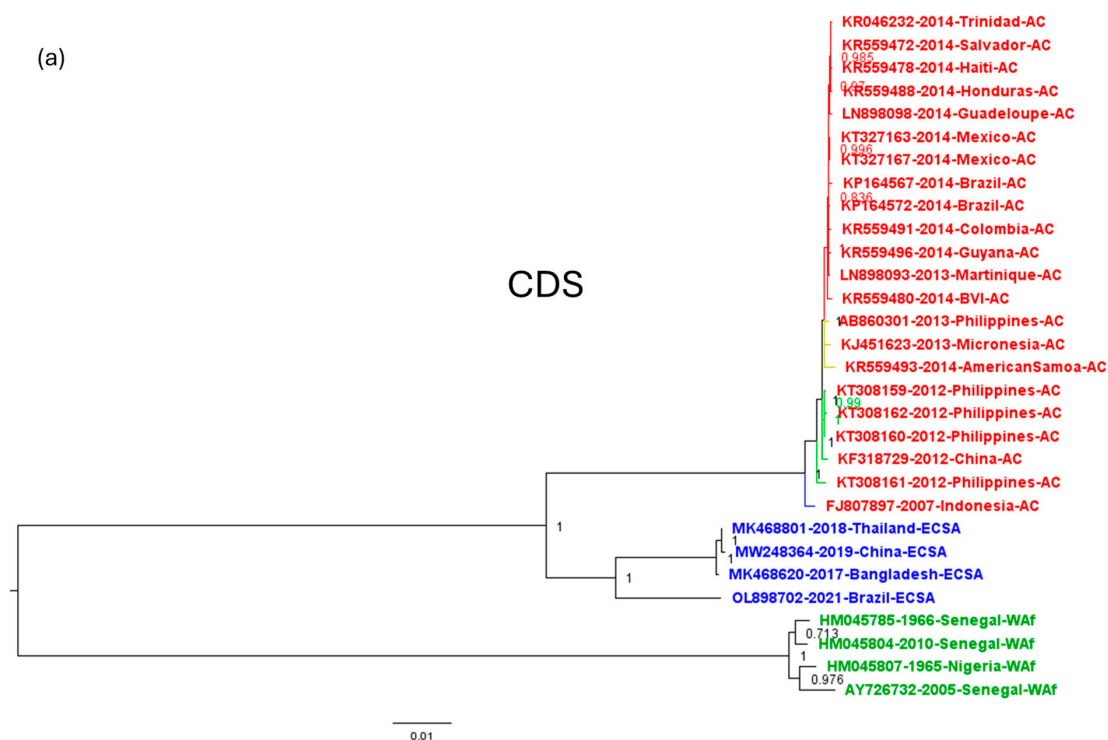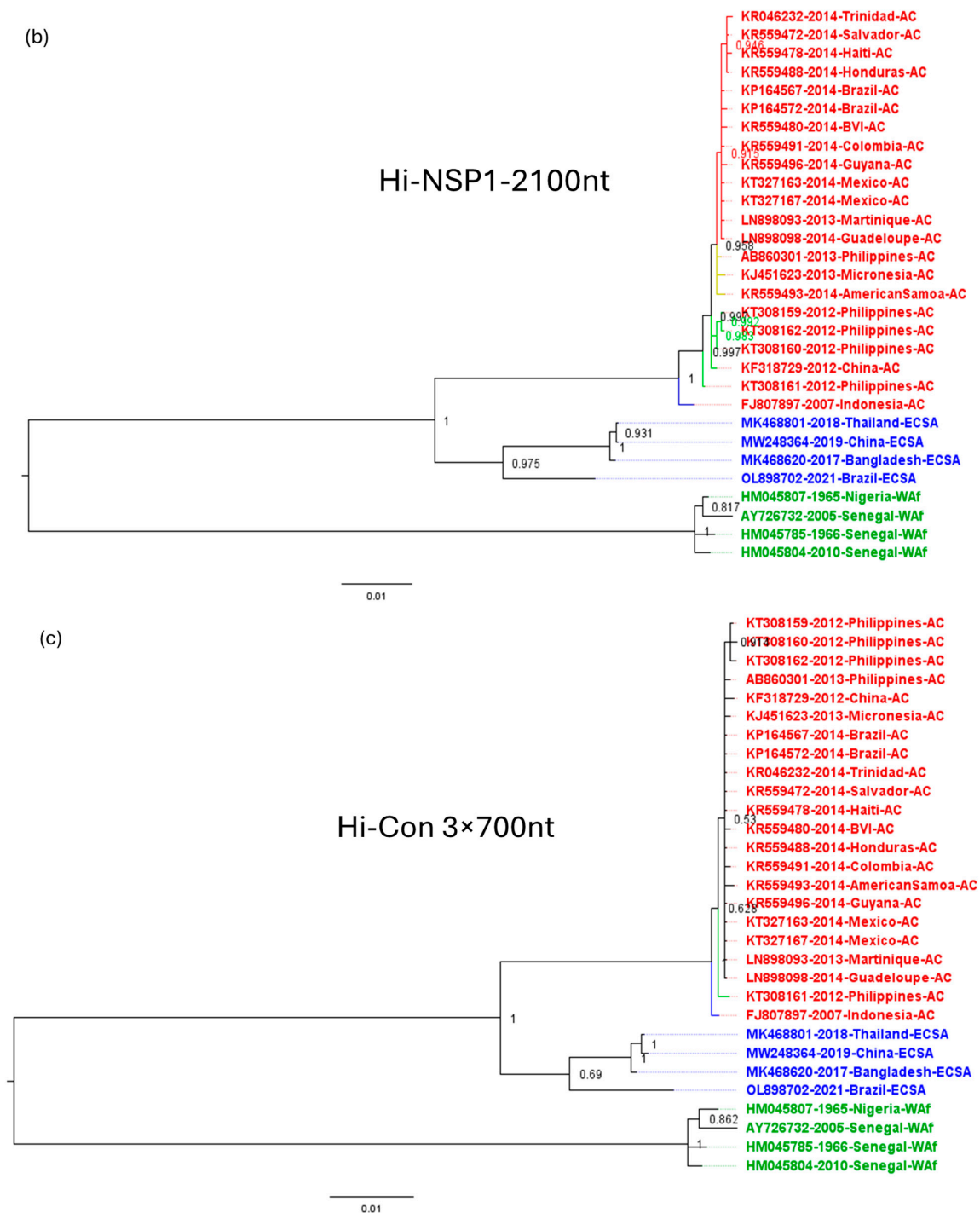

Figure 5. *Cont.*

Figure 5. Comparison of the accuracy of CDS trees and those constructed with the Hi-NSP1 region and the concatenated highly variable regions. The trees constructed with the different regions (Hi- and Low-variables) of 2100 nt length were compared with those constructed with their respective CDS trees to assess the congruence between the evolutionary relationships inferred from these distinct genotypes and geographic regions. Only the (**b**) Hi-NSP1 and (**c**) the three highly variable regions were able to adequately infer the evolutionary history of the American outbreak sequences obtained with the (**a**) CDS. The labels are color-coded to represent the three CHIKV genotypes: West Africa (Blue), ECSA (Green), and Asian and Caribbean (Red). Tree branches are colored to indicate lineages within the Asian and Caribbean genotype: Asian lineage from 2007 (Blue), Asian lineage prior to the 2012 outbreak (Green), Asian lineage with the adaptive mutations for dispersal in *Ae. Aegypti* (Yellow), and Caribbean lineage (Red).

The ability to differentiate the viral lineages of the different regions evaluated (all of 2100 nt length) varied. While all regions successfully distinguished the three major genotypes (Asian and Caribbean, ECSA, and West Africa), their discriminatory power varied when resolving finer-scale relationships within genotypes. Notably, Hi-NSP1 and the concatenated highly variable regions provided sufficient phylogenetic signal to differentiate the lineages within the Asian and Caribbean genotype (2007, 2012, 2013 Asian sequences, and American sequences) (Figure 5b,c). The Low-NSP1 region failed to differentiate the 2012 Asian sequences from the 2013 sequences preceding the American outbreak. Low-C, Low NSP-2, and concatenated low-variability regions grouped the pre-outbreak Asian sequences with the outbreak isolates in a single clade. Similarly, the Hi-E2 region could not distinguish the 2007 from the 2012 Asian sequences, and the Hi-NSP3 region did not differentiate the pre-outbreak Asian sequences from the American outbreak isolates (Figure S2).

4. Discussion

Chikungunya virus exhibits a remarkably rapid evolutionary rate, with mutations accumulating at a pace that closely aligns with its rapid transmission dynamics [39]. This translates to detectable genetic differences between lineages responsible for epidemics emerging within a short period, often on the scale of months [21,40]. Whole-genome sequencing (WGS) offers the most comprehensive view of CHIKV genetic variation, allowing for the identification of even minor mutations that might influence viral characteristics [41]. Although whole-genome sequencing is becoming more accessible, its costs are still a major barrier. Additionally, processing and analyzing the massive datasets requires sophisticated infrastructure, which is often lacking in resource-constrained settings [42]. NGS platforms hold considerable potential for chikungunya genome surveillance programs, enabling detailed characterization of viral strains. However, despite significant advancements in sequencing speed and technology, the cost of establishing and running a dedicated NGS facility [31] remains a substantial challenge for many chikungunya surveillance programs, particularly in low- and middle-income countries [31]. Additionally, individual sequencing runs incur costs depending on the type of analysis. For example, whole-genome sequencing can range from USD 100 to USD 1000 per sample based on the latest pricing trends [42]. Oxford Nanopore Technologies provides a more accessible and affordable form of genome sequencing [43], although with an error rate of around 14% [44]. This high error rate could hinder accurate assembly due to fragmented reads and ambiguous base calls, which could compromise downstream analyses, leading to misinterpretations of the viral sequence, and hinder understanding of the molecular epidemiology and evolutionary patterns of CHIKV [45]. This study proposes the use of highly variable genomic regions as a substitute for CDS. These strategically selected regions can effectively reproduce phylogenies derived from whole genomes and could be used, at lower cost, using PCR methods, for monitoring CHIKV evolution through ongoing genomic surveillance efforts. The PCR conditions and primers proposed for NSP1, NSP3, and E2 partial gene amplification are presented in Table S3.

The selection of appropriate genomic loci remains a crucial step in constructing robust phylogenetic trees, particularly when aiming at resolving relationships at different evolutionary scales. Evolutionary rate is widely recognized as a critical factor influencing the usefulness of any chosen locus [46–48]. The standard approach of studying molecular diversification using only a single structural gene of the virus [49,50] has been applied to CHIKV. However, the limited phylogenetic signal generated incongruent phylogenetic hypotheses lacking enough support [51]. For instance, a maximum likelihood analysis of the E1 gene failed to resolve the ECSA lineage as monophyletic [52], as demonstrated in whole-genome analyses and our findings. Genes evolving at a slow pace, compared to those of divergent lineages represented in a phylogenetic tree, may accumulate too few mutations to accurately resolve their evolutionary history.

The analysis of untranslated regions (UTRs) could greatly benefit the study of short-term adaptations in RNA viruses. These regions evolve rapidly, providing valuable insights [53]. However, this approach requires CDS sequencing, and presents significant challenges, because complex structures, low GC content, and repetitive elements can lead to errors during sequencing [54]. In addition, excessive rapid evolution can lead to a phenomenon known as saturation, where independent mutations obscure the true evolutionary history at a particular site [55]. This effect reduces the discriminatory power of the locus, hindering the ability to distinguish between closely related lineages [56]. Our study revealed different rates of nucleotide variation within the CHIKV genome, characterized by both highly variable and highly conserved regions. Notably, regions in genes encoding the NSP1, NSP3, and E2 proteins exhibited the highest degree of variability, potentially reflecting their roles in functions less constrained by selective pressure. Conversely, genome regions encoding NSP1, NSP2, and C proteins displayed the lowest variability, suggesting they might be under strong purifying selection to maintain essential viral functions.

Longer branches of phylogenetic trees indicate a greater number of mutations accumulated along the biological story of each lineage [48]. Our phylogenetic trees constructed using highly variable regions displayed longer branches compared to those generated with less variable regions [57]. However, the length of the genomic sequences employed in the analysis was important for the estimation of the accumulated variations, and trees built with short, highly variable regions did not accurately reflect distances among lineages. In this case, all evaluated regions, except for the highly variable NSP3 region (Hi-NSP3), resulted in shorter branches compared to the reference tree. The NS1 gene harbors a region with the highest combined rate of synonymous and non-synonymous mutations. However, the NSP3 gene exhibits the highest overall number of mutations, including a substantial proportion of both synonymous and non-synonymous changes. This pattern suggests a high degree of positive selection acting on the NSP3 region. Positive selection describes a process where mutations that confer an evolutionary advantage become more frequent over time within a population. In the context of viruses, these advantageous mutations might enhance essential functions for viral replication, transmission, or immune evasion [58]. For RNA viruses like CHIKV, with limited coding capacity, mutations can have a significant effect on protein structure and function [27]. Regions under positive selection, like the NSP3 gene in this case, might experience a higher proportion of non-synonymous mutations. These functionally relevant mutations likely contribute to an accelerated evolutionary rate in the NSP3 gene relative to other regions. This phenomenon, driven by positive selection, can lead to inflated mutation rates in specific regions. Consequently, branch lengths in phylogenetic trees may not accurately reflect the true evolutionary distances, leading to an overestimation of genetic distances between viral isolates. Our study suggests that using longer genomic regions or combining regions with varying levels of variability can mitigate this issue. The observed improvement with longer regions might be due to the mosaic nature of the CHIKV genome, where highly variable and conserved regions are interspersed. By increasing the analyzed region size, we captured a broader spectrum of the variability within the genome. This broader sampling provides a more accurate representation of the overall evolutionary rate across the CDS.

In contrast, low-variability regions exhibited a significantly diminished capacity to reproduce the CSD phylogeny branch length. Short branches potentially indicate a slower evolutionary rate within these regions, resulting in a limited number of accumulated mutations [59]. In the context of phylogenetic analysis, such limited variation presents a challenge. These regions might lack the necessary informative sites to accurately resolve relationships between lineages [60]. The limitations of low-variability regions in reconstructing the true evolutionary history (branch lengths and divergence times; BLD) likely influenced the performance in other evaluated parameters. This became evident in two key observations: First, shorter regions, which inherently captured less sequence variation, and yielded low K-tree score values. Low scores in this case suggest that the short, less variable regions lack sufficient informative sites to accurately reconstruct branch lengths,

potentially misleading estimates of evolutionary distances between lineages. Second, the mean posterior probability values, which measure the level of tree confidence, were consistently low across all region lengths. This determines that low-variability regions provide weak phylogenetic signals, making it difficult to confidently resolve relationships between lineages and unreliable tree inferences [38]. In contrast, highly variable regions contributed to more robust inferences across all evaluated parameters, showcasing their effectiveness in reconstructing CHIKV phylogenies. The concatenated highly variable regions achieved a level of accuracy statistically equivalent to the reference CDS phylogeny. This performance can be attributed to the lower susceptibility to substitution saturation of highly variable regions. Each gene region provides a unique evolutionary perspective, and combining this information increased the overall number of informative sites for analysis [61].

The capacity to distinguish between closely related CHIKV strains is vital for genomic surveillance. Genome surveillance programs provide comprehensive insights into the genetic makeup and evolutionary dynamics of pathogens [62]. This approach allows for a better understanding of disease spreading, aiding in the development of targeted prevention and control strategies [63,64]. Concatenated highly variable regions demonstrated remarkable performance in reconstructing the CSD phylogeny. Using a region size of 2100 nucleotides, the K-tree scores fell below 0.01, with no statistically significant differences to the reference CSD phylogeny. Concatenated highly variable regions of this length offered a robust approach for phylogenetic analysis.

To illustrate the potential of this approach, we investigated the applicability of 2100 nt regions for genomic analysis during the last outbreak in the Americas. For this, sequences from the three different CHIKV genotypes were chosen, including the Asian genotype and the Caribbean lineage, since distinguishing between closely related sequences during the outbreak was the main challenge in this analysis. The lineage responsible for the American outbreak is hypothesized to have undergone adaptive mutations in Asia, enabling efficient spread through *Ae. aegypti* [29], years before the outbreak itself. However, detecting these subtle mutations is vital for early intervention. All analyzed 2100-nucleotide regions successfully differentiated between the major genotypes (Asian/Caribbean, ECSA, West Africa), probably because of their substantial genetic distances. In addition, the Hi-NSP1 region and concatenated highly variable regions provided sufficient signal to differentiate lineages within the Asian/Caribbean genotypes. These regions, by distinguishing pre-outbreak Asian sequences from those isolated during the outbreak, captured more informative evolutionary changes compared to the others. By distinguishing pre-outbreak Asian sequences from those isolated during the outbreak, these regions offered valuable insights into the evolutionary trajectory of the virus and the emergence of potentially more transmissible variants.

Our study revealed that 2100 concatenated highly variable regions could be used to recapitulate CDS phylogenetic trees, but those shorter than 1500 nucleotides lacked sufficient resolution to decipher the intricate relationships between viral strains within the American outbreak. Focusing on informative regions with adequate length, like Hi-NSP1 or concatenated highly variable regions, can provide a balance between feasibility and detecting subtle evolutionary changes crucial for monitoring transmission patterns, identifying variants, and informing public health interventions. Future research could explore additional targeted regions and optimal combinations for concatenated approaches to maximize resolution while maintaining feasibility.

regions of CHIKV for outbreak tracking through genomic surveillance; Table S3: PCR conditions and primers proposed for NSP1, NSP3 and E2 partial gene amplification.

References

1. Chikungunya: Case Definitions for Acute, Atypical and Chronic Cases. Conclusions of an Expert Consultation, Managua, Nicaragua, 20–21 May 2015. *Wkly. Epidemiol. Rec.* **2015**, *90*, 410–414.
2. Powers, A.M.; Brault, A.C.; Shirako, Y.; Strauss, E.G.; Kang, W.; Strauss, J.H.; Weaver, S.C. Evolutionary Relationships and Systematics of the Alphaviruses. *J. Virol.* **2001**, *75*, 10118–10131. [CrossRef] [PubMed]
3. Rupp, J.C.; Sokoloski, K.J.; Gebhart, N.N.; Hardy, R.W. Alphavirus RNA Synthesis and Non-Structural Protein Functions. *J. Gen. Virol.* **2015**, *96*, 2483–2500. [CrossRef]
4. Ross, R.W. The Newala Epidemic. III. The Virus: Isolation, Pathogenic Properties and Relationship to the Epidemic. *J. Hyg.* **1956**, *54*, 177–191. [CrossRef] [PubMed]
5. Volk, S.M.; Chen, R.; Tsetsarkin, K.A.; Adams, A.P.; Garcia, T.I.; Sall, A.A.; Nasar, F.; Schuh, A.J.; Holmes, E.C.; Higgs, S.; et al. Genome-Scale Phylogenetic Analyses of Chikungunya Virus Reveal Independent Emergences of Recent Epidemics and Various Evolutionary Rates. *J. Virol.* **2010**, *84*, 6497–6504. [CrossRef] [PubMed]
6. Powers, A.M.; Brault, A.C.; Tesh, R.B.; Weaver, S.C. Re-Emergence of Chikungunya and o'nyong-Nyong Viruses: Evidence for Distinct Geographical Lineages and Distant Evolutionary Relationships. *J. Gen. Virol.* **2000**, *81*, 471–479. [CrossRef] [PubMed]
7. Arankalle, V.A.; Shrivastava, S.; Cherian, S.; Gunjikar, R.S.; Walimbe, A.M.; Jadhav, S.M.; Sudeep, A.B.; Mishra, A.C. Genetic Divergence of Chikungunya Viruses in India (1963–2006) with Special Reference to the 2005–2006 Explosive Epidemic. *J. Gen. Virol.* **2007**, *88*, 1967–1976. [CrossRef]
8. Rezza, G.; Nicoletti, L.; Angelini, R.; Romi, R.; Finarelli, A.C.; Panning, M.; Cordioli, P.; Fortuna, C.; Boros, S.; Magurano, F.; et al. Infection with Chikungunya Virus in Italy: An Outbreak in a Temperate Region. *Lancet* **2007**, *370*, 1840–1846. [CrossRef]
9. Grandadam, M.; Caro, V.; Plumet, S.; Thiberge, J.-M.; Souarès, Y.; Failloux, A.-B.; Tolou, H.J.; Budelot, M.; Cosserat, D.; Leparc-Goffart, I.; et al. Chikungunya Virus, Southeastern France. *Emerg. Infect. Dis.* **2011**, *17*, 910–913. [CrossRef]
10. Carey, D.E. Chikungunya and Dengue: A Case of Mistaken Identity? *J. Hist. Med. Allied Sci.* **1971**, *26*, 243–262. [CrossRef]
11. Mehdi, Z.; Shahbaz, H.; Owais, A.; Hasan, S.U.; Nasr, I.; Jahangir, A.; Zubair, N.; Abdul Khaliq, S.A.; Khalid, M.; Shahbaz, S.; et al. Frequency, Awareness, and Symptoms of Chikungunya Among Patients in a Tertiary Care Hospital of Karachi: A Cross-Sectional Study. *Cureus* **2019**, *11*, e4054. [CrossRef] [PubMed]
12. Leparc-Goffart, I.; Nougairede, A.; Cassadou, S.; Prat, C.; de Lamballerie, X. Chikungunya in the Americas. *Lancet* **2014**, *383*, 514. [CrossRef] [PubMed]
13. Paixão, E.S.; Teixeira, M.G.; Rodrigues, L.C. Zika, Chikungunya and Dengue: The Causes and Threats of New and Re-Emerging Arboviral Diseases. *BMJ Glob. Health* **2018**, *3*, e000530. [CrossRef] [PubMed]
14. Yactayo, S.; Staples, J.E.; Millot, V.; Cibrelus, L.; Ramon-Pardo, P. Epidemiology of Chikungunya in the Americas. *J. Infect. Dis.* **2016**, *214*, S441–S445. [CrossRef] [PubMed]
15. Vazeille, M.; Moutailler, S.; Coudrier, D.; Rousseaux, C.; Khun, H.; Huerre, M.; Thiria, J.; Dehecq, J.-S.; Fontenille, D.; Schuffenecker, I.; et al. Two Chikungunya Isolates from the Outbreak of La Reunion (Indian Ocean) Exhibit Different Patterns of Infection in the Mosquito, Aedes Albopictus. *PLoS ONE* **2007**, *2*, e1168. [CrossRef] [PubMed]

16. Tsetsarkin, K.A.; Vanlandingham, D.L.; McGee, C.E.; Higgs, S. A Single Mutation in Chikungunya Virus Affects Vector Specificity and Epidemic Potential. *PLoS Pathog.* **2007**, *3*, e201. [CrossRef] [PubMed]

17. Weaver, S.C. Arrival of Chikungunya Virus in the New World: Prospects for Spread and Impact on Public Health. *PLoS Negl. Trop. Dis.* **2014**, *8*, e2921. [CrossRef]

18. Stapleford, K.A.; Moratorio, G.; Henningsson, R.; Chen, R.; Matheus, S.; Enfissi, A.; Weissglas-Volkov, D.; Isakov, O.; Blanc, H.; Mounce, B.C.; et al. Whole-Genome Sequencing Analysis from the Chikungunya Virus Caribbean Outbreak Reveals Novel Evolutionary Genomic Elements. *PLoS Negl. Trop. Dis.* **2016**, *10*, e0004402. [CrossRef]

19. Lanciotti, R.S.; Valadere, A.M. Transcontinental Movement of Asian Genotype Chikungunya Virus. *Emerg. Infect. Dis.* **2014**, *20*, 1400–1402. [CrossRef] [PubMed]

20. Sahadeo, N.; Mohammed, H.; Allicock, O.M.; Auguste, A.J.; Widen, S.G.; Badal, K.; Pulchan, K.; Foster, J.E.; Weaver, S.C.; Carrington, C.V.F. Molecular Characterisation of Chikungunya Virus Infections in Trinidad and Comparison of Clinical and Laboratory Features with Dengue and Other Acute Febrile Cases. *PLoS Negl. Trop. Dis.* **2015**, *9*, e0004199. [CrossRef] [PubMed]

21. Rodríguez-Aguilar, E.D.; Martínez-Barnetche, J.; González-Bonilla, C.R.; Tellez-Sosa, J.M.; Argotte-Ramos, R.; Rodríguez, M.H. Genetic Diversity and Spatiotemporal Dynamics of Chikungunya Infections in Mexico during the Outbreak of 2014–2016. *Viruses* **2021**, *14*, 70. [CrossRef]

22. Díaz, Y.; Carrera, J.-P.; Cerezo, L.; Arauz, D.; Guerra, I.; Cisneros, J.; Armién, B.; Botello, A.M.; Araúz, A.B.; Gonzalez, V.; et al. Chikungunya Virus Infection: First Detection of Imported and Autochthonous Cases in Panama. *Am. J. Trop. Med. Hyg.* **2015**, *92*, 482–485. [CrossRef]

23. Mattar, S.; Miranda, J.; Pinzon, H.; Tique, V.; Bolanos, A.; Aponte, J.; Arrieta, G.; Gonzalez, M.; Barrios, K.; Contreras, H.; et al. Outbreak of Chikungunya Virus in the North Caribbean Area of Colombia: Clinical Presentation and Phylogenetic Analysis. *J. Infect. Dev. Ctries.* **2015**, *9*, 1126–1132. [CrossRef] [PubMed]

24. Nunes, M.R.T.; Faria, N.R.; de Vasconcelos, J.M.; Golding, N.; Kraemer, M.U.G.; de Oliveira, L.F.; Azevedo, R.d.S.d.S.; da Silva, D.E.A.; da Silva, E.V.P.; da Silva, S.P.; et al. Emergence and Potential for Spread of Chikungunya Virus in Brazil. *BMC Med.* **2015**, *13*, 102. [CrossRef]

25. Langsjoen, R.M.; Haller, S.L.; Roy, C.J.; Vinet-Oliphant, H.; Bergren, N.A.; Erasmus, J.H.; Livengood, J.A.; Powell, T.D.; Weaver, S.C.; Rossi, S.L. Chikungunya Virus Strains Show Lineage-Specific Variations in Virulence and Cross-Protective Ability in Murine and Nonhuman Primate Models. *mBio* **2018**, *9*, e02449-17. [CrossRef]

26. Sim, S.; Hibberd, M.L. Genomic Approaches for Understanding Dengue: Insights from the Virus, Vector, and Host. *Genome Biol.* **2016**, *17*, 38. [CrossRef] [PubMed]

27. Drake, J.W.; Holland, J.J. Mutation Rates among RNA Viruses. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13910–13913. [CrossRef] [PubMed]

28. Robishaw, J.D.; Alter, S.M.; Solano, J.J.; Shih, R.D.; DeMets, D.L.; Maki, D.G.; Hennekens, C.H. Genomic Surveillance to Combat COVID-19: Challenges and Opportunities. *Lancet Microbe* **2021**, *2*, e481–e484. [CrossRef]

29. Agarwal, A.; Sharma, A.K.; Sukumaran, D.; Parida, M.; Dash, P.K. Two Novel Epistatic Mutations (E1:K211E and E2:V264A) in Structural Proteins of Chikungunya Virus Enhance Fitness in Aedes Aegypti. *Virology* **2016**, *497*, 59–68. [CrossRef] [PubMed]

30. Sayers, E.W.; Bolton, E.E.; Brister, J.R.; Canese, K.; Chan, J.; Comeau, D.C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; et al. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2022**, *50*, D20–D26. [CrossRef] [PubMed]

31. Schwarze, K.; Buchanan, J.; Taylor, J.C.; Wordsworth, S. Are Whole-Exome and Whole-Genome Sequencing Approaches Cost-Effective? A Systematic Review of the Literature. *Genet. Med.* **2018**, *20*, 1122–1130. [CrossRef] [PubMed]

32. Rodríguez-Aguilar, E.D.; Martínez-Barnetche, J.; Rodríguez, M.H. Three Highly Variable Genome Regions of the Four Dengue Virus Serotypes Can Accurately Recapitulate the CDS Phylogeny. *MethodsX* **2022**, *9*, 101859. [CrossRef] [PubMed]

33. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [CrossRef] [PubMed]

34. Posada, D.; Crandall, K.A. MODELTEST: Testing the Model of DNA Substitution. *Bioinformatics* **1998**, *14*, 817–818. [CrossRef] [PubMed]

35. Huelsenbeck, J.P.; Ronquist, F. MRBAYES: Bayesian Inference of Phylogenetic Trees. *Bioinformatics* **2001**, *17*, 754–755. [CrossRef] [PubMed]

36. Kuhner, M.K.; Felsenstein, J. A Simulation Comparison of Phylogeny Algorithms under Equal and Unequal Evolutionary Rates. *Mol. Biol. Evol.* **1994**, *11*, 459–468. [CrossRef] [PubMed]

37. Soria-Carrasco, V.; Talavera, G.; Igea, J.; Castresana, J. The K Tree Score: Quantification of Differences in the Relative Branch Length and Topology of Phylogenetic Trees. *Bioinformatics* **2007**, *23*, 2954–2956. [CrossRef]

38. Alfaro, M.E.; Zoller, S.; Lutzoni, F. Bayes or Bootstrap? A Simulation Study Comparing the Performance of Bayesian Markov Chain Monte Carlo Sampling and Bootstrapping in Assessing Phylogenetic Confidence. *Mol. Biol. Evol.* **2003**, *20*, 255–266. [CrossRef] [PubMed]

39. Schuffenecker, I.; Iteman, I.; Michault, A.; Murri, S.; Frangeul, L.; Vaney, M.-C.; Lavenir, R.; Pardigon, N.; Reynes, J.-M.; Pettinelli, F.; et al. Genome Microevolution of Chikungunya Viruses Causing the Indian Ocean Outbreak. *PLoS Med.* **2006**, *3*, e263. [CrossRef]

40. Torres-Longoria, B.; Fragoso-Fonseca, D.E.; Núñez-León, A.; de la Luz Torres, M.; Vázquez-Pichardo, M.; Escobar-Escamilla, N.; Wong-Arámbula, C.; Ramírez-González, J.E.; Méndez-Tenorio, A.; Castro-Mussot, M.E.; et al. Epidemiological Surveillance of Chikungunya Fever in Mexico since Its Introduction in 2014–2016 and Identification of Circulating Genotypes. *Mol. Biol. Rep.* **2021**, *48*, 1967–1975. [CrossRef] [PubMed]

41. Popovich, K.J.; Snitkin, E.S. Whole Genome Sequencing-Implications for Infection Prevention and Outbreak Investigations. *Curr. Infect. Dis. Rep.* **2017**, *19*, 15. [CrossRef]

42. Mboowa, G.; Mwesigwa, S.; Kateete, D.; Wayengera, M.; Nasinghe, E.; Katagirya, E.; Katabazi, A.F.; Kigozi, E.; Kirimunda, S.; Kamulegeya, R.; et al. Whole-Genome Sequencing of SARS-CoV-2 in Uganda: Implementation of the Low-Cost ARTIC Protocol in Resource-Limited Settings. *F1000Res* **2021**, *10*, 598. [CrossRef] [PubMed]

43. Fox, E.J.; Reid-Bayliss, K.S.; Emond, M.J.; Loeb, L.A. Accuracy of Next Generation Sequencing Platforms. *Next Gener. Seq. Appl.* **2014**, *1*, 1000106. [CrossRef] [PubMed]

44. Wang, M.; Fu, A.; Hu, B.; Tong, Y.; Liu, R.; Liu, Z.; Gu, J.; Xiang, B.; Liu, J.; Jiang, W.; et al. Nanopore Targeted Sequencing for the Accurate and Comprehensive Detection of SARS-CoV-2 and Other Respiratory Viruses. *Small* **2020**, *16*, e2002169. [CrossRef] [PubMed]

45. Fritsch, H.; Giovanetti, M.; Xavier, J.; Adelino, T.E.R.; Fonseca, V.; de Jesus, J.G.; de Jesus, R.; Freitas, C.; Peterka, C.R.L.; Campelo de Albuquerque, C.F.; et al. Retrospective Genomic Surveillance of Chikungunya Transmission in Minas Gerais State, Southeast Brazil. *Microbiol. Spectr.* **2022**, *10*, e0128522. [CrossRef] [PubMed]

46. Betancur-R, R.; Naylor, G.J.P.; Ortí, G. Conserved Genes, Sampling Error, and Phylogenomic Inference. *Syst. Biol.* **2014**, *63*, 257–262. [CrossRef]

47. Doyle, V.P.; Young, R.E.; Naylor, G.J.P.; Brown, J.M. Can We Identify Genes with Increased Phylogenetic Reliability? *Syst. Biol.* **2015**, *64*, 824–837. [CrossRef]

48. Yang, Z. On the Best Evolutionary Rate for Phylogenetic Analysis. *Syst. Biol.* **1998**, *47*, 125–133. [CrossRef] [PubMed]

49. Díaz-Quiñonez, J.A.; Escobar-Escamilla, N.; Ortíz-Alcántara, J.; Vázquez-Pichardo, M.; de la Luz Torres-Rodríguez, M.; Nuñez-León, A.; Torres-Longoria, B.; López-Martínez, I.; Ruiz-Matus, C.; Kuri-Morales, P.; et al. Identification of Asian Genotype of Chikungunya Virus Isolated in Mexico. *Virus Genes* **2016**, *52*, 127–129. [CrossRef] [PubMed]

50. Laredo-Tiscareño, S.V.; Machain-Williams, C.; Rodríguez-Pérez, M.A.; Garza-Hernandez, J.A.; Doria-Cobos, G.L.; Cetina-Trejo, R.C.; Bacab-Cab, L.A.; Tangudu, C.S.; Charles, J.; De Luna-Santillana, E.J.; et al. Arbovirus Surveillance near the Mexico–U.S. Border: Isolation and Sequence Analysis of Chikungunya Virus from Patients with Dengue-like Symptoms in Reynosa, Tamaulipas. *Am. J. Trop. Med. Hyg.* **2018**, *99*, 191–194. [CrossRef] [PubMed]

51. Rokas, A.; Williams, B.L.; King, N.; Carroll, S.B. Genome-Scale Approaches to Resolving Incongruence in Molecular Phylogenies. *Nature* **2003**, *425*, 798–804. [CrossRef] [PubMed]

52. Muñoz-Medina, J.E.; Garcia-Knight, M.A.; Sanchez-Flores, A.; Monroy-Muñoz, I.E.; Grande, R.; Esbjörnsson, J.; Santacruz-Tinoco, C.E.; González-Bonilla, C.R. Evolutionary Analysis of the Chikungunya Virus Epidemic in Mexico Reveals Intra-Host Mutational Hotspots in the E1 Protein. *PLoS ONE* **2018**, *13*, e0209292. [CrossRef]

53. Zhang, S.; Yang, C.; Qiu, Y.; Liao, R.; Xuan, Z.; Ren, F.; Dong, Y.; Xie, X.; Han, Y.; Wu, D.; et al. Conserved Untranslated Regions of Multipartite Viruses: Natural Markers of Novel Viral Genomic Components and Tags of Viral Evolution. *Virus Evol.* **2024**, *10*, veae004. [CrossRef] [PubMed]

54. Chen, R.; Wang, E.; Tsetsarkin, K.A.; Weaver, S.C. Chikungunya Virus 3′ Untranslated Region: Adaptation to Mosquitoes and a Population Bottleneck as Major Evolutionary Forces. *PLoS Pathog.* **2013**, *9*, e1003591. [CrossRef]

55. Duchêne, D.A.; Mather, N.; Van Der Wal, C.; Ho, S.Y.W. Excluding Loci with Substitution Saturation Improves Inferences from Phylogenomic Data. *Syst. Biol.* **2022**, *71*, 676–689. [CrossRef] [PubMed]

56. Philippe, H.; Brinkmann, H.; Lavrov, D.V.; Littlewood, D.T.J.; Manuel, M.; Wörheide, G.; Baurain, D. Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. *PLoS Biol.* **2011**, *9*, e1000602. [CrossRef] [PubMed]

57. Binet, M.; Gascuel, O.; Scornavacca, C.; Douzery, E.J.P.; Pardi, F. Fast and Accurate Branch Lengths Estimation for Phylogenomic Trees. *BMC Bioinform.* **2016**, *17*, 23. [CrossRef] [PubMed]

58. Rodpothong, P.; Auewarakul, P. Positive Selection Sites in the Surface Genes of Dengue Virus: Phylogenetic Analysis of the Interserotypic Branches of the Four Serotypes. *Virus Genes* **2012**, *44*, 408–414. [CrossRef]

59. Mindell, D.P.; Thacker, C.E. Rates of Molecular Evolution: Phylogenetic Issues and Applications. *Annu. Rev. Ecol. Syst.* **1996**, *27*, 279–303. [CrossRef]

60. Som, A. Causes, Consequences and Solutions of Phylogenetic Incongruence. *Brief. Bioinform.* **2015**, *16*, 536–548. [CrossRef]

61. Gontcharov, A.A.; Marin, B.; Melkonian, M. Are Combined Analyses Better than Single Gene Phylogenies? A Case Study Using SSU rDNA and rbcL Sequence Comparisons in the Zygnematophyceae (Streptophyta). *Mol. Biol. Evol.* **2004**, *21*, 612–624. [CrossRef] [PubMed]

62. Meredith, L.W.; Aboualy, M.; Ochola, R.; Okwarah, P.; Ozel, M.; Abubakar, A.; Barakat, A. Key Aspects Defining the Development and Implementation of a Regional Genomic Surveillance Strategy for the Eastern Mediterranean Region. *Influenza Other Respir. Viruses* **2023**, *17*, e13205. [CrossRef] [PubMed]

63. Seib, K.L.; Dougan, G.; Rappuoli, R. The Key Role of Genomics in Modern Vaccine and Drug Design for Emerging Infectious Diseases. *PLoS Genet.* **2009**, *5*, e1000612. [CrossRef] [PubMed]

64. Armstrong, G.L.; MacCannell, D.R.; Carleton, H.A.; Neuhaus, E.B.; Bradbury, R.S.; Posey, J.E.; Taylor, J.; Gwinn, M. Pathogen Genomics in Public Health. *N. Engl. J. Med.* **2019**, *381*, 2569–2580. [CrossRef] [PubMed]