MDPI

Article

Updated phylogeny of Chikungunya virus suggests lineage-specific RNA architecture

Adriano de Bernardi Schneider [1,†], Roman Ochsenreiter [3,†], Reilly Hostager [1], Ivo L. Hofacker [3,4], Daniel Janies [2] and Michael T. Wolfinger [3,4,*]

1 AntiViral Research Center, Department of Medicine, University of California San Diego, San Diego, 92103, United States of America; {adeberna,rehostager}@ucsd.edu
2 Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, North Carolina, United States of America; djanies@uncc.edu
3 Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, 1090 Vienna, Austria; {romanoch,ivo}@tbi.univie.ac.at, michael.wolfinger@univie.ac.at
4 Research Group Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Währingerstraße 29, 1090 Vienna, Austria
* Correspondence: michael.wolfinger@univie.ac.at (MTW)
† These authors contributed equally to this work.

Version August 20, 2019 submitted to Viruses

Abstract: Chikungunya virus (CHIKV), a mosquito-borne alphavirus of the family Togaviridae, has recently emerged in the Americas from lineages from two continents, Asia and Africa. Historically, CHIKV circulated as at least four lineages worldwide with both enzootic and epidemic transmission cycles. To understand the recent patterns of emergence and the current status of the CHIKV spread, updated analyses of the viral genetic data and metadata are needed. Here, we performed phylogenetic and comparative genomics screens of CHIKV genomes, taking advantage of the public availability of many recently sequenced isolates. Based on these new data and analyses, we derive a revised phylogeny from nucleotide sequences in coding regions. Using this phylogeny, we uncover the presence of several distinct lineages in Africa that were previously considered a single one. In parallel, we performed thermodynamic modeling of CHIKV untranslated regions (UTRs), which revealed evolutionarily conserved structured and unstructured RNA elements in the 3'UTR. We provide evidence for duplication events in recently emerged American isolates of the Asian CHIKV lineage and propose the existence of a flexible 3'UTR architecture among different CHIKV lineages.

Keywords: RNA virus; evolution; epidemics; phylogeography; secondary structure

1. Introduction

Phylogeographic inquiries can reveal viral transmission patterns and pathogenesis of viral diseases. Dramatic (re)emergences of arthropod-borne viruses (arboviruses), such as Chikungunya (CHIKV), West Nile, Zika, and Dengue have increased in frequency and intensity, thus raising public health concerns [1]. Prolific expansion of arboviruses in countries where the viruses have previously not been recorded has drawn public attention. These trends are considered to be a function of climate change effects on the geographic range in which the vector can complete its life cycle and transmit CHIKV [2,3]. Here, we investigate the RNA architecture of CHIKV untranslated regions (UTRs) that underlies their ability to adapt and transmit in new environments and provide an updated large-scale phylogeny summarizing the geographic spread of CHIKV lineages.

CHIKV is a mosquito-borne virus that belongs to the *Alphavirus* genus, which encompasses mainly arboviruses, and is believed to have a marine origin [4,5]. The presence of CHIKV in sylvatic

mosquitoes, such as *Aedes africanus, Aedes furcifer* and *A. taylori*, and non-human primates in Uganda and Tanzania suggests that the virus originated in Central or East Africa [6,7].

The first isolation of CHIKV was in 1952 in the Newala district of Tanganyika, a British territory that is now part of Tanzania [8,9]. CHIKV has caused multiple outbreaks in all continents except Antarctica, most recently in Europe and Central and South America. In total, CHIKV has been present in over 100 countries [6,10–12].

CHIKV infects multiple human tissues, causing a febrile illness which is typically characterized by myalgia, polyarthralgia, polyarthritis, and rashes [13]. Several vaccine candidates are currently in preclinical or clinical development [14].

The virus had not been associated with life-threatening symptoms before an outbreak in 2005 on La Réunion island off the East African coast [15], which demonstrated a drastic expansion of the disease spectrum. In La Réunion and after, higher morbidity has been associated with CHIKV infection, including neurological issues, such as visual and hearing loss, paralysis, Guillain Barré Syndrome (GBS) and renal complications [13,16,17]. Recent mutations that change protein structure have been attributed to increased ability to establish infection (infectivity) and outbreaks as the virus becomes better adapted to vector species (e.g., *A. albopictus*, the only suitable agent for transmission in La Réunion island) [18]. Species of the *Aedes* genus, in particular *A. albopictus, A. aegypti, and A. polynesiensis*, are carriers of CHIKV, capable of transmission to humans [19].

1.1. Chikungunya lineages

CHIKV is a virus for which four major lineages have been proposed: Asian Urban (AUL), Indian Ocean (IOL), East, Central and South African (ECSA), and West African (WA) [20,21]. Despite these notions, the origin and the history of how CHIKV spreads inside of Africa remains elusive. This gap in our understanding is due to undersampling and confounding factors such as co-circulating Dengue and Yellow Fever viruses. The co-circulation of viruses diminishes the value of serology test which are not species specific. While the AUL has historically been constrained in Southeast Asia, there have been recent outbreaks in the Americas [18]. The IOL, circulating in India and surrounding regions, has been mainly restricted to cause outbreaks in South and Southeastern Asia, although it has caused recent outbreaks in Europe. The WA lineage is entirely isolated.

The ECSA lineage has a unique history as it gave rise to the IOL as well as has also been identified as the origin of recent outbreaks in Brazil and Haiti [12,22,23]. The vast territory in Africa, the lack of sampling, and increasing diversity of ECSA lineage fosters doubt on whether ECSA is a single lineage.

1.2. Conserved RNA structure as evolutionary trait

The CHIKV genome consists of a single-stranded, (+)-sense RNA of approximately 11.8kb, encoding nine proteins which are divided into structural and non-structural proteins [24]. The genome is organized as two open reading frames (ORFs), flanked by structured UTRs [25].

Populations of RNA viruses exhibit large genetic variability. This variation allows populations to have variants that are suited to changing conditions and immune escape. Viral variation results mainly from two mechanisms, mutation and recombination. Mutation occurs during replication by mis-introduction of nucleotides by error-prone RNA-dependent RNA polymerases that lack proof-reading capacity. Recombination occurs when co-circulating lineages co-infect a host and mixed segments are incorporated into the descendent viruses [26].

The role of functional RNA elements in pathogenesis is well established in flaviviruses [27,28], where they are typically conserved evolutionarily in UTRs [29–31]. In CHIKV, patterns of coupled historic recombination and mutation events are encoded in the 3′UTRs of different lineages [32]. The resulting variants have highly divergent 3′UTRs that vary in length. Characteristics of these 3′UTR variants are variable number of direct repeats (DRs) at the sequence level [2,33]. Distinctive CHIKV lineages also differ in their 3′UTR architecture, in particular in the copy number and arrangement of sequence repeats [34].

Evolutionary conservation in RNA viruses is typically maintained in secondary structure and manifested in rich base-pair conserving covariation patterns [35]. Relatively little is known about structured RNAs in alphaviruses, both in coding and in non-coding regions. Coding regions evolve under strong selection pressure to maintain protein function. Nevertheless, it has become apparent that structure within coding regions can also have biological function, as seen in the HIV RRE element [36]. In the context of alphaviruses, Kutchko *et al.* [37] performed RNA structure probing experiments, focusing on Sindbis virus (SINV) and Venezuelan equine encephalitis virus (VEEV), but could not find evidence for broad conservation of functional RNA structures.

The knowledge of structured RNAs in alphavirus UTRs, and CHIKV in particular, is limited. This limitation results from the fact that earlier studies focus on lineage-specific sequence repeats and could not reliably establish a relationship between all observed repeat patterns and RNA secondary structure [32,34].

A large number of genomic sequences and metadata have been deposited recently to public repositories. These data have raised the possibility of revising the lineages of CHIKV as a whole, reevaluating the historical global spread of the major epidemic clades and evaluating the RNA structural patterns among different clades. We also set out to structurally characterize the UTRs of all known CHIKV lineages, and found evidence for lineage-specific structured and unstructured repeat elements. We hypothesize such RNA elements to be involved in viral replication, vector adaptation and specificity, host factor binding, and pathogenicity.

2. Materials and Methods

2.1. Taxon sampling

We obtained viral genome data from the National Center for Biotechnology Information (NCBI) `Genbank` database (https://www.ncbi.nlm.nih.gov/genbank/). We downloaded 598 whole genome sequences of CHIKV. We compiled all geographic metadata available in these genome records to create data sets for the analyses described below. We removed eight sequences from the data set due to missing geographic location metadata or designation as a vaccine sequence. The metadata associated with geographic location was associated with the United Nations geoscheme for consistency. Therefore, "Central Africa" is labeled "Middle Africa".

2.2. Phylogenetic tree search

Two complete ORFs of CHIKV were extracted from the complete viral sequences for the phylogenetic analysis. Multiple sequence alignments of the individual ORFs were computed using `MAFFT` [38] under default settings and visualized in `AliView` [39]. In order to examine the possibility of recombinant sequences, we used RDP and methods therein (RDP, GENECONV, Bootscan, MaxChi, Chimaera, SiScan and PhylPro)[40].

We conducted phylogenetic tree searches with partitioning to separate the two ORFs. We used tree search method under the optimality criterion of maximum likelihood with substitution model testing as implemented in `IQ-Tree` [41].

We executed a 1000-replicate ultrafast bootstrap analysis of clade frequencies [42] and a SH-aLRT measure of support [43]. The ultrafast bootstrap analysis is a variation of the traditional bootstrap using heuristics and constraints to speed the search for optimal tree topologies. The SH-aLRT is an approximation of the likelihood ratio, a direct measure of how much the evidence supports the hypothesis.

According to recently published phylogenies of CHIKV [20,22,44], viruses from the Western African clade form a sister group to all other CHIKV clades. Given these results and the absence of co-circulating Western African strains with any other CHIKV lineage, we selected this clade as the outgroup for the phylogeny of CHIKV. de Bernardi Schneider [44] observed Western Africa as a monophyletic clade sister to all other clades by utilizing an external outgroup (O'nyong nyong virus),

thus the selection of this group should not yield an inherent bias on the tree topology at the same time that it results in a cleaner molecular sequence alignment given the outgroup sequences to belong to the same virus species as the ingroup. We performed the same procedure for a subset containing the 111 complete 3'UTR sequences from the 590 whole genome sequence data set.

The final trees with metadata annotation were rendered using custom Python scripts from the `ETE3` framework [45].

2.3. Global spread of virus

We extracted the list of taxa of the three major epidemic clades, Asian Urban Americas (AUL-Am), ECSA-IOL, and ECSA-Middle African South American (ECSA-MASA), from the phylogenetic tree generated from all CHIKV sequences. We created a csv table including all locations from the dataset with corresponding latitude and longitude extracted from getlatlong.net. We performed individual alignments for each epidemic clade on concatenated ORFs and ran these on an R script pipeline (described here). Our script performs five steps:

1. Tree search under the neighbor-joining algorithm with 100 bootstrap replicates [46,47];
2. Collapsing of branches with a bootstrap value under 80% [48];
3. Replacing tree tip names with metadata information (i.e. geographic location);
4. Parsimony ancestry reconstruction using the *asr_max_parsimony* function from Castor R Package [49,50] to build an edge list (i.e. origin to destination geographic state transitions);
5. Render the geographic state transitions reconstructed on the previous step on a map using geographic coordinates for each location [51,52].

In order to evaluate the historical spread of CHIKV using a parsimonious ancestral reconstruction, the tree must be rooted to infer event directionality. Thus, we selected a root for each clade based on the complete CHIKV phylogeny. For each individual clade, the sister taxon to all other taxa was selected. When a clade had multiple sister taxa, we randomly selected one as the root. We selected HM045788.1 (India) as the outgroup for AUL-Am, KF283986.1 (Comoros) for ECSA-IOL and KX262996.1 (Cameroon) for ECSA-MASA.

This script is an adaptation from `StrainHub` (https://github.com/abschneider/StrainHub) [49], a web-based application to reconstruct pathogen transmission networks, to visualize the spread of the individual lineages/epidemic clades. The scripts and data sets along with a detailed explanation of the functions and packages used for the analyses described here are available online at https://github.com/abschneider/paper-chikungunya-phylogeny.

2.4. Conserved RNA structures in CHIKV 3'UTR

Evolutionary conservation of functional RNAs is typically achieved by nature at the level of secondary structures. Comparative genomics approaches aim at detecting these conserved, structured RNAs as patterns of covariation in genome sequence alignments, typically constructed from phylogenetically narrow subgroups. To this end, Covariance Models (CMs) are statistical models of RNA structure formation that extend traditional Hidden-Markov-Models (HMMs) by simultaneously representing sequence and secondary structure information [53]. CMs provide a straightforward approach to find homologous RNAs in large RNA sequence databases.

All CHIKV 3'UTRs were processed and curated from full genomes by custom Perl scripts based on the `ViennaNGS` toolbox [54]. Structural RNA alignments were computed with `mlocarna` [55], and consensus structures were generated with `RNAalifold` [56]. Locally stable secondary structures were computed using `RNALalifold` from the `ViennaRNA` package [57]. Homologies of both sequences and secondary structures were inferred by HMMs and CMs as implemented in the `Infernal` package [58] as outlined previously [35]. All secondary structure plots were rendered using the `RNAplot` utility [57].

3. Results

3.1. Phylogeny of Chikungunya indicates multiple lineages in Africa

No significant recombination events (P-value of < 0.05) were detected in the aligned data sets, therefore we kept all 590 sequences in the data sets. The phylogenetic tree search performed with `IQ-Tree` resulted in a best-scoring heuristic maximum likelihood tree for CHIKV with 590 genomic sequences (Figure 1). Deep ancestral branches that define the lineages had high (>90) SH-aLRT support/ultrafast bootstrap values. More recent branches had lower SH-aLRT support/ultrafast bootstrap values. The phylogenetic hypothesis for CHIKV with branch values is available as supplementary dataset SD1 at https://github.com/abschneider/paper-chikungunya-phylogeny

The four viral strains grouped as expected according to reports in literature [20,22], however the increased number of taxa in the tree allowed us to better investigate relationships of clades and infer source-sink events (Figure 2). In addition to the four major lineages AUL, ECSA, IOL and WA, our decision to split the concept of ECSA into its three regions, Eastern Africa, Central (Middle) Africa, and Southern Africa, rose from the presence of distinct clades in this large-scale phylogeny. The epidemic strain from the ECSA-MASA clade was first found in Middle Africa and later identified in South America. The AUL lineage gave rise to an extensively nested American lineage which spans North, South, and Central America, as well as the Caribbean.

Thus, we extrapolate the names for these discrete lineages based on geographic location: Indian Ocean, Eastern African, South American, Middle African, African/Asian, Asian Urban, American and Western African (denoted as "leaves," Figure 1). Some of these lineages belong to the three major epidemic clades: AUL-Am (Asian Urban + American lineages), ECSA-MASA (Middle Africa / South American Lineage) and ECSA-IOL (Eastern Africa + Indian Ocean Lineage). In this paper, we distinguish the individual lineages, which are delimited by geographic region (e.g., IOL, WA), from epidemic clades, which may encompass more than a single lineage (e.g., AUL, ECSA-IOL, ECSA-MASA). Individual limited outbreaks (e.g., IOL in Europe) are not considered new lineages as the outbreak did not generate a new endemic region for the CHIKV lineage (Table 1).

Previous Work	**This Study**	
Lineages	Lineages	Epidemic Clades
Indian Ocean (IOL)	Indian Ocean	ECSA-IOL
East Central and South African (ECSA)	Eastern African	
	South American	ECSA-MASA
	Middle African	
	African/Asian	N/A
Asian Urban (AUL)	Asian Urban	Asian Urban American (AUL-Am)
	American	
West African (WA)	Western African	Western African (WA)

Table 1. Comparison between literature and proposed nomenclature of Chikungunya lineages and epidemic clades based on updated phylogeny.

3.2. Spread of major Chikungunya epidemic lineages

Our reconstruction of the spread of the three distinct epidemic clades is in agreement with the current literature (Figure 2). The Asian Urban lineage has spread from Southeastern Asia to the Pacific Islands and Central America, with traveler cases reaching the Continental United States (AUL-Am clade) [7,59]. A strain from Eastern Africa spread to India and is currently endemic in South and Southeast Asia (ECSA-IOL clade) [60]. More recently, a lineage from Middle Africa was introduced in South America (ECSA-MASA clade) [61,62].

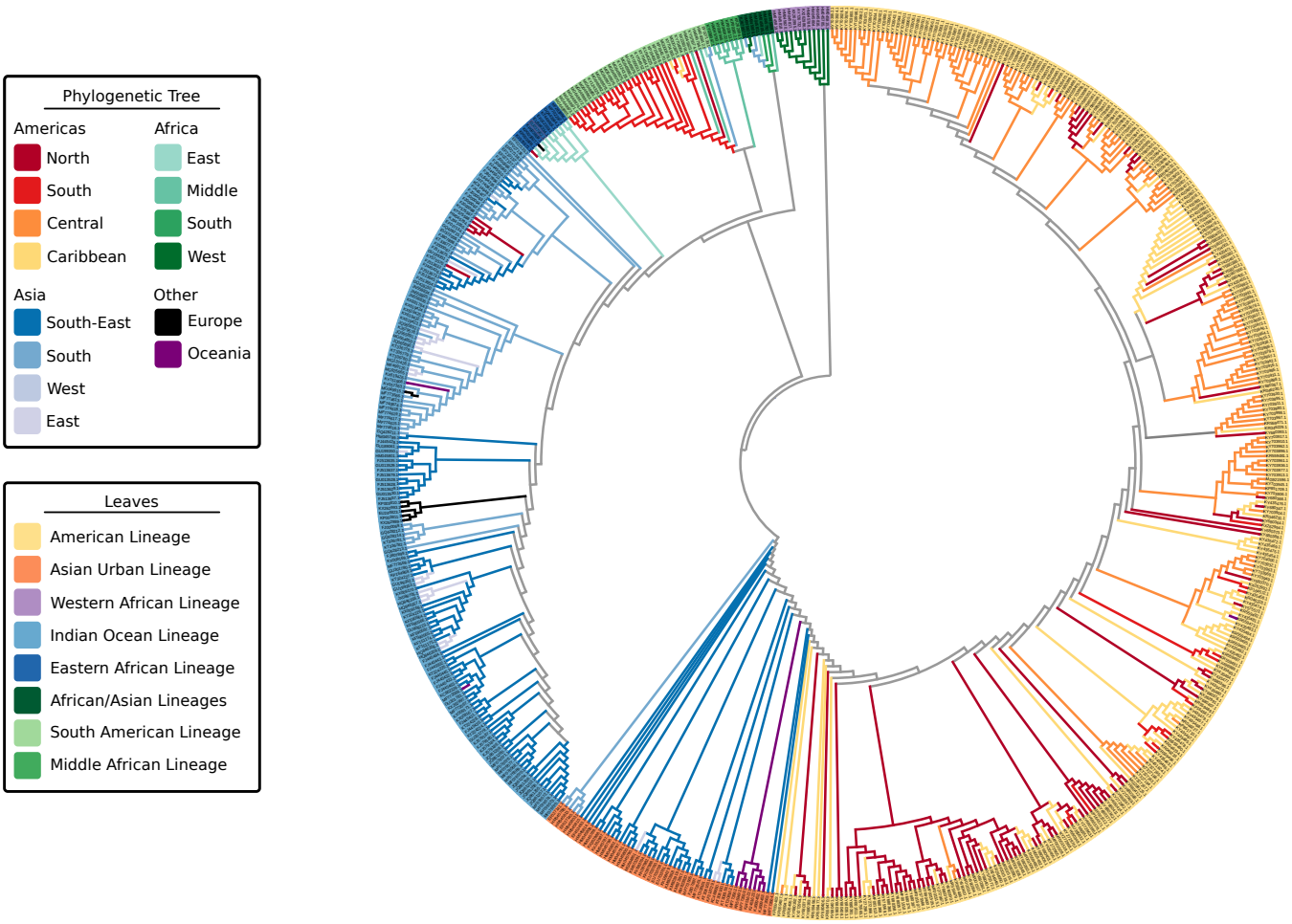


Figure 1. Maximum-Likelihood phylogenetic tree of 590 Chikungunya virus genomic sequences highlighting the geographic locations (tree branches) and lineages (tree leaves). Tree was computed using `MAFFT` alignment of individual ORFs and partitioned tree search with model test using IQ-Tree. Leaves are colored by lineage association. Branches of the tree are colored by geographic location of a sample. Colors of ancestral nodes and branches do not imply geographic origin and therefore are depicted in gray.

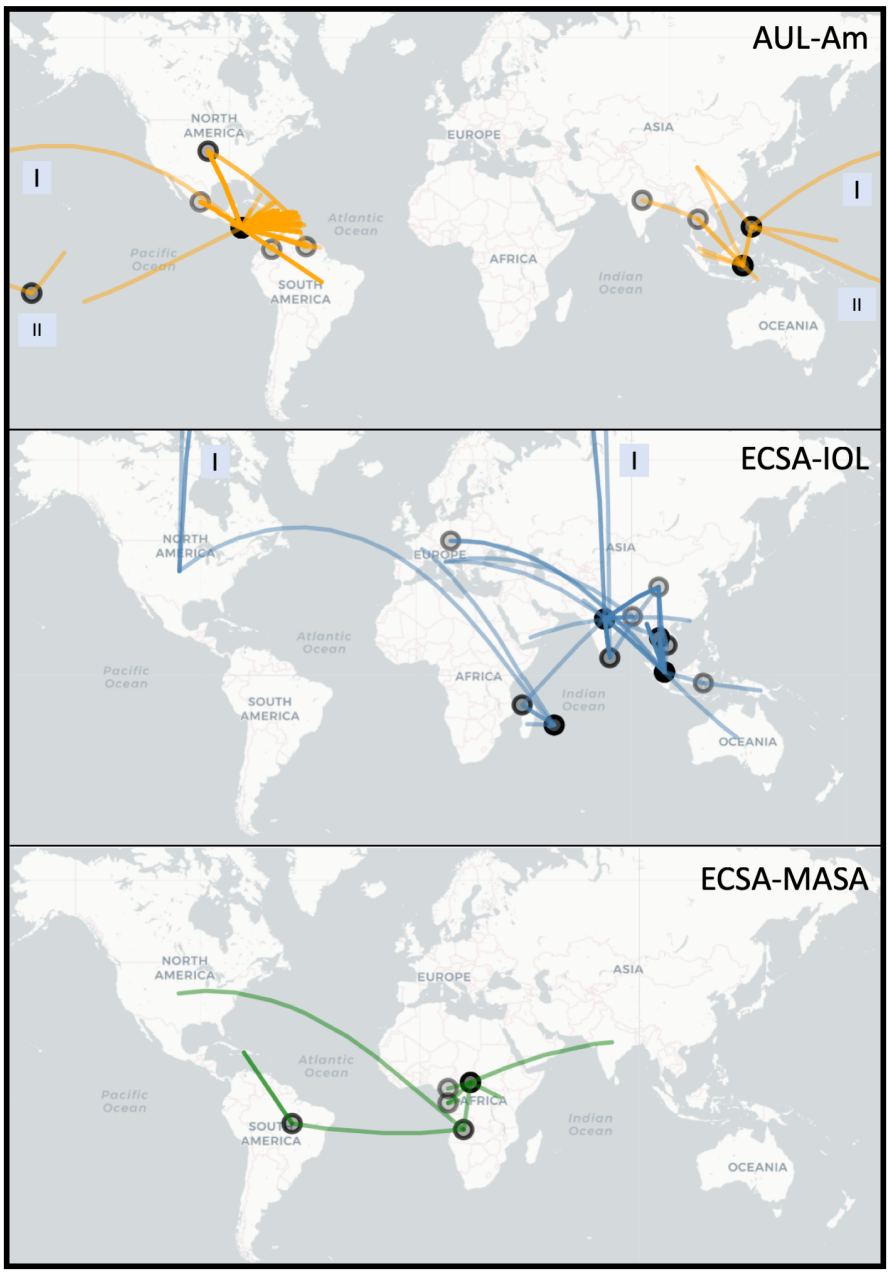


Figure 2. Worldwide spread of major epidemic Chikungunya clades: Asian Urban American (AUL-Am), ECSA Indian Ocean Lineage (ECSA-IOL) and ECSA Middle Africa South America (ECSA-MASA). This figure is rendered using the R package Leaflet with the underlying data based on parsimony ancestral reconstruction of metadata for place of isolation on a the tree resulting from neighbor-joining tree search with collapsed branches under 80% bootstrap threshold (see methods for more details). The circles on map represent locations from which a transmission event originates. The color transparency represents the number of times a transition between characters occur on a location (node) or between locations (edge) (darker lines = more transitions). Boxes labeled "I" and "II" indicate that line crossed the International Date Line (e.g. AUL) or the shortest physical path between the two points (for mapping purposes only) was through the North Pole (e.g. ECSA-IOL).

3.3. Lineage-specific 3'UTR organization

Composition and length of the full 3'UTRs analyzed in this study differ substantially between lineages and range from 510 to 930 nucleotides. Earlier studies reported cascades of repeated sequence elements (RSEs) [33] and a structurally conserved sequence element (CSE) required for replication [2] in CHIKV 3'UTR. Chen *et al.* [34] proposed distinct direct repeat elements 1 and 2 (DR1, DR2), together with a structured Y-shaped element, which all appear in lineage-specific duplication patterns. Filomatori *et al.* [32] adopted this classification in a recent study, however they were not able to see *in silico* secondary structure signals in DR sequences except for a Y-shaped element. Here, we follow up on previous knowledge, using the full set of presently available sequence data (i.e. 196 complete CHIKV 3'UTRs), and conclusively infer hitherto unresolved structural properties of all conserved elements.

From a multiple sequence alignment of all available 3'UTRs from all lineages, we manually characterized five highly conserved regions into four structured and one non-structured elements (Figure 3). We inferred sequence and structural homology from searches in the complete UTR set of all lineages, thus characterizing all instances of each conserved element. We used structural alignment and consensus structure prediction to find the conserved structures of the four structured elements.

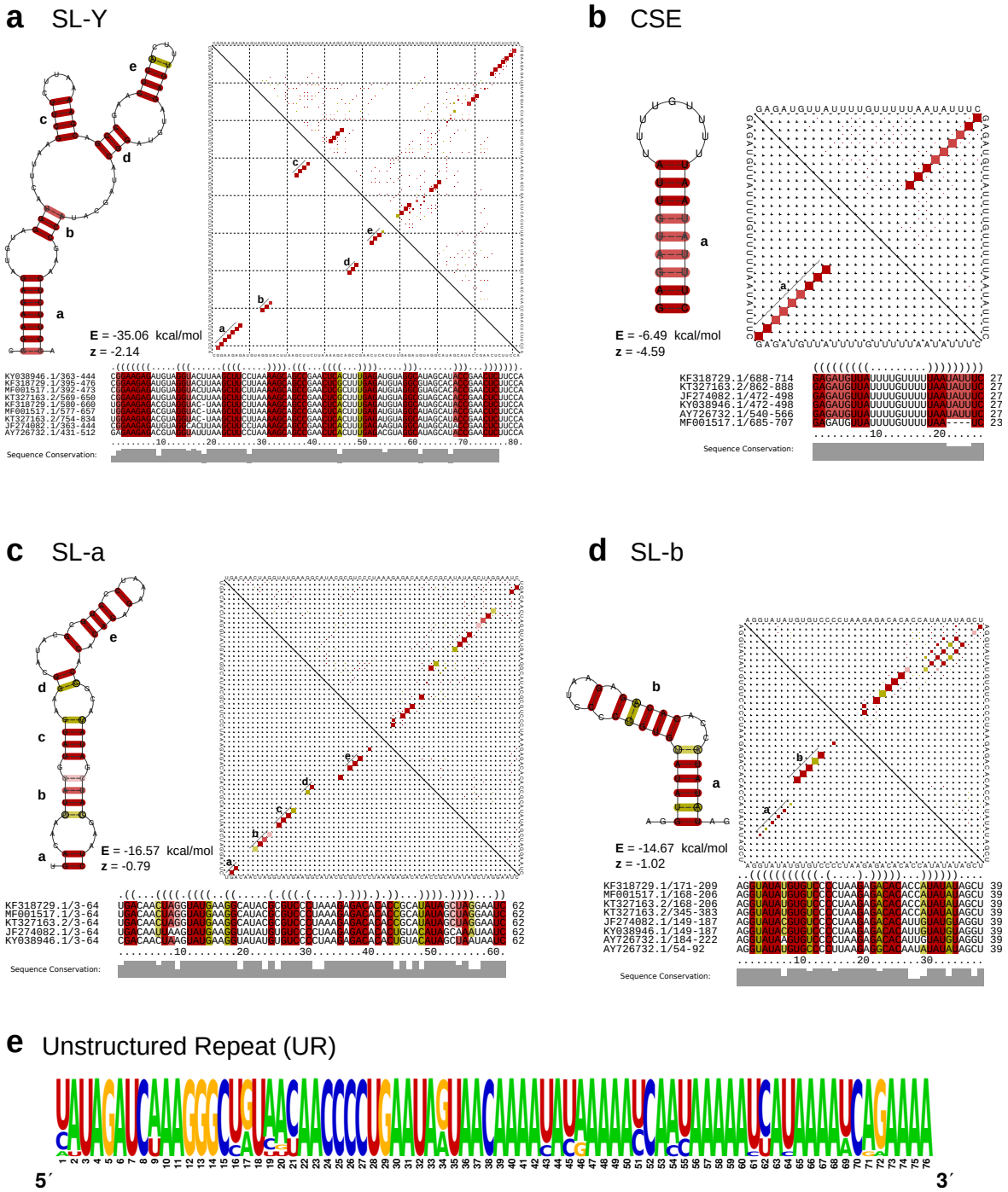
We characterized two stem-loop (SL) elements, termed SL-a and SL-b, the latter being conserved among all lineages (Figure 4). While SL-a is only conserved among the ECSA-derived and AUL lineages, comparison of the two SL elements indicates that SL-b is likely a shortened variant of SL-a and comprised of its innermost hairpin (Figure 3, c, d). Interestingly, this characteristic can also be observed in O'Nyong-Nyong virus (ONNV), which shares common ancestry with CHIKV. The genomic coordinates of both elements approximately correspond to the previously reported DRs, with DR1a spanning over the downstream portion of SL-a and DR1b being consistent with SL-b. The Y-shaped element (SL-Y) is conserved among all lineages but varies in copy number between ECSA-derived/WA (one copy) and AUL (two copies), and shows strong sequence conservation (Figure 3, a). The previously described CSE [63] folds into a short hairpin loop and is found upstream of the poly-A tail in all CHKIV isolates (Figure 3, b).

The structured elements SL-a, SL-b, and SL-Y are separated by one or two repeats of an intrinsically unstructured repeat (UR) element (Figure 3, e). The total copy number of UR elements varies from three to five between lineages (Figure 4). This sequence repeat element lacks structure, as marked by its incapability to fold into any stable RNA secondary structure. Both single sequence predictions of individual repeats, as well as consensus structure prediction using an alignment of all instances did not yield a stable minimum free energy structure, or alternative base pairing probabilities.

Annotation of all available 3'UTR sequences with CMs and HMMs revealed that different CHIKV lineages have evolved distinct patterns of structured and unstructured functional RNA elements. This lineage-specificity of potentially functional RNAs has led to a varied 3'UTR architecture among CHIKV strains. This is particularly pronounced in the AUL and its descendants, which show a characteristic 3'UTR organization. Several geographically diverse isolates of the AUL-Am (also referenced as Caribbean lineage, AUL-Cbn) appear to have undergone a partial duplication event in the 3'UTR [65]. As a result, an additional copy of the non-structural sequence repeat UR and the structural element SL-b are found immediately downstream to the first copy of SL-b and UR (Figure 4).

4. Discussion

Studies on CHIKV phylogenetics have provided phylogenomic means of understanding disease spread in the context of correlating disease epidemiology. The phylogenetic analyses we present are in agreement with the current literature, although the increased number of sequences presented provide insights on the hidden history of ECSA. Interestingly, the phylogeny based on ORFs is in agreement with the phylogeny based on complete UTRs, despite the intuitive idea that different evolutionary forces should act on coding and non-coding regions.



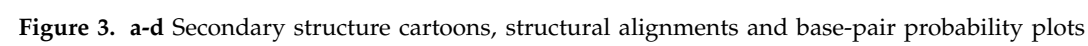


Figure 3. **a-d** Secondary structure cartoons, structural alignments and base-pair probability plots (dot-plots) of the structured CHIKV 3′UTR elements SL-Y, CSE, SL-a, and SL-b. Colored squares in the dot-plot report a pairing of two bases, with full squares corresponding to 100% pairing probability. The base pairings pattern of the corresponding most stable, minimum free energy structures are shown in the lower triangles, while the upper triangles show the base pairing probabilities considering all possible alternative, suboptimal structures. Colors in structure drawings, alignments and dot-plots indicate how many different types of base combinations are found for a particular base pair (red: 1; yellow: 2). All structure plots show the consensus structures as calculated by `RNAalifold` [56]. For each structure, the free energy (E) and the `RNAz` [64] z-score (z) are provided. Negative z-scores indicate a structure is more stable than it would be expected for other RNAs of the same dinucleotide content. **e** Sequence logo of the intrinsically unstructured repeat element UR, computed from all HMM-derived homologous sequences within the data set. Structure cartoons are not shown since the overall base pairing probability is essentially zero for this sequence.

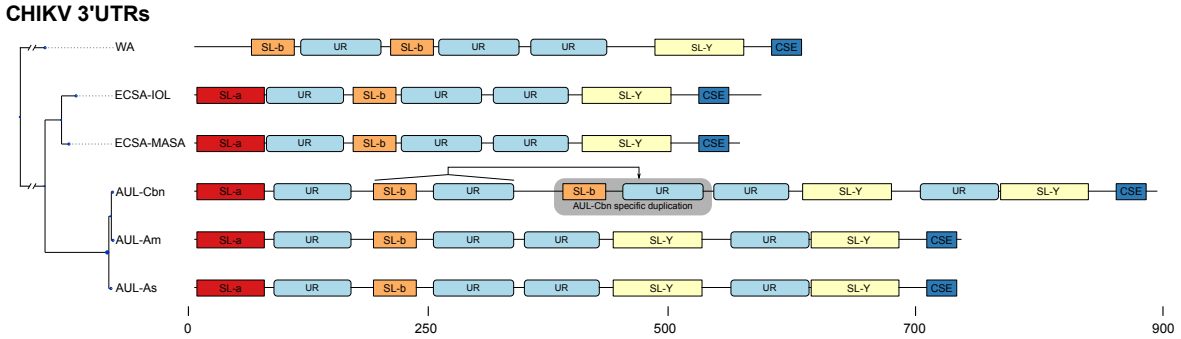


Figure 4. Schematic of the 3'UTR architecture of all major CHIKV strains. The phylogeny on the left is derived from a Maximum-Likelihood tree based on a genomic alignment of six representative sequences. For each strain, the sequence with the longest UTR was selected as a representative (WA: AY726732.1, ECSA-IOL: JF274082.1, ECSA-MASA: KY038946.1, AUL-Cbn: KT327163.2 (Mexico), AUL-Am: MF001517.1 (USA), AUL-Asia: KF318729.1 (China). All boxes correspond to exact coordinates of annotated structural/non-structural elements. A scale bar is provided for the nucleotide coordinates.

4.1. Chikungunya virus nomenclature

It is critical to emphasize the significance of the definition of "lineage" and "epidemic clade" when discussing the evolution of a virus. Much has been proposed as to how species should be delineated. In 1977, Wiley [66] proposed that the definition of species should be that "A species is a single lineage of ancestral descendant populations of organisms which maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate". Peterson [67] criticizes the current system of viral taxonomy, mentioning that one criterion alone should dominate viral classification: the evolutionary independence of evolving lineages.

Assessing each CHIKV taxon as an individual, and bearing in mind the concept of phylogenetic independence, we cannot conclusively state that CHIKV sequences isolated in the Americas are divergent enough from ancestral sequences in Africa and Asia to be considered two new American lineages. This logic however, is limited when considering the similar phylogenetic structure of IOL rising from Eastern African taxa. This non-uniformity represents the difficulty in qualifying lineages based on subjective interpretation.

Further, IOL and AUL have region-specific derived clades circulating newly endemic regions. Peterson [67] warns against the use of endemic regions to classify species. We acknowledge the flaw in solely using geography as a naming mechanism as the increased movement between continents due to airport passenger can substantially increase the number of co-circulating lineages named based on geography.

In order to remain consistent with the current literature we kept the naming scheme based on geography, but defined lineage as the pool of individual taxa found in a clade that are not isolated geographically. We define epidemic clade as a monophyletic group that may encompass one or more lineages regardless of geography.

4.2. Lineage diversity in Chikungunya virus

Multiple waves of CHIKV infection have occurred over the past two decades, with outbreaks traced from Eastern and Middle Africa to Southeastern Asia (e.g., Thailand, Malaysia), Southern Asia (India) and the Americas [20]. Clades of CHIKV observed in the analyses are consistent with those found in the literature with an important exception. Similar to White *et al.* [23], who alluded to the presence of ECSA subgroups, we observed that ECSA has multiple co-circulating lineages which gave rise to the current outbreaks in Southern and Southeastern Asia and the Americas. With the large amount of new sequences we analyzed, what was once called ECSA, East Central South African

lineage, is no longer a single lineage because it is not monophyletic. Figure 1 demonstrates that Middle African taxa gave rise to the outbreak in South America, while Eastern African strains gave origin to IOL. IOL has remained nearly exclusively on the Asian continent, with exception of a few African and European outbreaks. AUL, a sister clade to all ECSA derived lineages, has reached the American continent with extensive nesting diversity (Figure 1). In Africa, AUL is known to have emerged in 2004 and spread explosively in the Comoros islands and incited a severe epidemic in Kenya [18]. In the Americas, strains from AUL are observed as coming from the US to Central America (Figure 2), given the absence of locally acquired cases in the US, the indication of an ancestral sequence originating a transmission event to Central America could be due to lack of sampling the ancestral strain sequence and the close similarity of it with strain sequences found in Central America.

Asian Urban and IOL are the main lineages of CHIKV in Asia [68,69]. Moreover, Asian Urban has since emerged in Central America in 2013 and South America in 2014[70] forming the AUL-Am clade. The Americas faced a devastating outbreak in 2013-2014 after the AUL lineage was introduced to the Caribbean [71], causing nearly 3 million estimated cases. In 2014, the ECSA lineage was detected in Brazil as well [70]. Based on our phylogeny, the additional sister group lineage to both ECSA-IOL and ECSA-MASA, which encompasses Southern African isolates and is related to a few Asian isolates (of the African/Asian lineage) has no major outbreaks reported.

We propose that the ECSA genotype is too reductive in light of the rapid expansion and increased taxonomic sampling, and should be separated into their respective regions of the African continent: Eastern, Central, and Southern. Nevertheless, it is important to raise the attention to the lack of sampling in Africa which obscures the Chikungunya virus dynamics between the distinct regions within Africa. The current but limited amount of African sequences utilized in this study indicate that there may be multiple co-circulating lineages and this results should serve as justification for the importance of understanding the virus dynamics in Africa. As more isolates are sequenced, a more nuanced history of CHIKV evolution will emerge which will aid in our understanding of the dynamics of these outbreaks.

4.3. Lineage-specific 3'UTR architecture

Earlier work in alphavirus 3'UTRs reported a varied architecture of repeat sequence elements (RSEs) [2,33,72], which appear to influence virus replication in mosquito cells without affecting infectivity in mammalian cells [34]. Interestingly, molecular evolution experiments with engineered CHIKV 3'UTRs in mosquito and mammalian cells suggest that improved fitness was acquired mainly due to mutations in protein coding regions [73]. Still, the specific biological functions of individual 3'UTR RSEs have not been studied in detail. The comparative genomics approach employed in this study confirms the previously reported repeat organization in CHIKV 3'UTR sequences. It is now possible to conclusively associate repeat sequences with distinct RNA secondary structures that are conserved among all lineages. Importantly, the organization of all CHIKV 3'UTRs follows a common pattern characterized by an alternating arrangement of structured, i.e. SL-a, SL-b and SL-Y, and unstructured regions. The latter are defined by a lack of any RNA structure, as predicted by single sequence and consensus folding algorithms. This suggests that these regions of the CHIKV genome exist in a predominantly unpaired state, rendering them highly accessible.

We hypothesize that such intrinsically unstructured repeat regions act as structural insulators that separate functional secondary structures from each other. Unstructured stretches can mediate the formation of structure in adjacent regions by prohibiting base pairs that are in conflict with the element structure. Likewise, single-stranded, unstructured RNAs can serve as easily accessible protein binding sites. In particular, the enrichment of `A` nucleotides in the unstructured region not only indicates the lack of secondary structure, but might also facilitate the sponging of RNA-binding proteins [74]. Earlier studies suggested an involvement of alphavirus 3'UTRs with cellular HuR and La proteins [75,76]. While HuR binding in CHIKV appears impaired [77], stabilization of viral RNA through host factors is an evolutionary trait to counteract viral genome degradation by host nuclease activity [25]. Conversely,

other RNA-binding proteins, such as the Musashi family of translational regulators have been shown to efficiently interact with viral 3′UTRs [31,78]. The canonical Musashi binding motif, UAG, is present at multiple loci within the unstructured repeat of all CHIKV 3′UTR sequences, supporting the biological role of intrinsic lack of structure. Besides protein-aided regulation, RNA interference is a crucial mechanism that controls viral replication in mosquitoes. Host-encoded microRNAs (miRNAs) were shown to interact with the 3′UTRs of Eastern equine encephalitis virus (EEEV) [79] and CHIKV [80,81]. The miRNA binding site reported by Dubey *et al.* [80] overlaps with the structured Y-shaped element (SL-Y) characterized here. Whether or not the CHIKV genome encodes for endogenous viral miRNAs, as observed in flaviviruses [82], remains to be established.

The structural aspects of alphavirus RNA is understudied when compared to other arboviruses. While repeat motifs were shown to be a crucial architectural pattern in alphavirus UTRs, many studies focused on the role and nature of these sequence repeats, without asking how far RNA structure might be the evolutionary driving force behind UTR evolution. A recent study employing SHAPE structure probing (selective 2′-hydroxyl acylation analyzed by primer extension) experiments could not find evidence for large-scale structural conservation of RNA within alphaviruses [37]. Work in the 5′UTRs of VEEV [83] and SINV [84] could associate RNA structure elements with replication efficiency and antiviral response [85]. We identified structured elements in the 5′UTR of CHIKV, however did not observe strong covariation patterns (data not shown). Likewise, the structured elements SL-a and SL-b in the 3′UTR exhibit only a few covariations. This observed lack of covariation might be due to the high degree of sequence conservation in the 3′UTR. Alternatively, it could also be a consequence of a sampling bias, as many samples in our dataset originate from individual outbreaks and therefore only represent a fraction of the CHIKV population. In contrast to other arboviruses, such as flaviviruses carrying exoribonuclease-resistant RNAs (xrRNAs) [35], the lack of covariations makes it difficult to assert to what extent secondary structure conservation is a driving force of CHIKV UTR evolution. This is underlined by a weak support of structure-associated z-scores for elements SL-a and SL-b (Figure 3 c,d). Contrary, the SL-Y and CSE elements have very stable structures as indicated by strongly negative z-scores. One of the most striking features of the CHIKV 3′UTR is presence of multiple copies of the unstructured UR region separating structured regions. Our findings indicate that the evolutionary forces acting on CHIKV 3′UTR manifest as a process that can rapidly re-arrange structured and unstructured RNA elements, as observed in a recently emerged partial duplication of CHIKV strains isolated from the Caribbean [65]. Consequently, our predictions encourage further experimental validation to assess the structural and functional role of these proposed elements. In particular, it would be highly informative to elucidate whether or not these elements mediate the production of standalone non-coding RNAs. Additionally, experiments could reveal whether these conserved elements act in cis or trans.

To obtain a more comprehensive picture of the 3′UTR phylogeny in the context of diverged CHIKV lineages, we computed a separate phylogeny based exclusively on 3′UTR sequences (Figure S1). Although coding and non-coding parts of the viral genome are subject to different evolutionary pressures, mapping lineage association onto this tree reproduces the topology of the full CDS tree (Figures 1). This is to some extent surprising, as the nature of evolutionary forces acting on both entities is supposed to be different, best illustrated by the requirement to maintain a reading frame. This means that the 3′UTR evolution is not disentangled from CDS evolution in CHIKV, as expected intuitively. On a broader scale, our data provide evidence that CHIKV CDS and 3′UTRs show coherent evolutionary patterns, with leads to the hypothesis that the 3′UTR might represent an evolutionary scaffold that can rapidly replace functional elements, as required for adaptation to environmental change, e.g. by means of recombination [32]. Still, CHIKV 3′UTRs do not exhibit the high degree of plasticity known from other arboviruses with functional 3′UTRs, such as flaviviruses [28,35].

Author Contributions: MTW and ABS conceived the study; ABS, RO, RH and MTW analyzed the data; DJ supported the phylogenetic analysis; ILH supported the RNA structure analysis; All authors contributed to writing the paper and approved of the submitted version.

Funding: This work was supported by the National Institutes of Health (NIH) National Institute of Allergy and Infectious Diseases (grant number AI135992), by the Doktoratskolleg RNA Biology at University of Vienna and the Austrian Science Fund (SFB F43, I-1303).

Acknowledgments: We thank Adiba Hassan, MPH for the valuable discussions during the development of this work. We acknowledge the support of the Department of Bioinformatics and Genomics, The College of Computing and Informatics, and the Graduate School of the University of North Carolina at Charlotte.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gould, E.; Pettersson, J.; Higgs, S.; Charrel, R.; de Lamballerie, X. Emerging arboviruses: Why today? *One Health* **2017**, *4*, 1–13.
2. Pfeffer, M.; Kinney, R.M.; Kaaden, O.R. The Alphavirus 3′-Nontranslated Region: Size Heterogeneity and Arrangement of Repeated Sequence Elements. *Virology* **1998**, *240*, 100–108.
3. Campbell, L.P.; Luther, C.; Moo-Llanes, D.; Ramsey, J.M.; Danis-Lozano, R.; Peterson, A.T. Climate change influences on global distributions of dengue and chikungunya virus vectors. *Philos T Roy Soc B* **2015**, *370*, 20140135.
4. 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.
5. Forrester, N.; Palacios, G.; Tesh, R.; Savji, N.; Guzman, H.; Sherman, M.; Weaver, S.; Lipkin, W. Genome-scale phylogeny of the alphavirus genus suggests a marine origin. *J Virol* **2012**, *86*, 2729–2738.
6. Zeller, H.; Van Bortel, W.; Sudre, B. Chikungunya: its history in Africa and Asia and its spread to new regions in 2013–2014. *J Infect Dis* **2016**, *214*, S436–S440.
7. Chen, R.; Puri, V.; Fedorova, N.; Lin, D.; Hari, K.L.; Jain, R.; Rodas, J.D.; Das, S.R.; Shabman, R.S.; Weaver, S.C. Comprehensive genome scale phylogenetic study provides new insights on the global expansion of Chikungunya virus. *J Virol* **2016**, *90*, 10600–10611.
8. Ross, R. The Newala epidemic: III. The virus: isolation, pathogenic properties and relationship to the epidemic. *Epidemiol Infect* **1956**, *54*, 177–191.
9. Mason, P.; Haddow, A. An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952–1953: An additional note on Chikungunya virus isolations and serum antibodies. *T Roy Soc Trop Med H* **1957**, *51*, 238–240.
10. Rezza, G.; Nicoletti, L.; Angelini, R.; Romi, R.; Finarelli, A.; Panning, M.; Cordioli, P.; Fortuna, C.; Boros, S.; Magurano, F.; others. Infection with Chikungunya virus in Italy: an outbreak in a temperate region. *Lancet* **2007**, *370*, 1840–1846.
11. Van Bortel, W.; Dorleans, F.; Rosine, J.; Blateau, A.; Rousset, D.; Matheus, S.; Leparc-Goffart, I.; Flusin, O.; Prat, C.; Cesaire, R.; others. Chikungunya outbreak in the Caribbean region, December 2013 to March 2014, and the significance for Europe. *Eurosurveillance* **2014**, *19*, 20759.
12. Teixeira, M.G.; Andrade, A.M.; Maria da Conceição, N.C.; Castro, J.S.; Oliveira, F.L.; Goes, C.S.; Maia, M.; Santana, E.B.; Nunes, B.T.; Vasconcelos, P.F. East/Central/South African genotype Chikungunya virus, Brazil, 2014. *Emerg Infect Dis* **2015**, *21*, 906.
13. Weaver, S.C.; Lecuit, M. Chikungunya virus and the global spread of a mosquito-borne disease. *New Engl J Med* **2015**, *372*, 1231–1239.
14. Rezza, G.; Weaver, S.C. Chikungunya as a paradigm for emerging viral diseases: Evaluating disease impact and hurdles to vaccine development. *PLoS Neglect Trop D* **2019**, *13*, e0006919.
15. Borgherini, G.; Poubeau, P.; Staikowsky, F.; Lory, M.; Moullec, N.L.; Becquart, J.P.; Wengling, C.; Michault, A.; Paganin, F. Outbreak of Chikungunya on Reunion Island: early clinical and laboratory features in 157 adult patients. *Clin Infect Dis* **2007**, *44*, 1401–1407.
16. Pulmanausahakul, R.; Roytrakul, S.; Auewarakul, P.; Smith, D.R. Chikungunya in Southeast Asia: understanding the emergence and finding solutions. *Int J Infect Dis* **2011**, *15*, e671–e676.
17. Suhrbier, A.; La Linn, M. Clinical and pathologic aspects of arthritis due to Ross River virus and other alphaviruses. *Curr Opin Rheumatol* **2004**, *16*, 374–379.
18. Wahid, B.; Ali, A.; Rafique, S.; Idrees, M. Global expansion of chikungunya virus: mapping the 64-year history. *Int J Infect Dis* **2017**, *58*, 69–76.

19. Diallo, M.; Thonnon, J.; Traore-Lamizana, M.; Fontenille, D. Vectors of Chikungunya virus in Senegal: current data and transmission cycles. *Am J Trop Med Hyg* **1999**, *60*, 281–286.
20. Weaver, S.C.; Forrester, N.L. Chikungunya: Evolutionary history and recent epidemic spread. *Antivir Res* **2015**, *120*, 32–39.
21. 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.
22. 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.; Maharaj, P.D.; Brault, A.C.; Weaver, S.C. Genome-Scale Phylogenetic Analyses of Chikungunya Virus Reveal Independent Emergences of Recent Epidemics and Various Evolutionary Rates. *J Virol* **2010**, *84*, 6497–6504.
23. White, S.K.; Mavian, C.; Salemi, M.; Morris Jr, J.G.; Elbadry, M.A.; Okech, B.A.; Lednicky, J.A.; Dunford, J.C. A new "American" subgroup of African-lineage Chikungunya virus detected in and isolated from mosquitoes collected in Haiti, 2016. *PloS One* **2018**, *13*, e0196857.
24. Li, X.F.; Jiang, T.; Deng, Y.Q.; Zhao, H.; Yu, X.D.; Ye, Q.; Wang, H.J.; Zhu, S.Y.; Zhang, F.C.; Qin, E.D.; others. Complete genome sequence of a Chikungunya virus isolated in Guangdong, China. *J Virol* **2012**, *86*, 8904–8905.
25. Hyde, J.L.; Chen, R.; Trobaugh, D.W.; Diamond, M.S.; Weaver, S.C.; Klimstra, W.B.; Wilusz, J. The 5′ and 3′ ends of alphavirus RNAs–non-coding is not non-functional. *Virus Res* **2015**, *206*, 99–107.
26. Casal, P.E.; Chouhy, D.; Bolatti, E.M.; Perez, G.R.; Stella, E.J.; Giri, A.A. Evidence for homologous recombination in Chikungunya Virus. *Mol Phylogenet Evol* **2015**, *85*, 68–75.
27. Roby, J.A.; Pijlman, G.P.; Wilusz, J.; Khromykh, A.A. Noncoding subgenomic flavivirus RNA: multiple functions in West Nile virus pathogenesis and modulation of host responses. *Viruses* **2014**, *6*, 404–427.
28. Villordo, S.M.; Carballeda, J.M.; Filomatori, C.V.; Gamarnik, A.V. RNA Structure Duplications and Flavivirus Host Adaptation. *Trends Microbiol* **2016**, *24*, 270–283.
29. Chapman, E.G.; Moon, S.L.; Wilusz, J.; Kieft, J.S. RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. *eLife* **2014**, *3*, e01892.
30. MacFadden, A.; O'Donoghue, Z.; Silva, P.A.; Chapman, E.G.; C, O.R.; Sterken, M.G.; Pijlman, G.P.; Bredenbeek, P.J.; Kieft, J.S. Mechanism and structural diversity of exoribonuclease-resistant RNA structures in flaviviral RNAs. *Nat Commun* **2018**, *9*, 119.
31. de Bernardi Schneider, A.; Wolfinger, M.T. Musashi binding elements in Zika and related Flavivirus 3′ UTRs: A comparative study *in silico*. *Sci Rep* **2019**, *9*, 6911.
32. Filomatori, C.V.; Bardossy, E.S.; Merwaiss, F.; Suzuki, Y.; Henrion, A.; Saleh, M.C.; Alvarez, D.E. RNA recombination at Chikungunya virus 3'UTR as an evolutionary mechanism that provides adaptability. *PLoS Pathog* **2019**, *15*, e1007706.
33. Ou, J.H.; Trent, D.W.; Strauss, J.H. The 3′-non-coding regions of alpharivus RNAs contain repeating sequences. *J Mol Biol* **1982**, *156*, 719–730.
34. 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.
35. Ochsenreiter, R.; Hofacker, I.L.; Wolfinger, M.T. Functional RNA Structures in the 3'UTR of Tick-Borne, Insect-Specific and No-Known-Vector Flaviviruses. *Viruses* **2019**, *11*, 298.
36. Fernandes, J.; Jayaraman, B.; Frankel, A. The HIV-1 Rev response element: an RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex. *RNA Biol* **2012**, *9*, 6–11.
37. Kutchko, K.M.; Madden, E.A.; Morrison, C.; Plante, K.S.; Sanders, W.; Vincent, H.A.; Cruz Cisneros, M.C.; Long, K.M.; Moorman, N.J.; Heise, M.T.; Laederach, A. Structural divergence creates new functional features in alphavirus genomes. *Nucleic Acids Res* **2018**, *46*, 3657–3670.
38. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **2013**, *30*, 772–780.
39. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **2014**, *30*, 3276–3278.
40. Martin, D.P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol* **2015**, *1*.

41. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **2014**, *32*, 268–274.
42. Minh, B.Q.; Nguyen, M.A.T.; von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* **2013**, *30*, 1188–1195.
43. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **2010**, *59*, 307–321.
44. de Bernardi Schneider, A. Arboviruses: The hidden path of an imminent threat. PhD thesis, The University of North Carolina at Charlotte, 2018.
45. Huerta-Cepas, J.; Serra, F.; Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol* **2016**, *33*, 1635–1638.
46. Schliep, K.P. phangorn: phylogenetic analysis in R. *Bioinformatics* **2010**, *27*, 592–593.
47. Charif, D.; Lobry, J.R. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural approaches to sequence evolution*; Springer, 2007; pp. 207–232.
48. Paradis, E.; Claude, J.; Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **2004**, *20*, 289–290.
49. de Bernardi Schneider, A.; Ford, C.T.; Hostager, R.; Williams, J.; Cioce, M.; Çatalyürek, Ü.V.; Wertheim, J.O.; Janies, D. StrainHub: A phylogenetic tool to construct pathogen transmission networks. *bioRxiv* **2019**, [https://www.biorxiv.org/content/early/2019/05/31/650283.full.pdf]. doi:10.1101/650283.
50. Louca, S.; Doebeli, M. Efficient comparative phylogenetics on large trees. *Bioinformatics* **2017**, *34*, 1053–1055.
51. Graul, C. LeafletR: Interactive Web-Maps Based on the Leaflet JavaScript Library. R-package version 0.4-0, 2016.
52. Hijmans, R.J. geosphere: spherical trigonometry. R package version 1.5-5. *See https://cran.r-project.org/package=geosphere* **2016**.
53. Eddy, S.R.; Durbin, R. RNA sequence analysis using covariance models. *Nucleic Acids Res* **1994**, *22*, 2079–2088.
54. Wolfinger, M.T.; Fallmann, J.; Eggenhofer, F.; Amman, F. ViennaNGS: A toolbox for building efficient next-generation sequencing analysis pipelines. *F1000Research* **2015**, *4*. doi:10.12688/f1000research.6157.1.
55. Will, S.; Reiche, K.; Hofacker, I.L.; Stadler, P.F.; Backofen, R. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput Biol* **2007**, *3*, e65.
56. Bernhart, S.H.; Hofacker, I.L.; Will, S.; Gruber, A.R.; Stadler, P.F. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* **2008**, *9*, 474.
57. Lorenz, R.; Bernhart, S.H.; Zu Siederdissen, C.H.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. ViennaRNA Package 2.0. *Algorithm Mol Biol* **2011**, *6*, 26.
58. Nawrocki, E.P.; Eddy, S.R. Infernal 1.1: 100-fold Faster RNA Homology Searches. *Bioinformatics* **2013**, *29*, 2933–2935. doi:10.1093/bioinformatics/btt509.
59. Lindsey, N.P.; Staples, J.E.; Fischer, M. Chikungunya virus disease among travelers — United States, 2014–2016. *Am J Trop Med Hyg* **2018**, *98*, 192–197.
60. Hapuarachchi, H.; Bandara, K.; Sumanadasa, S.; Hapugoda, M.; Lai, Y.L.; Lee, K.S.; Tan, L.K.; Lin, R.T.; Ng, L.F.; Bucht, G.; others. Re-emergence of Chikungunya virus in South-east Asia: virological evidence from Sri Lanka and Singapore. *J Gen Virol* **2010**, *91*, 1067–1076.
61. da Costa, A.C.; Thézé, J.; Komninakis, S.C.V.; Sanz-Duro, R.L.; Felinto, M.R.L.; Moura, L.C.C.; de Oliveira Barroso, I.M.; Santos, L.E.C.; de Lemos Nunes, M.A.; Moura, A.A.; others. Spread of Chikungunya virus east/central/south African genotype in northeast Brazil. *Emerg Infect Dis* **2017**, *23*, 1742.
62. Souza, T.M.A.; Azeredo, E.L.; Badolato-Corrêa, J.; Damasco, P.V.; Santos, C.; Petitinga-Paiva, F.; Nunes, P.C.G.; Barbosa, L.S.; Cipitelli, M.C.; Chouin-Carneiro, T.; others. First report of the East-Central South African genotype of Chikungunya virus in Rio de Janeiro, Brazil. *PLoS Currents* **2017**, *9*.
63. Hardy, R.W.; Rice, C.M. Requirements at the 3′ end of the *Sindbis virus* genome for efficient synthesis of minus-strand RNA. *J Virol* **2005**, *79*, 4630–4639.
64. Gruber, A.R.; Findeiß, S.; Washietl, S.; Hofacker, I.L.; Stadler, P.F. RNAz 2.0: improved noncoding RNA detection. In *Biocomp 2010*; World Scientific, 2010; pp. 69–79.

65. Stapleford, K.A.; Moratorio, G.; Henningsson, R.; Chen, R.; Matheus, S.; Enfissi, A.; Weissglas-Volkov, D.; Isakov, O.; Blanc, H.; Mounce, B.C.; others. Whole-Genome Sequencing Analysis from the Chikungunya Virus Caribbean Outbreak Reveals Novel Evolutionary Genomic Elements. *PLoS Neglect Trop D* **2016**, *10*, e0004402.
66. Wiley, E.O. The evolutionary species concept reconsidered. *Syst Zool* **1978**, *27*, 17–26.
67. Peterson, A.T. Defining viral species: making taxonomy useful. *Virol J* **2014**, *11*, 131.
68. Yang, C.F.; Su, C.L.; Hsu, T.C.; Chang, S.F.; Lin, C.C.; Huang, J.C.; Shu, P.Y. Imported Chikungunya Virus Strains, Taiwan, 2006–2014. *Emerg Infect Dis* **2016**, *22*, 1981.
69. Maha, M.S.; Susilarini, N.K.; Nur Ika Hariastuti, S. Chikungunya virus mutation, Indonesia, 2011. *Emerg Infect Dis* **2015**, *21*, 379.
70. Nunes, M.R.T.; Faria, N.R.; de Vasconcelos, J.M.; Golding, N.; Kraemer, M.U.; de Oliveira, L.F.; da Silva Azevedo, R.d.S.; da Silva, D.E.A.; da Silva, E.V.P.; da Silva, S.P.; others. Emergence and potential for spread of Chikungunya virus in Brazil. *BMC Med* **2015**, *13*, 102.
71. 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.
72. Strauss, J.H.; Strauss, E.G. The Alphaviruses: Gene Expression, Replication, and Evolution. *Microbiol Mol Biol R* **1994**, *58*, 491–562.
73. Morley, V.J.; Noval, M.G.; Chen, R.; Weaver, S.C.; Vignuzzi, M.; Stapleford, K.A.; Turner, P.E. Chikungunya virus evolution following a large 3′ UTR deletion results in host-specific molecular changes in protein-coding regions. *Virus Evol* **2018**, *4*, vey012.
74. Charley, P.A.; Wilusz, J. Sponging of cellular proteins by viral RNAs. *Cur Opin Virol* **2014**, *9*, 14–18.
75. Pardigon, N.; Strauss, J.H. Mosquito homolog of the La autoantigen binds to Sindbis virus RNA. *J Virol* **1996**, *70*, 1173–1181.
76. Sokoloski, K.J.; Dickson, A.M.; Chaskey, E.L.; Garneau, N.L.; Wilusz, C.J.; Wilusz, J. Sindbis Virus Usurps the Cellular HuR Protein to Stabilize its Transcripts and Promote Productive Infections in Mmammalian and Mmosquito Cells. *Cell Host Microbe* **2010**, *8*, 196–207.
77. Dickson, A.M.; Anderson, J.R.; Barnhart, M.D.; Sokoloski, K.J.; Oko, L.; Opyrchal, M.; Galanis, E.; Wilusz, C.J.; Morrison, T.E.; Wilusz, J. Dephosphorylation of HuR protein during alphavirus infection is Associated with HuR relocalization to the cytoplasm. *J Biol Chem* **2012**, pp. jbc–M112.
78. Chavali, P.L.; Stojic, L.; Meredith, L.W.; Joseph, N.; Nahorski, M.S.; Sanford, T.J.; Sweeney, T.R.; Krishna, B.A.; Hosmillo, M.; Firth, A.E.; Bayliss, R.; Marcelis, C.L.; Goodfellow, I.; Woods, C.G.; Gergely, F. Neurodevelopmental protein Musashi 1 interacts with the Zika genome and promotes viral replication. *Science* **2017**, p. eaam9243.
79. Trobaugh, D.W.; Gardner, C.L.; Sun, C.; Haddow, A.D.; Wang, E.; Chapnik, E.; Mildner, A.; Weaver, S.C.; Ryman, K.D.; Klimstra, W.B. RNA viruses can hijack vertebrate microRNAs to suppress innate immunity. *Nature* **2014**, *506*, 245.
80. Dubey, S.K.; Shrinet, J.; Sunil, S. Aedes aegypti microRNA, miR-2944b-5p interacts with 3′UTR of Chikungunya virus and cellular target vps-13 to regulate viral replication. *PLOS Negl Trop D* **2019**, *13*, e0007429.
81. Yen, P.S.; Chen, C.H.; Sreenu, V.; Kohl, A.; Failloux, A.B. Assessing the Potential Interactions between Cellular miRNA and Arboviral Genomic RNA in the Yellow Fever Mosquito, *Aedes aegypti*. *Viruses* **2019**, *11*, 540.
82. Hussain, M.; Torres, S.; Schnettler, E.; Funk, A.; Grundhoff, A.; Pijlman, G.P.; Khromykh, A.A.; Asgari, S. West Nile virus encodes a microRNA-like small RNA in the 3′ untranslated region which up-regulates GATA4 mRNA and facilitates virus replication in mosquito cells. *Nucleic Acids Res* **2012**, *40*, 2210–2223.
83. Kulasegaran-Shylini, R.; Atasheva, S.; Gorenstein, D.G.; Frolov, I. Structural and Functional Elements of the Promoter Encoded by the 5′ Untranslated Region of the Venezuelan Equine Encephalitis Virus Genome. *J Virol* **2009**, *83*, 8327–8339.
84. Fayzulin, R.; Frolov, I. Changes of the Ssecondary Sstructure of the 5′ end of the Sindbis Virus Genome Inhibit Virus Growth in Mosquito Cells and Lead to Accumulation of Adaptive Mutations. *J Virol* **2004**, *78*, 4953–4964.

85. Hyde, J.L.; Gardner, C.L.; Kimura, T.; White, J.P.; Liu, G.; Trobaugh, D.W.; Huang, C.; Tonelli, M.; Paessler, S.; Takeda, K.; others. A viral RNA structural element alters host recognition of nonself RNA. *Science* **2014**, *343*, 783–787.

© 2019 by the authors. Submitted to *Viruses* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).